



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

⑪ Número de publicación: **2 296 176**

⑤① Int. Cl.:
G10L 11/00 (2006.01)

⑫

TRADUCCIÓN DE PATENTE EUROPEA

T3

⑧⑥ Número de solicitud europea: **05745748 .3**

⑧⑥ Fecha de presentación : **09.05.2005**

⑧⑦ Número de publicación de la solicitud: **1745464**

⑧⑦ Fecha de publicación de la solicitud: **24.01.2007**

⑤④ Título: **Dispositivo y procedimiento para analizar una señal de información.**

③⑩ Prioridad: **10.05.2004 DE 10 2004 023 436**

④⑤ Fecha de publicación de la mención BOPI:
16.04.2008

④⑤ Fecha de la publicación del folleto de la patente:
16.04.2008

⑦③ Titular/es: **m2any GmbH
Lichtenbergstrasse 8
85748 Garching, DE**

⑦② Inventor/es: **Herre, Jürgen;
Allamanche, Eric;
Hellmuth, Oliver y
Kastner, Thorsten**

⑦④ Agente: **Arizti Acha, Mónica**

ES 2 296 176 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

DESCRIPCIÓN

Dispositivo y procedimiento para analizar una señal de información.

5 La presente invención se refiere al análisis de señales y especialmente al análisis de señales para identificar un contenido de señal.

10 Para archivar las existencias cada vez más grandes de material de audio y vídeo, crear bases de datos de búsqueda sencilla o distribuir las por diferentes vías de venta, se requieren sistemas de reconocimiento de información automáticos que ayudan a identificar material de audio y vídeo o expresado en general, identificar material de información de manera unívoca, basándose en el contenido.

15 Una aplicación para ello es el denominado "Broadcast-Monitoring" o la supervisión por radio. Con ayuda de un sistema de supervisión de audio-vídeo de este tipo debe garantizarse por ejemplo que sólo se distribuyan contenidos legítimos o que los tantos por ciento respectivos para los titulares de los derechos del material de audio y vídeo se ajusten correctamente.

20 Otra aplicación es por ejemplo el reconocimiento de material de audio que debe intercambiarse entre partes mediante las redes de punto a punto.

25 Otra aplicación es la posibilidad de control de la industria de la publicidad, para controlar un emisor de televisión o de radio en la medida en que los tiempos de publicidad reservados también se hayan emitido realmente, o en que sólo se hayan emitido partes de los porcentajes de publicidad reservados, o en que partes de las emisiones de publicidad se han interferido durante la transmisión, lo que por ejemplo podría ser responsabilidad del canal de televisión o radio. En este punto se hace referencia a que en especial los costes para la publicidad televisiva en programas muy extendidos en horas de emisión favorables son tan elevados, que la industria de la publicidad especialmente con respecto a estos costes elevados tiene un interés vital en una posibilidad de control, para no tener que confiar sólo en las indicaciones de los canales de radio. Hasta el momento, como posibilidad de control se utilizan "oyentes de prueba" o "videntes de prueba" pagados, que observan continuamente un determinado programa de televisión y por ejemplo registran las horas exactas a las que se retransmite un anuncio comercial y que además supervisan, si durante la retransmisión no se ha producido ninguna interferencia, o si se ha retransmitido correctamente la totalidad del anuncio comercial, es decir, si no ha tenido lugar ninguna distorsión de la imagen, etc.

35 Las desventajas de este concepto son evidentes. Así, por un lado, los costes son considerables, y por otro lado, la fiabilidad o la fuerza probatoria de las afirmaciones de los oyentes de prueba o videntes de prueba son problemáticas, especialmente cuando se aumentan considerablemente las demandas de reembolso, que con respecto a su posibilidad de demostración sólo depende de observadores de prueba.

40 Para la supervisión de radio automatizada pueden utilizarse diferentes sistemas conocidos. Así, el documento WO 02/11123 A2 o la publicación técnica "Invited Talk: An Industrial-Strength Audio Search Algorithm", Avery Wang, ISMIR 2003, Baltimore, octubre del 2003, da a conocer sistemas y procedimientos para reconocer señales de tono y de música en un entorno de fuerte ruido y grandes distorsiones. A este respecto se estudia en primer lugar, si existe una coincidencia entre valores hash de un objeto de audio de referencia y el valor hash determinado actualmente del objeto de audio aún sin identificar. Si este es el caso, entonces se almacena el desplazamiento de tiempo correspondiente, esto es, la distancia relativa desde el comienzo del objeto de audio, del valor hash en el objeto de audio aún sin identificar y el desplazamiento de tiempo del valor hash en el objeto de referencia con la identificación correspondiente del objeto de audio de referencia. Una vez procesados todos los valores hash de entrada, entonces comienza una denominada fase de exploración. En ésta se estudia cuántos pares de desplazamiento de tiempo por cada tiempo de objeto de audio de referencia coinciden de manera continua. En caso de establecer un número determinado, entonces se parte de una identificación del objeto de audio de referencia correspondiente. Los pares de desplazamiento de tiempo se consideran continuos en el tiempo, es decir correspondientes entre sí con respecto al tiempo, precisamente cuando en un diagrama de dispersión bidimensional con un desplazamiento de tiempo como coordenada x y el otro como coordenada y forman una recta.

55 Otro dispositivo conocido para la identificación automática de composiciones musicales se da a conocer en la publicación *BATTLE E ET AL*: "Automatic Song Identification in Noisy Broadcast Audio" PROCEEDINGS OF THE FOURTH IASTED INTERNATIONAL CONFERENCE SIGNAL AND IMAGE PROCESSING ACTA PRESS ANAHEIM (EE.UU.), 2002, páginas 230-235, XP002337265.

60 En la publicación técnica "Robust Audio Hashing for Content Identification" de J. Haitzma, T. Kalker, J. Oostveen, en Proceedings of the Content-Based Multimedia Indexing, 2001, url: citeseer.ist.psu.edu/haitzma01robust.html, se presenta un sistema para el *audio-hashing* robusto para una identificación de contenido. Para reconocer música basándose en el contenido se utiliza una función hash, que asocia una secuencia de bits con un segmento de una señal de audio, y concretamente en una forma que para la percepción humana del sonido señales de audio acústicamente parecidas también generan una secuencia de bits parecida. Para calcular un valor hash, en primer lugar se aplica una función ventana a la señal de audio y se le somete a una transformación, para finalmente llevar a cabo una clasificación del resultado de transformación en bandas de frecuencia con un ancho de banda logarítmico. Para estas bandas de frecuencia los signos de las diferencias se determinan en la dirección del tiempo y la frecuencia. La secuencia de

ES 2 296 176 T3

bits que resulta de los signos forma el valor hash. Siempre se calcula un valor hash para una longitud de la señal de audio de 3 segundos. En caso de que para un segmento de este tipo la distancia de Hamming se encuentre entre un valor hash de referencia y un valor hash de prueba que se estudia por debajo de un umbral s , entonces se supone una coincidencia y el segmento de prueba se asigna al elemento de referencia.

5

Para realizar un reconocimiento de material de audio, la señal de audio se divide típicamente en pequeñas unidades de la longitud Δt . Estas unidades individuales se analizan en cada caso individualmente, para tener al menos una determinada resolución temporal.

10

De aquí resultan varios problemas.

Los resultados de reconocimiento de los pequeños segmentos de tiempo analizados de la señal de audio deben componerse de tal manera, que para un segmento de tiempo largo pueda llegarse a una conclusión correcta, unívoca acerca de la señal de audio reconocida.

15

Para analizar un flujo de datos de audio continuo deben reconocerse correctamente las transiciones de un elemento de audio a otro, es decir, un cambio de una composición A musical a una composición B musical.

20

Además existe la situación en la que de una composición musical hay varias versiones, que por ejemplo comienzan igual y sólo después de un cierto tiempo empiezan a distinguirse. A este respecto se piensa por ejemplo en versiones cortas o versiones maxi de una canción. De manera alternativa también existen situaciones en las que composiciones musicales, que se refieren a la misma canción, por ejemplo son diferentes al principio, tienen una parte central idéntica y que hacia el final de al menos una de las dos composiciones musicales vuelven a distinguirse entre sí. Para la concesión de los tantos por ciento a los titulares de los derechos de autor puede ser absolutamente importante si por ejemplo, a cambio de un canon mayor, puede sonar la versión maxi de una canción, si para un canon medio sólo puede sonar una versión normal, o si para un canon reducido ya puede sonar la versión corta de una canción. En este caso debería poder distinguirse de manera fiable entre diferentes versiones de una canción.

25

30

El presente estado de la técnica es insuficiente en la medida en que se producen errores de registro cuando los resultados de los reconocimientos individuales se componen de manera sencilla. En especial no se proporcionan indicaciones con respecto a si y en qué forma puede analizarse un flujo de datos de audio continuo a partir de varios objetos de audio diferentes y cómo pueden detectarse cambios correspondientes entre diferentes objetos de audio. Además, es cierto que especialmente en el estado de la técnica mencionado en último lugar se menciona la ambigüedad de valores hash de referencia. Sin embargo, no se menciona ninguna solución explícita para el problema para determinar un candidato unívoco. Una vez identificado un objeto de audio para un valor hash, entonces para el valor hash inmediatamente siguiente sólo se comprueba si éste va bien con el objeto de audio identificado. En caso contrario, entonces vuelve a buscarse, incluyendo todos los objetos de audio de referencia.

35

40

En el estado de la técnica no se conoce ninguna solución especialmente para distinguir diferentes versiones de una misma canción.

El objetivo de la presente invención se basa en crear un concepto fiable para analizar una señal de información.

45

Este objetivo se resuelve mediante un dispositivo para analizar una señal de información según la reivindicación 1 de patente, un procedimiento para analizar una señal de información según la reivindicación 20 de patente o un programa informático según la reivindicación 21 de patente.

50

La presente invención se basa en el conocimiento de que una identificación de contenido fiable se consigue porque no sólo se consideran resultados de reconocimiento individuales en sí mismos, sino más allá de un cierto periodo de tiempo. Así también existe en la secuencia de resultados de reconocimiento individuales para una secuencia de huellas digitales una información considerable, que puede utilizarse para el reconocimiento. Por tanto, según la invención, a partir de una secuencia de huellas digitales, que representa una secuencia de bloques de una señal de información, se realiza una formación de al menos dos hipótesis diferentes, siendo una primera hipótesis una suposición para la pertenencia de la secuencia de bloques a una primera entidad de información, y siendo la segunda hipótesis una suposición para la pertenencia de la secuencia de bloques a la segunda entidad de información. Ahora se estudian las al menos dos hipótesis y se someten a una valoración en la medida en que se llega a una conclusión acerca de la señal de información basándose en un resultado de estudio. La conclusión podría consistir por ejemplo en determinar que la secuencia de bloques representa una entidad de información que tiene una hipótesis que es la más probable. De manera alternativa o adicional, la hipótesis podría consistir en que una unidad de información esté finalizada con la huella digital, que como la última en el tiempo de la secuencia de huellas digitales contribuye a la hipótesis más probable.

60

65

De manera preferible, las hipótesis se estudian en la medida en que para huellas digitales hay al menos dos resultados de identificación diferentes, así como en que para cada uno de los dos resultados de identificación diferentes hay una medida de fiabilidad, pudiendo consistir esta medida de fiabilidad en un número concreto. Sin embargo, esta medida de fiabilidad puede darse también de manera implícita en la medida en que por el simple hecho de que por ejemplo se proporcionen dos resultados de identificación, se señalice una fiabilidad de por ejemplo 1/2 y que este número no se indique de manera explícita.

ES 2 296 176 T3

Para valorar si una hipótesis es más probable que la otra hipótesis, preferiblemente se agrupan medidas de fiabilidad de los reconocimientos individuales para el número en cuestión de bloques sucesivos en el tiempo, consistiendo esta agrupación preferiblemente en una sumación. A continuación, la hipótesis, que proporciona la medida de fiabilidad agrupada más alta se valora como la más probable.

5 En un ejemplo de realización preferido de la presente invención, como medio para proporcionar resultados de identificación sucesivos se utiliza una base de datos de huellas digitales, en la que están depositadas un número de huellas digitales de referencia en cada caso asignadas a un resultado de identificación. A continuación con la huella digital generada a partir de un bloque de la señal de información que va a analizarse se lleva a cabo una búsqueda en la base de datos, para buscar una huella digital de referencia en la base de datos, que proporciona una coincidencia con la huella digital de prueba. Según la configuración de la base de datos como resultado de búsqueda sólo se emite el mejor acierto, es decir el acierto con una distancia mínima de la base de datos como resultado de identificación. También se prefieren bases de datos que no sólo proporcionan resultados de acierto cualitativamente, sino que también proporcionan un resultado de acierto cuantitativo en la medida en que se emite un número de posibles aciertos con una medida de fiabilidad asignada, de manera que por ejemplo todos los aciertos con una medida de fiabilidad, que es superior o igual a un umbral determinado, como por ejemplo del 20%, se emiten por la base de datos. En el ejemplo de realización preferido de la presente invención se comienza una nueva hipótesis cuando aparece un nuevo resultado de identificación para el que todavía no existe ninguna hipótesis. Este procedimiento se realiza para un determinado número de bloques, para a continuación estudiar con respecto al pasado, si una hipótesis determinada que ha resultado fiable ya está finalizada, para a continuación reconocer esta hipótesis como la hipótesis más probable.

Una ventaja de la presente invención consiste en que el concepto trabaja de una manera fiable y aún así es tolerante con respecto a los errores especialmente con respecto a errores de transmisión. Así, no se intenta obtener una decisión mediante un único bloque, sino que en cierta medida se observa y evalúa en conjunto una secuencia de bloques sucesivos mediante la formación de hipótesis, de manera que interferencias de transmisión no a corto plazo o una aparición en general de ruido hacen que todo el proceso de reconocimiento sea inválido.

Además, el concepto según la invención proporciona automáticamente un registro de la calidad de transmisión desde el principio hasta el final por ejemplo de un anuncio comercial. Incluso cuando una hipótesis se ha reconocido como la hipótesis más probable, es decir, cuando se determina que existía un determinado anuncio comercial, entonces aún así pueden entenderse variaciones de la calidad en el anuncio comercial mediante las medidas de fiabilidad. Además, con ello puede reproducirse y registrarse especialmente la continuidad temporal completa de un anuncio comercial como ejemplo para una entidad de información, y en concreto especialmente en la medida en que no se ha repetido continuamente cualquier cosa del anuncio comercial, sino que todo el anuncio comercial se ha emitido de manera continua desde el principio del anuncio comercial hasta el final de anuncio comercial.

La presente invención es además ventajosa en la medida en que mediante la formación de hipótesis se reconoce automáticamente el final de una entidad de información y el comienzo de una entidad de información. Esto se debe a que una asignación a una entidad de información será en general unívoca. Esto significa, que durante un instante determinado no pueden reproducirse varias entidades de información en conjunto, sino que al menos para el número excesivo de contenidos de programa sólo se obtiene una entidad de información en la señal de información en un instante. El estudio de hipótesis y la valoración de las hipótesis debido al estudio de hipótesis proporcionan automáticamente un instante en el que ha finalizado una entidad de información previa y en el que comienza una nueva entidad de información. Esto se debe a la asignación de bloques conservada hasta en las hipótesis. Así, como anteriormente, una secuencia de huellas digitales corresponde a una secuencia de bloques, y a su vez una secuencia de resultados de identificación corresponde a una secuencia de huellas digitales, de manera que una hipótesis está asignada a la señal de información original con respecto al tiempo de manera unívoca.

El concepto según la invención es además ventajoso en la medida en que no se producen situaciones “empate” entre dos hipótesis, incluso cuando entidades de información tienen aún así por tramos el material de audio idéntico, como por ejemplo en el caso de versiones cortas o versiones largas de una misma canción.

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

55 la figura 1, un diagrama de bloques de un dispositivo según la invención;

la figura 2, un diagrama de bloques de una base de datos que puede utilizarse para el ejemplo de realización mostrado en la figura 1;

60 la figura 3, una representación esquemática de un resultado de emisión para una secuencia de huellas digitales para una secuencia de intervalos de tiempo así como las hipótesis asignadas;

las figuras 4a a 4c, un escenario de ejemplo para ejemplos de aplicación posteriores;

65 las figuras 5a a 5d, una representación esquemática de diferentes evaluaciones de errores;

la figura 6, un diagrama de bloques de un ejemplo de realización preferido de la presente invención;

ES 2 296 176 T3

las figuras 7a a 7c, una representación de la funcionalidad del concepto según la invención para el escenario de partida representado en las figuras 4a a 4c;

5 la figura 8, una representación esquemática de una señal de información con unidades de información, bloques de unidades de información y entidades de información con una pluralidad de bloques;

la figura 9, un escenario conocido para crear una base de datos de huellas digitales; y

10 la figura 10, un escenario conocido para la identificación de audio por medio de una base de datos de huellas digitales cargada según la figura 9.

La figura 1 muestra un diagrama de bloques de un dispositivo para analizar una señal de información según un ejemplo de realización preferido de la presente invención. Una señal de información a modo de ejemplo se representa con 800 en la figura 8. La señal 800 de información está compuesta por una secuencia 802 de bloques sucesivos en el tiempo de unidades de información, pudiendo ser las unidades 804 de información individuales por ejemplo muestras de audio, píxeles de vídeo o coeficientes de transformación de vídeo, etc. Una pluralidad de bloques de la secuencia 802 siempre forma conjuntamente una entidad 806 de información. En el ejemplo de realización mostrado en la figura 8 los primeros seis bloques forman la primera entidad de información y los bloques 7, 8, 9, 10 forman la segunda entidad de información. A partir de los bloques 11 a n, en la figura 8 está representada por ejemplo una tercera entidad de información. Una entidad de información podría ser por ejemplo una composición musical, un fragmento hablado, una imagen de vídeo o también por ejemplo una parte de una imagen de vídeo. Una entidad de información también podría ser sin embargo un texto o por ejemplo una página de un texto cuando la señal de información también comprende datos de texto.

25 El dispositivo mostrado en la figura 1 está configurado para trabajar utilizando una secuencia de huellas FA1, FA2, FA3,..., FAi digitales, que se generan a partir de la secuencia de bloques 802, o que por ejemplo se recuperan de una memoria, cuando las huellas digitales ya se han generado antes del análisis o a lo mejor incluso se proporcionan junto con la señal de información, según la forma de realización. Debe hacerse referencia a que en la formación de bloques también pueden utilizarse técnicas de solapamientos de bloques, como por ejemplo se conocen por la codificación de audio.

En cualquier caso, el dispositivo trabaja para analizar la señal de información utilizando una secuencia de huellas digitales para la secuencia de bloques, de manera que la secuencia de bloques 802 está representada por la secuencia de huellas FA1, FA2, FA3, FA4,..., FAi digitales. La secuencia de huellas digitales se alimenta en una entrada de huellas digitales a un medio 12 para proporcionar resultados de identificación para huellas digitales sucesivas. El medio 12 para proporcionar resultados de identificación sucesivos es eficaz para proporcionar resultados de identificación sucesivos para huellas digitales sucesivas, representando un resultado de identificación una pertenencia de un bloque de unidades de información a una entidad de información predeterminada. Cuando por ejemplo se parte de que una canción tiene una duración en el tiempo, que aproximadamente corresponde a seis bloques, entonces si bien los seis bloques proporcionan diferentes huellas digitales, sin embargo en el medio 12 para proporcionar se señala que todos estos seis bloques pertenecen a la entidad de información predeterminada, es decir a la canción mencionada.

Según la forma de realización, el medio 12 para proporcionar una huella digital proporcionará uno o varios resultados de identificación. Éste o estos resultados de identificación se suministran a un medio 14 para formar al menos dos hipótesis a partir de los resultados de identificación para las huellas digitales sucesivas. En especial, una primera hipótesis representa una suposición para la pertenencia de la secuencia de bloques a una primera entidad de información, y la segunda hipótesis es una suposición para la pertenencia de la secuencia de bloques a la segunda entidad de información. Las diferentes hipótesis H1, H2,... se suministran a un medio 16 para estudiar las hipótesis, estando configurado el medio 16 para trabajar según un algoritmo de estudio que puede ajustarse, para finalmente proporcionar un resultado de estudio en una salida 18 de resultados de estudio.

Este resultado de estudio en la línea 18 se suministra a continuación a un medio 20 para llegar a una conclusión acerca de la señal de información. El medio 20 para llegar a una conclusión acerca de la señal de información está configurado para emitir una información acerca de la señal de información basándose en el resultado de estudio y puede tener una pluralidad de ajustes.

60 Todos los ajustes tienen en común, que se llega a la conclusión acerca de la señal de información basándose en el resultado 18 de estudio. Los ejemplos de diferentes conclusiones acerca de la señal de información se basan en determinar que la secuencia de bloques representa una entidad de información, que tiene una hipótesis, que es la más probable. Las conclusiones alternativas consisten en que una entidad de información está finalizada con la huella digital que como la última huella digital en el tiempo contribuye a la hipótesis más probable. Una conclusión alternativa, a la que puede llegarse a través del medio 20 se basa en que se establece que en la señal de información existe una entidad de información o no.

65 El procesamiento posterior según la invención, que se proporciona especialmente a través de los medios 14, 16 y 20, es decir, la formación de al menos dos hipótesis, el estudio de las hipótesis y el llegar a una conclusión basándose en un resultado de estudio posibilita de este modo no sólo la identificación de una composición en una señal de

información en sí desconocida, es decir, que ha de analizarse, sino que posibilita también, independientemente de la identificación de una composición en sí, el reconocimiento del final de una primera composición, es decir de una primera entidad de información, y la detección del comienzo de una segunda entidad de información que sigue a la primera entidad de información.

5 Sin embargo, con respecto a la supervisión de la publicidad, el concepto de procesamiento posterior según la invención proporciona también la posibilidad de registrar si una composición determinada existía o no en la señal de información. A este respecto, las huellas digitales obtenidas a partir de la señal de información sólo se compararían con un conjunto de huellas digitales, concretamente el conjunto de huellas digitales, que representan la entidad de información predeterminada, es decir un anuncio comercial determinado. Así, esta conclusión no debe considerarse en principio en la medida en que se identifica una entidad de información, o que se registran el final de una entidad de información y el comienzo de una entidad de información siguiente, sino que se basa en registrar si existe o no una determinada entidad de información en una señal de información en sí desconocida que ha de analizarse.

15 La figura 2 muestra una forma de realización preferida especial del medio 12 para proporcionar resultados de identificación para huellas digitales sucesivas. En un ejemplo de realización preferido el medio 12 comprende, tal como muestra la figura 2, una base de datos, que comprende diferentes huellas FARj digitales de referencia, que están almacenadas con asignación a un resultado de identificación, es decir IDk. En el ejemplo de realización preferido, las huellas FAi digitales se procesan sucesivamente, es decir de manera secuencial en el tiempo. Así, una huella FAi digital se almacena en la base de datos a través de una línea 24 de entrada. En la base de datos se compara a continuación la huella FAi digital almacenada con todas las huellas FARj digitales de referencia. En el ejemplo de realización preferido, la base de datos no es una base de datos cualitativa, que establece que una huella digital de entrada coincide o no con una huella digital de referencia almacenada, sino que la base de datos es una base de datos cuantitativa que puede proporcionar una medida de distancia o de fiabilidad para los resultados emitidos. Así, la base 22 de datos en el ejemplo de realización preferido mostrado en la figura 2 proporcionaría en su salida 26 por ejemplo el resultado que se representa en una tabla 28 de resultados. Así, la base de datos indicaría por ejemplo que la huella FAi digital indica a un resultado IDx de identificación, esto es a una composición musical por ejemplo x con una fiabilidad ZV₁ del 60%. Simultáneamente, sin embargo, la base de datos también indicará que la huella FAi digital indica a una composición con el resultado IDy de identificación con una fiabilidad del 50%. Finalmente, la base de datos también podría dar como resultado que la huella FAi digital indica con una medida ZV₃ de fiabilidad de por ejemplo el 40% a otra composición con la identificación IDz.

Según la forma de realización, la totalidad de la tabla 28 de resultados puede suministrarse al medio 14 para formar al menos dos hipótesis de la figura 1. Sin embargo, alternativamente, la propia base 22 de datos podría ya tomar una decisión y siempre sólo el valor más probable, esto es, en el presente caso, suministrar el resultado IDx al medio 14 para formar al menos dos hipótesis. En este caso no necesariamente debería suministrarse la medida ZV₁ de fiabilidad también al medio 14 para formar al menos dos hipótesis. Más bien podría prescindirse de la transmisión adicional de las medidas ZV₁ de fiabilidad. De manera alternativa el medio 12 para proporcionar los resultados de identificación, que simultáneamente también proporciona las medidas de fiabilidad, también podría estar configurado para suministrar las medidas ZV₁ de fiabilidad en el orden correspondiente con asignación a los bloques no al medio 14 para formar al menos dos hipótesis, sino al medio 16 para estudiar las hipótesis, porque este medio 16 sólo requiere las medidas de fiabilidad, para por ejemplo encontrar la hipótesis más probable.

Por la base 22 de datos de la figura 2 puede observarse, que un resultado de identificación, como por ejemplo ID1 puede tener varias huellas FAR11, FAR12, FAR13 digitales asignadas, lo que indica que la composición, que se identifica con ID1 tiene varios bloques. Sin embargo, según la implementación una única huella digital larga también puede estar almacenada para la composición con la identificación ID1, que sin embargo se compone de las huellas FAR11, FAR12, FAR13,... digitales individuales. A continuación, la base de datos correlacionaría la huella FAi digital alimentada, que depende de la longitud de bloque y normalmente es mucho más corta que la huella digital larga, con la huella digital larga en cada fila de la base de datos para establecer, si un fragmento de la huella digital de referencia almacenada larga coincide o no con la huella FAi digital de referencia alimentada en la línea 24. La medida de fiabilidad se obtendría en este caso automáticamente en cierta medida, en concreto simplemente mediante una evaluación cuantitativa del resultado de correlación.

55 Además, con respecto a la figura 2, ya se hace referencia a las dos últimas filas, que están designadas con los resultados ID108 e ID109 de identificación. ID108 designa una versión larga de la composición musical, tal como se explicará haciendo referencia a la figura 4a, mientras que ID109 identifica una versión corta de la misma composición musical, tal como se representa en la figura 4b.

60 Tal como ya se ha explicado, la base 22 de datos, es decir, esta implementación del medio 12 para proporcionar resultados de identificación para huellas digitales sucesivas puede estar configurada de tal manera, que siempre proporcione sólo el resultado de identificación más probable. De manera alternativa, la base 22 de datos también podría estar configurada sin embargo para por ejemplo proporcionar siempre sólo los resultados de identificación, cuya probabilidad es superior a un umbral mínimo, como por ejemplo un umbral del 5%. Esto daría lugar a una variación del número de las filas de la tabla de huella digital a huella digital. De nuevo alternativamente, la base 22 de datos también podría estar implementada sin embargo para suministrar para cada huella FAi digital de entrada un número determinado de candidatos más probables, como por ejemplo los "Diez primeros", como los diez candidatos más probables del medio 14 para formar al menos dos hipótesis.

ES 2 296 176 T3

A continuación, mediante la figura 3, se representa una forma de realización de la base 22 de datos, en la que la base de datos siempre suministra los tres resultados de identificación más probables junto con valores de fiabilidad pertenecientes al medio 14 para formar hipótesis, es decir en cierta medida comprende una implementación de los “Tres primeros”. Así, por la figura 3 puede observarse, que para la huella FA1 digital se proporcionan resultados ID1, ID2, ID3 de identificación, y concretamente con las medidas de fiabilidad respectivas del 40%, 60% o 30%. Para el intervalo de tiempo Δt_2 , es decir para la huella FA2 digital volverán a proporcionarse los resultados ID1, ID2, ID3 de identificación, ahora sin embargo con otra probabilidad correspondiente, es decir con otra medida de fiabilidad correspondiente, que sólo se representa a modo de ejemplo en la figura 3 como porcentaje. Este procedimiento se realiza para todas las huellas FA1 a FA8 digitales de entrada. Al medio 14 para formar al menos dos hipótesis, tal como se representa en la figura 1, se proporcionan estos resultados de identificación. El medio 14 para formar al menos dos hipótesis está configurado para comenzar una nueva hipótesis siempre que se proporcione un nuevo resultado de identificación por el medio 12 para proporcionar los resultados de identificación. Esto se hace evidente a partir de la figura 3, porque en el instante Δt_1 se comienzan las hipótesis H1, H2, H3 con ID1, ID2 o ID3, y porque en el intervalo Δt_7 de tiempo vuelven a comenzarse nuevas hipótesis con ID108, ID109, ID4, y porque en el intervalo Δt_8 de tiempo debido al hecho de que en ese caso en el ejemplo mostrado ID8 aparece por primera vez, se comienza otra hipótesis H4 para ID8.

El medio 14 para formar al menos dos hipótesis es por tanto eficaz para ver para cada huella digital nueva si se produce un nuevo resultado de identificación, para comenzar una nueva hipótesis y para continuar una hipótesis ya comenzada previamente en la medida en que cuando para un periodo Δt_i de tiempo para la hipótesis ya comenzada previamente en los “Tres primeros” o “X primeros” está contenido un elemento, que, aunque con una menor probabilidad aún así proporciona un resultado de identificación para una hipótesis comenzada en este momento. Este procedimiento continúa durante un cierto tiempo. Entonces, por ejemplo en instantes predeterminados o, activado por un usuario, etc. el medio 16 para estudiar las hipótesis estudiará las hipótesis formadas para el pasado y para el caso mostrado en la figura 3 sumará por ejemplo las medidas de fiabilidad de las hipótesis H1, H2, H3 para los periodos Δt_1 a Δt_6 de tiempo. El medio 16 para estudiar al menos dos hipótesis establecería entonces, que la composición es lo más probable ID1, que por tanto la hipótesis H1 para el periodo Δt_1 a Δt_6 de tiempo es la hipótesis más probable, dado que la medida de fiabilidad alcanza un valor de 420, mientras que para la segunda hipótesis sólo se alcanza una medida de fiabilidad de 230, y mientras que para la tercera hipótesis sólo se alcanza una medida de fiabilidad de 135.

En el caso mostrado en la figura 3 las tres hipótesis se inician simultáneamente y las tres hipótesis finalizan simultáneamente. Sin embargo, esto no debe ser así obligatoriamente. Así, la hipótesis H1 podría finalizar por ejemplo antes, esto es, por ejemplo en el instante Δt_5 . La medida de fiabilidad de ID1 debería reducirse en este caso en 90, con lo que se llegaría a un valor de 330. En este caso se obtendría, que la hipótesis H1 es aún así la más probable, aunque la hipótesis H2 existe durante un periodo de tiempo más largo, aunque en total con una menor probabilidad. Por el ejemplo mostrado en la figura 3 puede verse además, que la hipótesis H1, a pesar del hecho de que para Δt_1 era menos probable que la hipótesis H2, al final “está por delante”.

Por la figura 3 puede verse además, que una hipótesis también podría tener “huecos” en la medida en que por ejemplo en el intervalo Δt_4 de tiempo por cualquier motivos, por ejemplo por la interferencia en un canal de transmisión, etc. sólo se proporcionan ID2 e ID3 con una probabilidad razonable, no así ID1. Entonces el valor de fiabilidad para ID1 debería reducirse en 60, lo que a su vez daría como resultado que la fiabilidad total sería 360 en vez de 420, de manera que también en este caso la hipótesis H1 es la hipótesis más probable.

Por los escenarios anteriormente descritos puede verse por tanto, que el concepto según la invención que trabaja basándose en un procesamiento posterior con hipótesis y por un lado considera la secuencia y por otro lado las medidas de fiabilidad de las operaciones de identificación de huellas digitales individuales, es extraordinariamente robusto frente a errores de transmisión y también frente a funcionalidades problemáticas en la base de datos o también frente a huellas digitales que para algunas entidades de información, tales como por ejemplo composiciones musicales, imágenes de vídeo, textos, etc. a lo mejor no se diferencian entre sí con tanta intensidad como se desearía.

En un ejemplo de realización preferido una hipótesis es un protocolo almacenado (figura 3: H1, H2, H3,...), preferiblemente en forma de una lista almacenada, que por un lado presenta una indicación a la entidad de información, para la que se crea la hipótesis, y por otro lado presenta una indicación a huellas digitales o bloques de unidades de información, para los que se plantea la hipótesis. Preferiblemente el protocolo contiene además para un bloque o huella digital una medida de fiabilidad.

Por la figura 3 puede reconocerse además que la primera entidad de información sólo se extiende por el periodo Δt_1 a Δt_6 de tiempo y que a partir de Δt_7 comienza una nueva entidad. Esto también puede observarse especialmente porque las tres hipótesis finalizan simultáneamente o, porque incluso cuando la hipótesis H3 se hubiera extendido además por ejemplo por Δt_7 , ahora aparecen valores de identificación completamente diferentes con una probabilidad muy alta, concretamente ID108 e ID109 con probabilidades de 90 y 85 y de este modo “sustituyen” a los “claros ganadores” del periodo de tiempo anterior.

Al final de la figura 3 se representan las diferentes conclusiones, a las que puede llegarse a modo de ejemplo, concretamente que la entidad de información en el periodo Δt_1 a Δt_6 de tiempo es la composición musical identificada por ID1. De manera alternativa, la conclusión también podría indicar que un cambio de entidad de información tiene

lugar entre Δt_6 y Δt_7 . De manera alternativa, sin embargo, una conclusión también podría indicar que la composición musical, identificada por ID1, está contenida en la señal de información.

A continuación, con referencia a las figuras 9 y 10 en primer lugar se hará referencia más en general a sistemas de bases de datos, tal como pueden utilizarse en relación a la presente invención de manera ventajosa. La presente invención se basa por tanto en un sistema para la identificación de material de audio, como por ejemplo música. El sistema conoce dos fases de operación. En la fase de entrenamiento, que se representa mediante la figura 9, el sistema de reconocimiento aprende las composiciones que han de identificarse posteriormente. En la fase de identificación, que se representa en la figura 10, pueden volver a reconocerse las composiciones de audio entrenadas anteriormente.

Para identificar una composición musical, o también cualquier otra señal de tono, se extrae de la misma un conjunto de datos compacto y único, que también se denomina huella digital (*fingerprint*) o signatura. Esta extracción tiene lugar en una extracción 900 de características de bloque. En la fase de entrenamiento o de aprendizaje se crean tales huellas digitales a partir de un conjunto de objetos de audio conocidos y se depositan en una base 902 de datos de huellas digitales. Preferiblemente, el medio 900 de extracción de características está configurado para utilizar como característica la característica SFM, significando SFM "medida de planeidad espectral" (*Spectral Flatness Measure*). Naturalmente, también pueden utilizarse otros sistemas de generación de huellas digitales o resultados de extracción de características. Sin embargo se ha demostrado, que las características referidas a la tonalidad y especialmente la característica SFM tienen por un lado un poder de diferenciación especialmente bueno y por otro lado, una compacidad especialmente buena. Para este fin, en primer lugar cada bloque se somete a una conversión de tiempo-frecuencia para a continuación calcular con los valores generados a partir de la conversión de tiempo-frecuencia un SFM para un bloque según la ecuación siguiente.

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

En esta ecuación, $X(n)$ representa el cuadrado de un valor absoluto de una componente espectral con el índice n , representando N el número total de los coeficientes espectrales de un espectro. Por la ecuación puede verse, que la medida SFM es igual al cociente a partir del valor medio geométrico de las componentes espectrales y el valor medio aritmético de las componentes espectrales. Se conoce que el valor medio geométrico es siempre más pequeño o como máximo igual al valor medio aritmético, de manera que el SFM tiene un intervalo de valores entre 0 y 1. En este contexto, un valor cerca de 0 indica una señal tonal y un valor cerca de 1 indica una señal muy de tipo ruido con una curva espectral plana. Se indica que el valor medio aritmético y el valor medio geométrico sólo son iguales, cuando todos los $X(n)$ son idénticos, lo que corresponde a una señal completamente atonal, es decir una señal de tipo ruido o de tipo pulso. Sin embargo, cuando en un caso extremo, sólo una componente espectral tiene un valor muy alto, mientras que otras componentes $X(n)$ espectrales tienen valores muy pequeños, la medida SFM tendrá un valor cerca de 0, lo que indica una señal muy tonal.

El concepto SFM así como otros conceptos de extracción de características, para generar huellas digitales, se representan por ejemplo en el documento WO 03/007185.

En la fase de identificación, que se representa en la figura 10, también tiene lugar típicamente la misma extracción 900 de características que en la fase de entrenamiento. Especialmente la huella digital, que se extrajo del objeto de audio en la entrada de audio para un periodo Δt de tiempo, se compara con las huellas digitales de referencia de la base 902 de datos de huellas digitales por medio de un comparador 904, estando contenido típicamente el comparador en el medio 12 para proporcionar resultados de identificación, tal como se ha representado mediante la figura 1. A continuación, al establecer una coincidencia mediante un criterio determinado se obtiene un resultado de reconocimiento para el periodo Δt de tiempo. Por tanto, si se establece una coincidencia mediante un criterio determinado, entonces, la huella digital desconocida y así el fragmento del objeto de audio desconocido, pueden asignarse a un material de referencia en la base de datos, esto es, a una lista de resultados IDi, IDi+1, ... de identificación con diferentes valores de fiabilidad.

Según la invención ahora no sólo se asigna un objeto de audio desconocido en la entrada exactamente a un objeto de audio de referencia en la base de datos de referencia, y concretamente sólo para un instante Δt , sino que se trabaja de manera continua, sin interrupción del flujo de datos en la entrada. Según la invención se realiza una asignación de diferentes fragmentos de objetos de audio en cada caso a los objetos de audio correctos de la base de datos de referencia. De este modo se obtiene una sucesión sin huecos, es decir, un protocolo, de los objetos de audio identificados en la entrada.

A continuación, mediante las figuras 4a a 5d se representa una dificultad especial del análisis continuo de un flujo de datos de audio continuo. Concretamente debe dividirse el objeto de audio en segmentos de longitud Δt_x , esto es

ES 2 296 176 T3

en bloques individuales, para poder realizar una asignación a un elemento de referencia en la base de datos para el segmento parcial del flujo de datos de audio. Ahora puede ocurrir que esta asignación de un fragmento individual del flujo de datos de audio no siempre sea unívoca y sólo se haga unívoca en relación con asignaciones anteriores y posteriores. Cuando se realizan asignaciones individuales y sólo se agrupan en una etapa posterior, se obtienen protocolos de reconocimiento con errores, lo que se representa a continuación.

La figura 4a representa una versión larga de una composición XY musical que también se representa mediante una huella digital larga representada en la figura 4a, estando asignado a esta huella digital el resultado ID108 de identificación. La figura 4b muestra lo mismo para una versión corta de la misma composición XY musical. ID109 remite por tanto a una versión corta de la composición XY musical, mientras que ID108 remite a una versión larga de esta composición musical. Después de que la versión corta sea más corta que la versión larga, la huella digital en la figura 4b también es más corta que la huella digital en la figura 4a. De la manera en la que ambos bloques están representados unos respecto a otros, las composiciones musicales y con ello también las huellas digitales ID108 e ID109 contienen material de audio idéntico o datos de huella digital idénticos. ID109 es por tanto una cantidad parcial de ID108. De este modo, por la figura 4c puede observarse, que la versión larga tiene un segmento inicial en el periodo Δt_0 de tiempo, que no existe en la versión corta. En el segmento medio entre t_1 y t_5 la versión larga y la versión corta son idénticas, mientras que la versión larga vuelve a tener un fragmento musical entre el instante t_5 y t_7 , que no existe en la versión corta, identificada por ID109.

A continuación, mediante las figuras 5a a 5d se representa cómo con un agrupamiento sencillo, es decir sin formación de hipótesis, con las identificaciones individuales pueden surgir protocolos de reconocimiento con errores. Se supone, que en la entrada del sistema en el instante t_0 se graba la composición ID108 musical. Además, la base de datos es eficaz en la medida en que identifica los elementos mostrados en la figura 5a para los periodos Δt_x de tiempo. Se indica que la identificación en la figura 5a es en principio correcta, aunque sin embargo en los periodos Δt_1 a Δt_4 de tiempo podría emitirse tanto ID108 como ID109. En última instancia, la determinación de los resultados de identificación en estos intervalos es ambigua, porque la base de datos en caso de no existir interferencia emitirá tanto ID109 como ID108 y debido a diferencias de cálculo por ejemplo siempre se decantará por el valor más probable, de manera que siempre, debido a algún ruido uno de los dos resultados ID108 o ID109 de identificación tendrá una medida de fiabilidad ligeramente superior. En el protocolo de reconocimiento, que se representa en la figura 5b, se realiza de este modo una identificación incorrecta en la medida en que en ningún instante se ha reproducido la composición identificada por ID109, sino que sólo se ha reproducido la composición identificada por ID108.

A continuación, mediante las figuras 5c y 5d, se representa otra alternativa. Se parte del hecho de que la base de datos emite la situación mostrada en la figura 5c. En el protocolo de reconocimiento vuelve a darse un agrupamiento incorrecto, concretamente en la medida en que ID109 existía entre T_1 y T_5 , mientras que esto naturalmente no se cumple. En su lugar se reprodujo la versión larga de la composición musical, esto es ID108 de t_0 a t_7 .

Además son concebibles otros protocolos de reconocimiento erróneo, que surgirán por la ambigüedad de los reconocimientos individuales para un segmento del flujo de datos de audio en el periodo Δt_x de tiempo.

Según la invención ahora se recurre al concepto general representado en la figura 6, en el que los resultados de reconocimiento obtenidos para un periodo Δt_x de tiempo, esto es las señales de salida del medio 12 de la figura 1, que según la implementación puede juntar los medios 900, 904, 902, se someten a un procesamiento posterior, que fundamentalmente corresponde al medio para formar al menos dos hipótesis y al medio para estudiar las hipótesis de la figura 1. A continuación, utilizando el procesamiento posterior, esto es, utilizando los resultados de estudio obtenidos en el procesamiento posterior, se llega a una conclusión acerca de la señal de información en forma de una sucesión de reconocimiento o de un protocolo de reconocimiento.

En la etapa de procesamiento posterior, se supone que la probabilidad para la transición de un objeto de audio de referencia identificado para el periodo Δt_x de tiempo a otro objeto de audio de referencia cualquiera para el periodo Δt_{x+1} de tiempo es igual. A partir de ello se forman diferentes hipótesis consideradas en primer lugar en paralelo para segmentos de audio relacionados a partir de los reconocimientos individuales. Ha de tenerse en cuenta, que los reconocimientos individuales se aúnan para formar una hipótesis cuando se refieren a una misma señal de audio de referencia y están relacionadas de manera continua en el tiempo. El protocolo de reconocimiento se obtiene aunando las hipótesis más probables en cada caso considerando el progreso en el tiempo. A continuación se representa con detalle un algoritmo preferido.

En primer lugar se forman diferentes hipótesis para segmentos de audio relacionados a partir de los reconocimientos individuales para los periodos Δt_x de tiempo (con $x = N, N+1, N+2, \dots$; representando t_N el instante de inicio para la hipótesis correspondiente) para en cada caso un objeto de audio de referencia reconocido.

A continuación, los reconocimientos individuales se agrupan para dar una hipótesis, cuando los reconocimientos individuales son sucesivos de manera continua en el tiempo.

La continuidad en el tiempo es otro elemento que sirve para establecer si se continúa una hipótesis ya existente, o si se comienza una nueva hipótesis. Así se considera el escenario, en el que por ejemplo un determinado solo de guitarra aparece en una composición en la versión corta de la composición muy al principio de la composición y en una versión larga de la composición más bien en el medio de la composición.

ES 2 296 176 T3

En un ejemplo de realización preferido la base de datos, esto es, el medio para proporcionar resultados de identificación no sólo emite una identificación de huella digital sino también un valor temporal, que se obtiene porque la huella digital de identificación en la base de datos tiene una longitud, y la huella digital (corta) introducida sólo coincide con una parte de la huella digital (larga) en la base de datos.

En el escenario anteriormente descrito, la base de datos para el solo de guitarra proporcionaría a lo mejor dos resultados ID (versión corta y versión larga), aunque con dos índices de tiempo diferentes. El índice de tiempo para el resultado ID para la versión corta es a este respecto menor que el índice de tiempo para la versión larga. Basándose en el índice de tiempo el medio para formar las hipótesis puede ahora continuar hipótesis (cuando una continuidad en el tiempo está entre el índice de tiempo y el último índice de tiempo en la hipótesis), o comenzar hipótesis nuevas, cuando no existe ninguna continuidad en el índice de tiempo obtenido actualmente y un último índice de tiempo de una hipótesis.

Cada discontinuidad en el tiempo con respecto a un objeto de audio de referencia genera una nueva hipótesis, cuando el siguiente elemento es con respecto al tiempo una distancia mayor que una distancia T_a en el tiempo que ha de establecerse, o cuando el siguiente elemento se encuentra en el tiempo antes del anterior.

Para el estudio de las hipótesis se realiza para cada hipótesis una suma de las medidas de confianza, esto es, de los valores de fiabilidad o de las medidas para la plausibilidad de los reconocimientos individuales.

Comenzando con el periodo Δt_0 de tiempo, a continuación se valora la hipótesis con la mayor medida de confianza como cierta y se asume en el protocolo de reconocimiento. Para el siguiente periodo de tiempo, que sigue a la primera hipótesis, vuelve a valorarse la hipótesis con la mayor medida de confianza como cierta y se asume en el protocolo de reconocimiento, y así sucesivamente.

Para el ejemplo anteriormente indicado se obtiene así una sucesión, que se representa mediante las figuras 7a a 7c. Para el periodo Δt_0 de tiempo, la base de datos proporciona, tal como se indica por ejemplo en la figura 2, sólo un resultado de identificación, concretamente ID108, que tiene una probabilidad o una medida de fiabilidad, que está por encima de un umbral. En el intervalo Δt_1 de tiempo, esto es, para el bloque de unidades de información, que se extienden por el intervalo Δt_1 de tiempo, la base de datos proporciona dos resultados, que tienen una medida de fiabilidad, que está por encima de un umbral. Los dos resultados también se obtienen para los bloques entre los instantes t_2 a t_5 . Entonces, para el periodo t_5 a t_7 de tiempo la base de datos proporciona a su vez sólo un único resultado de identificación, cuya medida de fiabilidad está por encima de un umbral.

El medio 14 (figura 1) para formar al menos dos hipótesis está configurado para iniciar una primera hipótesis en el instante t_0 debido al resultado ID108 de identificación, y para iniciar una nueva hipótesis, concretamente la hipótesis H2, en el instante t_1 debido al nuevo resultado ID109 de identificación añadido.

En algún momento tras el instante t_7 se considera entonces la situación de hipótesis mostrada en la figura 7a con las hipótesis H1 y H2, para luego, debido al estudio de las hipótesis, que puede desarrollarse tal como se representa en la figura 7b, calcular para cada hipótesis las funciones para las medidas de confianza de los reconocimientos individuales, es decir, x_{H1} y x_{H2} .

Si se parte del hecho de que entre t_1 y t_5 aparecen los resultados ID108 e ID109 de identificación con la misma probabilidad, así, en el ejemplo de realización mostrado en la figura 7a sólo la primera hipótesis H1 ganará, porque si bien la hipótesis entre t_1 y t_5 era igual de probable que la hipótesis H2, porque la hipótesis H1 vale sin embargo en el periodo Δt_0 de tiempo y en el periodo Δt_5 de tiempo y en el periodo Δt_6 de tiempo, esto es, contribuye a una medida de fiabilidad para un reconocimiento individual, que no se da para la hipótesis H2. Por tanto, esto significa para el protocolo de reconocimiento el caso correcto mostrado en la figura 7c, que concretamente la composición designada con ID108 se ha reproducido desde el instante t_0 hasta el instante t_7 .

Comenzando en t_0 se selecciona por tanto la hipótesis H1, puesto que hasta t_7 no hay ninguna hipótesis con una medida de confianza mayor. La hipótesis H2 se rechaza pudiéndose rechazar en principio todas las hipótesis que existen en paralelo a otra hipótesis, que se ha seleccionado como la más probable.

Por tanto, según la invención, se registra exactamente aquella secuencia, en este caso en el ejemplo un elemento, concretamente ID108, que realmente se reprodujo en la entrada de audio.

Se indica que para determinar el final de una hipótesis existen diferentes posibilidades. Así, independientemente de la situación de hipótesis, puede determinarse un final de entidad de información por ejemplo a partir de la propia señal de audio, cuando por ejemplo aparece una pausa con una longitud mínima determinada. Sin embargo, después de que este criterio no sea eficaz cuando entre dos entidades de información se produce una atenuación ("fading") o cuando dos composiciones se suceden tan rápidamente que no se encuentra ninguna pausa perceptible, se prefiere determinar un final de entidad de información debido a las hipótesis consideradas en el pasado. Esto puede tener lugar por ejemplo en la medida en que una hipótesis se considera como finalizada cuando por ejemplo dos o más bloques ya no suministran ningún resultado de identificación más con un valor de fiabilidad por encima de un umbral mínimo determinado al medio 14 para formar hipótesis. De manera alternativa, por ejemplo para el caso mostrado en la figura 3, también puede comenzarse simplemente por sumar en algún instante con referencia al pasado los valores de las

ES 2 296 176 T3

hipótesis para un número predeterminado de bloques, para a continuación ver qué hipótesis tenía al final, esto es, después de un número determinado de por ejemplo 20 bloques para determinados bloques el mayor valor y de este modo ha sobrevivido y ha “desplazado” a las otras hipótesis. En el ejemplo mostrado en la figura 3, esto significaría que las hipótesis, que es la entidad ID1 o ID2 o ID3 de información, también se continuarían para los periodos Δt_7 y Δt_8 de tiempo, no cambiando esto sin embargo nada con respecto al reconocimiento de ID1, puesto que se comienzan nuevas hipótesis, esto es, la hipótesis para ID108, ID109, ID4 e ID8 sólo considerablemente más tarde, esto es, para los bloques de Δt_7 y Δt_8 o por encima y con ello tales valores de fiabilidad combinados altos sólo se alcanzan mucho más tarde o ni siquiera.

10 A partir de lo anterior se hace evidente que el final de una hipótesis no tiene que determinarse obligatoriamente de manera activa, sino que este final puede obtenerse automáticamente a partir del análisis del pasado, esto es, de las hipótesis comenzadas. Por tanto, preferiblemente, siempre que aparezca un nuevo resultado de identificación con una medida de fiabilidad por encima de un umbral de significancia, se comienza una nueva hipótesis, haciendo entonces en algún momento una retrospectiva para ver qué hipótesis sobrevive para un periodo de tiempo determinado, no
15 teniendo que determinar explícitamente para ello un final de una hipótesis, puesto que se obtiene automáticamente por sí mismo.

En función de la circunstancia, el procedimiento según la invención puede implementarse en hardware o en software. La implementación puede realizarse en un medio de almacenamiento digital, especialmente un disquete o CD con señales de control legibles electrónicamente, que así pueden cooperar con un sistema informático programable de tal manera, que se realiza el procedimiento. En general, la invención consiste así también en un producto de programa informático con un código de programa almacenado en un soporte identificable por máquina para realizar el procedimiento según la invención, cuando el producto de programa informático se ejecuta en un ordenador. En otras palabras, la invención puede realizarse así como un programa informático con un código de programa para la realización del
25 procedimiento, cuando el programa informático se ejecuta en un ordenador.

30

35

40

45

50

55

60

65

REIVINDICACIONES

5 1. Dispositivo para analizar una señal de información, que presenta una secuencia (802) de bloques de unidades (804) de información, representando una pluralidad de bloques sucesivos de la secuencia de bloques una entidad (806) de información, utilizando una secuencia de huellas (FAi) digitales para la secuencia de bloques, de manera que la secuencia de bloques está representada por la secuencia de huellas digitales, con las características siguientes: un medio (12) para proporcionar resultados (IDi) de identificación para huellas digitales sucesivas, representando un resultado de identificación una pertenencia de un bloque de unidades de información a una entidad de información predeterminada, y existiendo para cada resultado de identificación una medida de fiabilidad, estando configurado el medio (12) para proporcionar para generar un primer resultado de identificación para una primera huella digital y para generar para un bloque siguiente un segundo resultado de identificación que se diferencia del primer resultado de identificación;

15 un medio (14) para formar al menos dos hipótesis a partir de los resultados de identificación para las huellas digitales sucesivas, siendo la primera hipótesis una suposición para la pertenencia de la secuencia de bloques a una primera entidad de información, y siendo una segunda hipótesis una suposición para la pertenencia de la secuencia de bloques a una segunda entidad de información, estando configurado el medio (14) para formar para comenzar la primera hipótesis en respuesta al primer resultado de identificación o para continuar la primera hipótesis ya existente y para comenzar la segunda hipótesis en respuesta al segundo resultado de identificación o continuar la segunda hipótesis ya existente;

20 un medio (16) para estudiar las al menos dos hipótesis agrupando las medidas de fiabilidad de las hipótesis, para obtener un resultado (18) de estudio; y

25 un medio (20) para llegar a una conclusión acerca de la señal de información basándose en el resultado de estudio.

2. Dispositivo según la reivindicación 1, en el que el medio (16) para estudiar está configurado para estudiar las hipótesis con respecto a información de probabilidad válida para las hipótesis.

30 3. Dispositivo según la reivindicación 1 ó 2, en el que el medio (20) para llegar a una conclusión está configurado para determinar que la secuencia de bloques representa una entidad de información, que tiene una hipótesis, que es la más probable, o que una entidad de información finaliza con la huella digital, que como la última en el tiempo contribuye a la hipótesis más probable, o que existe o no una entidad de información en la señal de información.

35 4. Dispositivo según una de las reivindicaciones anteriores, en el que el medio (12) para proporcionar está configurado para generar dos resultados de identificación diferentes para una huella digital.

40 5. Dispositivo según la reivindicación 4, en el que el medio (12) para proporcionar está configurado para generar una medida de fiabilidad para cada uno de los dos resultados de identificación diferentes.

6. Dispositivo según la reivindicación 4 ó 5, en el que el medio (14) para formar está configurado para asignar a la primera hipótesis uno primero de los dos resultados de identificación y a la segunda hipótesis uno segundo de los dos resultados de identificación.

45 7. Dispositivo según una de las reivindicaciones 3 a 6, en el que el medio (16) para estudiar está configurado para determinar la hipótesis, que tiene una mayor medida de fiabilidad agrupada.

50 8. Dispositivo según una de las reivindicaciones anteriores, en el que el medio (14) para formar está configurado para finalizar la primera o segunda hipótesis, cuando un número predeterminado de bloques ni obtiene un resultado de identificación, que indica a la primera entidad de información, ni un resultado de identificación, que indica a la segunda entidad de información.

9. Dispositivo según una de las reivindicaciones anteriores, en el que el medio (14) para formar está configurado para finalizar la primera o la segunda hipótesis cuando aparece un evento detectado en la señal de información.

55 10. Dispositivo según la reivindicación 9, en el que existe un detector de eventos que está configurado para detectar un nivel de energía en un bloque de unidades de información, que se encuentra por debajo de un nivel umbral, como el evento.

60 11. Dispositivo según una de las reivindicaciones anteriores, en el que el medio (12) para proporcionar está configurado para emitir para cada huella digital sólo el resultado de identificación más fiable sin o con medida de fiabilidad, para emitir para una huella digital un número predeterminado de huellas digitales más fiables en cada caso con o sin medida de fiabilidad, o para emitir para una huella digital sólo los resultados de identificación con o sin medidas de fiabilidad que tienen una medida de fiabilidad por encima de un umbral.

65 12. Dispositivo según una de las reivindicaciones anteriores, en el que el medio (16) para estudiar está configurado para sumar medidas de fiabilidad explícitas o implícitas que pertenecen a una hipótesis, para obtener una medida de fiabilidad agrupada.

ES 2 296 176 T3

13. Dispositivo según una de las reivindicaciones anteriores, en el que el medio (12) para proporcionar está configurado para realizar con una huella digital una búsqueda en una base de datos, en la que están almacenadas huellas digitales de entidades de información de referencia, y para proporcionar un número de resultados de identificación así como una medida de distancia para cada resultado de identificación como indicación a una medida de fiabilidad para cada resultado de identificación.

14. Dispositivo según la reivindicación 13, en el que el medio (12) para proporcionar está configurado para comenzar una nueva hipótesis para cada resultado de identificación, para el que todavía no existe una hipótesis, cuando una medida de distancia para el resultado de identificación tiene una relación con un umbral, que indica a una distancia menor que una distancia umbral.

15. Dispositivo según una de las reivindicaciones anteriores, en el que el medio (16) para estudiar está configurado para finalizar en respuesta a una determinación todas las hipótesis para las huellas digitales sucesivas, que se han formado para las huellas digitales, que recoge la hipótesis más probable.

16. Dispositivo según una de las reivindicaciones anteriores, en el que la señal de información comprende una señal de audio, en el que la unidad de información son muestras de audio en el dominio de tiempo o de frecuencia, y en el que una entidad de información comprende una composición musical, una secuencia hablada o un segmento de ruido.

17. Dispositivo según una de las reivindicaciones anteriores, en el que una huella digital para un bloque se determina mediante una conversión de tiempo-frecuencia y/o mediante el cálculo de una medida de planeidad espectral para un resultado de la conversión de tiempo-frecuencia.

18. Dispositivo según una de las reivindicaciones anteriores, en el que una huella digital para un bloque se genera de tal manera que la huella digital tiene una cantidad de datos que es menor que una cantidad de datos del bloque.

19. Dispositivo según una de las reivindicaciones anteriores, en el que el medio (12) para proporcionar resultados de identificación está configurado para además de un resultado de identificación proporcionar también un índice temporal nuevo para el resultado de identificación, y en el que el medio (14) para formar hipótesis está configurado para continuar una hipótesis, cuando existe una continuidad entre un índice temporal más actual en la hipótesis y el índice temporal nuevo, o para iniciar una hipótesis cuando no existe la continuidad.

20. Procedimiento para analizar una señal de información que presenta una secuencia (802) de bloques de unidades (804) de información, representando una pluralidad de bloques sucesivos de la secuencia de bloques una entidad (806) de información, utilizando una secuencia de huellas (FAi) digitales para la secuencia de bloques, de manera que la secuencia de bloques está representada por la secuencia de huellas digitales, con las etapas siguientes:

proporcionar (12) resultados (IDi) de identificación para huellas digitales sucesivas, representando un resultado de identificación una pertenencia de un bloque de unidades de información a una entidad de información predeterminada, y existiendo para cada resultado de identificación una medida de fiabilidad, generándose en la etapa de proporcionar un primer resultado de identificación para una primera huella digital y generándose para un bloque siguiente un segundo resultado de identificación que se diferencia del primer resultado de identificación;

formar (14) al menos dos hipótesis a partir de los resultados de identificación para las huellas digitales sucesivas, siendo una primera hipótesis una suposición para la pertenencia de la secuencia de bloques a una primera entidad de información, y siendo la segunda hipótesis una suposición para una pertenencia de la secuencia de bloques a una segunda entidad de información, presentando la etapa de formar las etapas siguientes:

en respuesta al primer resultado de identificación, comenzar la primera hipótesis o continuar la primera hipótesis ya existente, y, en respuesta al segundo resultado de identificación, comenzar la segunda hipótesis o continuar la segunda hipótesis ya existente;

estudiar (16) las al menos dos hipótesis agrupando las medidas de fiabilidad de las hipótesis, para obtener un resultado (18) de estudio; y

llegar (20) a una conclusión acerca de la señal de información basándose en el resultado de estudio.

21. Programa informático con un código de programa adaptado para realizar un procedimiento según la reivindicación 20 de patente cuando el programa se ejecuta en un ordenador.

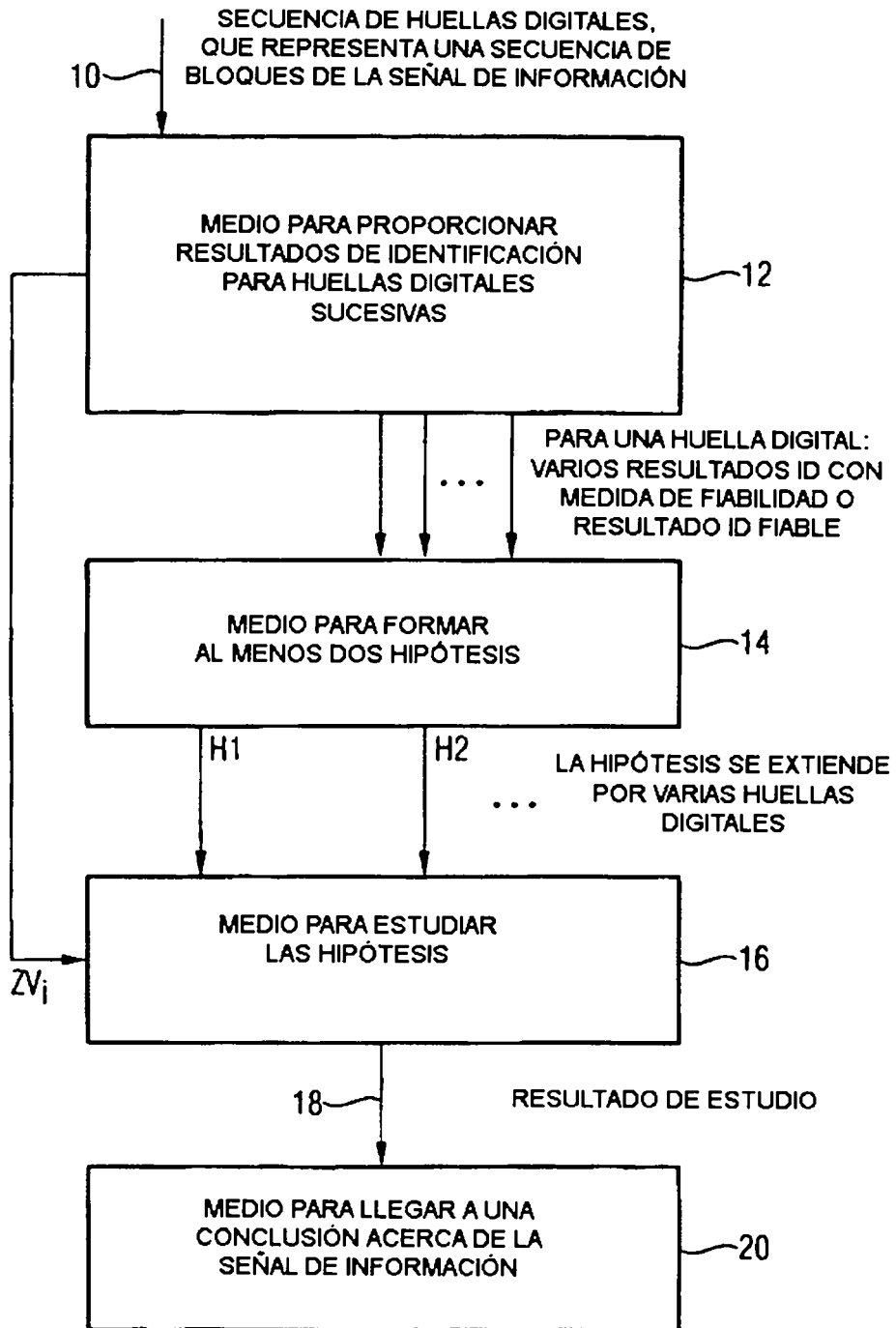


FIG. 1

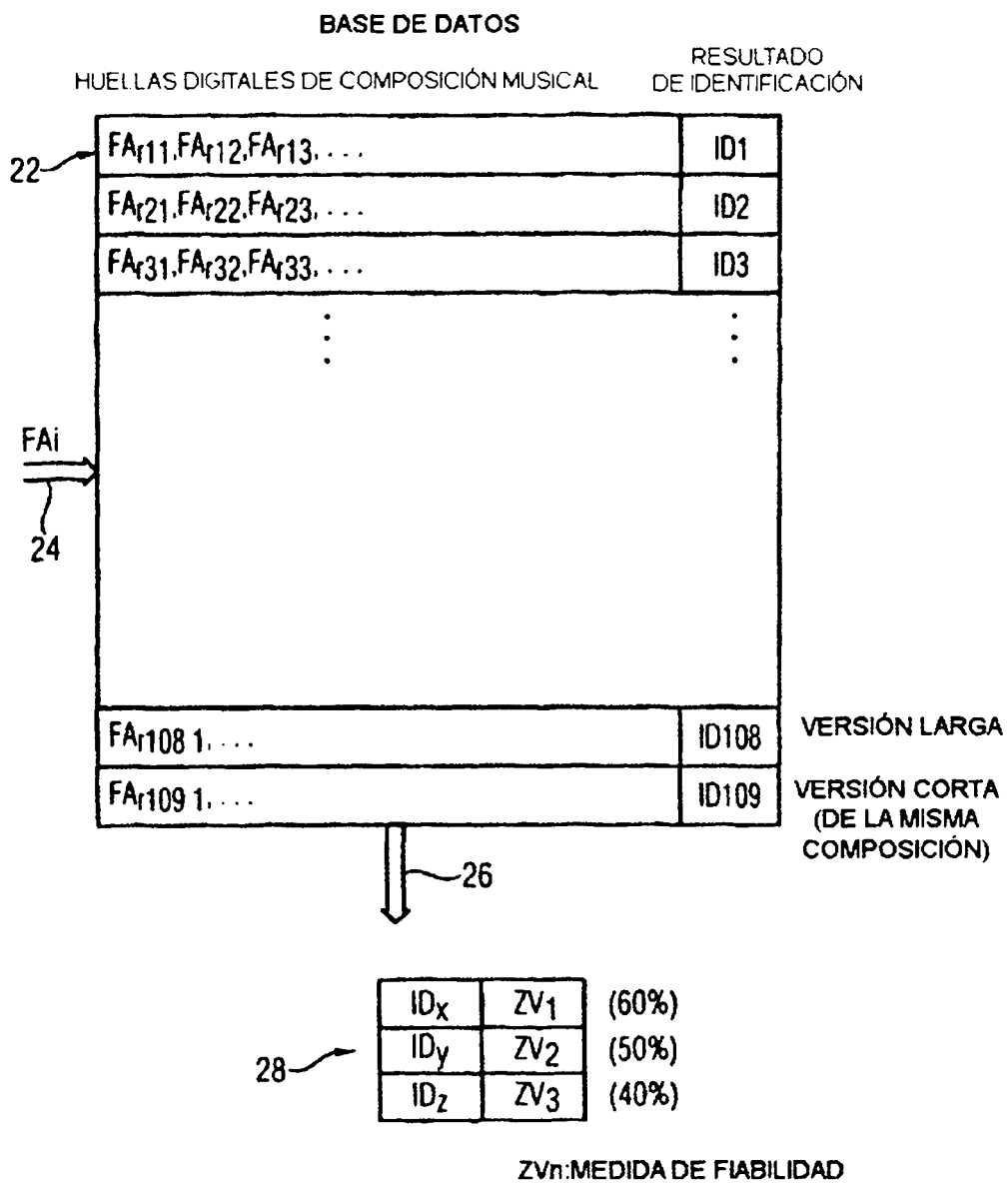


FIG. 2

	Δt_1	Δt_2	Δt_3	Δt_4	Δt_5	Δt_6	Δt_7	Δt_8
	FA1	FA2	FA3	FA4	FA5	FA6	FA7	FA8
	ID1 40%	ID1 60	ID1 80	ID1 60	ID1 90	ID1 90	ID 108 90	ID 108 80
	ID2 60	ID2 70	ID2 30	ID2 40	ID2 20	ID2 10	ID 109 85	ID 109 70
	ID3 30	ID3 20	ID3 30	ID3 40	ID3 10	ID3 5	ID 4 10	ID 4 10
H1	ID1	ID1	ID1	ID1	ID1	ID1	ID108	ID108
H2	ID2	ID2	ID2	ID2	ID2	ID2	ID109	ID109
H3	ID3	ID3	ID3	ID3	ID3	ID3	ID4	ID4
H4								ID8

PARA $\Delta t_1 - \Delta t_6$

Σ ID1: 420

Σ ID2: 230

Σ ID3: 135

TRANSICIÓN DE
LA COMPOSICIÓN MUSICAL

⇒ CONCLUSIÓN: - ENTIDAD DE INFORMACIÓN DE. $\Delta t_1 - \Delta t_6$.ES
COMPOSICIÓN MUSICAL ID1, H1 ERA CORRECTA, Y/O
- Δt_6 ES EL FINAL DE LA 1. ENTIDAD DE INFORMACIÓN
O Δt_7 ES EL INICIO DE LA 2. ENTIDAD DE INFORMACIÓN
; Y/O
- ID1 ESTÁ CONTENIDA EN LA SEÑAL DE INFORMACIÓN

FIG. 3

ID 108 ES UNA VERSIÓN LARGA DE UNA COMPOSICIÓN XY MUSICAL
(LONGITUD POR EJEMPLO 7:32, REPRESENTADO SIMBÓLICAMENTE
POR LA SIGUIENTE HUELLA DIGITAL)

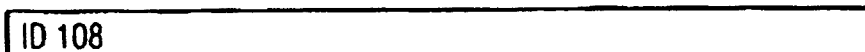


FIG. 4A

ID 109 ES UNA VERSIÓN CORTA DE LA MISMA COMPOSICIÓN XY MUSICAL
(POR EJEMPLO UNA DENOMINADA "VERSIÓN DE RADIO" O
"EDICIÓN DE RADIO", LONGITUD DE POR EJEMPLO 4:38),
REPRESENTADO SIMBÓLICAMENTE POR LA SIGUIENTE HUELLA DIGITAL:



FIG. 4B

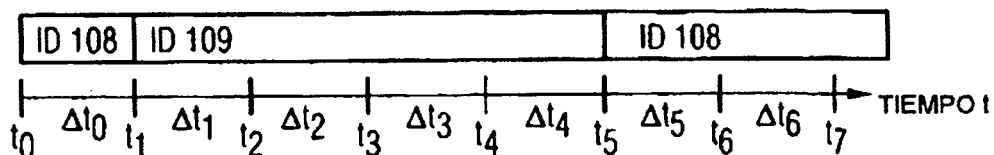


FIG. 4C

ELEMENTOS IDENTIFICADOS PARA LOS PERIODOS DE TIEMPO Δt_x (CONx=0,1,2,3,4,5,6)

ID 108	ID 109	ID 108	ID 109	ID 109	ID 108	ID 108
Δt_0	Δt_1	Δt_2	Δt_3	Δt_4	Δt_5	Δt_6

FIG. 5A

AGRUPACIÓN EN EL PROTOCOLO DE RECONOCIMIENTO

1. t_0 A t_1 : ID 108 (Δt_0)
2. t_1 A t_2 : ID 109 (Δt_1)
3. t_2 A t_3 : ID 108 (Δt_2)
4. t_3 A t_5 : ID 109 ($\Delta t_3, \Delta t_4$)
5. t_5 A t_7 : ID 108 ($\Delta t_5, \Delta t_6$)

FIG. 5B

ELEMENTOS IDENTIFICADOS PARA LOS PERIODOS DE TIEMPO Δt_x (CONx=0,1,2,3,4,5,6)

ID 108	ID 109	ID 109	ID 109	ID 109	ID 108	ID 108
--------	--------	--------	--------	--------	--------	--------

FIG. 5C

AGRUPACIÓN EN EL PROTOCOLO DE RECONOCIMIENTO

1. t_0 A t_1 : ID 108 (Δt_0)
2. t_1 A t_5 : ID 109 ($\Delta t_1 - \Delta t_4$)
3. t_5 A t_7 : ID 108 ($\Delta t_5 - \Delta t_6$)

FIG. 5D

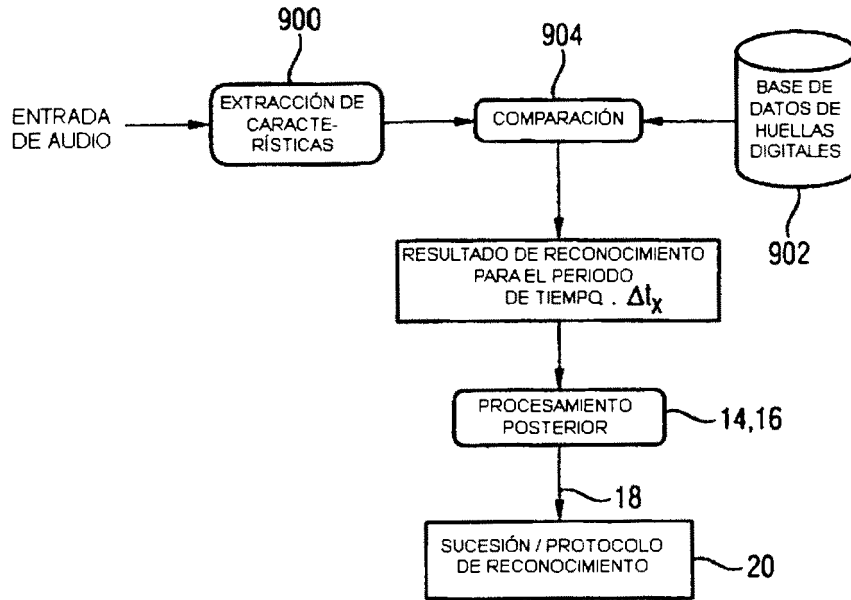


FIG 6

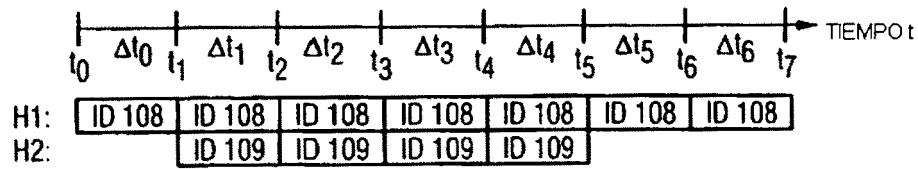


FIG 7A

HIPÓTESIS 1 (H1): t_0 A T_7 : ID 108, MEDIDA DE CONFIANZA $x_{H1}\%$
 HIPÓTESIS 2 (H2): t_1 A T_5 : ID 109, MEDIDA DE CONFIANZA $x_{H2}\%$
 CON $x_{H1}\% > x_{H2}\%$, SIENDO $x_{H1}\% > Y x_{H2}\%$ EN CADA CASO UNA FUNCIÓN DE LAS MEDIDAS DE CONFIANZA DE LOS RECONOCIMIENTOS INDIVIDUALES

FIG 7B

ESTO SIGNIFICA PARA EL PROTOCOLO DE RECONOCIMIENTO:

1. t_0 A t_7 : ID 108 (Δt_0 - Δt_6)

FIG 7C

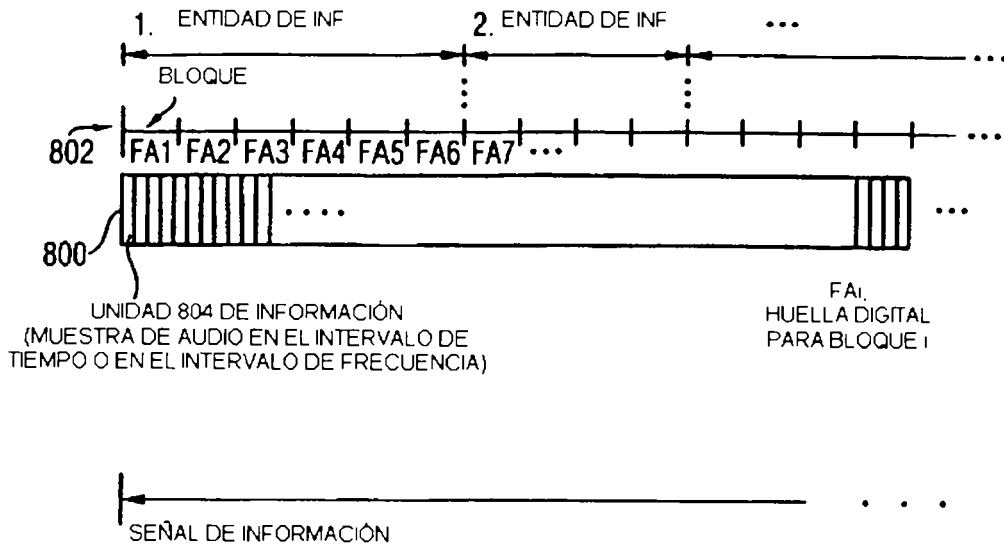


FIG. 8

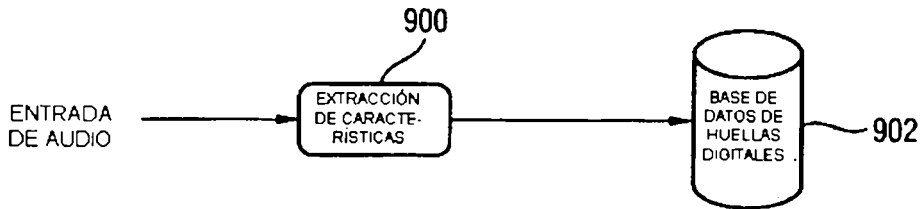


FIG. 9

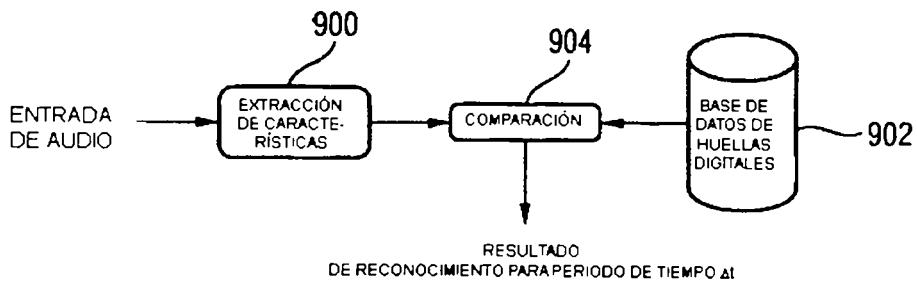


FIG. 10