(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 569 423**

(51) Int. CI.:

G10L 25/54 (2013.01)
H04H 60/58 (2008.01)
H04H 60/37 (2008.01)
G06K 9/00 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Fecha de presentación y número de la solicitud europea: **08.02.2006 E 11186629 (9)**

(97) Fecha y número de publicación de la concesión europea: **03.02.2016 EP 2437255**

(54) Título: **Identificación automática de material repetido en señales de audio**

(30) Prioridad:

08.02.2005 US 651010 P

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
10.05.2016

(73) Titular/es:

SHAZAM INVESTMENTS LIMITED (100.0%)
26-28 Hammersmith Grove
London W6 7HA, GB

(72) Inventor/es:

DE BUSK, DAVID L.;
BRIGGS, DARREN P.;
KARLINER, MICHAEL;
CHEONG TANG, RICHARD WING y
LI-CHUN WANG, AVERY

(74) Agente/Representante:

IZQUIERDO BLANCO, María Alicia

Identificación automática de material repetido en señales de audio

Descripción

Referencia cruzada a solicitudes relacionadas

Esta solicitud reivindica prioridad a la solicitud de patente provisional de Estados Unidos Nº 60/651.010 presentada el 8 de febrero, 2005, titulada IDENTIFICACIÓN AUTOMÁTICA DE MATERIAL REPETIDO EN SEÑALES DE AUDIO

Campo técnico

Esta invención se refiere al reconocimiento e identificación de patrones en archivos de medios, y más particularmente a identificar material referido en señales de medios, particularmente señales de audio, a través de uno o más flujos de medios.

Antecedentes de la invención

Los propietarios de derechos de autor, tales como para el contenido de música o vídeos, generalmente tienen derecho a compensación por cada ocasión en la que se reproduce su canción o vídeo. Para propietarios de derechos de autor de música en particular, determinar cuándo sus canciones se reproducen en cualquiera de los cientos de emisoras de radio, ya sea a través del aire o en internet, es una tarea abrumadora. Tradicionalmente, los propietarios de derechos de autor han entregado una serie de derechos de autor en estas circunstancias a terceras compañías que cargan a entidades que reproducen música para fines comerciales una tasa de suscripción para compensar su catálogo de propietarios de derechos de autor. Estas tasas después de distribuyen a los propietarios de derechos de autor en base a modelos estadísticos diseñados para compensar a aquellos propietarios de derechos de autor cuyas canciones reciben la mayor parte de las reproducciones. Estos métodos estadísticos solamente han sido cálculos muy irregulares de ocasiones reales de reproducción en base a tamaños pequeños de muestras.

La patente de Estados Unidos 6.990.453 publicada el 4 de enero, 2006, describe un sistema y método para comparar muestras no conocidas de medios de un flujo de medios tal como una señal de emisora de radio con una base de datos de archivos conocidos de medios tales como canciones, con el fin de seguir las ocasiones de reproducción de canciones conocidas. Desafortunadamente, mucho del contenido de un flujo de medios no se conoce previamente por una variedad de razones. Por ejemplo, audios únicos tales como programas de entrevistas, las conversaciones o presentaciones de disc-jockeys, o DJs y otros audios similares, representan audios únicos que no serán reconocibles.

Sin embargo, puede haber otros audios no reconocidos que pueden ser de interés para un sistema para controlar flujos de audio, y de hecho, pueden asociarse con un propietario de derechos de autor que debería ser compensado. Tales audios no reconocidos de interés podrían ser una canción no indexada, o un anuncio que puede usar música con derechos de autor u otros segmentos de audio reconocidos y repetidos. Estos segmentos de audio no reconocidos pueden repetirse dentro de un único flujo de medios o pueden repetirse a lo largo de múltiples flujos de medios, tal como un anuncio regional que se reproduce en múltiples emisoras de radio.

Una técnica para identificar y segmentar automáticamente objetos que se repiten en un flujo de medios se desvela por ejemplos en US 2003/0231868 A1.

Lo que se necesita es un sistema y método para reconocer segmentos o muestras repetidas en uno o más flujos de medios no reconocidos de otra manera, donde el sistema y método son capaces de comparar muestras con muestras con huellas digitales o indexadas previamente para encontrar casos de medios repetidos no reconocidos.

Breve resumen de la invención

De acuerdo con la presente invención, se proporciona un sistema como el definido por la reivindicación 1.

Breve descripción de los dibujos

Para una comprensión más completa de la presente invención, y las ventajas de la misma, se hace referencia a las siguientes descripciones tomadas junto con los dibujos acompañantes, en los que:

La FIG. 1 es un diagrama de bloques de una realización de un sistema para crear una base de datos de artículos de interés en flujos de audio no reconocido (ANR);

La FIG. 2 es un diagrama de flujo de un método ejemplar para crear segmentos repetidos coincidentes de ANR;

La FIG. 3 es un diagrama de bloques de un sistema informático ejemplar para implementar un sistema de toma de huellas digitales y puntos de referencia como el aquí descrito;

La FIG. 4 es un diagrama de flujo de un ejemplo de un método para construir un índice de base de datos de archivos de sonido;

La FIG. 5 ilustra esquemáticamente puntos de referencia o huellas digitales calculadas para una muestra de sonido tal como un segmento de ANR; y

La FIG. 6 es un diagrama de flujo de un ejemplo de un método para emparejar muestras o segmentos ANR con muestras o segmentos ANR previamente marcadas con huellas digitales o indexadas.

Descripción detallada de la invención

Ahora es una práctica común usar métodos automatizadas para identificar material pre-grabado contenido en señales de audio tales como grabaciones de emisiones de radio o TV, o grabaciones de actuaciones de material en lugares públicos tales como discotecas. Independientemente de las técnicas usadas, estos métodos requieren un acceso previo al material que se identificará, para que la señal pueda compararse con contenido conocido en una base de datos de referencia. Para mucho material esto no es un problema, ya que puede haber estado disponible comercialmente durante algún tiempo, como en el caso de CDs de música. Sin embargo, un porcentaje significativo de señales de audio consisten en material que no está fácilmente disponible, tal como música antes de la fecha de lanzamiento comercial, material de publicidad o música escrita para fines de identificación de emisoras de radio, por ejemplo,

Esto crea dos problemas para aquellos que participan en cuantificar con precisión el contenido de señales de audio:
(1) que el material que debería identificarse no está porque no está contenido en la base de datos de referencia, y
(2) partes sustanciales de una señal, mientras que no contienen material de interés, no pueden eliminarse de examen manual porque el método automático no lo identifica positivamente como no interesante.

El audio puede identificarse y segmentarse directamente, por ejemplo usando la técnica por Wang (solicitud de patente provisional de Estados Unidos Nº de serie 60/563.372), en la que el audio que se identifica fuera de una base de datos se segmenta en regiones conocidas, dejándose el resto de las regiones como audio no reconocido (ANR). El método de Wang está limitado a reconocer contenido que ya está en una base de datos, y no puede identificar material no contenido en una base de datos.

Para superar esta limitación, se usa un método “filtro” para examinar segmentos no reconocidos de audio (ANR) de flujos de medios controlados y para comprobar si coinciden con otros segmentos o partes de segmentos del pasado reciente. Una vez que se han encontrados tales coincidencias, pueden segmentarse e incluirse en la base de datos de referencia, permitiendo así que se creen bases de datos sin ningún acceso previo al material diana de interés.

En la Figura 1 se muestra un sistema para crear una base de datos de artículos de interés en flujos de ANR. El sistema 10 toma flujos ANR 13 de fuentes de audio no reconocido 11 e identifica segmentos repetidos en el ANR que pueden ser de interés. Los segmentos de ANR se envían a un administrador candidato 13 que recoge y marca cada instante de datos en el flujo de medios con un identificador único. El administrador candidatos 13 después envía el ANR a un generador de huella digital 13 donde los segmentos de audio en bruto del ANR se procesan para extraer características de huellas digitales y se indexan en una base de datos disponible para su consulta. El motor de búsqueda de audio 16 responde a las peticiones de búsqueda de audio del administrador candidato 14 y usa huellas digitales de ANR 15 del generador de huellas digitales 14 para comparar segmentos de ANR con la base de datos de segmentos de ANR previamente indexados. El motor de búsqueda de audio 16 registra coincidencias de segmentos de ANR con segmentos de ANR indexados. Cuando un segmento particular de ANR acumula un cierto número umbral de coincidencias, lo que significa que el sistema 10 ha visto el mismo contenido de audio múltiples veces a lo largo de uno o más flujos de medios, se determina que ese segmento de audio es de suficiente interés significativo para garantizar la identificación positiva. Esto se consigue publicando el contenido significativo, huellas digitales publicadas 18 y añadiéndolo a otros motores de búsqueda reconocidos 19.

Puede necesitarse que los segmentos de ANR significativos, aquellos que tienten múltiples coincidencias en los flujos de medios controlados, se envíen para identificarse positivamente y catalogarse. La identificación de segmentos de ANR significativo puede requerir su envío a un operario humano que escuchará el audio para hacer una identificación positiva. El operario humano identificará el audio e introducirá la información necesaria para permitir que el segmento de ANR se añada a la base de datos de contenido de audio conocido como se describe en otras solicitudes.

Un método para reconocer material repetido se describe con respecto a la Figura 2. El método recoge audio desconocido (ANR) de uno o más flujos de medios para su filtrado, donde cada instante de datos de audio tiene una referencia única de sello de tiempo (tal referencia de sello de tiempo se incrementa con el tiempo y puede aumentar con un identificador de flujo).

Se crea una base de datos de cribado que contiene material de programa de audio desconocido (ANR) a partir de flujos de medios controlados en el pasado reciente para su cribado, donde las grabaciones de audio en bruto se procesan para extraer características de huellas digitales y se indexan en una base de datos disponible para su consulta. Los métodos para preparar tal base de datos se desvelan en Wang y Smith (descritos en la publicación internacional número WO 02/11123 A2, titulada “Sistema y método para reconocer señales de sonido y música en ruido alto y distorsión” y reivindicando prioridad para la solicitud provisional de Estados Unidos Nº 60/222.023 presentada el 31 de julio, 200 y la solicitud de Estados Unidos Nº de serie 09/839.476, presentada el 20 de abril, 2001 (a partir de ahora “Wang y Smith”); o Wang y Cubert (descrito en la publicación internacional Nº WO 03/091990 titulada “Comparación de patrón de audio robusto e invariante” y reivindicando prioridad para la solicitud provisional de Estados Unidos 60/376.055 presentada del 25 de abril, 2002 (a partir de ahora “Wang y Culbert”). El uso de estos métodos particulares “Wang y Smith” o “Wang y Culbert” son ilustrativos y no deben interpretarse como limitativos.

Para procesar la segmentación automática, fragmentos cortos de rastreo del material de flujo de medios desconocido se envían para reconocimiento al motor de búsqueda de audio 16 de la Figura 1 representado una técnica de identificación tal como “Wang y Smith” o “Wang y Culbert”, incorporando la base de datos de cribado, y como se muestra en el proceso 21. En el proceso 22 se toma una determinación si ANR coincide con cualquier candidato al que previamente se han tomado las huellas digitales. Los segmentos coincidentes se identifican después fuera de los flujos de medios desconocidos y se añade el reconocimiento a los candidatos existentes en el proceso 23.

Si se reconoce un fragmento de rastreo $F_0(0)$, los resultados que se coinciden $F_{0,k}(0)$ (donde k es un índice de coincidencia) a partir de la base de datos de cribado se agrupan en una lista de fragmentos que coinciden. La tarea es descubrir los límites y calidad de coincidencias de segmentos que coinciden. Con este fin, los fragmentos de rastreo adyacentes $F_0(t)$, donde t es una tiempo compensado de F0(0), se envían para su reconocimiento y se recuperan sus correspondientes listas de fragmentos que coinciden. Las correspondientes listas de fragmentos que coinciden se escanean para continuidad, esto es, donde fragmentos de rastreo adyacentes se aplican en fragmentos que coinciden sustancialmente adyacentes. Una cadena de fragmentos de rastreo adyacentes puede aplicarse a una o más cadenas paralelas de fragmentos que coinciden. Tal lote de cadenas paralelas forma un grupo candidato. Cada cadena paralela que coincide se extiende lo máximo posible en ambas direcciones en tiempo para formar una cadena paralela que coincida lo máximo. Un punto de segmentación candidato es donde una pluralidad de cadenas paralelas acaba o empieza sustancialmente simultáneamente.

El conjunto de cadenas paralelas que coinciden lo máximo podrían tener diferentes longitudes y también diferentes extremos. Esto puede deberse a la incorporación jerárquica de segmentos repetidos de programa. La segmentación diferencial podría deberse a diferentes versiones de una canción o anuncio. Alternativamente, algún material repetido podría incorporarse en programas repetidos: los programas de radio con frecuencia se transmiten múltiples veces a lo largo del día.

Una manera para determinar una jerarquía de segmentación es ponderar cadenas paralelas en cada fragmento de acuerdo con el número de elementos que coinciden en su lista de fragmento que coincide. El conjunto de cadenas paralelas con la mayor ponderación tiene más posibilidades de ser un segmento atómico de programa, tal como una canción o un anuncio. Las cadenas paralelas con las siguientes ponderaciones más altas podrían deberse a programas repetidos de radio que contienen los segmentos atómicos de programa, tal como para telediarios cada hora o emisiones cíclicas de los 40 éxitos musicales, por ejemplo. Las cadenas paralelas con mayor ponderación son buenas candidatas para ser segmentos atómicos de programa y pueden promocionarse para publicarse en una base de datos de referencia para reconocimiento de anuncios o canciones. Los criterios para su publicación pueden incluir tales parámetros como el número de candidatos en el grupo (cuántas veces se ha repetido el material), y la exactitud de la correlación entre los candidatos, por ejemplo, elegir el segmento con los mejores resultados mutuos totales por parejas contra otros elementos de sus listas de fragmentos que coinciden. Una vez publicado, los flujos de medios de fuente que proporcionaron las muestras originales de audio que corresponden al mejor ejemplar coincidente de material repetido puede copiarse para proporcionar una muestra contigua de audio. La razón por la que hay que identificar un ejemplo “mejor” es típicamente porque algún material repetido, como una pista musical, puede estar superpuesta con material no repetido, como un presentador de un programa hablando sobre música. El “mejor” candidato será el que tenga la menor cantidad de este contenido falso.

Volviendo a la Figura 2, el proceso 24 evalúa candidatos que se han reconocido para determinar si cumplen el umbral para su publicación, como se muestra en el proceso 25. Si el candidato está preparado, se publica como se muestra en el proceso 29. Si el candidato no está preparado para la publicación en el proceso 25, se añade cualquier información adicional a la base de datos del motor de búsqueda en el proceso 26. Volviendo al proceso 22,

si el segmento de ANR no se reconoce, el método salta al proceso 26 donde el segmento se añade a la base de datos del motor de búsqueda.

El método 20, en el proceso 27, determina después si hay algo de ANR antiguo que no haya coincidido que esté preparado para su eliminación. Aunque el flujo de audio sea audio único que nunca coincidirá, tales como programas de entrevistas en directo, promociones radiofónicas, o simplemente la conversación de personalidades de la radio o DJs, esta información necesita eliminarse del sistema para hacer sitio a nuevo ANR que se está procesando. Si hay ANR preparado para su eliminación, como se determina a partir del sello de tiempo, la disponibilidad de memoria para nuevo contenido de ANR, o una combinación de estos u otros factores, el método pasa al proceso 28 que elimina el ANR antiguo. Aquellos expertos en la técnica reconocerán que el método 20 es un proceso continuo que intenta constantemente reconocer ANR nuevo cuando el ANR procesado pasa a través de los otros procesados en el método.

El proceso anterior puede interpretarse como activo en un lote fijo de material de flujo de audio desconocido. Sin embargo, puede mejorarse para procesar datos en una base incremental. Cuando se captura contenido nuevo de flujo de medios, se añade a la base de datos de cribado. El material nuevo también se usa para formar fragmentos de rastreo y se escanea para material repetido como se ha descrito anteriormente. Además, el material antiguo puede retirarse de la base de datos de cribado, previniendo así su crecimiento sin restricciones. Una manera de hacer esto, de acuerdo con “Wang y Smith”, es regenerar continuamente la base de datos usando una ventana móvil de material de flujo de medio desconocido cuando los datos nuevos llegan y los datos antiguos se retiran.

En referencia a las Figuras 3-6, se describe un ejemplo para tomar huellas e indexar segmentos de ANR.

Aunque la invención no se limita a ningún sistema hardware particular, un ejemplo de un sistema informático 30, que puede o no distribuirse, para su uso en la tomar huellas digitales y puntos de referencia en segmentos de medios, tal como un segmento de ANR se ilustra esquemáticamente en la Fig. 2. Los procesadores 32a-32f del sistema 30 conectados por una arquitectura de bus multiprocesador 34 o un protocolo de red tal como el protocolo clúster Beowulf, o una mezcla de los dos. En tal disposición, el índice de la base de datos se almacena preferentemente en una memoria de acceso aleatorio (RAM) en al menos un nodo 32a en el grupo, lo que asegura que la búsqueda de huellas digitales sea rápida. Los nodos computacionales correspondientes a los otros objetos, tales como nodos de puntos de referencia 32c y 32 f, nodos de huellas digitales 32b y 32e, y nodo de escaneo de alineación 32d, no requieren tanto peso de RAM como lo hacen el nodo o nodos 32a que soportan el índice de la base de datos. El número de nodos computaciones asignados a cada objeto pueden por lo tanto modificar su escala de acuerdo con su necesidad de manera que ningún objeto se vuelva un cuello de botella. La red computacional es por lo tanto paralelizable y pueden procesar además múltiples consultas simultáneas de reconocimiento de señal que se distribuyen entre los recursos computacionales disponibles.

En un ejemplo alternativo, ciertos de los objetos funcionales están acoplados más firmemente, mientras permanecen acoplados menos firmemente a otros objetos. Por ejemplo, la toma de puntos de referencia y huellas digitales de objetos puede residir en una localización físicamente separada del resto de los objetos computacionales. Un ejemplo de esto es una asociación firme de la toma de puntos de referencia y señales digitales de objetos con el proceso de captura de señal. En esta disposición, la toma de puntos de referencia y huellas digitales de objetos puede incorporarse como hardware o software adicional incorporado, por ejemplo, en un teléfono móvil, protocolo de aplicación inalámbrica (WAP), navegador, asistente digital personal (PDA), u otra terminal remota, tal como un servicio de identificación de contenido, la toma de puntos de referencia y huellas digitales de objetos puede incorporarse en la aplicación navegadora del cliente como un conjunto enlazado de instrucciones software o un módulo software adicional como una biblioteca de enlace dinámico (DLL) de Microsoft. En estos ejemplos, la captura de señal combinada, la toma de puntos de referencia y huellas digitales de un objeto constituyen el cliente final del servicio. El cliente final envía un resumen extraído de la aplicación de la muestra de señal capturada que contienen parejas de puntos de referencia y huellas digitales al servidor final, que realiza el reconocimiento. El envío de este resumen extraído de la aplicación al servidor, en lugar de la señal capturada en bruto, es ventajoso porque la cantidad de datos se reduce en gran medida, a menudo por un factor de 500 o superior. Tal información puede enviarse en tiempo real en un canal lateral con ancho de banda bajo junto con o en lugar de, por ejemplo, un flujo audio transmitido al servidor. Esto posibilita la realización del ejemplo sobre redes públicas de comunicación, que ofrecen anchos de bandas de tamaño relativamente pequeño para cada usuario.

Un ejemplo de un método para toma de huellas digitales y puntos de referencia se describirá ahora con detalle con referencia a muestras de audio, que pueden ser muestras o segmentos de ANR, y segmentos de ANR ya indexados en una base de datos tal como la base de datos 14 de la Figura 4.

Antes de realizar el reconocimiento, debe construirse un índice de base de datos de sonidos disponible para su consulta. Como aquí se usa, una base de datos es una colección indexada de datos, y no se limita a las bases de datos comercialmente disponibles. En el índice de la base de datos, lo elementos de datos relacionados se asocian entre sí, y los elementos individuales pueden usarse para recuperar datos asociados. El índice de base de datos de sonidos que contiene un conjunto índice para cada archivo o grabación en la colección o biblioteca de

grabaciones seleccionada, que puede incluir habla, música, anuncios, firmas de sonar y otros sonidos. Cada grabación también tiene un identificador único, sonido_ID. La propia base de datos de sonidos no almacena necesariamente los archivos de audio para cada grabación, pero los sonidos_IDs pueden usarse para recuperar los archivos de audio desde cualquier otro sitio. SE espera que el índice de base de datos de sonidos sea muy grande, conteniendo índices para millones o incluso billones de archivos. Preferentemente se añaden nuevas grabaciones progresivamente al índice de la base de datos.

En la Fig. 4 se muestra un diagrama de bloques de un método preferente 40 para construir el índice de base de datos de sonidos disponible para su consulta de acuerdo con un primer ejemplo. En este ejemplo, primero se calculan puntos de referencia, y después se calculan huellas digitales en o cerca de los puntos de referencia. Como será aparente para un experto en la técnica, pueden idearse métodos alternativos para construir el índice de la base de datos. En particular, muchas de las etapas enumeradas más abajo son opcionales, pero sirven para generar un índice de base de datos que se busca con más eficacia. Mientras la eficiencia de búsqueda es importante para reconocimiento de sonido en tiempo real a partir de bases de datos grandes, las bases de datos pequeñas pueden buscarse relativamente rápido incluso si no se han clasificado óptimamente.

Para indexar las bases de datos, cada grabación en la colección se somete a un análisis de toma de puntos de referencia y huellas digitales que genera un conjunto índice para cada archivo de audio. La Figura 5 ilustra esquemáticamente un segmento de una grabación de sonido para la que se han calculado puntos de referencia (PR) y huellas digitales (HD). Los puntos de referencia ocurren en puntos específicos en el tiempo del sonido y tienen valores en las unidades de tiempo compensadas de inicio del archivo, mientras las huellas digitales caracterizan el sonido en o cerca de un punto de referencia particular. Así, en este ejemplo. Cada punto de referencia para un archivo particular es único, mientras que la misma huella dactilar puede ocurrir numerosas veces en un único archivo o en múltiples archivos.

En la etapa 42, se toman puntos de referencia en cada grabación de sonido usando métodos para encontrar localizaciones distintivas y reproducibles en la grabación de sonido. Un algoritmo preferente para toma de puntos de referencia es capaz de marcar los mismos puntos en el tiempo en una grabación de sonido a pesar de la presencia de ruido y otras distorsiones lineales y no lineales. Algunos métodos de toma de puntos de referencia son conceptualmente independientes del proceso de toma de huellas digitales descrito más abajo, peor pueden elegirse para optimizar la actuación de éste último. La toma de puntos de referencia da como resultado una lista de puntos en el tiempo {punto de referencia} en la grabación de sonido en la que posteriormente se calculan las huellas digitales. Un buen diseño de toma de puntos de referencia marca aproximadamente 5-10 puntos de referencia por segundo de grabación de sonido; por supuesto, la densidad de la toma de puntos de referencia depende de la cantidad de actividad en la grabación de sonido.

Son posibles una variedad de técnicas para calcular puntos de referencia, todas ellas dentro del alcance del presente ejemplo. Los procesos técnicos específicos usados para implementar los diseños de toma de puntos de referencia del ejemplo son bien conocidos en la técnica y se analizarán con más detalle. Una técnica simple de toma de puntos de referencia, conocida como Norma Power, es calcular la potencia instantánea en cada punto en el tiempo posible en la grabación y seleccionar los máximos locales. Una manera de hacer esto es calcular la curva rectificando y filtrando la forma de onda directamente.

Otra manera es calcular la transformada de Hilbert (cuadratura) de la señal y usar la suma de las magnitudes cuadradas de la transformada de Hilbert y la señal original.

El método Norma Power de toma de puntos de referencia es bueno encontrando estados transitorios en la señal de sonido. Norma Power es realmente un caso especial de la Norma general espectral Lp en la que p=2. La norma general espectral Lp se calcula en cada momento a lo largo de la señal de sonido calculando un espectro de tiempo corto, por ejemplo por medio de una Transformada rápida de Fourier con ventana de Hanning (FFT). Un ejemplo preferente usa una velocidad de muestreo de 8000 Hz, un tamaño de marco FFT de 1024 muestras, y un paso de 64 muestras para cada porción de tiempo. La norma Lp para cada porción de tiempo se calcula después como la suma de la potencia p^ de los valores absolutos de los componentes espectrales, opcionalmente seguido por la toma de la raíz -p^. Como antes, los puntos de referencia se eligen como los máximos locales de los valores resultantes con el paso del tiempo. Un ejemplo de la norma espectral Lp se muestra en la Figura 5, un gráfico de la norma L4, como una función de tiempo para una señal particular. Las líneas discontinuas en los máximos locales indican la localización de los puntos de referencia elegidos.

Cuando p=oo, la norma Lao es efectivamente la norma máxima. Es decir, el valor de la norma es el valor absoluto del mayor componente espectral en la porción espectral. Esta norma da como resultado puntos de referencia robustos y una buena actuación de reconocimiento total, y es preferentemente para música tonal. Alternativamente, los puntos de referencia espectrales “multi-porciones” pueden calcularse tomando la suma de potencias p01 de valores absolutos de componentes espectrales sobre múltiples porciones de tiempo en compensaciones fijas o variables entre sí, en lugar de una única porción. Encontrar los máximos locales de eta suma extendida permite la optimización de la colocación de huellas digitales multi-porciones, como se describe más abajo.

Una vez que se han calculado los puntos de referencia, se calcula una huella digital en cada punto en el tiempo del punto de referencia en la grabación en la etapa 44. La huella digital es generalmente un valor o conjunto de valores que resume un conjunto de características en la grabación en o cerca del punto en el tiempo. En un ejemplo actualmente preferente, cada huella digital es un único valor numérico que es una función hash de múltiples características. Los tipos posibles de huellas digitales incluyen huellas digitales de porciones espectrales, huellas digitales multi-porciones, coeficientes LPD y coeficientes cepstrales. Por supuesto, cualquier tipo de huella digital que caracteriza la señal o características de la señal cerca de un punto de referencia está dentro del alcance del presente ejemplo. Las huellas digitales pueden calcularse mediante cualquier tipo de procesamiento de señal digital o análisis de frecuencia de la señal.

Para generar huellas digitales de porción espectral, se realiza un análisis de frecuencia en el área de cada punto en el tiempo del punto de referencia para extraer los varios picos espectrales superiores. Un simple valor de huella digital es solamente el valor de frecuencia única del pico espectral más fuerte. El uso de tal pico simple da como resultado un reconocimiento sorprendentemente bueno en la presencia de ruido; sin embargo, las huellas digitales de porción espectral con única frecuencia tienden a generar más falsos positivos que otros diseños de huellas digitales porque no son únicos. El número de falsos positivos puede reducirse usando huellas digitales consistentes en una función de los dos a tres picos espectrales más fuertes. Sin embargo, puede haber una mayor susceptibilidad al ruido si el segundo pico espectral más fuerte no es lo suficientemente fuerte para distinguirse de sus competidores en presencia de ruido. Esto es, el valor calculado de huella digital no puede ser lo suficientemente robusto como para poder reproducirse de manera fiable. A pesar de esto, la actuación de este caso es también buena.

Con el fin de tomar ventaja de la evolución en el tiempo de muchos sonidos, se determina un conjunto de porciones de tiempo añadiendo un conjunto de compensaciones de tiempo a un punto en el tiempo en un punto de referencia. En cada porción de tiempo resultante, se calcula una huella digital de porción espectral. El conjunto resultante de información de huella digital se combina después para formar una huella digital de múltiples tonos o múltiples porciones. Cada huella digital multi-porción es mucho más única que la huella digital con única porción espectral, porque rastrea la evolución temporal, dando como resultado menos coincidencias falsas en la búsqueda del índice de base de datos descrito más abajo. Los experimentos indican que debido a su mayor singularidad, las huellas digitales multi-porción calculadas a partir del único pico espectral más fuerte, en cada dos porciones de tiempo da como resultado un cálculo mucho más rápido (aproximadamente 100 veces más rápido) en la posterior búsqueda en el índice de la base de datos, pero con algunas degradación en el porcentaje de reconocimiento en presencia de ruido significativo.

Alternativamente, en lugar de usar una compensación fija o compensaciones desde una porción dada de tiempo para calcular una huella digital multi-porción, pueden usarse compensaciones variables. La compensación variable para la porción elegida es la compensación para el siguiente punto de referencia, o un punto de referencia en un cierto rango de compensaciones desde el punto de referencia “ancla” para la huella digital. En este caso, la diferencia de tiempo entre los puntos de referencia también está codificada en la huella digital, junto con información multi-frecuencia. Al añadir más dimensiones a las huellas digitales, se convierten en más únicas y tienen menos posibilidades de tener falsas coincidencias.

Además de los componentes espectrales, otras características espectrales pueden extraerse y usarse como huellas digitales. El análisis de codificación predictiva lineal (CPL) extrae las características linealmente predecible de una señal, tales como picos espectrales, así como forma espectral. CPL es bien conocido en la técnica de procesamiento de señales digitales. Para el presente ejemplo, los coeficientes de CPL de porciones de forma de onda anclados en las posiciones de puntos de referencia pueden usarse como huellas digitales realizando la función hash de coeficientes CPL cuantificados en un valor índice.

Los coeficientes cepstrales son útiles como una medición de periodicidad y pueden usarse para caracterizar señales que son armónicas, tales como voces o muchos instrumentos musicales. El análisis cepstral es bien conocido en la técnica de procesamiento de señales digitales. Para el presente ejemplo, se realiza la función hash a un número de coeficientes juntos en un índice y se usan como una huella digital.

En la Figura 6 se muestra un diagrama de bloques que ilustra conceptualmente las etapas generales de un ejemplo de un método 60 para comparar segmentos de ANR con huellas digitales de ANR, como por el motor de búsqueda de audio 16 en la Figura 1. Las etapas individuales se describen con más detalle más abajo. El método identifica una huella digital de ARN coincidente cuyas localizaciones relativas de huellas digitales características que más coinciden con las localizaciones relativas de las mismas huellas digitales de la muestra exógena de ANR. Después de capturar una muestra exógena en la etapa 62, los puntos de referencia y las huellas digitales se calculan en la etapa 64. Los puntos de referencia ocurren en localizaciones particulares, por ejemplo, puntos en el tiempo, en la muestra. Preferentemente, la propia muestra determina la localización en la muestra de los puntos de referencia, es decir, es dependiente de las mismas cualidades y es reproducible. Esto es, se calculan los mismos puntos de referencia para la misma señal cada vez que se repite el proceso. Para cada punto de referencia, se obtiene una huella digital que caracteriza una o más características de la muestra o cerca de la muestra. La cercanía de una característica a un punto de referencia se define por el método de toma de huellas digitales usado. En

algunos casos, se considera una característica cercana a un punto de referencia si corresponde claramente con el punto de referencia y no a un punto de referencia anterior o posterior. En otros casos, las características corresponden a múltiples puntos de referencia adyacentes. Por ejemplo, huellas digitales de texto pueden ser cadenas de palabras, huellas digitales de audio pueden ser componentes espectrales y huellas digitales de imágenes pueden ser valores RGB de pixeles. Más abajo se describen dos ejemplos generales de la etapa 64, uno en el que los puntos de referencia y las huellas digitales se calculan secuencialmente, y otro en el que se calculan simultáneamente.

En la etapa 66, las huellas digitales de muestra se usan para recuperar conjuntos de huellas digitales coincidentes almacenadas en un índice de base de datos 68, en la que las huellas digitales coincidentes se asocian con puntos de referencia e identificadores de un conjunto de huellas digitales de ANR. El conjunto de identificadores de archivos y valores de puntos de referencia recuperados se usan después para generar parejas de correspondencia (etapa 70) que contiene los puntos de referencia de la muestra (calculados en la etapa 64) y puntos de referencia de archivos recuperados en los que se calcularon las mismas huellas digitales. Las parejas de correspondencia resultantes se clasifican después mediante un identificador, generando conjuntos de correspondencias entre puntos de referencia de muestra y puntos de referencia de archivo para cada archivo aplicable. Cada conjunto se escanea para alineación entre los puntos de referencia de archivo y los puntos de referencia de muestra. Esto es, se identifican las correspondencias lineales en las parejas de puntos de referencia, y el conjunto se almacena de acuerdo con el número de parejas que se relacionan linealmente. Una correspondencia lineal ocurres cuando un gran número de localizaciones correspondientes de muestra y localizaciones de archivo pueden describirse con sustancialmente la misma ecuación lineal, dentro de una tolerancia permitida. Por ejemplo, si las curvas de número de ecuaciones que describen un conjunto de parejas de correspondencia varían en un ±5%, entonces el conjunto entero de correspondencias se considera linealmente relacionado. Por supuesto, puede seleccionarse cualquier tolerancia adecuada. El identificador del conjunto con el mayor resultado, esto es, con el mayor número de correspondencias linealmente relacionadas, es el identificador ganador de huella digital de ANR, que se localiza y regresa a la etapa 72.

Como se describe además más abajo, el reconocimiento se realiza con un componente de tiempo proporcional al logaritmo del número de entradas en la base de datos. El reconocimiento puede realizarse esencialmente en tiempo real, incluso con una base de datos muy grande. Esto es, una muestra puede reconocerse cuando se está obteniendo, con un pequeño retraso de tiempo. El método puede identificar un sonido en base a segmentos de 5-10 segundos e incluso tan bajos como 1-3 segundos. En un ejemplo preferente, el análisis de toma de puntos de referencia y huellas digitales, etapa 64, se realiza en tiempo real cuando las muestras se están capturando en la etapa 62. Las consultas en la base de datos (etapa 66) se realizan cuando las huellas digitales de muestra están disponibles, y los resultados de correspondencia se acumulan y escanean periódicamente para correspondencias lineales. Así, todas las etapas del método ocurren simultáneamente, y no en la manera lineal secuencial sugerida en la Fig. 6. Hay que señalar que el método es en parte análogo a un motor de búsqueda de texto: un usuario envía una muestra de consulta, y se devuelve un archivo coincidente indexado en la base de datos de sonidos.

Como se ha descrito anteriormente, este método identifica automáticamente material repetido, con una granularidad de tiempo que es dependiente de la longitud de las muestras de audio originalmente presentadas. Aunque esto es por sí mismo útil con los refinamientos del motor de reconocimiento de audio enumerado anteriormente, la granularidad puede mejorarse sustancialmente. El método para una mejor resolución de tiempo de material candidatos es el mismo que anteriormente, excepto que el motor de reconocimiento de audio devuelve la posición y longitud de una coincidencia en una muestra de audio, permitiendo así que el sistema esté libre de la granularidad de la muestra de audio (solicitud de patente referencia “Un método para caracterizar la superposición de dos segmentos de medios”). La técnica ahí desvelada mira a la densidad de soporte de un número de características coincidentes superpuestas alineadas en el tiempo de los datos de audio. Una región de superposición “coincidente” entre dos fragmentos de muestra de audio tiene una alta densidad; sin embargo, las regiones no coincidentes tienen baja densidad. Se elige un punto candidato de segmentación en una compensación de tiempo en un fragmento de muestra coincidente de medio desmarcando una transición entre superposición de características de alta y baja densidad. Este refinamiento produce extremos de segmentos en 100-200 milisegundos.

El sistema y método aquí desvelados se implementan típicamente mientras el software está funcionando en el sistema informático, con etapas individuales implementadas más eficientemente con módulos independientes de software. El código informático de instrucciones para los diferentes objetos se almacena en una memoria de uno o más ordenadores y se ejecuta por uno o más procesadores. En un ejemplo, los objetos del código se agrupan juntos en un único sistema informático, tal como un ordenador personal con base Intel u otras terminales de trabajo. En un ejemplo preferente, el método puede implementarse por un grupo en red de unidades de procesamiento central (CPUs), en las que diferentes procesadores ejecutan diferentes objetos de software con el fin de distribuir la carga computacional. Alternativamente, cada CPU puede tener una copia de todos los objetos de software, permitiendo una red homogénea de elementos idénticamente configurados. En esta segunda configuración, cada CPU tiene un subconjunto del índice de base de datos y es responsable de buscar su propio subconjunto de archivos de medios.

Aunque la presente invención y sus ventajas se han descrito con detalle, debería entenderse que pueden hacerse varios cambios, sustituciones y alteraciones sin partir de la invención como la definen las reivindicaciones adjuntas. Además, el alcance de la presente solicitud no pretende limitarse a los ejemplos particulares del proceso, máquina, fabricación, composición de materia, medios, métodos y etapas descritas en la especificación. Como se apreciará fácilmente a partir de la divulgación, pueden utilizarse los procesos, máquinas, fabricación, composición de materia, medios, métodos o etapas, existentes en el presente o que se desarrollarán más tarde que realizan sustancialmente la misma función o consiguen sustancialmente el mismo resultado que los ejemplos correspondientes aquí descritos.

Reivindicaciones

1. Un sistema para crear una base de datos de artículos de interés en flujos de audio no reconocido (ANR), tomando los flujos de audio no reconocido (ANR) de fuentes de audio no reconocido e identificando segmentos repetidos en el audio no reconocido (ANR) que pueden ser de interés, comprendiendo el sistema:

un administrador candidato, configurado para recibir segmentos de audio no reconocido (ANR), para recoger y marcar cada instante de datos en los segmentos de audio no reconocido (ANR) con un único identificador;
un generador de huellas digitales, configurado para recibir los segmentos de audio no reconocido (ANR) del administrador candidato, para procesar segmentos de audio en bruto del audio no reconocido (ANR) para extraer características de huellas digitales para los segmentos de audio no reconocido (ANR) y para indexarlos en una base de datos disponible para su consulta; y
un motor de búsqueda de audio configurado para responder a peticiones de búsquedas del administrador candidato, para usar huellas digitales de segmentos de audio no reconocido (ANR) del generador de huellas digitales para comparar los segmentos de audio no reconocido (ANR) con los segmentos de audio no reconocido (ANR) previamente indexados en la base de datos disponible para su búsqueda, para registrar coincidencias de los segmentos de audio no reconocido (ANR) con los segmentos de audio no reconocido (ANR) indexados y cuando un segmento particular de audio no reconocido (ANR) acumula un cierto número umbral de coincidencias, para determinar que el segmento particular de audio no reconocido (ANR) es de interés suficientemente significativo para garantizar una identificación positiva publicando el contenido significante de huellas digitales, y añadiendo las huellas digitales publicadas a otro motor de búsqueda.

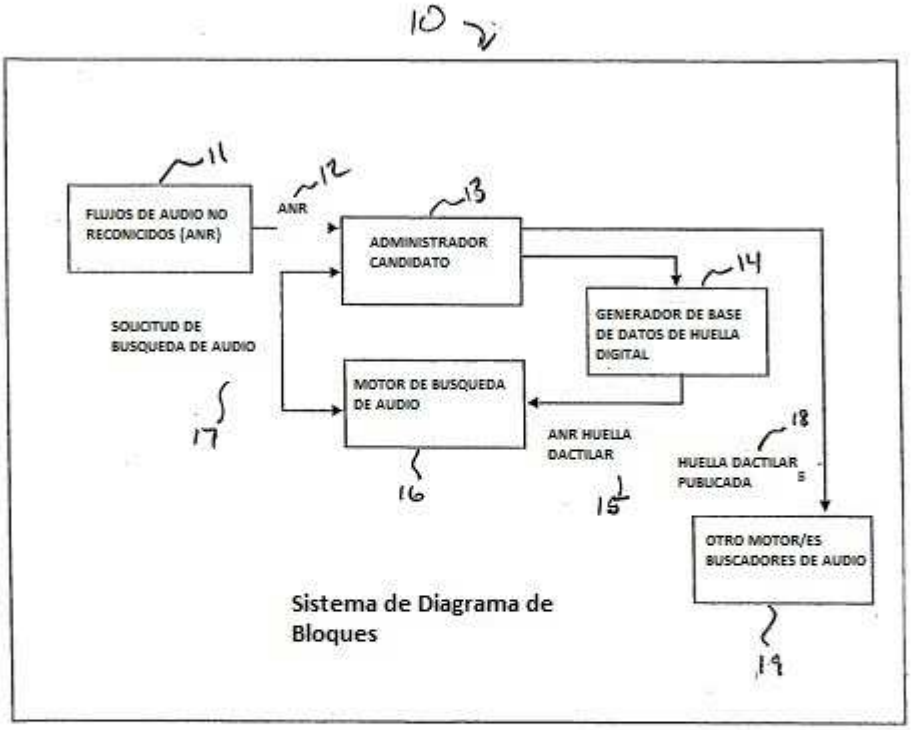
10
11
FLUJOS DE AUDIO NO RECONOCIDOS (ANR)
12
ANR
13
ADMINISTRADOR CANDIDATO
14
GENERADOR DE BASE DE DATOS DE HUELLA DIGITAL
SOLICITUD DE BUSQUEDA DE AUDIO
17
MOTOR DE BUSQUEDA DE AUDIO
16
ANR HUELLA DACTILAR
15
18
HUELLA DACTILAR PUBLICADA
OTRO MOTOR/ES BUSCADORES DE AUDIO
19
Sistema de Diagrama de Bloques

FIGURA 1

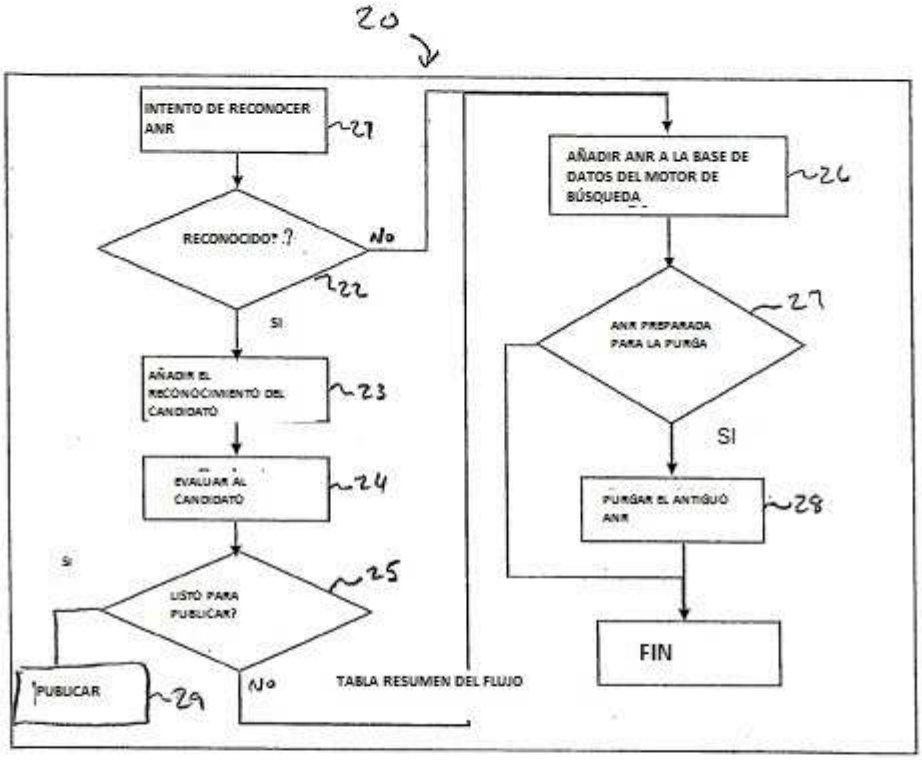
20
INTENTO DE RECONOCER ANR
21
RECONOCIDO? ?
No
22
SI
AÑADIR EL RECONOCIMIENTO DEL CANDIDATO
23
EVALUAR AL CANDIDATO
24
Si
LISTO PARA PUBLICAR?
25
No
PUBLICAR
29
TABLA RESUMEN DEL FLUJO
AÑADIR ANR A LA BASE DE DATOS DEL MOTOR DE BÚSQUEDA
26
ANR PREPARADA PARA LA PURGA
27
SI
PURGAR EL ANTIGUO ANR
28
FIN

FIGURA 2

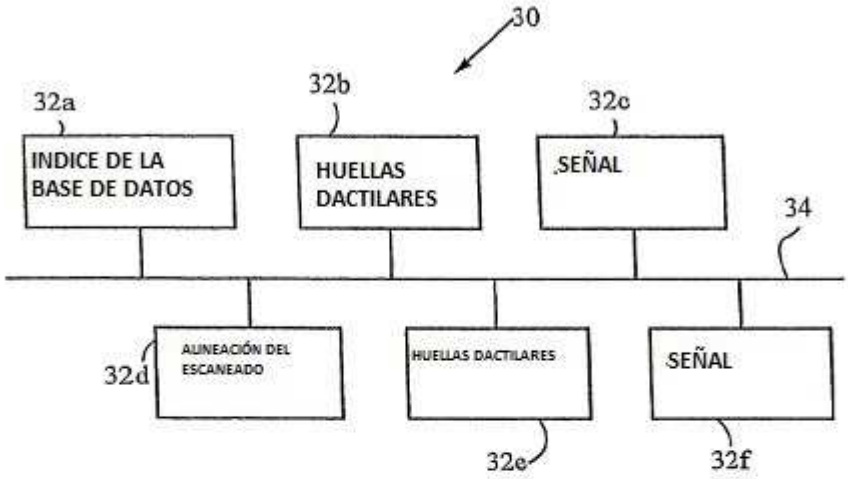
30
32a
INDICE DE LA BASE DE DATOS
32b
HUELLAS DACTILARES
32c
SEÑAL
34
32d
ALINEACIÓN DEL ESCANEADO
HUELLAS DACTILARES
SEÑAL
32e
32f

FIG. 3

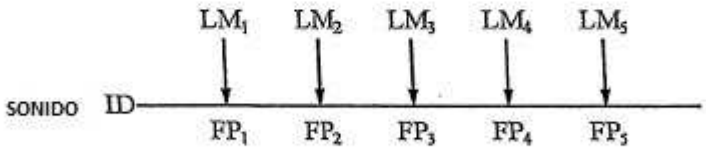
LM1
LM2
LM3
LM4
LM5
SONIDO
ID
FP1
FP2
FP3
FP4
FP5

FIG. 5

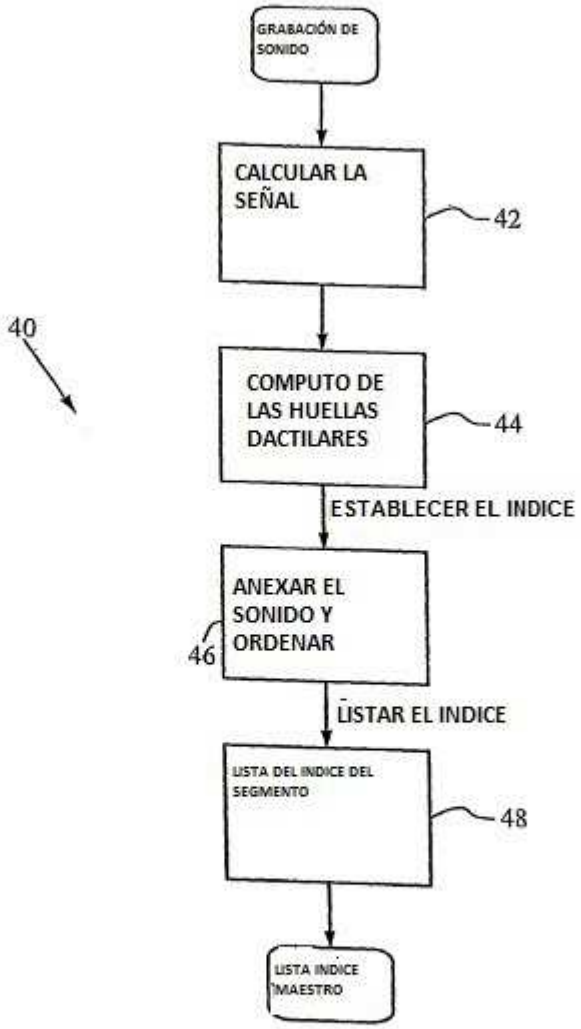
GRABACIÓN DE SONIDO
CALCULAR LA SEÑAL
42
40
COMPUTO DE LAS HUELLAS DACTILARES
44
ESTABLECER EL INDICE
ANEXAR EL SONIDO Y ORDENAR
46
LISTAR EL INDICE
LISTA DEL INDICE DEL SEGMENTO
48
LISTA INDICE MAESTRO

FIG. 4

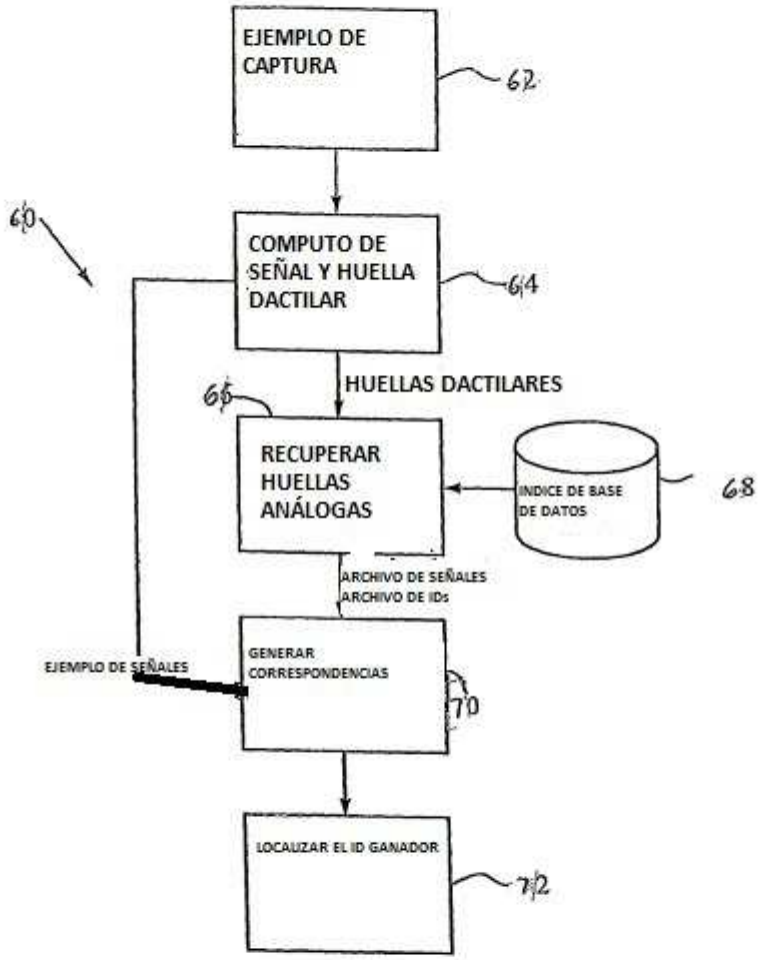
EJEMPLO DE CAPTURA
62
60
COMPUTO DE SEÑAL Y HUELLA DACTILAR
64
HUELLAS DACTILARES
66
RECUPERAR HUELLAS ANÁLOGAS
INDICE DE BASE DE DATOS
68
ARCHIVO DE SEÑALES
ARCHIVO DE IDs
EJEMPLO DE SEÑALES
GENERAR CORRESPONDENCIAS
70
LOCALIZAR EL ID GANADOR
72

FIG. 6