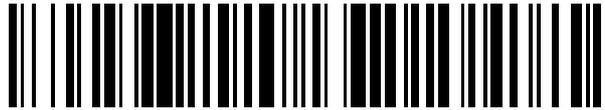


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 488 719**

51 Int. Cl.:

G10L 25/51 (2013.01)

G10L 25/18 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **02.06.2011 E 11726480 (4)**

97 Fecha y número de publicación de la concesión europea: **14.05.2014 EP 2580750**

54 Título: **Sistema y método para el reconocimiento de medios de audio**

30 Prioridad:

09.06.2010 US 352904 P

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

28.08.2014

73 Titular/es:

**ADELPHOI LIMITED (100.0%)
26 Litchfield Street
London WC2H 9TZ, GB**

72 Inventor/es:

**SELBY, ALEXANDER PAUL y
OWEN, MARK ST JOHN**

74 Agente/Representante:

PÉREZ BARQUÍN, Eliana

ES 2 488 719 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

Sistema y método para el reconocimiento de medios de audio

5 **Campo técnico**

La invención se refiere a sistemas y métodos de reconocimiento de audio para el reconocimiento automático de contenidos de medios de audio.

10 **Antecedentes**

Son conocidos diversos sistemas y métodos de reconocimiento de audio para procesar un flujo de audio entrante (un 'programa') y buscar en una base interna de datos de música y efectos de sonido ('pistas') para identificar usos de esas pistas dentro del programa.

15 En la vida real, la música es a menudo solamente una de las capas de audio de un programa. Uno de los retos para el reconocimiento del audio es reconocer la identidad de la música incluso en circunstancias donde hay otras capas de audio, tales como efectos de sonido, voz sobre-grabada, sonidos de fondo, etc., que ocurren simultáneamente. Otras distorsiones incluyen la igualación (ajuste de magnitudes globales relativas de agudos y bajos en una pista) y el cambio de ritmo y / o de tono.

20 Algunas técnicas de reconocimiento de audio están basadas en llevar a cabo directamente una búsqueda en la vecindad cercana de valores de troceo calculados, usando un algoritmo estándar. Allí donde el espacio en que se busca tiene un gran número de dimensiones, tales algoritmos estándar no rinden muy eficazmente.

25 Un artículo titulado "Un sistema de dactiloscopia de audio sumamente robusto", de J. Haitsma et al., de Philips Research, publicado en los Anales de la 3ª Conferencia Internacional sobre Extracción de Información Musical, 2002, describe un sistema de dactiloscopia de medios para comparar objetos de multimedia. El artículo describe que las huellas de un gran número de objetos de multimedia, junto con los meta-datos asociados (por ejemplo, el nombre del artista, el título y el álbum), son almacenados en una base de datos de modo que las huellas sirvan como un índice para los meta-datos. El contenido de multimedia no identificado puede ser luego identificado calculando una huella y usando esto para consultar la base de datos. El artículo describe un algoritmo de búsqueda de dos fases que está basado en realizar solamente comparaciones de huellas completas en posiciones candidatas, preseleccionadas por una búsqueda de sub-huellas. Las posiciones candidatas son localizadas usando una tabla de troceo, o de consulta, que tiene sub-huellas de 32 bits como entrada. Cada entrada apunta a una lista con punteros a las posiciones en las listas de huellas reales, donde están ubicadas las respectivas sub-huellas de 32 bits.

40 El documento US 2003/086341 revela la identificación de copias de grabaciones de sonidos originales, extrayendo características de la copia, creando un vector de esas características y comparando ese vector con una base de datos de vectores. La identificación puede ser realizada para copias de grabaciones de sonidos que hayan sido sometidas a la compresión y a otra manipulación, de modo que no sean réplicas exactas del original. La eficacia de cálculo permite que sean atendidos muchos cientos de consultas al mismo tiempo. Los vectores pueden ser de menos de 100 octetos, por lo que muchos millones de vectores pueden ser almacenados en un dispositivo portátil.

45 El documento WO 02/11123 revela un método para reconocer una muestra de audio, localizando el fichero de audio que más estrechamente coincida con la muestra de audio, desde una base de datos que indiza un gran conjunto de grabaciones originales. Cada fichero de audio indizado está representado en el índice de la base de datos por un conjunto de instantes de hito y las huellas asociadas. Los hitos ocurren en ubicaciones reproducibles dentro del fichero, mientras que las huellas representan características de la señal en, o cerca de, los instantes linderos. Para efectuar el reconocimiento, los hitos linderos y las huellas se calculan para la muestra desconocida y se usan para extraer huellas coincidentes desde la base de datos. Para cada fichero que contiene huellas coincidentes, los hitos linderos son comparados con los hitos linderos de la muestra en la cual se calcularon las mismas huellas. Si un gran número de correspondientes hitos linderos están linealmente relacionados, es decir, si huellas equivalentes de la muestra y del fichero extraído tienen la misma evolución temporal, entonces el fichero es identificado con la muestra.

55 El método puede ser usado para cualquier tipo de sonido o música, y es particularmente efectivo para señales de audio sometidas a distorsión lineal y no lineal, tales como el ruido de fondo, las distorsiones de compresión o las caídas de transmisión. La muestra puede ser identificada en un tiempo proporcional al logaritmo del número de entradas en la base de datos; dada la suficiente potencia de cálculo, el reconocimiento puede ser realizado en tiempo casi real según se está muestreando el sonido.

60 El documento US 2006/2229878 revela un método para extraer huellas de ondas (por ejemplo, señales musicales), junto con un aparato ejemplar para hacer lo mismo. El método se basa en el principio de calcular características en base a cambios de frecuencia en las bandas a lo largo del tiempo, además de cambios de amplitud en las bandas a lo largo del tiempo.

65 Sin embargo, persiste la necesidad de un aparato, un sistema y un método para la identificación más eficaz y más

fiable del contenido de medios de audio.

Resumen

5 Los aspectos de la invención están definidos en las reivindicaciones.

En un aspecto definido en la Reivindicación 1, se proporciona un método de reconocimiento automático para el reconocimiento automático de contenidos de medios de origen, a partir de una señal de origen, por comparación con contenidos de medios de referencia, incluyendo el método: generar un espectrograma a partir de la señal de origen, aplicando una transformación de Fourier a la señal de origen, incluyendo el espectrograma una pluralidad de columnas, siendo cada columna representativa de una tajada temporal, e incluyendo una pluralidad de depósitos de frecuencias, cada uno representativo de una respectiva gama de componentes de frecuencia para la tajada temporal de la señal de origen; generar al menos un vector de origen para una tajada temporal de la señal de origen, calculando razones de magnitudes de depósitos de frecuencia seleccionados desde la columna para la tajada temporal y cuantizando las razones para generar dígitos de un vector de origen, en donde la generación de al menos un vector para una tajada temporal incluye, para al menos un depósito de frecuencia seleccionado de una tajada temporal, calcular razones de ese depósito y de un depósito de frecuencia adyacente, o casi adyacente, a partir de la columna para la tajada temporal, y dividir las razones en gamas para generar al menos un dígito seleccionado para cada razón; realizar una evaluación primaria por correlación exacta de dígitos de primeros vectores con entradas en una tabla de consulta, en donde cada entrada en la tabla de consulta está asociada a un grupo de segundos vectores y en donde el número de dígitos de los primeros vectores usados para realizar la correlación exacta difieren entre las entradas en la tabla de consulta; y realizar una evaluación secundaria para determinar un grado de similitud entre los primeros vectores y cada uno del grupo de segundos vectores, para identificar segundos vectores cualesquiera que sean candidatos para correlacionar el contenido de medios de origen con el contenido de medios de referencia, en donde una base de datos almacena la tabla de consulta y los segundos vectores, y en donde los primeros vectores son vectores de origen o bien vectores de referencia, y los segundos vectores son los otros vectores de origen y vectores de referencia, representando cada vector de referencia una tajada temporal del contenido de medios de referencia.

En otro aspecto definido en la Reivindicación 6, se proporciona un método de reconocimiento automático para el reconocimiento automático de contenidos de medios de origen a partir de una señal de origen, por comparación con contenidos de medios de referencia, incluyendo el método: generar un espectrograma a partir de la señal de origen, aplicando una transformación de Fourier a la señal de origen, incluyendo el espectrograma una pluralidad de columnas, siendo cada columna representativa de una tajada temporal e incluyendo una pluralidad de depósitos de frecuencia, cada uno representativo de una respectiva gama de componentes de frecuencia para la tajada temporal de la señal de origen; generar al menos un vector de origen para una tajada temporal de la señal de origen, calculando razones de magnitudes de depósitos de frecuencia seleccionados a partir de la columna para la tajada temporal y cuantizando las razones para generar dígitos de un vector de origen; realizar una evaluación primaria por correlación exacta de dígitos de primeros vectores con entradas en una tabla de consulta, en donde cada entrada en la tabla de consulta está asociada a un grupo de segundos vectores, y en donde el número de dígitos de los primeros vectores usados para realizar la correlación exacta difiere entre las entradas en la tabla de consulta; y realizar una evaluación secundaria para determinar un grado de similitud entre los primeros vectores y cada uno del grupo de segundos vectores, para identificar segundos vectores cualesquiera que sean candidatos para correlacionar el contenido de medios de origen con el contenido de medios de referencia, en donde una base de datos almacena la tabla de consulta y los segundos vectores, y en donde los primeros vectores son vectores de origen o bien vectores de referencia, y los segundos vectores son los otros vectores de origen y vectores de referencia, representando cada vector de referencia una tajada temporal del contenido de medios de referencia; incluyendo adicionalmente el método generar un vector de origen adicional para una tajada temporal: generando un espectrograma adicional a partir de la primera señal, aplicando una transformación de Fourier a la señal de origen, incluyendo el espectrograma adicional una pluralidad de columnas, siendo cada columna representativa de una tajada temporal, e incluyendo una pluralidad de depósitos de frecuencia, cada uno representativo de una respectiva gama de componentes de frecuencia para la tajada temporal de la primera señal; aplicando una transformación adicional de Fourier a los respectivos depósitos de frecuencia provenientes de la columna para la tajada temporal, para generar un respectivo conjunto de coeficientes; generando el vector de origen adicional de modo que, para un conjunto de N coeficientes en una columna para una tajada temporal, para cada uno de los elementos 2 a N-1 del vector de origen adicional, sea formado un n-ésimo elemento por el cuadrado del n-ésimo coeficiente dividido entre el producto del (n-1)-ésimo coeficiente y el (n+1)-ésimo coeficiente, y cuantizando los elementos del vector resultante para generar al menos un dígito para cada elemento.

En un aspecto adicional definido en la Reivindicación 18, se proporciona un aparato para proporcionar el reconocimiento automático del contenido de medios de origen, a partir de una señal de origen, por la comparación con contenido de medios de referencia, incluyendo el aparato: un generador de espectrogramas operable para generar un espectrograma a partir de la señal de origen, aplicando una transformación de Fourier a la señal de origen, incluyendo el espectrograma una pluralidad de columnas, siendo cada columna representativa de una tajada temporal, e incluyendo una pluralidad de depósitos de frecuencia, cada uno representativo de una respectiva gama de componentes de frecuencia para la tajada temporal de la señal de origen; un generador de vectores operable

para generar al menos un vector de origen para una tajada temporal de la señal de origen, calculando razones de magnitudes de depósitos de frecuencia seleccionados a partir de la columna para la tajada temporal, y para cuantizar las razones para generar dígitos de un vector de origen, en donde, para generar al menos un vector para una tajada temporal, el generador de vectores es operable, para al menos un depósito de frecuencia seleccionado de una tajada temporal, para calcular una razón de magnitudes de ese depósito y de un depósito de frecuencia adyacente, o uno casi adyacente, a partir de la columna para la tajada temporal, y para dividir las razones en gamas, para generar al menos un dígito seleccionado para cada razón; un evaluador primario, operable para realizar una evaluación primaria, realizando una correlación exacta de dígitos de primeros vectores con entradas en una tabla de consulta, en donde cada entrada en la tabla de consulta está asociada a un grupo de segundos vectores, y en donde el número de dígitos de los primeros vectores usados para realizar la correlación exacta difiere entre las entradas en la tabla de consulta; un evaluador secundario operable para realizar una evaluación secundaria, para determinar un grado de similitud entre los primeros vectores y cada uno del grupo de segundos vectores, para identificar segundos vectores cualesquiera que sean candidatos para correlacionar el contenido de medios de origen con el contenido de medios de referencia; y una base de datos que comprende la tabla de consulta y los segundos vectores, en donde los primeros vectores son vectores de origen o bien vectores de referencia, y los segundos vectores son los otros vectores de origen y vectores de referencia, representando cada vector de referencia una tajada temporal del contenido de medios de referencia.

Para generar al menos un vector para una tajada temporal, el generador de vectores puede ser operable: para al menos un depósito de frecuencia seleccionado de una tajada temporal, para calcular razones de ese depósito y de un depósito de frecuencia adyacente, o casi adyacente, a partir de la columna para la tajada temporal; y para dividir las razones en gamas, para generar dos dígitos binarios para cada razón. Las gamas pueden diferir entre razones seleccionadas, para proporcionar una distribución esencialmente igual de valores de razones entre las gamas. El generador de vectores puede ser operable para generar uno de dichos vectores de origen usando depósitos de frecuencia seleccionados entre una banda de frecuencia desde 400 Hz a 1.100 Hz, y / o usando depósitos de frecuencia seleccionados entre una banda de frecuencia desde 1.100 Hz a 3.000 Hz. Para generar un vector de origen adicional para una tajada temporal: el generador de espectrogramas es operable para generar un espectrograma adicional aplicando una transformación de Fourier a la señal de origen, incluyendo el espectrograma adicional una pluralidad de columnas, siendo cada columna representativa de una tajada temporal, e incluyendo una pluralidad de depósitos de frecuencia, cada uno representativo de una respectiva gama de componentes de frecuencia para la tajada temporal de la señal de origen, y para aplicar una transformación adicional de Fourier a los respectivos depósitos de frecuencia a partir de la columna para la tajada temporal, para generar un respectivo conjunto de coeficientes; y el generador de vectores es operable para generar el vector de origen adicional de modo que, para un conjunto de N coeficientes en una columna para una tajada temporal, para cada uno de los elementos 2 a N-1 del vector de origen adicional, sea formado un n-ésimo elemento por el cuadrado del n-ésimo coeficiente, dividido entre el producto del (n-1)-ésimo coeficiente y el (n+1)-ésimo coeficiente; y para cuantizar los elementos del vector resultante, para generar al menos un dígito para cada elemento.

En un aspecto adicional más, definido en la Reivindicación 19, se proporciona un aparato para proporcionar el reconocimiento automático de contenido de medios de origen a partir de una señal de origen, por comparación con contenidos de medios de referencia, incluyendo el aparato: un generador de espectrogramas operable para generar un espectrograma a partir de la señal de origen, aplicando una transformación de Fourier a la señal de origen, incluyendo el espectrograma una pluralidad de columnas, siendo cada columna representativa de una tajada temporal, e incluyendo una pluralidad de depósitos de frecuencia, cada uno representativo de una respectiva gama de componentes de frecuencia para la tajada temporal de la señal de origen; un generador de vectores operable para generar al menos un vector de origen para una tajada temporal de la señal de origen, calculando razones de magnitudes de depósitos de frecuencia seleccionados a partir de la columna para la tajada temporal, y para cuantizar las razones, para generar dígitos de un vector de origen; un evaluador primario operable para realizar una evaluación primaria, realizando una correlación exacta de dígitos de primeros vectores con entradas en una tabla de consulta, en donde cada entrada en la tabla de consulta está asociada a un grupo de segundos vectores, y en donde el número de dígitos de los primeros vectores usados para realizar la correlación exacta difiere entre las entradas en la tabla de consulta; un evaluador secundario operable para realizar una evaluación secundaria, para determinar un grado de similitud entre los primeros vectores y cada uno del grupo de segundos vectores, para identificar segundos vectores cualesquiera que sean candidatos para correlacionar el contenido de medios de origen con el contenido de medios de referencia; y una base de datos que comprende la tabla de consulta y los segundos vectores, en donde los primeros vectores son vectores de origen o bien vectores de referencia, y los segundos vectores son los otros vectores de origen y vectores de referencia, representando cada vector de referencia una tajada temporal del contenido de medios de referencia; en donde, para generar un vector de origen adicional para una tajada temporal: el generador de espectrogramas es operable para generar un espectrograma adicional aplicando una transformación de Fourier a la señal de origen, incluyendo el espectrograma adicional una pluralidad de columnas, siendo cada columna representativa de una tajada temporal, e incluyendo una pluralidad de depósitos de frecuencia, cada uno representativo de una respectiva gama de componentes de frecuencia para la tajada temporal de la señal de origen, y para aplicar una transformación adicional de Fourier a los respectivos depósitos de frecuencia a partir de la columna para la tajada temporal, para generar un respectivo conjunto de coeficientes; y el generador de vectores es operable para generar el vector de origen adicional, de modo que, para un conjunto de N coeficientes en una columna para una tajada temporal, para cada uno de los elementos 2 a N-1 del vector de origen adicional, sea

formado un n-ésimo elemento por el cuadrado del n-ésimo coeficiente, dividido entre el producto del (n-1)-ésimo coeficiente y el (n+1)-ésimo coeficiente; y para cuantizar los elementos del vector resultante, para generar al menos un dígito para cada elemento.

5 La señal de origen puede ser una señal de audio y las frecuencias de los depósitos del espectrograma pueden ser adjudicadas de acuerdo a una escala logarítmica. La tabla de consulta puede estar organizada como un árbol de profundidad variable que lleva a las hojas, estando la tabla indizada por un primer vector, formando cada hoja una entrada en la tabla de consulta, asociada a un respectivo grupo de segundos vectores, estando el número de dígitos que llevan a cada hoja determinado para proporcionar grupos, de tamaño esencialmente igual, de segundos vectores para cada hoja. El número de dígitos que llevan a cada hoja puede formar el número de dígitos del primer vector usado para realizar la correlación exacta para una hoja dada. Cada hoja de la tabla de consulta puede identificar un grupo de segundos vectores con d dígitos coincidentes, en donde d corresponde a la profundidad del árbol hasta esa hoja. El evaluador secundario puede ser operable para realizar la evaluación secundaria, usando una métrica de distancia para determinar el grado de similitud entre el primer vector y cada uno del grupo de segundos vectores. Puede proporcionarse un evaluador terciario para realizar una evaluación terciaria para cualquier segundo vector identificado como un candidato, siendo el evaluador terciario operable para determinar un grado de similitud entre uno o más primeros vectores adicionales y uno o más segundos vectores adicionales, correspondientes al segundo vector candidato identificado en la evaluación secundaria. Los primeros vectores adicionales y los segundos vectores adicionales pueden estar separados en el tiempo, respectivamente, del primer vector y del segundo vector candidato. La señal de origen puede ser una señal de programa recibida. Un generador de registros puede ser operable para generar un registro del contenido de medios correlacionado de la señal de programa. Un generador de hojas de indicaciones puede ser operable para generar una hoja de indicaciones que identifica el contenido de medios correlacionado. Los segundos vectores pueden ser los vectores de origen y el aparato puede ser configurado para generar la base de datos a partir de los vectores de origen. El aparato puede incluir al menos un procesador, y almacenamiento y software de ordenador operable para implementar el generador de espectrogramas, el generador de vectores de troceo y los evaluadores.

En una realización ejemplar, se proporciona el reconocimiento automático del contenido de medios de muestra. Un espectrograma es generado para sucesivas tajadas temporales de una señal de audio. Uno o más vectores de muestra son generados para una tajada temporal, calculando razones de magnitudes de respectivos depósitos de frecuencia a partir de una columna para la tajada temporal. En una etapa de evaluación primaria (etapa de prueba primaria) se realiza una correlación exacta de bits del vector de muestra con entradas en una tabla de troceo, para identificar un grupo de uno o más vectores de referencia. En una etapa de evaluación secundaria (etapa de prueba secundaria) se realiza un grado de similitud entre el vector de muestra y cada uno del grupo de vectores de referencia, para identificar vectores de referencia cualesquiera que sean candidatos para correlacionar el contenido de medios de muestra, representando cada vector de referencia una tajada temporal del contenido de medios de referencia. Los vectores también pueden ser diversamente descritos como "troceos", "vectores de troceo", "firmas" o "huellas".

40 Una realización de la invención puede proporcionar ajustabilidad a escala y eficacia de operación. Una realización de la invención puede funcionar eficazmente y fiablemente con una base de datos muy grande de pistas de referencia.

45 Una realización de la invención puede emplear troceos con buena potencia discriminatoria (un montón de 'entropía') de modo que un troceo generado a partir de audio de programa tienda a no coincidir con demasiados troceos en la base de datos. Una realización de la invención puede emplear un gran número de mediciones del espectro de la señal de audio. Cada medición puede ser en forma de un número binario de 2 bits, por ejemplo, que sea relativamente robusto ante las distorsiones. Conjuntos de troceos espectrales pueden ser generados a partir de estas mediciones, que dependen de partes restringidas del espectro.

50 Una realización de la invención usa un método que combina una búsqueda en base de datos de correlación exacta, en una etapa primaria, con etapas de refinación que usan información adicional almacenada en una estructura de árbol de profundidad variable. Esto produce un efecto similar al de una búsqueda de vecindad cercana, pero logra aumentos en la velocidad de procesamiento, en varios órdenes de magnitud sobre una búsqueda convencional de vecindad cercana. Las búsquedas de correlación exacta pueden ser llevadas a cabo eficazmente en un ordenador, y permiten que se realice un reconocimiento más rápido. Una realización permite el reconocimiento preciso en entornos distorsionados, cuando se usan muy grandes bases de datos de huellas de origen, con requisitos de procesamiento reducido en comparación con enfoques anteriores. Una realización permite que sea creada una firma (o huella) correspondiente a un momento en el tiempo, de tal manera que la entropía de la parte de la firma que participa en una correlación exacta sencilla esté cuidadosamente controlada, en lugar de usar una correlación aproximada sin tal control cuidadoso de la entropía de la firma. Esto puede permitir precisión y ajustabilidad a escala a un coste de procesador muy reducido.

65 En lugar de tomar un gran número de mediciones de un espectrograma, una realización ejemplar tiene en cuenta las distintas potencias de diversos troceos, variando el número de bits del troceo que se requieren para correlacionar exactamente. Por ejemplo, solamente los primeros 27 bits de un troceo potente pueden ser correlacionados

exactamente, mientras que un número mayor, por ejemplo, los primeros 34 bits, pueden ser correlacionados para un troceo más débil. Una realización de la invención puede usar una estructura de árbol de profundidad variable para permitir que estas operaciones de correlación sean llevadas a cabo eficazmente.

5 Una realización ejemplar puede proporcionar el reconocimiento preciso en entornos ruidosos, y puede hacerlo incluso si el audio a reconocer es de muy breve duración (por ejemplo, menos de tres segundos, o menos de dos segundos o menos de un segundo). Una realización ejemplar puede proporcionar el reconocimiento ante una muy grande base de datos, origen de contenido dotado de huellas (por ejemplo, por encima de un millón de canciones).
10 Una realización ejemplar puede ser implementada en un ordenador autónomo convencional, o en un sistema de ordenadores en red. Una realización ejemplar puede mejorar significativamente la calidad de los resultados de los sistemas de reconocimiento existentes, y mejorar los costes de las implementaciones a gran escala de tales sistemas.

Breve descripción de los dibujos

15 Las realizaciones se describen a continuación en la presente memoria, a modo de ejemplo solamente, con referencia a los dibujos adjuntos.

20 La figura 1 es un diagrama de bloques esquemáticos de un aparato ejemplar.

La figura 2 es un diagrama de flujo que da un panorama general de un método de procesamiento de señales de audio.

25 La figura 3 es una representación esquemática que ilustra un ejemplo del establecimiento de niveles de cuantización en distintas frecuencias.

La figura 4 ilustra una distribución ejemplar de distancias entre vectores de prueba.

30 La figura 5 es una representación esquemática de un sistema de ordenador para implementar una realización del método de la figura 2.

La figura 6 ilustra una estructura de base de datos del sistema de ordenador de la figura 5 en más detalle.

Descripción detallada

35 Una realización ejemplar de la invención proporciona un sistema de reconocimiento de audio que procesa un flujo de audio entrante (un 'programa') y busca en una base interna de datos de música y efectos de sonido ('pistas') para identificar los usos de esas pistas dentro del programa. Un ejemplo de una salida de una realización ejemplar puede ser en forma de una hoja de indicaciones que enumera las secciones de las pistas usadas y dónde aparecen en el programa.
40

Una realización ejemplar puede funcionar con una base de datos, por ejemplo, de diez millones de segundos de música. Sin embargo, otras realizaciones son ajustables a escala para funcionar con una base de datos mucho mayor, por ejemplo, una base de datos de un millardo de segundos de música, y son capaces de reconocer fragmentos de una duración del orden, por ejemplo, de tres segundos o menos, por ejemplo, un segundo, y pueden operar a una velocidad de alrededor de diez veces superior al tiempo real en un ordenador servidor convencional al procesar audio de una típica estación de radio musical.
45

Las siguientes son definiciones de algunos de los términos usados en este documento:

50 Una 'pista' es un fragmento de audio a reconocer en algún momento posterior. Todas las pistas disponibles son procesadas y combinadas en una base de datos.

55 Un "programa" es un trozo de audio por reconocer. Se supone que un programa incluye algunas pistas unidas entre sí y sometidas a diversas distorsiones, intercaladas con otro material.

Una "distorsión" es algo que le ocurre a una pista que compone un programa. Son ejemplos de distorsiones:

- 60 • Ruido: la mezcla de ruido aleatorio con la pista;
- Voz sobre-grabada: la mezcla del habla con la pista;
- Tono: el cambio de tono mientras se mantiene la temporización subyacente;
- 65 • Ritmo: el cambio de ritmo mientras se mantiene el tono;

- Velocidad: el cambio tanto de tono como de ritmo (por ejemplo, reproduciendo una cinta más rápidamente).

Ha de observarse que el tono, el ritmo y la velocidad están relacionados y que dos cualesquiera pueden ser combinados para producir el tercero.

5 Un "troceo" es un trozo pequeño de información, obtenido de una parte específica (tajada temporal) de una pista o programa que, idealmente, no está modificado por la distorsión.

10 La figura 1 es un diagrama de bloques esquemáticos de un ejemplo de un aparato 110 que conforma una realización de la presente invención.

15 Un origen 102 de señal puede ser en forma, por ejemplo, de un micrófono, una radio o un receptor de programas de Internet, o similares, para recibir un programa de medios, por ejemplo, un programa de audio, y para proporcionar una señal 104 de origen.

20 Un generador 112 de espectrogramas puede ser operable para generar un espectrograma a partir de la señal 104 de origen, aplicando una transformación de Fourier a la señal de origen, incluyendo el espectrograma una pluralidad de columnas, siendo cada columna representativa de una tajada temporal, e incluyendo una pluralidad de depósitos de frecuencia, cada uno representativo de una respectiva gama de componentes de frecuencia para la tajada temporal de la señal de origen.

25 Un generador 114 de vectores puede ser operable para generar al menos un vector de origen para una tajada temporal de la señal de origen, calculando razones de magnitudes de los respectivos depósitos de frecuencia a partir de la columna para la tajada temporal, y cuantizando las razones para generar dígitos de un vector de origen.

Una base de datos 46 incluye vectores de referencia, representando cada vector de referencia una tajada temporal de un contenido de medios de referencia.

30 Un evaluador 116 de contenido puede incluir, respectivamente, los evaluadores 118, 120 y 122, primario, secundario y terciario.

35 Un evaluador 118 primario puede ser operable para realizar una evaluación primaria realizando una correlación exacta de dígitos de vectores de origen con entradas en una tabla 66 de consulta de la base de datos 46, en donde cada entrada en la tabla de consulta está asociada a un grupo de vectores de referencia, y en donde el número de dígitos de los vectores de origen usados para realizar la correlación exacta puede diferir entre las entradas en la tabla 66 de consulta. La tabla 66 de consulta puede ser organizada como un árbol de profundidad variable que lleva a las hojas, en donde cada hoja forma una entrada en la tabla de consulta asociada a un respectivo grupo de vectores de referencia. El número de dígitos que llevan a cada hoja puede ser determinado para proporcionar grupos de tamaños esencialmente iguales de vectores de referencia para cada hoja. El número de dígitos que llevan a cada hoja puede formar el número de dígitos del vector de origen usado para realizar la correlación exacta para una hoja dada. Cada hoja de la tabla 66 de consulta puede identificar un grupo de vectores de referencia con d dígitos idénticos, donde d corresponde a la profundidad del árbol hasta esa hoja.

45 Un evaluador 120 secundario puede ser operable para realizar una evaluación secundaria, para determinar un grado de similitud entre un vector de origen y cada uno del grupo de vectores de referencia en la base de datos 46, para identificar vectores de referencia cualesquiera que sean candidatos para correlacionar el contenido de medios de origen con el contenido de medios de referencia. El evaluador 120 secundario puede ser operable para realizar la evaluación secundaria usando una métrica de distancia para determinar el grado de similitud entre el vector de origen y cada uno de los vectores de referencia en el grupo de vectores de referencia.

50 Un evaluador 122 terciario puede ser operable para realizar una evaluación terciaria para cualquier vector de referencia identificado como un candidato. El evaluador 122 terciario puede ser operable para determinar un grado de similitud entre uno o más vectores de origen adicionales y uno o más vectores de referencia adicionales, correspondientes al vector de referencia candidato, identificado en la evaluación secundaria, en donde cada uno de los vectores de origen adicionales y de los vectores de referencia adicionales puede estar separado en el tiempo del vector de origen y del vector de referencia candidato identificado.

60 Un generador 124 de salida puede ser operable para generar un registro de salida, por ejemplo, una hoja de indicaciones, que identifica el contenido de medios correlacionado de la señal de origen.

65 La figura 2 es un diagrama 10 de flujo que presenta un panorama general de las etapas de un método de una realización ejemplar de la invención. El aparato de la figura 1 y el método de la figura 2 pueden ser implementados por uno o más sistemas de ordenador y por uno o más productos de programa de ordenador, funcionando en uno o más sistemas de ordenador. El producto, o los productos, de programa de ordenador puede(n) ser almacenado(s) en medios adecuados cualesquiera legibles por ordenador, por ejemplo, discos de ordenador, cintas, almacenamiento de estado sólido, etc. En diversos ejemplos, varias de las etapas del proceso pueden ser realizadas

por programas individuales de ordenador y / o sistemas individuales de ordenador. Por ejemplo, la generación de un espectrograma, según lo descrito más adelante, puede ser realizada por un programa de ordenador y / o un sistema de ordenador, independientes de uno o más programas de ordenador y / o sistemas de ordenador usados para realizar la generación de troceos y / o la prueba de bases de datos y / o la generación de hoja de indicaciones.

5 Además, una o más de las partes del aparato de la figura 1 o del proceso de la figura 2 pueden ser implementadas usando hardware de propósito especial, por ejemplo, circuitos integrados de propósito especial, configurados para proporcionar la funcionalidad descrita en más detalle en la siguiente descripción.

10 Sin embargo, solamente por motivos de facilidad de explicación, se supone que los procesos descritos en lo que sigue con referencia a la figura 2, procesos estos que incluyen la generación 12 de espectros, la generación 14 de vectores, la evaluación 16 de señales (incluso las etapas 18, 20 y 22 primaria, secundaria y terciaria) y la generación 24 de salidas, son realizados por un aparato que comprende un sistema servidor de ordenador, que incluye uno o más procesadores y almacenamiento, y que está controlado por uno o más programas. Las etapas de proceso descritas más adelante, incluyendo la generación 12 de espectros, la generación 14 de vectores, la evaluación 16 de contenidos (incluyendo las etapas 18, 20 y 22 primaria, secundaria y terciaria) y la generación 24 de salidas, también corresponden, respectivamente, a las funciones realizadas por el generador 112 de espectros, el generador 114 de vectores, el evaluador 116 de contenidos (incluyendo los de los evaluadores 118, 120 y 122 primario, secundario y terciario) y el generador 124 de salidas de la figura 1.

20 Generación 12 de espectros

En este ejemplo, una señal de origen, en forma de una señal de audio, es procesada para generar un espectrograma, por ejemplo, aplicando una Transformación Rápida de Fourier (FFT) a la señal de audio.

25 En una realización ejemplar, la señal de audio debería ser formateada de manera congruente con un método de generación de la base de datos con respecto a la cual ha de ser comparada la señal de audio. En una realización ejemplar, la señal de audio puede ser convertida a un formato plano .WAV, muestreado, por ejemplo, a 12 kHz, en estéreo si es posible, o en mono si no lo es, y con, por ejemplo, 16 bits por muestra. En una realización ejemplar, el audio en estéreo que comprende un canal izquierdo y un canal derecho está representado como canales de suma (izquierdo más derecho) y de diferencia (izquierdo menos derecho), a fin de brindar mayor resistencia a la voz sobregrabada y distorsiones similares. El fichero de audio es luego procesado para generar un espectrograma.

30 Los parámetros aplicados al espectrograma están basados, a grandes rasgos, en la percepción del sonido del oído humano, dado que la clase de distorsiones por las que es probable que pase el sonido son aquellas que preservan la percepción de un ser humano. El espectrograma incluye una serie de columnas de información para sucesivos intervalos de muestreo (tajadas temporales). Cada tajada temporal corresponde, por ejemplo, a un valor entre 1 y 50 ms (por ejemplo, aproximadamente 20 ms). Los sucesivos segmentos pueden solaparse en una proporción significativa de su longitud, por ejemplo, entre el 90 y el 99%, por ejemplo, alrededor del 97% de su longitud. Como resultado, el carácter del sonido tiende a cambiar solo lentamente entre segmento y segmento. Una columna para una tajada temporal puede incluir una pluralidad de depósitos de frecuencia dispuestos en una escala logarítmica, teniendo cada depósito, por ejemplo, un ancho de aproximadamente un semitono.

40 Puede proporcionarse un número significativo de depósitos de frecuencia para cada tajada temporal, o columna, del espectro. Por ejemplo, pueden generarse del orden de entre 40 y cien o más depósitos de frecuencia. En un ejemplo específico, se proporcionan 92 depósitos de frecuencia.

45 Generación 14 de vectores

50 Una segunda etapa 14 es la generación de uno o más vectores de troceo, o troceos. En una realización ejemplar, se genera un cierto número de distintos tipos de troceos. Una o más secuencias de vectores de baja dimensión, que forman los troceos (o 'huellas', 'firmas'), son diseñadas para que sean robustas ante los diversos tipos de distorsiones que puedan encontrarse.

55 En una realización ejemplar, a fin de dar resistencia al ruido añadido y señales similares, los valores medidos pueden ser cuantizados groseramente antes de generar un troceo. Hay conflicto entre el deseo de cuantizar groseramente y la necesidad de obtener entropía suficiente desde el audio de origen. A fin de realizar la entropía obtenida, la cuantización puede ser realizada no linealmente, de modo que, para cualquier medición dada, los valores cuantizados tiendan a ser igualmente probables, haciendo más uniforme la distribución de troceos, según se muestra en la figura 3. Los umbrales de cuantización pueden ser seleccionados independientemente en cada frecuencia, para hacer más uniforme la distribución de troceos. Para maximizar la robustez, cada medición puede ser seleccionada para depender solamente de dos puntos en el espectrograma.

60 En una realización ejemplar, un troceo básico es obtenido a partir de una única columna del espectrograma, calculando la razón de las magnitudes de depósitos de frecuencia adyacentes, o casi adyacentes. En un ejemplo, un vector puede ser generado determinando una razón del contenido de depósitos de frecuencia adyacentes en la columna, y dividiendo la razón en una entre cuatro gamas.

Por ejemplo, para cada uno de los depósitos 0 a 91, determinar una razón como:

- valor del depósito i / valor del depósito $i+1$

5

- y determinar dentro de cuál de las cuatro gamas 00, 01, 10 y 11 cae la razón.

En términos simplistas, consideremos que la gama 00 corresponde a las razones entre 0 y 0,5, la gama 01 corresponde a las razones entre 0,5 y 1, la gama 10 corresponde a las razones entre 1 y 5 y la gama 11 corresponde a las razones entre 5 y el infinito. Puede verse, por lo tanto, que, para cada par de depósitos comparados, puede ser generado un número de dos bits. En otro ejemplo, puede usarse un número distinto de gamas para generar un número distinto de bits, o uno o más dígitos, de acuerdo a una base distinta.

10

Un vector de ese tipo puede ser esencialmente invariable con respecto a cambios globales de amplitud en la señal original, y robusto con respecto a la igualación (impulso o corte de frecuencias altas o bajas). Las gamas 00, 01, 10 y 11 pueden ser distintas para cada depósito, y pueden ser obtenidas empíricamente recogiendo valores de las razones procedentes de un conjunto de prueba de audio, y dividiendo la distribución resultante en cuatro partes iguales.

15

En una realización ejemplar, se generan luego dos troceos. Un troceo es generado usando una banda de frecuencia entre alrededor de 400 Hz y alrededor de 1.100 Hz (un 'troceo de tipo 0') y el otro, usando una banda de frecuencia entre alrededor de 1.100 Hz y alrededor de 3.000 Hz (un 'troceo de tipo 1'). Estas bandas de frecuencia relativamente alta son más robustas ante la distorsión producida por el agregado de una voz sobre-grabada a una pista.

20

25

En una realización ejemplar, se genera un tipo adicional de troceo ('troceo de tipo 2') que está diseñado para que sea robusto ante la variación del tono (tal como ocurre cuando una secuencia de muestras de audio es reproducida de nuevo, más rápido o más despacio que la velocidad nominal de muestreo). Se genera un conjunto similar de depósitos de espectrograma de frecuencia logarítmica para el troceo básico. Se toma la amplitud de cada depósito de espectrograma y se aplica una segunda transformación de Fourier. Este enfoque genera un conjunto de coeficientes semejante a un 'cepstro de frecuencia logarítmica'. Un desplazamiento del tono en el audio original corresponderá a una traslación en la columna del espectrograma de frecuencia logarítmica y, por tanto (ignorando los efectos marginales), a un desplazamiento de fase en los coeficientes resultantes. Los coeficientes resultantes son luego procesados para formar un nuevo vector, cuyo n -ésimo elemento se obtiene tomando el cuadrado del n -ésimo coeficiente dividido entre el producto de los coeficientes $(n-1)$ -ésimo y $(n+1)$ -ésimo. Esta cantidad es invariable con respecto al desplazamiento de fase en los coeficientes y, por tanto, también al desplazamiento del tono en la señal original. También es invariable ante el cambio de volumen en la señal original.

30

35

Según los segmentos sucesivos se solapan en una proporción significativa de su longitud, el carácter del sonido tiende a cambiar solo lentamente entre segmento y segmento, por lo cual los troceos tienden a cambiar solamente en uno o dos bits, o dígitos, entre segmento y segmento.

40

Como todos estos troceos solamente inspeccionan una columna del espectrograma, en principio son invariables ante la variación de ritmo (dilatación temporal o compresión sin desplazamiento de tono). Dado que puede hallarse que algunos algoritmos de cambio de ritmo provocan alguna distorsión de los componentes de audio de menor frecuencia, los troceos basados en componentes de mayor frecuencia, según lo descrito anteriormente, son más robustos.

45

Una realización ejemplar puede proporcionar robustez con respecto a la voz sobre-grabada en audio de programas. El efecto general del agregado de voz sobre-grabada a una pista es el cambio de un espectrograma en áreas que tienden a estar localizadas en el tiempo y en la frecuencia. El uso de troceos que dependen solamente de una única columna del espectrograma, lo que corresponde a una sección muy breve del audio, brinda robustez con respecto a la voz sobre-grabada. Esto da una buena oportunidad de reconocimiento de una pista si la voz sobre-grabada se interrumpe aunque sea brevemente (tal vez, incluso en el medio de una palabra). El uso de troceos que están localizados, al menos parcialmente, en la frecuencia también ayuda a mejorar la resistencia a la voz sobre-grabada, así como ante otras ciertas clases de distorsión.

50

55

Además, el hecho de que cada troceo dependa solamente de una sección muy breve de audio da el potencial para reconocer secciones muy breves de una pista.

60

La resistencia a una transposición en el tono (con o sin un cambio de ritmo que acompañe) puede ser lograda generando troceos en base a un cálculo de cepstro modificado.

Etapas de prueba (Evaluación de contenido) 16

65

En una realización ejemplar, el audio del programa es reconocido luego comparando los troceos con troceos pre-

calculados de las pistas en una base de datos. El objetivo del proceso de consulta es realizar una consulta aproximada, o búsqueda del 'vecino más cercano', sobre la totalidad de la base de datos de música, por ejemplo, usando el vector obtenido de una columna del espectrograma. Esta es una búsqueda de alta dimensión, con un gran número de posibles objetos de destino obtenidos de la base de datos de música.

5 En una realización ejemplar, esto se hace como un proceso 16 de prueba de múltiples etapas.

Etapa de prueba primaria (Evaluación primaria) 18

10 Una etapa 18 de prueba primaria se realiza usando una consulta de correlación exacta. En una realización ejemplar, esto se efectúa con los troceos como un vector binario simple con un pequeño número de bits, para realizar una consulta en una tabla de troceo. Como resultado del uso de un número pequeño de bits, cada consulta devuelve habitualmente un gran número de aciertos en la base de datos. Por motivos que quedarán claros más tarde, el conjunto de aciertos en la base de datos, extraídos en respuesta a la consulta primaria para una clave dada, se denomina una 'hoja'.

15 En la práctica, los bits que son extraídos del espectrograma para construir la clave no son independientes y no es igualmente probable que sean '0' o '1'. En otras palabras, la entropía por bit del vector (con respecto a una muestra dada de música) es menor que uno.

20 La entropía por bit para algunas clases de vectores es mayor que la de otros. Otra manera de decirlo es que algunas claves son mucho más comunes que otras. Por lo tanto, si se usa una clave de tamaño fijo para acceder a la base de datos, se hallará a veces un gran número de aciertos y a veces se hallará un pequeño número de aciertos. Si se escoge una clave al azar, la probabilidad de que caiga en una hoja dada es proporcional al número de entradas en esa hoja, y la cantidad de trabajo adicional implicado en la comprobación de cada una de esas entradas para determinar si es efectivamente una buena correlación también es proporcional al número de entradas en esa hoja. Como resultado, la cantidad total de trabajo que se espera hacer para esa clave es entonces proporcional al promedio de los cuadrados de los tamaños de las hojas. A la vista de esto, en una realización, este valor es minimizado (es decir, las prestaciones del sistema son maximizadas) haciendo los tamaños de las hojas tan iguales como sea posible.

25 En una realización, por lo tanto, se escoge una estructura de base de datos que está orientada a la igualación de los tamaños de las hojas.

30 Pueden obtenerse bits de un troceo a partir de funciones continuas del espectrograma, si se desea: por ejemplo, una cantidad continua puede ser cuantizada en uno entre ocho valores distintos, y el resultado codificado en el troceo como tres bits. En tales casos, es ventajoso no usar un esquema uniforme de cuantización, sino escoger en cambio (a partir de un ejemplo basado en el análisis de una gran muestra de música) umbrales de cuantización de modo que cada posible valor cuantizado tienda a tener igual probabilidad de ocurrencia. Los niveles de cuantización usados al crear la base de datos son los mismos que los usados al crear troceos del programa, a consultar en la base de datos.

35 Los bits en el troceo también pueden ser dispuestos de modo que los de más probable robustez (por ejemplo, los bits más significativos de cantidades continuas cuantizadas) estén situados hacia el extremo más significativo del troceo, y los bits menos robustos, hacia el extremo menos significativo del troceo.

40 En una realización, la base de datos está dispuesta en forma de un árbol binario. Una profundidad en el árbol corresponde a la posición de un bit en el troceo. El árbol es recorrido desde abajo hacia arriba, consumiendo un bit del troceo clave (el más significativo, es decir, el más robusto, primero) para determinar si se selecciona el vástago izquierdo o derecho en cada punto, hasta que se halle un nodo terminal (u 'hoja'), digamos, en la profundidad d. La hoja contiene información acerca de aquellas pistas en la base de datos que incluyen un troceo cuyos d bits más significativos coincidan con los del troceo clave.

45 Las hojas están a diversas profundidades, escogiéndose las profundidades de modo que cada una de las hojas del árbol contenga un número de entradas del mismo orden, por ejemplo, aproximadamente el mismo número de entradas. Debería observarse que en otros ejemplos el árbol podría estar basado en otra base numérica que la de un árbol binario (por ejemplo, un árbol terciario).

50 En la etapa de prueba primaria, por lo tanto, se busca una correlación exacta entre los bits seleccionados del troceo del audio del programa y los troceos almacenados para pistas de referencia. El número de dígitos que son correlacionados depende del tamaño de la base de datos y de cuán común sea ese troceo entre las pistas, en general, por lo que menos bits son correlacionados para troceos más raros. El número de bits que son correlacionados puede variar, por ejemplo, entre 10 y alrededor de 30 bits en el caso de un árbol binario, según el tamaño de la base de datos de pistas.

65 Además, como los troceos consecutivos del mismo tipo cambian habitualmente en solamente uno o dos bits,

también pueden ser obtenidas, en general, correlaciones exactas para los bits correlacionados, incluso si los momentos en el programa en los cuales son generados los troceos no están exactamente sincronizados con los momentos para los cuales fueron generados los troceos para la base de datos de pistas de referencia.

5 Etapa de prueba secundaria (Evaluación secundaria) 20

10 En una realización, una etapa 20 de prueba secundaria implica buscar un troceo de programa en la base de datos, por medio de un acceso de fichero aleatorio. Esto captura el contenido de una única hoja, que contiene un gran número, habitualmente unos pocos cientos, por ejemplo, del orden de 200, de correlaciones de troceos. Cada correlación corresponde a un punto en una de las pistas originales, que es superficialmente similar al troceo del programa.

15 Cada una de estas entradas está acompañada por 'información de prueba secundaria', esto es, datos que contienen información adicional obtenida del espectrograma. Los troceos de tipo 0 y de tipo 1 están acompañados por información de espectrograma cuantizada proveniente de aquellas partes del espectrograma no implicadas en la creación del troceo original; los troceos de tipo 2 están acompañados por bits adicionales, obtenidos de los coeficientes de estilo cepstro. Las entradas también incluyen información que permite la ubicación de una pista original correspondiente a un troceo, y la posición en esa pista.

20 El propósito de la prueba secundaria es obtener una idea estadísticamente más potente en cuanto a si coinciden o no las muestras del programa y una entrada de la base de datos, aprovechando el hecho de que esta etapa del proceso ya no está constreñida a la búsqueda de correlaciones exactas. En una realización ejemplar, puede usarse una métrica Manhattan de distancia, o alguna otra métrica de distancia, para determinar un grado de similitud entre dos vectores de información de prueba secundaria.

25 En una realización ejemplar, cada prueba secundaria que se aprueba comporta un acceso aleatorio de fichero adicional a la base de datos, para obtener información para una prueba terciaria, según lo descrito más adelante. Teniendo esto en mente, en una realización ejemplar, se dispone un umbral para aprobar la prueba secundaria, de modo que, en promedio, alrededor de una de las entradas de base de datos en una hoja apruebe la prueba secundaria. En otras palabras, la probabilidad de aprobar una prueba secundaria debería ser más o menos el valor recíproco del tamaño de hoja.

30 La figura 4 ilustra una distribución ejemplar de distancias entre dos vectores de prueba secundaria seleccionados al azar entre una gran base de datos de música, una curva para cada uno de tres tipos de troceo. Se escoge por ello un umbral para un tipo dado de prueba secundaria, escogiendo un punto en la curva adecuada, de modo que el área bajo la cola a la izquierda de ese punto, como fracción del área total bajo la curva, sea aproximadamente igual al valor recíproco del tamaño de hoja.

40 Por tanto, en la etapa de prueba secundaria, cada acierto primario se somete a una 'prueba secundaria' que implica comparar la información de troceo generada a partir del mismo segmento de audio con la pista candidata en el punto de coincidencia.

Etapa de prueba terciaria (Evaluación terciaria) 22

45 Como se ha indicado anteriormente, la información almacenada en la hoja permite la ubicación de una pista original correspondiente al troceo y la posición en esa pista. Cuando una prueba secundaria es aprobada, se capturan datos de prueba terciaria correspondientes a una breve sección de pista alrededor del punto de coincidencia. La información de prueba terciaria incluye una serie de troceos de la pista original. Los troceos del programa son luego comparados con los datos de prueba terciaria. Este proceso no está constreñido a la búsqueda de correlación exacta, por lo que una métrica de distancia, por ejemplo, una métrica Manhattan de distancia, puede ser usada para determinar cuán similares son los troceos del programa a los datos de prueba terciaria. En una realización ejemplar, la métrica implica un cálculo probabilístico completo basado en tablas de probabilidades empíricamente determinadas, para determinar un grado de similitud entre los troceos del programa y los datos de prueba terciaria.

55 Tanto la secuencia de troceos de programa como la secuencia de troceos de prueba terciaria están acompañadas por información de sello temporal. Normalmente, estos deberían alinearse: en otras palabras, los sellos temporales de troceos de programa deberían tener un desplazamiento constante a partir de los sellos temporales coincidentes de pruebas terciarias. Sin embargo, si el programa ha sido extendido en el tiempo (una 'distorsión de ritmo'), este desplazamiento se desalineará gradualmente. Cuando mayor sea la distorsión de ritmo, más rápida será la desalineación. Para detectar esta desalineación, la prueba terciaria puede ser realizada en un cierto número de ritmos distintos de prueba, y puede seleccionarse el mejor resultado como la estimación de ritmo para el valor coincidente. Dado que las distorsiones de ritmo son relativamente raras, en una realización ejemplar, este proceso de selección está sesgado hacia la creencia de que no ha ocurrido ninguna distorsión de ritmo.

65 En la prueba terciaria, se realiza un barrido hacia atrás y hacia adelante a partir del punto de coincidencia, evaluando la similitud entre los troceos de programa y los troceos de prueba terciaria, y usando la estimación de

ritmo para determinar la velocidad relativa a la cual se realiza el barrido en el programa y en los datos de prueba terciaria. Mientras continúe habiendo buenas correlaciones por encima de una cierta tasa, esto se toma como evidencia de que el programa contiene la pista sobre ese periodo. Cuando ya no se ven más buenas correlaciones, esto se toma como evidencia de que se ha hallado el inicio o el final de ese uso de la pista.

5 Es improbable que la estimación inicial de ritmo sea exacta. Durante el barrido, por lo tanto, se prueban troceos de programa levemente por delante, y levemente por detrás, de la posición nominal calculada. Si estos coinciden con la información de prueba terciaria mejor que los troceos en la posición nominal, se aplica una corrección al ritmo estimado. De ese modo se asimila el rastreo de pequeñas magnitudes de desalineación en el ritmo.

10 Dado que los troceos usados en una realización ejemplar dependen de una única columna del espectrograma, son intrínsecamente resistentes a un cambio en el ritmo. Se realza la eficacia en cuanto a que el análisis o la búsqueda con respecto a los cambios de ritmo se pospone hasta la etapa de prueba terciaria, y en esa etapa hay solamente unos pocos candidatos para examinar y por ello es viable, en términos de cálculo, una búsqueda exhaustiva sobre los posibles desplazamientos del ritmo.

15 En consecuencia, en la fase de prueba terciaria se usa una segunda base de datos que puede contener una versión sumamente comprimida de los espectrogramas de las pistas originales. En una realización ejemplar, la base de datos se basa en troceos similares a los de la base de datos primaria, con el agregado de alguna información lateral extra. Estos datos están dispuestos para que sean rápidamente accesibles por pista y por posición dentro de esa pista. El sistema puede ser dispuesto de modo que los índices quepan dentro de una RAM de ordenador. Durante la prueba terciaria, el audio de programa a cada lado de una correlación candidata, que ha aprobado la prueba secundaria, es comparado con respecto a la base de datos, usando un cálculo probabilístico completo. Esta prueba es capaz de rechazar falsos positivos que hayan aprobado la prueba secundaria, y halla simultáneamente los puntos de inicio y de final dentro del programa donde se usa el material de la pista.

20 En resumen, cada troceo que aprueba la prueba secundaria se somete a la prueba terciaria en base a una alineación del material del programa y el material de la pista, implícito por la etapa de prueba secundaria. En la prueba terciaria, esa alineación se extiende hacia atrás y hacia adelante en el tiempo, desde el punto donde ocurrió el acierto primario, comparando el programa y la pista candidata, usando una base de datos que contiene troceos, junto con otra información, para permitir que se haga una comparación precisa. Si la correlación no puede ser extendida satisfactoriamente en cualquier dirección, se descarta; en caso contrario, se informa la gama de tiempos de programa sobre los cuales ha sido hallada una correlación satisfactoria (como un 'punto de entrada' y un 'punto de salida'), junto con la identidad de la pista coincidente y la gama de tiempos de pista que han sido correlacionados.

25 En una realización ejemplar, esto forma una entrada candidata en una hoja de indicaciones de salida.

Etapa 22 de salida

30 Como se ha mencionado anteriormente, una aplicación del proceso de reconocimiento de audio es la generación de una hoja de indicaciones. El resultado de la prueba terciaria es una serie de correlaciones candidatas del material del programa con respecto a pistas en la base de datos original. Cada correlación incluye los puntos de inicio y de final del programa, el número de identificación de la pista, el punto de inicio y de final dentro de la pista y una medida global de la calidad de la correlación. Si la calidad de correlación es suficientemente alta, entonces esta correlación es una candidata para la entrada en la hoja de indicaciones.

35 Cuando se halla una nueva entrada candidata de la hoja de indicaciones, se compara con las entradas que ya están en la hoja de indicaciones. Si no hay un solapamiento significativo en el tiempo del programa con una entrada existente, se añade a la hoja de indicaciones. Si hay un solapamiento significativo con otra entrada, entonces la entrada es desalojada si su calidad de correlación es mayor y, en caso contrario, la candidata será descartada.

40 Cuando todos los troceos de programa han sido procesados, puede emitirse una hoja de indicaciones completada.

45 Como se ha indicado anteriormente, el proceso que ha sido descrito es realizado automáticamente por uno o más programas de ordenador funcionando en uno o más sistemas de ordenador, y puede ser integrado en un único proceso que es realizado en tiempo real, o bien puede ser separado entre uno o más procesos individuales realizados en momentos distintos por uno o más programas de ordenador que funcionan en uno o más sistemas distintos de ordenador. Detalles adicionales del funcionamiento del sistema son descritos en los siguientes pasajes.

50 En el ejemplo presente, se supone que el sistema, según se muestra en la figura 5, es un sistema servidor 30 de ordenador que recibe como entrada un programa 32 de audio y emite una hoja 34 de indicaciones. El sistema de ordenador incluye uno o más procesadores 42, la memoria de acceso aleatorio (RAM) 44 para programas y datos, y una base de datos 46, así como otras características convencionales de un sistema de ordenador, incluyendo interfaces de entrada / salida, fuentes de alimentación, etc., que no se muestran en la figura 5.

55 La base de datos 46 de referencia

La base de datos 46 se construye a partir de una colección de ficheros de música de origen, en un cierto número de etapas.

En una realización ejemplar, la base de datos es generada por los siguientes procesos:

- 5 1. Cada fichero de música de origen es convertido a un formato WAV plano, muestreado, por ejemplo, a 12 kHz, en estéreo si es posible, o en mono si no lo es, por ejemplo, con 16 bits por muestra. El audio en estéreo, que comprende un canal izquierdo y un canal derecho, es convertido en canales de suma (izquierdo más derecho) y de diferencia (izquierdo menos derecho).
- 10 2. Se crea un fichero (por ejemplo, llamado listaorigen) que contiene una lista numerada de los nombres de ficheros de origen. Cada línea del fichero puede contener un único número identificador (un 'Identificador de pista' o 'Identificador de segmento'), seguido por un espacio, seguido por el nombre del fichero.
- 15 3. Los troceos son generados a partir de las pistas de música de origen, para crear un fichero (por ejemplo, llamado infosegbruto) que contiene los troceos de las pistas de origen. Se genera un fichero auxiliar (por ejemplo, llamado infosegbruto.aux) que contiene la información de nombres de pistas, a partir de la listaorigen.
- 20 4. Los troceos son clasificados en orden de Identificador de pista y tiempo.
- 25 5. Se generan los datos de prueba terciaria y se crean índices para ellos, para formar un fichero infosegbruto catalogado.
- 30 6. El fichero infosegbruto catalogado es clasificado por orden ascendente del valor de troceo.
- 35 7. Se genera un primer índice de racimo (véase descripción del formato más adelante).
- 40 8. Se genera un fichero de datos auxiliares (por ejemplo, llamado datosaux), siendo usado el fichero de datos auxiliares para exhibir nombres de fichero en la hoja de indicaciones emitida.
- 45 9. Los diversos ficheros son luego ensamblados en la base de datos.

Para una realización ejemplar del sistema, diseñada para funcionar con una base de datos de diez millones de segundos de audio, diversos parámetros del sistema, a exponer a continuación, se fijan de la siguiente manera.

- Tamaño máximo de hoja = 400

- Primera profundidad de racimo = 20

- 40 Debería observarse, sin embargo, que estos son ejemplos solamente de los parámetros del sistema, y que distintas realizaciones emplearán distintos parámetros. Por ejemplo, para mayores bases de datos la primera profundidad de racimo podría aumentarse, por ejemplo, hasta alrededor de 23 o 24 bits para cien millones de segundos de audio, y alrededor de 26 o 27 bits para un millardo de segundos de audio. En el ejemplo descrito en más detalle a continuación, se supone una primera profundidad de racimo de 24 bits.

- 45 En una realización ejemplar, a fin de mantener manejables los tamaños de ficheros, diversas estructuras de datos usadas son empaquetadas en octetos y bits para su almacenamiento como parte de la base de datos.

Troceo en bruto

- 50 En una realización ejemplar, un troceo en bruto es almacenado como seis octetos, o 48 bits. Los bits más significativos son los usados para la consulta de la base de datos primaria.

Hojas de base de datos e Infosegbruto

- 55 Cada hoja en la base de datos contiene una secuencia de estructura de infosegbruto. Un programa por analizar también es convertido en una secuencia de estructuras de infosegbruto antes de que se hagan consultas en la base de datos.

- 60 Cada estructura de infosegbruto aloja un troceo en bruto, junto con información acerca de su procedencia (su Identificador de pista y su posición dentro de esa pista, almacenados como cuatro octetos para cada uno) y un campo de 16 octetos de información de prueba secundaria.

- 65 Cuando se genera inicialmente, la información de posición se fija para indicar el tiempo del troceo con respecto al inicio de la pista, medido en unidades de aproximadamente 20 milisegundos. Durante el procedimiento de construcción de la base de datos, este valor es reemplazado por un desplazamiento directo en los datos de prueba

terciaria (el infosegbruto 'catalogado').

Las estructuras de datos de infosegbruto son almacenadas secuencialmente, en el orden de troceo, en una estructura de fichero plano llamada el BFF ('gran fichero plano'). Cada hoja es una sub-sección contigua del BFF, que consiste precisamente en aquellas estructuras de datos de infosegbruto cuyos troceos tienen iguales sus primeros d bits ('profundidad'), donde d se escoge en cada caso de modo que el número de estructuras de datos de infosegbruto dentro de la hoja no sea mayor que el parámetro de sistema aplicable del 'máximo tamaño de hoja'. La selección del valor de profundidad puede ser realizada dividiendo primero el BFF en hojas, cada una con el valor de profundidad fijado en el valor del parámetro de sistema 'profundidad de primer racimo'. Luego, cualquier hoja con un valor d de profundidad, cuyo tamaño supere el parámetro de sistema 'máximo tamaño de hoja', puede ser dividida en dos hojas, cada una con un valor de profundidad d más uno; siendo repetido este procedimiento de división hasta que no queden hojas cuyo tamaño supere el parámetro de sistema 'máximo tamaño de hoja'.

La figura 6 es un diagrama esquemático que da un panorama general de la estructura de la base de datos 46 y de las consultas asociadas a cada troceo obtenido a partir del audio del programa.

Hay dos niveles de índice para las hojas de la base de datos.

Como se ha expuesto anteriormente, la base de datos 46 toma la forma de un árbol binario de profundidad no uniforme.

Para simplificar la indización de la base de datos, cada hoja tiene una profundidad de al menos el parámetro 62 de profundidad del primer racimo, digamos, 24 bits. La parte del árbol por encima de un nodo en la profundidad del racimo se conoce como un 'racimo'. Hay 2^F racimos, donde F = la profundidad del primer racimo, y cada uno de estos racimos corresponde a una sección contigua del BFF 74, el cual, a su vez, contiene un cierto número de hojas 72.

Un troceo 60 de programa se muestra en el extremo superior izquierdo de la figura 6. Un cierto número de los bits más significativos (fijado por un parámetro PROFUNDIDADDELPRIMERRACIMO 62) es usado como un desplazamiento en un índice 66 basado en RAM (el 'índice de primer racimo'), que contiene información acerca de la forma de un árbol de profundidad variable. El nivel superior 68 del índice 66 de la base de datos contiene una entrada por racimo. Simplemente apunta a un registro 70 (de longitud variable) en el segundo índice, que contiene información acerca de ese racimo. Se usan bits adicionales del troceo del programa para recorrer los pocos nodos finales del árbol formado por el segundo índice. En el ejemplo ilustrado, se toman tres bits adicionales ('101'). Siguiendo la estructura de árbol mostrada en la figura 6, si el primero de estos bits hubiera sido un cero, se habría tomado un total de solamente dos bits. La información almacenada en el índice de primer racimo basado en RAM es suficiente para hallar directamente el correspondiente registro de base de datos para una hoja 72.

De tal modo, el índice de segundo nivel describe la forma del árbol binario en un racimo y los tamaños de las hojas dentro de él. Una entrada consiste en lo siguiente:

(i) Un desplazamiento dentro del BFF 74, donde comienzan los datos para este racimo.

(ii) una codificación de la forma del árbol binario en el racimo. Este es un flujo de bits con un bit para cada nodo (interior y hoja) del árbol, considerado en el orden encontrado en un recorrido, con la profundidad como primer índice, del árbol. El bit es un cero si el nodo es interior, y 1 si es una hoja. El flujo de bits es rellenado con bits 0 al final del último octeto, si es necesario.

(iii) El tamaño de cada hoja 72 en el racimo, en el orden encontrado en un recorrido, con la profundidad como primer índice, del árbol, codificado en forma comprimida, de modo que la mayoría de los tamaños estén expresados en un único octeto.

En el pequeño número de casos en que un racimo contiene solamente troceos con poca entropía (es decir, en que el racimo es relativamente grande), un valor indicador especial puede reemplazar (ii) y (iii) en lo que antecede, y las correspondientes entradas del BFF no son indizadas.

En una realización ejemplar, ambos niveles del índice 66 / 70 están diseñados para caber en la memoria RAM en el sistema servidor, permitiendo que el contenido de cualquier hoja de la base de datos sea capturado con un único acceso aleatorio al BFF.

En el BFF, junto con cada troceo coincidente, se almacena información adicional obtenida del espectrograma, de manera similar a la descrita anteriormente con respecto a los troceos del programa. Dado que solamente unos pocos cientos de correlaciones han de ser consideradas en la etapa de prueba secundaria, puede usarse una métrica de distancia para determinar si efectivamente hay una buena correlación entre el programa y una pista de referencia identificada en la etapa de prueba primaria. La evaluación de una métrica de ese tipo sobre la base de datos entera sería prohibitivamente cara en tiempo de cálculo. Como se ha indicado anteriormente, el umbral para

esta prueba se fija de modo que solamente apruebe un número muy pequeño de correlaciones potenciales, tal vez tan pequeño como uno o dos.

5 Para aumentar adicionalmente el valor extraído del único acceso aleatorio al disco de la base de datos, la información de prueba secundaria puede ser comprimida usando un algoritmo de compresión adecuado.

10 La información de prueba terciaria consiste en una secuencia de estructuras de datos 76 de prueba terciaria, ordenados por Identificador de pista y desplazamiento temporal dentro de esa pista. Cada una de estas contiene un desplazamiento temporal (en unidades de aproximadamente 20 milisegundos) con respecto a la entrada anterior, almacenado como un único octeto, y un troceo en bruto.

15 La base de datos 46 incluye un índice 78 en los datos 76 de prueba terciaria, que da el punto de partida de cada pista. Este índice está diseñado para que sea lo bastante pequeño para caber en memoria RAM y, por lo tanto, permitir que cualquier elemento deseado de los datos de prueba terciaria sea capturado con un único acceso aleatorio al fichero de la base de datos. Los datos 80 que definen una entrada para el índice 76 de datos de prueba terciaria son proporcionados, con los datos 87 de prueba secundaria, en el BFF 74.

20 A fin de reducir los tiempos de acceso a la base de datos, la base de datos se mantiene, ventajosamente, en discos de estado sólido, en lugar de discos rígidos tradicionales, ya que los tiempos de acceso aleatorio (o 'búsqueda') para un disco de estado sólido son habitualmente del orden de cien veces más rápidos que los de un disco rígido tradicional. Allí donde el tamaño de la base de datos lo permite, toda la información puede ser almacenada en la memoria RAM de un ordenador. Además, según lo indicado, con una estructura de árbol de profundidad variable, pueden tomarse tantos bits de un troceo como se requiera para reducir el número de pruebas secundarias realizadas por debajo de un umbral fijado, por ejemplo, de unos pocos cientos.

25 Aunque realizaciones ejemplares específicas han sido descritas en lo que antecede, se conciben modificaciones y agregados en otras realizaciones.

30 Funciones de troceo

Por ejemplo, las funciones de troceo pueden ser adaptadas para proporcionar diversos grados de robustez, por ejemplo, para escoger el orden de los bits dentro del troceo, para maximizar su robustez con respecto a la consulta de base de datos de correlación exacta. Podrían usarse otras fuentes de entropía, invariables con respecto al desplazamiento del tono, con la base de datos en gran escala, además de los coeficientes de troceo de tipo cepstral.

35 Árbol de la base de datos

40 En el ejemplo anterior, la estructura 70 de árbol de la base de datos está organizada sobre una base binaria. Sin embargo, en otros ejemplos, el número de vástagos de un nodo podría ser un número distinto a dos y, en efecto, podría variar sobre el árbol. Este enfoque podría usarse para facilitar adicionalmente la igualación de los tamaños de las hojas. Como alternativa, o adicionalmente, puede usarse una estructura de árbol donde un troceo puede ser almacenado para cada uno de los vástagos de un nodo, por ejemplo, para ambos vástagos izquierdo y derecho de un nodo en un árbol binario (conocido como un 'árbol reguero').

45 Identificación de pistas duplicadas

50 Optativamente, se podría buscar, en la base de datos de pistas, secciones duplicadas de música. Las secciones únicas (que llamaremos 'segmentos') serían entonces almacenadas en la base de datos e identificadas según lo descrito anteriormente; y una posterior etapa de procesamiento convertirá la lista de segmentos reconocidos en una lista de pistas. Un enfoque de ese tipo implicaría un pre-procesamiento adicional, pero reduciría los requisitos de almacenamiento de la base de datos y podría acelerar el procesamiento en tiempo real.

Información de tiempo absoluto

55 En la realización anteriormente descrita, un tiempo absoluto para una entrada de datos de prueba terciaria es determinado barriendo hacia adelante hasta la misma desde el inicio de ese segmento, acumulando diferencias de tiempos. Optativamente, podrían incluirse marcadores de tiempo absoluto en una secuencia de entradas de datos de prueba terciaria.

60 Reducción de la base de datos

65 A fin de reducir el tamaño de la base de datos de prueba secundaria, puede usarse la reducción de la base de datos. Esto implica calcular un 'troceo de un troceo' para descartar una fracción fija de troceos de manera determinística. Por ejemplo, para reducir la base de datos en un factor de tres, pueden emplearse las siguientes modificaciones. Para cada troceo generado, aquellos bits que necesitarán correlacionarse exactamente en la base de datos son considerados como un entero. Si este entero no es exactamente divisible entre tres, se descarta el troceo, es decir,

no se incluye en la base de datos construida a partir del material de pistas de origen. Análogamente, si un troceo que incumple este criterio se encuentra al procesar material del programa, se sabe inmediatamente que no estará en la base de datos y, por lo tanto, no se realizará ninguna consulta. Se usa un criterio determinístico que es una función de los bits implicados en la correlación exacta para aceptar o rechazar troceos, en lugar de aceptar o rechazar sencillamente al azar con una probabilidad fija, ya que este último enfoque tendría un efecto adverso mucho mayor sobre la tasa de aciertos del troceo, especialmente para mayores tasas de reducción.

Realizaciones alternativas

10 Las realizaciones descritas anteriormente son solamente a modo de ejemplo. Pueden concebirse realizaciones alternativas dentro del espíritu y el alcance de las reivindicaciones.

15 Por ejemplo, en las realizaciones ejemplares descritas con respecto a las figuras, la evaluación primaria incluye realizar una correlación exacta de dígitos de un vector de origen con entradas en la tabla de consulta, en donde cada entrada en la tabla de consulta está asociada a un grupo de vectores de referencia. La evaluación secundaria incluye entonces determinar un grado de similitud entre el vector de origen y cada uno del grupo de vectores de referencia, para identificar vectores de referencia cualesquiera que sean candidatos para correlacionar el contenido de medios de origen con el contenido de medios de referencia. La evaluación terciaria implica luego determinar un grado de similitud entre uno o más vectores de origen adicionales y uno o más vectores de referencia adicionales, estando cada uno de los vectores de origen adicionales y de los vectores de referencia adicionales separado en el tiempo, respectivamente, del vector de origen y del vector de referencia candidato. Las evaluaciones secundarias y terciarias implican accesos aleatorios al almacenamiento que aloja la base de datos de vectores de referencia. Ha de observarse que la base de datos de vectores de referencia puede tener un tamaño considerable, por ejemplo, del orden de 10 teraocetos, o mayor.

25 Allí donde el procesamiento es realizado usando un aparato que está formado por un sistema de ordenador autónomo o en red, por ejemplo, un sistema de ordenador con uno o más procesadores y almacenamiento compartido, es ventajoso que la base de datos se aloje en dispositivos de memoria de estado sólido (SSD) para aumentar la velocidad de procesamiento y, por lo tanto, acelerar las etapas de procesamiento secundarias y terciarias. Sin embargo, tal almacenamiento es caro en la actualidad. El procesamiento puede ser realizado de esta manera usando dispositivos de almacenamiento más lentos, y de menor coste, tales como el almacenamiento en disco, pero esto puede frenar el proceso de reconocimiento, especialmente allí donde la base de datos de referencia sea grande.

30 Otra alternativa es usar un aparato que emplee un enfoque de formaciones, o un enfoque en la nube, para el procesamiento, donde las tareas de procesamiento son distribuidas a múltiples sistemas de ordenador, por ejemplo, que funcionan como tareas de trasfondo, siendo coordinados los resultados del procesamiento en la nube en un sistema de ordenador anfitrión.

40 También se concibe un enfoque adicional en el que una base de datos de origen, de vectores de origen, es generada a partir de un programa de origen, y luego los medios de referencia de una base de datos de referencia son correlacionados con respecto a la base de datos de origen, de manera lineal o en forma de flujo. Esto tiene la ventaja de que una base de datos de origen, de vectores de origen, por ejemplo, de la programación de un día de una estación de radio podría ser alojada en unos pocos gigaoctetos de memoria de acceso aleatorio, y luego la base de datos de referencia podría ser transmitida como flujo desde un almacenamiento de bajo coste, por ejemplo, un disco o cinta, y el proceso de comparación podría ser realizado en forma de lotes de bajo coste. En consecuencia, usando un enfoque de ese tipo, una base de datos de medios de origen, de vectores de origen para el material del programa de origen (por ejemplo, de un programa de radio, o un periodo adecuado de programación (digamos, una hora, una parte o la totalidad de un día, etc.)), podría ser generada de la manera descrita para la base de datos de medios de referencia, de vectores de referencia, de la figura 6. Los vectores de origen podrían ser almacenados en memoria de acceso aleatorio, clasificados en orden de valores crecientes de troceo, en una tabla de troceo, o en una estructura de base de datos similar a la descrita para la base de datos de medios de referencia, de vectores de referencia, de la figura 6. Los vectores de referencia podrían luego ser comparados con la base de datos de medios de origen, transmitiendo secuencialmente por flujo vectores de referencia desde la base de datos de medios de referencia (lo que es mucho más rápido que los accesos aleatorios en el caso de un almacenamiento de bajo coste, tal como el disco o la cinta). Este proceso podría incluir una evaluación primaria de la realización de una correlación exacta de dígitos de cada vector de referencia, con respecto a entradas en la tabla de la base de datos de origen, en donde cada entrada en la tabla de la base de datos de origen está asociada a un grupo de vectores de origen. La evaluación secundaria podría entonces incluir la determinación de un grado de similitud entre el vector de referencia actual y cada uno de los grupos de vectores de origen, para identificar vectores de origen cualesquiera que sean candidatos para correlacionar el contenido de medios de origen con el contenido de medios de referencia. La evaluación terciaria podría entonces implicar la determinación de un grado de similitud entre uno o más vectores de origen adicionales y uno o más vectores de referencia adicionales, estando cada uno de los vectores de origen adicionales y de los vectores de referencia adicionales separado en el tiempo, respectivamente, del vector de origen y del vector de referencia candidato. Las evaluaciones secundarias implicarían accesos aleatorios al almacenamiento que aloja la base de datos de vectores de origen, pero como esta es relativamente pequeña, puede

ser alojada en memoria de acceso aleatorio. Las evaluaciones terciarias implicarían accesos al almacenamiento que aloja la base de datos de vectores de origen y la base de datos de vectores de referencia. En una realización, la base de datos de vectores de referencia es almacenada en el orden natural, es decir, pista por pista, y con los vectores almacenados en orden temporal dentro de cada pista. En esta realización, las consultas implicadas en las evaluaciones terciarias se referirán a entradas adyacentes en la base de datos y, por ello, los accesos secuenciales al almacenamiento pueden ser usados para almacenar para reducir los tiempos de acceso. En una realización alternativa, la base de datos de vectores de referencia es almacenada en orden de valores crecientes de troceo, con fines de realización de pruebas secundarias, y el conjunto de candidatos para la evaluación terciaria sería recogido y clasificado por número de pista, para permitir que se usen accesos secuenciales al almacenamiento, con fines de realización de pruebas terciarias.

REIVINDICACIONES

1. Un método de reconocimiento automático para el reconocimiento automático de contenido de medios de origen, procedente de una señal de origen, por comparación con contenido de medios de referencia, incluyendo el método:

5 generar (12) un espectrograma a partir de la señal de origen, aplicando una transformación de Fourier a la señal de origen, incluyendo el espectrograma una pluralidad de columnas, siendo cada columna representativa de una tajada temporal, e incluyendo una pluralidad de depósitos de frecuencia, cada uno representativo de una respectiva gama de componentes de frecuencia para la tajada temporal de la señal de origen;

10 generar (14) al menos un vector de origen para una tajada temporal de la señal de origen, calculando razones de magnitudes de depósitos de frecuencia seleccionados, a partir de la columna para la tajada temporal, y cuantizando las razones para generar dígitos de un vector de origen, en donde la generación de al menos un vector para una tajada temporal incluye, para al menos un depósito de frecuencia seleccionado de una tajada temporal, calcular razones de ese depósito y de un depósito de frecuencia adyacente, o casi adyacente, a partir de la columna para la tajada temporal, y dividir las razones en gamas para generar al menos un dígito seleccionado para cada razón;

15 realizar una evaluación (18) primaria por la correlación exacta de dígitos de primeros vectores con entradas en una tabla (66) de consulta, en donde cada entrada en la tabla de consulta está asociada a un grupo de segundos vectores, y en donde el número de dígitos de los primeros vectores usados para realizar la correlación exacta difiere entre las entradas en la tabla de consulta; y

20 realizar una evaluación (20) secundaria para determinar un grado de similitud entre los primeros vectores y cada uno del grupo de segundos vectores, para identificar segundos vectores cualesquiera que sean candidatos para correlacionar el contenido de medios de origen con el contenido de medios de referencia,

25 en el cual una base de datos (146) almacena la tabla de consulta y los segundos vectores, y

30 en el cual los primeros vectores son vectores de origen, o bien vectores de referencia, y los segundos vectores son los otros vectores de origen y los otros vectores de referencia, representando cada vector de referencia una tajada temporal del contenido de medios de referencia.

2. El método de la reivindicación 1, en el cual la generación de al menos un vector para una tajada temporal incluye:

35 para al menos un depósito de frecuencia seleccionado de una tajada temporal, calcular razones de ese depósito y un depósito de frecuencia adyacente, o casi adyacente, a partir de la columna para la tajada temporal; y

dividir las razones en gamas para generar dos dígitos binarios para cada razón.

40 3. El método de la reivindicación 1 o la reivindicación 2, en el cual:

las gamas difieren entre los depósitos de razones seleccionadas, para igualar la distribución de los valores de las razones entre las gamas.

45 4. El método de una cualquiera de las reivindicaciones 1 a 3, que incluye:

generar uno de dichos vectores de origen usando depósitos de frecuencia seleccionados entre al menos una entre una banda de frecuencia entre 400 Hz y 1.100 Hz, y una banda de frecuencia entre 1.100 Hz y 3.000 Hz.

50 5. El método de una cualquiera de las reivindicaciones 1 a 4, que incluye generar un vector de origen adicional para una tajada temporal:

55 generando un espectrograma adicional a partir de la primera señal, aplicando una transformación de Fourier a la señal de origen, incluyendo el espectrograma adicional una pluralidad de columnas, siendo cada columna representativa de una tajada temporal, e incluyendo una pluralidad de depósitos de frecuencia, cada uno representativo de una respectiva gama de componentes de frecuencia para la tajada temporal de la primera señal;

60 aplicando una transformación adicional de Fourier a los respectivos depósitos de frecuencia, a partir de la columna para la tajada temporal, para generar un respectivo conjunto de coeficientes;

65 generando el vector de origen adicional de modo que, para un conjunto de N coeficientes en una columna para una tajada temporal, para cada uno de los elementos 2 a N-1 del vector de origen adicional, sea formado un n-ésimo elemento por el cuadrado del n-ésimo coeficiente, dividido entre el producto del (n-1)-ésimo coeficiente y el (n+1)-ésimo coeficiente, y cuantizando los elementos del vector resultante para generar al menos un dígito para cada elemento.

6. Un método de reconocimiento automático para el reconocimiento automático de contenido de medios de origen, a partir de una señal de origen, por comparación con contenido de medios de referencia, incluyendo el método:

5 generar (12) un espectrograma a partir de la señal de origen, aplicando una transformación de Fourier a la señal de origen, incluyendo el espectrograma una pluralidad de columnas, siendo cada columna representativa de una tajada temporal, e incluyendo una pluralidad de depósitos de frecuencia, cada uno representativo de una respectiva gama de componentes de frecuencia para la tajada temporal de la señal de origen;

10 generar (14) al menos un vector de origen para una tajada temporal de la señal de origen, calculando razones de magnitudes de depósitos de frecuencia seleccionados a partir de la columna para la tajada temporal, y cuantizando las razones para generar dígitos de un vector de origen;

15 realizar una evaluación (18) primaria por la correlación exacta de dígitos de primeros vectores con entradas en una tabla (66) de consulta, en donde cada entrada en la tabla de consulta está asociada a un grupo de segundos vectores, y en donde el número de dígitos de los primeros vectores usados para realizar la correlación exacta difiere entre las entradas en la tabla de consulta; y

20 realizar una evaluación (20) secundaria para determinar un grado de similitud entre los primeros vectores y cada uno del grupo de segundos vectores, para identificar segundos vectores cualesquiera que sean candidatos para correlacionar el contenido de medios de origen con el contenido de medios de referencia,

en el cual una base de datos (146) almacena la tabla de consulta y los segundos vectores, y

25 en el cual los primeros vectores son vectores de origen, o bien vectores de referencia, y los segundos vectores son los otros vectores de origen y los otros vectores de referencia, representando cada vector de referencia una tajada temporal del contenido de medios de referencia;

incluyendo además el método generar un vector de origen adicional para una tajada temporal:

30 generando un espectrograma adicional a partir de la primera señal, aplicando una transformación de Fourier a la señal de origen, incluyendo el espectrograma adicional una pluralidad de columnas, siendo cada columna representativa de una tajada temporal, e incluyendo una pluralidad de depósitos de frecuencia, cada uno representativo de una respectiva gama de componentes de frecuencia para la tajada temporal de la primera señal;

35 aplicando una transformación adicional de Fourier a los respectivos depósitos de frecuencia, a partir de la columna para la tajada temporal, para generar un respectivo conjunto de coeficientes;

40 generando el vector de origen adicional de modo que, para un conjunto de N coeficientes en una columna para una tajada temporal, para cada uno de los elementos 2 a N-1 del vector de origen adicional, sea formado un n-ésimo elemento por el cuadrado del n-ésimo coeficiente, dividido entre el producto del (n-1)-ésimo coeficiente y el (n+1)-ésimo coeficiente, y cuantizando los elementos del vector resultante para generar al menos un dígito para cada elemento.

45 7. El método de una cualquiera de las reivindicaciones 1 a 6, en el cual la señal de origen es una señal de audio y las frecuencias de los depósitos del espectrograma son adjudicadas de acuerdo a una escala logarítmica.

8. El método de una cualquiera de las reivindicaciones 1 a 7, en el cual:

50 la tabla de consulta está organizada como un árbol de profundidad variable que lleva a las hojas, estando la tabla indizada por el primer vector;

cada hoja conforma una entrada en la tabla de consulta, asociada un respectivo grupo de segundos vectores;

55 el número de dígitos que llevan a cada hoja está determinado para igualar el tamaño de los grupos de segundos vectores para cada hoja.

9. El método de la reivindicación 8, en el cual:

60 el número de dígitos que llevan a cada hoja forma el número de dígitos del primer vector usado para realizar la correlación exacta para una hoja dada.

10. El método de la reivindicación 8 o la reivindicación 9, en el cual cada hoja de la tabla de consulta identifica un grupo de segundos vectores con d dígitos coincidentes, en donde d corresponde a la profundidad del árbol hasta esa hoja.

65 11. El método de una cualquiera de las reivindicaciones 1 a 10, que incluye realizar la evaluación secundaria usando

una métrica de distancia para determinar el grado de similitud entre el primer vector y cada uno del grupo de segundos vectores.

5 12. El método de una cualquiera de las reivindicaciones 1 a 11, que incluye realizar una evaluación (22) terciaria para cualquier segundo vector identificado como candidato, incluyendo la evaluación terciaria determinar un grado de similitud entre uno o más primeros vectores adicionales y uno o más segundos vectores adicionales, correspondientes al segundo vector candidato identificado en la evaluación secundaria.

10 13. El método de la reivindicación 12, en el cual los primeros vectores adicionales y los segundos vectores adicionales están separados en el tiempo, respectivamente, del primer vector y del segundo vector candidato.

14. El método de una cualquiera de las reivindicaciones 1 a 13, en el cual la señal de origen es una señal de programa recibida.

15 15. El método de la reivindicación 14, que incluye generar un registro del contenido de medios correlacionado de la señal del programa.

20 16. El método de la reivindicación 15, que incluye generar una hoja de indicaciones que identifica el contenido de medios correlacionado.

17. El método de una cualquiera de las reivindicaciones 1 a 16, en el cual los segundos vectores son los vectores de origen y el aparato está configurado para generar la base de datos a partir de los vectores de origen.

25 18. Un producto de programa de ordenador que incluye instrucciones de programa, operables para llevar a cabo el método de una cualquiera de las reivindicaciones 1 a 17.

19. Aparato para llevar a cabo el método de la Reivindicación 1 o de cualquier Reivindicación dependiente de la misma, incluyendo el aparato:

30 un generador (112) de espectrogramas, operable para generar (12) un espectrograma a partir de la señal de origen, aplicando una transformación de Fourier a la señal de origen, incluyendo el espectrograma una pluralidad de columnas, siendo cada columna representativa de una tajada temporal, e incluyendo una pluralidad de depósitos de frecuencia, cada uno representativo de una respectiva gama de componentes de frecuencia para la tajada temporal de la señal de origen;

35 un generador (114) de vectores, operable para generar (14) al menos un vector de origen para una tajada temporal de la señal de origen, calculando razones de magnitudes de depósitos de frecuencia seleccionados, a partir de la columna para la tajada temporal, y para cuantizar las razones para generar dígitos de un vector de origen, en donde, para generar al menos un vector para una tajada temporal, el generador de vectores es operable, para al menos un depósito de frecuencia seleccionado de una tajada temporal, para calcular una razón de magnitudes de ese depósito y un depósito adyacente, o casi adyacente, a partir de la columna para la tajada temporal, y dividir las razones en gamas para generar al menos un dígito seleccionado para cada razón;

45 un evaluador (118) primario, operable para realizar una evaluación (18) primaria, realizando una correlación exacta de dígitos de primeros vectores con entradas en una tabla de consulta, en donde cada entrada en la tabla (66) de consulta está asociada a un grupo de segundos vectores, y en donde el número de dígitos de los primeros vectores usados para realizar la correlación exacta difiere entre entradas en la tabla de consulta;

50 un evaluador (120) secundario, operable para realizar una evaluación secundaria (20), para determinar un grado de similitud entre los primeros vectores y cada uno del grupo de segundos vectores, para identificar segundos vectores cualesquiera que sean candidatos para correlacionar el contenido de medios de origen con el contenido de medios de referencia; y

55 una base de datos (146) que comprende la tabla de consulta y los segundos vectores, en el cual los primeros vectores son vectores de origen, o bien vectores de referencia, y los segundos vectores son los otros vectores de origen, y los otros vectores de referencia, representando cada vector de referencia una tajada temporal del contenido de medios de referencia.

60 20. Aparato para llevar a cabo el método de la Reivindicación 6, o de cualquier Reivindicación dependiente de la misma, incluyendo el aparato:

65 un generador (112) de espectrogramas, operable para generar (12) un espectrograma a partir de la señal de origen, aplicando una transformación de Fourier a la señal de origen, incluyendo el espectrograma una pluralidad de columnas, siendo cada columna representativa de una tajada temporal, e incluyendo una pluralidad de depósitos de frecuencia, cada uno representativo de una respectiva gama de componentes de frecuencia para la tajada temporal

de la señal de origen;

5 un generador (114) de vectores, operable para generar (14) al menos un vector de origen para una tajada temporal de la señal de origen, calculando razones de magnitudes de depósitos de frecuencia seleccionados, a partir de la columna para la tajada temporal, y para cuantizar las razones para generar dígitos de un vector de origen;

10 un evaluador (118) primario, operable para realizar una evaluación (18) primaria, realizando una correlación exacta de dígitos de primeros vectores con entradas en una tabla de consulta, en donde cada entrada en la tabla de consulta está asociada a un grupo de segundos vectores, y en donde el número de dígitos de los primeros vectores, usados para realizar la correlación exacta, difiere entre las entradas en la tabla de consulta;

15 un evaluador (120) secundario, operable para realizar una evaluación (20) secundaria para determinar un grado de similitud entre los primeros vectores y cada uno del grupo de segundos vectores, para identificar segundos vectores cualesquiera que sean candidatos para correlacionar el contenido de medios de origen con el contenido de medios de referencia; y

una base de datos (146) que comprende la tabla de consulta y los segundos vectores,

20 en el cual los primeros vectores son vectores de origen, o bien vectores de referencia, y los segundos vectores son los otros vectores de origen y los otros vectores de referencia, representando cada vector de referencia una tajada temporal del contenido de medios de referencia;

en el cual, para generar un vector de origen adicional para una tajada temporal:

25 el generador de espectrogramas es operable para generar un espectrograma adicional, aplicando una transformación de Fourier a la señal de origen, incluyendo el espectrograma adicional una pluralidad de columnas, siendo cada columna representativa de una tajada temporal, e incluyendo una pluralidad de depósitos de frecuencia, cada uno representativo de una respectiva gama de componentes de frecuencia para la tajada temporal de la señal de origen, y para aplicar una transformación adicional de Fourier a los respectivos depósitos de frecuencia, a partir de la columna para la tajada temporal, para generar un respectivo conjunto de coeficientes; y

30 el generador de vectores es operable para generar el vector de origen adicional, de modo que, para un conjunto de N coeficientes en una columna para una tajada temporal, para cada uno de los elementos 2 a N-1 del vector de origen adicional, sea formado un n-ésimo elemento por el cuadrado del n-ésimo coeficiente, dividido entre el producto del (n-1)-ésimo coeficiente y el (n+1)-ésimo coeficiente, y para cuantizar los elementos del vector resultante para generar al menos un dígito para cada elemento.

40 21. El aparato de la Reivindicación 19 o la Reivindicación 20, que incluye adicionalmente un evaluador (122) terciario, siendo el evaluador terciario para realizar una evaluación (22) terciaria para cualquier segundo vector identificado como candidato, y siendo el evaluador terciario operable para determinar un grado de similitud entre uno o más primeros vectores adicionales y uno o más segundos vectores adicionales, correspondientes al segundo vector candidato, identificado en la evaluación secundaria.

45 22. El aparato de una cualquiera de las Reivindicaciones 19 a 21, incluyendo el aparato al menos un procesador (42) y almacenamiento (44, 46), y software de ordenador, operable para implementar el generador de espectrogramas, el generador de vectores de troceo y los evaluadores.

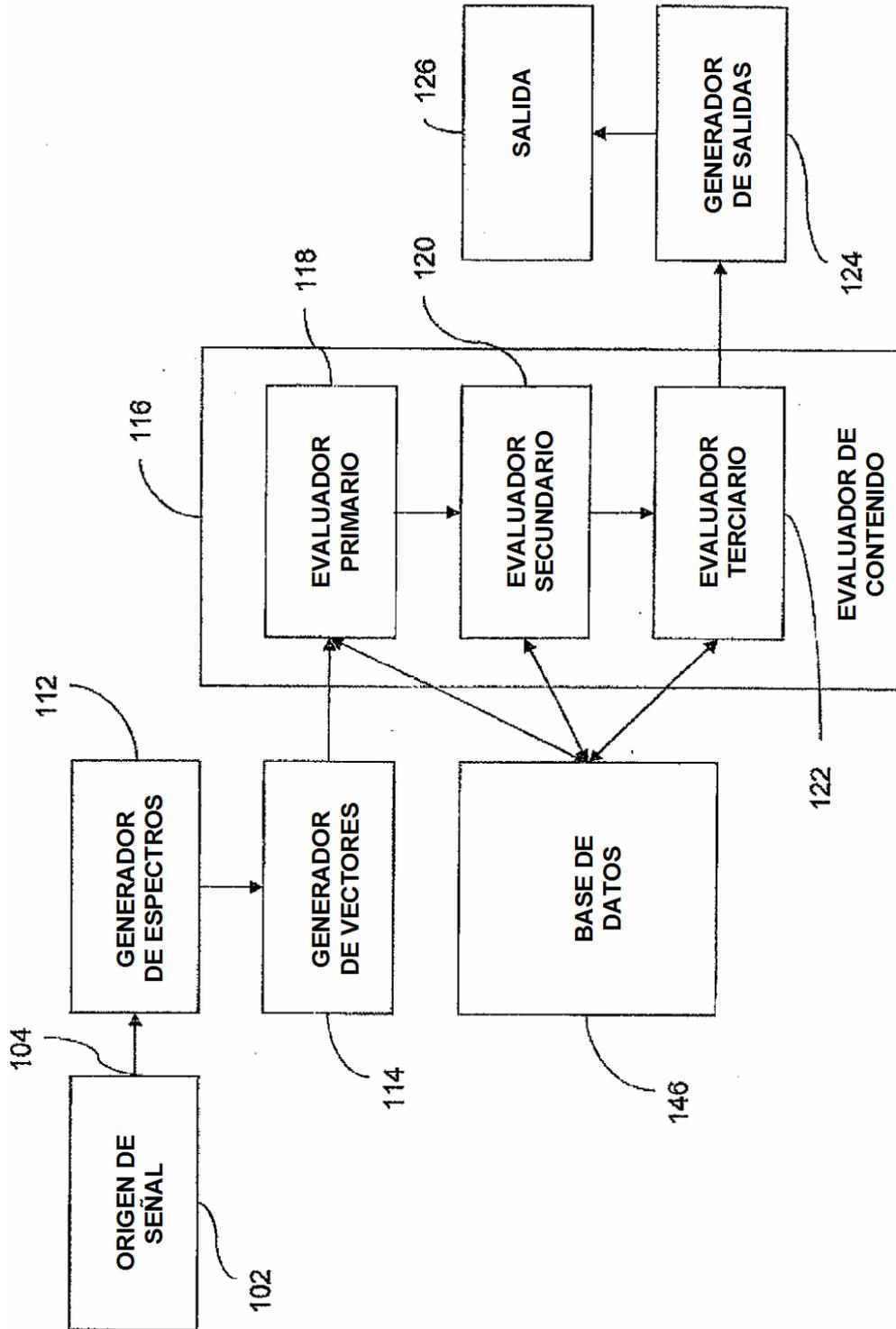
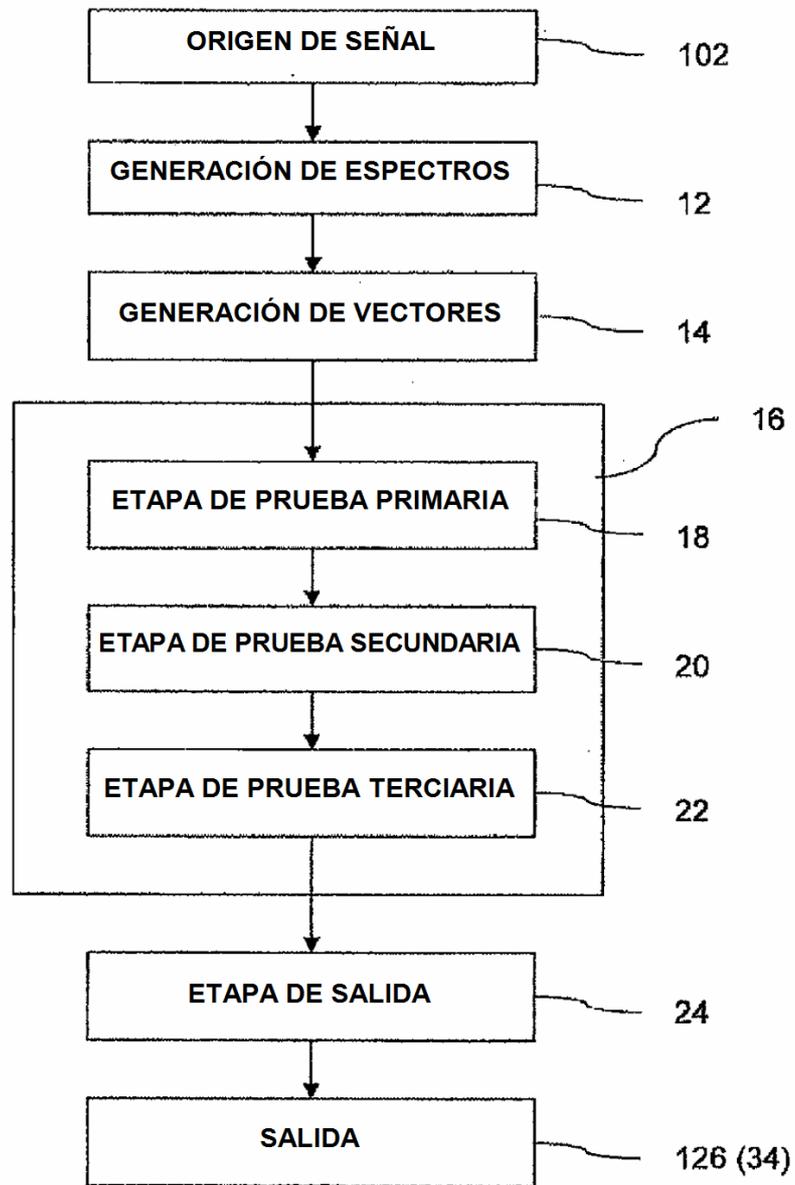


FIG. 1



10

FIG. 2

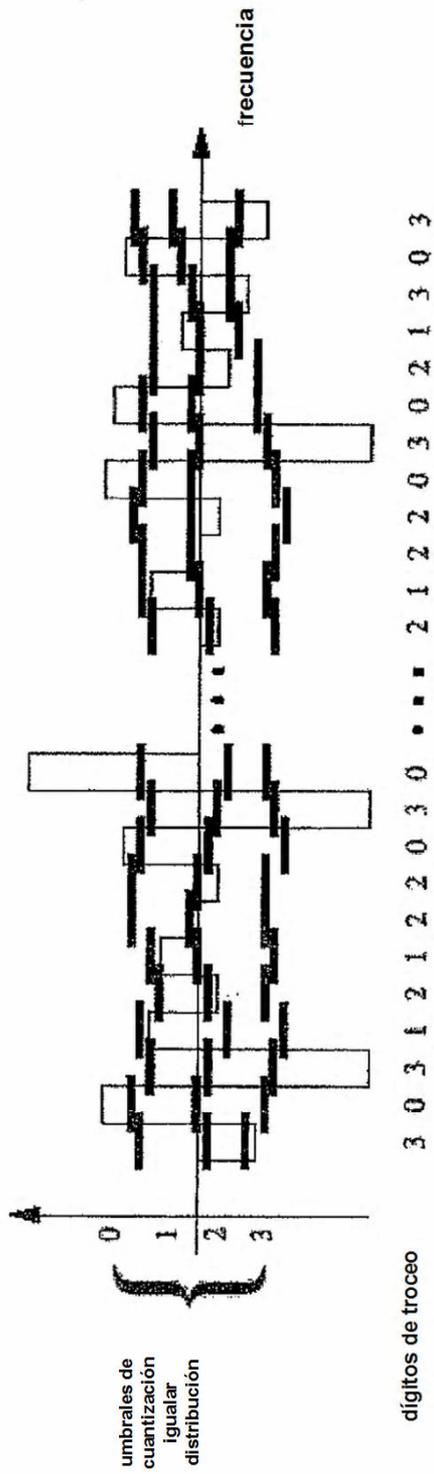


FIG. 3

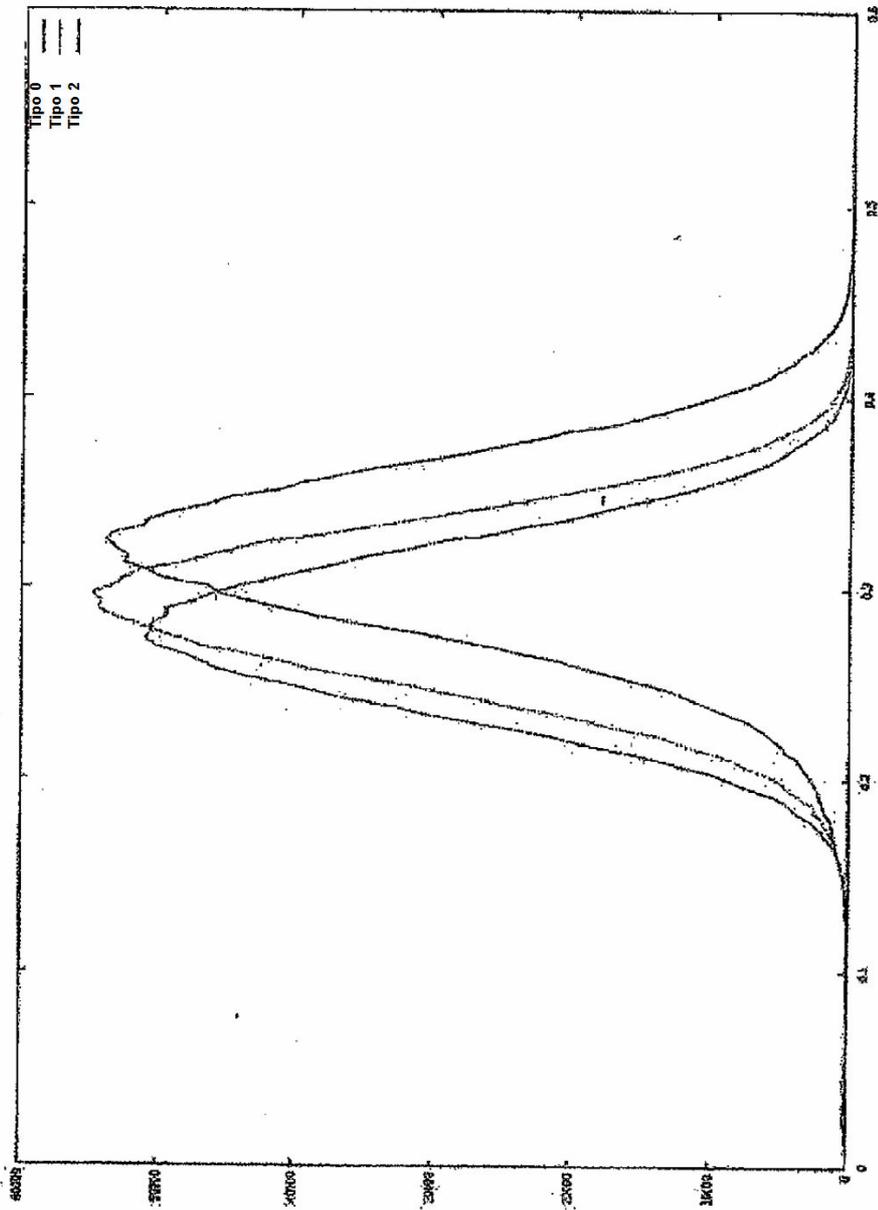


FIG. 4

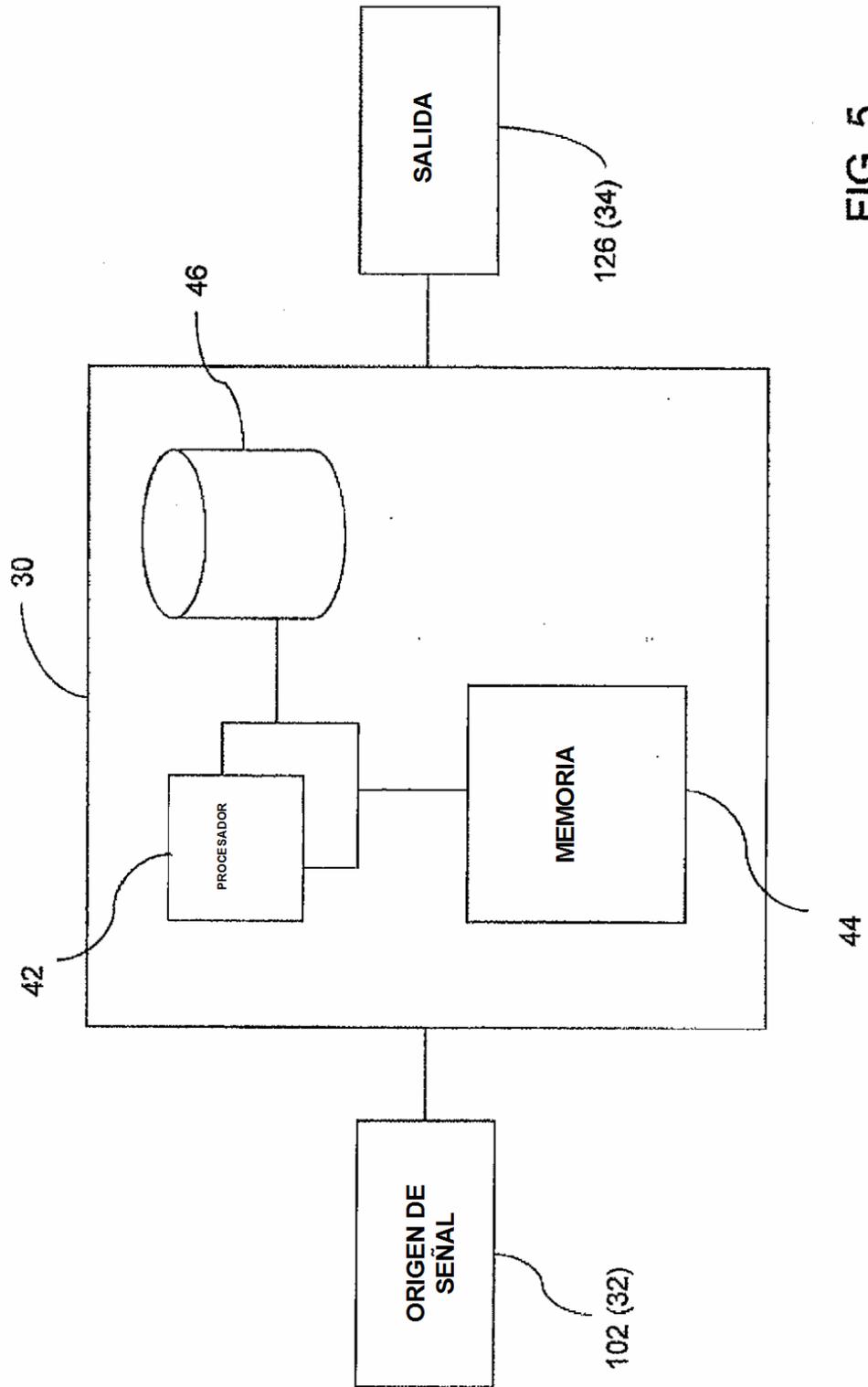


FIG. 5

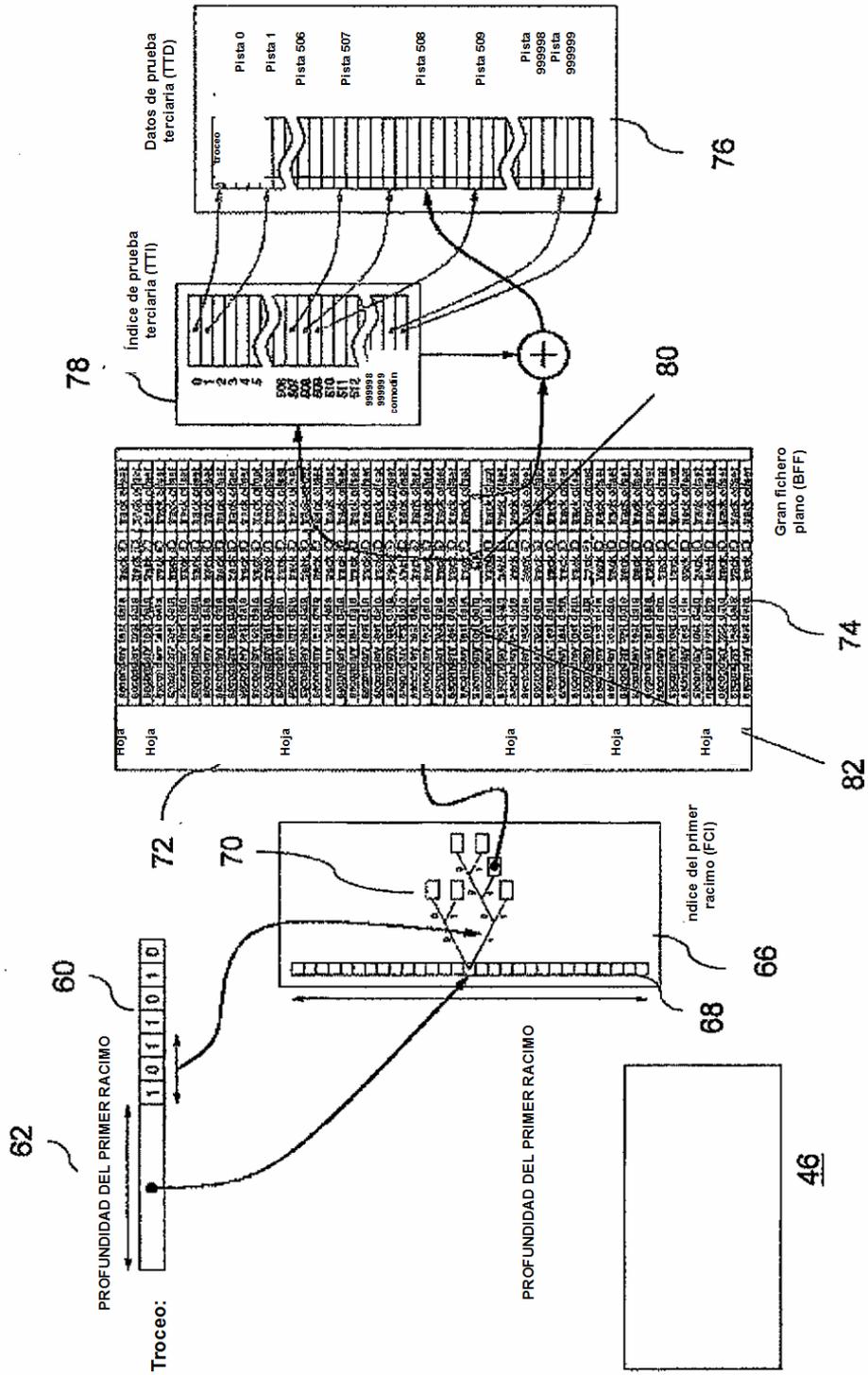


FIG. 6