

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 776 705**

51 Int. Cl.:

**G10L 25/54** (2013.01)

**G06F 16/683** (2009.01)

**G10H 1/00** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **25.02.2014 PCT/FR2014/050399**

87 Fecha y número de publicación internacional: **04.09.2014 WO14131984**

96 Fecha de presentación y número de la solicitud europea: **25.02.2014 E 14713179 (1)**

97 Fecha y número de publicación de la concesión europea: **25.12.2019 EP 2962301**

54 Título: **Generación de una firma de una señal de audio musical**

30 Prioridad:

**27.02.2013 FR 1351752**

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

**31.07.2020**

73 Titular/es:

**INSTITUT MINES-TELECOM (100.0%)  
46 Rue Barrault  
75634 Paris, FR**

72 Inventor/es:

**FENET, SÉBASTIEN;  
GRENIER, YVES y  
RICHARD, GAËL**

74 Agente/Representante:

**VEIGA SERRANO, Mikel**

ES 2 776 705 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**

Generación de una firma de una señal de audio musical

**5 Sector de la técnica**

La presente invención se refiere a la generación de firma para la identificación de señales de audio.

10 Se refiere, más precisamente, a la identificación de señales de audio musicales por extracción de huellas (comúnmente llamada "audio-fingerprinting", "huella digital de audio").

**Estado de la técnica**

15 La identificación de señales de audio tiene como finalidad recuperar unos metadatos, con el fin de asociarlos a un extracto de audio desconocido *a priori*. Los terminales móviles, tales como unos teléfonos móviles de tipo Smartphone (teléfono inteligente), tienden a integrar unas aplicaciones de identificación de este tipo. A estos efectos, el usuario captura un extracto de audio por medio del micrófono de su Smartphone (teléfono inteligente) y la aplicación devuelve un conjunto de metadatos, tales como el título de una canción, el nombre de un artista, el nombre de un álbum, etc.

20 La identificación de señales de audio posee, igualmente, unas aplicaciones en el control automático de derecho de autor, en la segmentación y la anotación automáticas de flujos multimedia, en la detección de jingle, etc.

25 La extracción de huella es la herramienta más extendida para efectuar la identificación de señales de audio de solicitud cuando no está vinculado ningún metadato a la señal de audio desconocida. Consiste en extraer, de señales de audio de referencia, una representación compacta (huella o firma) que se almacena, a continuación, en una base de datos. Durante la identificación de una señal de audio de solicitud, se extrae su firma. Entonces, se busca la mejor correspondencia con la firma extraída en la base de datos que comprende las firmas de las señales de audio de referencia.

30 Se conocen unos numerosos ejemplos de modelos de huellas (o firmas), así como unos numerosos métodos de búsqueda de la mejor correspondencia.

En particular, generalmente, se ponen de relieve dos objetivos en los métodos conocidos:

- 35 - la señal de audio capturada (de solicitud) se debe reconocer incluso si ha experimentado una serie de distorsiones;
- el algoritmo implementado debe poder gestionar una base de datos que contiene unas cantidades muy importantes de señales de audio de referencia.

40 En el estado de la técnica, la robustez a la distorsión se define, generalmente, como que es la robustez a diferentes postprocesamientos de la señal de audio. En efecto, la problemática que se considera, generalmente, es la identificación de una señal que se ha truncado, reigualado, comprimido en amplitud, dilatado temporalmente, desviado al nivel del tono y/o grabado en un entorno con un micrófono malo. Dicho de otra manera, se trata de distorsiones introducidas en el canal de transmisión durante la difusión de la señal de audio.

45 Según los métodos, la robustez cubre más o menos estos tipos de distorsiones, pero estas distorsiones corresponden todas a unos postprocesamientos aplicados a una señal de referencia. La señal de audio de solicitud corresponde, entonces, a una de las señales de audio de referencia a la que se ha aplicado una serie de distorsiones y la identificación consiste en volver a encontrar la señal de audio de referencia. Una identificación de este tipo se llama "identificación cuasi exacta".

50 En audio, se definen dos señales de audio similares como que son dos señales de audio que presentan las mismas características armónicas, rítmicas y melódicas. Intuitivamente, estas son dos señales de audio que un ser humano considerará como próximas.

55 Sin embargo, los métodos por extracción de huellas tradicionales no abordan, generalmente, el caso en el que dos señales de audio son similares. Esto significa, en concreto, que si el sistema ha aprendido un título (entonces, almacenado en una base de datos en forma de una señal de audio de referencia) y que se requiere la identificación de una reposición de este título (señal de audio de solicitud), el sistema será incapaz de indicar la correspondencia.

60 De manera más molesta, un mismo título, con la misma instrumentación, que está regrabado por el mismo artista, no se considerará como correspondiente al título original. En efecto, aunque, para el oído humano, las dos grabaciones parecen ser las mismas, las ligeras variaciones que existen entre las dos versiones (ligeros desvíos rítmicos, canto ligeramente modificado, ejecución instrumental un poco diferente) hacen de ello unas señales radicalmente diferentes de un punto de vista informático.

65 La identificación de dos señales de audio similares se llama "identificación aproximada".

Unas limitaciones de este tipo se deben, principalmente, al hecho de que los métodos por extracción de huellas tradicionales utilizan unas informaciones de bajo nivel para caracterizar la señal. Estas informaciones, que integran muy poca significación musical, no se conservan de una versión a otra de un mismo título.

5 De este modo, los métodos por extracción de huellas son poco utilizados en los campos de la similitud musical o de la detección de reposiciones.

10 Por otra parte, se han desarrollado unos enfoques que permiten poner en correspondencia dos versiones diferentes de una misma canción. Sin embargo, unos enfoques de este tipo son incapaces de procesar unas bases de datos de señales de referencia de gran tamaño. En efecto, estos métodos consumen recursos y requieren comparar una por una las señales de audio de referencia con la señal de audio de solicitud.

15 Por lo tanto, existe una necesidad de definir una firma de una señal de audio que permita a la vez la identificación cuasi exacta y la identificación aproximada de señales de audio, permitiendo al mismo tiempo gestionar unas bases de datos de gran tamaño.

20 La técnica anterior, el documento US 2009/234649 A1, describe un aparato de generación de firma de audio que comprende un análisis de la energía de la señal, una detección de eventos y un generador de firma. Esta técnica anterior comprende, igualmente, un aparato para la comparación de las firmas.

25 El documento US 2008/263360 permite obtener unos troceados robustos para unos contenidos multimedia, tales como unas secuencias de audio. La secuencia de audio está subdividida en tramas sucesivas. Se calcula una propiedad robusta que se representa por un bit de troceado correspondiente. De este modo, una secuencia de audio estará representada por una concatenación de palabras de troceado binario, a razón de una por trama.

### Objeto de la invención

30 La presente invención viene a mejorar la situación.

Según un primer aspecto de la invención, se describe un procedimiento de generación de una firma de una señal de audio musical de una duración dada, comprendiendo el procedimiento las siguientes etapas:

- 35 - modelado de la señal de audio musical para obtener, para cada banda de frecuencias de un conjunto de  $n$  bandas de frecuencias, un diagrama que representa la energía de la señal de audio para la banda de frecuencias, en función del tiempo durante la duración dada;
- determinación de instantes  $t_k$  de transición musical de la señal de audio durante la duración dada;
- asociación de cada instante  $t_k$  de transición musical con una información local que comprende un vector de  $n$  valores respectivamente representativos de la energía de la señal de audio en cada uno de los  $n$  diagramas obtenidos entre el instante  $t_k$  de transición musical y un instante siguiente  $t_{k+1}$  de transición musical y/o un vector de  $n$  valores respectivamente representativos de la energía de la señal de audio en cada uno de los  $n$  diagramas obtenidos entre el instante  $t_k$  de transición musical y un instante anterior  $t_{k-1}$  de transición musical;
- 40 - determinación, en función de la información local asociada a cada instante  $t_k$  de transición musical, de una clave asociada al instante de transición musical, formando las claves determinadas un primer conjunto de claves de la señal de audio;
- 45 - generación de una firma de la señal de audio musical que comprende unos pares, comprendiendo cada par una clave del primer conjunto de claves y un instante  $t_k$  de transición musical asociado. Las claves del primer conjunto de claves se obtienen por cuantificación de los vectores de  $n$  valores respectivamente representativos de la energía de la señal en cada banda de frecuencias del conjunto de  $n$  bandas de frecuencias en un período comprendido entre el instante  $t_k$  de transición musical y un instante siguiente  $t_{k+i}$  de transición musical o por cuantificación de los vectores de  $n$  valores respectivamente representativos de la energía de la señal en cada banda de frecuencias del conjunto de  $n$  bandas de frecuencias en un período comprendido entre el instante  $t_k$  de transición musical y un instante anterior  $t_{k-i}$  de transición musical.

55 El procedimiento permite, de este modo, generar una firma que es representativa de informaciones rítmicas y armónicas de la señal de audio musical. De este modo, las claves obtenidas según el procedimiento citado anteriormente pueden ser comunes a dos versiones de un mismo título de audio y la firma se puede utilizar en el contexto de la identificación aproximada. Por lo demás, siendo la firma compacta, permite la comparación eficaz sobre una duración importante de señal, lo que es necesario, en concreto, en el marco de la identificación aproximada.

60 Se entiende por señal de audio musical una señal de audio de solicitud (que se busca identificar) o una señal de audio de referencia (para almacenamiento en una base de datos y para comparación ulterior con una señal de audio de solicitud).

65 La cuantificación utilizada para generar las claves permite bajar la selectividad operada basándose en estas claves. Cuantificando unos vectores de valores de energía procedentes de dos señales similares, pero diferentes, es posible

obtener la misma clave dos veces y, de este modo, poder identificar una señal de solicitud de manera aproximada.

Según un modo de realización de la invención, cada clave del primer conjunto de claves puede ser un vector de  $n$  bits, estando cada bit de índice  $i$  determinado por comparación entre el valor de energía de índice  $i$  en el vector de  $n$  valores a partir del que se determina la clave y la media de los valores del vector de  $n$  valores.

De este modo, la cuantificación aplicada a los valores de los vectores es máxima, lo que permite bajar la selectividad de las claves, conservando al mismo tiempo un significado de un punto de vista musical (utilización de un vector de  $n$  bits asociado a un instante de transición musical representativo).

Como complemento, si el valor de energía de índice  $i$  en el vector de  $n$  valores a partir del que se determina la clave es superior a la media de los valores de dicho vector de  $n$  valores, entonces, el bit de índice  $i$  puede ser igual a 1, siendo el bit de índice  $i$  igual a 0 si no.

Según un modo de realización de la invención, los instantes  $t_k$  de transición musical corresponden a los máximos locales de una función de inicio calculada a partir de la señal de audio musical.

Se entiende por función de inicio cualquier función cuyo papel es estimar los ataques de notas y los ataques de percusión en una señal de audio. A título de ejemplo, una función de este tipo se describe en la publicación "Accurate Tempo Estimation Based on Harmonic + Noise Decomposition", de Alonso y Al, 2007.

Una realización de este tipo permite asegurar que los tiempos de transición musical son significativos de un punto de vista rítmico.

Un segundo aspecto de la invención se refiere a un procedimiento de identificación de una señal musical, de solicitud, de entre unas señales musicales, de referencia, estando la señal de solicitud y las señales de referencia representadas por unas firmas respectivas obtenidas por la implementación del procedimiento según el primer aspecto de la invención, comprendiendo, de este modo, cada firma de señal de referencia un segundo conjunto de claves a las que están asociados unos instantes respectivos  $t_k$  de transición musical, comprendiendo el procedimiento las etapas:

- para cada firma de señal de referencia que tiene al menos una clave del segundo conjunto en común con el primer conjunto de claves de la firma de la señal de solicitud, constitución de un tercer conjunto de claves que comprende la al menos una clave del segundo conjunto en común con el primer conjunto, estando cada clave del tercer conjunto asociada a la vez al instante de transición musical asociado a la clave en la firma de la señal de referencia y al instante de transición musical asociado a la clave en la firma de la señal de solicitud;
- determinación de una puntuación en función de una correspondencia entre cada pareja de instantes de transición musical asociada a cada clave del tercer conjunto; y
- selección de al menos una firma de señal de referencia que tiene una de las puntuaciones más elevadas, en calidad de firma de una señal de referencia candidata para corresponder a la señal de solicitud.

Este modo de realización permite, de este modo, obtener con una complejidad reducida un subconjunto de firmas de las señales de audio de referencia que son susceptibles de identificar de manera cuasi exacta o aproximada la señal de audio de solicitud.

Como complemento, se puede seleccionar un juego de firmas de señales de referencia que tienen unas puntuaciones más elevadas y se puede aplicar un postprocesamiento:

- a unas señales de referencia candidatas, representadas por las firmas de dicho juego y
- a la señal de solicitud,

y una etapa de identificación de la señal de solicitud como una de las señales de referencia.

De este modo, es posible aplicar un postprocesamiento, que implementa unos cálculos, generalmente, complejos, a un conjunto reducido de señales de referencia candidatas, lo que permite disminuir los recursos requeridos para identificar la señal de solicitud, en comparación con el mismo postprocesamiento aplicado al conjunto de la base de datos.

Como variante o como complemento, la determinación de la puntuación de una firma de señal de referencia comprende las siguientes etapas:

- determinación de una función afín que pone en correspondencia un mayor número de instantes de transición musical de la señal de referencia asociados a las claves del tercer conjunto con los instantes de transición musical de la señal de solicitud asociados a las claves respectivas del tercer conjunto.

La puntuación atribuida a la firma de la señal de referencia puede, entonces, ser igual al mayor número citado anteriormente.

Una realización de este tipo permite seleccionar con una complejidad reducida unas firmas de señales de referencia que tienen una puntuación elevada, siendo la puntuación representativa de una similitud musical entre las señales de audio.

5 Como complemento, la función afín se puede obtener por aplicación de una transformada de Hough al conjunto de parejas de instantes de transición musical asociadas a las claves del tercer conjunto.

10 Según un modo de realización, la determinación de la puntuación de una señal de audio de referencia puede comprender las siguientes etapas:

- 15 - construcción de una nube de puntos, correspondiendo cada punto a una clave del tercer conjunto, siendo la abscisa del punto el instante de transición musical de la señal de referencia asociada a la clave del tercer conjunto y correspondiendo la ordenada del punto al instante de transición musical de la señal de solicitud asociada a la clave del tercer conjunto;
- determinación de una recta que conecta un máximo de puntos de la nube de puntos construida;
- determinación de la puntuación de la firma de la señal de referencia como que es igual al número de puntos conectados por la recta determinada en la nube de puntos.

20 Como complemento o como variante, el postprocesamiento se aplica sucesivamente para cada señal de referencia correspondiente a una firma seleccionada y a la señal de solicitud modificada en función de la función afín determinada para la obtención de la puntuación de la firma de la señal de referencia.

25 De este modo, este modo de realización permite aprovechar informaciones procedentes de la etapa de obtención de una función afín (que puede ser la recta en la nube de puntos introducida anteriormente), en el marco del postprocesamiento, que mejora, de este modo, la eficacia del procedimiento de identificación.

30 Como complemento, la señal de solicitud se modifica por puesta a escala de la señal de solicitud en función de una pendiente de la función afín y por reencuadre temporal de la señal de solicitud en función de una ordenada en el origen de la función afín.

Un modo de realización de este tipo permite asegurar que la señal de audio de referencia y la señal de audio de solicitud modificada se comparen con una misma dilatación temporal y que parte del mismo instante de partida.

35 Un tercer aspecto de la invención se refiere a un programa de ordenador que incluye unas instrucciones para la implementación del procedimiento según el primer aspecto de la invención, cuando este programa se ejecuta por un procesador.

40 Un cuarto aspecto de la invención se refiere a un programa de ordenador que incluye unas instrucciones para la implementación del procedimiento según el segundo aspecto de la invención, cuando este programa se ejecuta por un procesador.

Un quinto aspecto de la invención se refiere a un dispositivo de generación de una firma de una señal de audio musical de una duración dada, comprendiendo el dispositivo:

- 45 - una unidad de modelado de la señal de audio musical para obtener, para cada banda de frecuencias de un conjunto de  $n$  bandas de frecuencias, un diagrama que representa la energía de la señal de audio para la banda de frecuencias, en función del tiempo durante la duración dada;
- 50 - una primera unidad de determinación de instantes  $t_k$  de transición musical de la señal de audio durante la duración dada;
- una unidad de asociación de cada instante  $t_k$  de transición musical con una información local que comprende un vector de  $n$  valores respectivamente representativos de la energía de la señal de audio en cada uno de los  $n$  diagramas obtenidos entre el instante  $t_k$  de transición musical y un instante siguiente  $t_{k+1}$  de transición musical y/o un vector de  $n$  valores respectivamente representativos de la energía de la señal de audio en cada uno de los  $n$  diagramas obtenidos entre el instante  $t_k$  de transición musical y un instante anterior  $t_{k-1}$  de transición musical;
- 55 - una segunda unidad de determinación, en función de la información local asociada a cada instante  $t_k$  de transición musical, de una clave asociada al instante de transición musical, formando las claves determinadas un primer conjunto de claves de la señal de audio;
- 60 - una unidad de generación de una firma de la señal de audio musical que comprende unos pares, comprendiendo cada par una clave del primer conjunto de claves y un instante  $t_k$  de transición musical asociado.

Un sexto aspecto de la invención se refiere a un dispositivo de identificación de una señal musical, de solicitud, de entre unas señales musicales, de referencia, estando la señal de solicitud y las señales de referencia representadas por unas firmas respectivas obtenidas por la implementación del procedimiento según el primer aspecto de la invención, comprendiendo, de este modo, cada firma de señal de referencia un segundo conjunto de claves a las que

están asociados unos instantes respectivos  $\tau_k$  de transición musical, comprendiendo el dispositivo:

- 5 - una unidad de constitución, para cada firma de señal de referencia que tiene al menos una clave del segundo conjunto en común con el primer conjunto de claves de la firma de la señal de solicitud, de un tercer conjunto de claves que comprende la al menos una clave del segundo conjunto en común con el primer conjunto, estando cada clave del tercer conjunto asociada a la vez al instante de transición musical asociado a la clave en la firma de la señal de referencia y al instante de transición musical asociado a la clave en la firma de la señal de solicitud;
- 10 - una unidad de determinación de una puntuación en función de una correspondencia entre cada pareja de instantes de transición musical asociada a una clave del tercer conjunto; y
- una unidad de selección de al menos una firma de señal de referencia que tiene una de las puntuaciones más elevadas, en calidad de firma de una señal de referencia candidata para identificar la señal de solicitud.

### Descripción de las figuras

- 15 Otras características y ventajas de la invención se mostrarán con el examen de la descripción detallada a continuación y de los dibujos adjuntos en los que:
- la figura 1 ilustra un procedimiento de generación de una firma de una señal de audio musical de una duración dada según la invención;
  - 20 - la figura 2 presenta una superposición de una función de inicio calculada a partir de una señal de audio musical y de un diagrama de croma determinado a partir de la misma señal de audio musical, según la invención;
  - la figura 3 ilustra un procedimiento de identificación de una señal de audio musical, de solicitud, de entre unas señales de audio musicales, de referencia, estando la señal de audio de solicitud y las señales de audio de referencia representadas por unas firmas respectivas obtenidas por la implementación del procedimiento según la figura 1;
  - 25 - la figura 4 es un diagrama de tiempo/tiempo que representa una nube de puntos según un modo de realización de la invención;
  - la figura 5 ilustra un dispositivo de generación de una firma de una señal de audio musical de una duración dada, según una realización de la invención;
  - 30 - la figura 6 presenta un dispositivo de identificación 600 de una señal de audio musical, de solicitud, de entre unas señales de audio musicales, de referencia, estando la señal de audio de solicitud y las señales de referencia representadas por unas firmas respectivas obtenidas por la implementación del procedimiento según la figura 1, según la invención.

### 35 Descripción detallada de la invención

La figura 1 ilustra un procedimiento de generación de una firma de una señal de audio musical de una duración dada según la invención.

- 40 En una etapa 101, una señal de audio musical es recibida por un dispositivo de generación de firma según la invención, que se ilustrará ulteriormente con referencia a la figura 5. Una señal de audio musical de este tipo puede corresponder a una señal de audio para la que al usuario le gustaría acceder a unos metadatos, tales como el nombre del artista o el título. Una señal de audio de este tipo se llama señal de audio de solicitud. La señal de audio musical puede corresponder, igualmente, a una señal de audio de referencia cuya firma está destinada a estar almacenada en una base de datos con vistas a una comparación ulterior con una firma de señal de audio de solicitud (como se detalla en lo que sigue con referencia en la figura 3) con vistas a una identificación de la señal de audio de solicitud.

- 50 Como se ha explicado anteriormente, la invención tiene como propósito permitir no solamente la identificación cuasi exacta (a saber, la búsqueda eficaz de un extracto de audio en una base importante con robustez a las distorsiones), sino, igualmente, la identificación aproximada (a saber, la aproximación de un extracto de audio con una señal de audio de referencia que le es similar dentro de una base importante).

- 55 A partir de la señal de audio musical recibida, se efectúa una etapa de determinación de estados de la señal de audio musical con referencia a la etapa 102.

- La etapa 102 comprende, en una etapa 103, la determinación de instantes  $t_k$  de transición musical de la señal de audio musical, durante la duración dada de la señal de audio musical.

- 60 En paralelo de la etapa 103, en una etapa 104, la etapa de determinación de estados comprende el modelado de la señal de audio musical para obtener, para cada banda de frecuencias de un conjunto de  $n$  bandas de frecuencias, un diagrama que representa la energía de la señal de audio de solicitud para la banda de frecuencias en cuestión, en función del tiempo durante la duración dada citada anteriormente.

- 65 Las etapas 103 y 104 se comprenderán mejor con referencia a la figura 2, que presenta una superposición de una función de inicio 201 calculada a partir de la señal de audio musical y de un diagrama de croma 205 determinado a partir de la señal de audio musical. Hay que anotar que la determinación del diagrama de croma 205 y de la función

de inicio 201 pueden requerir ambos dos la aplicación de una transformada de Fourier a la señal de audio de solicitud, con unas ventanas diferentes. Con el fin de evitar cualquier desvío temporal entre la función de inicio 201 y el diagrama de croma 205, la invención puede prever sincronizar las transformadas de Fourier.

5 La presente invención puede prever, por ejemplo, detectar los instantes  $t_k$  de transición musical como correspondientes a los máximos locales 202 de la función de inicio.

10 El diagrama de croma 205 está constituido por los  $n$  diagramas 206.i, variando  $i$  de 1 a  $n$ , que ilustra la energía de la señal de audio musical para cada banda de frecuencias. Cuanto más importante es la energía, más claro es el diagrama de croma.

El procedimiento según la invención comprende, a continuación, una etapa 105 de definición de estados de la señal de audio musical. Un estado  $s_k$  está definido por la asociación de un instante  $t_k$  con una información local  $i_k$ :

15 
$$S_k = (t_k, i_k).$$

La información local  $i_k$  puede estar dada por las siguientes características:

20 
$$i_k = (C_{k,l}, C_{k,r})$$

en el que  $c_{k,l}$  es un vector cromático medio a la izquierda de  $t_k$  y  $c_{k,r}$  es un vector cromático medio a la derecha de  $t_k$ .

25 El vector  $c_{k,r}$  es un vector de  $n$  valores respectivamente representativos de la energía de la señal de audio musical en cada uno de los  $n$  diagramas 206 obtenidos en el período 203 entre el instante  $t_k$  de transición musical y un instante siguiente  $t_{k+1}$  de transición musical. El vector  $c_{k,l}$  es un vector de  $n$  valores respectivamente representativos de la energía de la señal de audio de solicitud en cada uno de los  $n$  diagramas 206.i obtenidos en el período 204 entre el instante  $t_k$  de transición musical y un instante anterior  $t_{k-1}$  de transición musical. Por ejemplo, el componente de índice  $i$  del vector  $c_{k,r}$  (anotado  $c_{k,r,i}$  a continuación) corresponde a la media de la energía de la banda de frecuencias representada por el diagrama 206.i sobre el período 203 y el componente de índice  $i$  del vector  $c_{k,l}$  (anotado  $c_{k,l,i}$  a continuación) corresponde a la media de la energía de la banda de frecuencia representada por el diagrama 206.i sobre el período 204.

35 Hay que anotar que la información local  $i_k$  puede corresponder únicamente al vector cromático medio  $c_{k,r}$  a la derecha de  $t_k$  o al vector cromático medio  $c_{k,l}$  a la izquierda de  $t_k$ .

40 La ventaja de la definición de estados presentada más arriba es que la señal de audio de solicitud está representada en una forma compacta y representativa de informaciones rítmicas y armónicas de la señal de audio musical. De este modo, la presente invención permite unas aplicaciones de identificación cuasi exacta, pero, igualmente, de identificación aproximada de una señal de audio de solicitud.

El procedimiento según la invención comprende, a continuación, una etapa 106 de determinación, en función de la información local asociada a cada instante  $t_k$  de transición musical, de una clave asociada al instante  $t_k$  de transición musical, formando las claves determinadas un primer conjunto de claves de la señal de audio musical.

45 Es conveniente generar un primer conjunto de claves que sean invariantes a las variaciones de la señal de audio musical y lo suficientemente discriminatorias. La invariancia a las variaciones debe cubrir las distorsiones clásicas de postprocesamiento (igualación, desvío de tono, etc.), pero, igualmente, las grabaciones de un mismo título en unas condiciones diferentes (correspondencia de señales de audio similares).

50 Según la invención, se puede determinar una clave para cada estado  $s_k$  determinado de la señal de audio musical y puede corresponder a una versión cuantificada de uno de los vectores  $c_{k,l}$  y  $c_{k,r}$ . En lo que sigue, se toma el ejemplo de una versión cuantificada del vector  $c_{k,r}$ . Cada componente  $c_{k,r,i}$  del vector  $c_{k,r}$  se puede, por ejemplo, cuantificar con la ayuda de un solo bit  $\tilde{c}_{k,r,i}$ , de la siguiente manera:

$$\tilde{c}_{k,r,i} = \begin{cases} 1 & \text{si } c_{k,r,i} > \bar{c}_{k,r} \\ 0 & \text{si no} \end{cases}$$

55 con  $\bar{c}_{k,r}$  correspondiente a la media de los  $c_{k,r,i}$  para  $i$  que varía entre 1 y  $n$ .

60 De este modo, la cuantificación del vector  $c_{k,r}$  es determinista. La invención puede prever, igualmente, con el fin de limitar el riesgo de modificar una clave en esta fase del procedimiento, generar un juego probabilístico de vectores  $\tilde{c}_{k,r}$  binarios, asociados a sus probabilidades respectivas.

Cada vector binario  $\tilde{c}_{k,r}$  constituye una clave de la señal de audio musical y está almacenado en asociación con el tiempo  $t_k$  de transición musical.

5 El primer conjunto de claves generadas de este modo y los instantes  $t_k$  respectivamente asociados a estas claves forman, entonces, una firma de la señal de audio musical, en una etapa 107.

10 Una firma de este tipo según la invención contiene unas informaciones rítmicas y armónicas sobre la señal de audio musical, que permite, de este modo, la identificación aproximada, así como la identificación cuasi exacta. Un procedimiento de generación de firma de este tipo se aplica a la vez a la señal de audio de solicitud, pero, igualmente, a las señales de audio de referencia con vistas al almacenamiento de sus firmas respectivas en una base de datos.

15 Cuando el procedimiento descrito anteriormente se aplica a una señal de audio de referencia, se prevé, además, la definición de una función de índice. Se llama función de índice, una función que toma una clave como entrada y que devuelve un juego de valores.

El interés de una operación de este tipo es que es potencialmente realizable con una complejidad  $O(1)$ . En nuestro caso, los valores son unos punteros sobre las señales de audio de referencia de la base de datos y las claves son unas características de audio que caracterizan estas señales de audio de referencia.

20 De este modo, cuando el procedimiento de generación de firma se ha aplicado a todas las señales de audio de referencia, las claves de todas las referencias se extraen y almacenan en el índice. De este modo, los valores de una clave corresponden a los identificadores de las señales de audio de referencia, así como los tiempos  $t_k$  de transición musical correspondientes a la clave en estas señales de audio de referencia.

25 La figura 3 ilustra un procedimiento de identificación de una señal musical, de solicitud, de entre unas señales musicales, de referencia, estando la señal de solicitud y las señales de referencia representadas por unas firmas respectivas obtenidas por la implementación del procedimiento según la figura 1.

30 En una etapa 301, una señal de audio de solicitud es recibida por un dispositivo de identificación que se presentará ulteriormente con referencia a la figura 6.

35 En una etapa 302, se obtiene una firma de la señal de audio de solicitud, por aplicación del procedimiento según la figura 1 a la señal de audio de solicitud. La firma comprende, de este modo, un primer conjunto de claves  $K$  asociadas respectivamente a unos instantes  $t_k$  de transición musical (hay que anotar que el índice  $k$  de los instantes de transición musical, es independiente del símbolo  $K$  que representa una clave).

40 En una etapa 303, se accede a una base de datos que comprende las firmas de las señales de audio de referencia, con vistas a constituir, con la ayuda de la función de índice y para cada firma de señal de referencia que tiene al menos una clave del segundo conjunto en común con el primer conjunto de claves de la firma de la señal de solicitud, un tercer conjunto de claves que comprende las claves del segundo conjunto en común con el primer conjunto. Hay que anotar que una clave  $K$  del tercer conjunto puede estar asociada a varios instantes de transición musical en el primer conjunto y/o a varios instantes de transición musical en el segundo conjunto.

45 El procedimiento comprende, a continuación, una etapa 304 de determinación de una puntuación en función de una correspondencia entre cada pareja de instantes de transición musical asociada a una clave del tercer conjunto. La determinación de la puntuación se efectúa para cada firma de señal de referencia que comprende al menos una clave en común con el primer conjunto de claves de la firma de la señal de audio de solicitud. La determinación de la puntuación puede consistir, por ejemplo, en la determinación de una función afín que pone en correspondencia un mayor número de instantes de transición musical de la señal de referencia asociados a las claves del tercer conjunto con los instantes de transición musical de la señal de solicitud asociados a las mismas claves del tercer conjunto, siendo la puntuación atribuida a la firma de la señal de referencia igual al mayor número citado anteriormente.

55 Un ejemplo detallado de determinación de la puntuación de una firma de señal de referencia, a partir del tercer conjunto de claves común a esta firma de señal de referencia y a la firma de la señal de solicitud, se ilustra con referencia a la figura 4.

60 La figura 4 es un diagrama de tiempo/tiempo que representa una nube de puntos 403, correspondiendo cada punto a una clave del tercer conjunto, siendo la abscisa 401 del punto el instante de transición musical de la señal de referencia asociada a la clave del tercer conjunto y correspondiendo la ordenada 402 del punto al instante de transición musical de la señal de solicitud asociada a la clave del tercer conjunto. De este modo, se construye una nube de puntos por firma de señal de referencia que comparte al menos una clave en común con la firma de la señal de solicitud.

65 A partir de esta nube de puntos, se determina una recta 404 que conecta un máximo de puntos de la nube de puntos construida. La determinación de la recta 404 se puede efectuar por aplicación de un algoritmo de Hough al conjunto de parejas de instantes de transición musical asociadas a las claves del tercer conjunto.



- La puntuación de la firma de la señal de referencia para la que se ha construido la nube de puntos puede ser, entonces, igual al número de puntos conectados por la recta 404 determinada en la nube de puntos. En la figura 4, se observa que se obtiene una recta 404 que comprende un gran número de puntos. El hecho de que unos numerosos puntos estén situados fuera de la recta se explica por el hecho de que una clave del primer conjunto puede estar presente
- 5 varias veces en el segundo conjunto. Esto traduce la repetición de una clave en el título representado por la señal de audio de referencia. Además, hay que tener en cuenta que, debido a la cuantificación muy fuerte aplicada para determinar las claves, estas últimas no son muy discriminatorias (lo que es voluntario y útil para permitir la identificación aproximada de la señal de audio de solicitud).
- 10 En una etapa 305, se selecciona al menos una firma de señal de referencia que tiene una de las puntuaciones más elevadas, en calidad de firma de una señal de referencia candidata para corresponder a la señal de solicitud. Por ejemplo, la invención puede prever seleccionar M firmas de señales de referencia candidatas, pudiendo M estar predeterminado o pudiendo depender del número de claves del primer conjunto de claves de la firma de la señal de solicitud, por ejemplo. No hay ninguna restricción vinculada al valor de M.
- 15 En una etapa 306, se selecciona un juego de firmas de señales de referencia que tienen unas puntuaciones más elevadas y se aplica un postprocesamiento:
- a unas señales de referencia candidatas, representadas por las firmas del juego y
  - 20 - a la señal de solicitud,
- con vistas a identificar la señal de solicitud como una de las señales de referencia. El juego de firmas puede comprender, por ejemplo, las M firmas citadas anteriormente.
- 25 Una etapa de postprocesamiento de este tipo puede implementar unos cálculos complejos, sin, por ello, movilizar unos recursos demasiado importantes, en la medida en que la base de datos de señales de referencia se ha reducido a M señales de referencia candidatas.
- Un ejemplo de postprocesamiento se describe a continuación. No obstante, la invención no está restringida en modo alguno a este ejemplo y cualquier método conocido se podrá aplicar, con el fin de seleccionar una de las señales de audio de referencia candidatas como que identifica la señal de audio de solicitud. El estado de la técnica propone unas numerosas técnicas a estos efectos.
- 30 En esta fase del procedimiento, es conveniente comparar las M señales de audio de referencia seleccionadas con la señal de audio de solicitud.
- 35 A título de ejemplo, a continuación, se propone un enfoque de programación dinámica que se basa en el modelado de señales de audio por estados, tal como se ha descrito anteriormente.
- 40 En primer lugar, para cada firma de señal de referencia que posee unas claves en común con el primer conjunto de claves, la señal de audio de solicitud se puede modificar en función de la función afín determinada para la obtención de la puntuación de la firma de la señal de referencia.
- En efecto, la recta 404 define una función afín caracterizada por una pendiente con referencia 405 en la figura 4 y anotada a y por una ordenada en el origen con referencia 406 y anotada b. La recta 404 significa que una clave extraída de la señal de audio de solicitud en un instante t, corresponderá a una clave en la señal de audio de referencia en el instante a(t-b).
- 45 Lo inverso de a representa, entonces, una dilatación temporal de la señal de audio de solicitud, mientras que b representa el instante en la señal de audio de referencia en el que comienza la señal de audio de solicitud. De este modo, la señal de audio de solicitud se puede modificar por puesta a escala de la señal de solicitud en función de la pendiente a y por reencuadre temporal de la señal de audio de solicitud en función de la ordenada en el origen b.
- 50 La señal de audio de solicitud modificada de este modo está representada por la secuencia de estados  $\{s_1...s_i...s_m\}$ , siendo m el número de instantes de transición musical determinado anteriormente en el intervalo de tiempo correspondiente a la duración de la señal de audio de la solicitud. Para representar los estados de las señales de referencia candidatas, se utiliza en lo que sigue la anotación  $\{\sigma_1... \sigma_j... \sigma_p\}$ , siendo p el número de instantes de transición musical en la señal de referencia candidata considerada, siendo p dependiente de la señal de referencia candidata.
- 55 Con el fin de alinear dinámicamente dos secuencias de estados, es conveniente definir tres tipos de penalizaciones:
- $f^d(s_i)$ : penalización asignada para la supresión del estado  $s_i$ ;
  - $f^i(s_i)$ : penalización asignada a la inserción del estado  $s_i$ ;
  - $f^s(s_i, \sigma_j)$ : penalización asignada a la sustitución del estado  $s_i$  por el estado  $\sigma_j$ .
- 60 A título de ejemplo, las funciones  $f^d$  y  $f^i$  pueden ambas dos tomar un valor constante igual a 0,3.
- 65

Sabiendo que  $s_i = (t_i, (C_{i,l}, C_{i,r}))$  y  $\sigma_j = (T_j, (\gamma_{j,l}, \gamma_{j,r}))$ , anotando  $T_j$  el instante de transición musical de índice  $j$  en la señal de referencia,  $\gamma_{j,l}$  el vector cromático medio a la izquierda de  $T_j$  y  $\gamma_{j,r}$  el vector cromático medio a la derecha de  $T_j$ , se define  $f^s$  por:

5

$$f^s(s_i, \sigma_j) = \cos(C_{i,l}, \gamma_{j,l}) \cdot \cos(C_{i,r}, \gamma_{j,r}) \cdot e^{-\frac{|t_i - T_j|}{c}}$$

10 El primer término de coseno penaliza la semejanza del vector cromático medio a la izquierda de  $t_i$  con el vector cromático medio a la izquierda de  $T_j$ . El segundo término de coseno penaliza la semejanza del vector cromático medio a la derecha de  $t_i$  con el vector cromático medio a la derecha de  $T_j$ . El término exponencial penaliza el error temporal entre la ocurrencia del estado  $s_i$  en la señal de solicitud modificada y la ocurrencia del estado  $\sigma_j$  en la señal de referencia.

15 Una programación dinámica consiste en llenar de manera iterativa una matriz  $D$ . Para cada par  $(i, j) \in \{1..m\} \times \{1..p\}$ ,  $D(i, j)$  contiene la puntuación de la alineación de la subsecuencia  $\{s_1..s_i\}$  con la subsecuencia  $\{\sigma_1.. \sigma_j\}$ .  $D$  se calcula de la siguiente manera:

$$D(i, j) = \max \left\{ \begin{array}{l} D(i, j-1), f^d(\sigma_j) \\ D(i-1, j-1), f^s(s_i, \sigma_j) \\ D(i-1, j), f^i(s_i) \end{array} \right\}$$

20 La puntuación de la alineación de  $\{s_1..s_m\}$  con  $\{\sigma_1.. \sigma_p\}$  está dada finalmente por  $D(m, p)$ .

25 En una etapa 307, al término del postprocesamiento, la señal de audio de solicitud se identifica de manera cuasi exacta o aproximada como correspondiente a la señal de audio de referencia que tiene la puntuación  $D(m, p)$  más elevada. Una identificación de este tipo puede, por otra parte, estar condicionada por una etapa de comparación con un umbral predeterminado: si la puntuación  $D(m, p)$  más elevada es inferior a este umbral, no se obtiene ninguna identificación para la señal de audio de solicitud.

30 La figura 5 presenta un dispositivo 500 de generación de una firma de una señal de audio musical de una duración dada.

A estos efectos, el dispositivo 500 comprende una unidad de recepción 501 de una señal de audio musical. Como se ha descrito anteriormente, la señal de audio musical puede ser una señal de audio de solicitud o de referencia.

35 El dispositivo 500 comprende, además, una unidad de modelado 502 de la señal de audio musical para obtener, para cada banda de frecuencias de un conjunto de  $n$  bandas de frecuencias, un diagrama que representa la energía de la señal de audio para la banda de frecuencias, en función del tiempo durante la duración dada. Los diagramas obtenidos pueden estar agrupados en forma del diagrama de croma 205 ilustrado en la figura 2. En particular, la unidad de modelado 502 es adecuada para implementar la etapa 104 del procedimiento ilustrado en la figura 1.

40 El dispositivo 500 comprende una primera unidad de determinación 502 de instantes  $t_k$  de transición musical de la señal de audio durante la duración dada, que es adecuada para implementar la etapa 103 del procedimiento ilustrado en la figura 1.

45 Con el fin de sincronizar las operaciones efectuadas por las unidades 502 y 503, el dispositivo 500 comprende una unidad de sincronización 504 adecuada para implementar la etapa 104 del procedimiento ilustrado en la figura 1.

50 El dispositivo 500 comprende una unidad de asociación 505 de cada instante  $t_k$  de transición musical con una información local que comprende un vector de  $n$  valores respectivamente representativos de la energía de la señal de audio en cada uno de los  $n$  diagramas obtenidos entre el instante  $t_k$  de transición musical y un instante siguiente  $t_{k+1}$  de transición musical y/o un vector de  $n$  valores respectivamente representativos de la energía de la señal de audio en cada uno de los  $n$  diagramas obtenidos entre el instante  $t_k$  de transición musical y un instante anterior  $t_{k-1}$  de transición musical. La unidad de asociación 505 es adecuada para implementar la etapa 105 del procedimiento ilustrado en la figura 1.

55 El dispositivo 500 comprende, además, una segunda unidad de determinación 506, en función de la información local asociada a cada instante  $t_k$  de transición musical, de una clave asociada al instante de transición musical, formando las claves determinadas un primer conjunto de claves de la señal de audio. La unidad de determinación 506 es

adecuada para implementar la etapa 106 del procedimiento ilustrado en la figura 1.

5 El dispositivo 500 comprende una unidad de generación 507 de una firma de la señal de audio musical que comprende unos pares de claves del primer conjunto de claves y de instantes  $t_k$  de transición musical asociados, que es adecuada para implementar la etapa 107 del procedimiento ilustrado en la figura 1. La unidad de generación 507 puede, además, ser adecuada para transmitir la señal de audio musical, así como la firma generada a una entidad remota.

10 Por ejemplo, cuando la señal de audio musical es una señal de audio de solicitud, se puede transmitir con la firma generada a un servidor remoto a cargo de identificar la señal de audio de solicitud o, más generalmente, a un dispositivo según la figura 6. En este caso, el dispositivo 500 puede estar insertado en un terminal de usuario, tal como un terminal móvil de telecomunicaciones, un ordenador portátil o fijo, etc.

15 Como variante, cuando la señal de audio musical es una señal de audio de referencia, se puede transmitir con la firma generada a una base de datos de referencia adecuada para almacenar la señal de audio de referencia en asociación con la firma generada.

20 La figura 6 ilustra un dispositivo de identificación 600 de una señal musical, de solicitud, de entre unas señales musicales, de referencia, estando la señal de solicitud y las señales de referencia representadas por unas firmas respectivas obtenidas por la implementación del procedimiento según la figura 1. A estos efectos, el dispositivo 600 puede estar conectado al dispositivo 500 y, de este modo, puede recibir por una unidad de recepción 601 una señal de audio de solicitud acompañada de una firma. En este caso, una unidad 602 extrae la firma de la señal de audio de solicitud para comunicación a una unidad de constitución 603, descrita en lo que sigue. Como variante, la unidad de recepción 601 puede recibir una señal de audio de solicitud sin firma, estando la firma generada por una unidad 602. La unidad 602 puede, de este modo, corresponder al dispositivo de generación de firma 500 y puede transmitir la firma generada a la unidad de constitución 603.

25 Además, el dispositivo 600 puede estar conectado a una base de datos 604 que almacena unas firmas asociadas respectivamente a unas señales de audio de referencia. En la base de datos 604, cada firma de señal de referencia comprende, de este modo, un segundo conjunto de claves a las que están asociados unos instantes respectivos  $T_k$  de transición musical. Según la invención, la base de datos 604 puede estar integrada en el dispositivo 600.

30 El dispositivo 600 comprende la unidad de constitución 603, que accede a la base de datos 604 y que, para cada firma de señal de referencia que tiene al menos una clave del segundo conjunto en común con el primer conjunto de claves de la firma de la señal de solicitud, constituye un tercer conjunto de claves que comprende las claves del segundo conjunto en común con el primer conjunto, estando cada clave del tercer conjunto asociada a la vez al instante de transición musical asociado a la clave en la firma de la señal de referencia y al instante de transición musical asociado a la clave en la firma de la señal de solicitud. La unidad de constitución 603 es adecuada para implementar la etapa 303 del procedimiento según la figura 3.

35 El dispositivo 600 comprende, además, una unidad de determinación 605 de una puntuación en función de una correspondencia entre cada pareja de instantes de transición musical asociada a una clave del tercer conjunto, que es adecuada para implementar la etapa 304 del procedimiento según la figura 3.

40 El dispositivo 600 comprende una unidad de selección 606 de al menos una firma de señal de referencia que tiene una de las puntuaciones más elevadas, en calidad de firma de una señal de referencia candidata para identificar la señal de solicitud. La unidad de selección 606 es adecuada para implementar la etapa 305 del procedimiento según la figura 3.

45 El dispositivo comprende, además, una unidad de postprocesamiento 607 adecuada para implementar la etapa 306 del procedimiento según la figura 3, con el fin de determinar de entre un conjunto de señales de referencia candidatas, la señal de audio de referencia que identifica la señal de audio de solicitud. Además, la unidad 607 puede no devolver ningún resultado cuando la puntuación más elevada de una señal de audio de referencia es inferior al umbral introducido anteriormente.

50 El dispositivo 600 comprende una unidad de transmisión 608 para devolver la señal de audio de referencia que identifica la señal de audio de solicitud, a un terminal de usuario o a una unidad de visualización del dispositivo 600, no presentada en la figura 6.

55 La invención prevé, igualmente, el caso en que el dispositivo 600 no comprende la unidad 306. En este caso, la unidad de transmisión 608 transmite las señales de solicitudes de candidatos a una entidad remota adecuada para comprender la unidad 306, lo que es particularmente ventajoso en los casos en que el dispositivo 600 posee unas capacidades de cálculo reducidas.

60 Hay que anotar que el dispositivo 500 según la figura 5 y el dispositivo 600 según la figura 6 pueden estar agrupados dentro de un mismo dispositivo global (cuando el dispositivo 500 sustituye a la unidad 602 del dispositivo 600). El dispositivo global se puede utilizar, entonces, en un terminal de usuario o en un servidor remoto accesible desde un

terminal de usuario. En este último caso, el terminal de usuario transmite al servidor remoto una señal de solicitud.

La presente invención prevé, igualmente, un programa de ordenador que incluye unas instrucciones para la implementación del procedimiento ilustrado en la figura 1, cuando este programa se ejecuta por un procesador y un  
5 programa de ordenador que incluye unas instrucciones para la implementación del procedimiento ilustrado en la figura 3, cuando este programa se ejecuta por un procesador.

**REIVINDICACIONES**

1. Procedimiento de generación de una firma de una señal de audio musical de una duración dada, comprendiendo el procedimiento las siguientes etapas:

- 5 - modelado (104) de dicha señal de audio musical para obtener, para cada banda de frecuencias de un conjunto de n bandas de frecuencias, siendo n estrictamente superior a 1, un diagrama que representa la energía de la señal de audio para dicha banda de frecuencias, en función del tiempo durante dicha duración dada;
- 10 - determinación (103) de instantes  $t_k$  de transición musical de la señal de audio durante la duración dada, los instantes  $t_k$  de transición musical corresponden a los máximos locales de una función de inicio calculada a partir de la señal de audio musical;
- 15 - asociación (105) de cada instante  $t_k$  de transición musical con una información local que comprende un vector de n valores respectivamente representativos de la energía de la señal de audio en cada uno de los n diagramas obtenidos entre el instante  $t_k$  de transición musical y un instante siguiente  $t_{k+1}$  de transición musical y/o un vector de n valores respectivamente representativos de la energía de la señal de audio en cada uno de los n diagramas obtenidos entre el instante  $t_k$  de transición musical y un instante anterior  $t_{k-1}$  de transición musical;
- 20 - determinación (106), en función de la información local asociada a cada instante  $t_k$  de transición musical, de una clave asociada a dicho instante de transición musical, formando las claves determinadas un primer conjunto de claves de la señal de audio;
- generación (107) de una firma de la señal de audio musical que comprende unos pares, comprendiendo cada par una clave del primer conjunto de claves y un instante  $t_k$  de transición musical asociado,

en el que las claves del primer conjunto de claves se obtienen por cuantificación de los vectores de n valores respectivamente representativos de la energía de la señal en cada banda de frecuencias del conjunto de n bandas de frecuencias en un período comprendido entre el instante  $t_k$  de transición musical y un instante siguiente  $t_{k+1}$  de transición musical o por cuantificación de los vectores de n valores respectivamente representativos de la energía de la señal en cada banda de frecuencias del conjunto de n bandas de frecuencias en un período comprendido entre el instante  $t_k$  de transición musical y un instante anterior  $t_{k-1}$  de transición musical.

2. Procedimiento según la reivindicación 1, en el que cada clave del primer conjunto de claves es un vector de n bits, estando cada bit de índice i determinado por comparación entre el valor de energía de índice i en el vector de n valores a partir del que se determina dicha clave y la media de los valores de dicho vector de n valores.

3. Procedimiento según la reivindicación 2, en el que, si el valor de energía de índice i en el vector de n valores a partir del que se determina la clave es superior a la media de los valores de dicho vector de n valores, entonces, el bit de índice i es igual a 1, siendo el bit de índice i igual a 0 si no.

4. Procedimiento de identificación de una señal musical, de solicitud, de entre unas señales musicales, de referencia, estando la señal de solicitud y las señales de referencia representadas por unas firmas respectivas obtenidas por la implementación del procedimiento según una de las reivindicaciones 1 a 3, comprendiendo, de este modo, cada firma de señal de referencia un segundo conjunto de claves a las que están asociados unos instantes respectivos  $t_k$  de transición musical, comprendiendo dicho procedimiento las etapas:

- para cada firma de señal de referencia que tiene al menos una clave del segundo conjunto en común con el primer conjunto de claves de la firma de la señal de solicitud, constitución (303) de un tercer conjunto de claves que comprende la al menos una clave del segundo conjunto en común con el primer conjunto, estando cada clave del tercer conjunto asociada a la vez al instante de transición musical asociado a la clave en la firma de la señal de referencia y al instante de transición musical asociado a la clave en la firma de la señal de solicitud;
- determinación (304) de una puntuación en función de una correspondencia entre cada pareja de instantes de transición musical asociada a una clave del tercer conjunto; y
- selección (305) de al menos una firma de señal de referencia que tiene una de las puntuaciones más elevadas, en calidad de firma de una señal de referencia candidata para corresponder a la señal de solicitud.

5. Procedimiento según la reivindicación 4, en el que se selecciona un juego de firmas de señales de referencia que tienen unas puntuaciones más elevadas y se aplica un postprocesamiento (306):

- a unas señales de referencia candidatas, representadas por las firmas de dicho juego y
- a la señal de solicitud,

y que comprende una etapa de identificación de la señal de solicitud como una de las señales de referencia.

6. Procedimiento según una de las reivindicaciones 4 y 5, en el que la determinación (304) de la puntuación de una firma de señal de referencia comprende las siguientes etapas:

- determinación de una función afín que pone en correspondencia un mayor número de instantes de transición musical de la señal de referencia asociados a las claves del tercer conjunto con los instantes de transición musical

de la señal de solicitud asociados a las claves respectivas del tercer conjunto;

en el que la puntuación atribuida a la firma de la señal de referencia es igual a dicho mayor número.

5 7. Procedimiento según la reivindicación 6, en el que la función afín se obtiene por aplicación de una transformada de Hough al conjunto de parejas de instantes de transición musical asociados a las claves del tercer conjunto.

8. Procedimiento según una de las reivindicaciones 4 a 7, en el que la determinación (304) de la puntuación de una señal de audio de referencia comprende las siguientes etapas:

10 - construcción de una nube de puntos (403), correspondiendo cada punto a una clave del tercer conjunto, siendo la abscisa (401) del punto el instante de transición musical de la señal de referencia asociada a la clave del tercer conjunto y correspondiendo la ordenada (402) del punto al instante de transición musical de la señal de solicitud asociada a la clave del tercer conjunto;

15 - determinación de una recta (404) que conecta un máximo de puntos de la nube de puntos construida;  
- determinación de la puntuación de la firma de la señal de referencia como que es igual al número de puntos conectados por dicha recta determinada en la nube de puntos.

20 9. Procedimiento según la reivindicación 6 o 7, en el que el postprocesamiento (306) se aplica sucesivamente para cada señal de referencia correspondiente a una firma seleccionada y a la señal de solicitud modificada en función de la función afín determinada para la obtención de la puntuación de la firma de la señal de referencia.

25 10. Procedimiento según la reivindicación 9, en el que la señal de solicitud se modifica por puesta a escala de dicha señal de solicitud en función de una pendiente (405) de la función afín y por reencuadre temporal de dicha señal de solicitud en función de una ordenada en el origen (406) de la función afín.

11. Programa de ordenador que incluye unas instrucciones para la implementación del procedimiento según una de las reivindicaciones 1 a 3, cuando este programa se ejecuta por un procesador.

30 12. Programa de ordenador que incluye unas instrucciones para la implementación del procedimiento según una de las reivindicaciones 4 a 10, cuando este programa se ejecuta por un procesador.

13. Dispositivo de generación de una firma de una señal de audio musical de una duración dada, comprendiendo el dispositivo (500):

35 - una unidad de modelado (503) de dicha señal de audio musical para obtener, para cada banda de frecuencias de un conjunto de  $n$  bandas de frecuencias, siendo  $n$  estrictamente superior a 1, un diagrama que representa la energía de la señal de audio para dicha banda de frecuencias, en función del tiempo durante dicha duración dada;

40 - una primera unidad de determinación (502) de instantes  $t_k$  de transición musical de la señal de audio durante la duración dada, los instantes  $t_k$  de transición musical corresponden a los máximos locales de una función de inicio calculada a partir de la señal de audio musical;

45 - una unidad de asociación (505) de cada instante  $t_k$  de transición musical con una información local que comprende un vector de  $n$  valores respectivamente representativos de la energía de la señal de audio en cada uno de los  $n$  diagramas obtenidos entre el instante  $t_k$  de transición musical y un instante siguiente  $t_{k+1}$  de transición musical y/o un vector de  $n$  valores respectivamente representativos de la energía de la señal de audio en cada uno de los  $n$  diagramas obtenidos entre el instante  $t_k$  de transición musical y un instante anterior  $t_{k-1}$  de transición musical;

50 - una segunda unidad de determinación (506), en función de la información local asociada a cada instante  $t_k$  de transición musical, de una clave asociada a dicho instante de transición musical, formando las claves determinadas un primer conjunto de claves de la señal de audio;

- una unidad de generación (507) de una firma de la señal de audio musical que comprende unos pares, comprendiendo cada par una clave del primer conjunto de claves y un instante  $t_k$  de transición musical asociado,

en el que las claves del primer conjunto de claves se obtienen por cuantificación de los vectores de  $n$  valores respectivamente representativos de la energía de la señal en cada banda de frecuencias del conjunto de  $n$  bandas de frecuencias en un período comprendido entre el instante  $t_k$  de transición musical y un instante siguiente  $t_{k+1}$  de transición musical o por cuantificación de los vectores de  $n$  valores respectivamente representativos de la energía de la señal en cada banda de frecuencias del conjunto de  $n$  bandas de frecuencias en un período comprendido entre el instante  $t_k$  de transición musical y un instante anterior  $t_{k-1}$  de transición musical.

60 14. Dispositivo de identificación de una señal musical, de solicitud, de entre unas señales musicales, de referencia, estando la señal de solicitud y las señales de referencia representadas por unas firmas respectivas obtenidas por la implementación del procedimiento según una de las reivindicaciones 1 a 3, comprendiendo, de este modo, cada firma de señal de referencia un segundo conjunto de claves a las que están asociados unos instantes respectivos  $\tau_k$  de transición musical, comprendiendo dicho dispositivo (600):

65 - una unidad de constitución (603), para cada firma de señal de referencia que tiene al menos una clave del

segundo conjunto en común con el primer conjunto de claves de la firma de la señal de solicitud, de un tercer conjunto de claves que comprende la al menos una clave del segundo conjunto en común con el primer conjunto, estando cada clave del tercer conjunto asociada a la vez al instante de transición musical asociado a la clave en la firma de la señal de referencia y al instante de transición musical asociado a la clave en la firma de la señal de solicitud;

5 - una unidad de determinación (604) de una puntuación en función de una correspondencia entre cada pareja de instantes de transición musical asociada a una clave del tercer conjunto; y

- una unidad de selección (605) de al menos una firma de señal de referencia que tiene una de las puntuaciones más elevadas, en calidad de firma de una señal de referencia candidata para identificar la señal de solicitud.

10

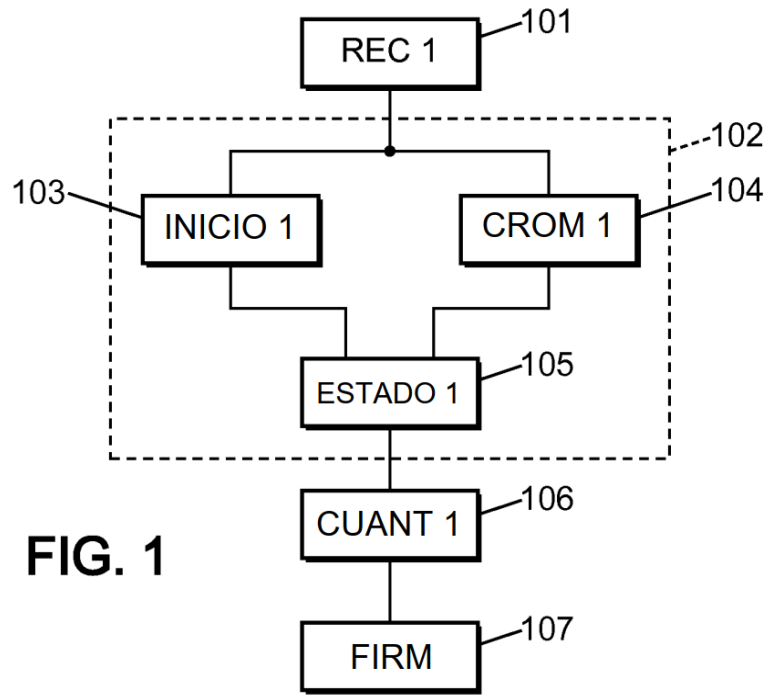
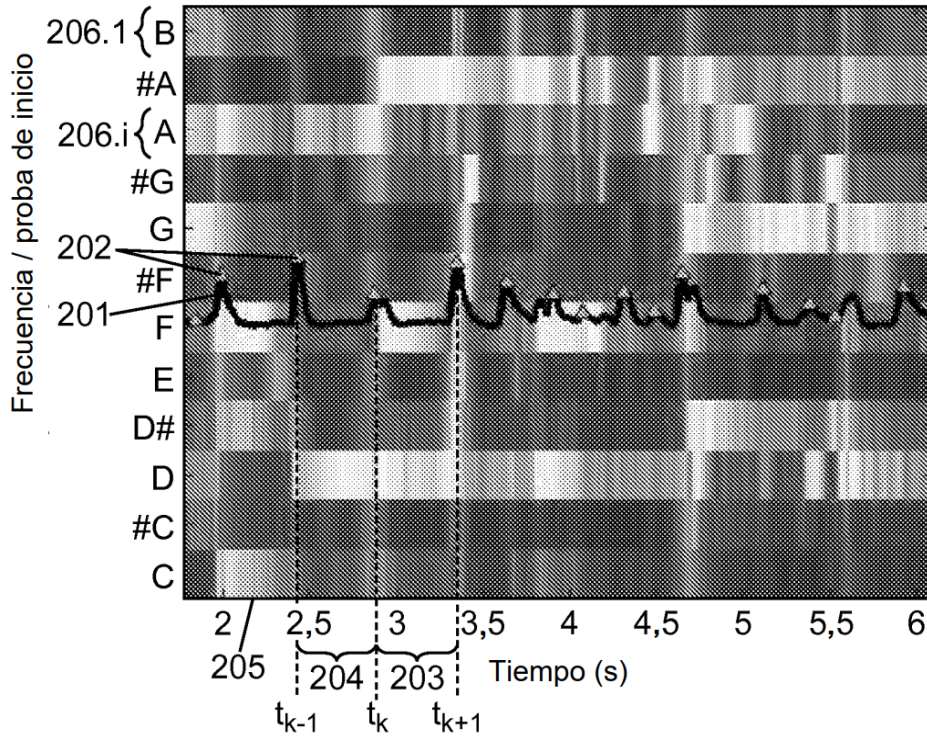


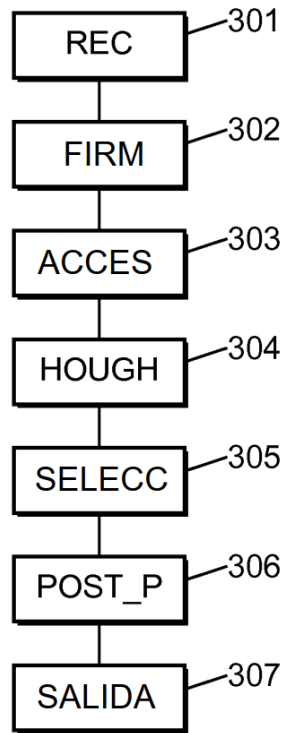
FIG. 1

FIG. 2

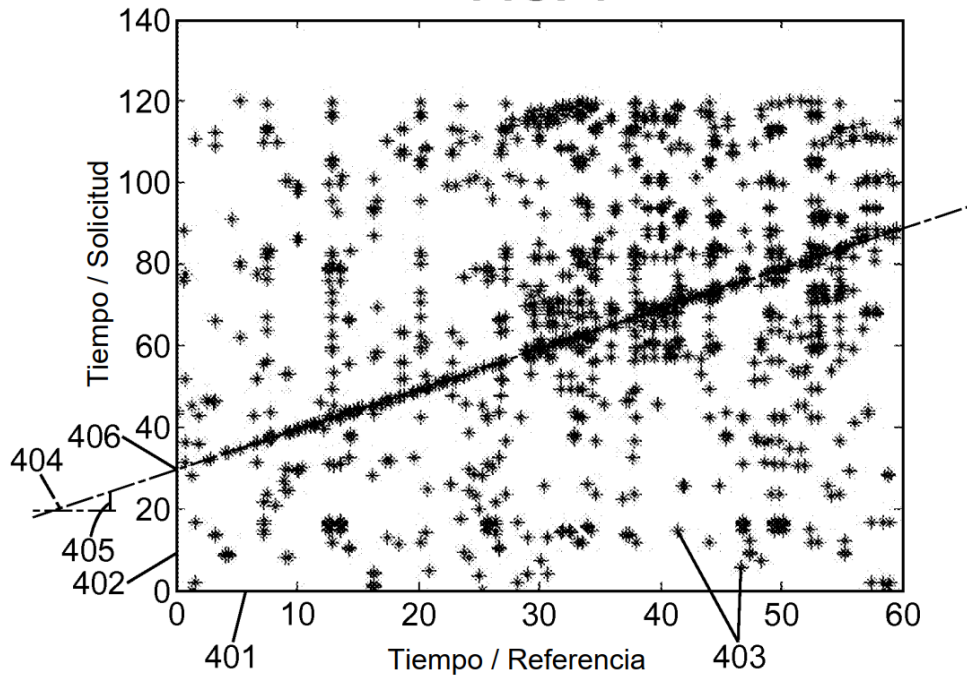




**FIG. 3**



**FIG. 4**



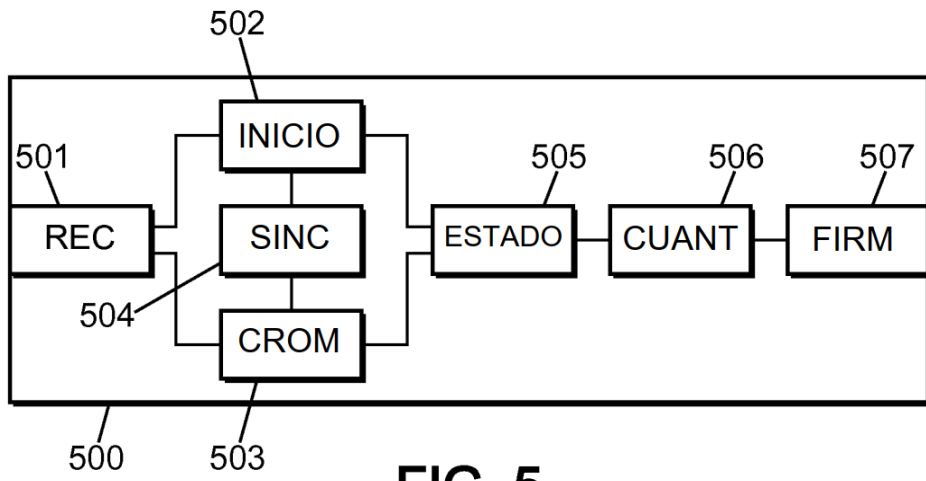


FIG. 5

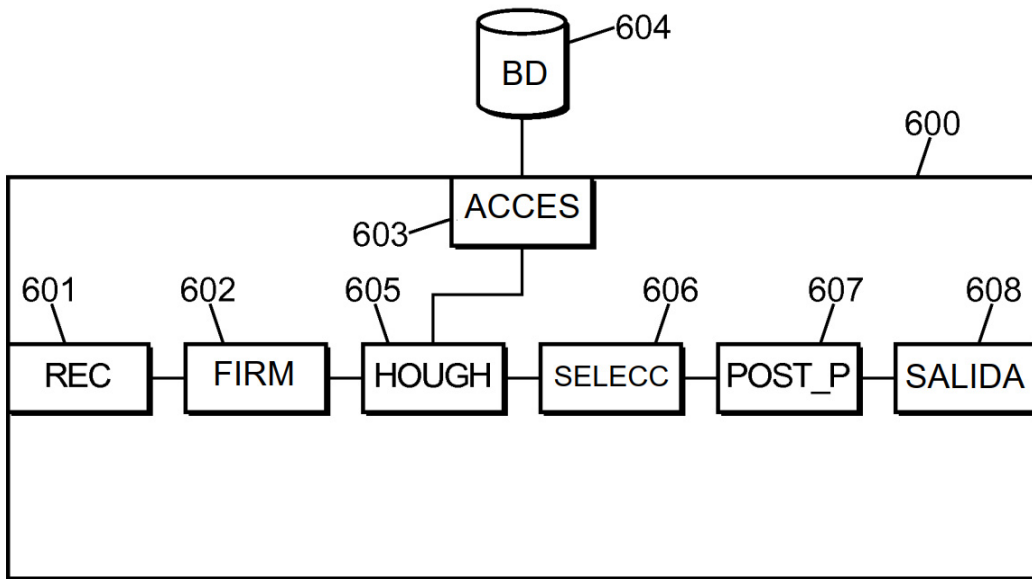


FIG. 6