

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **3 010 466**

51 Int. Cl.:

G10H 1/00 (2006.01)

G10H 1/36 (2006.01)

G10H 1/38 (2006.01)

G10H 1/44 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **28.04.2023 E 23170706 (8)**

97 Fecha y número de publicación de la concesión europea: **06.11.2024 EP 4270373**

54 Título: **Método para identificar una canción**

30 Prioridad:

28.04.2022 FI 20227059

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
03.04.2025

73 Titular/es:

**YOUSICIAN OY (100.00%)
Siltasaarekatu 16
00530 Helsinki, FI**

72 Inventor/es:

**THÜR, CHRISTOPH;
KLAPURI, ANSSI y
KESTI, VELI-JUSSI**

74 Agente/Representante:

ALMARAZ PALMERO, Agustín

ES 3 010 466 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

Método para identificar una canción

5 Campo técnico

La presente divulgación se refiere en general a métodos y sistemas implementados por ordenador. Más concretamente, la presente divulgación se refiere a un método implementado por ordenador para identificar una canción que el usuario desea interpretar y a un sistema asociado.

10

Antecedentes

Esta sección ilustra información de antecedentes útil sin admitir ninguna técnica descrita en el presente documento que sea representativa del estado de la técnica.

15

En una solución convencional, el usuario selecciona una canción y, a continuación, opera la interfaz de usuario para que la solución, tal como una aplicación móvil, proporcione al usuario la notación musical de la canción o empiece a reproducir la canción o la pista de acompañamiento de la canción. A continuación, el usuario puede tocar con la pista de acompañamiento. Se conoce el documento WO 2018/112260 A1, que divulga un sistema de búsqueda y emparejamiento en el que una melodía tocada en un piano se empareja dentro de una base de datos, y el sistema selecciona en consecuencia un ejercicio musical o un acompañamiento musical. El sistema registra además datos históricos de ejercicios reproducidos previamente y propone un modo en el que solo se utilizan ejercicios vocales o acompañamientos musicales reproducidos previamente para el algoritmo de búsqueda/concordancia.

20

25

Tocar una canción es, sin embargo, a menudo, una experiencia algo personal, emocional y expresiva. Idealmente, el intérprete conoce la canción de memoria y es capaz de tocar con el corazón y cantar. Lo siguiente mejor es tocar con la ayuda de la letra y de tablas de acordes o tablatura. Las cosas que distraen de esta experiencia son, por ejemplo, desplazarse por las letras de las canciones mientras se toca, lo que hace que el usuario retire la mano del instrumento, o navegar por listas de canciones que obligan al usuario a cambiar de la mentalidad de tocar a la de analizar.

30

Sin embargo, la detección automática de la canción a partir de una interpretación realizada libremente por el usuario, tal como un usuario tocando un instrumento a su propio tempo, y la elección de armonías sin adherirse a los detalles de un arreglo/notación predefinidos de la canción es difícil. Cuando toca libremente, el intérprete suele leer una notación musical basada en acordes que describe la progresión armónica de la canción con nombres de acordes, símbolos de acordes o diagramas de acordes sin detallar la disposición nota a nota de la canción ni los patrones rítmicos. Tocar libremente es muy común cuando se toca la parte de un instrumento de acompañamiento en una canción. El reto de detectar una canción concreta a partir de esa reproducción libre es que la cantidad de información contenida en unos pocos acordes normalmente no es suficiente para desambiguar la canción de entre una gran colección de canciones.

35

40

Sumario de la invención

Las reivindicaciones adjuntas definen el alcance de la protección. Cualquiera de los ejemplos y descripciones técnicas de aparatos, los productos y/o métodos de la descripción y/o dibujos no cubiertos por las reivindicaciones no se presentan como realizaciones de la invención, sino como técnica anterior o ejemplos útiles para comprender la invención.

45

Un objetivo de la invención es que un usuario que utilice el sistema y el método de acuerdo con la presente divulgación no tenga que hacer clic en una interfaz de usuario o en menús de una aplicación para elegir la canción o canciones que desea interpretar. En su lugar, el usuario puede simplemente empezar a tocar, y el sistema y el método reconocen la canción a partir del historial de reproducción del usuario y de la interpretación de reproducción actual del usuario, que puede utilizarse para mostrar automáticamente al usuario música escrita y/o letras de canciones. Además, el sistema y el método pueden proporcionar audio de pista de acompañamiento referente a la canción reconocida al tempo del usuario, y en la posición en la que el usuario se encuentra en la canción.

50

55

De acuerdo con un primer aspecto de ejemplo, se proporciona un método implementado por ordenador para identificar una canción que comprende:

- proporcionar información sobre la notación musical y la progresión armónica de las canciones,
- recibir una señal de audio en tiempo real de un usuario que interpreta en un instrumento,
- detectar la actividad de reproducción en segmentos sucesivos del usuario en segmentos sucesivos a partir de la señal de audio en tiempo real,
- detectar notas y/o acordes a partir de la señal de audio en tiempo real,
- almacenar información del historial de reproducción del usuario que comprende información sobre las canciones que un usuario ha tocado anteriormente y la frecuencia con la que el usuario ha tocado dichas canciones, opcionalmente, adicionalmente, en relación con el momento en que el usuario ha tocado dichas canciones,

60

65

- basándose en la información del historial de reproducción, calcular una primera probabilidad para al menos una canción,
- basándose en un número de primeras probabilidades para un número de canciones y, basándose en la actividad de reproducción detectada y en las notas y/o acordes detectados, estimar la canción que el usuario está interpretando, en donde

5
10 o dicha estimación comprende el cálculo de una segunda probabilidad para un número de canciones diferentes, en donde las segundas probabilidades están definidas por la señal de audio en tiempo real correspondiente a una canción particular del historial de reproducción en combinación con la primera probabilidad asociada a dicha canción, y

- proporcionar la canción que el usuario está interpretando o información referente a la canción.

De acuerdo con un segundo aspecto de ejemplo se proporciona un sistema o aparato que comprende:

15 un almacenamiento para mantener un documento musical que defina cómo deben tocarse las diferentes partes en una pieza musical;
una pantalla configurada para mostrar una parte del documento musical o información referente a una canción o versión de una canción cuando un usuario toca la pieza musical;
20 una entrada para recibir una señal de audio en tiempo real de música tocada por el usuario;
al menos un procesador configurado para realizar al menos:

-- proporcionar un conjunto de datos de audio que comprenda información de notación musical para canciones,

- recibir una señal de audio en tiempo real de un usuario que interpreta en un instrumento,
- detectar la actividad de reproducción en segmentos sucesivos del usuario en segmentos sucesivos a partir de la señal de audio en tiempo real,
- detectar notas y/o acordes a partir de la señal de audio en tiempo real,
- almacenar información del historial de reproducción del usuario que comprende información sobre las canciones que un usuario ha tocado anteriormente y la frecuencia con la que el usuario ha tocado dichas canciones, opcionalmente, adicionalmente, en relación con el momento en que el usuario ha tocado dichas canciones,
- basándose en la información del historial de reproducción, calcular una primera probabilidad para al menos una canción,
- basándose en un número de primeras probabilidades para un número de canciones y, basándose en la actividad de reproducción detectada y en las notas y/o acordes detectados, estimar la canción que el usuario está interpretando, en donde

40 o dicha estimación comprende el cálculo de una segunda probabilidad para un número de canciones diferentes, en donde las segundas probabilidades están definidas por la señal de audio en tiempo real correspondiente a una canción particular del historial de reproducción en combinación con la primera probabilidad asociada a dicha canción, y

- proporcionar la canción que el usuario está interpretando o información referente a la canción.

45 El aparato puede ser o comprender un teléfono móvil.

El aparato puede ser o comprender un reloj inteligente.

50 El aparato puede ser o comprender un ordenador de tipo tableta.

El aparato puede ser o comprender un ordenador portátil.

El aparato puede ser o comprender un reloj inteligente.

55 El aparato puede ser o comprender un ordenador de tipo tableta.

El aparato puede ser o comprender un ordenador portátil.

60 El aparato puede comprender un amplificador de instrumentos inteligente, tal como un amplificador de guitarra inteligente.

El aparato puede comprender un altavoz inteligente, tal como un altavoz proporcionado por asistente virtual.

65 El aparato puede ser o comprender un ordenador de sobremesa.

El aparato puede ser o comprender un ordenador.

De acuerdo con un tercer aspecto de ejemplo, se proporciona un programa informático que comprende un código de programa ejecutable por ordenador que, cuando es ejecutado por al menos un procesador, hace que un aparato realice al menos el método del primer aspecto de ejemplo.

De acuerdo con un cuarto aspecto de ejemplo, se proporciona un producto de programa de ordenador que comprende un medio legible por ordenador no transitorio que tiene almacenado el programa de ordenador del tercer aspecto de ejemplo.

De acuerdo con un quinto aspecto de ejemplo, se proporciona un aparato que comprende medios para realizar el método del primer aspecto de ejemplo.

Cualquier medio de memoria anterior puede comprender un almacenamiento digital de datos tal como un disco o disquete de datos; almacenamiento óptico; almacenamiento magnético; almacenamiento holográfico; almacenamiento opto-magnético; memoria de cambio de fase; memoria resistiva de acceso aleatorio; memoria magnética de acceso aleatorio; memoria de electrolito sólido; memoria ferroeléctrica de acceso aleatorio; memoria orgánica; o memoria de polímero. El medio de memoria puede estar formado en un dispositivo sin otras funciones sustanciales más que almacenar memoria, o puede estar formado como parte de un dispositivo con otras funciones, que incluyen, aunque no de forma limitativa, una memoria de un ordenador; un conjunto de chips; y un subconjunto de un dispositivo electrónico.

La expresión "un número de" se refiere en el presente documento a cualquier número entero positivo a partir de uno (1), por ejemplo, a uno, dos o tres.

La expresión "una pluralidad de" se refiere en el presente documento a cualquier número entero positivo a partir de dos (2), por ejemplo, a dos, tres, o cuatro.

En lo anterior se han ilustrado diferentes aspectos y realizaciones de ejemplo no vinculantes. Las realizaciones anteriores se utilizan simplemente para explicar aspectos o etapas seleccionados que pueden utilizarse en diferentes implementaciones. Algunas realizaciones pueden presentarse solo con referencia a ciertos aspectos de ejemplo. Debe apreciarse que las realizaciones correspondientes pueden aplicarse también a otros aspectos de ejemplo.

Breve descripción de los dibujos

Se describirán algunas realizaciones de ejemplo, haciendo referencia a los dibujos adjuntos, en los que:

- la Figura 1 muestra esquemáticamente un sistema de acuerdo con una realización de ejemplo;
- la Figura 2 muestra un diagrama de bloques de un aparato de acuerdo con una realización de ejemplo;
- las Figuras 3a-d muestran un diagrama de flujo de acuerdo con una realización de ejemplo; y
- la Figura 4 muestra una visión general de una realización de ejemplo.

Descripción detallada de los dibujos

En la siguiente descripción, los signos de referencia similares denotan elementos o etapas similares.

La Figura 1 muestra esquemáticamente un sistema 100 de acuerdo con una realización de ejemplo. El sistema comprende un instrumento musical 114 y un aparato 112, tal como un teléfono móvil, un ordenador de tipo tableta, amplificador de instrumentos inteligente, altavoz inteligente, o un ordenador portátil. El escenario puede ser, por ejemplo, un usuario tocando un instrumento 114 y utilizando un aparato de usuario 112 en su casa.

La Figura 2 muestra un diagrama de bloques de un aparato 200 de acuerdo con una realización de ejemplo. El aparato 200 comprende una interfaz de comunicación 210; un procesador 220; una interfaz de usuario 230; y una memoria 240.

La interfaz de comunicación 210 comprende, en una realización, una circuitería de comunicación por cable y/o inalámbrica, tal como Ethernet; LAN inalámbrica; Bluetooth; GSM; CDMA; WCDMA; LTE; y/o circuitería 5G. La interfaz de comunicación puede estar integrada en el aparato 200 o proporcionarse como parte de un adaptador, tarjeta o similar, que es acoplable al aparato 200. La interfaz de comunicación 210 puede admitir una o más tecnologías de comunicación diferentes. El aparato 200 puede comprender también, o como alternativa, más de una de las interfaces de comunicación 210.

En el presente documento, un procesador puede referirse a una unidad central de procesamiento (CPU); un microprocesador; un procesador de señales digitales (DSP); una unidad de procesamiento de gráficos; un circuito integrado de aplicación específica (ASIC); una matriz de puertas programables en campo; un microcontrolador; o una combinación de tales elementos.

La interfaz de usuario 230 puede comprender una circuitería para recibir entradas de un usuario del aparato 200, por ejemplo, a través de un teclado; interfaz gráfica de usuario mostrada en la pantalla del aparato 200; circuitería de reconocimiento de voz; o un dispositivo accesorio; tal como un micrófono, auriculares o una conexión de entrada de línea de audio 250 para recibir la señal de audio de la interpretación; y para proporcionar salida al usuario a través de, por ejemplo, una interfaz gráfica de usuario o un altavoz.

La memoria 240 comprende una memoria de trabajo y una memoria persistente configuradas para almacenar código de programa de ordenador y datos. La memoria 240 puede comprender uno cualquiera o más de: una memoria de solo lectura (ROM); una memoria de solo lectura programable (PROM); una memoria de solo lectura programable y borrrable (EPROM); una memoria de acceso aleatorio (RAM); una memoria flash; un disco de datos; un almacenamiento óptico; un almacenamiento magnético; una tarjeta inteligente; una unidad de estado sólido (SSD); o similar. El aparato 200 puede comprender una pluralidad de las memorias 240. La memoria 240 puede construirse como parte del aparato 200 o como un accesorio que se inserta en una ranura; puerto; o similar del aparato 200 por un usuario o por otra persona, o por un robot. La memoria 240 puede servir al único propósito de almacenar datos o estar construida como parte de un aparato 200 que sirve a otros propósitos, tal como el procesamiento de datos.

Un experto aprecia que, además de los elementos mostrados en la Figura 2, el aparato 200 puede comprender otros elementos, tales como micrófonos; pantallas; así como circuitería adicional tal como circuitería de entrada/salida (E/S); chips de memoria; circuitos integrados de aplicación específica (ASIC); circuitería de procesamiento para propósitos específicos, tal como circuitería de codificación/decodificación de fuentes; circuitería de codificación/decodificación de canales; circuitería de cifrado/descifrado; y similares. Adicionalmente, el aparato 200 puede comprender una batería desechable o recargable (no mostrada) para alimentar el aparato 200 si no se dispone de una fuente de alimentación externa.

Las Figuras 3a-d muestran un diagrama de flujo de acuerdo con una realización de ejemplo. Las Figs. 3a-d ilustran un proceso que comprende diversas etapas posibles incluyendo algunas etapas opcionales, mientras que también pueden incluirse otras etapas, y/o algunas de las etapas pueden realizarse más de una vez:

- 300. proporcionar un conjunto de información sobre notación musical o progresión armónica de las canciones,
- 301. recibir una señal de audio en tiempo real de un usuario que interpreta en un instrumento,
- 302. detectar la actividad de reproducción en segmentos sucesivos del usuario en segmentos sucesivos a partir de la señal de audio en tiempo real,
- 303. detectar notas y/o acordes a partir de la señal de audio en tiempo real,
- 304. almacenar información del historial de reproducción del usuario, que comprende información sobre las canciones que el usuario ha tocado anteriormente y con qué frecuencia y cuándo las ha tocado, opcionalmente, adicionalmente, en relación con el momento en que el usuario ha tocado dichas canciones,
- 305. basándose en la información del historial de reproducción, calcular una primera probabilidad para al menos una canción,
- 306. basándose en un número de primeras probabilidades para un número de canciones y, basándose en la actividad de reproducción detectada y en las notas y/o acordes detectados, estimar la canción que el usuario está interpretando, en donde dicha estimación comprende el cálculo de una segunda probabilidad para un número de canciones diferentes, en donde las segundas probabilidades están definidas por la señal de audio en tiempo real correspondiente a una canción particular del historial de reproducción en combinación con la primera probabilidad asociada a dicha canción y
- 307. proporcionar la canción que el usuario está interpretando o información referente a la canción.

El método puede también comprender uno o más de:

- 308. empezar a reproducir la canción o el audio de acompañamiento relacionado con la canción;
- 309. mostrar información referente a la canción, tal como la notación musical de la canción;
- 310. estimar el tempo del usuario e incluir información de los tempos a los que un usuario ha tocado diferentes canciones a la información del historial de reproducción;
- 311. utilizar la información de tempo para estimar la canción que el usuario está interpretando.
- 312. detectar la signatura de tiempo a partir de la señal de audio en tiempo real y utilizar la signatura de tiempo reconocida para estimar la canción que el usuario está interpretando.
- 313. detectar la posición temporal a partir de la señal de audio en tiempo real y almacenar información de la posición temporal en la que el usuario normalmente empieza a tocar cada una de las canciones en la información del historial de reproducción, y utilizar dicha posición temporal para estimar la canción que el usuario está

interpretando;

314. empezar a reproducir la canción o el audio de la pista de acompañamiento en la posición de la canción;

315. empezar a reproducir la canción o el audio de la pista de acompañamiento al tempo del usuario;

5 316. incluir el estilo en el que el usuario ha tocado previamente cada una de las canciones a la información del historial de reproducción e incluir dicho estilo para estimar la canción que el usuario está interpretando, en donde el estilo comprende información de los patrones rítmicos aplicados y la disposición de los tonos de las notas individuales;

10 317. detectar la afinación del instrumento que el usuario está tocando a partir de la señal de audio en tiempo real e incluir la afinación aplicada en cada uno de los historiales de reproducción de las canciones y utilizar dicha información de afinación para estimar la canción que el usuario está interpretando;

318. recibir una señal de audio en tiempo real de un usuario que interpreta en un instrumento, en donde la señal de audio en tiempo real comprende una fuente de audio;

15 319. recibir una señal de audio en tiempo real de un usuario que interpreta en un instrumento, en donde la señal de audio en tiempo real comprende al menos dos fuentes de audio;

320. utilizar una o ambas fuentes de audio para detectar notas y/o acordes;

321. recibir una señal de audio en tiempo real de un usuario que interpreta en un instrumento, en donde, si la fuente de audio comprende voces, detectar la letra de la canción a partir de la señal de audio en tiempo real y usar dicha información de la letra para estimar la canción que el usuario está interpretando;

20 322. calcular la segunda probabilidad mediante:

a. seguir el tempo y el compás de la interpretación del usuario y subdividir la señal de audio en tiempo real en segmentos sucesivos utilizando cada posición de compás como límite de segmento, en donde, opcionalmente, se pueden colocar límites de segmento adicionales a medio camino entre cada dos posiciones de compás, para duplicar la resolución temporal;

25 b. si se supone que el usuario toca de acuerdo con una notación musical detallada (nota por nota), el cálculo de una matriz de probabilidades para todas las notas musicales diferentes, tal como basado en el sistema de teclado de piano de 88 notas, en cada segmento; y

30 c. si el usuario está tocando libremente (basándose en la lectura o el recuerdo de un documento musical basado en acordes), el cálculo de una matriz de probabilidades para los distintos acordes que pueden aparecer en el catálogo de canciones que mantiene el sistema; en donde

35 d. en ambos casos formando una matriz de probabilidades en la que una dimensión es el tiempo (en unidades de compás musical) y la otra dimensión es la nota/acorde;

40 e. es decir, cotejar la matriz de probabilidades en distintas posiciones temporales de los documentos musicales que representan distintas canciones en el historial de reproducción, en donde dicha concordancia puede comprender la determinación de un grado de concordancia mediante el cálculo del coeficiente de correlación entre la matriz de probabilidades de notas/acordes estimada a partir de la señal de audio con la matriz de valores binarios de la matriz de presencia de notas/acordes extraída del documento musical.

323. calcular la segunda probabilidad mediante:

45 a. El método de cualquier reivindicación anterior, en donde la segunda probabilidad se calcula mediante:

b. estimación de las probabilidades variables en el tiempo de la actividad de reproducción a partir de la señal de audio en tiempo real, tal como rastreando la probabilidad de que el usuario esté tocando cualquier nota sonora en segmentos sucesivos de la señal de audio;

50 c. seguimiento del tempo variable del usuario, en donde varios candidatos/hipótesis de tempo son seguidos uno al lado del otro, con probabilidades asociadas a cada candidato de tempo;

55 d. estimación de las probabilidades de que suenen diferentes notas/acordes en la señal de audio en tiempo real:

i. si se supone que el usuario toca a partir de una notación musical detallada (nota por nota), la estimación de las probabilidades variables en el tiempo de todas las notas musicales admisibles, tal como basado en el sistema de teclado de piano de 88 notas;

60 ii. si se supone que el usuario toca libremente, el cálculo de una matriz de probabilidades para los distintos acordes que pueden aparecer en el catálogo de canciones que mantiene el sistema;

65 e. uso de las mediciones variables en el tiempo, incluida la probabilidad estimada de actividad, tempo y probabilidades de nota/acorde para seguir la posición de reproducción del usuario en varias canciones diferentes simultáneamente;

f. calcular un valor para la segunda probabilidad de una canción a partir de la fiabilidad de la posición de reproducción estimada en la canción;

5 324. determinar la fiabilidad de la posición de reproducción estimada basándose en uno o más de los siguientes elementos:

10 a. dada la trayectoria de las posiciones de reproducción de los usuarios en función del tiempo para una cierta canción S, comparar con las notas/acordes que aparecen en un documento musical referente a una canción en las posiciones indicadas por la trayectoria de posiciones de reproducción, tras lo cual comparar las probabilidades estimadas para esas notas/acordes en los momentos correspondientes de la señal de audio, y calcular la probabilidad total de haber observado esas notas/acordes a lo largo del tiempo;

15 b. controlar la estabilidad de la posición de reproducción estimada - si se producen saltos repentinos en la posición estimada y la estabilidad de la velocidad de avance de la posición de reproducción estimada, en donde una mayor estabilidad indica una mayor fiabilidad; y

20 c. calcular la relación X/Y entre la duración X del acorde interpretado y la duración Y del acorde escrito, en donde la duración del acorde ejecutado se calcula contando la cantidad de compases musicales desde un cambio de acorde hasta el siguiente cambio de acorde, y en donde la duración del acorde escrito se refiere a la cantidad de compases musicales entre esos dos cambios de acorde en el documento musical;

25 325. utilizar la probabilidad variable en el tiempo estimada de actividad de reproducción para ponderar las probabilidades calculadas de los diferentes acordes/notas de tal manera que se dé más importancia a los puntos temporales en los que el intérprete realmente toca algo;

326. proporcionar varias canciones coincidentes simultáneamente en la pantalla del dispositivo cuando la canción interpretada sigue siendo ambigua;

327. permitir al usuario indicar la correcta haciendo clic o continuar la reproducción hasta reconocer la canción correcta;

30 328. almacenar información de la actividad de reproducción del usuario disponiéndose en modo de escucha continua, donde se puede grabar la interpretación del usuario para saber cuántas veces y cuándo el usuario ha tocado esas canciones;

329. permitir al usuario añadir manualmente canciones al historial de reproducción, sin tocarlas realmente;

330. estimar la canción que el usuario está interpretando fuera del historial de reproducción del usuario;

35 331. recuperar información de sistemas externos referente a la actividad de escucha de canciones del usuario;

332. cotejar la presente interpretación con canciones que estén fuera del historial de reproducción del usuario y coincidan con el "gusto musical" del usuario.

40 La Figura 4 ilustra una vista general de una realización de ejemplo. Se muestra al usuario tocando un instrumento, en concreto una guitarra en este caso, uso de un aparato móvil con micrófono o entrada de línea para seguir la interpretación del usuario, es decir, tocar el instrumento, y detectar qué notas y acordes toca el usuario. El usuario puede empezar a tocar libremente con su instrumento y el sistema y el método del mismo reconocen la canción, tras lo cual se puede proporcionar al usuario la música escrita y la letra referentes a la canción a través de la pantalla del aparato móvil. Además, el sistema y el método pueden, tras el reconocimiento de la canción, empezar a reproducir una pista de acompañamiento referente a la canción reconocida a través del aparato móvil, opcionalmente en el lugar y tempo al que el usuario está interpretando la canción.

50 La interpretación del usuario puede comprender, por ejemplo, reproducción de una pluralidad de acordes de abertura de una canción o de alguna otra posición o parte de una canción. Adicionalmente, el historial de reproducción del usuario se combina con la actuación del usuario detectada, de modo que la canción que corresponde a la interpretación del usuario pueda detectarse con mayor eficacia a partir de un gran conjunto de canciones. El aparato móvil está provisto de un historial de reproducción del usuario, información de la canción y datos de audio de la pista de acompañamiento desde una disposición de servidor externo o nube. Como se ha mencionado, tal información de la canción puede comprender la notación de la canción, pero también información adicional, tal como información de práctica o versiones alternativas de la canción y/o diversos tipos de visualización de la canción o la interpretación de la misma por parte del usuario. Como alternativa, el aparato móvil puede almacenar al menos parte del historial de reproducción del usuario, información de la canción y datos de la pista de acompañamiento. De este modo, el usuario puede utilizar el sistema y el método con muy poca intervención por parte del usuario u operación de la interfaz de usuario y centrarse en la libre expresión con su instrumento, que se detecta automáticamente y se coteja con una canción a la que, a continuación, se puede proporcionar automáticamente información sobre la canción para, por ejemplo, acompañar la interpretación del usuario o ayudarle a tocar o practicar la canción.

60 Se conocen muchas técnicas de estimación del tempo y pueden usarse, ya que es un tema ampliamente conocido en la técnica anterior. Se analizan a continuación ejemplos de estimación de la actividad del usuario, posición de interpretación y tempo, que se obtienen analizando la señal de audio de la interpretación en tiempo real.

Las características de actividad indican cuándo el usuario está tocando realmente en lugar de no producir momentáneamente ninguna nota sonora del instrumento. Esto último puede deberse a cualquier motivo, tal como un descanso (punto de silencio) en el patrón rítmico aplicado, o debido a que el intérprete hace una pausa en su interpretación. Por consiguiente, las características de actividad desempeñan dos papeles en nuestro sistema: 1) Permiten ponderar las probabilidades calculadas de diferentes acordes de tal manera que se da más importancia a los momentos de la interpretación en los que el intérprete toca realmente algo (es decir, donde esté presente la información sobre la interpretación). 2) Las características de actividad permiten al método mantener fija la posición estimada cuando el intérprete hace una pausa y seguir avanzando la posición cuando se reanuda la interpretación. Para artistas aficionados, no es raro dudar e incluso pararse un momento para averiguar la posición de la mano en el instrumento, por ejemplo. También, al interpretar en casa, no es raro dejar de actuar durante un rato para charlar con otra persona, por ejemplo. Más técnicamente, las características de actividad describen, en una realización, la probabilidad de que suene cualquier nota en un segmento de audio dado: $p(\text{NotasSonando} | \text{SegmentoAudio}(t))$ como un número real entre 0 y 1.

Las características tonales controlan el contenido tonal de la interpretación del usuario. Como se ha descrito anteriormente, cuando se interpreta a partir de una partitura, no conocemos de antemano las notas exactas que tocará el usuario ni su sincronización: el arreglo/textura de la música se desconoce de antemano. Por esta razón, empleamos en su lugar una serie de modelos que representan diferentes acordes que pueden aparecer en las partituras. Los modelos permiten calcular una "coincidencia" o "puntuación" de esos acordes: la probabilidad de que el acorde correspondiente esté sonando en un segmento dado del audio de la interpretación. Obsérvese que el sistema puede ser incluso totalmente agnóstico de las notas componentes de cada acorde - por ejemplo cuando el modelo de cada acorde se entrena a partir de datos de audio, dándole ejemplos donde el acorde suena/no suena. El vector de características de tonalidad se obtiene calculando una correspondencia entre un segmento dado de audio de interpretación y todos los acordes únicos que aparecen en la canción. Más técnicamente: probabilidades de que suenen diferentes acordes en un determinado segmento de audio t : $p(\text{Acorde}(i) | \text{SegmentoAudio}(t))$, donde el índice de acordes $i = 1, 2, \dots$, <número de acordes únicos en la canción>. Las características de tonalidad nos ayudan a estimar la probabilidad de que el intérprete se encuentre en distintas partes de la canción. Los intérpretes aficionados a veces saltan hacia atrás en la interpretación para repetir un segmento corto o corregir un error de interpretación. También son posibles los saltos hacia delante. El contenido armónico de la interpretación del usuario permite al método "anclar" la posición del usuario en la canción incluso en presencia de tales saltos.

Las características de tiempo se utilizan para estimar el tiempo (o, velocidad de interpretación) del intérprete en tiempo real. En muchas canciones, hay segmentos en los que el acorde no cambia durante mucho tiempo. Dentro de tales segmentos, el tiempo estimado del usuario adelanta la posición del intérprete. En otras palabras, incluso en ausencia de cambios de acordes (cambios armónicos), el tener una estimación del tiempo del usuario nos permite ir actualizando la posición del intérprete. Más técnicamente: probabilidades de diferentes tempos (velocidades de interpretación) dado el segmento de audio de interpretación t , $p(\text{Tempo}(j) | \text{SegmentoAudio}_{0, 1, 2, \dots, t})$, donde el índice j abarca todos los valores de tiempo entre un tiempo mínimo y un tiempo máximo de interés.

Combinando la información de las tres características mencionadas anteriormente y la información de la pista de acompañamiento, podemos abordar los diversos retos que plantea el seguimiento de la posición $x(t)$ y el tiempo de interpretación de un intérprete aficionado y establecer una pista de acompañamiento que corresponda a la posición y el tiempo de interpretación en donde:

1. Las características de actividad ayudan a detectar los momentos en los que está presente la información sobre la interpretación, en otras palabras, donde el intérprete produce realmente algunas notas sonoras. También captan la situación cuando el usuario hace una pausa en la interpretación.
2. Las características de tonalidad indican las posibles posiciones (en una escala temporal mayor) en las que podría encontrarse el usuario en la canción. Esta función ayuda a resolver los casos en los que el usuario salta hacia delante o hacia atrás en la canción.
3. Las funciones de tiempo adelantan localmente la posición del usuario, dentro de segmentos en los que la tonalidad se mantiene igual durante algún tiempo. La posición del usuario $x(t)$ en el tiempo t puede extrapolarse a partir de la posición anterior $x(t-1)$ y la velocidad de interpretación $v(t)$. Sin embargo, a veces el usuario puede saltar hacia atrás o hacia delante dentro de la canción. En ese caso, las funciones de tonalidad ayudan a detectar el salto y a "restablecer" esta extrapolación lineal local de la posición del intérprete.

Cualquiera de los métodos, etapas del método, o combinaciones de los mismos descritos anteriormente, pueden controlarse o realizarse mediante hardware; software; firmware; o cualquier combinación de los mismos. El software y/o el hardware pueden ser locales; distribuidos; centralizados; virtualizados; o cualquier combinación de los mismos. Asimismo, cualquier forma de informática, incluida la inteligencia computacional, puede usarse para controlar o ejecutar cualquiera de los métodos, etapas del método o combinaciones de los mismos descritos anteriormente. La inteligencia computacional puede referirse a, por ejemplo, cualquiera de inteligencia artificial; redes neuronales; lógica difusa; aprendizaje automático; algoritmos genéticos; computación evolutiva; o cualquier combinación de los mismos.

Se han presentado diversas realizaciones. Debería apreciarse que, en el presente documento, las palabras

comprender; incluir; y contener se utilizan cada una como expresiones abiertas sin exclusividad prevista.

5 La descripción anterior ha proporcionado, a modo de ejemplos no limitativos de realizaciones particulares, una descripción completa e informativa del mejor modo actualmente contemplado por los inventores para llevar a cabo la invención. Sin embargo, es evidente para un experto en la materia que la invención no se restringe a los detalles de las realizaciones presentadas en lo anterior, pero que puede implementarse en otras realizaciones utilizando diferentes combinaciones de realizaciones sin desviarse de las características de la invención.

10 Así mismo, algunas de las características de las realizaciones de ejemplo divulgadas anteriormente pueden usarse ventajosamente sin el correspondiente uso de otras características. Como tal, la descripción anterior se considerará meramente ilustrativa de los principios de la presente invención, y no limitativa de la misma. En consecuencia, el alcance de la invención solo está restringido por las reivindicaciones de patente adjuntas.

REIVINDICACIONES

1. Un método implementado por ordenador para identificar una canción que comprende:

- 5 - proporcionar un conjunto de información sobre notación musical o progresión armónica de las canciones,
- recibir una señal de audio en tiempo real de un usuario que interpreta en un instrumento,
- detectar la actividad de reproducción en segmentos sucesivos del usuario en segmentos sucesivos a partir de la señal de audio en tiempo real,
- detectar notas y/o acordes a partir de la señal de audio en tiempo real,
- 10 - almacenar información del historial de reproducción del usuario que comprende información sobre las canciones que un usuario ha tocado anteriormente y la frecuencia con la que el usuario ha tocado dichas canciones, opcionalmente, adicionalmente, en relación con el momento en que el usuario ha tocado dichas canciones,
- basándose en la información del historial de reproducción, calcular una primera probabilidad para al menos una canción,

15 estando el método **caracterizado por**:

- basándose en un número de primeras probabilidades para un número de canciones y, basándose en la actividad de reproducción detectada y en las notas y/o acordes detectados, estimar la canción que el usuario está interpretando, en donde

20 o dicha estimación comprende el cálculo de una segunda probabilidad para un número de canciones diferentes, en donde las segundas probabilidades están definidas por la señal de audio en tiempo real correspondiente a una canción particular del historial de reproducción en combinación con la primera probabilidad asociada a dicha canción, y

- proporcionar la canción que el usuario está interpretando o información referente a la canción.

2. El método de la reivindicación 1, en donde proporcionar un conjunto de datos de audio y empezar a reproducir la canción o el audio de acompañamiento relacionado con la canción.

3. El método de cualquier reivindicación anterior, en donde mostrar información referente a la canción, tal como la notación musical de la canción.

35 4. El método de cualquier reivindicación anterior, en donde estimar adicionalmente el tempo del usuario e incluir información de los tempos a los que un usuario ha tocado diferentes canciones a la información del historial de reproducción.

40 5. El método de la reivindicación 4, en donde utilizar la información de tempo para estimar la canción que el usuario está interpretando.

45 6. El método de cualquier reivindicación anterior, en donde detectar adicionalmente la signatura de tiempo a partir de la señal de audio en tiempo real y utilizar la signatura de tiempo reconocida para estimar la canción que el usuario está interpretando.

50 7. El método de cualquier reivindicación anterior, en donde detectar adicionalmente la posición temporal a partir de la señal de audio en tiempo real y almacenar información de la posición temporal en la que el usuario normalmente empieza a tocar cada una de las canciones en la información del historial de reproducción, y utilizar dicha posición temporal para estimar la canción que el usuario está interpretando.

8. El método de la reivindicación 7, en donde empezar a reproducir la canción o el audio de la pista de acompañamiento en la posición de la canción.

55 9. El método de la reivindicación 8, en donde empezar a reproducir la canción o el audio de la pista de acompañamiento al tempo del usuario.

60 10. El método de cualquier reivindicación anterior, en donde incluir adicionalmente el estilo en el que el usuario ha tocado previamente cada una de las canciones a la información del historial de reproducción e incluir dicho estilo para estimar la canción que el usuario está interpretando, en donde el estilo comprende información de los patrones rítmicos aplicados y la disposición de los tonos de las notas individuales.

65 11. El método de cualquier reivindicación anterior, en donde detectar la afinación del instrumento que el usuario está tocando a partir de la señal de audio en tiempo real e incluir la afinación aplicada en cada uno de los historiales de reproducción de las canciones, y utilizar dicha información de afinación para estimar la canción que el usuario está interpretando.

ES 3 010 466 T3

12. El método de cualquier reivindicación anterior, en donde la señal de audio en tiempo real comprende una fuente de audio.
- 5 13. El método de cualquier reivindicación anterior, en donde la señal de audio en tiempo real comprende al menos dos fuentes de audio.
14. El método de cualquiera de la reivindicación 13, en donde se utilizan una o ambas fuentes de audio para detectar notas y/o acordes.
- 10 15. El método de cualquier reivindicación anterior, en donde, si la fuente de audio comprende voces, detectar la letra de la canción a partir de la señal de audio en tiempo real y utilizar dicha información de la letra para estimar la canción que el usuario está interpretando.
- 15 16. La utilización del método de la reivindicación 1 en un terminal o dispositivo móvil que comprende medios de procesamiento, medios de almacenamiento de datos y medios de visualización.
17. Un sistema que comprende una entidad de procesamiento dispuesta para al menos almacenar, proporcionar y procesar información para ejecutar el método de la reivindicación 1.
- 20 18. Un programa informático almacenado en un medio legible por ordenador no transitorio, que comprende un código de programa ejecutable por ordenador que, cuando es ejecutado por al menos un procesador, hace que un aparato realice al menos el método de la reivindicación 1.

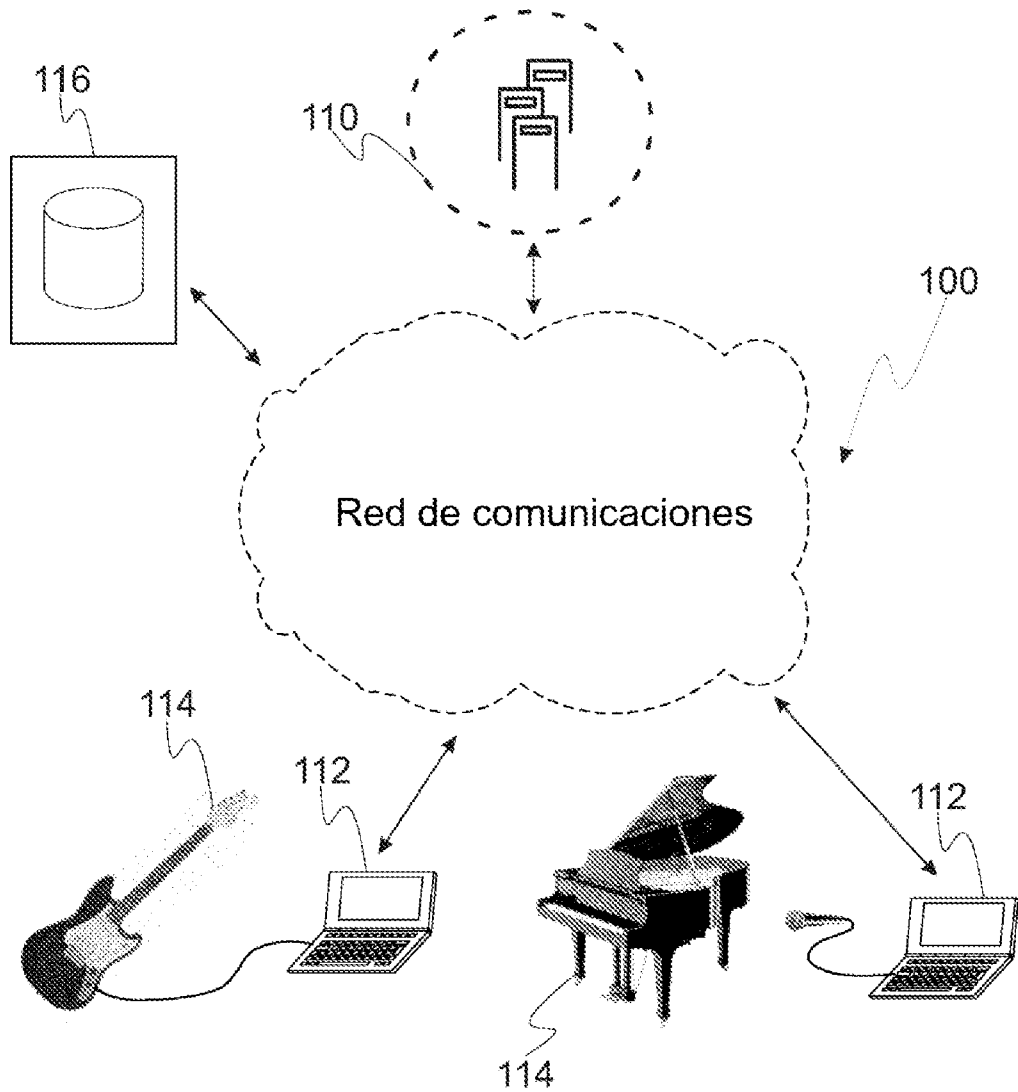


Figura 1

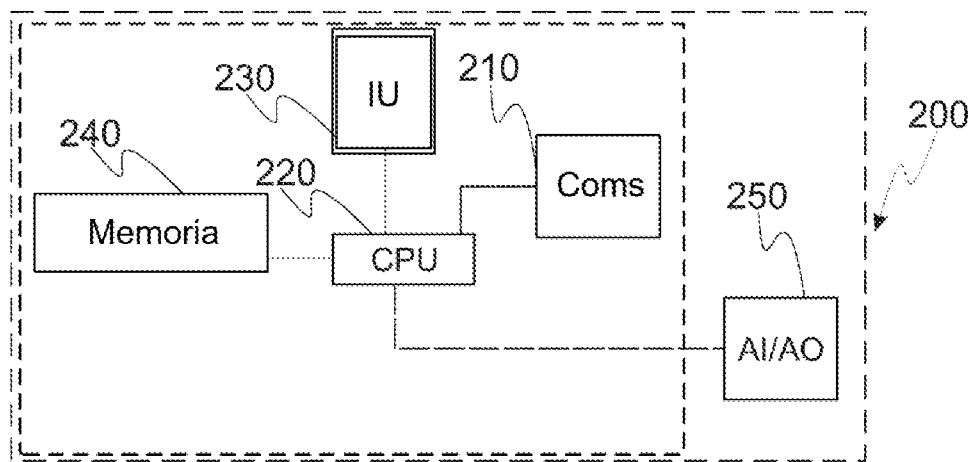


Figura 2

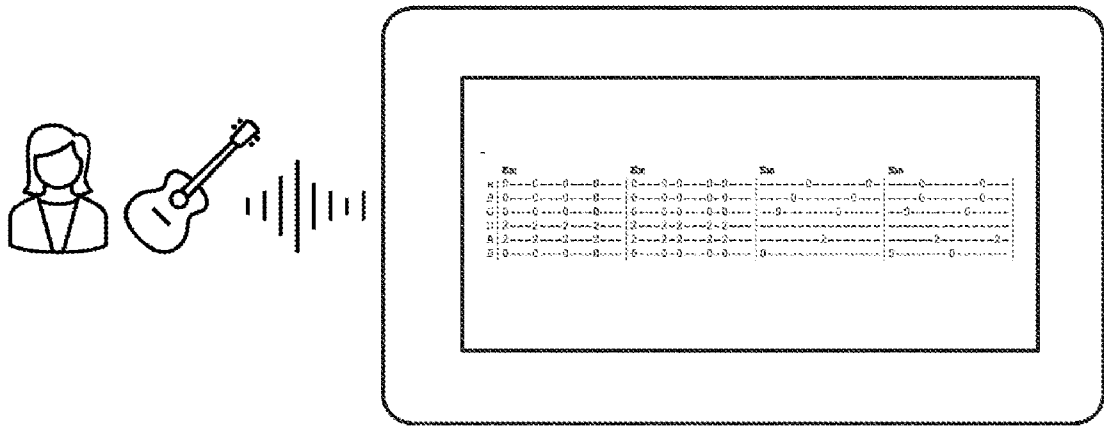


Figura 4

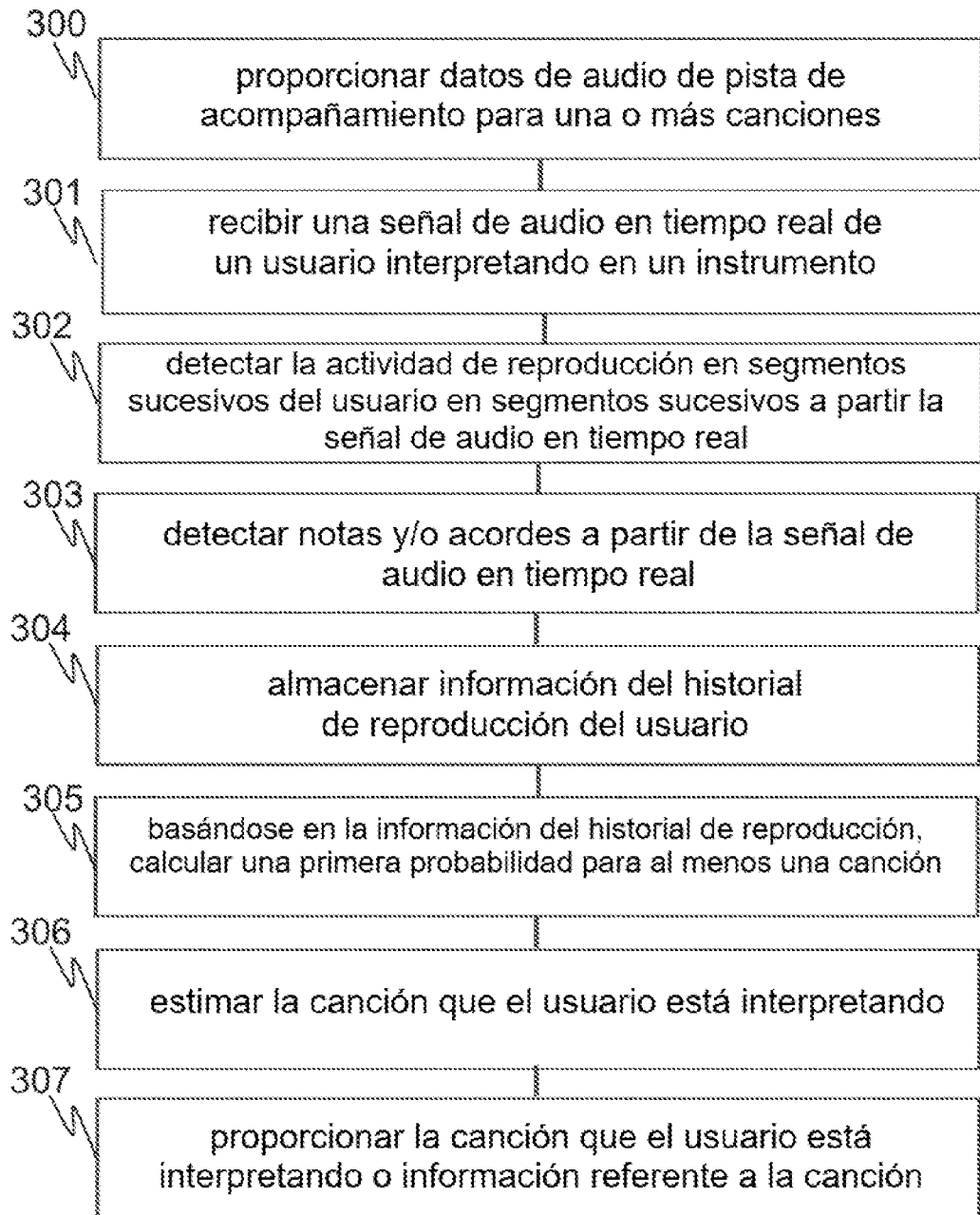


Figura 3a

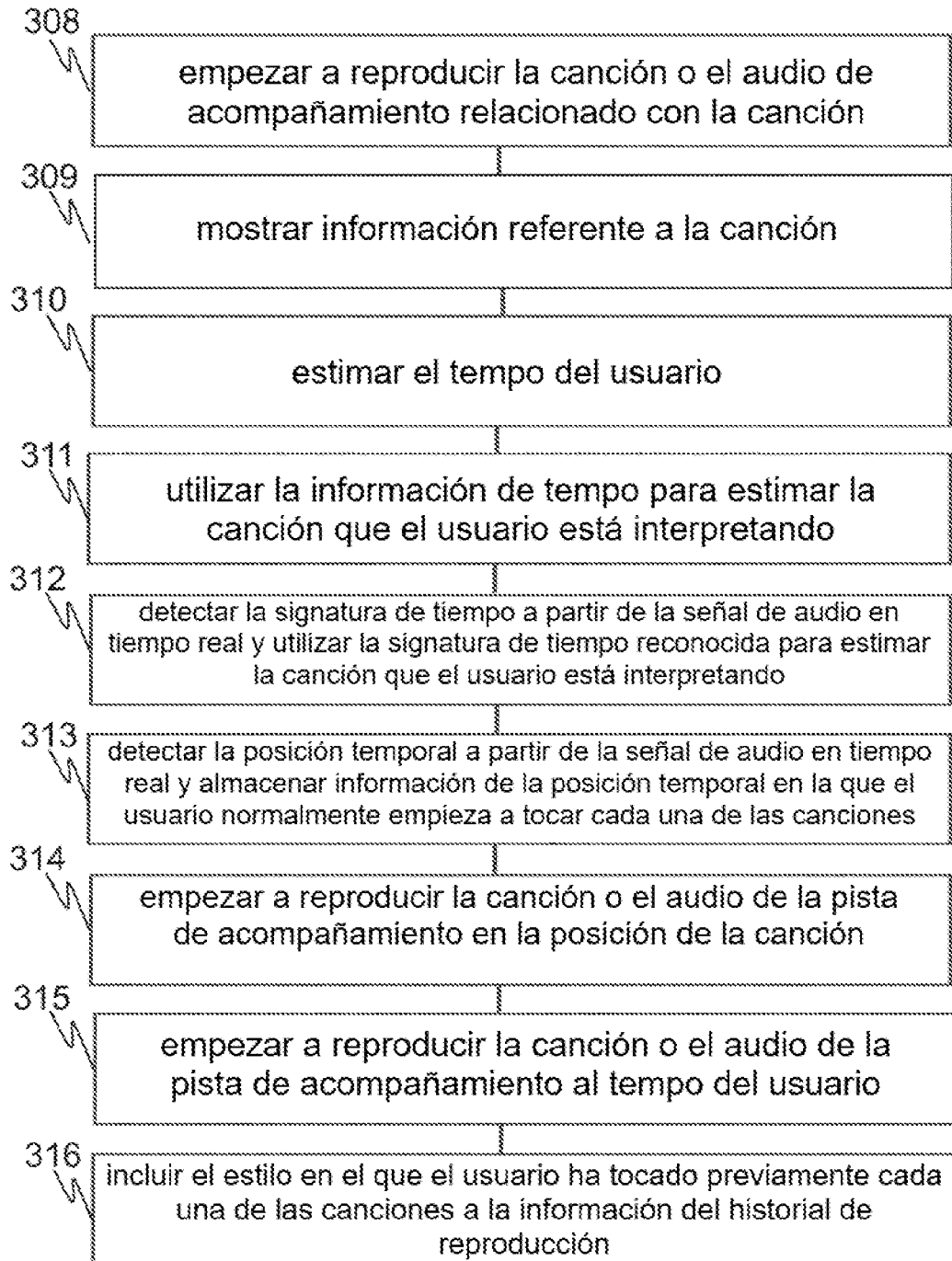


Figura 3b

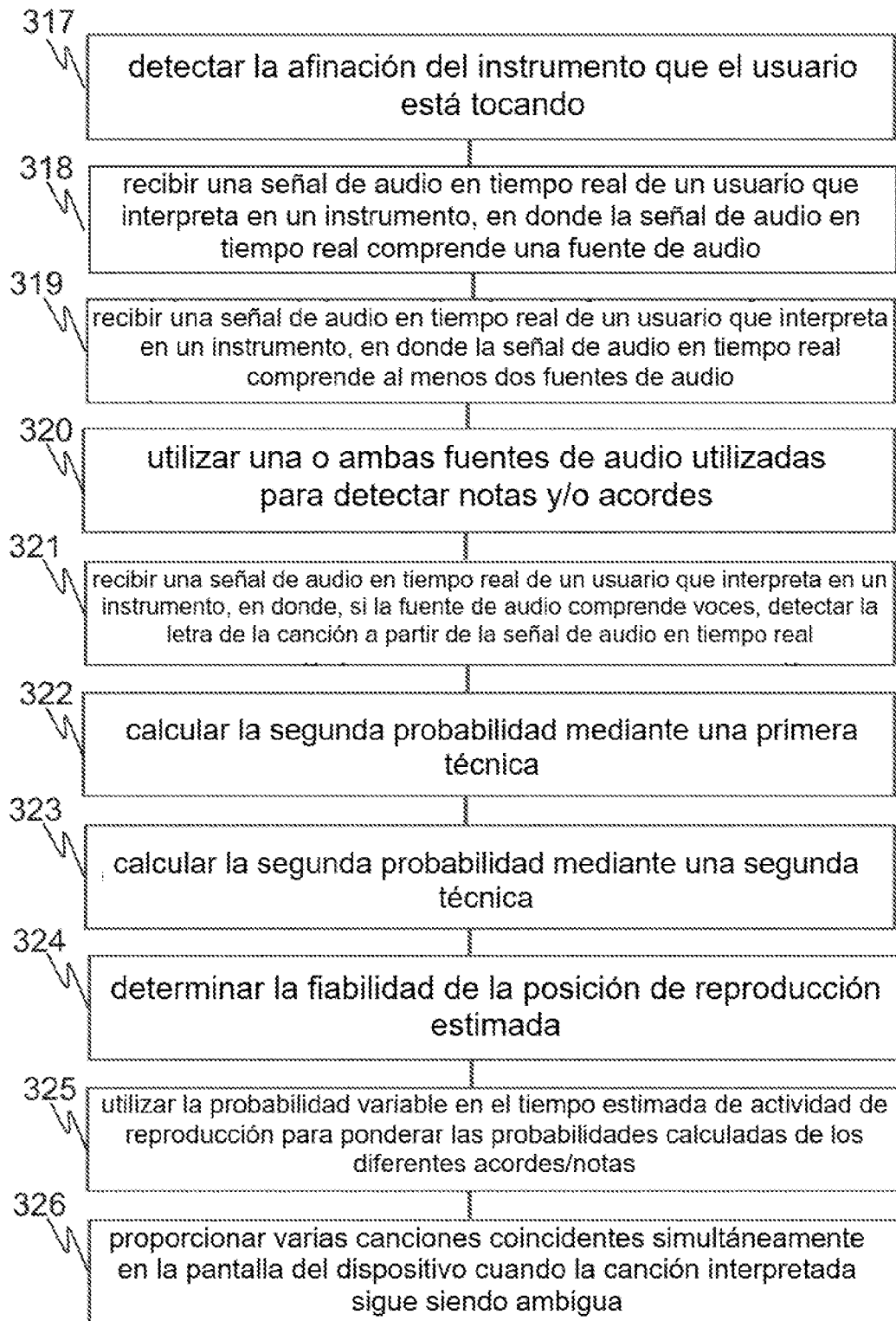


Figura 3c

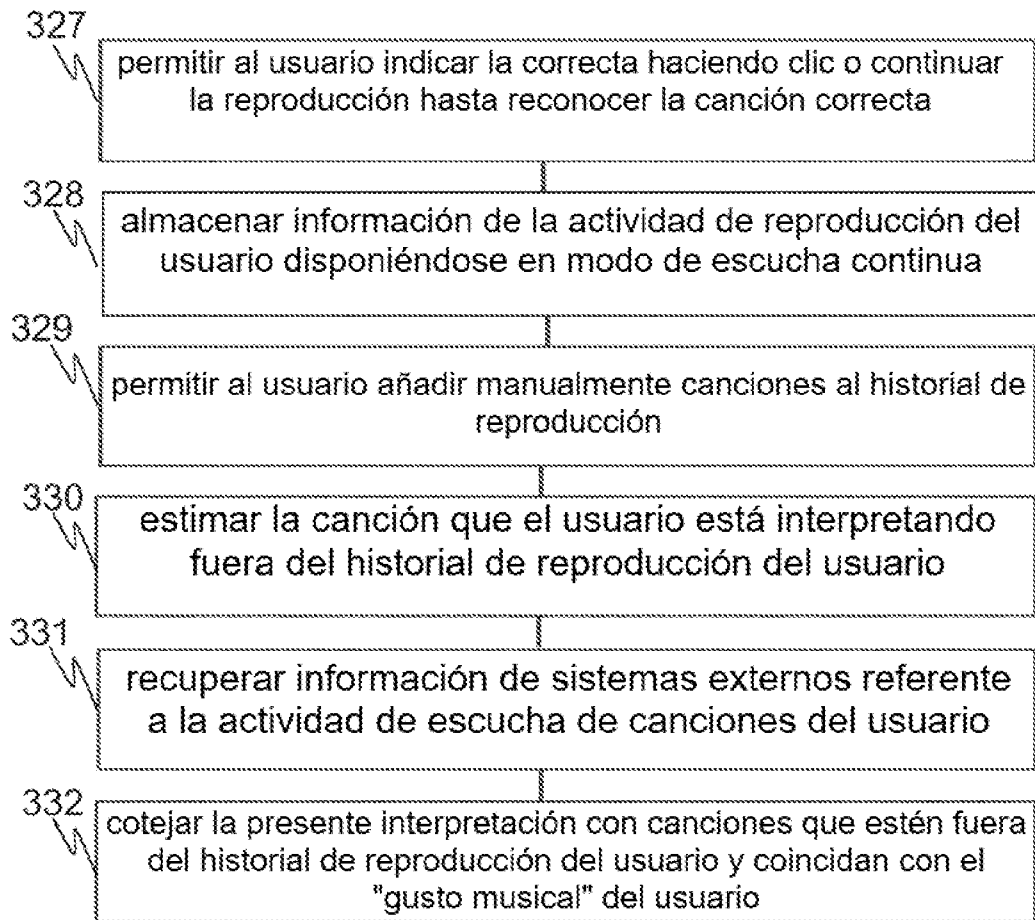


Figura 3d