

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 539 813**

51 Int. Cl.:

G10H 1/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **01.02.2008 E 08728874 (2)**

97 Fecha y número de publicación de la concesión europea: **25.03.2015 EP 2115732**

54 Título: **Transcripción de música**

30 Prioridad:

01.02.2007 US 887738 P

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

06.07.2015

73 Titular/es:

**MUSEAMI, INC. (100.0%)
100 Canal Pointe Blvd, Suite 117
Princeton, NJ 08540, US**

72 Inventor/es:

**TAUB, ROBERT D. y
CABANILLA, J. ALEXANDER**

74 Agente/Representante:

CURELL AGUILÁ, Mireia

ES 2 539 813 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

Transcripción de música.

5 **Antecedentes**

La presente invención se refiere a aplicaciones de audio en general y, en particular, a la descomposición de audio y la generación de partituras.

10 Puede resultar deseable proporcionar una conversión precisa, en tiempo real, de señales de entrada de audio sin procesar, en datos de partituras para su transcripción. Por ejemplo, un intérprete musical (por ejemplo, en vivo o grabado, que haga uso de voces y/u otros instrumentos) puede desear transcribir automáticamente una interpretación para generar texto musical o para convertir la interpretación en un archivo de partitura digital editable. Muchos elementos pueden formar parte de la interpretación musical, incluyendo notas, timbres, modos, dinámica, ritmos y pistas. Puede que el intérprete requiera que todos estos elementos se extraigan de manera fiable del archivo de audio para generar una partitura precisa.

20 En general los sistemas convencionales proporcionan solamente capacidades limitadas en estas áreas, e incluso dichas capacidades generalmente proporcionan salidas con una precisión y una oportunidad temporal limitadas. Por ejemplo, muchos sistemas convencionales requieren que el usuario proporcione datos al sistema (diferentes de una señal de audio) para ayudar a este último a convertir una señal de audio en datos de partitura útiles. Una de las limitaciones resultantes es que el suministro de datos, diferentes de la señal de audio sin procesar, al sistema puede consumir mucho tiempo o puede ser no deseable. Otra limitación resultante es que el usuario puede no conocer gran parte de los datos requeridos por el sistema (por ejemplo, puede que el usuario no esté familiarizado con la teoría musical). Todavía otra limitación resultante es que el sistema puede tener que proporcionar capacidades exhaustivas de interfaz de usuario para permitir la provisión de datos requeridos al sistema (por ejemplo, puede que el sistema tenga que disponer de un teclado, una pantalla, etcétera).

30 La publicación "Fast labelling of notes in music signals", de Paul M Brossier *et al* (ISMIR 2004), da a conocer un sistema para la estimación de atributos de notas a partir de una fuente de música monofónica en vivo. El inicio de las notas musicales se detecta basándose en criterios tanto de amplitud como de altura tonal. En la detección del inicio basada en la amplitud, se obtiene una envolvente de contenido de alta frecuencia a partir de la señal de entrada.

35 La publicación "*Real-Time temporal segmentation of note objects in music signals*", de Brossier *et al*, ICMC 2004, detalla adicionalmente una estrategia para la detección de inicios basada en la amplitud, implementando una captación de picos de la señal envolvente de contenido de alta frecuencia con el uso de una señal diferencial basada en una versión promediada de la envolvente de contenido de alta frecuencia.

40 Por lo tanto, puede que resulte deseable proporcionar capacidades mejoradas para extraer de manera automática y precisa datos de partitura a partir de un archivo de audio sin procesar.

Sumario

45 Se describen métodos, sistemas y dispositivos para extraer de manera automática y precisa datos de partitura a partir de una señal de audio. Se identifica un cambio en la información de frecuencia de la señal de audio de entrada, que supera un primer umbral, y se identifica un cambio en la información de amplitud de la señal de entrada de audio, que supera un segundo umbral. Se genera un evento de inicio de nota de tal manera que cada evento de inicio de nota representa una localización temporal en la señal de entrada de audio de por lo menos uno de un cambio identificado en la información de frecuencia, que supera el primer valor de umbral, o un cambio identificado de la información de amplitud, que supera el segundo valor de umbral. Las técnicas descritas en la presente se pueden implementar en métodos, sistemas y soportes de almacenamiento legibles por ordenador que tienen incorporado en ellos un programa legible por ordenador.

55 La invención se define en las reivindicaciones adjuntas 1 a 15. En un aspecto de la invención, se recibe una señal de audio desde una o más fuentes de audio. La señal de audio se procesa para extraer información de frecuencia y de amplitud. La información de frecuencia y de amplitud se usa para detectar eventos de inicio de notas (es decir, posiciones de tiempo en las que se determina que comienza una nota musical). Para cada evento de inicio de nota, se generan datos de envolvente, datos de timbre, datos de altura tonal, datos dinámicos, y otros datos. Examinando datos de entre conjuntos de eventos de inicio de notas, se generan datos de tempo, datos de métrica, datos de tonalidades, datos de dinámica global, datos de instrumentación y pistas, y otros datos. A continuación, los diversos datos se usan para generar una salida de partitura.

65 En un ejemplo de forma de realización, se generan datos de tempo a partir de una señal de audio y se determina un conjunto de tempos de referencia. Se determina un conjunto de duraciones de notas de referencia, representando cada duración de nota de referencia un espacio de tiempo que dura un tipo de nota predeterminado en cada tempo de referencia, y se determina una ventana de extracción de tempos, que representa una parte contigua de la señal

de audio que se extiende desde una primera localización temporal hasta una segunda localización temporal. Se genera un conjunto de eventos de inicio de nota localizando los eventos de inicio de nota que se producen dentro de la parte contigua de la señal de audio; generando una separación entre notas para cada evento de inicio de nota, de manera que cada separación entre notas representa el intervalo de tiempo entre el evento de inicio de nota y el siguiente evento de inicio de nota sucesivo en el conjunto de eventos de inicio de nota; generando un conjunto de valores de error, estando asociado cada valor de error a un tempo de referencia asociado, incluyendo la generación del conjunto de valores de error dividir cada separación entre notas por cada una del conjunto de duraciones de notas de referencia, redondear cada resultado de la etapa de división a un múltiplo más próximo de la duración de nota de referencia usada en la etapa de división, y evaluar el valor absoluto de la diferencia entre cada resultado de la etapa de redondeo y cada resultado de la etapa de división; identificando un valor de error mínimo del conjunto de valores de error; y determinando un tempo extraído asociado a la ventana de extracción de tempos, siendo el tempo extraído el tempo de referencia asociado que está asociado al valor de error mínimo. Se pueden generar además datos de tempo determinando un conjunto de segundas duraciones de notas de referencia, de manera que cada duración de nota de referencia representa un espacio de tiempo que dura cada uno de un conjunto de tipos de nota predeterminados en el tempo extraído; generando una duración de nota recibida para cada evento de inicio de nota; y determinando un valor de nota recibido para cada duración de nota recibida, de manera que el valor de nota recibido representa la segunda duración de nota de referencia que se aproxima mejor a la duración de nota recibida.

En otro ejemplo de forma de realización, una técnica para generar datos de tonalidades a partir de una señal de audio incluye determinar un conjunto de funciones de coste, estando asociada cada función de coste a una tonalidad y representando un ajuste de cada una de un conjunto de frecuencias predeterminadas a la tonalidad asociada; determinar una ventana de extracción de tonalidades, que representa una parte contigua de la señal de audio que se extiende desde una primera localización temporal hasta una segunda localización temporal; generar un conjunto de eventos de inicio de nota localizando los eventos de inicio de nota que se producen dentro de la parte contigua de la señal de audio; determinar una frecuencia de nota para cada uno del conjunto de eventos de inicio de nota; generar un conjunto de valores de error de tonalidad basándose en la evaluación de las frecuencias de notas con respecto a cada una del conjunto de funciones de coste; y determinar una tonalidad recibida, siendo la tonalidad recibida la tonalidad asociada a la función de coste que generó el valor de error de tonalidad más bajo. En algunas formas de realización, el método incluye además generar un conjunto de alturas tonales de referencia, representando cada altura tonal de referencia una relación entre una del conjunto de alturas tonales predeterminadas y la tonalidad recibida; y determinar una designación de altura tonal de tonalidad para cada evento de inicio de nota, representando la designación de altura tonal de tonalidad la altura tonal de referencia que se aproxima mejor a la frecuencia de nota del evento de inicio de nota.

Todavía en otro ejemplo de forma de realización, una técnica para generar datos de pistas a partir de una señal de audio incluye generar un conjunto de eventos de inicio de nota, estando caracterizado cada evento de inicio de nota por al menos un conjunto de características de nota, incluyendo el conjunto de características de nota una frecuencia de nota y un timbre de nota; identificar un número de pistas de audio presentes en la señal de audio, estando caracterizada cada pista de audio por un conjunto de características de pista, incluyendo el conjunto de características de pista por lo menos uno de un mapa de alturas tonales o un mapa de timbres; y asignar una pista presupuesta para cada conjunto de características de nota para cada evento de inicio de nota, siendo la pista presupuesta la pista de audio caracterizada por el conjunto de características de pista que se corresponde de forma más ajustada con el conjunto de características de nota.

A partir de la siguiente descripción de formas de realización preferidas que ilustran, a título de ejemplo, los principios de la invención, deberían ponerse de manifiesto otras características y ventajas de la presente invención.

Breve descripción de los dibujos

Se puede obtener una comprensión más exhaustiva de la naturaleza y las ventajas de la presente invención en referencia a los siguientes dibujos. En las figuras adjuntas, componentes o características similares pueden presentar la misma indicación de referencia. Además, varios componentes del mismo tipo se pueden diferenciar colocando a continuación de la indicación de referencia un guion y una segunda indicación que diferencia entre los componentes similares. Si en la memoria descriptiva se usa únicamente la primera indicación de referencia, la descripción es aplicable a uno cualquiera de los componentes similares que presentan la misma primera indicación de referencia con independencia de la segunda indicación de referencia.

La figura 1A proporciona un diagrama de bloques simplificado, de alto nivel, de un sistema según la presente invención.

La figura 1B proporciona un diagrama de bloques simplificado, de nivel inferior, de un sistema como el mostrado en la figura 1 según la presente invención.

La figura 2 proporciona un diagrama de flujo de un método ejemplificativo para convertir datos de señales de audio en datos de partituras según formas de realización de la invención.

La figura 3 proporciona un diagrama de flujo de un método ejemplificativo para la detección de la altura tonal según formas de realización de la invención.

5 La figura 4A proporciona un diagrama de flujo de un método ejemplificativo para la generación de eventos de inicio de notas según formas de realización de la invención.

La figura 4B proporciona un diagrama de flujo de un método ejemplificativo para determinar un evento de ataque según formas de realización de la invención.

10 La figura 5 proporciona una ilustración de una señal de audio con varias envolventes para su uso en la generación de eventos de inicio de notas según formas de realización de la invención.

La figura 6 proporciona un diagrama de flujo de un método ejemplificativo para la detección de la duración de notas según formas de realización de la invención.

15 La figura 7 proporciona una ilustración de una señal de audio con varias envolventes para su uso en la detección de duración de notas según formas de realización de la invención.

20 La figura 8 proporciona un diagrama de flujo de un método ejemplificativo para la detección de pausas según unas formas de realización de la invención.

La figura 9 proporciona un diagrama de flujo de un método ejemplificativo para la detección del tempo según ejemplos de realizaciones.

25 La figura 10 proporciona un diagrama de flujo de un método ejemplificativo para la determinación de valores de notas según ejemplos de realización.

La figura 11 proporciona una gráfica de datos ejemplificativos que ilustran este método ejemplificativo de detección de tempos.

30 La figura 12 proporciona datos ejemplificativos adicionales que ilustran el método ejemplificativo de detección de tempos mostrado en la figura 11.

35 La figura 13 proporciona un diagrama de flujo de un método ejemplificativo para la detección de tonalidades según ejemplos de realización.

Las figuras 14A y 14B proporcionan ilustraciones de dos funciones de coste de tonalidad, ejemplificativas, usadas en la detección de tonalidades según ejemplos de realización.

40 La figura 15 proporciona un diagrama de flujo de un método ejemplificativo para la determinación de la designación de alturas tonales de tonalidades según ejemplos de realización.

La figura 16 proporciona un diagrama de bloques de un sistema computacional 1600 para implementar ciertas formas de realización de la invención.

45 **Descripción detallada**

Esta descripción proporciona únicamente formas de realización ejemplificativas, y no pretende limitar el alcance, la aplicabilidad, o la configuración de la invención. Por el contrario, la consiguiente descripción de las formas de realización proporcionará a aquellos versados en la materia una descripción capacitadora para implementar formas de realización de la invención. Se pueden realizar varios cambios en la función y la disposición de elementos sin desviarse con respecto al alcance de la invención.

55 De este modo, varias formas de realización pueden omitir, sustituir, o añadir varios procedimientos o componentes según resulte apropiado. Por ejemplo, debería apreciarse que en formas de realización alternativas, los métodos se pueden llevar a cabo en un orden diferente al descrito, y las diversas etapas se pueden añadir, omitir, o combinar. Además, características descritas con respecto a ciertas formas de realización se pueden combinar en otras diversas formas de realización. Diferentes aspectos y elementos de las formas de realización se pueden combinar de una manera similar.

60 Debería apreciarse también que los siguientes sistemas, métodos y software pueden ser de manera individual o conjunta componentes de un sistema de mayor tamaño, pudiendo otros procedimientos tener prioridad sobre su aplicación o alternativamente pueden modificar esta última. Además, puede que se requieran varias etapas antes, después, o simultáneamente con las siguientes formas de realización.

65 La figura 1A muestra un diagrama de bloques simplificado, de alto nivel, de un sistema construido según la

invención, para extraer de manera automática y precisa datos de partitura a partir de una señal de audio según la invención. El sistema 100 recibe una señal de entrada de audio 104 en una unidad receptora de audio 106 y hace pasar la señal a través de una unidad procesadora de señales 110, una unidad procesadora de notas 130, y una unidad procesadora de partituras 150. A continuación, la unidad procesadora de partituras 150 puede generar una salida de partitura 170.

Según algunas formas de realización de la invención, el sistema 100 puede recibir una composición o interpretación en forma de una señal de entrada de audio 104 y generar la representación de partitura musical correspondiente 170 de la interpretación. La señal de entrada de audio 104 puede provenir de una interpretación en vivo o puede incluir una reproducción de una interpretación grabada, y puede involucrar tanto instrumentos musicales como la voz humana. Se pueden producir representaciones de partituras musicales 170 para cada uno de los diferentes instrumentos y voces que constituyen una señal de entrada de audio 104. La representación de partitura musical 170 puede proporcionar, por ejemplo, altura tonal, ritmo, timbre, dinámica, y/o cualquier otra información de partitura útil.

En algunas formas de realización, se discernirán instrumentos y voces, de manera individual o combinados, con respecto a los otros según las frecuencias a las cuales estén interpretando los instrumentos y voces (por ejemplo, a través de una diferenciación de registros) o distinguiendo entre diferentes timbres. Por ejemplo, en una orquesta, se pueden identificar y distinguir entre sí músicos individuales o grupos de músicos (por ejemplo, primeros violines o segundos violines, o violines y violonchelos) que estén interpretando en intervalos de frecuencia diferentes. De manera similar, se pueden usar matrices de micrófonos u otros detectores de audio para mejorar la resolución de la señal de entrada de audio recibida 104, para incrementar el número de pistas de audio o instrumentos incluidos en la señal de entrada de audio 104, o para proporcionar otra información para la señal de entrada de audio 104 (por ejemplo, información espacial o profundidad).

En una forma de realización, un micrófono o matriz de micrófonos 102 recibe en tiempo real una composición y la misma se transduce en una señal de entrada de audio, eléctrica, analógica 104 para su recepción por la unidad receptora de audio 106. En otras formas de realización, la señal de entrada de audio 104 puede comprender datos digitales, tales como un archivo musical grabado, adecuado para la reproducción. Si la señal de entrada de audio 104 es una señal analógica, la misma es convertida por la unidad receptora de audio 106 en una representación digital como preparación para el procesamiento de señales digitales por parte de la unidad procesadora de señales 110, la unidad procesadora de notas 130, y la unidad procesadora de partituras 150. Puesto que la señal de entrada se recibe en tiempo real, puede que no haya manera de predeterminar la duración total de la señal de entrada de audio 104. Como tal, la señal de entrada de audio 104 se puede recibir y almacenar en intervalos predeterminados (por ejemplo, una cantidad de tiempo transcurrido, número de muestras digitales, cantidades de memoria usadas, etcétera), y se puede procesar en consecuencia. En otra forma de realización, el receptor de audio 106 recibe un fragmento de sonido grabado y el mismo se digitaliza, presentando por lo tanto una duración de tiempo fija.

En algunas formas de realización, se puede usar una matriz de micrófonos para la detección de múltiples instrumentos que están tocando simultáneamente. Cada micrófono de la matriz se colocará de manera que esté más cerca de un instrumento particular que de cualquiera de los otros, y, por lo tanto, la intensidad de las frecuencias producidas por ese instrumento será mayor para ese micrófono que para cualquiera de los otros. La combinación de la información proporcionada por los cuatro detectores sobre el sonido recibido y completo, y el uso de las señales grabadas por todos los micrófonos, puede dar como resultado una representación abstracta digital de la composición, que podría imitar una representación MIDI de la grabación con la información sobre los instrumentos en este caso. La combinación de información incluirá información referente a la secuencia de alturas tonales o notas, con duración temporal de frecuencias (ritmo), series de sobretonos asociadas a la frecuencia fundamental (timbre: tipo de instrumento o voz específica), e intensidad relativa (dinámica). Alternativamente, se puede usar un único micrófono para recibir la salida de múltiples instrumentos u otras fuentes simultáneamente.

En varias formas de realización, información extraída de la señal de entrada de audio 104 se procesa para generar automáticamente una representación de partitura musical 170. Puede haber disponibles bibliotecas y paquetes de software convencionales para producir texto musical a partir de la representación de partitura musical 170. Muchas de estas herramientas aceptan una entrada en forma de una representación de la composición en un formato predeterminado, tal como la Interfaz Digital para Instrumentos Musicales (MIDI) o similares. Por lo tanto, algunas formas de realización del sistema generan una representación de partitura musical 170 que está sustancialmente en conformidad con la normativa MIDI para garantizar compatibilidad con dichas herramientas convencionales. Una vez que se ha creado la representación de partitura musical 170, las aplicaciones potenciales son múltiples. En varias formas de realización, a la partitura musical se le aplica una de las siguientes opciones: se visualiza en una pantalla del dispositivo, se imprime, se importa hacia programas de adición musical, se almacena, o se comparte con otros (por ejemplo, para un proyecto musical en colaboración).

Se apreciará que según la invención son posibles muchas implementaciones del sistema 100. En algunas formas de realización, el sistema 100 se implementa en forma de un dispositivo dedicado. El dispositivo puede incluir uno o más micrófonos internos, configurados para detectar presión acústica y convertirla en una señal de entrada de audio 104 para su uso por el sistema 100. Alternativamente, el dispositivo puede incluir uno o más puertos de entrada de audio para comunicación por interfaz con micrófonos externos, dispositivos de medios, medios de almacenamiento

de datos, u otras fuentes de audio. En ciertas formas de realización de las mencionadas, el dispositivo puede ser un dispositivo de mano o portátil. En otras formas de realización, el sistema 100 se puede implementar en un dispositivo multifuncional o de funcionalidad general (por ejemplo, en forma de módulos de software almacenados en un soporte legible por ordenador para su ejecución por parte de un ordenador). En ciertas formas de realización de las mencionadas, la fuente de audio 102 puede ser una tarjeta de sonido, un micrófono externo, o un archivo de audio almacenado. A continuación, se genera la señal de entrada de audio 104 y la misma se suministra al sistema 100.

Otras formas de realización del sistema 100 se pueden implementar como una versión simplificada o monoaural para su funcionamiento como dispositivo de dictado musical, el cual recibe audio de usuarios que tocan un instrumento o cantan una cierta canción o melodía o una parte de la misma a un micrófono. En la disposición de un solo micrófono, el sistema 100 traduce posteriormente la música grabada desde el micrófono a la partitura musical correspondiente. Esto puede proporcionar un equivalente musical al software de texto-a-voz que traduce palabras y frases habladas en texto legible por ordenador. Como conversión de sonido-a-notas, la canción o melodía se registrará como si se estuviera tocando un instrumento.

Se apreciará que diferentes implementaciones del sistema 100 también pueden incluir diferentes tipos de interfaces y funciones en relación con la compatibilidad con usuarios y otros sistemas. Por ejemplo, se pueden proporcionar puertos de entrada para entradas de nivel de línea (por ejemplo, provenientes de un sistema estereofónico o un amplificador de guitarra), entradas de micrófono, entradas de red (por ejemplo, de Internet), u otros componentes de audio digitales. De manera similar, se pueden proporcionar puertos de salida para dar salida hacia altavoces, componentes de audio, ordenadores, y redes, etcétera. Además, en algunas implementaciones, el sistema 100 puede proporcionar entradas de usuario (por ejemplo, teclados físicos o virtuales, correderas, botones, conmutadores, etcétera) y/o salidas de usuario (por ejemplo, pantallas, altavoces, etcétera). Por ejemplo, se pueden proporcionar capacidades de interfaz para permitir que un usuario escuche grabaciones o datos extraídos de las grabaciones por el sistema 100.

En la figura 1B se proporciona un diagrama de bloques de nivel inferior de una forma de realización del sistema 100. Se pueden usar una o más fuentes de audio 102 para generar una señal de entrada de audio. La fuente de audio 102 puede ser cualquier cosa con capacidad de proporcionar una señal de entrada de audio 104 al receptor de audio 106. En algunas formas de realización, como fuentes de audio 102 se usan uno o más micrófonos, transductores, y/u otros sensores. Los micrófonos pueden convertir ondas de presión o electromagnéticas provenientes de una interpretación en vivo (o la reproducción de una interpretación grabada) en una señal eléctrica para ser usada como señal de entrada de audio 104. Por ejemplo, en una interpretación de audio en vivo, se puede usar un micrófono para captar y convertir audio de un cantante, mientras que se pueden usar "captadores" ("pick-ups") electromagnéticos para captar y convertir audio proveniente de una guitarra y un bajo. En otras formas de realización, las fuentes de audio 102 pueden incluir dispositivos analógicos o digitales configurados para proporcionar una señal de entrada de audio 104 ó un archivo de audio del cual se puede leer una señal de entrada de audio 104. Por ejemplo, archivos de audio digitales se pueden almacenar en soportes de almacenamiento en un formato de audio y pueden ser suministrados por los soportes de almacenamiento en forma de una señal de entrada de audio 104 al receptor de audio 106.

Se apreciará que, en función de la fuente de audio 102, la señal de entrada de audio 104 puede tener diferentes características. La señal de entrada de audio 104 puede ser monofónica o polifónica, puede incluir múltiples pistas de datos de audio, puede incluir audio de muchos tipos de instrumentos, y puede incluir un cierto formato de archivo, etcétera. De manera similar, se apreciará que el receptor de audio 106 puede ser cualquier cosa con capacidad de recibir la señal de entrada de audio 104. Además, el receptor de audio 106 puede incluir uno o más puertos, decodificadores, u otros componentes necesarios para comunicarse por interfaz con las fuentes de audio 102, o recibir o interpretar la señal de entrada de audio 104.

El receptor de audio 106 puede proporcionar funcionalidad adicional. En una forma de realización, el receptor de audio 106 convierte señales de entrada de audio analógicas 104 en señales de entrada de audio digitales 104. En otra forma de realización, el receptor de audio 106 está configurado para convertir en sentido descendente la señal de entrada de audio 104 a una frecuencia de muestreo inferior con el fin de reducir la carga computacional para el sistema 100. En una forma de realización, la señal de entrada de audio 104 se submuestra a aproximadamente entre 8 y 9 kHz. Esto puede proporcionar una resolución frecuencial mayor de la señal de entrada de audio 104, y puede reducir ciertas restricciones sobre el diseño del sistema 100 (por ejemplo, especificaciones de los filtros).

Todavía en otra forma de realización, el receptor de audio 106 incluye un componente de detección de umbrales, configurado para comenzar a recibir la señal de entrada de audio 104 (por ejemplo, iniciar la grabación) al producirse la detección de niveles de audio que superen ciertos umbrales. Por ejemplo, el componente de detección de umbrales puede analizar el audio durante un periodo de tiempo especificado para detectar si la amplitud de la señal de entrada de audio 104 permanece por encima de un umbral predeterminado durante cierta cantidad de tiempo predeterminada. El componente de detección de umbrales puede estar configurado además para dejar de recibir la señal de entrada de audio 104 (por ejemplo, detener la grabación) cuando la amplitud de la señal de entrada de audio 104 cae por debajo de un umbral predeterminado durante una cantidad de tiempo predeterminada. Todavía en otra forma de realización, el componente de detección de umbrales se puede usar para generar una bandera para el

sistema 100, que represente la condición de que la amplitud de la señal de entrada de audio 104 supera o cae por debajo de un umbral durante una cantidad de tiempo, en lugar de comenzar o finalizar realmente la recepción de la señal de entrada de audio 104.

5 **Procesado de señales y notas**

Según la figura 1B, el receptor de audio 106 traslada la señal de entrada de audio 104 a la unidad procesadora de señales 110, la cual incluye una unidad de extracción de amplitud 112 y una unidad de extracción de frecuencia 114. La unidad de extracción de amplitud 112 está configurada para extraer de la señal de entrada de audio 104 información relacionada con la amplitud. La unidad de extracción de frecuencia 114 está configurada para extraer de la señal de entrada de audio 104 información relacionada con la frecuencia.

En una forma de realización, la unidad de extracción de frecuencia 114 transforma la señal del dominio del tiempo al dominio de la frecuencia usando un algoritmo de transformada. Por ejemplo, mientras se encuentra en el dominio del tiempo, la señal de entrada de audio 104 se puede representar como cambios de amplitud con respecto al tiempo. No obstante, después de aplicar un algoritmo de Transformada Rápida de Fourier (FFT), la misma señal de entrada de audio 104 se puede representar en forma de una gráfica de las amplitudes de cada uno de sus componentes de frecuencia (por ejemplo, la intensidad o contribución relativa de cada banda de frecuencias en un intervalo de frecuencias, como una serie de sobretonos, sobre el cual se procesará la señal). Para obtener eficiencia en el procesamiento, puede resultar deseable limitar el algoritmo a un cierto intervalo de frecuencias. Por ejemplo, el intervalo de frecuencias puede cubrir únicamente el espectro audible (por ejemplo, aproximadamente de 20 Hz a 20 kHz).

En varias formas de realización, la unidad procesadora de señales 110 puede extraer información relacionada con la frecuencia, de otras maneras. Por ejemplo, muchos algoritmos de transformada dan salida a una señal en "cubos" de frecuencia lineales de anchura fija. Esto puede limitar la potencial resolución o eficacia frecuencial de la transformada, teniendo en cuenta especialmente que la señal de audio puede ser de naturaleza inherentemente logarítmica (en lugar de lineal). En la técnica se conocen muchos algoritmos para extraer de la señal de entrada de audio 104 información relacionada con la frecuencia.

A continuación, la información relacionada con la amplitud, extraída por la unidad de extracción de amplitud 112, y la información relacionada con la frecuencia, extraída por la unidad de extracción de frecuencia 114, pueden ser usadas por diversos componentes de la unidad de procesado de notas 130. En algunas formas de realización, la unidad de procesado de notas 130 incluye la totalidad o parte de una unidad detectora de inicios de notas 132, una unidad detectora de duración de notas 134, una unidad detectora de alturas tonales 136, una unidad detectora de pausas 144, una unidad detectora de envolventes 138, una unidad detectora de timbres 140, y una unidad detectora de dinámicas de notas 142.

La unidad detectora de inicios de notas 132 está configurada para detectar el inicio de una nota. El inicio (o comienzo) de una nota se manifiesta típicamente en la música, en forma de un cambio de altura tonal (por ejemplo, un ligado), un cambio de amplitud (por ejemplo, una parte de incorporación de una envolvente), o alguna combinación de un cambio de altura tonal y amplitud. Como tal, la unidad detectora de inicios de notas 132 se puede configurar para generar un evento de inicio de nota cada vez que se produzca un cierto tipo de cambio de frecuencia (o altura tonal) y/o amplitud, tal como se describe de forma más detallada posteriormente en relación con las figuras 4 y 5.

Las notas musicales también se pueden caracterizar por su duración (por ejemplo, la cantidad de tiempo que dura una nota en segundos o número de muestras). En algunas formas de realización, la unidad de procesado de notas 130 incluye una unidad detectora de duración de notas 134, configurada para detectar la duración de una nota marcada por un evento de inicio de nota. La detección de duración de una nota se describe de forma más detallada posteriormente en relación con las figuras 6 y 7.

Merece la pena indicar que ciertas características de la música son psicoacústicas, por contraposición a atributos puramente físicos de una señal. Por ejemplo, la frecuencia es una propiedad física de una señal (por ejemplo, que representa el número de ciclos por segundo recorridos por una onda sinusoidal), pero la altura tonal es un fenómeno psicoacústico más complejo. Uno de los motivos es que una nota de una sola altura tonal tocada por un instrumento está constituida habitualmente por varias frecuencias, cada una con una amplitud diferente, conocidas como timbre. El cerebro puede captar una de esas frecuencias (por ejemplo, típicamente la frecuencia fundamental) como "altura tonal", mientras que capta las otras frecuencias meramente como la adición de "color armónico" a la nota. En algunos casos, la altura tonal de una nota experimentada por un oyente puede ser una frecuencia que esté ausente, en su mayoría o completamente, de la señal.

En algunas formas de realización, la unidad de procesado de notas 130 incluye una unidad detectora de alturas tonales 136, configurada para detectar la altura tonal de una nota marcada por un evento de inicio de nota. En otras formas de realización, la unidad detectora de alturas tonales 136 está configurada para realizar un seguimiento de la altura tonal de la señal de entrada de audio 104, en lugar de (o además de) realizar el seguimiento de las alturas tonales de notas individuales. Se apreciará que la unidad detectora de alturas tonales 136 puede ser usada por la

unidad detectora de inicios de notas 132 en algunos casos para determinar un cambio de altura tonal de la señal de entrada de audio 104 que supera un valor de umbral.

5 Ciertas formas de realización de la unidad detectora de alturas tonales 136 procesan además alturas tonales de manera que sean más compatibles con una representación de partitura musical final 170. Se describen más detalladamente formas de realización de la detección de alturas tonales con respecto a la figura 3.

10 Algunas formas de realización de la unidad de procesado de notas 130 incluyen una unidad detectora de pausas 144 configurada para detectar la presencia de pausas dentro de la señal de entrada de audio 104. Una forma de realización de la unidad detectora de pausas 144 usa información relacionada con la amplitud, extraída por la unidad de extracción de amplitud 112 e información de confianza obtenida por la unidad detectora de alturas tonales 136. Por ejemplo, la información relacionada con la amplitud puede revelar que la amplitud de la señal de entrada de audio 104 es relativamente baja (por ejemplo, en el ruido de fondo o cerca de este último) durante cierta ventana de tiempo. Durante la misma ventana de tiempo, la unidad detectora de alturas tonales 136 puede determinar que existe una confianza muy baja con respecto a la presencia de cualquier altura tonal particular. Usando esta y otra información, la unidad detectora de pausas 144 detecta la presencia de una pausa, y una localización temporal en la que comenzó probablemente la pausa. Se describen adicionalmente formas de realización de la detección de pausas con respecto a las figuras 9 y 10.

20 En algunas formas de realización, la unidad de procesado de notas 130 incluye una unidad detectora de timbres 140. La información relacionada con la amplitud, extraída por la unidad de extracción de amplitud 112, y la información relacionada con la frecuencia, extraída por la unidad de extracción de frecuencia 114, pueden ser usadas por la unidad detectora de timbres 140 para detectar información de timbres para una parte de la señal de entrada de audio 104. La información de timbres puede revelar la composición armónica de la parte de la señal de audio 104. En algunas formas de realización, la unidad detectora de timbres 140 puede detectar información de timbres referente a una nota particular que comienza en un evento de inicio de nota.

30 En una forma de realización de la unidad detectora de timbres 140, la información relacionada con la amplitud y la información relacionada con la frecuencia se convolucionan con un filtro Gaussiano para generar un espectro filtrado. A continuación, el espectro filtrado se puede usar para generar una envolvente en torno a una altura tonal detectada por la unidad detectora de alturas tonales 136. Esta envolvente se puede corresponder con el timbre de la nota a esa altura tonal.

35 Algunas formas de realización, la unidad de procesado de notas 130 incluye una unidad detectora de envolventes 138. La información relacionada con la amplitud, extraída por la unidad de extracción de amplitud 112, puede ser usada por la unidad detectora de envolventes 138 para detectar información de envolventes correspondiente a una parte de la señal de entrada de audio 104. Por ejemplo, el toque de una tecla en un piano puede provocar que un macillo golpee un conjunto de cuerdas, dando como resultado una señal de audio con una amplitud de ataque grande. Esta amplitud pasa rápidamente a través de un decaimiento, hasta que se sostiene en una amplitud de un estado algo estable en el que las cuerdas resuenan (evidentemente, la amplitud puede disminuir lentamente a través de esta parte de la envolvente a medida que se consume la energía de las cuerdas). Finalmente, cuando se suelta la tecla del piano, un apagador aterriza sobre las cuerdas, provocando que la amplitud caiga rápidamente hasta cero. A este tipo de envolvente se le hace referencia típicamente como envolvente ADSR (ataque, decaimiento, sostenimiento, relajación). La unidad detectora de envolventes 138 se puede configurar para detectar parte o la totalidad de las partes de una envolvente ADSR, o cualquier otro tipo de información útil de la envolvente.

50 En varias formas de realización, la unidad de procesado de notas 130 incluye también una unidad detectora de dinámicas de notas 142. En ciertas formas de realización, la unidad detectora de dinámicas de notas 142 proporciona una funcionalidad similar a la unidad detectora de envolventes 138 para notas específicas que comienzan en ciertos eventos de inicio de notas. En otras formas de realización, la unidad detectora de dinámicas de notas 142 está configurada para detectar envolventes de notas que o bien son anómalas con respecto a un patrón de envolventes que son detectadas por la unidad detectora de envolventes 138 o bien encajan en un cierto patrón predefinido. Por ejemplo, una nota en staccato puede estar caracterizada por partes de ataque brusco y sostenimiento corto de su envolvente ADSR. En otro ejemplo, una nota acentuada puede estar caracterizada por una amplitud del ataque significativamente mayor que las de notas circundantes.

60 Se apreciará que la unidad detectora de dinámicas de notas 142 y otras unidades de procesado de notas se pueden usar para identificar otros muchos atributos de una nota lo cual puede resultar deseable como parte de una representación de partitura musical 170. Por ejemplo, las notas se pueden marcar como ligadas, como acentuadas, como staccato, como notas de adorno, etcétera. Se pueden extraer muchas otras características de las notas según la invención.

Procesado de partituras

65 Se puede usar información referente a múltiples notas o eventos de inicio de notas (incluyendo pausas) para generar otra información. Según la forma de realización de la figura 1B, varios componentes de la unidad de procesado de

5 notas 130 pueden estar en comunicación operativa con varios componentes de la unidad de procesado de partituras 150. La unidad de procesado de partituras 150 puede incluir la totalidad o parte de una unidad de detección de tempos 152, una unidad de detección de métricas 154, una unidad de detección de tonalidades 156, una unidad de identificación de instrumentos 158, una unidad de detección de pistas 162, y una unidad de detección de dinámicas globales 164.

10 En algunas formas de realización, la unidad de procesado de partituras 150 incluye una unidad de detección de tempos 152, configurada para detectar el tempo de la señal de entrada de audio 104 durante una ventana de tiempo. Típicamente, el tempo de una pieza musical (por ejemplo, la velocidad a la cual parece pasar psicoacústicamente la música) se puede ver afectado en parte por la presencia y duración de notas y pausas. Como tal, ciertas formas de realización de la unidad de detección de tempos 152 usan información de la unidad detectora de inicios de notas 132, la unidad detectora de duraciones de notas 134, y la unidad detectora de pausas 144 para determinar el tempo. Otras formas de realización de la unidad de detección de tempos 152 usan además el tempo determinado para asignar valores de nota (por ejemplo, negras, corcheas, etcétera) a notas y pausas. Se describen de forma más detallada operaciones ejemplificativas de la unidad de detección de tempos 152 con respecto a las figuras 11 a 15.

20 La métrica dictamina cuántos pulsos hay en cada compás musical, y qué valor de nota se considera un pulso único. Por ejemplo, una métrica de 4/4 representa que cada compás tiene cuatro pulsos (el numerador) y que un pulso individual se representa por medio de una negra (el denominador). Por este motivo, la métrica puede ayudar a determinar posiciones de notas y de barras de compas, y otra información que pueda ser necesaria para proporcionar una representación de partitura musical útil 170. En algunas formas de realización, la unidad de procesado de partituras 150 incluye una unidad de detección de métricas 154, configurada para detectar la métrica de la señal de entrada de audio 104.

25 En algunas formas de realización, se deducen métricas sencillas a partir de información del tempo y de valores de notas extraídos por la unidad de detección de tempos 152 y a partir de otra información (por ejemplo, información de dinámicas de notas extraída por la unidad detectora de dinámicas de notas 142). No obstante, habitualmente la determinación de la métrica es una tarea compleja que conlleva un reconocimiento de patrones complejos.

30 Por ejemplo, considérese que se extrae la siguiente secuencia de valores de notas a partir de la señal de entrada de audio 104: negra, negra, corchea, corchea, corchea, corchea. Esta secuencia sencilla se podría representar como un compás de 4/4, dos compases de 2/4, cuatro compases de 1/4, un compás de 8/8, o muchas otras métricas. Suponiendo que había un acento (por ejemplo, una amplitud de ataque aumentada) en la primera negra y la primera corchea, esto puede hacer que sea más probable que la secuencia sea una de entre dos compases de 2/4, dos compases de 4/8, o un compás de 4/4. Además, suponiendo que 4/8 es una métrica muy inhabitual puede bastar con eliminar la misma como conjetura. Aún más, el hecho de saber que el género de la entrada de audio 104 es canción popular puede aumentar la posibilidad de que 4/4 sea el candidato más probable para la métrica.

40 El ejemplo anterior ilustra las complejidades implicadas incluso con una secuencia de valores de nota muy sencilla. Muchas secuencias de notas son mucho más complejas, implicando muchas notas de diferentes valores, notas que abarcan múltiples compases, notas con puntillo y de adorno, síncope, y otras dificultades en la interpretación de la métrica. Por este motivo, los algoritmos informáticos tradicionales pueden encontrar dificultad en determinar de manera precisa la métrica. Como tal, varias formas de realización de la unidad de detección de métricas 154 usan una red neuronal artificial (ANN) 0160, entrenada para detectar esos patrones complejos. La ANN 0160 se puede entrenar proporcionando a la ANN 0160 muchas muestras de diferentes métricas y funciones de coste que mejoran la precisión con cada muestra. En algunas formas de realización, la ANN 0160 se entrena usando un paradigma de aprendizaje. El paradigma de aprendizaje puede incluir por ejemplo, algoritmos de aprendizaje supervisado, aprendizaje no supervisado, o aprendizaje por refuerzo.

50 Se apreciará que se pueden generar muchos tipos útiles de información para ser usados por la representación de partitura musical 170 usando información o bien de tempo o bien de métrica o ambas informaciones. Por ejemplo, la información puede permitir una determinación de en dónde agrupar notas por compases (por ejemplo, como conjuntos de corcheas) en lugar de designar las notas individualmente con corchetes; cuándo dividir una nota entre dos compases y fundirla en una; o cuándo designar conjuntos de notas como tresillos (o conjuntos de orden superior), notas de adorno, trinos o mordentes, *glissandos*; etcétera.

60 Otro conjunto de información que puede resultar útil en la generación de una representación de partitura musical 170 se refiere a la tonalidad de una sección de la señal de entrada de audio 104. La información de tonalidad puede incluir, por ejemplo, una altura tonal fundamental identificada y un modo asociado. Por ejemplo, "La menor" representa que la altura tonal fundamental de la tonalidad es "La" y el modo es menor. Cada tonalidad se caracteriza por una armadura de clave, que identifica las notas que se encuentran "en la tonalidad" (por ejemplo, parte de la escala diatónica asociada a la tonalidad) y "fuera de la tonalidad" (por ejemplo, alteraciones en el paradigma de la tonalidad). "La menor", por ejemplo, no contiene sostenidos ni bemoles, mientras que "Re mayor" contiene dos sostenidos y ningún bemol.

65 En algunas formas de realización, la unidad de procesado de partituras 150 incluye una unidad de detección de

tonalidades 156, configurada para detectar la tonalidad de la señal de entrada de audio 104. Algunas formas de realización de la unidad de detección de tonalidades 156 determinan la tonalidad basándose en la comparación de secuencias de alturas tonales con un conjunto de funciones de coste. Las funciones de coste pueden buscar, por ejemplo, la minimización del número de alteraciones en una pieza musical durante una ventana de tiempo especificada. En otras formas de realización, la unidad de detección de tonalidades 156 puede usar una red neuronal artificial para realizar determinaciones de tonalidades complejas o mejorar la precisión de estas últimas. Todavía en otras formas de realización, una secuencia de cambios de tonalidad se puede evaluar con respecto a funciones de coste para mejorar la precisión de determinaciones de la tonalidad. Aún en otras formas de realización, la información de tonalidad obtenida por la unidad de detección de tonalidades 156 se puede usar para atribuir a notas (o eventos de inicio de notas) designaciones particulares de alturas tonales de una tonalidad. Por ejemplo, un "Si" en Fa mayor se puede designar como "Si natural". Evidentemente, la información de tonalidad se puede usar para generar una armadura de clave u otra información para la representación de la partitura musical. En algunas formas de realización, la información de tonalidad se puede usar además para generar información de acordes u otra información armónica. Por ejemplo, se pueden generar acordes de guitarra en formato de tablatura, o se pueden proporcionar acordes de jazz. En relación con las figuras 13 a 15, se describen de forma más detallada operaciones ejemplificativas de la unidad de detección de tonalidades 156.

En otras formas de realización, la unidad de procesado de partituras 150 incluye también una unidad de identificación de instrumentos 158, configurada para identificar un instrumento que se está tocando en la señal de entrada de audio 104. Normalmente, se dice que un instrumento tiene un timbre particular. No obstante, puede haber diferencias de timbre en un instrumento individual en función de la nota que se esté tocando o de la manera que en la que se esté tocando la nota. Por ejemplo, el timbre de cada violín es diferente sobre la base de, por ejemplo, los materiales usados en su construcción, el tacto del intérprete, la nota que se esté tocando (por ejemplo, una nota tocada en una cuerda al aire tiene un timbre diferente con respecto a la misma nota tocada en una cuerda pulsada, y una nota baja en el registro del violín tiene un timbre diferente de una nota en el registro superior), si la nota se está tocando con arco o con pizzicato, etcétera. Sin embargo, aun así puede haber una similitud suficiente entre notas del violín para identificarlas como violines, por contraposición a otro instrumento.

Formas de realización de la unidad de identificación de instrumentos 158 se configuran para comparar características de notas individuales o múltiples con el fin de determinar el intervalo de alturas tonales que están siendo tocadas aparentemente por un instrumento de la señal de entrada de audio 104, el timbre que está siendo producido por el instrumento en cada una de esas alturas tonales, y/o la envolvente de amplitud de notas que están siendo tocadas en el instrumento. En una forma de realización, las diferencias de timbre se usan para detectar diferentes instrumentos mediante la comparación de timbres distintivos típicos de muestras de instrumentos con timbres detectados de la señal de entrada de audio 104. Por ejemplo, incluso cuando se esté tocando la misma nota con el mismo volumen durante el mismo tiempo, un saxofón y un piano pueden sonar muy diferente debido a sus timbres diferentes. Evidentemente, tal como se ha mencionado más arriba, las identificaciones basadas solamente en el timbre pueden tener una precisión limitada.

En otra forma de realización, se usan intervalos de alturas tonales para detectar instrumentos diferentes. Por ejemplo, un violonchelo típicamente puede reproducir notas que van desde aproximadamente dos octavas por debajo del Do central hasta aproximadamente una octava por encima del Do central. No obstante, un violín puede reproducir típicamente notas que van desde justo por debajo del Do central hasta aproximadamente cuatro octavas por encima del Do central. Así, aun cuando un violín y un violonchelo pueden tener timbres similares (los dos son instrumentos de cuerda frotada), los intervalos de sus alturas tonales pueden ser suficientemente diferentes para ser usados para la identificación. Evidentemente, puede haber posibilidad de errores dado que los intervalos se solapan en cierta medida. Además, otros instrumentos (por ejemplo, el piano) tienen intervalos mayores, los cuales se pueden solapar con muchos instrumentos.

Todavía en otra forma de realización, se usa la detección de envolvente para identificar instrumentos diferentes. Por ejemplo, una nota tocada en un instrumento percutado con macillo (por ejemplo, un piano) puede sonar diferente de la misma nota tocada en un instrumento de las maderas (por ejemplo, una flauta), de caña (por ejemplo, un oboe), de los metales (por ejemplo, una trompeta), o de cuerda (por ejemplo, un violín). No obstante, cada instrumento puede tener la capacidad de producir muchos tipos diferentes de envolvente, en función de cómo se toque la nota. Por ejemplo, un violín se puede tocar con pizzicato o con el arco, o una nota se puede tocar con ligado o staccato.

Como mínimo debido a las dificultades antes mencionadas, la identificación precisa de instrumentos puede requerir una detección de patrones complejos, lo cual conlleva múltiples características de la señal de entrada de audio 104 posiblemente durante múltiples notas. Como tal, algunas formas de realización de la unidad de identificación de instrumentos 158 utilizan una red neuronal artificial entrenada para detectar combinaciones de estos patrones complejos.

Algunas formas de realización de la unidad de procesado de partituras 150 incluyen una unidad de detección de pistas 162, configurada para identificar una pista de audio de entre la señal de entrada de audio 104. En algunos casos, la señal de entrada de audio 104 puede estar en un formato que ya esté separado por pistas. Por ejemplo, el audio de algunas Cintas de Audio Digital (DATs) se puede almacenar en forma de ocho pistas de audio digitales

independientes. En estos casos, la unidad de detección de pistas 162 se puede configurar simplemente para identificar las pistas de audio individuales.

No obstante, en otros casos, múltiples pistas se pueden almacenar en una única señal de entrada de audio 104 y es necesario identificarlas mismas extrayendo ciertos datos de la señal de entrada de audio. Como tal, algunas formas de realización de la unidad de detección de pistas 162 están configuradas para usar información extraída del archivo de entrada de audio 104 con el fin de identificar pistas de audio independientes. Por ejemplo, una interpretación puede incluir cinco instrumentos que suenan simultáneamente (por ejemplo, un quinteto de jazz). Puede resultar deseable identificar dichos instrumentos independientes como pistas independientes para poder representar de manera precisa la interpretación en una representación de partitura musical 170.

La detección de pistas se puede lograr de varias maneras diferentes. En una forma de realización, la unidad de detección de pistas 162 usa la detección de alturas tonales para determinar si diferentes secuencias de notas parecen restringidas a ciertos intervalos de alturas tonales. En otra forma de realización, la unidad de detección de pistas 162 usa información de identificación de instrumentos proveniente de la unidad de identificación de instrumentos 158 para determinar pistas diferentes.

Muchas partituras también contienen información referente a la dinámica global de una composición o interpretación. Dinámica global se refiere a dinámica que abarca más de una nota, por contraposición a la dinámica de una nota antes descrita. Por ejemplo, una pieza completa o sección de una pieza se puede señalar como *forte* (intenso) o *piano* (suave). En otro ejemplo, una secuencia de notas puede acrecentarse gradualmente en un *crescendo*. Para generar este tipo de información, algunas formas de realización de la unidad de procesamiento de partituras 150 incluyen una unidad de detección de dinámicas globales 164. Formas de realización de la unidad de detección de dinámicas globales 164 usan información de amplitud, en algunos casos incluyendo información de dinámicas de notas y/o información de envolventes, para detectar dinámicas globales.

En ciertas formas de realización, valores de umbral se predeterminan o generan de manera adaptativa a partir de la señal de entrada de audio 104 para ayudar en determinaciones de las dinámicas. Por ejemplo, el volumen medio de una interpretación rock se puede considerar *forte*. Las amplitudes que superan ese valor medio en cierta magnitud (por ejemplo, en un umbral, una desviación estándar, etcétera) se pueden considerar *fortissimo*, mientras que las amplitudes que caen por debajo de ese valor medio en cierta magnitud se pueden considerar *piano*.

Ciertas formas de realización pueden considerar además el espacio de tiempo durante el cual se producen cambios dinámicos. Por ejemplo, una pieza que comienza con dos minutos de notas discretas y repentinamente cambia a una sección *forte*. Por otro lado, una pieza discreta que va aumentando durante el transcurso de unas pocas notas, permanece en ese volumen más alto durante unas pocas notas más, y a continuación vuelve a la amplitud original, se puede considerar que tiene un *crescendo* seguido por un *decrescendo*.

La totalidad de los diversos tipos de información antes descritos, y cualquier otra información útil, se puede generar para su uso como representación de partitura musical 170. Esta representación de partitura musical 170 se puede memorizar o se puede dar salida a la misma. En ciertas formas de realización, a la representación de partitura musical 170 se le da salida hacia un software de generación de partituras, el cual puede transcribir los diversos tipos de información en un formato de partituras. El formato de las partituras se puede configurar para visionarlo, imprimirlo, transmitirlo electrónicamente, etcétera.

Se apreciará que las diversas unidades y componentes antes descritos se pueden implementar de varias maneras sin desviarse con respecto a la invención. Por ejemplo, ciertas unidades pueden ser componentes de otras unidades, o se pueden implementar como una funcionalidad adicional de otra unidad. Además, las unidades se pueden conectar de muchas maneras, y pueden fluir datos entre ellas de muchas maneras según la invención. Como tal, la figura 1B se debería considerar como ilustrativa, y no debería interpretarse como limitativa del alcance de la invención.

Métodos para el procesamiento de audio

La figura 2 proporciona un diagrama de flujo de un método ejemplificativo para convertir datos de señales de audio en datos de partituras según formas de realización de la invención. El método 200 comienza en el bloque 202 mediante la recepción de una señal de audio. En algunas formas de realización, la señal de audio se puede haber sometido a un preprocesado. Por ejemplo, la señal de audio se puede convertir de analógica a digital, se puede convertir en sentido descendente a una frecuencia de muestreo inferior, se puede transcodificar para obtener compatibilidad con ciertos codificadores o decodificadores, se puede analizar sintácticamente para obtener pistas de audio monofónicas, o se puede aplicar cualquier otro preprocesado útil.

En el bloque 204, se puede extraer información de frecuencia de la señal de audio y se pueden identificar ciertos cambios de frecuencia. En el bloque 206, se puede extraer información de amplitud de la señal de audio y se pueden identificar ciertos cambios de amplitud.

En algunas formas de realización, se obtiene información de altura tonal en el bloque 208 a partir de la información de frecuencia extraída de la señal de entrada de audio en el bloque 204. Se describen más detalladamente con respecto a la figura 3 formas de realización ejemplificativas de la detección de alturas tonales en el bloque 208. Además, en algunas formas de realización, la información extraída e identificada referente a frecuencia y amplitud se usa para generar eventos de inicio de notas en el bloque 210. En relación con las figuras 4 a 5 se escriben más exhaustivamente formas de realización ejemplificativas de la generación de eventos de inicio de notas en el bloque 210.

En algunas formas de realización del método 200, la información de frecuencia extraída en el bloque 204, la información de amplitud extraída en el bloque 206, y los eventos de inicio de notas generados en el bloque 210 se usan para extraer y procesar otra información de la señal de audio. En ciertas formas de realización, la información se usa para determinar duraciones de notas en el bloque 220, para determinar pausas en el bloque 230, para determinar tempos durante ventanas de tiempo en el bloque 240, para determinar tonalidades durante ventanas en el bloque 250, y para determinar la instrumentación en el bloque 260. En otras formas de realización, las duraciones de notas determinadas en el bloque 220, las pausas determinadas en el bloque 230, y los tempos determinados en el bloque 240 se usan para determinar valores de notas en el bloque 245; las tonalidades determinadas en el bloque 250 se usan para determinar designaciones de alturas tonales de la tonalidad en el bloque 255; y la instrumentación determinada en el bloque 260 se usa para determinar pistas en el bloque 270. En varias formas de realización, las salidas de los bloques 220 a 270 están configuradas para usarse con el fin de generar datos de representación de partituras musicales en el bloque 280. En referencia a las figuras 6 a 15 se describen de forma más detallada métodos ejemplificativos correspondientes a los bloques 220 a 255.

Detección de alturas tonales

La figura 3 proporciona un diagrama de flujo de un método ejemplificativo para la detección de la altura tonal según formas de realización de la invención. La percepción humana de la altura tonal es un fenómeno psicoacústico. Por lo tanto, algunas formas de realización del método 208 comienzan en el bloque 302 mediante el pre-filtrado de una señal de entrada de audio con un banco de filtros psicoacústicos. El pre-filtrado en el bloque 302 puede conllevar, por ejemplo, una escala de ponderación que simula el campo auditivo del oído humano. Dichas escalas de ponderación son conocidas para aquellos versados en la materia.

A continuación, el método 208 puede continuar en el bloque 304 dividiendo la señal de entrada de audio 104 en intervalos predeterminados. Estos intervalos se pueden basar en eventos de inicio de notas, frecuencia de muestreo de la señal, o cualquier otro intervalo útil. En función del tipo de intervalo, formas de realización del método 208 se pueden configurar, por ejemplo, para detectar la altura tonal de una nota marcada por un evento de inicio de nota o para realizar el seguimiento de cambios de altura tonal en la señal de entrada de audio.

Para cada intervalo, el método 208 puede detectar una frecuencia fundamental en el bloque 306. La frecuencia fundamental se puede asignar como una "altura tonal" del intervalo (o de la nota). La frecuencia fundamental es normalmente la frecuencia significativa más baja, y la frecuencia que presenta la mayor intensidad, aunque no siempre.

El método 208 puede procesar además las alturas tonales para que sean más compatibles con una representación de partitura musical final. Por ejemplo, la representación de partitura musical puede requerir un conjunto bien definido y finito de alturas tonales, representado por las notas que constituyen la partitura. Por lo tanto, formas de realización del método 208 pueden separar un espectro frecuencial en compartimentos asociados a notas musicales particulares. En una forma de realización, el método 208 calcula la energía de cada uno de los compartimentos e identifica el compartimento con la energía significativa más baja como frecuencia de la altura tonal fundamental. En otra forma de realización, el método 208 calcula una serie de sobretonos de la señal de entrada de audio basándose en la energía de cada uno de los compartimentos, y usa las series de sobretonos para determinar la frecuencia de la altura tonal fundamental.

En una forma de realización ejemplificativa, el método 208 utiliza un banco de filtros que tiene un conjunto de filtros de solape uniforme, con un ancho de dos octavas. Cada banco de filtros se aplica a una parte de la señal de entrada de audio. La salida de cada banco de filtros se analiza para determinar si la parte filtrada de la señal de entrada de audio es suficientemente sinusoidal para contener esencialmente una única frecuencia. De esta manera, el método 208 puede tener la capacidad de extraer la frecuencia fundamental de la señal de entrada de audio durante un cierto intervalo de tiempo, como altura tonal de la señal en ese intervalo. En ciertas formas de realización, el método 208 se puede configurar para obtener la frecuencia fundamental de la señal de entrada de audio durante un intervalo, incluso cuando la frecuencia fundamental falte en la señal (por ejemplo, usando relaciones geométricas entre las series de sobretonos de frecuencias presentes en la señal de entrada de audio durante esa ventana).

En algunas formas de realización, el método 208 usa una serie de salidas de bancos de filtros para generar un conjunto de muestras de audio en el bloque 308. Cada muestra de audio puede tener un registro de datos asociado, que incluye, por ejemplo, información referente a frecuencia estimada, valores de confianza, sellos de tiempo,

duraciones e índices de teclas de piano. Se apreciará que en la técnica se conocen muchas formas de extracción de esta información del registro de datos a partir de la señal de entrada de audio. En “Real time voice processing with audiovisual feedback: toward autonomous agents with perfect pitch”, *Advances in Neural Information Processing Systems (NIPS) 15*, págs. 1.205 a 1.212 (2002), de Lawrence Saul, Daniel Lee, Charles Isbell, y Yaun LeCun, que se incorpora a la presente memoria como referencia a todos efectos, se detalla un planteamiento ejemplificativo. La información del registro de datos para las muestras de audio se puede almacenar temporalmente y clasificar para determinar qué altura tonal oíría un oyente.

Algunas formas de realización del método 208 continúan en el bloque 310 determinando dónde se produjo el cambio de altura tonal. Por ejemplo, si las alturas tonales se separan en compartimentos musicales (por ejemplo, tonos de una escala), puede resultar deseable determinar en qué lugar cruzó la altura tonal de la señal de audio de un compartimento al siguiente. De lo contrario, el vibrato, el trémolo y otros efectos musicales se pueden identificar erróneamente como cambios de altura tonal. La identificación del comienzo de un cambio de altura tonal también puede ser útil en la determinación de eventos de inicio de notas, tal como se describe a continuación.

15 **Detección de inicios de notas**

Muchos elementos de una composición musical se caracterizan, por lo menos parcialmente, por los comienzos de notas. En una partitura, por ejemplo, puede resultar necesario saber dónde comienzan las notas para determinar la colocación temporal correcta de notas en compases, el tempo y la métrica de una composición, y otra información importante. Algunas interpretaciones musicales expresivas conllevan cambios de nota que implican determinaciones subjetivas de dónde comienzan las notas (por ejemplo, debido a ligados lentos de una nota a otra). No obstante, la generación de partituras puede imponer una determinación más objetiva de dónde comienzan y finalizan las notas. A estos comienzos de nota se les hace referencia en la presente como eventos de inicio de notas.

La figura 4A proporciona un diagrama de flujo de un método ejemplificativo para la generación de eventos de inicio de notas según formas de realización de la invención. El método 210 comienza en el bloque 410 mediante la identificación de eventos de cambio de altura tonal. En algunas formas de realización, los eventos de cambio de altura tonal se determinan en el bloque 410 sobre la base de cambios de información de frecuencia 402 extraída de la señal de audio (por ejemplo, como en el bloque 204 de la figura 2) que superan un primer valor de umbral 404. En algunas formas de realización del método 210, el evento de cambio de altura tonal se identifica usando el método descrito en referencia al bloque 208 de la figura 2.

Mediante la identificación de eventos de cambio de altura tonal en el bloque 410, el método 210 puede detectar eventos de inicio de notas en el bloque 450 siempre que se produzca un cambio suficiente de la altura tonal. De esta manera, incluso un ligado lento de una altura tonal a otra, sin ningún cambio detectable de amplitud, generaría un evento de inicio de nota en el bloque 450. No obstante, el uso de solamente la detección de alturas tonales no conseguiría detectar una altura tonal repetida. Si un intérprete fuera a tocar la misma altura tonal múltiples veces consecutivamente, no se produciría ningún cambio en la altura tonal para señalar un evento de cambio de la altura tonal en el bloque 410, y ninguna generación de un evento de inicio de nota en el bloque 450.

Por lo tanto, formas de realización del método 210 identifican también eventos de ataque en el bloque 420. En algunas formas de realización, los eventos de ataque se determinan en el bloque 420 basándose en cambios de la información de amplitud 406 extraída de la señal de audio (por ejemplo, como en el bloque 206 de la figura 2) que superan un segundo valor de umbral 408. Un evento de ataque puede ser un cambio de la amplitud de la señal de audio de un carácter tal que señalice el inicio de una nota. Mediante la identificación de eventos de ataque en el bloque 420, el método 210 puede detectar eventos de inicio de notas en el bloque 450 cada vez que se produzca un cambio característico de la amplitud. De esta manera, incluso una altura tonal repetida generaría un evento de inicio de nota en el bloque 450.

Se apreciará que son posibles muchas formas de detección de un evento de ataque. La figura 4B proporciona un diagrama de flujo de un método ejemplificativo para determinar un evento de ataque según formas de realización de la invención. El método 420 comienza usando información de amplitud 406 extraída de la señal de audio para generar una primera señal de envolvente en el bloque 422. La primera señal de envolvente puede representar una “envolvente rápida” que realiza un seguimiento de cambios, a nivel de envolvente, de la amplitud de la señal de audio.

En algunas formas de realización, la primera señal de envolvente se genera en el bloque 422 en primer lugar rectificando y filtrando la información de amplitud 406. En una forma de realización, se toma un valor absoluto de la amplitud de la señal, que a continuación se rectifica usando un rectificador de onda completa para generar una versión rectificada de la señal de audio. A continuación, la primera señal de envolvente se puede generar filtrando la señal rectificada con el uso de un filtro paso-bajo. Esto puede producir una primera señal de envolvente que contiene sustancialmente la forma global de la señal de audio rectificada.

En el bloque 424 se puede generar una segunda señal de envolvente. La segunda señal de envolvente puede representar una “envolvente lenta” que constituye una aproximación de la potencia media de la envolvente de la

señal de audio. En algunas formas de realización, la segunda señal de envolvente se puede generar en el bloque 424 calculando la potencia media de la primera señal de envolvente o bien de manera continua o bien sobre intervalos de tiempo predeterminados (por ejemplo, integrando la señal). En ciertas formas de realización, los segundos valores de umbral 408 se pueden obtener a partir de los valores de la segunda señal de envolvente en posiciones de tiempo dadas.

En el bloque 426, se genera una señal de control. La señal de control puede representar cambios direccionales más significativos en la primera señal de envolvente. En una forma de realización, la señal de control se genera en el bloque 426: (1) hallando la amplitud de la primera señal de envolvente en una primera localización temporal; (2) continuando en esa amplitud hasta una segunda localización temporal (por ejemplo, la primera y la segunda posiciones de tiempo están separadas por una cantidad de tiempo predeterminada); y (3) fijando la segunda localización temporal como localización temporal nueva y repitiendo el proceso (es decir, cambiando a la amplitud nueva en la segunda localización temporal y permaneciendo allí durante la cantidad de tiempo predeterminada).

A continuación, el método 420 identifica como evento de ataque en el bloque 428 cualquier localización en la que la señal de control se haga mayor que (por ejemplo, cruce en una localización positiva) la segunda señal de envolvente. De esta manera, únicamente se pueden identificar eventos de ataque cuando se produce un cambio significativo en la envolvente. En la figura 5 se muestra una ilustración ejemplificativa de este método 420.

La figura 5 proporciona una ilustración de una señal de audio con varias envolventes para su uso en la generación de eventos de inicio de notas según formas de realización de la invención. La gráfica ilustrativa 500 representa la amplitud con respecto al tiempo para la señal de entrada de audio 502, la primera señal de envolvente 504, la segunda señal de envolvente 506, y la señal de control 508. La gráfica ilustra también posiciones de eventos de ataque 510 en los que la amplitud de la señal de control 508 se hace mayor que en la amplitud de la segunda señal de envolvente 506.

Detección de duraciones de notas

Una vez que se ha identificado el comienzo de una nota generando un evento de inicio de nota, puede resultar útil determinar dónde finaliza la nota (o la duración de la misma). La figura 6 proporciona un diagrama de flujo de un método ejemplificativo para la detección de duración de notas según formas de realización de la invención. El método 220 comienza mediante la identificación de una primera localización de inicio de nota en el bloque 602. En algunas formas de realización, la primera localización de inicio de nota se identifica en el bloque 602 generando (o identificando) un evento de inicio de nota, tal como se ha descrito de forma más detallada con respecto a las figuras 4 a 5.

En algunas formas de realización, el método 220 continúa identificando una segunda localización de inicio de nota en el bloque 610. Esta segunda localización de inicio de nota se puede identificar en el bloque 610 de la misma manera o de una manera diferente con respecto a la identificación de la primera localización de inicio de nota identificada en el bloque 602. En el bloque 612, se calcula la duración de una nota asociada a la primera localización de inicio de nota determinando el intervalo de tiempo entre la primera localización de inicio de nota a la segunda localización de inicio de nota. Esta determinación del bloque 612 puede dar como resultado la duración de una nota en forma del tiempo transcurrido desde el inicio de una nota al inicio de la nota siguiente.

No obstante, en algunos casos una nota puede finalizar cierto tiempo antes del comienzo de la nota siguiente. Por ejemplo, a una nota le puede seguir una pausa, o la nota puede ser tocada al estilo staccato. En estos casos, la determinación del bloque 612 daría como resultado una duración de nota que supera la duración real de la misma. Merece la pena indicar que esta limitación potencial se puede corregir de muchas maneras detectando la localización de fin de nota.

Algunas formas de realización del método 220 identifican una localización de fin de nota en el bloque 620. En el bloque 622, se puede calcular a continuación la duración de una nota asociada a la primera localización de inicio de nota determinando el intervalo de tiempo entre la primera localización de inicio de nota y la localización de fin de nota. Esta determinación del bloque 622 puede dar como resultado la duración de una nota en forma del tiempo transcurrido desde el inicio de una nota al final de esa nota. Una vez que se ha determinado la duración de la nota o bien en el bloque 612 o bien en el bloque 622, la duración de la nota se puede asignar a la nota (o evento de inicio de nota) que comienza en la primera localización temporal en el bloque 630.

Se apreciará que son posibles muchas formas de identificación de una localización de fin de nota en el bloque 620 según la invención. En una forma de realización, la localización de fin de nota se detecta en el bloque 620 determinando si hay presente alguna pausa entre las notas, y restando la duración de las pausas con respecto a la duración de la nota (la detección de pausas y duraciones de pausas se describen más adelante). En otra forma de realización, se analiza la envolvente de la nota para determinar si la nota estaba siendo tocada de tal manera que cambiaba su duración (por ejemplo, al estilo staccato).

Todavía en otra forma de realización del bloque 620, la localización de fin de nota se detecta de manera a la

detección de la localización de inicio de nota del método 420 de la figura 4B. Usando información de amplitud extraída de la señal de entrada de audio, se pueden generar una primera señal de envolvente, una segunda señal de envolvente, y una señal de control. Las posiciones de fin de nota se pueden determinar identificando posiciones en las que la amplitud de la señal de control se hace menor que la amplitud de la segunda señal de envolvente.

5 Merece la pena indicar que en música polifónica, puede haber casos en los que las notas se solapan. Como tal, puede haber condiciones en las que el fin de una primera nota viene después del comienzo de una segunda nota, aunque antes del fin de la segunda nota. Por lo tanto, la simple detección del primer fin de nota después del
10 comienzo de una nota puede no dar como resultado la localización de fin apropiada para esa nota. Como tal, puede que sea necesario extraer pistas monofónicas (según se describe posteriormente) para identificar de manera más precisa relaciones de notas.

15 La figura 7 proporciona una ilustración de una señal de audio con varias envolventes para su uso en la detección de duraciones de notas según formas de realización de la invención. La gráfica ilustrativa 700 representa la amplitud con respecto al tiempo para la señal de entrada de audio 502, y la primera señal de envolvente 504, la segunda
20 señal de envolvente 506, y la señal de control 508. La gráfica ilustra también posiciones de inicio de nota 710 en las que la amplitud de la señal de control 508 se hace mayor que la amplitud de la segunda señal de envolvente 506, y posiciones de fin de nota 720 en las que la amplitud de la señal de control 508 se vuelve menor que la amplitud de la segunda señal de envolvente 506.

25 La gráfica 700 ilustra además dos formas de realización de detección de duración de notas. En una forma de realización, se determina una primera duración de nota 730-1 hallando el tiempo transcurrido entre una primera localización de inicio de nota 710-1 y una segunda localización de inicio de nota 710-2. En otra forma de realización, se determina una segunda duración de nota 740-1 hallando el tiempo transcurrido entre una primera localización de
30 inicio de nota 710-1 y una primera localización de fin de nota 720-1.

Detección de pausas

35 La figura 8 proporciona un diagrama de flujo de un método ejemplificativo para la detección de pausas según formas de realización de la invención. El método 230 comienza identificando una condición de amplitud baja en la señal de audio de entrada en el bloque 802. Se apreciará que son posibles muchas formas de identificación de una condición de amplitud baja según la invención. En una forma de realización, un nivel de umbral de ruido se fija a cierta amplitud por encima del ruido de fondo para la señal de audio de entrada. A continuación, una condición de amplitud baja se puede identificar como una región de la señal de audio de entrada durante la cual la amplitud de la señal
40 permanece por debajo del umbral de ruido durante cierta cantidad de tiempo predeterminada.

45 En el bloque 804, se analizan regiones en las que se produce una condición de amplitud baja en relación con la confianza de la altura tonal. La confianza de la altura tonal puede identificar la probabilidad de que haya presente una altura tonal (por ejemplo, como parte de una nota pretendida) en la región. Se apreciará que la confianza de la altura tonal se puede determinar de muchas maneras, por ejemplo, según se ha descrito en referencia a la detección de alturas tonales anteriormente.

50 Cuando la confianza de altura tonal está por debajo de cierto umbral de confianza de altura tonal en una región de amplitud baja de la señal, puede ser altamente improbable que haya presente cualquier nota. En ciertas formas de realización, se determina que las regiones en las que no hay presente ninguna nota incluyen una pausa en el bloque 806. Evidentemente, tal como se ha mencionado antes, otras condiciones musicales pueden dar como resultado la aparición de una pausa (por ejemplo, una nota en staccato). Como tal, en algunas formas de realización, se puede usar otra información (por ejemplo, información de envolventes, identificación de instrumentos, etcétera) para mejorar la precisión en la determinación de si hay presente una pausa.

Detección de tempos

55 Una vez que se conocen las posiciones de notas y pausas, puede que resulte deseable determinar el tempo. El tempo adapta el concepto musical adaptativo de pulso al concepto físico convencional de tiempo, proporcionando esencialmente una medida de la velocidad de una composición musical (por ejemplo, con qué rapidez debería interpretarse la composición). Normalmente, el tempo se representa en número de pulsos por minuto, donde un pulso se representa por cierto valor de una nota. Por ejemplo, una partitura musical puede representar un único pulso como una negra, y el tempo puede ser ochenta y cuatro pulsos por minuto (bpm). En este ejemplo, la interpretación de la composición con el tempo designado significaría tocar la composición a una velocidad en la que
60 se interpreta la música a ochenta y cuatro negras cada minuto.

65 La figura 9 proporciona un diagrama de flujo de un método ejemplificativo para la detección del tempo según ejemplos de la invención. El método 240 comienza con la determinación de un conjunto de tempos de referencia en el bloque 902. En un ejemplo, se pueden usar tempos de metrónomos convencionales. Por ejemplo, un metrónomo típico se puede configurar para marcar el ritmo de tempos que van desde 40 bpm a 208 bpm, en intervalos de 4 bpm (es decir, 40 bpm, 44 bpm, 48 bpm, ..., 208 bpm). En otras formas de realización, se pueden usar otros valores e

intervalos entre valores. Por ejemplo, el conjunto de tempos de referencia puede incluir todos los tempos que van desde 10 bpm a 300 bpm en intervalos de 1/4 de bpm (es decir, 10 bpm, 10,25 bpm, 10,5 bpm, ... 300 bpm).

A continuación, el método 240 puede determinar duraciones de notas de referencia para cada tempo de referencia. Las duraciones de notas de referencia pueden representar cuánto dura un cierto valor de nota con un tempo de referencia dado. En algunas formas de realización, las duraciones de notas de referencia se pueden medir en tiempo (por ejemplo, segundos), mientras que en otras formas de realización, las duraciones de notas de referencia se pueden medir en número de muestras. Por ejemplo, suponiendo que una negra representa un único pulso, la negra con 84 bpm durará aproximadamente 0,7143 segundos (es decir, 60 segundos por minuto divididos por 84 pulsos por minuto). De manera similar, suponiendo una frecuencia de muestreo de 44.100 muestras por segundo, la negra con 84 bpm durará 31.500 muestras (es decir, 44.100 muestras por segundo multiplicado por 60 segundos por minuto dividido por 84 pulsos por minuto). En ciertas formas de realización, se puede evaluar un número de valores de nota en cada tempo de referencia para generar el conjunto de duraciones de notas de referencia. Por ejemplo, se pueden evaluar semicorcheas, corcheas, negras y blancas. De esta manera, se pueden crear valores de nota idealizados para cada tempo de referencia.

En algunos ejemplos del método 240, en el bloque 906 se puede determinar una ventana de extracción de tempos. La ventana de extracción de tempos puede ser una ventana de tiempo predeterminada o adaptativa que abarque alguna parte contigua de la señal de entrada de audio. Preferentemente, la ventana de extracción de tempos es suficientemente ancha para cubrir un número elevado de eventos de inicio de notas. Como tal, ciertas formas de realización del bloque 906 adaptan la anchura de la ventana de extracción de tempos de manera que cubra un número predeterminado de eventos de inicio de notas.

En el bloque 908, se identifica o genera el conjunto de eventos de inicio de notas que se produce durante la ventana de extracción de tempos. En ciertas formas de realización, también se identifica o genera el conjunto de posiciones de inicio de pausa que se produce durante la ventana de extracción de tempos. En el bloque 910, se extraen separaciones entre inicios de nota. Las separaciones entre inicios de nota representan la cantidad de tiempo transcurrido entre el inicio de cada nota o pausa, y el inicio de la nota o pausa sucesiva. Tal como se ha descrito anteriormente, las separaciones entre inicios de notas pueden ser iguales o diferentes a las duraciones de las notas.

El método 240 continúa en el bloque 920 determinando valores de error para cada separación entre inicios de notas, extraída, con respecto a los valores de nota idealizados determinados en el bloque 904. En una forma de realización, cada separación entre inicios de notas se divide por cada duración de nota de referencia en el bloque 922. A continuación, el resultado se puede usar para determinar la duración de nota de referencia más próxima (o múltiplo de una duración de nota de referencia) a la separación entre inicios de nota en el bloque 924.

Por ejemplo, una separación entre inicios de nota puede ser 35.650 muestras. La división de la separación entre inicios de nota por las diversas duraciones de notas de referencia y la toma de valor absoluto de la diferencia pueden generar varios resultados, representando cada resultado un valor de error. Por ejemplo, el valor de error de la separación entre inicios de nota en comparación con una negra de referencia a 72 bpm (36.750 muestras) puede ser aproximadamente 0,03, mientras que el valor de error de la separación entre inicios de nota en comparación con una corchea de referencia a 76 bpm (17.408 muestras) puede ser aproximadamente 1.05. A continuación, el valor de error mínimo se puede usar para determinar la duración de la nota de referencia más próxima (por ejemplo, en este caso ejemplificativo una negra a 72 bpm).

En algunos ejemplos, se generan uno o más valores de error a través de múltiples eventos de inicio de nota. En una forma de realización, los valores de error de todos los eventos de inicio de nota en la ventana de extracción de tempos se combinan matemáticamente antes de determinar un valor de error compuesto mínimo. Por ejemplo, los valores de error de los diversos eventos de inicio de nota se pueden sumar, promediar, o combinar matemáticamente de otra manera.

Una vez que en el bloque 920 se han determinado los valores de error, en el bloque 930 se determina el valor de error mínimo. A continuación, el tempo de referencia asociado al valor de error mínimo se puede usar como tempo extraído. En el ejemplo anterior, el valor de error más bajo fue el resultado de la duración de una nota de referencia correspondiente a una negra a 72 bpm. Como tal, 72 bpm se pueden determinar como el tempo extraído sobre una ventana dada.

Una vez que se ha determinado el tempo, puede resultar deseable asignar valores de nota para cada nota o pausa identificada en la señal de entrada de audio (o por lo menos en una ventana de la señal). La figura 10 proporciona un diagrama de flujo de un método ejemplificativo para la determinación de un valor de nota según ejemplos de la invención. El método 245 comienza en el bloque 1002 mediante la determinación de un segundo conjunto de duraciones de notas de referencia para el tempo extraído en el bloque 930 de la figura 9. En algunos ejemplos, el segundo conjunto de duraciones de notas de referencia es igual al primer conjunto de duraciones de notas de referencia. En estos ejemplos, se apreciará que el segundo conjunto simplemente se puede extraer como un subconjunto de primer conjunto de duraciones de notas de referencia. En otros ejemplos, el primer conjunto de duraciones de notas de referencia incluye solamente un subconjunto de los posibles valores de notas, mientras que

el segundo conjunto de duraciones de notas de referencia incluye un conjunto más completo de posibles duraciones de notas para el tempo extraído.

5 En el bloque 1004, el método 245 puede generar o identificar las duraciones de notas recibidas para los eventos de inicio de nota en la ventana, según se extraen de la señal de entrada de audio. Las duraciones de notas recibidas pueden representar las duraciones reales de las notas y pausas que se producen durante la ventana, por contraposición a las duraciones idealizadas representadas por el segundo conjunto de duraciones de notas de referencia. En el bloque 1006, las duraciones de notas recibidas se comparan con las duraciones de notas de referencia para determinar la duración de nota de referencia más próxima (o múltiplo de una duración de nota de referencia).

15 A continuación, la duración de nota de referencia más próxima se puede asignar a la nota o pausa como su valor de nota. En un ejemplo, se determina que una duración de nota recibida es aproximadamente 1,01 negras de referencia, y se le puede asignar un valor de nota de una negra. En otro ejemplo, se determina que una duración de nota recibida es aproximadamente 1,51 corcheas de referencia, y se le asigna un valor de nota de una corchea con puntillo (o una corchea ligada a una semicorchea).

20 La figura 12 proporciona una gráfica de datos ejemplificativos que ilustran este método ejemplificativo de detección de tempos. La gráfica 1200 representa un valor de error compuesto con respecto al tempo en pulsos por minuto. Los puntos en forma de recuadro 1202 representan valores de error por el uso de negras de referencia, y los puntos en forma de rombo 1204 representan valores de error por el uso de corcheas de referencia. Por ejemplo, el primer punto en forma de recuadro 1202-1 de la gráfica 1200 ilustra que para un conjunto de separaciones entre inicios de nota en comparación con una negra de referencia a 72 bpm, se generó un valor de error de aproximadamente 3,3.

25 La gráfica 1200 ilustra que el error mínimo para las duraciones de referencia de negras 1210-1 y el error mínimo para las duraciones de referencia de corcheas 1210-2 se generaron ambos a 84 bpm. Esto puede indicar que durante la ventana de la señal de entrada de audio, el tempo extraído es 84 bpm.

30 La figura 11 proporciona datos ejemplificativos adicionales que ilustran el método ejemplificativo de detección de tempos mostrado en la figura 12. Se muestra una parte del conjunto de separaciones entre inicios de nota 1102, medida en número de muestras que van desde 7.881 a 63.012 muestras. Las separaciones entre inicios de nota 1102 se evalúan con respecto a un conjunto de duraciones de notas de referencia 1104. Las duraciones de notas de referencia 1104, tal como se muestra, incluyen duraciones tanto en segundos como en muestras (suponiendo una frecuencia de muestreo de 44.100 muestras por segundo) de cuatro valores de nota sobre ocho tempos de referencia. Tal como se muestra en la figura 12, se determina que el tempo extraído es 84 bpm. Las duraciones de notas de referencia referentes a un tempo de referencia de 84 bpm 1106 se extraen, y se comparan con las separaciones entre inicios de nota. Se identifican las duraciones de notas de referencia más próximas 1108. A continuación, estas duraciones se pueden usar para asignar valores de nota 1110 a cada separación entre inicios de nota (o la duración de cada nota que comienza en cada separación entre inicios de nota).

40

Detección de tonalidades

45 La determinación de la tonalidad de una parte de la señal de entrada de audio puede ser importante para generar una salida de partitura útil. Por ejemplo, la determinación de la tonalidad puede proporcionar la armadura de clave correspondiente a la parte de la composición y puede identificar cuándo las notas deberían identificarse como alteraciones. No obstante, la determinación de la tonalidad puede resultar difícil por diversas razones.

50 Una razón es que las composiciones normalmente se mueven entre tonalidades (por ejemplo, por modulación). Por ejemplo, una canción de rock puede tener estrofas en la tonalidad de Sol mayor, modular a la tonalidad de Do mayor para cada estribillo, y modular adicionalmente a Re menor durante el puente. Otra razón es que las composiciones normalmente contienen varias alteraciones (notas que no se encuentran "en la tonalidad"). Por ejemplo, una canción en Do mayor (que no contiene sostenidos ni bemoles) puede usar un sostenido o bemol para añadir color o tensión a una frase de notas musicales. Todavía otra de las razones es que las composiciones normalmente tienen periodos de transición entre tonalidades, donde las frases presentan una especie de tonalidad híbrida. En estos estados híbridos, puede resultar difícil determinar cuándo cambia la tonalidad, o qué partes de la música pertenecen a cada tonalidad. Por ejemplo, durante una transición de Do mayor a Fa mayor, una canción puede usar de manera repetida un Si bemol. Esto se manifestaría como una alteración en la tonalidad de Do mayor, pero no en la tonalidad de Fa. Por lo tanto, puede resultar deseable determinar dónde se produce el cambio de tonalidad, de manera que la representación de la partitura musical 170 ni refleje incorrectamente alteraciones ni alterne repetidamente entre tonalidades. Todavía otra de las razones por la que la determinación de la tonalidad puede resultar difícil es que múltiples tonalidades pueden tener armaduras de clave idénticas. Por ejemplo, no hay sostenidos ni bemoles ni en Do mayor, ni en La menor ni en Re dórico.

60

65 La figura 13 proporciona un diagrama de flujo de un método ejemplificativo para la detección de la tonalidad según ejemplos de la invención. El método 250 comienza determinando un conjunto de funciones de coste de tonalidad en el bloque 1302. Las funciones de coste pueden buscar, por ejemplo, la minimización del número de alteraciones en

una pieza musical sobre una ventana de tiempo especificada.

Las figuras 14A y 14B proporcionan ilustraciones de funciones de coste de tonalidad ejemplificativas usadas en la detección de tonalidades según ejemplos de la invención. En la figura 14A, la función de coste de tonalidad 1400 se basa en una serie de escalas diatónicas en diversas tonalidades. Se da un valor de "1" para todas las notas de la escala diatónica para esa tonalidad, y se da un valor de "0" para todas las notas que no están en la escala diatónica para esa tonalidad. Por ejemplo, la tonalidad de Do mayor contiene la siguiente escala diatónica: Do-Re-Mi-Fa-Sol-La-Si. Así, la primera fila 1402-1 de la función de coste 1400 muestra "1" para solamente esas notas.

En la figura 14B, la función de coste de tonalidad 1450 se basa también en una serie de escalas diatónicas en varias tonalidades. A diferencia de la función de coste 1400 de la figura 14A, la función de coste 1450 de la figura 14B asigna un valor de "2" para todos los primeros, terceros y quintos tonos de la escala en una tonalidad dada. Sin embargo, se da un valor de "1" para la totalidad del resto de notas de la escala diatónica para esa tonalidad, y se da un valor de "0" para todas las notas que no están en la escala diatónica de esa tonalidad. Por ejemplo, la tonalidad de Do mayor contiene la escala diatónica, Do-Re-Mi-Fa-Sol-La-Si, en la cual el primer tono de la escala es Do, el tercer tono de la escala es Mi, y el quinto tono de la escala es Sol. Así, la primera fila 1452-1 de la función de coste 1450 muestra 2-0-1-0-2-1-0-2-0-1-0-1.

Esta función de coste 1450 puede resultar útil por varias razones. Una razón es que en muchos géneros musicales (por ejemplo, popular, rock, clásico, etcétera), el primer, el tercer y el quinto tonos de la escala tienden a tener una relevancia psicoacústica en la creación de un sentido de una cierta tonalidad en un oyente. Como tal, la ponderación de la función de coste con una mayor intensidad hacia esas notas puede mejorar la precisión de la determinación de la tonalidad en ciertos casos. Otra de las razones para usar esta función de coste 1450 puede ser la diferenciación de tonalidades con armaduras de clave similares. Por ejemplo, Do mayor, Re dórico, Sol mixolidio, La menor, y otras tonalidades no contienen, ninguna de ellas, sostenidos o bemoles. No obstante, cada una de estas tonalidades tiene un primer, un tercer y/o un quinto tonos de escala diferentes, de cada una de las otras. Así, una misma ponderación de todas las notas en la escala puede revelar poca diferencia entre la presencia de estas tonalidades (aun cuando puede haber diferencias psicoacústicas significativas), pero una ponderación ajustada puede mejorar la determinación de la tonalidad.

Se apreciará que en las funciones de coste se pueden realizar otros ajustes por diferentes motivos. En un ejemplo, la función de coste se puede ponderar de manera diferente para reflejar un género de la señal de entrada de audio (por ejemplo, recibida de un usuario, de información de encabezamiento del archivo de audio, etcétera). Por ejemplo, una función de coste de blues puede ponderar notas con mayor intensidad en concordancia con las escalas pentatónicas, en lugar de diatónicas, de una tonalidad.

Volviendo a la figura 13, en el bloque 1304 se puede determinar una ventana de extracción de tonalidades. La ventana de extracción de tonalidades puede ser una ventana de tiempo predeterminada o adaptativa que abarque cierta parte contigua de la señal de entrada de audio. Preferentemente, la ventana de extracción de tonalidades es suficientemente ancha para cubrir un número elevado de eventos de inicio de nota. Como tal, ciertos ejemplos del bloque 1304 adaptan la anchura de la ventana de extracción de tempos de manera que cubra un número predeterminado de eventos de inicio de notas.

En el bloque 1306, se identifica o genera el conjunto de eventos de inicio de nota que se produce durante la ventana de extracción de tonalidades. A continuación, en el bloque 1308 se determina la altura tonal de nota para cada evento de inicio de nota. La altura tonal de la nota se puede determinar de una manera eficaz en el bloque 1308, incluyendo los métodos de determinación de alturas tonales antes descritos. Se apreciará que, debido a que un evento de inicio de nota representa una localización temporal, técnicamente no puede haber una altura tonal en esa localización temporal (la determinación de la altura tonal requiere cierta duración temporal). Como tal, la altura tonal en un inicio de nota se refiere en general a la altura tonal asociada a la duración de la nota que viene a continuación del evento de inicio de nota.

En el bloque 1310, se puede evaluar cada altura tonal de nota con respecto a cada función de coste para generar un conjunto de valores de error. Por ejemplo, supongamos que la secuencia de alturas tonales de notas para una ventana de la señal de entrada de audio es la siguiente: Do-Do-Sol-Sol-La-La-Sol-Fa-Fa-Mi-Mi-Re-Re-Do. La evaluación de esta secuencia con respecto a la primera fila 1402-1 de la función de coste 1400 en la figura 14A puede dar como resultado un valor de error de $1+1+1+1+1+1+1+1+1+1+1+1+1+1+1+1=14$. La evaluación de la secuencia con respecto a la tercera fila 1402-2 de la función de coste 1400 en la figura 14A puede dar como resultado un valor de error de $0+0+1+1+1+1+1+1+0+0+1+1+1+1+1+0=9$. Aún más importante, la evaluación de la secuencia con respecto a la cuarta fila 1402-3 de la función de coste 1400 en la figura 14A puede dar como resultado el mismo valor de error de 14 que cuando se usó la primera fila 1402-1. Usando estos datos, parece relativamente improbable que la secuencia de alturas tonales esté en la tonalidad de Re mayor, pero parece imposible determinar si Do mayor o La menor (que comparten la misma armadura de clave) es un candidato más probable.

El uso de la función de coste 1450 en la figura 14B produce resultados diferentes. La evaluación de la secuencia con respecto a la primera fila 1452-1 puede dar como resultado un valor de error de $2+2+2+2+1+1+2+1+1+2+2+1+1+2=$

22. La evaluación de la secuencia con respecto a la tercera fila 1452-2 puede producir un valor de error de $0+0+1+1+2+2+1+0+0+2+2+1+1+0 = 13$. Aún más importante, la evaluación de la secuencia con respecto a la cuarta fila 1452-3 puede dar como resultado un valor de error de $2+2+1+1+2+2+1+1+1+2+2+1+1+2 = 21$, uno menos que el valor de error de 22 alcanzado cuando se usó la primera fila 1452-1. Usando estos datos, sigue pareciendo relativamente improbable que la secuencia de alturas tonales esté en la tonalidad de Re mayor, aunque en este momento parece ligeramente más probable que la secuencia esté en Do mayor que en La menor.

Se apreciará que las funciones de coste antes descritas (por ejemplo, 1400 y 1450) producen resultados más altos cuando es más probable que las notas recibidas estén en una tonalidad dada debido al hecho de que a las notas dentro de la tonalidad se les asignan valores diferentes de cero. No obstante, otros ejemplos pueden asignar "0" a alturas tonales que sean la "mayoría en la tonalidad" según los criterios de la función de coste. El uso de estos otros ejemplos de funciones de coste puede producir números más altos para tonalidades que presentan una menor coincidencia, generando así lo que puede ser un valor de error más intuitivo (es decir, un valor de error más alto representa una coincidencia menor).

En el bloque 1312, los diversos valores de error para las diferentes funciones de coste de tonalidades se comparan para producir la tonalidad con la mejor coincidencia con la secuencia de alturas tonales de las notas. Tal como se ha mencionado anteriormente, en algunos ejemplos, esto puede implicar el hallazgo del resultado más alto (es decir, la mejor coincidencia), mientras que en otros ejemplos, esto puede implicar el hallazgo del resultado más bajo (es decir, el error de coincidencia menor), en función de la formulación de la función de coste.

Merece la pena indicar que son posibles otros métodos de determinación de tonalidades según la invención. En algunos ejemplos, se puede usar una red neuronal artificial para realizar o mejorar la precisión de determinaciones de tonalidades complejas. En otros ejemplos, se puede evaluar una secuencia de cambios de tonalidad con respecto a funciones de coste para mejorar la precisión de determinaciones de tonalidades. Por ejemplo, el método puede detectar una serie de tonalidades en la señal de audio de entrada del patrón Do mayor – Fa mayor – Sol mayor – Do mayor. No obstante, la confianza en la detección de Fa mayor puede estar limitada, debido a la detección de una serie de Si naturales (el 4 sostenido de Fa – una nota improbable en la mayoría de géneros musicales). Dado que la tonalidad identificada como Fa mayor precede a una sección en Sol mayor de una canción que comienza y acaba Do mayor, la presencia de Si naturales, incluso ocasionales, puede indicar que la determinación de la tonalidad se debería revisar para obtener una elección más ajustada (por ejemplo, Re dórico o incluso Re menor).

Una vez que se ha determinado la tonalidad, puede resultar deseable emparejar designaciones de alturas de la tonalidad con notas en cada evento de inicio de nota (por lo menos para aquellos eventos de inicio que se producen dentro de la ventana de extracción de tonalidades). La figura 15 proporciona un diagrama de flujo de un método ejemplificativo para la determinación de la designación de alturas de la tonalidad según ejemplos de realización. El método 255 comienza generando un conjunto de alturas tonales de referencia para la tonalidad extraída, en el bloque 1502.

Merece la pena señalar que las posibles alturas tonales pueden ser las mismas para todas las tonalidades (por ejemplo, considerando especialmente los patrones de afinación actuales). Por ejemplo, la totalidad de las doce notas cromáticas de cada octava de un piano se puede tocar en cualquier tonalidad. La diferencia puede ser cómo se representan dichas alturas tonales en una partitura (por ejemplo, tonalidades diferentes pueden asignar alteraciones diferentes a la misma altura tonal de la nota). Por ejemplo, las alturas tonales de una tonalidad para las "teclas blancas" de un piano en Do mayor se pueden designar como Do, Re, Mi, Fa, Sol, La, y Si. El mismo conjunto de alturas de la tonalidad en Re mayor se puede designar como Do natural, Re, Mi, Fa natural, Sol, La, y Si.

En el bloque 1504, se determina la altura tonal de referencia más próxima a cada altura tonal de una nota extraída y la misma se usa para generar la determinación de altura de la tonalidad para esa nota. A continuación, se puede asignar la determinación de la altura de la tonalidad a la nota (o evento de inicio de nota) en el bloque 1506.

Sistema de hardware ejemplificativo

Los sistemas y métodos descritos anteriormente se pueden implementar de varias maneras. Una de estas implementaciones incluye varios componentes electrónicos. Por ejemplo, unidades del sistema de la figura 1B se pueden implementar, de manera individual o en conjunto, con uno o más Circuitos Integrados de Aplicación Específica (ASICs) adaptados para llevar a cabo parte o la totalidad de las funciones aplicables en hardware. Alternativamente, las funciones se pueden realizar por parte de otra u otras unidades de procesado (o núcleos), en uno o más circuitos integrados. En otras formas de realización, se pueden usar otros tipos de circuitos integrados (por ejemplo, ASICs Estructurados/de Plataforma, Matrices de Puertas Programables in Situ (FPGAs), y otros ICs Semi-Personalizados), los cuales se pueden programar según cualquier manera conocida en la técnica. Las funciones de cada unidad también se pueden implementar, en su totalidad o parcialmente, con instrucciones incorporadas en una memoria, con un formato para ser ejecutadas por uno o más procesadores generales o de aplicación específica.

La figura 16 proporciona un diagrama de bloques de un sistema computacional 1600 para implementar ciertas formas de realización de la invención. En una forma de realización, el sistema computacional 1600 puede funcionar como el sistema 100 mostrado en la figura 1A. Debería indicarse que la figura 16 está destinada únicamente a proporcionar una ilustración generalizada de varios componentes, pudiéndose utilizar cualquiera o la totalidad de ellos según se crea apropiado. Por lo tanto, la figura 16 ilustra ampliamente cómo se pueden implementar elementos individuales del sistema de una manera relativamente separada o relativamente más integrada.

El sistema computacional 1600 se muestra de manera que comprende elementos de hardware que se pueden acoplar eléctricamente por medio de un bus 1626 (o se pueden situar en comunicación de otra manera, según se crea apropiado). Los elementos de hardware pueden incluir uno o más procesadores 1602, que incluyen, sin carácter limitativo, uno o más procesadores de función general y/o uno o más procesadores de función especial (tales como chips de procesamiento de señal digital, chips de aceleración para gráficos, y/o similares); uno o más dispositivos de entrada 1604, los cuales pueden incluir, sin carácter limitativo, un ratón, un teclado, y/o similares; y uno o más dispositivos de salida 1606, los cuales pueden incluir sin carácter limitativo un dispositivo de visualización, una impresora, y/o similares.

El sistema computacional 1600 puede incluir además (y/o estar en comunicación con) uno o más dispositivos de almacenamiento 1608, los cuales pueden comprender, sin carácter limitativo, medios de almacenamiento accesibles locales y/o en red y/o pueden incluir, sin carácter limitativo, una unidad de disco, un grupo de unidades de disco, un dispositivo de almacenamiento óptico, un dispositivo de almacenamiento de estado sólido, tal como una memoria de acceso aleatorio ("RAM"), y/o una memoria de solo lectura ("ROM"), que puede ser programable, actualizable de tipo *flash*, y/o similares. El sistema computacional 1600 también podría incluir un subsistema de comunicaciones 1614, el cual puede incluir, sin carácter limitativo, un módem, una tarjeta de red (inalámbrica o por cable), un dispositivo de comunicaciones por infra-rojos, un dispositivo de comunicaciones inalámbricas, y/o un conjunto de chips (*chip set*) (tal como un dispositivo Bluetooth, un dispositivo 802.11, un dispositivo WiFi, un dispositivo WiMax, utilidades de comunicación celular, etcétera), y/o similares. El subsistema de comunicaciones 1614 puede permitir el intercambio de datos con una red (tal como la red que se describe posteriormente, por nombrar un ejemplo), y/o cualesquiera otros dispositivos descritos en la presente. En muchas formas de realización, el sistema computacional 1600 comprenderá además una memoria de trabajo 1618, la cual puede incluir un dispositivo RAM o ROM, según se ha descrito anteriormente.

El sistema computacional 1600 también puede comprender unos elementos de software, mostrados de manera que actualmente están ubicados dentro de la memoria de trabajo 1618, incluyendo un sistema operativo 1624 y/u otro código, tal como uno o más programas de aplicación 1622, los cuales pueden comprender programas de ordenador de la invención, y/o pueden diseñarse para implementar métodos de la invención y/o configurar sistemas de la invención, según se ha descrito en la presente. Meramente a título de ejemplo, uno o más procedimientos descritos con respecto al(a los) método(s) expuesto(s) anteriormente se podrían implementar en forma de código y/o instrucciones ejecutables por un ordenador (y/o un procesador dentro de un ordenador). Un conjunto de estas instrucciones y/o código se podría almacenar en un soporte de almacenamiento legible por ordenador 1610b. En algunas formas de realización, el soporte de almacenamiento legible por ordenador 1610b es el(los) dispositivo(s) de almacenamiento 1608 descrito(s) anteriormente. En otras formas de realización, el soporte de almacenamiento legible por ordenador 1610b se podría incorporar dentro de un sistema de ordenador. Todavía en otras formas de realización, el soporte de almacenamiento legible por ordenador 1610b podría ser independiente del sistema de ordenador (es decir, un soporte extraíble, tal como un disco compacto, etcétera), y/o se podría proporcionar en un paquete de instalación, de tal manera que el soporte de almacenamiento se puede usar para programar un ordenador de función general con las instrucciones/código almacenados en el mismo. Estas instrucciones podían adoptar la forma de código ejecutable, el cual puede ser ejecutado por el sistema de ordenador 1600 y/o podrían adoptar la forma de código fuente y/o instalable, el cual, al producirse la compilación y/o instalación en el sistema de ordenador 1600 (por ejemplo, usando cualquiera de una variedad de compiladores, programas de instalación, utilidades de compresión/descompresión, etcétera, disponibles de forma general), adopta entonces la forma de código ejecutable. En estas formas de realización, el soporte de almacenamiento legible por ordenador 1610b puede ser leído por un lector de soportes de almacenamiento legibles por ordenador 1610a.

Se pondrá de manifiesto para aquellos versados en la materia que se pueden aplicar variaciones sustanciales según requisitos específicos. Por ejemplo, también se podría usar hardware personalizado, y/o elementos particulares se podrían implementar en hardware, software (incluyendo software portable, tales como miniaplicaciones, etcétera), o en las dos opciones. Además, se puede utilizar una conexión con otros dispositivos informáticos, tales como dispositivos de entrada/salida de red.

En algunas formas de realización, uno o más de los dispositivos de entrada 1604 se pueden acoplar a una interfaz de audio 1630. La interfaz de audio 1630 se puede configurar para comunicarse con un micrófono, un instrumento, un dispositivo de audio digital, u otra fuente de archivos o señales de audio, por ejemplo de manera física, óptica, electromagnética, etcétera. Además, en algunas formas de realización, uno o más de los dispositivos de salida 1606 se puede acoplar a una interfaz de transcripción de fuentes 1632. La interfaz de transcripción de fuentes 1632 se puede configurar para dar salida a datos de representación de partituras musicales generados por formas de realización de la invención, hacia uno o más sistemas con capacidad de gestionar estos datos. Por ejemplo, la

interfaz de transcripción de fuentes se puede configurar para comunicarse con software de transcripción de partituras, sistemas de edición de partituras, altavoces, etcétera.

En una forma de realización, la invención utiliza un sistema de ordenador (tal como el sistema computacional 1600) para llevar a cabo métodos de la invención. De acuerdo con un conjunto de formas de realización, algunos o la totalidad de los procedimientos de dichos métodos son realizados por el sistema computacional 1600 como respuesta a la ejecución, por parte del procesador 1602, de una o más secuencias de una o más instrucciones (que podrían estar incorporadas en el sistema operativo 1624 y/u otro código, tal como un programa de aplicación 1622) contenidas en la memoria de trabajo 1618. Dichas instrucciones se pueden leer en la memoria de trabajo 1618 desde otro soporte legible por máquina, tal como uno o más del(de los) dispositivo(s) de almacenamiento 1608 (ó 1610). Meramente a título de ejemplo, la ejecución de las secuencias de instrucciones contenidas en la memoria de trabajo 1618 podría provocar que el(los) procesador(es) 1602 realizase(n) uno o más procedimientos de los métodos descritos en la presente.

Las expresiones “soporte legible por máquina” y “soporte legible por ordenador”, tal como se usan en la presente, se refieren a cualquier soporte que participe en la provisión de datos que provoquen que una máquina funcione de una manera específica. En una forma de realización implementada usando el sistema computacional 1600, varios soportes legibles por máquina se podrían ver implicados en la provisión de instrucciones/código al(a los) procesador(es) 1602 para la ejecución y/o se podrían usar para almacenar y/o transportar dichas instrucciones/código (por ejemplo, en forma de señales). En muchas implementaciones, un soporte legible por máquina es un soporte de almacenamiento físico y/o tangible. Un soporte de este tipo puede adoptar muchas formas, incluyendo, aunque sin carácter limitativo, soportes no volátiles, soportes volátiles, y unos soportes de transmisión. Los soportes no volátiles incluyen, por ejemplo, discos ópticos o magnéticos, tales como el(los) dispositivo(s) de almacenamiento (1608 ó 1610). Los soportes volátiles incluyen, sin carácter limitativo, memoria dinámica, tal como la memoria de trabajo 1618. Los soportes de transmisión incluyen cables coaxiales, hilo de cobre, y fibra óptica, incluyendo los hilos metálicos que comprenden el bus 1626, así como los diversos componentes del subsistema de comunicaciones 1614 (y/o los soportes a través de los cuales el subsistema de comunicaciones 1614 proporciona comunicación con otros dispositivos). Por tanto, los soportes de transmisión también pueden adoptar la forma de ondas (incluyendo, sin carácter limitativo, ondas de radiocomunicaciones, acústicas, y/o de luz, tales como las generadas durante las comunicaciones de datos por ondas de radiocomunicaciones y por infra-rojos).

Las formas comunes de soportes legibles por ordenador, físicos y/o tangibles, incluyen, por ejemplo, un disco *floppy*, un disco flexible, un disco duro, cinta magnética, o cualquier otro soporte magnético, un CD-ROM, cualquier otro soporte óptico, tarjetas perforadas, cinta perforada, cualquier otro soporte físico con patrones de agujeros, una RAM, una PROM, una EPROM, una FLASH-EPROM, cualquier otro chip o cartucho de memoria, una onda portadora según se describe posteriormente en la presente, o cualquier otro soporte a partir del cual un ordenador pueda leer instrucciones y/o código.

En el transporte de una o más secuencias de una o más instrucciones al(a los) procesador(es) 1602 para su ejecución pueden verse implicadas varias formas de soportes legibles por máquina. Meramente a título de ejemplo, las instrucciones se pueden transportar inicialmente en un disco magnético y/o disco óptico de un ordenador remoto. Un ordenador remoto podría cargar las instrucciones en su memoria dinámica y enviar las instrucciones en forma de señales, a través de un soporte de transmisión, para ser recibidas y/o ejecutadas por el sistema computacional 1600. Estas señales, que podrían estar en forma de señales electromagnéticas, señales acústicas, señales ópticas y/o similares, son todas ellas ejemplos de ondas portadoras en las cuales se pueden codificar instrucciones, de acuerdo con varias formas de realización de la invención.

El subsistema de comunicaciones 1614 (y/o sus componentes recibirán generalmente las señales, y a continuación el bus 1626 podría transportar las señales (y/o los datos, instrucciones, etcétera, transportados por las señales) a la memoria de trabajo 1618, desde la cual el(los) procesador(es) 1602 recupera y ejecuta las instrucciones. Las instrucciones recibidas por la memoria de trabajo 1618 se pueden almacenar opcionalmente en un dispositivo de almacenamiento 1608 o bien antes o bien después de su ejecución por parte del(de los) procesador(es) 1602.

Otras capacidades

Se apreciará que son posibles muchas otras capacidades de procesado además de las descritas anteriormente. Un grupo de capacidades de procesado adicionales implica el aumento del nivel de capacidad de personalización que se proporciona a un usuario. Por ejemplo, formas de realización pueden permitir una capacidad de personalización mejorada de varios componentes y métodos de la invención.

En algunas formas de realización, los diversos umbrales, ventanas, y otras entradas en los componentes y métodos pueden ser ajustables, cada uno de ellos, por varios motivos. Por ejemplo, puede que el usuario tenga la capacidad de ajustar la ventana de extracción de tonalidades, en caso de que se ponga de manifiesto que las determinaciones

de las tonalidades se están realizando con demasiada frecuencia (por ejemplo, el usuario puede no desear que los desvíos breves con respecto a la tonalidad aparezcan como cambio de tonalidad en la partitura). Como ejemplo adicional, una grabación puede incluir un ruido de fondo que provenga de la alimentación de 60 Hz usada durante la interpretación en la grabación. Puede que el usuario desee ajustar varios algoritmos de filtrado para ignorar esta altura tonal de 60 Hz, de manera que no se represente como una nota baja en la partitura. Todavía en otro ejemplo, el usuario puede ajustar la resolución de compartimentos musicales en los que se cuantifican alturas tonales con el fin de ajustar la resolución de las alturas tonales de notas.

En otras formas de realización, se puede proporcionar una menor capacidad de personalización al usuario. En una forma de realización, puede que el usuario tenga la capacidad de ajustar un nivel de precisión de representación. El usuario puede introducir (por ejemplo, por medio de una corredera física o virtual, un botón, un interruptor, etcétera) si el sistema debería generar representaciones de partituras más precisas o menos precisas, basándose en uno o más parámetros, incluyendo la selección de la precisión para elementos individuales de representación de la partitura, como el tempo y la altura tonal.

Por ejemplo, una serie de ajustes internos pueden actuar conjuntamente de manera que el valor de nota mínimo sea una semicorchea. Ajustando la precisión de representación, se pueden detectar duraciones mayores o menores y las mismas se pueden representar como el valor mínimo. Esto puede resultar útil cuando un intérprete no está interpretando estrictamente con un ritmo constante (por ejemplo, no hay sección de percusión, ningún metrónomo, etcétera), y un sistema demasiado sensible puede producir representaciones no deseables (por ejemplo, notas con triple puntillo). Como ejemplo adicional, una serie de ajustes internos pueden actuar conjuntamente de manera que el cambio de altura tonal mínimo sea un semitono (es decir, notas de la escala cromática).

Todavía en otras formas de realización, al usuario se le puede proporcionar una capacidad de personalización todavía menor. En una forma de realización, el usuario puede introducir si se trata de un usuario novato o un usuario avanzado. En otra forma de realización, el usuario puede introducir si el sistema debería tener una sensibilidad alta o baja. En cualquiera de las formas de realización, muchos parámetros diferentes en muchos componentes o métodos se pueden ajustar conjuntamente para encajar con el nivel deseado. Por ejemplo, en uno de los casos, puede que un cantante desee transcribir de manera precisa toda vacilación de la altura tonal y la duración (por ejemplo, como ayuda práctica para encontrar errores, o para reproducir fielmente una interpretación específica con todas sus sutilezas estéticas); mientras que en otro caso, el cantante puede que desee generar una partitura para edición fácil de leer haciendo que el sistema ignore desviaciones pequeñas.

Otro conjunto de capacidades de procesamiento adicionales conlleva el uso de diferentes tipos de entrada para mejorar la precisión o influir de otra manera en el procesamiento de la señal de audio de entrada. Una de las formas de realización usa una o más redes neuronales artificiales (ANN's) entrenadas, para mejorar la precisión de ciertas determinaciones. Por ejemplo, las determinaciones psicoacústicas (por ejemplo, métrica, tonalidad, instrumentación, etcétera) pueden resultar muy adecuadas para el uso de ANN entrenadas.

Otro ejemplo proporciona al usuario la capacidad de formar capas de múltiples pistas (por ejemplo, un hombre orquesta). El usuario puede empezar con la interpretación de una pista de tambor, la cual se procesa en tiempo real usando el sistema de la invención. A continuación, el usuario puede interpretar sucesivamente una pista de guitarra, una pista de teclado, y una pista vocal, cada una de las cuales es procesada. En algunos casos, el usuario puede seleccionar múltiples pistas para su proceso conjunto, mientras que en otros casos, el usuario puede optar por hacer que cada pista se procese por separado. A continuación, la información de algunas pistas se puede usar para mejorar la precisión o dirigir el procesamiento y otras pistas. Por ejemplo, la pista del tambor se puede procesar independientemente para generar información de métrica y tempo de alta confianza. A continuación, la información de tempo y métrica se puede usar con las otras pistas para determinar de forma más precisa duraciones de notas y valores de las mismas. Como ejemplo alternativo, la pista de guitarra puede proporcionar muchas alturas tonales sobre ventanas de tiempo pequeñas, lo cual puede hacer que resulte más sencillo determinar la tonalidad. A continuación, la determinación de la tonalidad se puede usar para asignar determinaciones de alturas de la tonalidad a las notas en la pista de teclado. Todavía como ejemplo alternativo, las múltiples pistas se pueden alinear, cuantificar, o normalizar en una o más dimensiones (por ejemplo, las pistas se pueden normalizar para tener un tempo, un volumen medio, un intervalo de alturas tonales, una resolución de las alturas tonales, una duración de nota mínima, etcétera, iguales). Además, en algunas formas de realización del "hombre orquesta", el usuario puede usar un instrumento para generar la señal de audio, a continuación puede usar el sistema o métodos para convertirla en un instrumento o instrumentos diferentes (por ejemplo, reproducir la totalidad de las cuatro pistas de un cuarteto usando un teclado, y usar el sistema para convertir la entrada de teclado en un cuarteto de cuerda). En algunos casos, esto puede conllevar el ajuste de timbre, la transposición de las líneas musicales, y otro procesamiento.

Todavía otro ejemplo usa entradas extrínsecas a la señal de entrada de audio para mejorar la precisión o dirigir el procesamiento. En una forma de realización de ejemplo, se recibe información de género desde uno de los siguientes: un usuario, otro sistema (por ejemplo, un sistema de ordenador o Internet), o información de encabezamiento del archivo de audio digital, para mejorar la precisión de diversas funciones de coste. Por ejemplo, las funciones de coste de la tonalidad pueden ser diferentes para el blues, la música clásica india, la música popular, etcétera; o una instrumentación diferente puede ser más probable en diferentes géneros (por ejemplo, es muy probable que un

sonido de “tipo órgano” sea un órgano en música de himnos y un acordeón en música de polca).

5 Un tercer conjunto de capacidades de procesamiento adicionales conlleva el uso de información a través de múltiples componentes o métodos para mejorar la precisión de determinaciones complejas. En un ejemplo, la salida del método de identificación de instrumentos se usa para mejorar la precisión de determinaciones basándose en capacidades o limitaciones conocidas de los instrumentos identificados. Por ejemplo, supóngase que el método de identificación de instrumentos determina que es probable que una línea musical esté siendo tocada por un piano. No obstante, el método de identificación de alturas tonales determina que la línea musical contiene un vibrato ligero y rápido (por ejemplo, trino de la altura tonal dentro de solamente uno o dos semitonos de la designación de la altura detectada de la tonalidad). Debido a que típicamente este no es un efecto posible que se reproduzca en un piano, el sistema puede determinar que la línea está siendo tocada con otro instrumento (por ejemplo, un teclado electrónico o un órgano).

15 Se apreciará que son posibles muchas capacidades de procesamiento adicionales del tipo mencionado. Además, debe indicarse que los métodos, sistemas, y dispositivos expuestos anteriormente están destinados solo a ser ejemplos. Debe ponerse de relieve que varias formas de realización pueden omitir, sustituir, o añadir varios procedimientos o componentes según se crea apropiado. Por ejemplo, debe apreciarse que, en ejemplos alternativos, los métodos se pueden ejecutar en un orden diferente al descrito, y que se pueden añadir, omitir, o combinar varias etapas. Además, las características descritas con respecto a ciertas formas de realización se pueden combinar en otras formas de realización diversas. Diferentes aspectos y elementos de las formas de realización se pueden combinar de una manera similar. Además, debe recalarse que la tecnología evoluciona y que, por lo tanto, muchos de los elementos son ejemplos y no deben interpretarse como limitativos del alcance de la invención.

25 En la descripción se ofrecen detalles específicos para proporcionar una comprensión minuciosa de las formas de realización. No obstante, aquellos con conocimientos habituales en la materia entenderán que las formas de realización se pueden poner en práctica sin estos detalles específicos. Por ejemplo, se han mostrados circuitos, procesos, algoritmos, estructuras y técnicas bien conocidos, sin detalles innecesarios con el fin de evitar el enrevesamiento de las formas de realización. Además, los encabezamientos proporcionados en la presente pretenden únicamente ayudar a clarificar las descripciones de varias formas de realización, y no deben interpretarse como limitativos del alcance de la invención o la funcionalidad de alguna parte de la invención. Por ejemplo, ciertos métodos o componentes se pueden implementar como parte de otros métodos o componentes, aun cuando los mismos se hayan descrito bajo diferentes encabezamientos.

35 Además, debe indicarse que las formas de realización se pueden describir como un proceso que se representa en forma de un diagrama de flujo o diagrama de bloques. Aunque cada uno de ellos puede describir las operaciones como un proceso secuencial, muchas de las operaciones se pueden ejecutar en paralelo o simultáneamente. Adicionalmente, el orden de las operaciones se puede reorganizar. Un proceso puede tener etapas adicionales no incluidas en la figura.

REIVINDICACIONES

1. Sistema para generar datos de partituras a partir de una señal de audio, comprendiendo el sistema:

- 5 un receptor de audio que se puede hacer funcionar para procesar la señal de audio; y
- una unidad de identificación de notas que se puede hacer funcionar para recibir la señal de audio procesada y generar un evento de inicio de nota asociado a una localización temporal en la señal de audio procesada como respuesta a por lo menos una de entre:
- 10 identificar un cambio de frecuencia que supera un primer valor de umbral; e
- identificar un cambio de amplitud que supera un segundo valor de umbral;
- 15 comprendiendo la unidad de identificación de notas:
- un procesador de señales, que comprende:
- 20 una unidad detectora de frecuencias que se puede hacer funcionar para identificar el cambio de frecuencia de la señal de audio, que supera el primer valor de umbral,
- una unidad detectora de amplitudes que se puede hacer funcionar para identificar un cambio de amplitud de la señal de audio, que supera el segundo valor de umbral; y
- 25 un procesador de notas que incluye un generador de eventos de inicio de nota que está en comunicación operativa con la unidad detectora de frecuencias y la unidad detectora de amplitudes y que se puede hacer funcionar para generar el evento de inicio de nota; comprendiendo el procesador de notas asimismo:
- 30 un primer generador de envolventes que se puede hacer funcionar para generar una primera señal de envolvente de acuerdo con la amplitud de la señal de audio procesada;
- un segundo generador de envolventes que se puede hacer funcionar para generar una segunda señal de envolvente de acuerdo con un valor de potencia medio de la primera señal de envolvente; y
- 35 un generador de señales de control que se puede hacer funcionar para generar una señal de control como respuesta a un cambio en la primera señal de envolvente desde una primera dirección hasta una segunda dirección de tal manera que el cambio se prolongue durante un espacio de tiempo mayor que un tiempo de control predeterminado;
- 40 identificando la unidad detectora de amplitudes el cambio de amplitud de la señal de audio que supera el segundo valor de umbral como respuesta a que la amplitud de la señal de control presente un valor mayor que una amplitud de la segunda señal de envolvente.

45 2. Sistema según la reivindicación 1, en el que la generación de un evento de inicio de nota incluye indicar un valor de marca de tiempo de la señal de entrada de audio correspondiente al evento de inicio de nota, y/o

comprendiendo la primera función de envolvente una función que efectúa una aproximación de la amplitud de la señal de entrada de audio en cada valor de marca de tiempo y la segunda función de envolvente comprende una función que efectúa una aproximación de la potencia media de la primera función de envolvente durante un intervalo de promediado, y/o

incluyendo la generación de un evento de inicio de nota asimismo ajustar el intervalo de promediado de la segunda función de envolvente como respuesta a un valor de ajuste recibido.

55 3. Sistema según la reivindicación 2, en el que el valor de ajuste recibido se determina de acuerdo con:

- (a) una selección de clase de instrumento recibida desde una entrada de usuario, o
- 60 (b) una selección de género musical recibida desde una entrada de usuario.

4. Sistema según la reivindicación 1, que comprende asimismo:

65 una unidad detectora de duraciones de nota, en comunicación operativa con el generador de eventos de inicio de nota, y que se puede hacer funcionar para detectar una duración de nota determinando por lo menos el intervalo de tiempo entre un primer evento de inicio de nota y un segundo evento de inicio de nota, habiendo sido

generados el primer evento de inicio de nota y el segundo inicio de nota por el generador de eventos de inicio de nota de manera que el segundo evento de inicio de nota sea posterior en el tiempo al primer evento de inicio de nota; y

5 la asociación de la duración de nota al primer evento de inicio de nota, representando la duración de nota el intervalo de tiempo determinado, y/o

siendo el segundo inicio de nota el evento de inicio de nota más próximo posterior en el tiempo al primer evento de inicio de nota.

10 5. Sistema según la reivindicación 1, que comprende asimismo:

15 una unidad detectora de notas y eventos, que se puede hacer funcionar para generar una nota y un evento asociados a una localización temporal en la señal de audio cuando la amplitud de la señal de control se hace menor que la amplitud de la segunda señal de envolvente; y

una unidad detectora de duraciones de nota, en comunicación operativa con el generador de eventos de inicio de nota y la unidad detectora de eventos de fin de nota, y que se puede hacer funcionar para:

20 detectar una duración de nota por lo menos determinando el intervalo de tiempo entre un evento de inicio de nota y un evento de fin de nota de manera que el evento de fin de nota sea posterior en el tiempo al evento de inicio de nota; y

25 asociar la duración de nota al evento de inicio de nota, representando la duración de nota el intervalo de tiempo determinado.

6. Sistema según la reivindicación 1, que comprende asimismo:

30 una unidad detectora de pausas, que se puede hacer funcionar para detectar una pausa identificando una parte de la señal de audio que tiene una amplitud inferior a un umbral de detección de pausas.

35 7. Sistema según la reivindicación 6, en el que el detector de pausas se puede hacer funcionar asimismo para detectar una pausa determinando un valor de confianza de altura tonal menor que un umbral de confianza de altura tonal, representando el valor de confianza de altura tonal la probabilidad de que la parte de la señal de audio comprenda una altura tonal en relación con un evento de inicio de nota.

40 8. Sistema según la reivindicación 1, en el que la señal de audio se recibe desde una o más fuentes de audio, seleccionándose cada fuente de audio de entre el grupo constituido por un micrófono, un componente de audio digital, un archivo de audio, una tarjeta de sonido, y un reproductor de medios.

9. Método de generación de datos de partitura a partir de una señal de audio, comprendiendo el método:

45 identificar un cambio de información de frecuencia a partir de la señal de audio, que supera un primer valor de umbral;

identificar un cambio de información de amplitud a partir de la señal de audio, que supera un segundo valor de umbral; y

50 generar un evento de inicio de nota, representando cada evento de inicio de nota una localización temporal en la señal de audio, de por lo menos uno de entre un cambio identificado de la información de frecuencia, que supera el primer valor de umbral, o un cambio identificado de la información de amplitud, que supera el segundo valor de umbral; comprendiendo asimismo:

55 asociar un registro de nota al evento de inicio de nota, comprendiendo el registro de nota un conjunto de datos característicos de la nota, y

60 generar una primera señal de envolvente, realizando la primera señal de envolvente sustancialmente un seguimiento de un valor absoluto de la información de amplitud de la señal de audio; generar una segunda señal de envolvente, realizando la segunda señal de envolvente sustancialmente un seguimiento de una potencia media de la primera señal de envolvente; y generar una señal de control, realizando la señal de control sustancialmente un seguimiento de cambios direccionales de la primera señal de envolvente que duran más que un tiempo de control predeterminado, comprendiendo la identificación de un cambio de la información de amplitud identificar una primera localización de comienzo de nota que representa una localización temporal en la señal de audio, en la que una amplitud de la señal de control se hace mayor que una amplitud de la segunda señal de envolvente.

65

10. Método según la reivindicación 9, en el que la generación de un evento de inicio de nota incluye indicar un valor de marca de tiempo de la señal de entrada de audio correspondiente al evento de inicio de nota.
- 5 11. Método según la reivindicación 10, en el que la primera función de envolvente comprende una función que efectúa una aproximación de la amplitud de la señal de entrada de audio en cada valor de marca de tiempo, y la segunda función de envolvente comprende una función que efectúa una aproximación de la potencia media de la primera función de envolvente durante un intervalo de promediado.
- 10 12. Método según la reivindicación 11, en el que la generación de un evento de inicio de nota incluye asimismo ajustar el intervalo de promediado de la segunda función de envolvente como respuesta a un valor de ajuste recibido.
13. Método según la reivindicación 12, en el que el valor de ajuste recibido se determina de acuerdo:
- 15 (a) con un tipo de instrumento recibido a partir de una entrada de usuario, o
- (b) una selección de género musical recibida a partir de una entrada de usuario.
- 20 14. Método según la reivindicación 11, que comprende asimismo:
- 25 identificar una segunda localización de comienzo de nota que representa una localización temporal en la señal de audio, en la que la amplitud de la señal de control se hace mayor que la amplitud de la segunda señal de envolvente por vez primera tras la primera localización temporal; y asociar una duración al evento de inicio de nota, representando la duración el intervalo de tiempo desde la primera localización de comienzo de nota a la segunda localización de comienzo de nota, y/o
- 30 identificar una localización de fin de nota que representa una localización temporal en la señal de audio, en la que la amplitud de la señal de control se hace menor que la amplitud de la segunda señal de envolvente por primera vez tras la primera localización de comienzo de nota; y asociar una duración al evento de inicio de nota, representando la duración el intervalo de tiempo desde la primera localización de comienzo de nota hasta la localización de fin de nota.
- 35 15. Soporte de almacenamiento legible por ordenador, que tiene incorporado en el mismo un programa legible por ordenador para dirigir el funcionamiento de un sistema de generación de datos de partitura que incluye un receptor de audio configurado para recibir una señal de audio, un procesador de señales configurado para procesar la señal de audio según el método de la reivindicación 9.

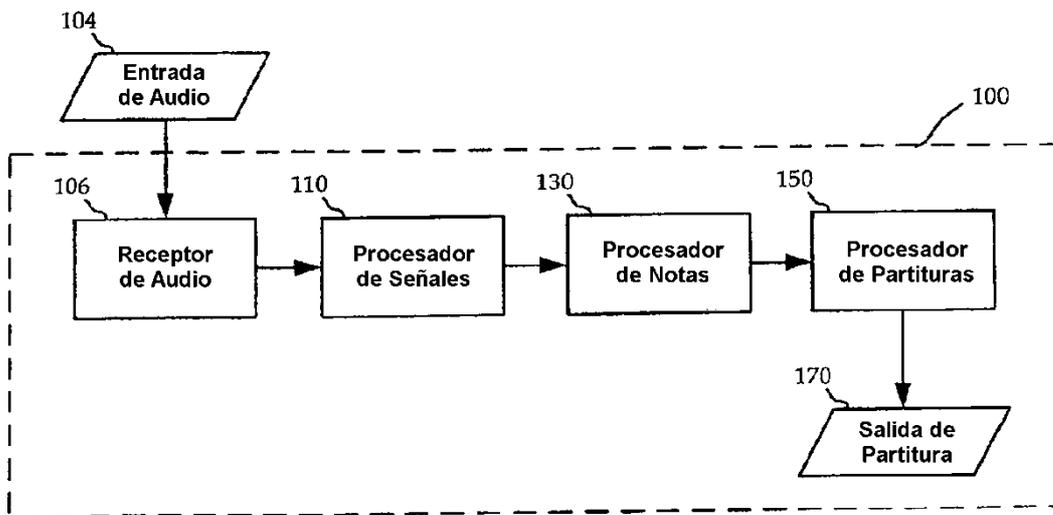


FIG. 1A

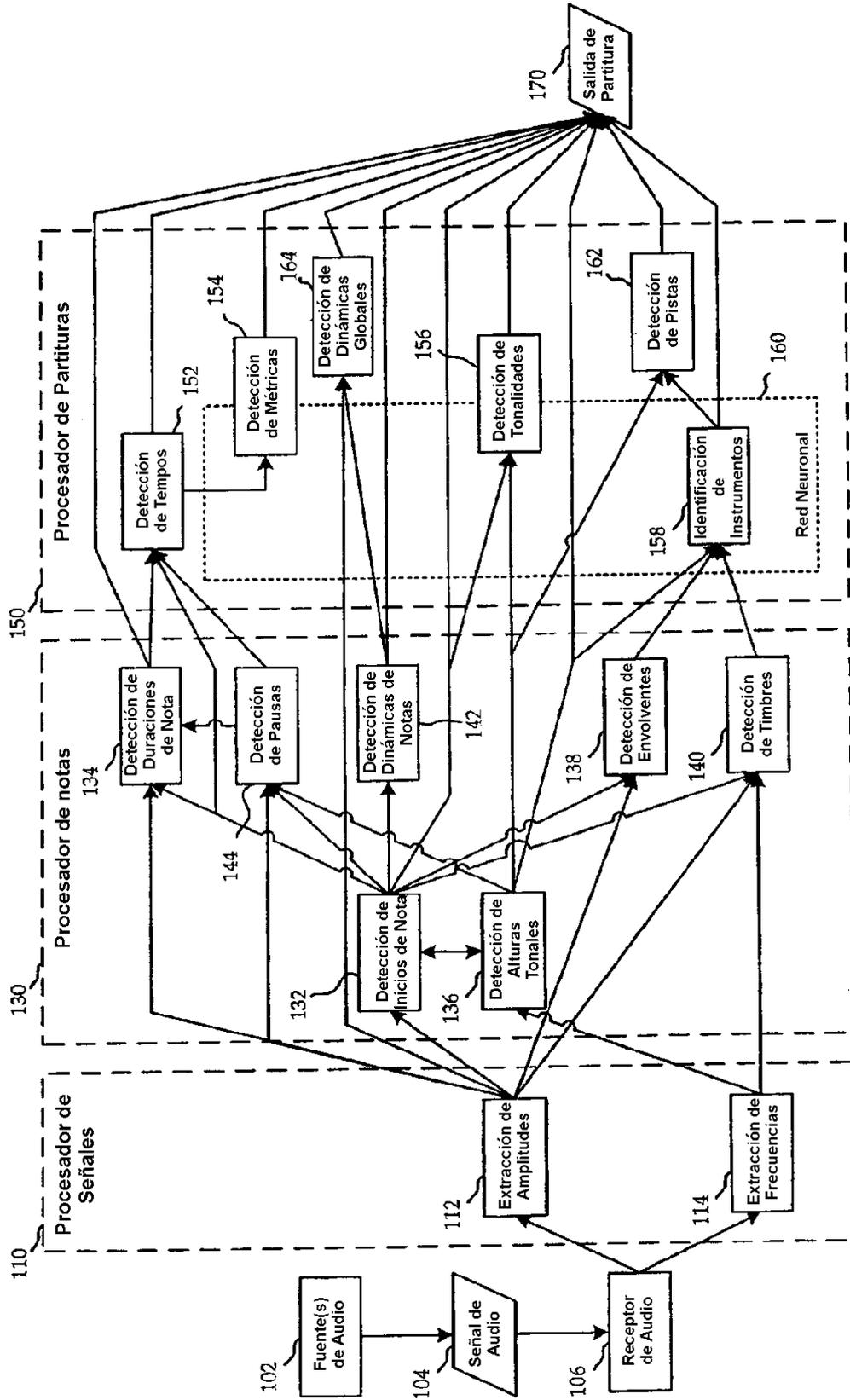


FIG. 1B

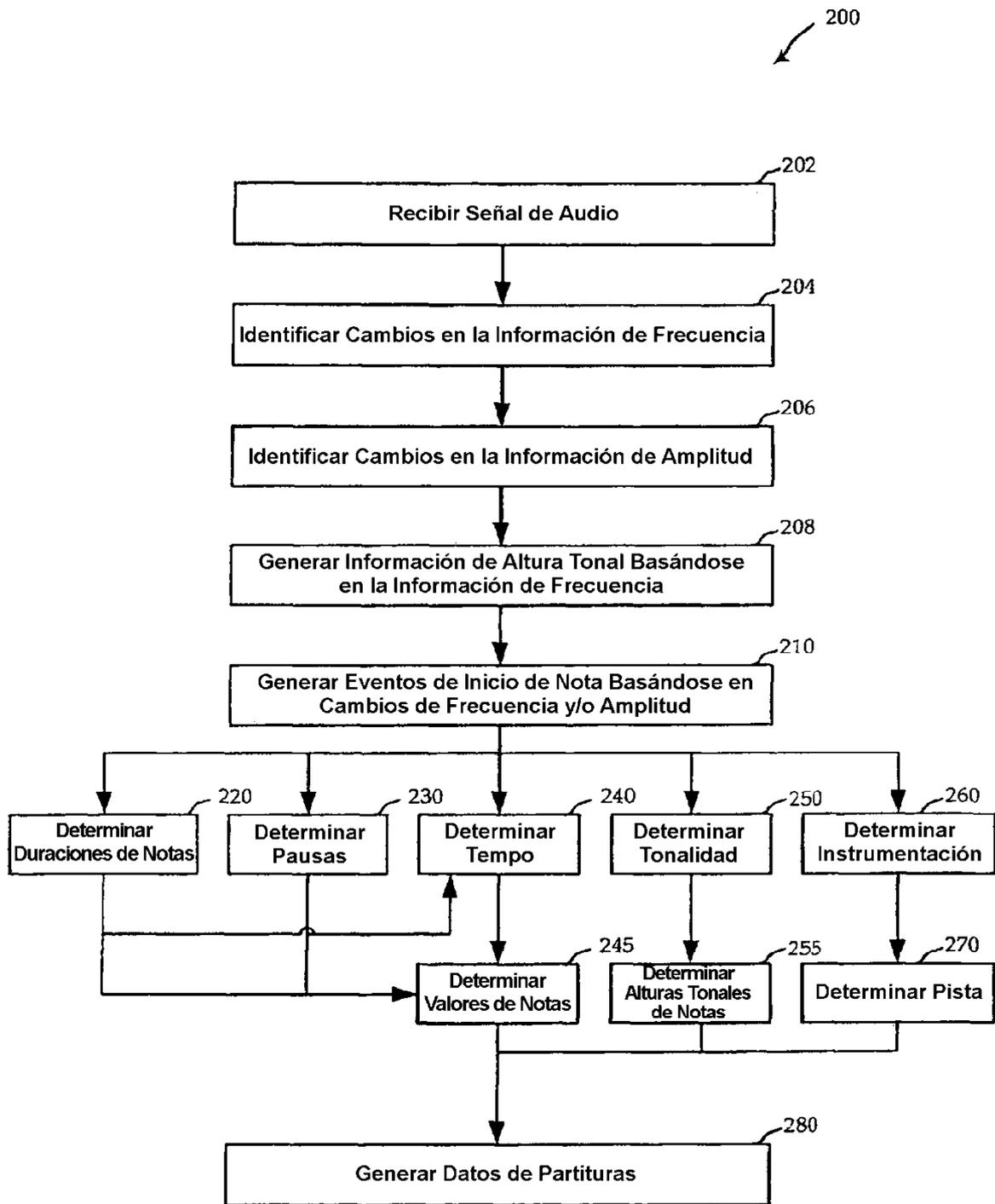


FIG. 2

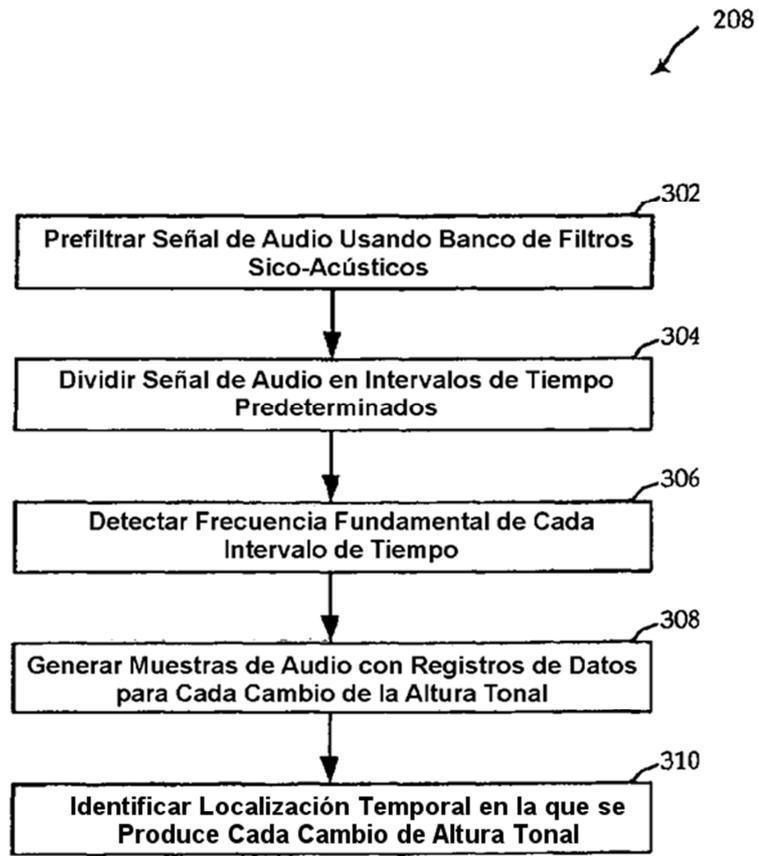


FIG. 3

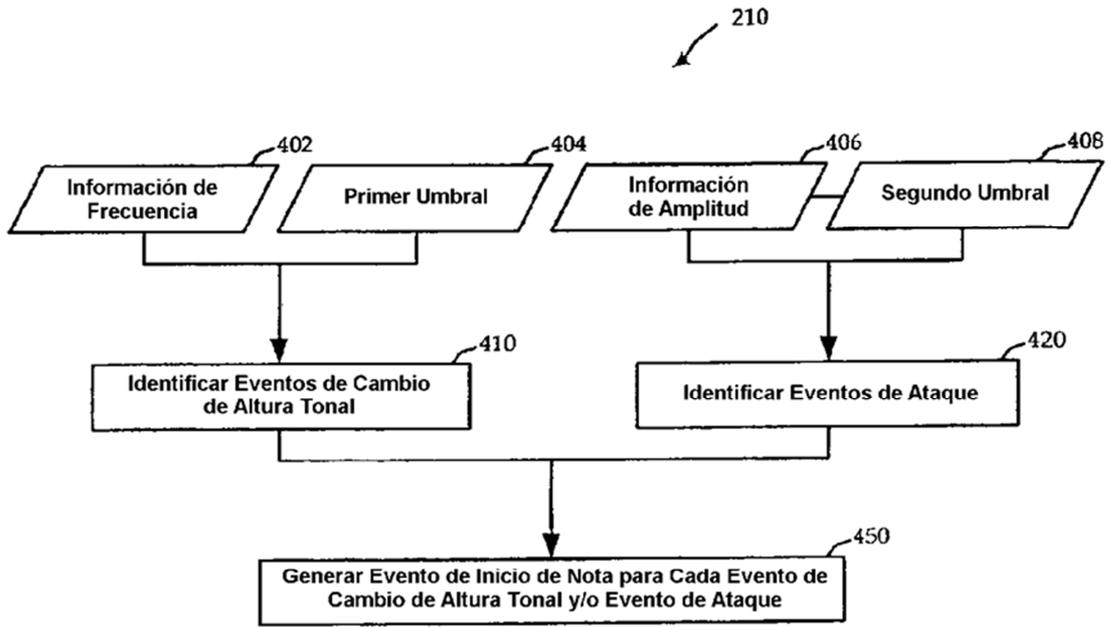


FIG. 4A

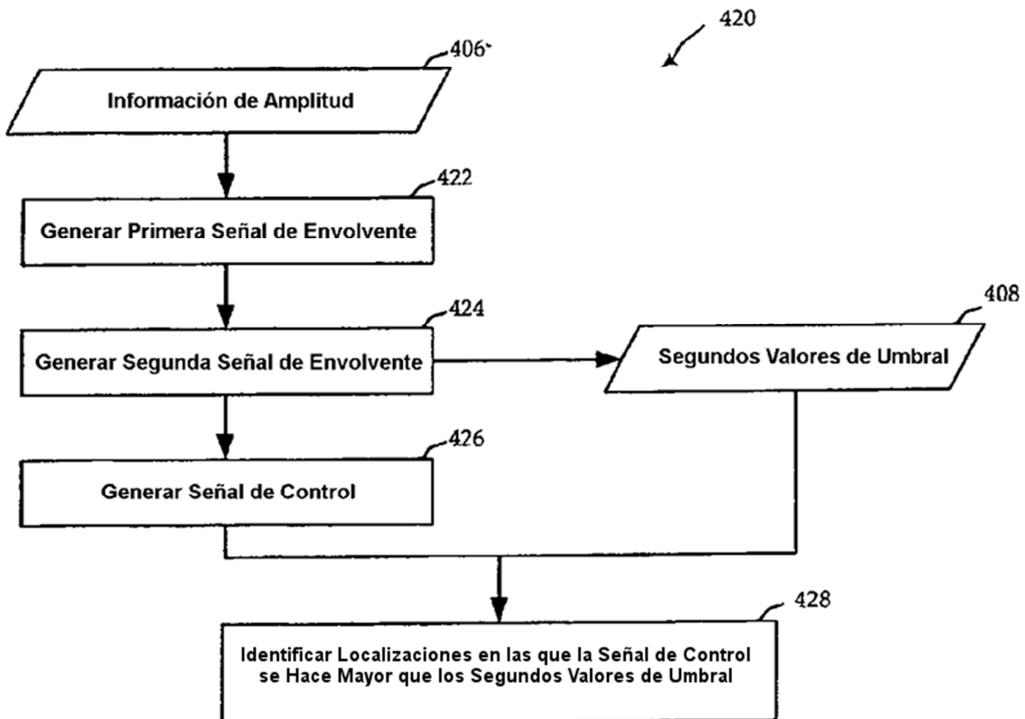


FIG. 4B

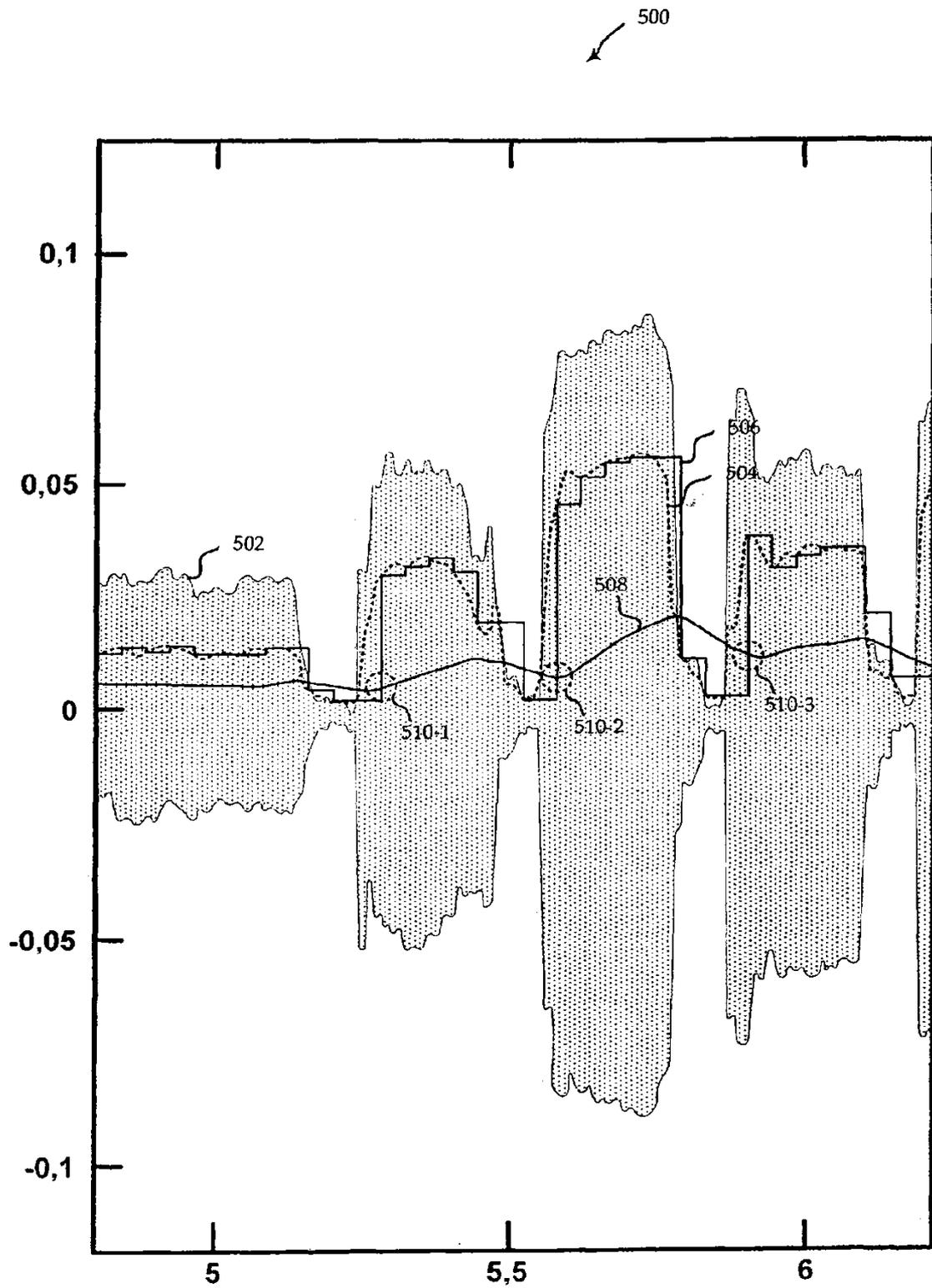


FIG. 5

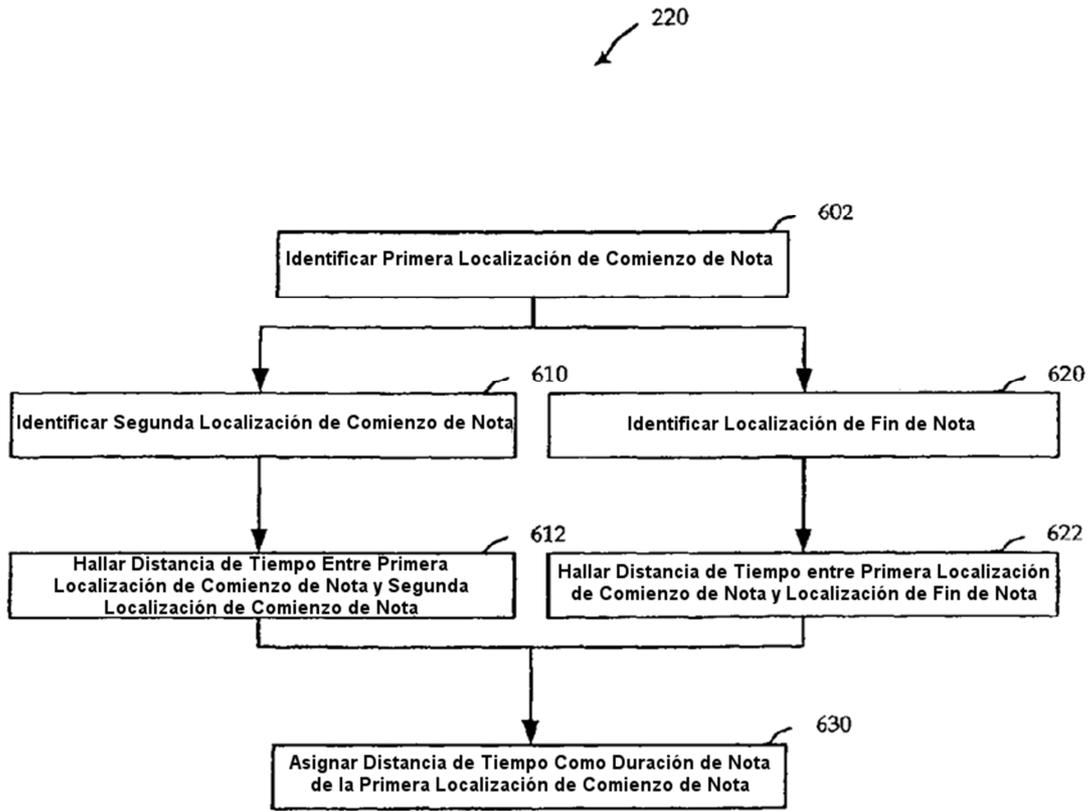


FIG. 6

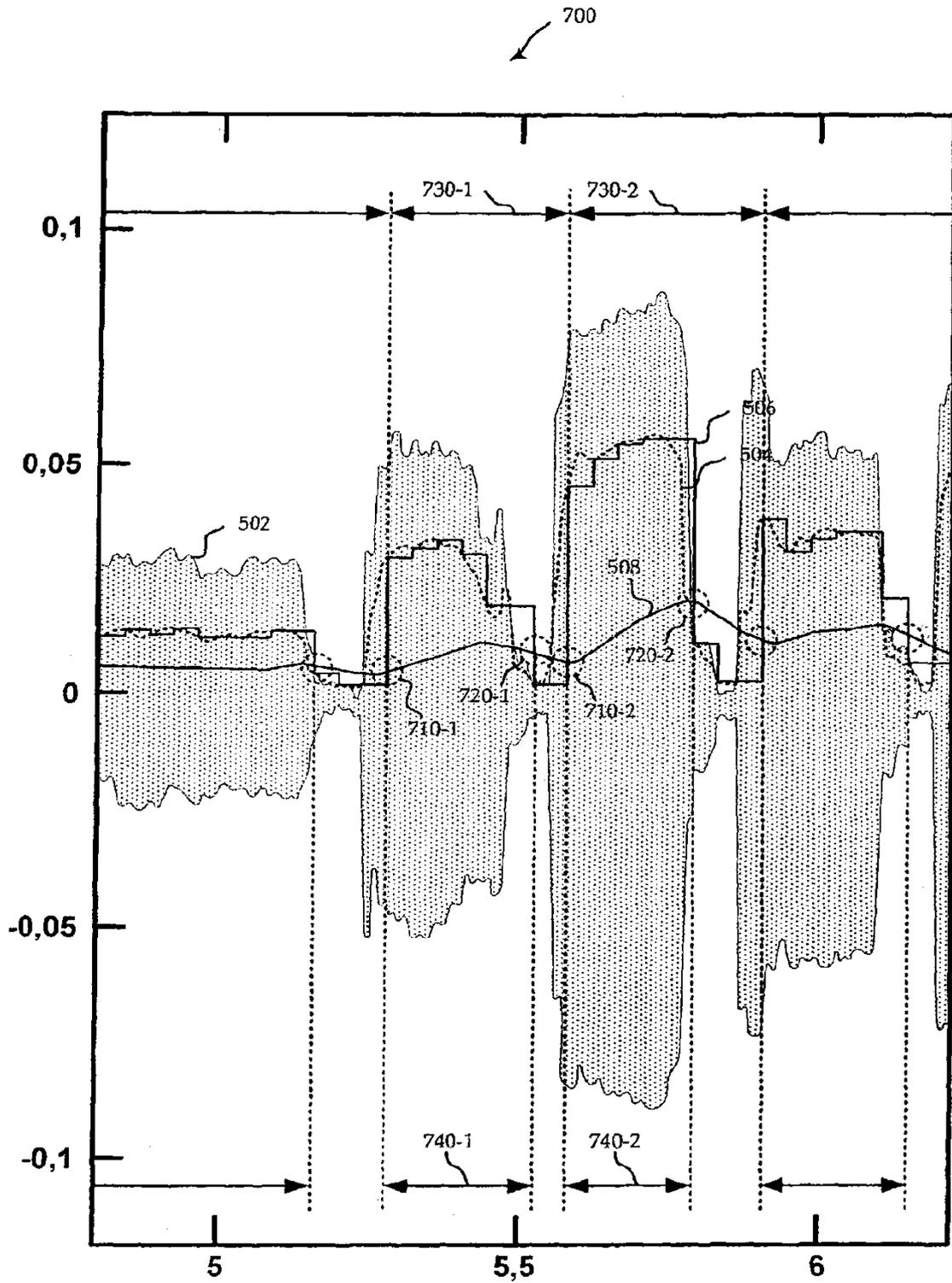


FIG. 7

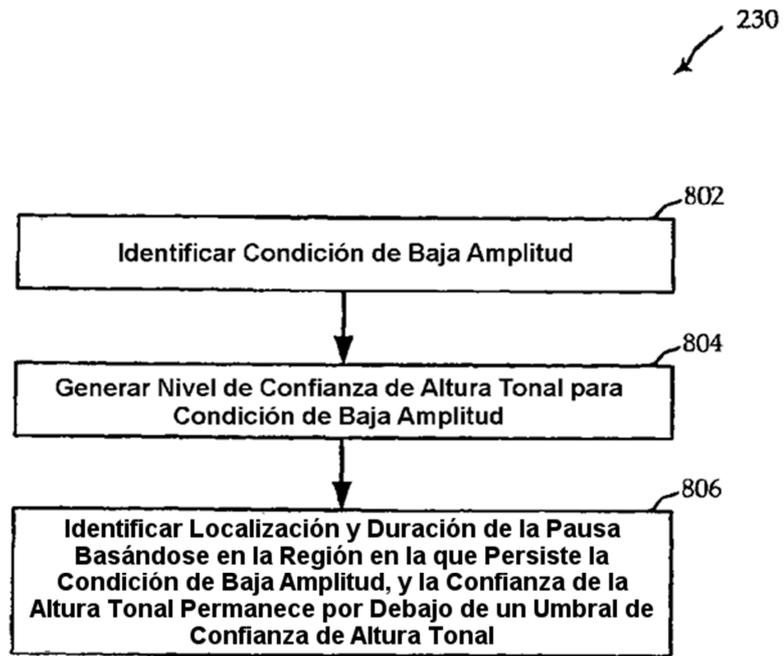


FIG. 8

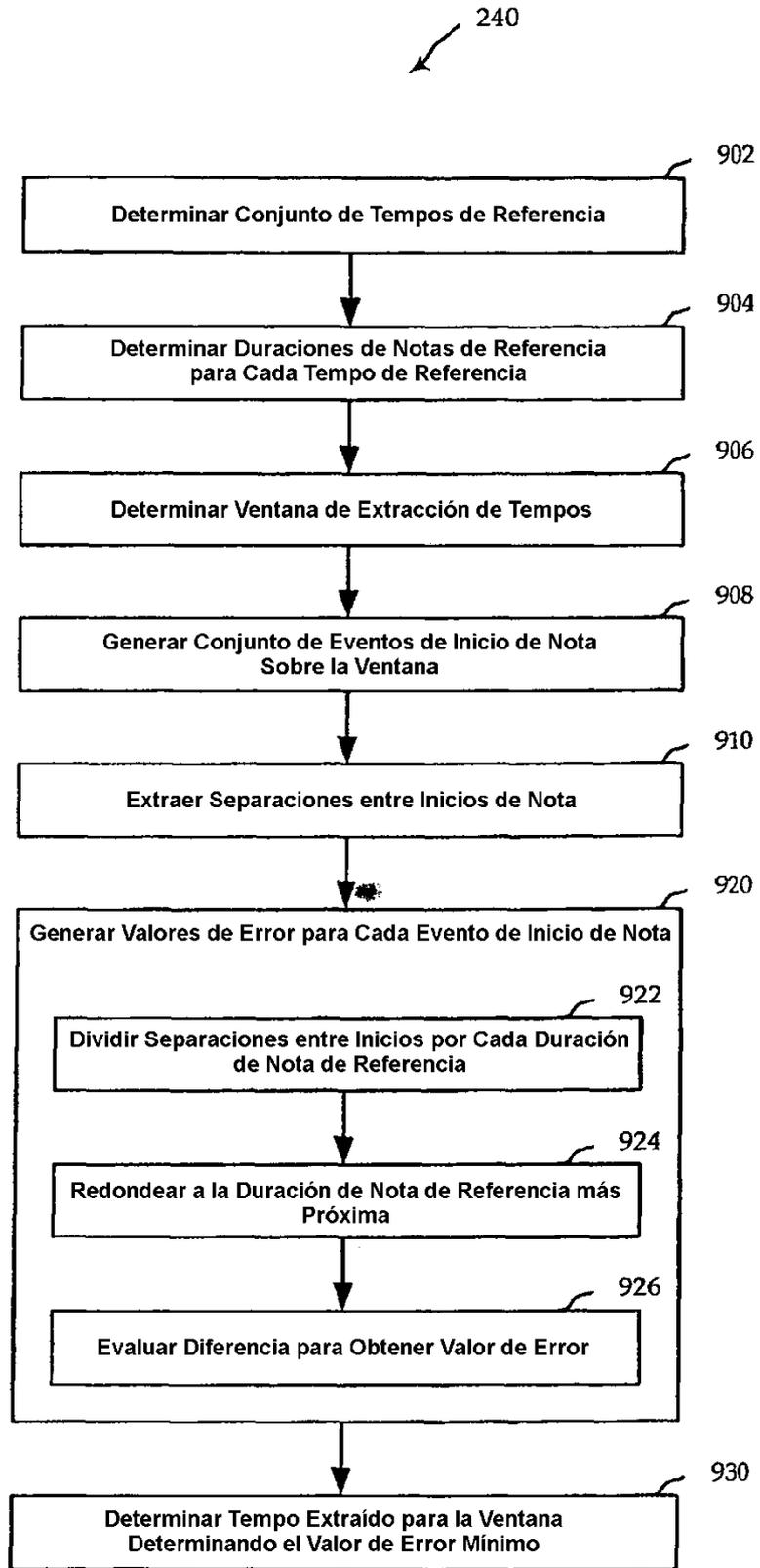


FIG. 9

245

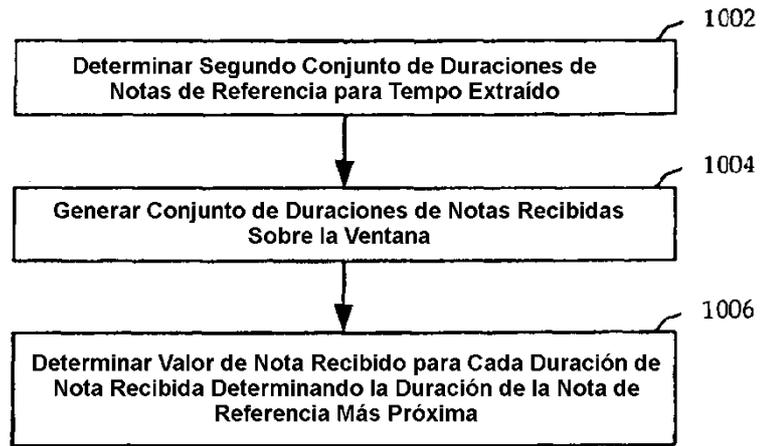


FIG. 10

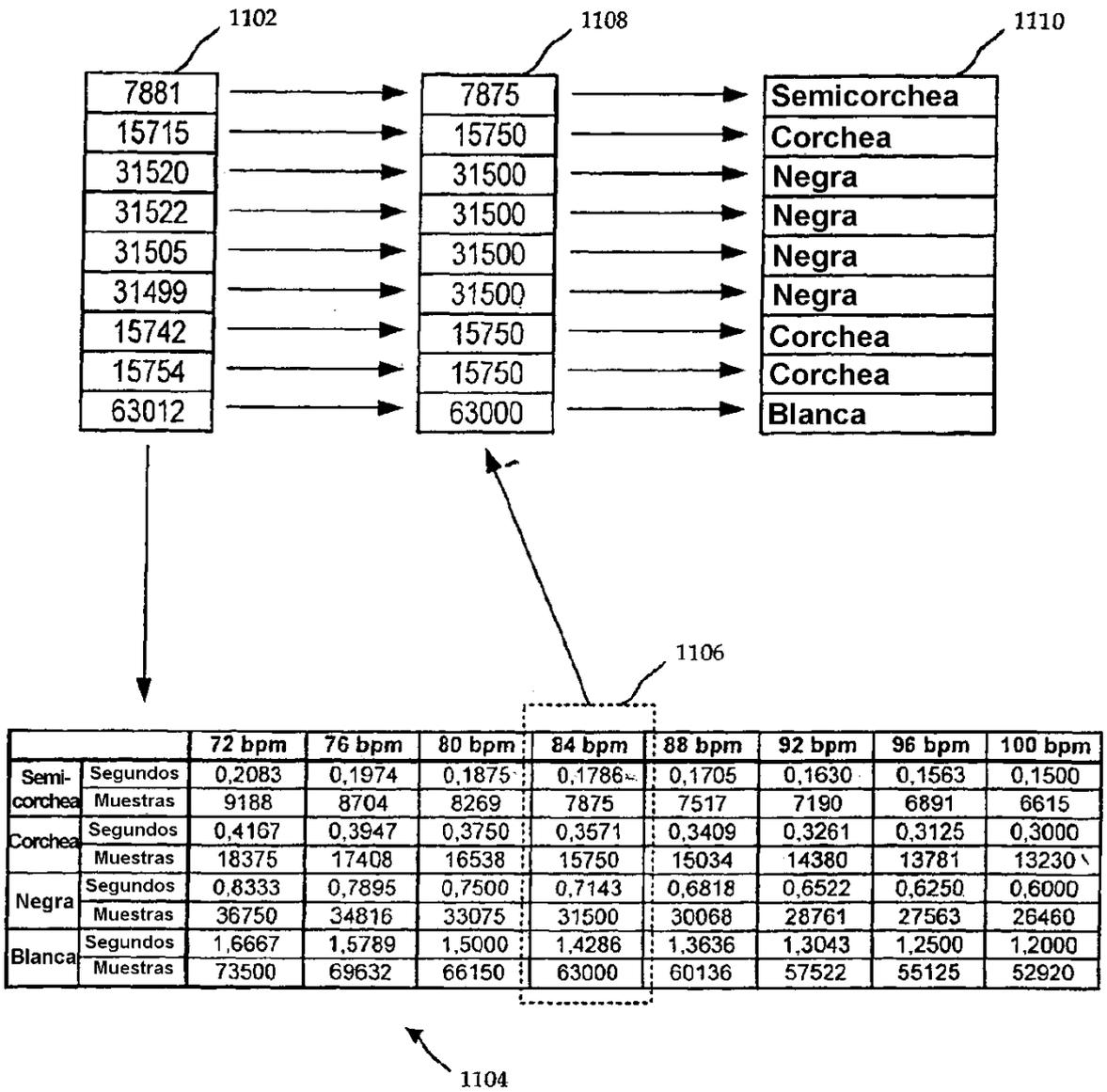


FIG. 11

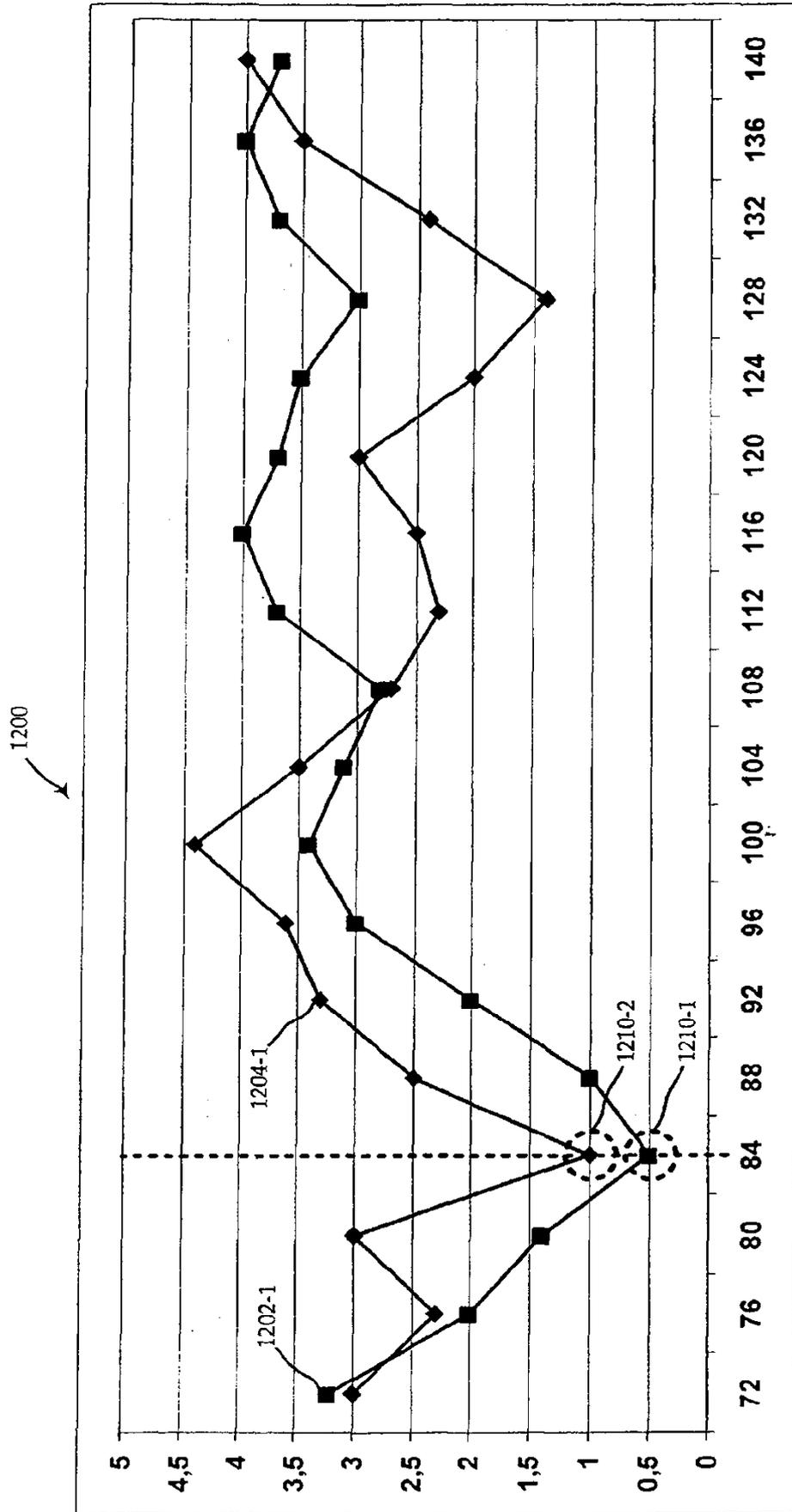


FIG. 12

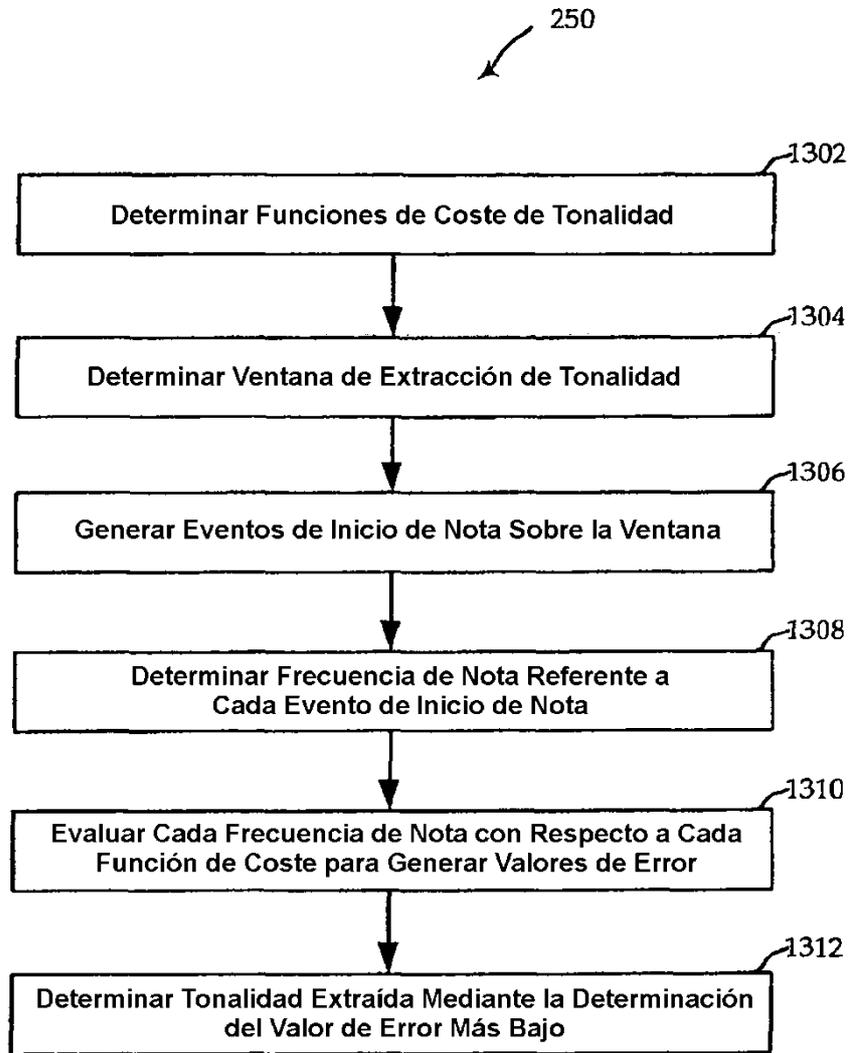


FIG. 13

1400

	Do	Do#/Re ^b	Re	Re#/Mi ^b	Mi	Fa	Fa#/Sol ^b	Sol	Sol#/La ^b	La	La#/Si ^b	Si
1402-1 Do Mayor	1	0	1	0	1	1	0	1	0	1	0	1
1402-2 Do Menor	1	0	1	1	0	1	0	1	1	0	1	0
1402-3 Re Mayor	0	1	1	0	1	0	1	1	0	1	0	1
La Menor	1	0	1	0	1	1	0	1	0	1	0	1

FIG. 14A

1450

	Do	Do#/Re ^b	Re	Re#/Mi ^b	Mi	Fa	Fa#/Sol ^b	Sol	Sol#/La ^b	La	La#/Si ^b	Si
1452-1 Do Mayor	2	0	1	0	2	1	0	2	0	1	0	1
1452-2 Do Menor	2	0	1	2	0	1	0	2	1	0	1	0
1452-3 Re Mayor	0	2	1	0	1	0	2	1	0	2	0	1
La Menor	2	0	1	0	2	1	0	1	0	2	0	1

FIG. 14B

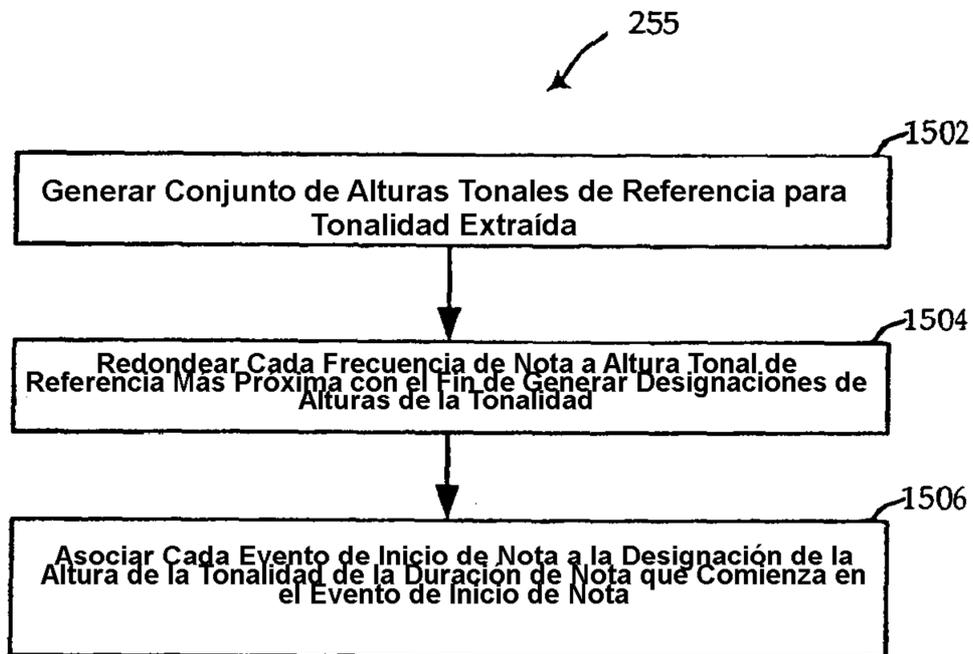


FIG. 15

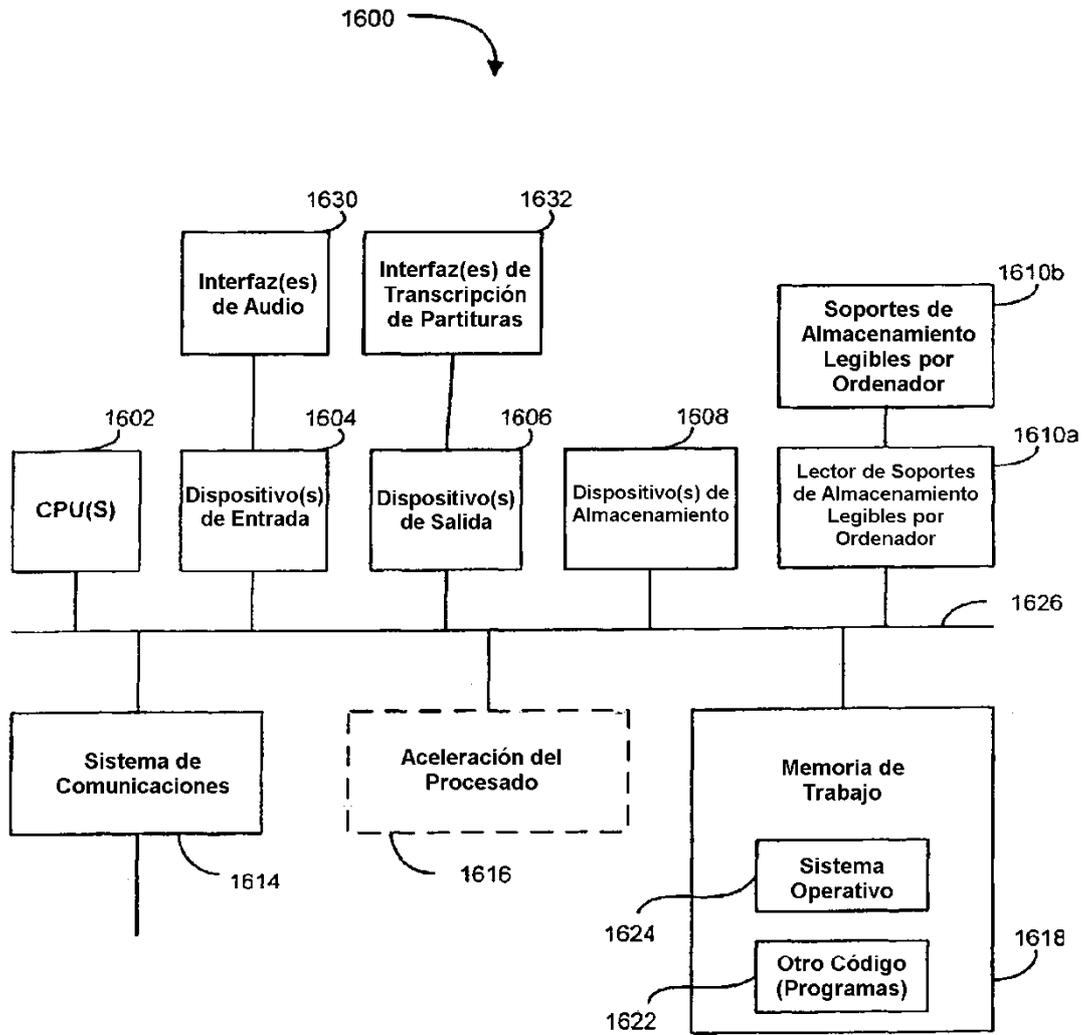


FIG. 16