

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 970 812**

51 Int. Cl.:

G10H 1/00 (2006.01)

G10H 1/057 (2006.01)

G10H 1/40 (2006.01)

G10H 1/36 (2006.01)

G10H 1/46 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **22.06.2016** **E 20156444 (0)**

97 Fecha y número de publicación de la concesión europea: **11.10.2023** **EP 3671725**

54 Título: **Sistema de aumento de contenido de medios y método para alinear transiciones en archivos de medios con eventos variables temporalmente**

30 Prioridad:

22.06.2015 GB 201510907

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
30.05.2024

73 Titular/es:

MASHTRAXX LIMITED (100.0%)
29 Wood Street
Stratford upon Avon,
Warwickshire CV37 6JG, GB

72 Inventor/es:

LYSKE, JOSEPH MICHAEL WILLIAM

74 Agente/Representante:

VALLEJO LÓPEZ, Juan Pedro

ES 2 970 812 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Sistema de aumento de contenido de medios y método para alinear transiciones en archivos de medios con eventos variables temporalmente

5

Antecedentes de la invención

Esta invención se refiere, en general, a un sistema de música y puede específicamente, pero no exclusivamente, aplicarse a un sistema y metodología para la sincronización contextual de secciones (o cortes) de una pista de audio digital en relación con eventos del mundo real. Las secciones se caracterizan en términos de una percepción del usuario y/o evaluación basada en máquina que clasifica cada sección de audio en términos de un "tema" percibido de acuerdo con lo sugerido al usuario o la máquina por un ritmo o pulso subyacente detectado o identificado de otra manera dentro de la sección de audio. En particular, la presente invención se refiere a un sistema de aumento de contenido de medios.

15

Sumario de la técnica anterior

La música, la industria del cine y los juegos, y en particular los aspectos relacionados con el suministro de contenido, están evolucionando. A este respecto, la venta o distribución de (por ejemplo) música o bandas sonoras como archivos digitales transmitidos o descargados se está volviendo dominante en estos mercados. Esto contrasta con la venta de tecnologías de discos compactos y DVD (o, históricamente, discos de vinilo) a través de los establecidos, pero ahora menguados, puntos de venta personalizados.

20

Mientras que las ventas de música son comerciales y de contenido perceptivo y de naturaleza estética, no existe un mecanismo simple y directo para identificar y asignar contenido de audio contextualmente relevante a múltiples escenarios del mundo real. De hecho, se ha reconocido que el contenido soporta un compromiso emocional, fisiológico y/o psicológico del oyente o usuario y, por lo tanto, promueve la experiencia sensorial del oyente o usuario. Por lo tanto, uno de los problemas que enfrenta la industria es la mejor manera de aumentar la experiencia del oyente/usuario, especialmente a nivel personal/individual. De hecho, hace tiempo que se reconoce que la relevancia contextual de o la relación entre una pieza de música y un evento produce reconocimiento o induce una respuesta emocional complementaria, por ejemplo, una sensación de temor o suspense durante una película o una asociación de productos que surgen en la publicidad televisiva. Otro ejemplo se relaciona con el entrenamiento por intervalos en las llamadas clases de ejercicio de ciclo estático "spin" que frecuentemente usan música de fondo continua intercalada con breves ráfagas de música asociadas con mayores tasas de trabajo cardiovascular y mayor emisión de energía a través de un mayor número de revoluciones de piernas por unidad de tiempo. En The encyclopedia of human-computing interaction, 2013, 2ª edición, su autor (M. Hassenzahl) describe cómo la "experiencia de despertar creada por un reloj despertador difiere sustancialmente de la experiencia creada por el amanecer y los pájaros felices", y a continuación se pregunta si es posible "crear tecnología que comprenda las características cruciales del amanecer y los pájaros y que logre ofrecer una experiencia similar, incluso cuando el sol se niega a brillar y las aves ya se han ido a África".

25

30

35

40

En consecuencia, La adaptación técnica del contenido digital y los formatos de archivo ofrece la posibilidad de influir, dar forma y/o diseñar experiencias de usuario (tanto singulares como colectivas) de formas nuevas e innovadoras. La adaptación técnica no lo hace, sin embargo, necesita restringirse a trabajos futuros, pero también puede aplicarse a catálogos posteriores de música almacenados en bases de datos accesibles para proporcionar una nueva oportunidad de vida a dicha música almacenada y similares, por ejemplo, dentro de un entorno licenciado controlado y dentro del contexto de una experiencia de vida personalizada por el usuario.

45

Existen numerosos algoritmos disponibles que realizan tareas relacionadas con la transcripción de música o la interpretación musical. Por ejemplo, los siguientes documentos técnicos identifican técnicas conocidas que pueden ser para la recuperación automatizada de información musical ("MIR") y, por lo tanto, uno o más de tales procesos pueden encontrar un uso como un componente dentro de una o más de las realizaciones o aplicaciones de la invención (como se describe en el presente documento):

50

i) los algoritmos de detección de tono se han tratado en el artículo "Blackboard system and top-down processing for the transcription of simple polyphonic music" por Bello, J.P., que aparece en la publicación técnica Digital Audio Effects, DAFX [2000]. Una segunda técnica se ha descrito en el artículo "Non-negative matrix factorization for polyphonic music transcription" por Smaragdís, P. y J.C. Brown, Taller de IEEE sobre aplicaciones de procesamiento de señales para audio y acústica [2003].

55

60

ii) La detección de inicio y desplazamiento se ha tratado en el artículo "Complex Domain Onset Detection For Musical Signals" por Duxbury, C., et al., 6ª Conferencia internacional sobre efectos de audio digital (DAFx-03) [2003], Londres, UK. Otro artículo en este campo es "Onset Detection Using Comb Filters" por Gainza, M., B. Lawlor y E. Coyle, Taller de IEEE sobre aplicaciones de procesamiento de señales para audio y acústica [2005].

65

iii) Estimación de armadura de clave tratada en "Musical key extraction from audio" por Pauws, S., Simposio internacional sobre recuperación de información musical, Barcelona [2004] y "Detection Of Key Change In Classical Piano Music" de Chai, W. y B. Vercoe, ISMIR, Londres [2005].

iv) Extracción de tempo como se ha tratado en "Tempo and Beat Analysis of Acoustic Musical Signals" por Scheirer, E., Diario de Acoust. Soc. Am., 1998. 103 (1): pág. 588-601. Se ha proporcionado otro enfoque en el artículo "Causal Tempo Tracking of Audio" por Davies, M.E.P. y M.D. Plumbley, Actas de la Conferencia internacional sobre recuperación de información musical, Instituto Audiovisual, Universitat Pompeu Fabra, Barcelona, España [2004].
 5 v) Detección de compás como se ha tratado en documentos escritos en coautoría o en colaboración con Gainza, M. y E. Coyle, especialmente: a) "Time Signature Detection by Using a Multi-Resolution Audio Similarity Matrix", 122ª Convención de la Sociedad de ingeniería de audio, Viena, Austria [2007]; b) "Automatic Bar Line Segmentation", Documento de la Convención de la Sociedad de ingeniería de audio, presentado en la 123ª Convención, Octubre, Nueva York, NY, Estados Unidos [2007] y c) "Time signature detection by using a multi resolution audio similarity matrix", 122ª. Convención de la Sociedad de ingeniería de audio, 5-8 de mayo, Viena, Austria [2007].
 10

Las películas publicadas por corporaciones cinematográficas, como Sony Entertainment, claramente tiene una banda sonora que se edita y, en general, se sincroniza con los fotogramas de acción a medida que se monta en la película.
 15 De hecho, las imágenes en movimiento (por ejemplo, en películas y televisión y juegos informáticos) a menudo necesitan una pista musical para acompañarlas. Este proceso se conoce como sincronización de música. Los titulares de derechos y los cineastas emplean a profesionales para realizar búsquedas en grandes catálogos de música creativa para encontrar piezas apropiadas para la sincronización. A continuación se realiza una laboriosa tarea de edición para ajustar la pieza de música al metraje; esta tarea requiere equipo especializado y no es personalizable por el usuario final/observador de películas. De hecho, la partitura se monta y aprueba subjetivamente por el estudio. Tampoco puede configurarse dinámicamente contra eventos en tiempo real.
 20

También, otro problema se relaciona con el acceso y la búsqueda de información almacenada en una gran cantidad de bases de datos diferentes que tienen diferentes formatos de almacenamiento. A este respecto, el lenguaje empleado dentro de la comunidad de la industria de sincronización es multivariado e involucra múltiples repertorios que a veces son conflictivos. De hecho, la terminología puede tener diferentes significados en diferentes contextos, y estos son dinámicos y negociados. El uso de una interfaz de lenguaje natural se traduciría en un sistema de búsqueda significativo para los usuarios. Se ha reconocido ampliamente que la expansión masiva en la música digital a lo largo de los últimos años ha presentado problemas complejos para los usuarios y requiere poderosas técnicas y herramientas de gestión del conocimiento.
 25
 30

En una línea similar a la sincronización de medios cinematográficos, las llamadas "clases de spinning" que usan bicicletas estáticas hacen uso de una música de acompañamiento, pero esto es solo un fondo que el instructor de PT usa como distracción y ayuda para que los miembros de la clase emprendan un entrenamiento intensivo o una recuperación activa. No hay capacidad para la pista de música de fondo y, de hecho, el sistema de radiodifusión de audio que reproduce la música es dinámico y reacciona a los cambios en tiempo real. Al igual que todos los sistemas de radiodifusión de audio o altavoces existentes, la pista de audio simplemente se reproduce y, por lo tanto, sigue con precisión la disposición de la pista de música pregrabada.
 35

El documento WO 2014/047322-Kelleltt et al./UJAM se refiere a un sistema para la reorganización automática de una composición musical. El sistema incluye un proceso de asignación de metadatos a una pieza de música existente para dividirla en secciones e identificar secciones del mismo tipo, y la lógica para eliminar y reorganizar las secciones para producir una reproducción personalizada con un duración deseada y opciones adicionales para incluir o eliminar secciones o instrumentos específicos bajo el control de un usuario.
 40
 45

El artículo técnico "Advanced Synchronization of Audio or Symbolic Musical Patterns: An Algebraic Approach", de Florent Berthaut et al. publicado en IEEE Sixth International Conference On Semantic Computing, 2012, IEEE, (19 de septiembre de 2012), ISBN 978-1-4673-4433-3, páginas 302 - 309, define un modelo genérico para patrones musicales de audio o simbólicos finitos que codifican estructuralmente un mecanismo de sincronización.
 50

El documento US 2015/018993 - Trivedi son antecedentes adicionales a la presente invención en el sentido de que este se refiere al procesamiento de audio para su reproducción, y más específicamente al procesamiento de archivos de audio para proporcionar una transición entre pistas de audio sucesivas durante la reproducción. De acuerdo con algunos ejemplos, un flujo incluye determinar, con un dispositivo informático, una primera característica de audio de una primera pista de audio y determinar, con el dispositivo informático, una segunda característica de audio de una segunda pista de audio. El flujo puede incluir además recibir, en el dispositivo informático, datos que representan una activación generada por el usuario. El flujo además puede determinar un parámetro de transición, sensible a la activación generada por el usuario, para la primera pista de audio y la segunda pista de audio basándose en una o más de entre la primera característica de audio y la segunda característica de audio. También, el flujo puede provocar la presentación de una transición desde la primera pista de audio a la segunda pista de audio.
 55
 60

Sumario de la invención

De acuerdo con un primer aspecto de la invención, se proporciona un sistema de aumento de contenido de medios de acuerdo con la reivindicación 1. En otro aspecto de la invención, se proporciona un método para alinear secciones de al menos un archivo de medios digital para componer un producto de medios, el método de acuerdo con la
 65

reivindicación 10. Diversos aspectos y realizaciones de la invención como se resume en las reivindicaciones adjuntas y la siguiente descripción pueden implementarse como una solución de hardware y/o como software.

Breve descripción de los dibujos

5 Las realizaciones a modo de ejemplo de la presente invención se describirán ahora haciendo referencia a los dibujos adjuntos en los que:

- 10 la figura 1 es un diagrama representativo de una relación que se establece entre un evento del mundo real y una pieza de audio, tal como una sección de música y en la que la relación se establece de acuerdo con la presente invención;
- la figura 2 es una representación esquemática de un sistema de acuerdo con una realización preferida de la presente invención;
- 15 la figura 3 es un diagrama funcional de la inteligencia de montaje de secciones empleada dentro del sistema de la figura 2;
- la figura 4 es un diagrama funcional para el montaje de pistas y la creación de metadatos de acuerdo con un proceso de creación preferido;
- la figura 5 es la representación en el dominio de tiempo de una estrategia de corte preferida entre un punto de salida y un punto de entrada a diferentes temas musicales dentro de las secciones de audio;
- 20 Las figuras 6a, 6b y 6c muestran una atenuación cruzada o corte entre las secciones de audio activo y de destino en relación con tanto una base de tiempo común como unos eventos de inicio respectivos, la atenuación cruzada o corte de acuerdo con una realización preferida;
- la figura 7, compuesta de las figuras 7a a 7c, es una representación espectral de diferentes secciones de música, teniendo las secciones unos puntos de salida y entrada determinados de acuerdo con una realización preferida de la presente invención;
- 25 la figura 8 es un diagrama de sincronización que muestra la sincronización relativa entre diferentes tipos de compás en diferentes secciones de música;
- la figura 9, compuesta de las figuras 9a a 9d, muestra un proceso por el que se determina un inicio, el proceso empleado por el sistema preferido para realizar la atenuación cruzada o corte ejemplificado en las figuras 6a a 6c;
- 30 y
- la figura 10 es una configuración a modo de ejemplo de una base de datos de pistas de acuerdo con una realización preferida.

Descripción detallada de una realización preferida

35 La figura 1 es un diagrama 10 representativo de una relación que se establece entre un evento del mundo real 12 y un tema musical 14 de una pista de audio, estando la relación establecida de acuerdo con la presente invención. Por ejemplo, el evento 12 puede monitorizarse o de otra modo detectarse (tal como a través de un sensor o dispositivo de telemetría usado por una persona que hace ejercicio), o puede pregrabarse y estar presente en un medio (tal como una sucesión de fotogramas de escenas de una película). Con el paso del tiempo, los eventos del mundo real 12 cambian, por lo que los eventos 12 pueden diferenciarse entre sí.

45 Tomando un ejemplo de ejercicio como puramente a modo de ejemplo (o de hecho una película de alguien haciendo ejercicio), la persona que hace ejercicio puede pasar por un estado de calentamiento entre, digamos, el comienzo (tiempo cero) y el final del minuto dos, una fase de ejercicio activo entre el final del minuto dos y el final del minuto seis, un período de recuperación entre el final del minuto seis y el final del minuto ocho y finalmente un período de enfriamiento entre el final del minuto ocho y el final del minuto nueve. Los diferentes períodos activos dentro de este régimen de ejercicio potencialmente, si no siempre, garantizan un tratamiento diferente en términos de acompañamiento musical contextual.

50 La duración de la actividad se describe en términos de minutos completos, pero esto es puramente a modo de ejemplo y podría basarse igualmente en segundos u otros períodos. El tiempo de las transiciones entre las actividades de ejercicio también se selecciona arbitrariamente en el ejemplo para facilitar la explicación y, por lo tanto, las transiciones podrían producirse en cualquier punto de activación apropiado. Por ejemplo, las transiciones pueden accionarse por un sensor que se monitoriza activamente e informando la frecuencia cardíaca de una persona durante un programa de ejercicio. En términos de una película, los puntos de transición seleccionados pueden estar al final de una escena o estar vinculados al aspecto de un personaje o al aspecto de un discurso.

60 En la figura 1, para proporcionar un efecto inmersivo, la actividad 12 se ve aumentada por una pista musical contigua 14 que sigue una ruta natural que refleja la grabación original del artista. Sin embargo, mientras que el inicio de la pista 16 puede comenzar inicialmente para coincidir libremente y aproximadamente con el comienzo mismo del estado de calentamiento en el momento cero, la pista se disocia rápidamente con el evento y, por lo tanto, solo el fondo, es decir, el audio natural es subóptimo con el fin de soportar o aumentar el evento del mundo real al que se asociado libremente.

65 Sin embargo, se ha reconocido que diferentes secciones de, por ejemplo, una pista de audio de música en realidad

puede tener una relevancia directa para diferentes actividades de eventos temporales y que se desean transiciones entre diferentes secciones de audio (basadas en un tema comúnmente identificado en la sección y el evento temporal) para aumentar una experiencia sensorial general. En otras palabras, hay temas presentes en las secciones 14 de una partitura musical, teniendo estas secciones diferentes duraciones y diferentes cualidades emotivas que un usuario o dispositivo puede clasificar basándose en una evaluación cualitativa (y/o cuantitativa). Por lo tanto, el término "tema" debería entenderse como una descripción genérica de las propiedades seleccionadas o los atributos percibidos que un usuario y/o un proceso MIR pueden asignar a una sección de música (u otra muestra de audio) basándose en un significado contextual percibido para un sección de audio identificada.

Por lo tanto, una realización de la presente invención opera para identificar secciones y almacenar segmentos de música muestreados (o similares) dentro de una base de datos. Cada sección 14, cada una de las cuales puede tener una duración diferente T_1 - T_4 , está codificada con metadatos que identifican la sección (y su tema) y/o específicamente las cualidades contextuales resolubles de la sección 14. El acceso y la referencia de los metadatos permite que las secciones se reordenen posteriormente a tiempo con el fin de permitir la alineación de las secciones con los eventos del mundo real relacionados, es decir, relevantes.

Preferentemente, el sistema nunca altera el archivo de audio original, sino que usa los metadatos montados que permiten que un motor de reproducción salte alrededor de un archivo codificado adecuadamente, editándolo y atenuándolo de manera cruzada casi en tiempo real (guardándolo en un búfer apropiado para permitir su procesamiento).

A modo de ejemplo no limitativo en el contexto del ejercicio, una pista de música puede incluir un tempo alegre para un coro, con el tempo alegre apropiado para una frecuencia cardíaca más rápida inducida por una mayor tasa de producción de energía durante una parte de sprint del ejercicio. Por el contrario, la misma pista de música puede incluir una sección de medio tiempo. Por lo tanto, la sección de medio tiempo sería más apropiada para un período de recuperación activa durante el ejercicio, pero no el sprint. Por lo tanto, la invención reivindicada actualmente produce varias secciones identificadas, es decir, cortes de una pista, que pueden moverse a tiempo con el fin de alinearse más apropiadamente con la actividad accionada externamente. Este movimiento en el tiempo puede provocar el reordenamiento de las secciones, es decir, una secuencia naturalmente contigua en una pista de música original puede producir una secuencia $T_1, T_2, T_3, T_1, T_5, T_6, T_1$ donde T_1 es un coro repetido y T_2 a T_6 son versos, temas o movimientos musicales. Este movimiento en el tiempo también puede hacer que la duración total de la salida de audio se extienda o se trunque en relación con la pista de música original. La identificación del tema y el almacenamiento de una caracterización relacionada de la sección en metadatos podría conducir a una ordenación realineada de la sección (en relación con una película) que produce una salida de audio que tiene una secuencia diferente $T_1, T_1, T_6, T_2, T_3, T_1, T_1, T_1$. Este reordenamiento se representa en la figura 1 por el espectro de audio "Mashtraxx" 20 que muestra este reordenamiento de las secciones de Mashtraxx. Preferentemente, las secciones adyacentes son contiguas entre sí, por ejemplo, no hay un cambio perceptible en el ritmo de la música y/o no hay un silencio perceptible entre las secciones adyacentes. Más adelante se describe un mecanismo de atenuación entre una sección de muestra activa de una pista y una sección almacenada (de la misma pista o de una pista diferente) teniendo en cuenta la figura 5 y la figura 7.

Las realizaciones de la presente invención permiten que las secciones musicales identificadas se sincronicen y se vinculen con un evento del mundo real. La sincronización lograda incluye una transición musical perfecta, lograda usando un desplazamiento de sincronismo, tal como un avance relativo, preferentemente para que coincida con precisión con una firma o pulso musical preexistente pero identificado, entre secciones preidentificadas potencialmente dispares que tienen diferentes temas emotivos definidos por sus respectivos tipos de compás, intensidades, claves, ritmos musicales y/o frases musicales. Por lo tanto, las realizaciones preferidas aumentan la experiencia sensorial general al cambiar entre secciones musicales (y, por lo tanto, temas diferentes) dentro del contexto de lo que se está produciendo en el entorno físico circundante, por ejemplo, durante las diferentes fases de un entrenamiento cardiovascular en una clase de step (plataforma rectangular), la velocidad y la intensidad de la música aumentan durante los períodos de sprint y disminuyen durante los períodos de recuperación. La música acompañante (o contenido audible, tal como el diseño de sonido no diegético y diegético o el sonido acústico) se selecciona automáticamente en tiempo real para acompañar un evento físico cambiante, que puede monitorizarse electrónicamente en tiempo real, tal como con un monitor de frecuencia cardíaca y, por lo tanto, forma parte integral pero complementaria de una experiencia sensorial general. En lugar de monitorizar en tiempo real, la muestra de audio acompañante, tal como un fragmento de música, puede preprogramarse para atarse y jugarse con un punto de activación, tal como un cambio de escena dentro de un videojuego.

Volviendo a la figura 2, se muestra una representación esquemática de un sistema 30 de acuerdo con una realización preferida de la presente invención.

Tal como se ha indicado anteriormente, como una entrada, el sistema 30 adquiere eventos externos 12 del mundo real. Estos pueden ser eventos detectados en tiempo real o eventos grabados almacenados en un medio y presentados al sistema a través de una interfaz de comunicaciones 32. Un entorno de procesamiento 34 normalmente incluye al menos un procesador que ejecuta código de programa almacenado en la memoria 34.

El entorno de procesamiento 34 puede soportarse en varias plataformas, incluido un servidor conectado a través de una red, un ordenador de tableta, un PC o un teléfono inteligente.

5 La memoria 36 puede ser local para el entorno de procesamiento 34 o localizada remotamente en un sistema distribuido. Las aplicaciones ("apps") almacenadas en la memoria 36 permiten el análisis automatizado de los eventos del mundo real y, asimismo, permiten la caracterización de datos de dominio de tiempo muestreados dentro de los eventos relativos a un vocabulario preidentificado de términos 38 (que también se almacenan en la memoria 346). En términos prácticos, una señal de entrada correspondiente a un evento externo monitorizado o grabado 12 puede contener múltiples características diferentes, por ejemplo, diferentes frecuencias cardíacas asociadas con diferentes fases del ejercicio y la recuperación activa (tal como se muestra en la figura 1) detectadas e informadas desde un monitor en tiempo real, o de lo contrario en el contexto de un video puede haber diferentes expresiones faciales identificables, diferentes intensidades de píxeles y/o tasas de movimiento de píxeles para un blob a medida que se mueve a través de una pantalla. Pueden definirse otras características que reflejan una emoción o actividad.

15 El hilo común es que las características identificables varían con el tiempo, ya que la actividad exterior cambia con el tiempo, ya que el estado emocional o físico del entorno o del individuo cambia con el tiempo. Por lo tanto, la entrada del evento, independientemente de la forma, puede considerarse como un conjunto de múltiples segmentos de eventos contiguos de datos de eventos muestreados que tienen temas diferentes pero identificables. Funcionalmente, el entorno de procesamiento 34 está configurado para hacer referencia a cada segmento de evento contra al menos una palabra de vocabulario prealmacenada, con esta referencia grabada de manera permanente o en una base transitoria con el segmento del evento. Las palabras de vocabulario para cada segmento de evento actúan como una clave o "elemento informativo" que puede usarse posteriormente para hacer una referencia cruzada y, en última instancia, seleccionar una sección de audio relevante que se haya caracterizado de manera similar con una palabra de vocabulario correspondiente o estrechamente relacionada que pueda almacenarse en metadatos para la pista de audio, almacenada en un archivo asociado y más preferentemente compartimentada dentro de un encabezado para cada sección de audio. A menos que el contexto específico requiera una interpretación más limitada, las expresiones "sección de audio", "segmento" y "fragmento de audio" deberían considerarse como equivalentes y representativas de la variación de muestras de longitud desde dentro de un archivo de audio digital.

20 Como entrada secundaria al entorno de procesamiento 34, puede proporcionarse una interfaz de usuario 40 para permitir que un usuario critique la entrada de evento externo 12. Esta función crítica puede ser independiente o complementaria al análisis automatizado que genera los elementos informativos. Por lo tanto, la crítica manual también atribuye y adjunta palabras de vocabulario a segmentos de eventos y, por lo tanto, proporciona un proceso alternativo o complementario para generar elementos informativos.

35 Volviendo a la figura 2, El análisis funcional de los eventos externos que se realiza por el entorno de procesamiento puede incluir:

40 i) sincronización relativa y/o real 42 dentro del evento externo 12. Por ejemplo, el análisis de sincronización puede identificar eventos que satisfacen un valor umbral mínimo, eventos/características que tienen una duración sostenida y/o puntos de éxito. Un punto de éxito es un momento dado en la línea de tiempo de un juego o película cuando se produce una acción que requiere que se produzca una característica musical, es decir, un "golpeo", al mismo tiempo. Otro nombre de la industria para un punto de éxito es "Mickey Mousing", por ejemplo, en una caricatura de Tom y Jerry, puede sonar una fuerte nota de bloque de madera después de que una bola de boliche rueda a lo largo de un estante y a continuación caiga sobre la cabeza de Tom, el gato. Como alternativa, a veces los puntos de éxito duran un tiempo determinado; esto se conoce más apropiadamente como una "característica sostenida". Por ejemplo, después de que las bolas de boliche hayan caído sobre la cabeza de Tom y después de que hayan rebotado, Tom tiembla incontrolablemente durante tres segundos. Por lo tanto, una característica sostenida relacionada podría realizarse por el sonido emitido por una regla cuando se hace sonar en un escritorio, con el sonido de la vibración sostenida durante un período de tres segundos antes de que se atenúe o termine.

50 ii) Intensidades 44 y la escala de intensidades dentro del evento externo. Por ejemplo, las intensidades pueden basarse en niveles de luz relativos o niveles de sonido o tasas de cambio en una función monitorizable (tal como la frecuencia cardíaca) recibida de la fuente exterior a través de la entrada 32. La intensidad también puede relacionarse con una intensidad percibida emocionalmente que identifica cómo una sección en particular en una pieza de audio es significativa o está subordinada en relación con otras secciones en toda la pieza de audio, que me puede dar una canción completa o una compilación de pistas de diferentes fuentes o grabaciones de audio, por ejemplo, diferentes pistas de diferentes CD de audio. Por lo tanto, la intensidad puede ser un conjunto de medidas subjetivas de acuerdo con el gusto o la preferencia de un usuario.

60 iii) Análisis de la función de audio 46, que incluye pero no se limita al análisis de pulso, análisis de tipo de compás, identificación del movimiento climático, aumento o disminución de intensidad, caída dramática donde el audio cesa de repente, análisis estático o rítmico, subrayado e identificación de un riff.

Una vez que los segmentos del evento se han codificado y se han producido elementos informativos, estos segmentos de eventos pueden transmitirse a través de una red o almacenarse de otra manera 50 para su posterior recuperación y uso por la inteligencia de montaje de secciones 52. Esta inteligencia de montaje de secciones 52 se basa en el soporte de procesamiento de una inteligencia artificial ("AI") y, como alternativa, puede denominarse "marco de trabajo

Mash-traxx". La inteligencia de montaje de secciones 52 está configurada para proporcionar un procesamiento adicional y la generación de una nueva muestra de medios (o "archivo de medios compuesto" 54) en el que el nuevo audio, tal como una nueva pista de música, mejora los segmentos de eventos del evento del mundo real de una manera perfecta y secuenciada. El nuevo audio puede, de hecho, complementar o sustituir las muestras de audio capturadas del evento del mundo real.

La inteligencia de montaje de secciones 52 responde a entradas adicionales, es decir, una pista de audio mejorada 58 que contiene metadatos de Mashtraxx 58 para segmentos o fragmentos de la pista de audio original. La pista de audio puede, de hecho, ser muestras múltiples de múltiples pistas seleccionadas por el usuario o de un solo archivo de audio grabado (que no necesita ser música). La pista de audio 56 podría, por ejemplo, proporcionarse por una biblioteca de iTunes® o transmitirse o adquirirse de un repositorio de música. La generación de metadatos de Mashtraxx se describirá posteriormente.

La generación de metadatos de Mashtraxx se describirá con más detalle a continuación y en relación específica con la figura 4. Sin embargo, en general, Los metadatos de Mashtraxx proporcionan una definición de segmento de audio, incluyendo el o los tipos de compás de pista y otras propiedades y/o funciones musicales, suministrados como datos asociados con la pista de audio. Un archivo de audio digital modificado para una pista puede incluir etiquetas de metadatos que están casadas con eventos musicales en la pista, tal como unas caídas de címbalos y unos pulsos de tambor. Como alternativa, puede accederse o transmitirse el acceso a los metadatos de Mashtraxx desde un repositorio/base de datos gestionado externamente al que se accede a través de una aplicación local que ejecuta un dispositivo. En este último aspecto, una aplicación basada en dispositivo está preferentemente dispuesta para extraer metadatos de Mashtraxx de la base de datos remota basándose en un estado registrado del trabajo artístico original con derechos de autor, es decir, la pista de audio, con el dispositivo/usuario local. En el caso de que la pista de audio original se considere una copia ilegal, puede imponerse una prohibición de acceso a los metadatos de Mashtraxx y/o, en el caso contrario, la aplicación puede marcar la copia ilegal o deshabilitar la reproducción de la pista original hasta que la pista/usuario se valide por una licencia autenticada para la pista original.

En una realización antifalsificación preferida, una aplicación instalada localmente verifica el contenido local mediante huellas digitales del audio; esto puede basarse en cualquier cantidad de técnicas conocidas, incluyendo metadatos de pista en el encabezado de la pista y/o el análisis dentro de la pista a través del muestreo de audio. A continuación, la aplicación se configura para verificar una base de datos central que contiene datos de Mashtraxx, revelando la verificación si existen datos de Mashtraxx para la pista identificada. Si es así, la aplicación se configura para presentar los datos de Mashtraxx al usuario como una opción de Mashtraxx dentro de la aplicación.

La adquisición de un estado con licencia se entiende bien y, en general, se basa en un proceso de registro. El registro para el uso de derechos de autor no está, sin embargo, relacionado con los problemas fundamentales de la presente invención y, especialmente, la creación de etiquetas de metadatos que se usan para aumentar una experiencia sensorial inmersiva asociada con el audio coordinado que complementa los cambios en los eventos del mundo real.

La figura 3 es un diagrama funcional de la inteligencia de montaje de secciones 52 empleada dentro del sistema de la figura 2.

Para las secciones de audio que se van a insertar y que se han caracterizado anteriormente en términos de sus temas inherentes (para los que se ha colocado una etiqueta en los metadatos correspondientes de las secciones de audio, como se explicará más adelante), la inteligencia de montaje de secciones 52 selecciona 59 y ordena 60 las secciones de audio para que correspondan al o los eventos de palabras reales. Esto da como resultado un conjunto ordenado de secciones de audio 62 que pueden incluir pasajes de música contiguos que se repiten. La selección de la sección de audio apropiada requiere la consideración y el análisis 64 de la etiqueta de metadatos y, opcional pero preferentemente, una evaluación 66 de si la sección de audio seleccionada 62 en sí misma contiene material suficiente para proporcionar su inserción. Una determinación de que no hay material suficiente conduce a uno o más rechazos de la sección de audio, una decisión de repetir, es decir, un bucle, el segmento de audio con el fin de completar la totalidad de la duración del evento del mundo real, y/o asesorar al usuario a través de la interfaz de usuario (de, por ejemplo, la figura 2 cuando normalmente se realiza para incluir una interfaz gráfica de usuario "GUI"). El bucle puede implicar un bucle de intrasección reflejado por el prealmacenamiento de una etiqueta de metadatos. Este bucle intrasección permite el establecimiento de un bucle interior dentro de una sección y, por lo tanto, define puntos de corte que son internamente autoconsistentes.

La selección del segmento de audio puede dictarse por la entrada directa del usuario de una nota informativa en el sistema a través de una interfaz de control. La nota informativa indica unos parámetros variables que incluyen, pero sin limitarse a, duración, la naturaleza de los puntos de éxito dentro de una pista y el estado de ánimo de la pista. Como alternativa, puede presentarse al usuario múltiples opciones de seguimiento alternativas, teniendo cada una un tema común, y a continuación pedir indicar una preferencia. Un sistema totalmente automatizado es una alternativa.

Independientemente del proceso que realmente ejecuta un atenuación y/o un corte directo desde una pista de audio activa a una pista de destino (como se trata con particular atención en las figuras 6a a 6c), la inteligencia de montaje de fotogramas 52 se configura para evaluar si las secciones en diferentes pistas son fundamentalmente compatibles

y, preferentemente, compatibles en la medida en que puedan empalmarse o cortarse para producir una transición audiblemente perfecta entre las mismas. El proceso preferido se describirá ahora haciendo referencia a las figuras 7, 8 y 10.

5 La figura 7 es una representación espectral de diferentes secciones de música, teniendo las secciones unos puntos de salida y entrada determinados de acuerdo con una realización preferida de la presente invención. La figura 7, que muestra un proceso de punto de edición que puede ejecutarse de manera independiente para alinear los temas relacionados en secciones de audio seleccionadas, es un precursor para reorganizar las secciones de audio seleccionadas en un formato de datos consolidado que, finalmente, se alinea con los eventos de palabras reales para
10 aumentar la experiencia sensorial reforzando una contribución realizada a una etapa de sonido acompañante.

En la figura 7a, un espectro de audio 151 para una sección de audio ("sección de audio 1") donde el cambio de envolvente se simplifica mucho y se muestra como bloques con transiciones bruscas. El espectro 151, como se entenderá, tiene una representación de dominio de tiempo que varía con el tiempo. Una base de tiempo, en forma de
15 compases musicales ("Compás 1" a "Compás 8" para la sección 1), particiona la sección de audio entre un punto de entrada "abstracto" nominalmente identificado 152 y un punto de salida "abstracto" nominalmente identificado 154. Por lo tanto, el punto de entrada abstracto 152 y el punto de salida abstracto 154 definen la sección de audio, con estos puntos abstractos determinados y seleccionados por una función MIR y/o entrada del usuario para identificar una transición caracterizable entre diferentes temas dentro de la totalidad de, por ejemplo, una canción completa. Por
20 lo tanto, los "temas" son aspectos contextualmente diferentes con una pista de audio o un compuesto montado. Los temas pueden clasificar el contexto de una sección en términos de similitudes o diferencias percibidas/identificadas. Por lo tanto, los temas se codifican dentro de una base de datos relacionada con las secciones de la pista y pueden codificarse directamente en metadatos atribuidos.

25 Desafortunadamente, se ha reconocido que, debido a la naturaleza del audio, el punto de entrada abstracto 152 y el punto de salida abstracto 154 probablemente no coincidan con los puntos de salida 162, 164 y los puntos de entrada 166, 168 efectivos, es decir, reales, para una sección de audio. Por ejemplo, dentro de la sección 1 (número de referencia 160 de la figura 7a), puede haber uno o más puntos de salida abstractos posibles 154, 155, teniendo cada punto de salida abstracto 154, 155 uno o más puntos de salida reales 162, 164. En este contexto a modo de ejemplo,
30 puede identificarse una transición temática que se produzca en la transición entre el Compás 4 y el Compás 5, mientras que los puntos de salida reales 162, 164 (que tienen en cuenta una base de tiempo y pulso) podrían producirse en diferentes momentos dentro del Compás 4.

Antes de poder hacer la transición entre secciones, por lo tanto, es necesario buscar la compatibilidad entre las
35 diferentes secciones y, asimismo, las transiciones de tiempo que proporcionan una transición audible sustancialmente perfecta. En la primera instancia, las etiquetas de metadatos identificadas y aplicadas pueden proporcionar una indicación de una correlación aceptable entre diferentes temas. Por ejemplo, suponiendo una escala numérica del tema entre uno y ocho, una primera sección de audio puede tener una caracterización de un tema que podría ser representativa de un escenario lento y melancólico. Por el contrario, una segunda sección de audio puede tener una
40 caracterización de dos temas que podría ser representativa de una atmósfera ligeramente más rápida (en relación con la primera sección de audio) y contextualmente más intensa. Una tercera sección de audio puede tener una percepción muy energética y, por lo tanto, se caracteriza por tener una intensidad (es decir, un nivel definido por el usuario) escalada en una caracterización de nivel ocho de N (donde N es un número entero). En este ejemplo, una transición entre un tema de la primera sección a otro tema en una sección de música diferente parece eminentemente alcanzable desde
45 la perspectiva de una transición audible perfecta. También puede ser posible realizar una transición perfectamente entre la primera sección y la segunda sección dada la aparente cercanía en las caracterizaciones. Sin embargo, en realidad, es posible que no sea factible realizar una transición desde la primera sección a la tercera sección de audio basándose solo en una evaluación de la caracterización del tema subyacente a las secciones de audio respectivas. La realización preferida, sin embargo, proporciona una solución a este problema de transición audible perfecta.

50 Por lo tanto, es necesario, de acuerdo con la presente invención, investigar adicionalmente la naturaleza de los puntos de entrada y salida reales entre una sección de audio activa y una sección de audio de destino propuesta que se va a cortar en el audio. Los puntos de salida reales se seleccionan a partir de al menos uno y, por lo general, a partir de una pluralidad de puntos de salida adecuados que se unen en una anacrusa identificada. Se identifican los puntos de
55 entrada y salida adecuados (y se almacenan en una base de datos) para cada sección de audio antes de la edición, de tal manera que el entorno de procesamiento inteligente 52 requiere un almacenamiento en búfer mínimo para montar un audio editado que sea complementario a un flujo resoluble de estímulos de eventos externos. Por lo tanto, la transición entre audios puede lograrse sustancialmente en tiempo real, sin pérdida perceptible de continuidad de audio, independientemente de si las secciones se seleccionan dinámicamente a medida que se despliegan los eventos del mundo real.
60

La figura 7b es una vista ampliada del Compás 4 de la figura 7a. De nuevo, por el bien de la claridad, el espectro de audio del Compás 4 se muestra como una forma de bloque simplificada que tiene oscilaciones de señal que varían con el tiempo. El compás 4 se divide en cuatro pulsos ("Pulso 1" a "Pulso 4"), con cada pulso dividido además en un
65 número igual de fracciones (f_1 a f_4), en este caso cuatro fracciones iguales por pulso. Con respecto al punto de salida abstracto 155, los puntos de salida posibles reales corresponderán a un punto en o inmediatamente antes o después

de una anacrusa (alternativa e indistintamente denominada como "pastilla" o "inicio"), como se muestra en la figura 7b en relación con "Posible salida real B 164". Una anacrusa es una nota o secuencia de notas que precede al primer pulso en un compás de música. Por lo tanto, a través del MIR y/o una entrada de usuario, las anacrusas para los puntos de salida reales se identifican próximas al punto de salida abstracto 155. En el caso a modo de ejemplo de la figura 7b, la anacrusa precede al punto de salida abstracto por: i) en el caso de un primer posible punto de salida 162, una distancia de menos de un pulso y cero fracciones; y ii) en el caso de un segundo posible punto de salida 164, una distancia de cero pulsos y menos dos fracciones.

Debe realizarse una evaluación similar para identificar un punto de entrada en una sección de audio. A continuación, se hace referencia a la figura 7c. Los puntos de salida reales se seleccionan a partir de al menos uno y, por lo general, a partir de una pluralidad de puntos de salida adecuados que se unen en una anacrusa identificada. De nuevo, por el bien de la claridad, un espectro de audio de Compás 0 y Compás 1 se muestra como una forma de bloque simplificado en el que las oscilaciones de señal varían con el tiempo. El Compás 0 y el Compás 1 se han dividido en cuatro pulsos ("Pulso 1" a "Pulso 4"), con cada pulso dividido además en un número igual de fracciones (f_1 a f_4), en este caso cuatro fracciones iguales por pulso. Con respecto al punto de entrada abstracto 152, los posibles puntos de entrada reales 166, 168 corresponderán a un punto en o inmediatamente antes o después de una anacrusa. Por lo tanto, a través del MIR y/o una entrada de usuario, las anacrusas para los puntos de entrada reales 166, 168 se identifican próximas al punto de entrada abstracto 152. En el caso a modo de ejemplo de la figura 7c, la anacrusa se extiende sobre el punto de entrada abstracto por: i) en el caso de un primer posible punto de entrada 166, una distancia de más cero pulsos y cero fracciones; y ii) en el caso de un segundo posible punto de entrada 168, una distancia de menos un pulso y menos cero fracciones.

Como resumen del proceso para identificar localizaciones precisas y reales de puntos de entrada y salida dentro de una sección de audio, la realización preferida particiona las secciones de audio (identificadas en términos de temas) en unidades de tiempo recurrentes, por ejemplo, pulsos rítmicos (que pueden tener una duración variable). Cada pulso puede tener su propio "tempo", entendiéndose que el término "tempo" significa el número de unidades de tiempo por minuto. A continuación, estos pulsos se dividen o se "fraccionan" en al menos una pluralidad de divisiones de tiempo iguales y separadas uniformemente (denominadas "fracciones"). El número de fracciones dentro de los pulsos adyacentes puede variar. La duración de las fracciones entre unidades de tiempo adyacentes ("pulsos") puede, sin embargo, ser diferente debido a que cada unidad de tiempo tiene potencialmente su propio tiempo. Por lo tanto, con respecto a la primera base de tiempo relacionada con los pulsos y una segunda medida cuántica fraccional en una segunda base de tiempo, una localización aproximada de un punto de entrada/salida potencialmente adecuado pero significativo, es decir, una anacrusa, se determina en relación con un punto de inicio/fin de una sección identificada. Es este desplazamiento medido en el tiempo, expresado en términos de pulsos y fracciones, para la anacrusa lo que permite la integración perfecta de una sección de audio en otra sección de audio diferente. Más específicamente, para realizar una transición audiblemente perfecta, la inteligencia de procesamiento del sistema busca un punto de salida y un punto de entrada, expresado en términos de pulsos y fracciones, que tenga el mismo desplazamiento medido (ya que se expresa en términos de pulsos y fracciones).

Una base de datos 200, como se muestra en la figura 10, se monta y por lo tanto se pone a disposición de un dispositivo que opera dentro del sistema de la figura 2. La base de datos 200, por lo tanto, correlaciona pistas de música con metadatos de pista relacionados. Más especialmente, la base de datos asigna las secciones de audio 202 a los temas identificados 204 de las pistas 206 y cómo se definen estos temas con respecto a una sincronización definida 208 dentro de la sección. Por último, cada tema 202 se divide en al menos uno y, en general, una multiplicidad de puntos de entrada 210 y puntos de salida 212 dignos de edición que se relacionan con eventos de anacrusa (expresados en términos de pulsos y fracciones).

Volviendo a las figuras 7b y 7c, el corte entre secciones se realizaría, por lo tanto, desde la sección de audio activa en el Compás 4 y al comienzo del Pulso 4 (número de referencia 162) hasta la sección (X-1), Compás cero, posición de Pulso 4 (que tiene un desplazamiento de anacrusa de un pulso, cero fracciones) para el punto de entrada 168. Por supuesto, se producirán otras combinaciones de desplazamientos de pulsos y fracciones para la anacrusa, con estos accionados por el audio. El almacenamiento en búfer se usa para compensar el retraso o el avance de sincronización, según sea necesario, para alinear el corte entre secciones diferentes.

El objetivo de la figura 7 es, por lo tanto, soportar una función de edición efectiva y, más especialmente, permitir que se identifiquen los puntos de edición en relación con las distancias cuantificadas específicas de las transiciones de tipo de compás, relacionadas con cualquier tipo de compás, y los pulsos y fracciones de un pulso dentro del ritmo del audio.

La base de datos de la figura 10 almacena, como mínimo, los puntos de entrada y salida haciendo referencias cruzadas para las secciones de audio, necesitando estos tener la sincronización de alineación correcta si se van a empalmar perfectamente. De hecho, los criterios de coincidencia deben cumplirse antes de calcular los puntos de edición. Los pulsos se calculan a partir de la detección de inicio (detección de pulsos), las fracciones se calculan dividiendo equitativamente el tiempo entre los pulsos de "entrada/salida real" y la "entrada/salida abstracta" no coinciden con las "posiciones de fracción abstracta" y los "pulsos reales detectados"; no existe tal relación.

La figura 8 es un diagrama de sincronización 250 que muestra la sincronización relativa entre diferentes tipos de compás 252, 254 en diferentes secciones de, por ejemplo, una pista de música. Por ejemplo, el tipo de compás en una primera sección de audio puede ser un tiempo de 6/8 compuesto, mientras que la sincronización en una segunda sección de audio puede ser un tiempo de 4/4 simple. Debido a la discrepancia entre los tipos de compás, es necesario que el sistema de una realización preferida evalúe si, de hecho, corresponde realmente un desplazamiento registrado (expresado en términos de pulsos y fracciones) entre los puntos de salida y entrada. En otras palabras, el corte entre secciones debe tener en cuenta los tipos de compás variables entre secciones.

En una realización preferida, esta evaluación se logra estableciendo una tercera base de tiempo de pulsos que tienen una longitud que depende de un múltiplo común más bajo de fracciones dentro de los compases respectivos para diferentes secciones, con los compases de las secciones respectivas particionados a continuación en un número igual de pulsos de longitud fija. El número de pulsos en una fracción puede, por lo tanto, diferir entre fracciones de diferentes secciones. Por lo tanto, el sistema aplica un coeficiente para alinear diferentes tipos de compás, siendo este coeficiente una relación entre pulsos dentro de las diferentes secciones.

En el ejemplo de la figura 8, en el tiempo de 6/8 compuesto habría seis fracciones entre pulsos adyacentes. En el tiempo de 4/4 simple, habría cuatro fracciones entre pulsos adyacentes. Por lo tanto, un múltiplo común más bajo resultante para el producto proporciona un conteo de pulsos de seis. En consecuencia, siguiendo la correlación, un coeficiente de 1,5 identifica un punto de corte (relacionado con una anacrusa adecuada) para estar presente en el Compás X, pulso 1, corchea 2, fracción 2 del tiempo compuesto (en este ejemplo), pero en el Compás Y, pulso 3, fracción 3 del tiempo simple.

El coeficiente de multiplicación resulta en la anacrusa de tiempo en un tipo de compás que puede intercambiarse con otros, que pueden ser complejos.

Expresamente la figura 8 técnicamente pero desde una perspectiva musical:

el tipo de compás simple es 4/4/(16). La notación 4/4 es convencional, mientras que el 16 significa fracciones de semicorchea; por lo tanto, están presentes cuatro fracciones de semicorchea en un pulso de negra. La salida real para el tipo de compás simple del 4/4/16 es menos un pulso y menos dos 2 fracciones desde el final del compás. El final del compás, en este caso, corresponde a la salida abstracta.

El tipo de compás compuesto es 6/8/(16). El 6/8 significa que hay 2 pulsos de una negra con puntillo, cada uno en un compás. Esto significa que hay tres corcheas en cada pulso, en comparación con el tipo de compás simple en el que solo hay corcheas. Análogamente, hay seis fracciones de 16 en comparación con cuatro fracciones de 16.

En este ejemplo de la figura 8, esto significa que para editar a mitad de un pulso, se necesita multiplicar por 1,5 el número de fracciones para realizar la edición equivalente en un compás de 6/8 da partir de un compás de 4/4. Por lo tanto, para alinear el punto de edición entre el pulso, la salida de tipo de compás simple se multiplica por 1,5, o viceversa, el componente más bajo del tipo de compás compuesto se divide por 1,5. Esto significa que una cuadrícula de 24 pulsos, de los cuales unos 16 en el tiempo simple son iguales a tres pulsos, y los 16 en el tiempo compuesto son iguales a dos pulsos. Por lo tanto, el sistema puede medir todas las entradas y salidas en tales pulsos y multiplicar el número por el coeficiente dado relacionado con el tipo de compás. Se observa que los 24 pulsos son válidos en este ejemplo, pero puede ser diferente en otras situaciones con otros tipos de compás. En la práctica, la inteligencia de sistema dice: "Tengo una salida de tiempo simple en menos un pulso y menos dos fracciones, por lo tanto, $-4 -2 = -6$ fracciones de 16. Esto da dieciocho pulsos y proporciona tres pulsos por fracción. Para un tiempo complejo, la salida precalculada se produce a menos un pulso, menos una corchea, menos una fracción. Por lo tanto, $-6 -2 -1 = -9$ fracciones de 16, dando estos dieciocho pulsos y proporcionando tres pulsos por fracción, *pero dividido por el coeficiente compuesto de 1,5*. Ya que ambos están ahora normalizados a una longitud de dieciocho pulsos, existe una salida a esta entrada.

La inteligencia de montaje de secciones 52 está configurada para emprender un análisis de cualquier entrada de nota informativa suministrada, cargado o de otro modo introducido por el usuario a través de una interfaz de usuario. La inteligencia de montaje de secciones 52 opera para identificar y calcular segmentos de música necesarios para cumplir con la nota informativa, con la identificación basada en los metadatos relacionados con la sección. La nota informativa puede actualizarse dinámicamente durante la salida de una pista para reflejar los cambios en los requisitos notificados por el usuario.

A continuación, las secciones de audio disponibles 62 se ponen a disposición 68 para la secuenciación completa en un secuenciador 70 que, preferentemente, se basa en la lógica de IA. El secuenciador 70, que normalmente se implementa mediante programación, está configurado para seleccionar y alinear segmentos de audio apropiados con los eventos correspondientes. La secuencia se basa en la correlación de la etiqueta colocada en los metadatos de la sección de audio con el elemento informativo, es decir, la palabra de vocabulario de Mashtraxx, asignada a cada parte del evento externo a aumentar de manera audible.

Funcionalmente, el secuenciador 70 está configurado o dispuesto para identificar y hacer coincidir el intervalo o función de música 72; esto es necesario para una transición perfecta entre las secciones de audio. Adicionalmente, el

secuenciador 70 opera para hacer coincidir los puntos de entrada y salida 74 entre diferentes temas identificados, en el que se reproduce una sección de audio de un primer tema y una sección de audio de un segundo tema debe cortarse en el primer tema a medida que el primer tema se atenúa (rápidamente). De nuevo, esto es necesario para una transición perfecta entre las secciones de audio. Por último, el secuenciador 70 coincide con la sincronización musical 5 76. Estas funcionalidades pueden basarse en algoritmos "MIR" de recuperación de información musical (tal es como los indicados anteriormente), aunque las ineficiencias o el análisis automatizado erróneo en las evaluaciones MIR se abordan mediante el suministro de una interfaz que proporciona al usuario la capacidad de introducir, sobrescribir o definir etiquetas aplicadas en los metadatos de sección de audio para definir el tema de la sección de audio/música. De nuevo, esto es necesario para una transición perfecta entre las secciones de audio. La definición de un segmento, 10 como se ha indicado anteriormente, se realiza en términos de propiedades percibidas para un oyente/usuario y/o parámetros cuantitativos y/o cualitativos (medidos a través del software MIR) del audio dentro de la sección específica.

El "tema" puede representarse en una forma codificada en binario o en un lenguaje natural basado en palabras que incluye más o menos información que el ejemplo proporcionado inmediatamente antes. Por ejemplo, la heurística 15 definida por el usuario puede usarse para proporcionar una mayor granularidad y definición para la nota informativa y, por lo tanto, una selección y un montaje más dirigidos de los segmentos de audio. La heurística puede relacionarse, por ejemplo, con i) la secuencia de la canción y, por lo tanto, el grado de reordenamiento de la salida de audio resultante en relación con el orden de los versos y los coros en la canción o banda sonora original, ii) la singularidad del segmento, por ejemplo, una indicación sobre la cantidad o permisibilidad de segmentos repetidos, y iii) los puntos 20 de bucle que establecen un punto para un "relleno" necesario para alargar un segmento de audio específico.

En el límite, la etiqueta de metadatos para el segmento de audio podría ser simplemente una palabra que se correlaciona con el vocabulario Mash-traxx usado para definir la caracterización del evento externo actual y en curso 12.

Un generador de audio 80 está acoplado al secuenciador 70. El generador de audio 80 está configurado para gestionar la atenuación cruzada 82 entre un segmento de audio activo que se está reproduciendo actualmente y un segmento de destino que es el siguiente segmento de audio programado y seleccionado que se reproducirá cuando cambie el 30 evento externo. Esta función incluye la identificación de una transición apropiada asociada con una anacrusa (alternativa e intercambiablemente denominada como "pastilla" o "inicio"), Por ejemplo, una nota o secuencia de notas que precede al primer pulso en un compás de música. En términos de las realizaciones de la invención reivindicada actualmente, el inicio, tal como un pulso, es importante porque puede discernirse su omisión o falta de claridad y es representativo de una interrupción no natural en el audio, mientras que su presencia proporciona continuidad y flujo. Por lo tanto, las realizaciones preferidas se empalman entre segmentos de audio actuales y segmentos de audio 35 subsiguientes en una anacrusa inmediatamente antes de un inicio temprano en la pista actualmente activa o en la pista de destino, permitiendo de este modo que diferentes segmentos de audio que transmiten diferentes temas se hagan coincidir entre sí. Por lo tanto, el sistema opera para fijar y documentar una anacrusa en un punto apropiado en un segmento de audio.

El generador de audio 80 incluye preferentemente una función de almacenamiento en búfer 84 para suavizar las transmisiones de audio y minimizar los retrasos. Como su salida 86, el generador de audio 80 puede transmitir 88 el segmento de audio montado o enviar el segmento de audio a un archivo 90.

La figura 4 es un diagrama funcional para el montaje de pistas y la creación de metadatos de acuerdo con un proceso de creación preferido. La figura 4 representa, por lo tanto, los procesos aplicados, por ejemplo, a una pista musical original 56 para crear un archivo multimedia compuesto (número de referencia 54 de la figura 1) que contiene 45 metadatos de Mashtraxx para la pista musical original 56.

Para una pista de audio cargada o puesta a disposición del sistema de la figura 3, la pista de audio está sujeta a algoritmos MIR basados en AI, como se ha indicado anteriormente. A través de una disposición preferida de una GUI, pueden introducirse una revisión humana de la composición espectral de la pista de audio, un tipo o tipos de compás 102, incluyendo el tempo y otros eventos musicales importantes para definir o refinar (más) los temas y los segmentos de audio. Este proceso basado en humanos, aunque simplemente preferido, puede abordar por lo tanto las deficiencias o errores asociados con la interpretación automática de la caracterización de la pista de audio. Por ejemplo, El punto 50 de inicio de una envolvente de señal deseada para un segmento de audio puede ajustarse manualmente a través de una GUI con el fin de alinear con mayor precisión el inicio del muestreo de datos de audio con el comienzo de un evento musical específico (y, por lo tanto, un tema identificado relevante para el evento externo que debe aumentarse por el proceso de la presente invención).

El análisis del tipo o tipos de compás proporciona una determinación de compases y pulsos y, asimismo, una variación de estos a medida que se producen entre las muestras que componen las secciones de audio seleccionadas. Este análisis proporciona la base para el mecanismo descrito (de la realización preferida) para la integración perfecta de una sección en otra, independientemente de los diferentes tipos de compás y anacrusas.

Adicionalmente, dentro de cada sección de audio, el sistema está configurado para detectar 104 inicios en diferentes puntos. El corte entre una fuente, es decir, un corte de reproducción activo y un corte de destino, es decir, el segmento

de audio que se ha seleccionado para ser el próximo reproducido con el fin de cambiar el tema de audio, se describirá con más detalle a continuación y haciendo referencia a la figura 5, mientras que la figura 9 muestra un proceso preferido por el que puede determinarse un inicio.

5 La determinación del inicio explicada con respecto a la figura 9 puede implementarse independientemente en otras realizaciones y aspectos como se describe en el presente documento, aunque su inclusión en el sistema proporciona una solución integral. Un inicio musical es el momento en el que comienza un evento musical (u otro evento audible significativo).

10 En la figura 9a, se muestra que una señal de audio típica 300 comprende unos componentes 302, 304 de amplitud variable. Para la detección del inicio, un enfoque preferido separa en primer lugar la parte transitoria (es decir, los aspectos que evolucionan rápidamente de la señal de audio) de la parte estacionaria (es decir, los aspectos estables de la señal de audio) de la señal de audio 300; esto produce una representación de dominio de tiempo 310 como se muestra en la figura 9b. Los transitorios se asocian más normalmente con frecuencias más altas y donde estas
15 frecuencias más altas demuestran aumentos significativos en la energía. A continuación, se determina una envoltura de amplitud 320 de la parte transitoria; esto se basa normalmente en una función de convolución como lo entenderá un experto en el procesamiento de señales de audio.

20 Posteriormente, en relación con un umbral de potencia 330, se detectan picos en la envoltura de amplitud 320. Los picos corresponden a los puntos máximos de la envolvente en una región localizada. A continuación, para alcanzar la posición exacta del inicio, el sistema está configurado para funcionar a través de la señal transitoria desde el pico hacia atrás en el tiempo para encontrar el punto X en el que una característica de un período de la señal antes de ese punto X es diferente de forma máxima con respecto a la misma característica de un período de la señal después de ese punto X. Las características incluyen, pero sin limitación, la desviación convencional de la primera derivada de la
25 señal transitoria antes y después del punto X. En otras palabras, la relación más grande en las potencias adyacentes, como se muestra en toda la envolvente de amplitud transitoria, proporciona el punto más apropiado para el inicio; esto se ilustra en la figura 9d. Esto puede expresarse de nuevo de manera diferente, un comienzo de inicio se detecta buscando preferentemente identificar el punto donde la desviación convencional de la tasa de cambio de una característica identificada durante un período anterior a ese punto de muestreo es diferente de forma máxima para la
30 misma característica después de ese punto.

Los períodos de muestreo pueden estar en la región de aproximadamente diez milisegundos a aproximadamente cincuenta milisegundos. Como ejemplo de un inicio, es probable que un estruendo de címbalo represente un inicio ya que es probable que tenga un cambio máximo en los transitorios en el punto de su aparición en la señal de audio.

35 Volviendo nuevamente a la figura 4 y el proceso para generar metadatos de pistas, el análisis automatizado adicional de pistas (que nuevamente pueden revisarse a la luz de la entrada manual) funciona para resolver la posición de compás/pulso 106. Dada la relación entre el nivel y por lo tanto el inicio, La detección de compás y pulso puede obtenerse a partir de o basarse en la detección de inicio 104.

40 Con respecto al procesamiento de señal en curso dentro del sistema, el análisis de tipos de compás 102, la detección de inicio 104 y la detección de compás/pulso 106 (así como cualquier entrada supervisada de un usuario 110) se usa para cortar o segmentar la pista de audio en temas, de duración variable, que se correlacionan con el vocabulario de Mashtraxx aplicado para clasificar la naturaleza del evento externo y, por lo tanto, la naturaleza de la mejora de audio.
45 A este respecto, una metaetiqueta 112, preferentemente realizada usando lógica de AI, compila una pista de audio mejorada 58 que incluye etiquetas de metadatos para las secciones de audio seleccionadas. De acuerdo con lo indicado o inferido, estos metadatos identifican al menos la función e intensidad musical, así como los puntos de salida y entrada para cortes y atenuaciones para los temas asociados con la sección. El etiquetado de metadatos significa que una trama de audio muestreado y entregado, es decir, un corte o segmento, puede hacerse que coincida con
50 precisión con la sincronización de un evento externo identificado 12, tal como una identificada mediante la superación del umbral de activación o la identificación de un tema de una palabra asignada del vocabulario de Mashtraxx.

Los tipos de compás o la generación de la línea de tiempo permiten que las secciones de pista se describan en notación musical, lo que es esencial para la correcta coincidencia de las secciones de pista. La identificación precisa
55 del tipo de compás permite una línea de tiempo móvil que soporta el ajuste automático al pulso más cercano.

El sistema genera una línea de tiempo creada preferentemente en pulsos y compases para una pista de audio digital, garantizando que se mantenga una notación musicalmente fiel para el proceso de edición. Es esto lo que sustenta el corte, el empalme y la remezcla exitosos del contenido de tal manera que no sea perjudicial para la integridad musical
60 de la pieza. La inteligencia de sistema permite la construcción de un enfoque flexible que puede adaptarse de acuerdo con la entrada de audio a modificar. Por ejemplo, Si se detecta una disminución en el tempo, la línea de tiempo puede ajustarse en consecuencia. Esto es importante en la deconstrucción de las formas musicales con el fin de remezclar y reeditar. Esto soporta una comprensión tangible de cómo el audio se relaciona musicalmente con la línea de tiempo y permite al usuario mantener un sentido real de la estructura musical dentro de un formato digital.

65 Por lo tanto, la creación de una sección de audio es el proceso de dividir una pista en secciones que pueden

reorganizarse para formar una nueva composición. Preferentemente, las secciones se ajustan automáticamente a los compases. Cada sección puede pertenecer a un grupo que puede definirse por el usuario, pero preferentemente predeterminado a uno de una introducción, un verso, un puente, un coro, Central 8 o fila de salida (de acuerdo con lo identificado por la metaetiqueta). Cada sección permite propiedades de metadatos personalizables, incluyendo pero no limitado a la intensidad, la función musical, los instrumentos, las voces y las etiquetas.

El sistema de la presente invención hace uso del software para mantener el análisis rítmico compositivo tradicional en términos de compases, pulsos y fracciones de los compases. Esto permite al usuario involucrarse en el corte y comprensión de las composiciones de compás y pulso de una manera tradicional. Esto mantendrá la intuición compositiva, haciendo que la deconstrucción y la reorganización de ideas sea musicalmente sencilla. El sistema también incorpora datos de pulso, que explotan todo el potencial del audio digital, permitiendo que la fracción compositiva más pequeña se corte en pulsos adicionales. La función principal de los pulsos es ofrecer la granularidad más pequeña necesaria para hacer frente a diferentes tipos de compás y coeficientes utilizados para cortar entre diferentes tipos de compás en diferentes secciones.

De acuerdo con la presente invención, los metadatos para una pista completa contienen, por lo tanto, un conjunto de descripciones de secciones de audio, así como una explicación de cómo pueden hacer la transición hacia o desde cualquier otra sección. Por lo tanto, cada sección de audio incluye metadatos integrales que permiten construir una secuencia de corte dada una breve descripción del usuario que se correlaciona con una caracterización de un evento externo. Opcionalmente, los metadatos pueden contener una referencia a la pista de audio original, permitiendo localizar la pista correcta a partir de una biblioteca de música en línea, tal como Spotify. Esto permite que los metadatos se almacenen independientemente de la pista original y que tanto los metadatos como la pista se transmitan o se extraigan por la inteligencia de montaje de secciones 52.

Por lo tanto, el marco de trabajo de metadatos permite etiquetar música de acuerdo con ciertas categorías definidas por el usuario. Por ejemplo, estas categorías podrían incluir, por ejemplo, género, estado anímico, estilo y tempo, y podrían usarse para crear una base de datos de sonido en la que pueda buscarse. En efecto, el sistema ofrece el potencial de buscar música no solo por título y artista, sino por el sonido, el género y el estilo musical. Por lo tanto, la configuración de la base de datos de la figura 10 permite la búsqueda de las metaetiquetas de las secciones individuales, y no solo en un enfoque global para la entrega de metadatos. Por ejemplo, se prefiere que el sistema de metaetiquetado pueda cumplir requisitos de información tales como: "Me gustaría una canción que sea optimista, animadamente feliz, también necesita tener un solo de trompeta en algún momento, y una sección para relajarse con fondos de cuerda en algún otro lugar". Los sistemas existentes no pueden proporcionar tal riqueza y flexibilidad de búsqueda.

Preferentemente, el esquema de metadatos no solo permite que los datos se atribuyan a toda la pista, sino que facilita la anotación fina de secciones específicas, proporcionando de este modo datos descriptivos de las características más destacadas de una pieza de música. Una o más (y normalmente múltiples) caracterizaciones musicales aplicadas son comunes a cada pieza de música y, por lo tanto, permiten que el mecanismo de IA edite un archivo de audio digital dependiendo de los requisitos de sincronización con los eventos entrantes identificados (del mundo real). Las metaetiquetas emocionales del audio son, por lo tanto, independientes de las caracterizaciones aplicadas al tema de cada sección. La búsqueda de diferentes estados de ánimo emocionales permitirá intercambiar diferentes secciones de audio para cumplir con los requisitos de sincronización. Por lo tanto, algunas características clave incluyen puntos de éxito musicales y descriptores de funciones relacionados con:

- Clímax: donde la música alcanza un pico
- Sección estática: la música no aumenta ni disminuye en la intensidad percibida del nivel contextual o real
- Abandono dramático: la música cambia repentinamente para volverse muy escasa
- Caída de atenuación: la música disminuye gradualmente en intensidad
- Subida: la música aumenta gradualmente en intensidad
- Poner de relieve: relleno de audio de bajo nivel bajo diálogo y actuación
- Género
- Tipo de instrumento

Cada sección de una pieza de música puede describirse en términos de su función de forma, es decir, cómo opera la sección dentro de la totalidad de la pieza, por ejemplo, poner de relieve, caída de atenuación y subida. Por lo tanto, independientemente de las metaetiquetas contextuales relacionadas emocionalmente aplicadas a una sección, si los datos de evento entrante pueden describirse en términos de las funciones de forma musicales descritas, puede usarse cualquier pieza de música y cumplirá los requisitos de sincronización independientemente de cualquier subjetividad emocional que surja con la caracterización de una pieza específica por parte del usuario. Esto es significativo para la funcionalidad del procesamiento descrito aplicado en las diversas realizaciones de la invención.

El metaetiquetado no necesita limitarse a la música, y podría aplicarse a otras fuentes de sonido (tales como patear una pelota de fútbol) para que un escenario de sonido pueda aumentarse con una sección de audio insertada que se etiqueta a un evento externo identificado.

La figura 5 es la representación dominio de tiempo 130 de una estrategia de corte preferida entre un punto de salida y un punto de entrada para diferentes temas musicales dentro de las muestras de pista.

Dentro de la figura 5 hay dos espectrales de dominio de tiempo: un primer segmento de audio activo "en reproducción" 132 y un segundo segmento de audio "de destino" 134 que se usa para introducir un cambio en el tema y el segundo segmento que debe cortarse, es decir, para reemplazar, el primer segmento activo como una nueva salida de audio correlacionada. La realización preferida logra esta atenuación/corte de una manera casi perfecta.

Ambos espectros de dominio de tiempo incluyen características comunes, tales como los inicios 136a-136f.

En términos de un tipo de compás o base de tiempo 150, esto se representa en el diagrama de líneas de puntos verticales que dividen la envolvente de audio en segmentos que pueden tener una duración correspondiente a una semicorchea, croché o algún otro sincronismo seleccionado. En la figura 5, una representación a modo de ejemplo proporciona cuatro fracciones de un pulso 3.1, 3.2, 3.3 y 3.4 que dividen la parte relevante de cada envolvente en cuatro sectores de tiempo iguales. Estas cuatro fracciones pueden, de hecho, representar un pulso musical o pulsos múltiples o algún otro período en la pista de audio.

Haciendo referencia a las figuras 6a, 6b y 6c, en términos de la salida del primer segmento de audio de reproducción activo 132 al segundo segmento de audio de destino 134, el proceso de atenuación/corte preferido en primer lugar identifica y a continuación contrasta los inicios relativos en el primer inicio activo de segmento de reproducción de audio A y el inicio del segundo segmento de audio de destino D , con la evaluación realizada en relación con la base de tiempo apropiada más cercana 150. Al tener dos inicios, el inicio A y en el inicio D reproducidos en estrecha proximidad de tiempo o un inicio esperado retrasado significativamente en el tiempo provocaría una notable discontinuidad en la fluidez del audio y la transición entre temas.

En consecuencia, cuando tiene lugar una transición temática, una realización preferida observa las diferencias de tiempo de muestreo (s_1 en relación con s_2 y s_3 en relación con s_4) entre el primer inicio (en cada uno de los segmentos activo y de destino) y selecciona la primera de las alternativas.

Independientemente de si el inicio se produce en la pista de reproducción activa o en la pista de destino, el inicio más temprano en el tiempo en relación con la base de tiempo 150 es el punto en el que hay una atenuación cruzada, y preferentemente un corte instantáneo 180, desde la sección de reproducción activa hasta la sección de destino. Para la atenuación cruzada, una atenuación cruzada típica puede tomar hasta aproximadamente cinco milisegundos. Se prefiere la atenuación cruzada a un corte instantáneo, ya que evita la posibilidad de una explosión de audio ya que es poco probable que la tensión en las dos señales coincida. A menos que un contexto específico requiera una interpretación más limitada, el término "corte" debería verse como un caso especial de atenuación cruzada que se produce más rápidamente.

En la figura 6a, el corte 180 desde la sección de audio activa 132 a la sección de destino 134 se activa simultáneamente con (o, si el almacenamiento en búfer está disponible, marginalmente antes) la muestra donde el inicio en la sección de audio activa 132 comienza a aumentar. La pista de audio de destino se reproduce y se convierte en la pista activa hasta que se requiera otro cambio de tema. En la figura 6b, el corte 180 desde la sección de audio activa 132 a la sección de destino 134 se activa simultáneamente con (o, si el almacenamiento en búfer está disponible, marginalmente antes) la muestra donde el inicio en la sección de audio de destino 132 comienza a aumentar. Habitualmente, las tasas de muestreo y las bases de tiempo operan en fracciones de segundo, por lo tanto, un corte absoluto (en lugar de un atenuación más lenta) es intrascendente y no se nota sustancialmente. Por lo tanto, el proceso permite la reconstrucción perfecta e interminable de una pieza de audio. Se prefiere el almacenamiento en búfer ya que proporciona el preprocesamiento.

En la figura 6c, se observa que el inicio en el audio de destino 134 se produce antes de la base de tiempo 150, mientras que en la sección de reproducción activa se detecta que el inicio se produce después de la base de tiempo. En la figura 6c, por lo tanto, el corte y la atenuación cruzada se sincroniza con el inicio detectado en la sección de destino 134, garantizando que el almacenamiento en búfer se mantenga en su sincronización.

Al separar los metadatos de la pista original y almacenar los metadatos en una base de datos segura, el sistema de la realización preferida mantiene opcionalmente la sustancia del contenido invisible, protegiendo de este modo los problemas de licencias y derechos de autor. Por lo tanto, el proceso de separación puede eliminar gran parte del acceso necesario para la compartición P2P (de igual a igual), ya que el contenido puede generarse dinámicamente cada vez que se necesite. De hecho, el sistema de la realización preferida puede proteger el contenido original, así como el nuevo contenido generado por la entrada de usuario y el proceso de sincronización.

Por lo tanto, el sistema y los procesos de la presente invención soportan la edición creativa a través de un proceso automatizado operable en tiempo real. Como tal, la música para juegos puede editarse dinámicamente para seguir la historia del juego: los altibajos, los logros y los fracasos. Ya que una experiencia en el juego varía invariablemente, esto significa que el marco de trabajo inteligente 52 (de la figura 2) puede operar para transmitir música validada a una clave de acceso para cada pista para eliminar la piratería para aquellos usuarios que no poseen una clave de acceso

válida. Adicionalmente, el uso del metaetiquetado de audio y la correlación con eventos del mundo real permite al usuario hacer uso de su música favorita, por ejemplo, un YouTube, un Video de Facebook y un Vimeo personales.

5 En resumen, La presente innovación proporciona un método para editar información de audio digital con características de composición similares a música que permiten la sincronización de música en tiempo real. El sistema proporciona detección de inicio, así como la identificación de un evento de anacrusa tanto en una sección de audio de salida como en una de entrada que proporciona una transición audible perfecta. Este sofisticado sistema usa DSP y algoritmos de AI especializados para analizar los datos de audio sin procesar y realizar manipulaciones editoriales que anteriormente exigían un conocimiento experto y profesionales de la música capacitados para su ejecución. Además, este sistema adaptativo segmenta los datos de audio, mientras que simultáneamente busca imbuir un conjunto de códigos de control que se ajustan fácilmente para permitir un cambio editorial. La presente invención sirve para transformar los parámetros musicalmente significativos de la información de audio digital. Si bien la forma de onda se ha convertido en la interfaz convencional de representación y navegación para la edición digital, descuida el análisis rítmico compositivo tradicional en términos de compases y pulsos. Esta información es esencial para la reorganización y mezcla de pistas de audio, y es integral para la sincronización automatizada. Asimismo, este sistema soporta la codificación de metadatos descriptivos adicionales en el formato de archivo relacionado con las texturas, trayectorias e intensidad de una pieza de música dada. Actualmente no existe un formato de metadatos para capturar las funciones narrativas de una pieza de música, y la presente proporcionará una forma estructurada para hacerlo y un sistema para convertir esta información de alto nivel en un formato de datos transferible de bajo nivel.

20 Desde la perspectiva de la funcionalidad del usuario final, el software está configurado para permitir a los usuarios finales proporcionar datos descriptivos que facilitarán la recuperación automática de una selección de pistas que se mezclan automáticamente, mediante la inteligencia de montaje de secciones 52, para ajustarse al escenario que requiere un acompañamiento musical. Los usuarios finales pueden proporcionar orientación a través de las interfaces de acceso descritas, orientación que puede basarse en el estado de ánimo que desean crear e incluso sugerir una narrativa que desean que la música logre. El sistema está configurado preferentemente para reducir la sobrecarga de información recuperando solo los elementos que se estiman relevantes para el usuario, basándose en una entrada simplista del usuario, tal como la anotación básica de un dato de película o información sobre el uso deseado de la música. El sistema está configurado para seleccionar pistas relevantes basándose en especificaciones globales y de intrapista decididas por un analista o anotador, tal como una entrada y correlacionadas con el evento externo a través del lenguaje natural o una palabra de vocabulario de Mashtraxx. Ejemplos de esto serían el estado de ánimo, el ritmo de paso/velocidad general de la música, los instrumentos clave (si tienen alguno en mente) y género.

35 A menos que las disposiciones específicas sean recíprocamente excluyentes entre sí, las diversas realizaciones descritas en el presente documento pueden combinarse para mejorar la funcionalidad del sistema y/o para producir funciones complementarias en la entrega efectiva del audio sincronizado relevante para los sentidos. Dichas combinaciones serán apreciadas fácilmente por los expertos en la materia proporcionada la totalidad de la descripción anterior. Análogamente, los aspectos de las realizaciones preferidas pueden implementarse en disposiciones independientes donde son apropiadas disposiciones funcionales más limitadas. De hecho, se entenderá que a menos que las características en las realizaciones preferidas específicas se identifiquen expresamente como incompatibles entre sí o el contexto circundante implique que son recíprocamente excluyentes y no fácilmente combinables en un sentido complementario y/o de apoyo, la totalidad de la presente divulgación contempla e imagina que las características específicas de esas realizaciones complementarias pueden combinarse selectivamente para proporcionar una o más soluciones técnicas exhaustivas, pero ligeramente diferentes.

45 La presente invención puede proporcionarse en una forma descargable o de otro modo en un medio legible por ordenador, tal como un CD ROM, que contiene el código de programa que, cuando se instancia, ejecuta la funcionalidad de incrustación de enlaces en un servidor web o similar. Por supuesto, se apreciará, que la descripción anterior se ha dado solo a modo de ejemplo y que pueden realizarse modificaciones en detalle dentro del alcance de las reivindicaciones.

50 Por ejemplo, mientras que la realización preferida se ha descrito en el contexto de un entorno de ejercicio, la tecnología también puede aplicarse para aumentar los sonidos, tales como los perfiles de sonido identificados asociados con golpear una pelota de tenis. Por lo tanto, el sistema de la presente invención puede organizarse para aumentar la experiencia ambiental con partidos deportivos en vivo sustituyendo una muestra de audio en la mezcla de sonido en el punto en el que se identifica que la pelota de tenis está golpeándose.

60 El método y el sistema de identificación de los inicios, como se describe especialmente con respecto a las figuras 9a a 9d, puede, de hecho, encontrar más y más aplicaciones más allá de la sincronización de datos de audio para aumentar los eventos monitorizados. Por lo tanto, el método y el sistema pueden aplicarse de manera más general para identificar los inicios en cualquier archivo de contenido de música o audio, independientemente de si la música se clasifica o no en términos de un tema contextual (ya sea detectado automáticamente o establecido a través de un usuario o una crítica experta).

65 Aunque las realizaciones preferidas tratan la adaptación del audio a eventos externos, lo contrario es cierto porque los metadatos de audio pueden activar o configurar eventos externos. Por ejemplo, un evento externo puede ser una

5 transición de video que se produce entre escenas contiguas en una película o una nueva fase dentro de un nivel de un juego informático (tal como cuando llegan nuevos personajes o cuando se adquieren nuevas habilidades y se presentan por primera vez al jugador durante el juego). Basándose en una comprensión de (por ejemplo) uno o más de entre las intensidades, las secciones, los compases, los pulsos y las fracciones de un pulso determinados en el contexto de las realizaciones preferidas, los metadatos de audio obtenidos a partir de las diversas realizaciones preferidas pueden usarse para accionar un cambio en la propiedad temporal observada de los eventos tal como se experimentan, por ejemplo, luces intermitentes o accesorios que se mueven dentro de una actuación en vivo o dentro de un entorno de CGI o un video grabado.

REIVINDICACIONES

1. Un sistema de aumento de contenido de medios que comprende:

5 una base de datos que contiene una multiplicidad de archivos de audio, conteniendo cada uno una o más secciones de audio y metaetiquetas asociadas que identifican temas contextuales asociados con cada sección de audio, identificando adicionalmente cada metaetiqueta al menos un punto de entrada abstracto y al menos un punto de salida abstracto que soportan transiciones de corte o atenuación perfectas entre diferentes secciones de audio y en donde cada uno de dicho al menos un punto de entrada abstracto y dicho al menos un punto de salida abstracto se relaciona con un inicio asociado con una anacrusa para la sección de audio y en donde la anacrusa se expresa en términos de pulsos y fracciones y cada anacrusa es una nota o secuencia de notas que precede al primer pulso en un compás de música en cada sección de audio;

10 un sistema de procesamiento acoplado a la base de datos y sensible a dichas metaetiquetas; y una entrada acoplada al sistema de procesamiento, la entrada receptiva a una entrada de datos (12) que define un evento variable temporalmente en forma de un evento físico monitorizado o detectado o, como alternativa, en forma de una sucesión de fotogramas de imagen y en donde el evento variable temporalmente incluye transiciones identificables entre actividades discerniblemente diferentes;

15 en donde el sistema de procesamiento está dispuesto:

20 para resolver cada una de dichas actividades diferentes en uno o más de una pluralidad de diferentes temas contextuales categorizados; y **caracterizado por**, basándose en una correlación identificada entre cada uno de dichos temas contextuales categorizados relacionados con la entrada de datos y metaetiquetas de dichas secciones de audio, seleccionar una primera sección de audio y una sección de audio de destino diferente que reflejan las diferentes actividades en el evento variable temporalmente y correlacionarlas con las mismas, y, después de eso, seleccionar anacrusas que son puntos de posible salida real y de posible entrada real para cortar o atenuar perfectamente entre sí dichas secciones de audio primera y de destino seleccionadas para alinearse con temas contextuales categorizados correspondientes de las diferentes actividades a medida que la entrada de datos varía con el tiempo, en donde:

25 dichas anacrusas seleccionadas tienen unos desplazamientos medidos correspondientes en el tiempo a unos puntos de salida abstractos y de entrada abstractos respectivamente en la primera sección de audio y la sección de audio de destino entre las que va a producirse el corte o la atenuación, dichas anacrusas seleccionadas en la primera sección de audio y la sección de audio de destino se alinean de tal manera que se produzcan en un momento común,

30 las diferencias de tiempo medidas de inicio en cada una de la primera sección de audio seleccionada y la sección de audio de destino seleccionada a partir de las anacrusas seleccionadas crean posiciones alternativas de atenuación cruzada o corte para las transiciones entre la primera sección de audio seleccionada y la sección de audio de destino seleccionada,

35 la posición de atenuación cruzada o corte se selecciona basándose en qué posición respectiva de atenuación cruzada o corte en la primera sección de audio y la segunda sección de audio de destino se evalúa que se produce más temprano en el tiempo con respecto a las anacrusas alineadas y en donde la primera posición de atenuación cruzada o corte tiene un inicio asociado, y la producción de una atenuación cruzada o corte entre la primera sección de audio y la sección de audio de destino en un punto en o antes de que dicho inicio asociado comience a aumentar para realizar una transición rítmica ininterrumpida en la salida de audio preservando la integridad de la sincronización musical entre la primera sección de audio y la sección de audio de destino; y

40 generar, como una salida, un archivo de medios o una salida en tiempo real en la que las transiciones entre diferentes secciones de audio se alinean correspondientemente con transiciones relacionadas entre diferentes actividades.

45

50

2. El sistema de aumento de contenido de medios de la reivindicación 1, en donde dicho evento variable temporalmente está relacionado con uno de:

55 un comienzo y un final en el caso de una característica sostenida para el tema contextual, en donde la característica sostenida es uno de una función de forma y un punto de éxito a lo largo del tiempo; y un momento específico para un punto de éxito.

60 3. El sistema de aumento de contenido de medios de acuerdo con cualquier reivindicación anterior, en donde el sistema de procesamiento selecciona la sección de audio sustancialmente en tiempo real para acompañar el evento físico monitorizado o detectado o la sucesión de fotogramas de imagen.

65 4. El sistema de contenido de medios de acuerdo con cualquier reivindicación anterior, en donde el sistema está dispuesto para aumentar una experiencia sensorial general de un usuario en el mundo real cambiando, reordenando o repitiendo dinámicamente y a continuación reproduciendo secciones de audio seleccionadas dentro del contexto de

dichos temas categorizados asociados con la naturaleza cambiante de los eventos presentados como la entrada de datos.

5 El sistema de aumento de contenido de medios de acuerdo con cualquier reivindicación anterior, en donde el sistema de procesamiento es un conjunto distribuido de sistemas de procesamiento.

6. El sistema de aumento de contenido de medios de acuerdo con la reivindicación 2, en donde la función de forma se selecciona a partir de uno del grupo que consiste en:

- 10 una sección estática;
una caída;
una atenuación en intensidad;
una subida de intensidad; y
15 una intensidad relativa de la sección.

7. El sistema de aumento de contenido de medios de acuerdo con la reivindicación 2, en donde el punto de éxito a lo largo del tiempo se selecciona a partir de uno del grupo que consiste en:

- 20 un clímax en donde la música alcanza un pico;
una sección estática en donde la música no aumenta ni disminuye en la intensidad percibida del nivel contextual o real;
una caída dramática: la música cambia repentinamente para volverse muy escasa;
una subida en donde la música aumenta gradualmente en intensidad;
una caída de atenuación en la que la música disminuye gradualmente en intensidad;
25 una puesta en relieve en la que hay un relleno de audio de bajo nivel bajo diálogo y actuación;
género; y
tipo de instrumento.

8. El sistema de aumento de contenido de medios de acuerdo con cualquier reivindicación anterior, en donde la entrada adopta la forma de una intensidad que surge de al menos uno de:

- 30 un cambio relativo en los niveles de luz;
un cambio relativo en los niveles de sonido; y
tasas de cambio en una función monitorizable.

9. El sistema de aumento de contenido de medios de acuerdo con cualquiera de las reivindicaciones 1 a 7, en donde la entrada adopta la forma de una escala de intensidades que surge de al menos uno de:

- 40 un cambio relativo en los niveles de luz;
un cambio relativo en los niveles de sonido; y
tasas de cambio en una función monitorizable.

10. Un método para aumentar una experiencia sensorial de un usuario componiendo un producto de medios mediante la alineación selectiva de secciones de audio de al menos un archivo de audio con un evento variable temporalmente en forma de un evento físico monitorizado o detectado o, como alternativa, en forma de una sucesión de fotogramas de imagen y en donde (a) el evento variable temporalmente incluye transiciones identificables entre actividades discerniblemente diferentes y (b) cada sección de audio tiene una metaetiqueta asociada que identifica al menos un tema contextual asociado con cada sección de audio y, para cada sección de audio, identificando adicionalmente la metaetiqueta asociada al menos un punto de entrada abstracto y al menos un punto de salida abstracto que soportan transiciones de corte o atenuación perfectas entre diferentes secciones de audio y en donde cada uno de dicho al menos un punto de entrada abstracto y dicho al menos un punto de salida abstracto se relaciona con un inicio asociado con una anacrusa para la sección de audio y en donde la anacrusa se expresa en términos de pulsos y fracciones y cada anacrusa es una nota o secuencia de notas que precede al primer pulso en un compás de música en cada sección de audio, comprendiendo el método:

- 55 recibir, como una entrada de datos (12) temporal, el evento variable temporalmente como un flujo de datos; acceder a una base de datos que contiene una multiplicidad de archivos de audio, conteniendo cada uno una o más secciones de audio y metaetiquetas asociadas que identifican temas contextuales asociados con cada sección de audio,
60 resolver cada una de dichas actividades diferentes en uno o más de una pluralidad de diferentes temas contextuales categorizados;
basándose en identificar una correlación entre cada uno de dichos temas contextuales categorizados relacionados con la entrada de datos y temas contextuales en metaetiquetas de secciones de audio,
caracterizado por seleccionar una primera sección de audio y una sección de audio de destino diferente que reflejan las diferentes actividades en el evento variable temporalmente y correlacionarlas con las mismas, y,
65 después de eso;

seleccionar anacrusas que son puntos de posible salida real y de posible entrada real para cortar o atenuar perfectamente entre sí dichas secciones de audio primera y de destino seleccionadas para alinearse con temas contextuales categorizados correspondientes de las diferentes actividades a medida que la entrada de datos varía con el tiempo, en donde:

5 dichas anacrusas seleccionadas tienen unos desplazamientos medidos correspondientes en el tiempo a unos puntos de salida abstractos y de entrada abstractos respectivamente en la primera sección de audio y la sección de audio de destino entre las que va a producirse el corte o la atenuación,
 10 dichas anacrusas seleccionadas en la primera sección de audio y la sección de audio de destino se alinean de tal manera que se produzcan en un momento común,
 las diferencias de tiempo medidas de inicio en cada una de la primera sección de audio seleccionada y la sección de audio de destino seleccionada a partir de las anacrusas seleccionadas crean posiciones alternativas de atenuación cruzada o corte para las transiciones entre la primera sección de audio seleccionada y la sección de audio de destino seleccionada,
 15 la posición de atenuación cruzada o corte se selecciona basándose en qué posición respectiva de atenuación cruzada o corte en la primera sección de audio y la segunda sección de audio de destino se evalúa que se produce más temprano en el tiempo con respecto a las anacrusas alineadas y en donde la primera posición de atenuación cruzada o corte tiene un inicio asociado, y
 20 la producción de una atenuación cruzada o corte entre la primera sección de audio y la sección de audio de destino en un punto en o antes de que dicho inicio asociado comience a aumentar para realizar una transición rítmica ininterrumpida en la salida de audio preservando la integridad de la sincronización musical entre la primera sección de audio y la sección de audio de destino; y
 25 generar, como una salida, un archivo de medios o una salida en tiempo real en la que las transiciones entre secciones de audio se alinean con transiciones entre diferentes actividades.

11. El método de la reivindicación 10, en donde dicho evento variable temporalmente está relacionado con uno de:

30 un comienzo y un final en el caso de una característica sostenida para el tema contextual, en donde la característica sostenida es uno de una función de forma y un punto de éxito a lo largo del tiempo; y
 un momento específico para un punto de éxito.

12. El método de la reivindicación 10, en donde el método aumenta una experiencia sensorial general de un usuario en el mundo real cambiando, reordenando o repitiendo dinámicamente y a continuación reproduciendo secciones de audio seleccionadas dentro del contexto de dichos temas categorizados asociados con la naturaleza cambiante de los eventos presentados como la entrada de datos.

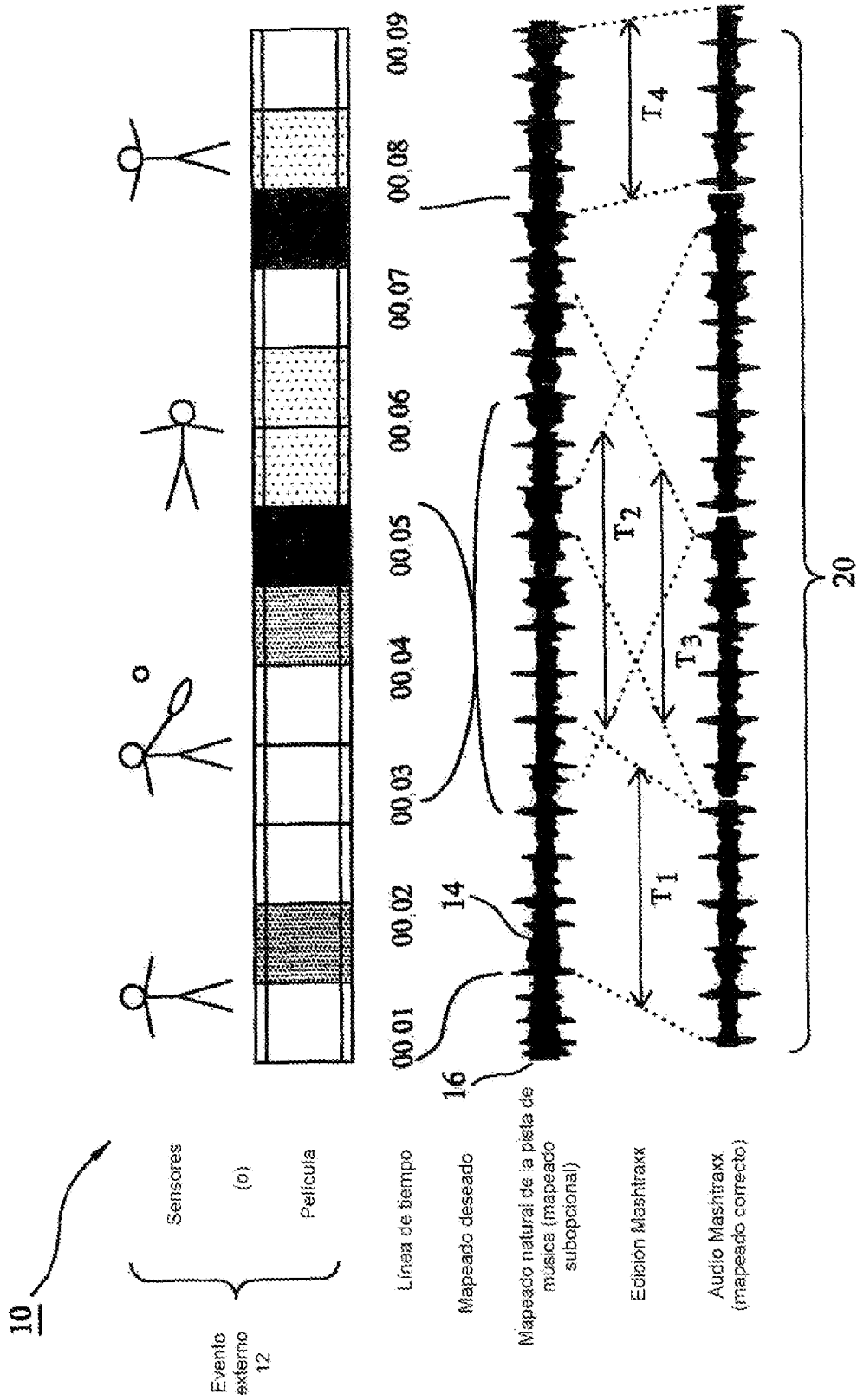


FIG. 1

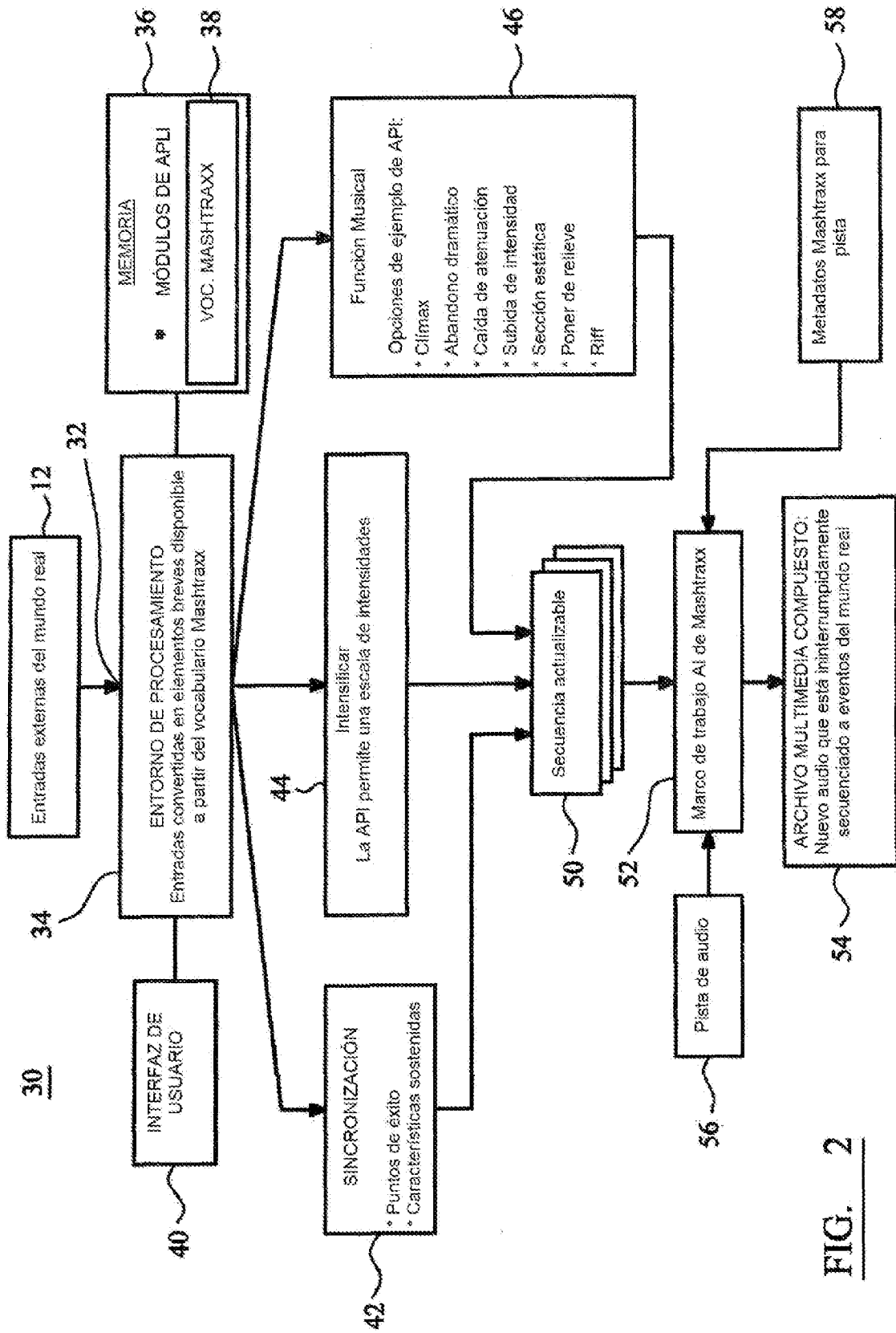


FIG. 2

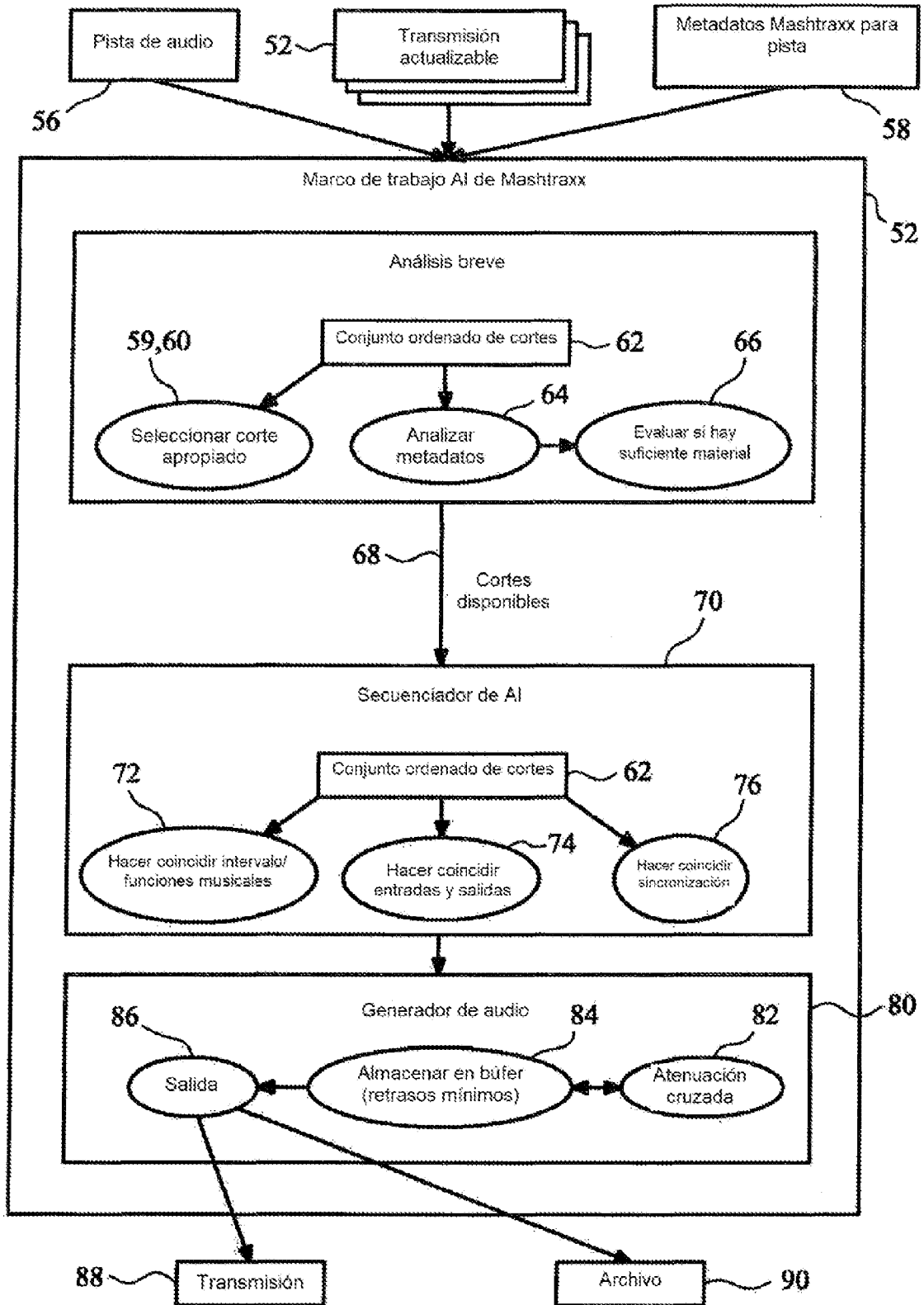


FIG. 3

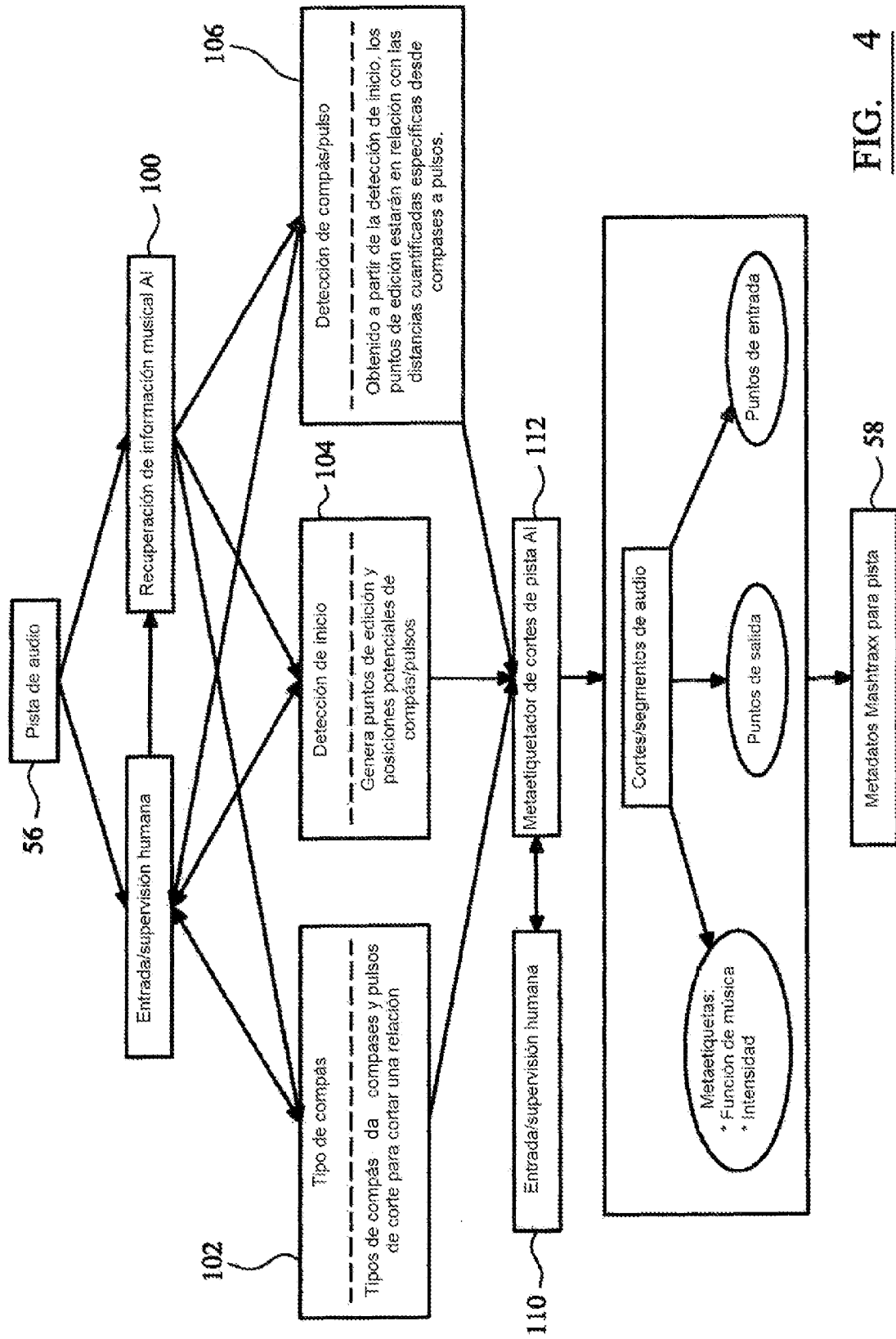


FIG. 4

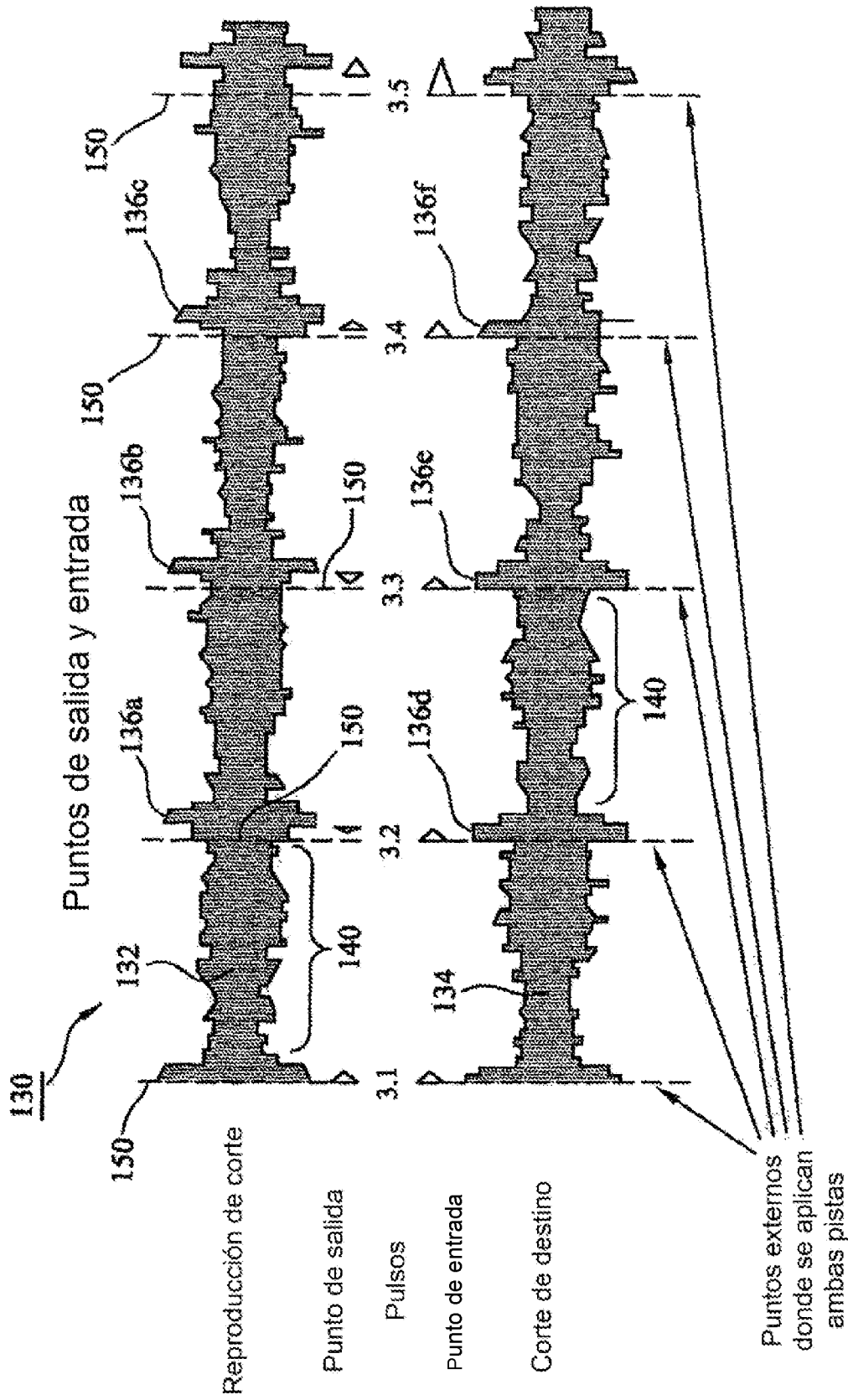


FIG. 5

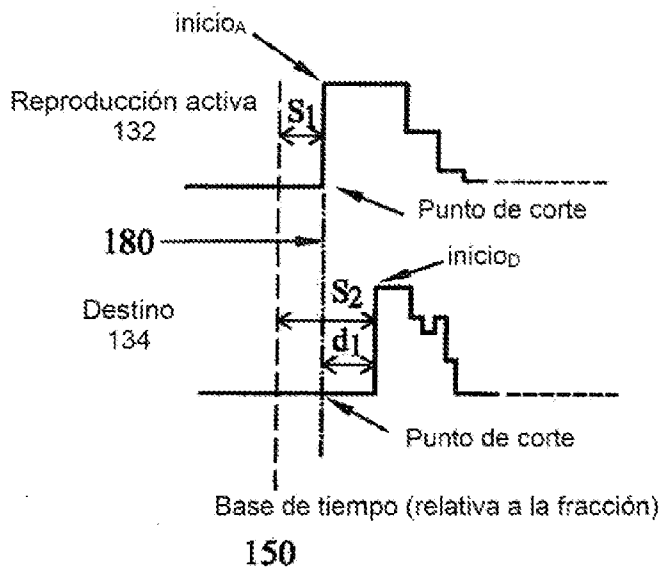


FIG. 6a

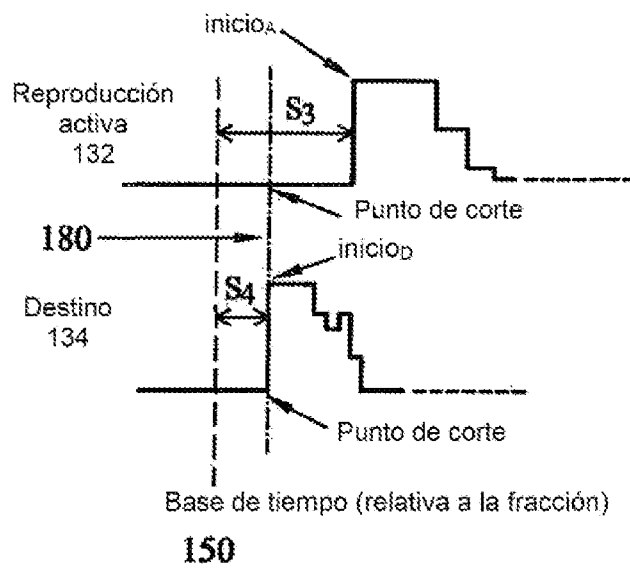


FIG. 6b

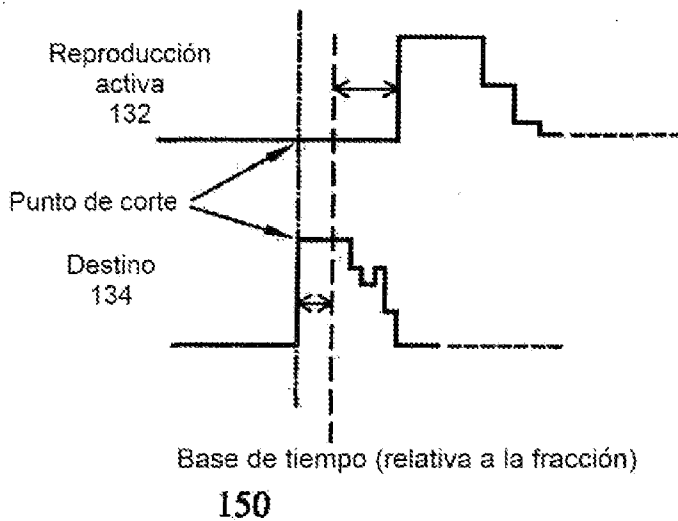
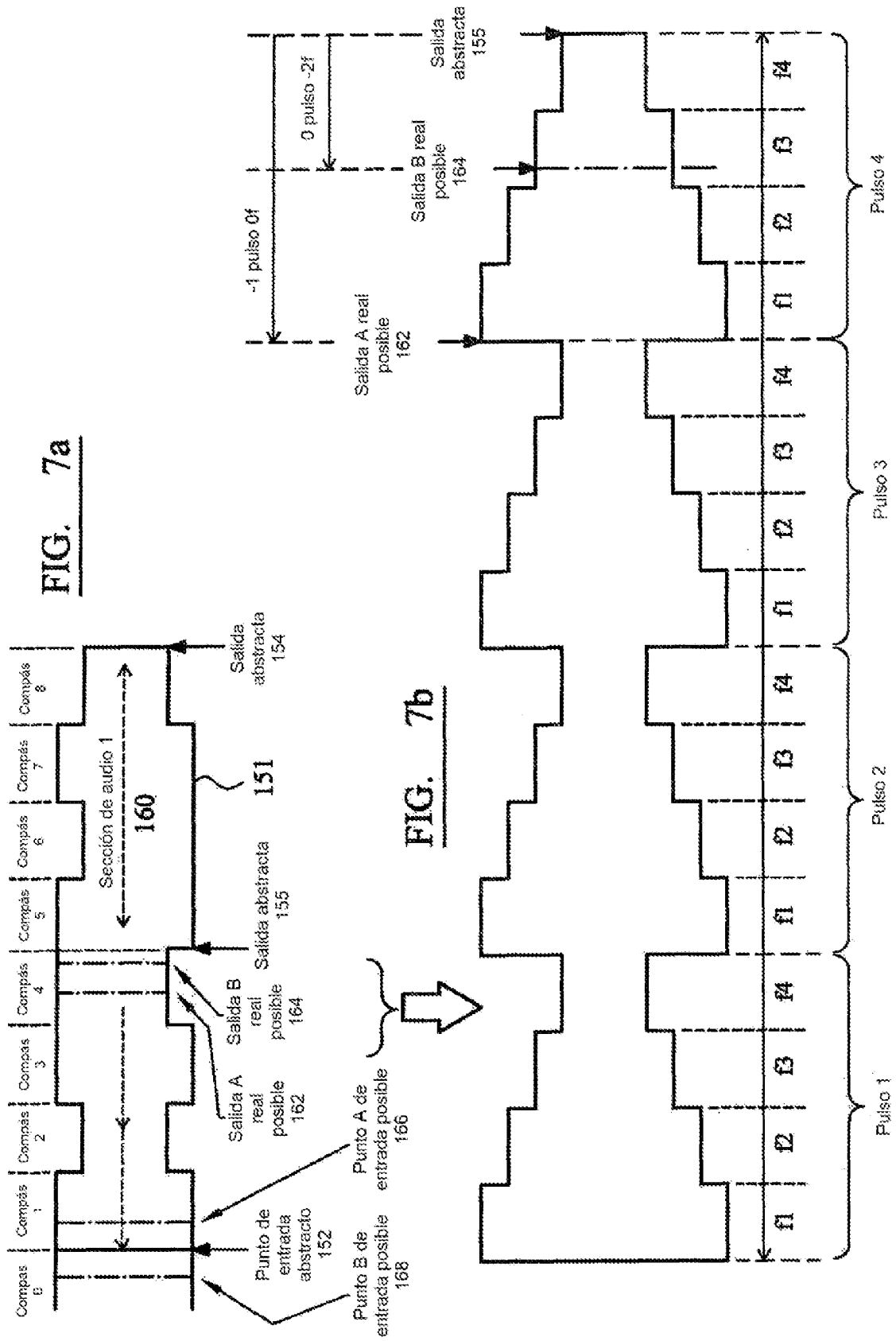


FIG. 6c



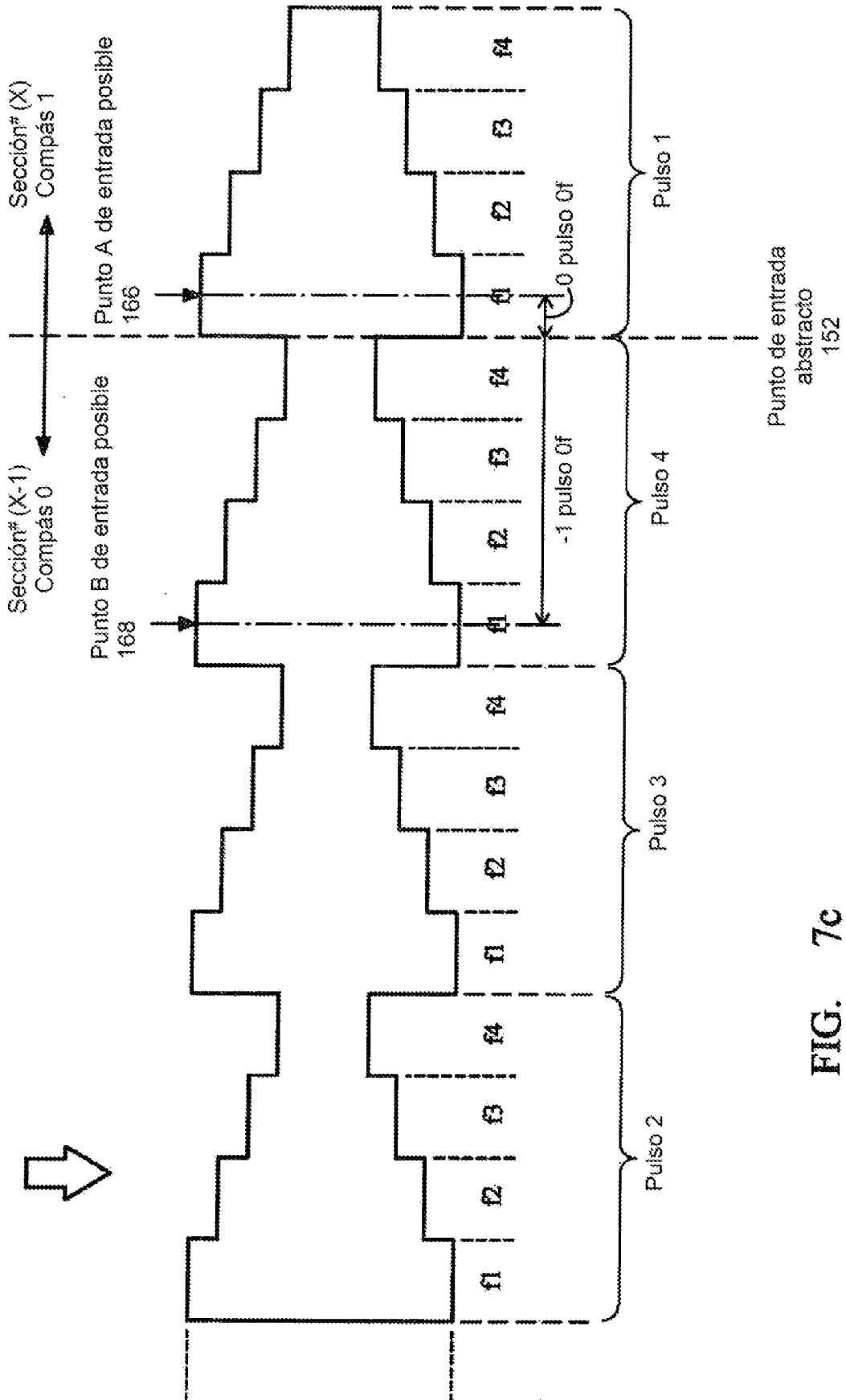


FIG. 7c

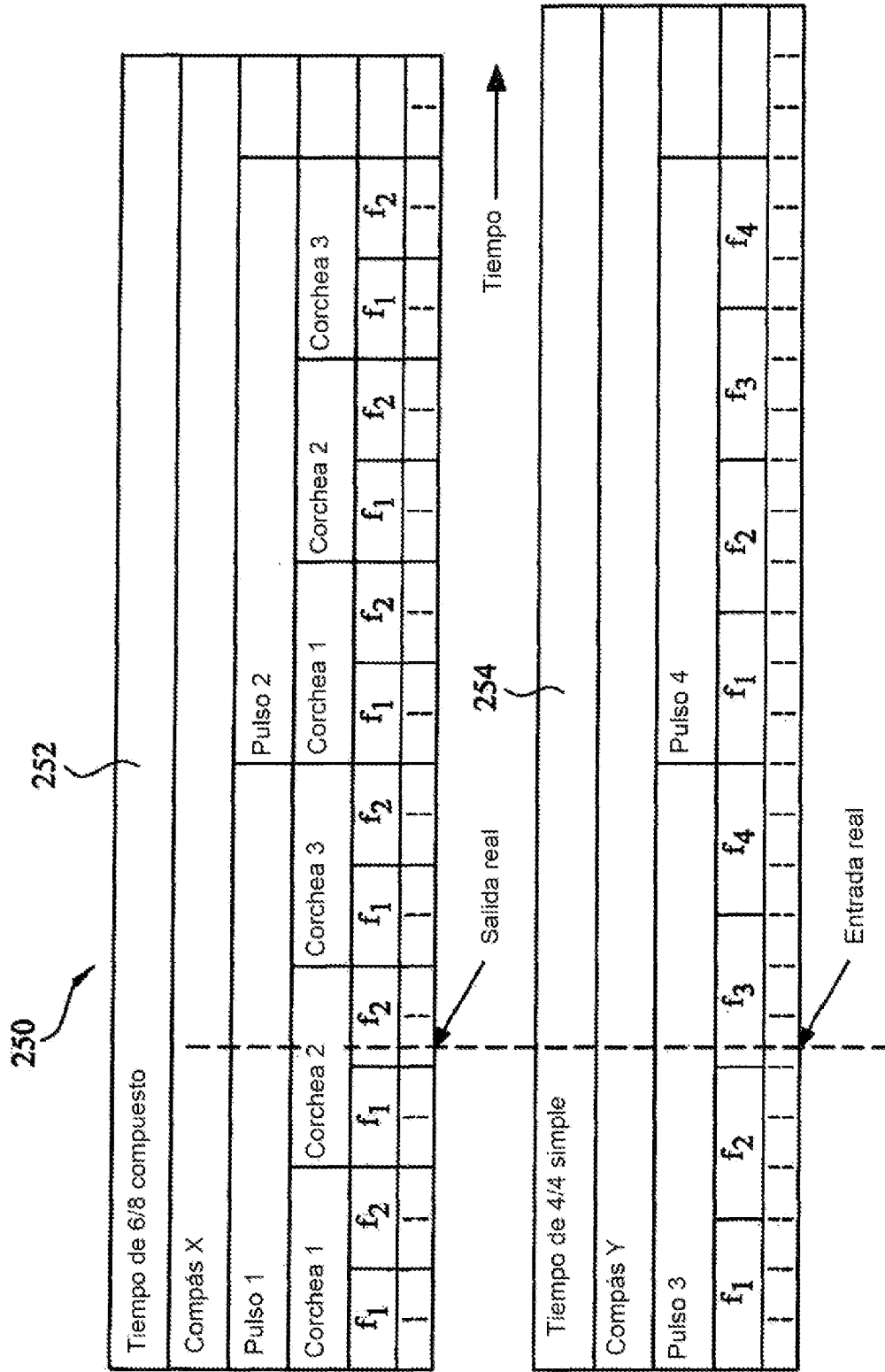


FIG. 8

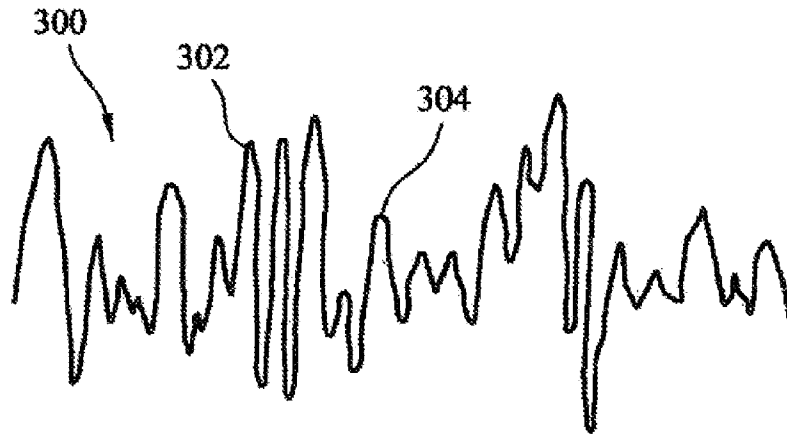


FIG. 9a

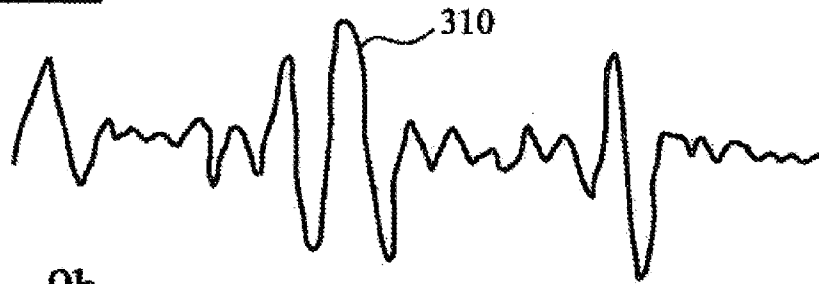


FIG. 9b

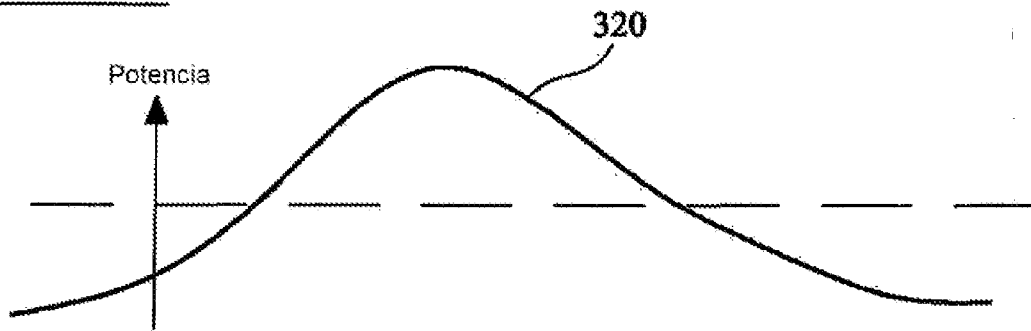


FIG. 9c

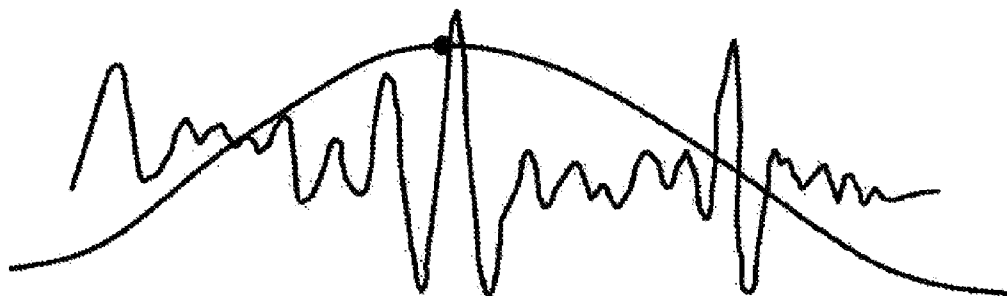


FIG. 9d

Base de datos					
Pista	Sección	Descripción	Posición de tiempo	Pto de entrada	Pto de salida
1	001	Coros	0,15 - 0,36		
	010	VERSO	0,36 - 0,59		
	011	Intro	0,00 - 0,15		
2	001	Coros	1,05 - 1,20		
	010	VERSO	1,30 - 1,48		
	011	Intro	0,00 - 0,10		
	001	Coros	2,05 - 2,20		
	011	xxx	3,19 - 3,49		
	101	FILA DE SALIDA	4,06 - 4,21		
N		Clasificación de temas 204	Sincronización de sección	Puntos de salida/ entrada Anacrusa Desplazamiento (Pulso + Fracción) en relación al inicio/ fin de una sección de audio	

FIG. 10