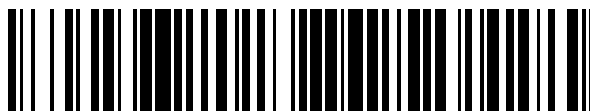


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 561 534**

51 Int. Cl.:

G10L 15/26 (2006.01)

H04H 60/04 (2008.01)

G10L 15/22 (2006.01)

G06F 9/44 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **11.01.2012 E 12700279 (8)**

97 Fecha y número de publicación de la concesión europea: **16.12.2015 EP 2659485**

54 Título: **Mezclador de pistas de audio semántico**

30 Prioridad:

03.02.2011 EP 11153211

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

26.02.2016

73 Titular/es:

**FRAUNHOFER GESELLSCHAFT ZUR
FÖRDERUNG DER ANGEWANDTEN
FORSCHUNG E.V. (100.0%)
Hansastraße 27C
80686 München, DE**

72 Inventor/es:

**UHLE, CHRISTIAN;
HERRE, JÜRGEN;
POPP, HARALD y
RIDDERBUSCH, FALKO**

74 Agente/Representante:

ARIZTI ACHA, Monica

ES 2 561 534 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).

Mezclador de pistas de audio semántico

DESCRIPCIÓN

5 El campo de la presente invención se refiere a un mezclador de audio para mezclar señales de múltiples pistas de acuerdo con especificaciones del usuario. Este se refiere al procesamiento de señales de audio, en particular a la tarea de mezclar una grabación de múltiples pistas de acuerdo con un conjunto de criterios definidos por el usuario. El campo de la invención se refiere adicionalmente a un método para mezclar una pluralidad de pistas de audio para dar una señal de mezcla. El campo de la invención también se refiere a un programa informático para dar instrucciones a un ordenador para que realice el método para mezclar una pluralidad de pistas de audio.

10 La disponibilidad continuamente creciente de contenido multimedia da al usuario nuevas maneras de disfrutar la música y de interactuar con la música. Estas posibilidades están acompañadas por el desafío de desarrollar las herramientas para ayudar al usuario en tales actividades.

15 Desde la perspectiva de la recuperación de información, este desafío se abordó hace ya más de una década, conduciendo a la emocionante área de investigación de la recuperación de información musical y a numerosas aplicaciones comerciales.

20 Un aspecto diferente que no se ha abordado hasta este punto es la interacción con contenido que está disponible en un formato de múltiples pistas. Un formato de múltiples pistas puede consistir en señales separadas y alineadas en el tiempo (también conocidas como pistas singulares (ST, *single track*) para cada objeto de sonido (SO, *sound object*) o grupos de objetos (subgrupos). De acuerdo con una definición, los subgrupos son los componentes individuales de una mezcla, guardados por separado (por lo general, en disco o en cinta) con el fin de su uso en una remezcla.

25 En el proceso tradicional de producción musical, se combinan múltiples pistas singulares de una manera sofisticada para dar una señal de mezcla (MS, *mixture signal*) la cual se envía a continuación al usuario final. La evolución en curso de las tecnologías de audio digital, por ejemplo, el desarrollo de nuevos formatos de audio para audio basándose en objetos paramétricos, permite la interacción con la música en una medida mucho mayor. El usuario tiene acceso a grabaciones de múltiples pistas y puede controlar activamente el proceso de mezclado. Algunos artistas han empezado a publicar los subgrupos para algunas de sus canciones, pretendiéndose que los oyentes puedan remezclar y reutilizar con libertad la música de cualquier manera deseada.

30 Una obra musical o de audio publicada en formato de múltiples pistas se puede usar de muchas maneras. El usuario puede controlar los parámetros de mezclado para las diferentes pistas, enfatizando de este modo pistas seleccionadas al tiempo que se atenúan otras pistas. Se pueden silenciar una o más pistas, por ejemplo, con el fin de hacer karaoke o acompañamiento. Se puede aplicar efectos de sonidos tales como eco, reverberación, distorsión, coros, etc., a unas pistas seleccionadas sin afectar a las otras pistas. Se pueden extraer una o más pistas del formato de múltiples pistas y se pueden usar en otra obra musical u otra forma de obra de audio, tal como un audiolibro, una ponencia, un *podcast*, etc. En la siguiente descripción, una aplicación de las enseñanzas que se divulgan en el presente documento, analiza de una forma a modo de ejemplo, la masterización de una obra musical grabada. Sin embargo, se debe entender que se pretende que el procesamiento de cualquier sonido grabado que involucre mezclar una pluralidad de pistas de audio singulares, sea abordado y esté cubierto de igual manera por las enseñanzas que se divulgan en el presente documento.

35 El mezclado automático ha sido, y todavía es, el objeto de una cantidad de proyectos de investigación. En 2009, Perez-Gonzalez y col. describieron un método para la ecualización automática de señales de múltiples pistas (E. Perez-Gonzalez y J. Reiss, "Ecualización Automática de Audio Multicanal Usando Métodos Adaptativos Cruzados" (*"Automatic Equalization of Multi-Channel Audio Using Cross-Adaptive Methods"*), Actas de la 127^a Conv. de AES, 2009). Los autores presentan un método para establecer de forma automática la atenuación para cada señal de una señal de múltiples pistas. Se determinan las ganancias de tal modo que la intensidad sonora de cada señal es igual a la intensidad sonora promedio de todas las señales. Otro artículo de los mismos autores abordó el "Control de Ganancia y Atenuador Automáticos para Mezclado en Vivo" (*"Automatic Gain and Fader Control for Live Mixing"*) y se publicó en las Actas de WASPAA, 2009.

40 HiFi Semántico es el nombre del Proyecto Europeo IST-507913 (H. Vinet y col., "Informe Final de HiFi Semántico" (*"Semantic HiFi Final Report"*), Informe Final de IST-507913 2006). Este se refiere principalmente a la recuperación, la exploración y la compartición de contenido multimedia. Este comprende la exploración y la navegación en bases de datos, la generación de listas de reproducción, la navegación intrapista (usando análisis estructural como identificación de versos - coros) y la compartición de metadatos. Este también aborda la interacción / creación de contenidos / edición: la generación de mezclas incluye la sincronización (es decir, la "concatenación" de señales de audio, no la mezcla de señales de múltiples pistas), la transformación de voz, la transformación de ritmo, instrumentos controlados por voz y efectos.

Se conoce otro proyecto bajo la designación "Audio Estructurado" ("*Structured Audio*") o MPEG 4. El Audio Estructurado permite la transmisión de señales de audio a bajas velocidades de transmisión de bits y la manipulación y el acceso de forma perceptual de datos sónicos usando una descripción simbólica y semántica de las señal (consúltese B. L. Vercoe y W. G. Gardner y E. D. Scheirer, "Audio Estructurado: Creación, Transmisión y Modelizado de Representaciones Paramétricas de Sonidos" ("*Structured Audio: Creation, Transmission, and Rendering of Parametric Sound Representations*"), Actas de IEEE, vol. 86, págs. 922 - 940, 1998). Este presenta una descripción de posproducción paramétrica de sonidos para mezclar múltiples transmisiones y añadir efectos de audio. Las descripciones paramétricas determinan cómo se sintetizan los sonidos. El audio estructurado se refiere a la sintetización de señales de audio.

En la solicitud de patente internacional publicada con el número de publicación internacional WO 2010/111373 A1, se divulga una interfaz y sistema controlado por habla y consciente del contexto. El sistema de interfaz de usuario dirigido por habla incluye por lo menos un altavoz para entregar una señal de audio a un usuario y por lo menos un micrófono para capturar expresiones verbales de un usuario. Un dispositivo de interfaz realiza la interconexión con e altavoz y el micrófono y proporciona una pluralidad de señales de audio al altavoz para que sean oídas por el usuario. Un circuito de control está acoplado de forma operativa con el dispositivo de interfaz y está configurado para seleccionar por lo menos una de la pluralidad de señales de audio como una señal de audio de primer plano para su entrega al usuario a través del altavoz. El circuito de control es accionable para reconocer expresiones verbales de un usuario y usar las expresiones verbales reconocidas para controlar la selección de la señal de audio de primer plano.

La publicación de solicitud de patente de EE. UU. con n.º US 2002/0087310 A1 divulga un método y sistema implementado por ordenador para la manipulación de un diálogo verbal con un usuario. Una entrada verbal procedente de un usuario contiene unas palabras que están orientadas hacia una pluralidad de conceptos. La entrada verbal del usuario contiene una solicitud para que se preste un servicio. El reconocimiento verbal de la entrada verbal del usuario se usa para generar palabras reconocidas. Se aplica una plantilla de diálogo a las palabras reconocidas. La plantilla de diálogo tiene unos nodos que están asociados con unos conceptos previamente determinados. Los nodos incluyen informaciones de procesamiento de solicitudes diferentes. Se identifican regiones conceptuales dentro de la plantilla de diálogo basándose en qué nodos están asociados con conceptos que coinciden aproximadamente con los conceptos de las palabras reconocidas. La solicitud del usuario se procesa mediante el uso de la información de procesamiento de solicitudes de los nodos que están contenidos dentro de las regiones conceptuales identificadas.

El artículo "Detección de Transitorios de Señales de Audio Basándose en un Filtro de Peine Adaptativo En el Dominio de la Frecuencia" ("*Transient Detection of Audio Signals Based on an Adaptive Comb Filter in the Frequency Domain*"), M. Kwong y R. Lefebvre, presenta un algoritmo de detección de transitorios adecuado para la selección de ritmos en señales musicales. En muchas señales de audio, los transitorios de baja energía están enmascarados por sonidos estacionarios de alta energía. Estos transitorios enmascarados, así como unos transitorios de mayor energía y más visibles, portan una información importante acerca de la segmentación temporal y el ritmo de la señal musical. El algoritmo de segmentación propuesto usa un modelo sinusoidal combinado con un filtrado de peine adaptativo en el dominio de la frecuencia para eliminar la componente estacionaria de una señal de sonido. Después del filtrado, la envolvente temporal de la señal residual se analiza para localizar las componentes transitorias. Los resultados muestran que el algoritmo propuesto puede detectar con precisión la mayor parte de los transitorios de baja energía.

La mezcla de una grabación de múltiples pistas por lo general es una tarea de creación de contenidos que, en general, es realizada por un experto, el ingeniero de mezclado. Los desarrollos actuales en multimedia, como formatos de audio interactivos, conducen a aplicaciones en las que es necesario mezclar grabaciones de múltiples pistas de una manera automatizada o de una manera semiautomatizada guiada por una persona no experta. Se desea que la señal de mezcla obtenida de forma automática tenga calidad de sonido subjetiva comparable con una señal de mezcla generada por una persona experta.

Las enseñanzas que se divulgan en el presente documento abordan este objetivo general. Las enseñanzas se refieren al procesamiento de señales de audio, en particular la tarea de mezclar múltiples pistas de acuerdo con un conjunto de criterios de grabación definidos por el usuario con el fin (futuro) de escuchar. Un mezclador de audio y un método para mezclar una pluralidad de pistas de audio para dar una señal de mezcla de acuerdo con las enseñanzas que se divulgan en el presente documento establecen una conexión entre una idea substancialmente estética de una persona no experta y la señal de mezcla resultante.

Por lo menos uno de estos objetivos y / u otros objetivos posibles se logran por medio de un mezclador de audio de acuerdo con la reivindicación 1 o la reivindicación 13, un método para mezclar una pluralidad de pistas de audio de acuerdo con la reivindicación 11 o la reivindicación 14 y un programa informático de acuerdo con la reivindicación 15.

De acuerdo con las enseñanzas que se divulgan en el presente documento, un mezclador de audio para mezclar una pluralidad de pistas de audio para dar una señal de mezcla comprende un intérprete de comandos semánticos, un procesador de pistas de audio y un combinador de pistas de audio. El intérprete de comandos semánticos está configurado para recibir un comando de mezclado semántico y para entregar una pluralidad de parámetros de mezclado para la pluralidad de pistas de audio a partir del comando de mezclado semántico. El procesador de pistas de audio está configurado para procesar la pluralidad de pistas de audio de acuerdo con la pluralidad de parámetros de mezclado. El combinador de pistas de audio está configurado para combinar la pluralidad de pistas de audio procesadas por el procesador de pistas de audio para dar la señal de mezcla.

El método para mezclar una pluralidad de pistas de audio para dar una señal de mezcla de acuerdo con las enseñanzas que se divulgan comprende: recibir un comando de mezclado semántico; obtener una pluralidad de parámetros de mezclado para la pluralidad de pistas de audio a partir del comando de mezclado semántico; procesar la pluralidad de pistas de audio de acuerdo con la pluralidad de parámetros de mezclado; y combinar la pluralidad de pistas de audio que resultan del procesamiento de la pluralidad de pistas de audio para formar la señal de mezcla.

El programa informático comprende o representa instrucciones para permitir que un ordenador o un procesador realice el método para mezclar una pluralidad de pistas de audio. El programa informático se puede realizar en un medio legible por ordenador que tiene almacenado en el mismo dicho programa informático para realizar, cuando se ejecuta en un ordenador, el método tal como se expresa en la reivindicación 14.

El comando de mezclado semántico puede estar basado en criterios definidos por el usuario que proporcionan una descripción semántica de la señal de mezcla resultante deseada. De acuerdo con las enseñanzas que se divulgan en el presente documento, se pueden incorporar el análisis de audio semántico, la psicoacústica y el procesamiento de señales de audio entre sí con el fin de obtener una señal de mezcla de forma automática basándose en las descripciones semánticas. Este proceso se puede denominar "Mezclado Semántico".

Se puede considerar el Mezclado Semántico como un método que permite que un ordenador mezcle una grabación de múltiples pistas de acuerdo con una especificación dada por un usuario. La especificación por lo general se da en forma de una descripción semántica. Dada esta descripción semántica, se pueden determinar los parámetros de mezclado teniendo en cuenta las características de la pista o pistas individuales y la audición humana.

Por lo tanto, el mezclador de audio de acuerdo con las enseñanzas que se divulgan en el presente documento por lo general comprende un ordenador o un procesador, o este interactúa con un ordenador / procesador. El procesador de pistas de audio y el combinador de pistas de audio se pueden combinar como una sola unidad.

La obtención de la pluralidad de parámetros de mezclado a partir del comando de mezclado semántico puede involucrar el análisis de un significado del comando de mezclado semántico, o de partes del mismo. Una parte del comando de mezclado semántico puede ser una expresión semántica, tal como una palabra o un grupo de palabras. La expresión o expresiones semánticas se pueden traducir entonces a un conjunto de parámetros de mezclado específicos para la pluralidad de pistas de audio. De este modo, el comando de mezclado semántico se implementa por medio de los parámetros de mezclado específicos que se corresponden con el significado del comando de mezclado semántico. La acción de traducir el comando de mezclado semántico y / o de sus expresiones semánticas constituyentes puede comprender evaluar una función de traducción o consultar una tabla de consulta, por ejemplo. Los parámetros de la función de traducción o los registros de datos de la tabla de consulta por lo general están predefinidos y representan una recopilación de conocimiento experto de, por ejemplo, ingenieros de mezclado con experiencia. El conocimiento experto se puede recopilar con el tiempo, por ejemplo, mediante el registro de las instrucciones orales dadas por un artista o un productor musical a su ingeniero de mezclado, así como los ajustes que sean realizados por el ingeniero de mezclado. De este modo, un ingeniero de mezclado experto puede entrenar la función de traducción y / o la tabla de consulta.

De acuerdo con un aspecto de las enseñanzas que se divulgan en el presente documento, el intérprete de comandos semánticos puede comprender una base de datos de vocabulario para identificar expresiones semánticas dentro del comando de mezclado semántico. Por medio de la base de datos de vocabulario, el intérprete de comandos semánticos puede identificar, por ejemplo, sinónimos. Además, puede ser posible establecer una correspondencia de una palabra o un grupo de palabras contenidas en el comando de mezclado semántico con un valor específico. Por ejemplo, se puede establecer una correspondencia de una palabra para identificar un instrumento ("guitarra") con un identificador o número de canal particular, sobre el cual se ha grabado el instrumento. La base de datos de vocabulario puede comprender además entradas que identifican una cierta parte de una parte musical, tal como el comienzo (por ejemplo, "Intro", el coro ("Coro"), o el final (por ejemplo, "Coda" u "Outro"). Aún otro posible uso de la base de datos de vocabulario es para reconocer y asignar parámetros o estilos de mezclado expresados de forma semántica, tales como "fuerte", "suave", "claro", "amortiguado", "distante", "cercano", etc.

En una realización de las enseñanzas que se divulgan en el presente documento, el mezclador de audio puede comprender además un identificador de pista de audio para identificar una pista de audio objetivo de entre la pluralidad de pistas de audio. La pista de audio objetivo se puede indicar dentro del comando de mezclado semántico por una expresión de identificación de pista de audio. El identificador de pista de audio puede ser útil si la pluralidad de pistas de audio no están claramente marcadas o identificadas en lo que respecta a qué parte o subgrupo contienen estas. Por ejemplo, las pistas de audio pueden estar simplemente numeradas como "pista 1", "pista 2", ... "pista N". El identificador de pista de audio puede analizar entonces cada una de la pluralidad de pistas de audio para determinar o bien ninguna, o bien una o bien varias pistas de audio que parecen coincidir con una pista de audio identificada por la expresión de identificación de pista.

El identificador de pista de audio puede estar configurado para recuperar un registro de datos que se corresponde con la expresión de identificación de pista de audio a partir de una base de datos de plantillas de pistas de audio, para realizar un análisis de por lo menos un nombre de pista, un identificador de pista, un timbre, una estructura rítmica, un rango de frecuencias, una muestra de sonido y una densidad de armónicos de por lo menos una pista de audio de entre la pluralidad de pistas de audio, para comparar un resultado del análisis con el registro de datos dando como resultado por lo menos una puntuación de coincidencia, y para determinar la pista de audio objetivo sobre la base de la por lo menos una puntuación de coincidencia entre la por lo menos una pista de audio y el registro de datos. La tarea que se va a realizar por el identificador de pista de audio es identificar la pista de audio objetivo de entre la pluralidad de pistas de audio. La pista de audio objetivo se corresponde con la expresión de identificación de pista de audio, es decir, si la expresión de identificación de pista de audio es "guitarra", entonces, a continuación de la identificación con éxito por el identificador de pista de audio, la pista de audio objetivo por lo general debería contener la parte de guitarra de una obra musical. La base de datos de plantillas de pistas de audio puede comprender un registro de datos que se corresponde con el instrumento "guitarra", comprendiendo el propio registro de datos unos valores y / o una información que son característicos para una guitarra. Por ejemplo, el registro de datos puede comprender un modelo de frecuencia del sonido de guitarra típico y / o un modelo de ataque - decaimiento del sonido de guitarra típico. El registro de datos también podría contener una muestra de sonido de una guitarra, que el identificador de pista de audio puede usar para un análisis de similitud.

De acuerdo con un aspecto de las enseñanzas que se divulgan en el presente documento, el mezclador de audio puede además comprender un identificador de sección de tiempo para identificar una sección de tiempo objetivo dentro de la pluralidad de pistas de audio, estando indicada la sección de tiempo objetivo dentro del comando de mezclado semántico por una expresión de identificación de sección de tiempo. En el caso de que el usuario desee mezclar una primera sección de una señal de audio (por ejemplo, una obra musical) de manera diferente de una segunda sección de la misma señal de audio, el mezclador de audio por lo general necesita saber dónde empiezan y terminan las diversas secciones de la señal de audio con el fin de aplicar parámetros de mezclado específicos a estas secciones de la señal de audio.

El identificador de sección de tiempo puede estar configurado para estructurar la pluralidad de pistas de audio en una pluralidad de secciones de tiempo. En especial las obras musicales tienen a menudo una cierta estructura que está influenciada por convenciones musicales, por ejemplo, la forma de la canción con sus secciones alternantes de verso y coro. Este conocimiento puede ser aprovechado por el identificador de sección de tiempo determinando en primer lugar si la señal de audio que está representada por la pluralidad de pistas de audio sigue una cierta estructura musical y para asignar entonces las secciones de tiempo de la señal de audio a las secciones de tiempo de la estructura musical. Con este fin, el identificador de sección de tiempo puede comprender un reconocedor de patrones para reconocer patrones recurrentes y / o similares dentro de la señal de audio. El reconocimiento de patrones se puede basar en un análisis de melodía, un análisis armónico y un análisis rítmico, por mencionar unos pocos.

El identificador de sección de tiempo puede estar configurado para realizar un análisis de la pluralidad de pistas de audio para determinar por lo menos un instante de tiempo en el que tiene lugar un cambio de una propiedad característica de una señal de audio que está representada por la pluralidad de pistas de audio, y para usar el por lo menos un instante de tiempo determinado como por lo menos un límite entre dos secciones de tiempo adyacentes.

El mezclador de audio puede comprender además una interfaz de metadatos para recibir metadatos en relación con la pluralidad de pistas de audio, siendo los metadatos indicativos de por lo menos uno de un nombre de pista, un identificador de pista, una información de estructura de tiempo, una información de intensidad, atributos espaciales de una pista de audio o una parte de la misma, características de timbre y características rítmicas. Los metadatos pueden haber sido generados por el productor de la pluralidad de pistas de audio y proporcionar una información útil para el mezclador de audio o el método para mezclar la pluralidad de pistas de audio. La disponibilidad de los metadatos guarda al mezclador de audio o al método de tener que realizar un análisis exhaustivo de la señal de audio con el fin de identificar las diversas pistas de audio y / o secciones de tiempo. La interfaz de metadatos también se puede usar para almacenar los resultados (instrumentos, estructura de tiempo, ...) de un análisis para su futura reutilización. De este modo, solo es necesario realizar una vez un análisis potencialmente extenso de la pluralidad de pistas de audio. Además, también se puede almacenar cualquier corrección manual en los resultados

de análisis determinados de forma automática, de tal modo que el usuario no tenga que corregir los mismos problemas una y otra vez. Al tener los resultados de análisis almacenados disponibles, el usuario puede producir diferentes versiones de mezcla a partir de la misma pluralidad de pistas de audio usando los mismos metadatos.

5 De acuerdo con un aspecto de las enseñanzas que se divulgan en el presente documento, el mezclador de audio puede comprender además una interfaz de comando para recibir el comando de mezclado semántico en un formato lingüístico. El formato lingüístico permite que el usuario exprese sus deseos en lo que respecta al resultado de la mezcla realizada por el mezclador de audio substancialmente por medio del lenguaje normal. El comando de mezclado semántico en el formato lingüístico se puede introducir en el mezclador de audio como lenguaje hablado usando un micrófono o lenguaje escrito usando, por ejemplo, un teclado.

15 De acuerdo con otro aspecto de las enseñanzas que se divulgan en el presente documento, el mezclador de audio puede comprender además una interfaz de ejemplo para recibir una señal de mezcla a modo de ejemplo, y un analizador de señal de mezcla para analizar la señal de mezcla a modo de ejemplo y para generar el comando de mezclado semántico basándose en el análisis de la señal de mezcla a modo de ejemplo. Usando la señal de mezcla a modo de ejemplo proporcionada por medio de la interfaz de ejemplo, el analizador de señal de mezcla puede determinar qué rasgos distintivos caracterizan la señal de mezcla a modo de ejemplo. Por ejemplo, el analizador de señal de mezcla puede reconocer un énfasis sobre la parte de tambor (fuertemente repetitiva) y la parte de bajo, mientras la melodía está menos acentuada. Estos rasgos distintivos detectados sugieren una así denominada Mezcla de Baile, es decir, un cierto estilo de mezcla. Esta información se puede proporcionar desde el analizador de señal de mezcla al intérprete de comandos semánticos. Basándose en esta información, el intérprete de comandos semánticos puede, por ejemplo, aumentar el volumen de la parte de tambor y la parte de bajo en relación con las otras partes. El intérprete de comandos semánticos podría incluso sustituir la parte de tambor con, por ejemplo, una parte de tambor sintetizada que, por lo general, se usa para el estilo de Mezcla de Baile deseado.

25 La interfaz de ejemplo puede estar configurada además para recibir una pluralidad de pistas de audio de ejemplo a partir de las cuales se obtuvo la señal de mezcla a modo de ejemplo. El analizador de señal de mezcla puede estar configurado para comparar las pistas de audio de ejemplo con la señal de mezcla a modo de ejemplo con el fin de determinar los parámetros de mezclado que se usaron para obtener la señal de mezcla a modo de ejemplo resultante. El comando de mezclado semántico producido por el analizador de señal de mezcla podría entonces comprender una descripción de cómo se modificaron las pistas de audio de ejemplo antes de que estas se mezclaran entre sí para formar la señal de mezcla a modo de ejemplo. Por ejemplo, el comando de mezclado semántico puede comprender una expresión tal como “tambores significativamente más altos; voces moderadamente más suaves, más distantes, filtradas con filtro de paso alto”. El intérprete de comandos semánticos puede entonces obtener la pluralidad de parámetros de mezclado a partir de este comando de mezclado semántico.

40 De acuerdo con otro aspecto de las enseñanzas que se divulgan en el presente documento, el intérprete de comandos semánticos puede comprender un procesador perceptual para transformar el comando de mezclado semántico en la pluralidad de parámetros de mezclado de acuerdo con un modelo perceptual de propiedades relacionadas con la audición de la señal de mezcla. El modelo perceptual por lo general implementa unas reglas psicoacústicas que describen cómo se deberían elegir ciertos parámetros de mezclado con el fin de lograr un efecto deseado para el oyente. Por ejemplo, para transmitir una impresión de distancia, pueden estar involucradas varias acciones de procesamiento de sonido, tales como reverberación, filtrado de frecuencias y atenuación. El modelo perceptual que, por lo general, está basado en hallazgos psicoacústicos, facilita la determinación de parámetros de mezclado adecuados para la realización de un efecto deseado.

50 De acuerdo con otro aspecto de las enseñanzas que se divulgan en el presente documento, el intérprete de comandos semánticos comprende un procesador de lógica difusa para recibir por lo menos una regla difusa que se obtiene a partir del comando de mezclado semántico mediante el intérprete de comandos semánticos, y para generar la pluralidad de parámetros de mezclado sobre la base de la por lo menos una regla difusa. El procesador de lógica difusa está bien adaptado para procesar el comando de mezclado semántico en la forma de la por lo menos una regla difusa. La por lo menos una regla difusa establece una correspondencia de una cantidad de entrada del procesador de lógica difusa con una cantidad de salida del procesador de lógica difusa en un dominio substancialmente semántico, es decir, un establecimiento de correspondencia de una cantidad de un primer formato semántico a una cantidad de un segundo formato semántico.

60 El procesador de lógica difusa puede estar configurado para recibir por lo menos dos reglas difusas concurrentes preparadas por el intérprete de comandos semánticos, y en el que el mezclador de audio además comprende un selector aleatorio para seleccionar una regla difusa concurrente de entre las por lo menos dos reglas difusas concurrentes. Mediante la aleatorización de la selección de la regla difusa de entre dos o más reglas difusas concurrentes, se puede crear una ilusión de libertad artística de tal modo que las señales de mezcla producidas por el mezclador de audio no tiendan a sonar substancialmente similares, por lo menos en lo que se refiere al estilo de mezclado, lo que de lo contrario podría ser el caso, cuando el mezclador de audio sigue un esquema más rígido con respecto a las reglas difusas.

En términos del método para mezclar la pluralidad de pistas de audio, se puede consultar una base de datos de vocabulario para identificar expresiones semánticas dentro del comando de mezclado semántico.

5 El método puede comprender, adicionalmente o como alternativa, una identificación de una pista de audio objetivo de entre la pluralidad de pistas de audio, estando indicada la pista de audio objetivo dentro del comando de mezclado semántico por una expresión de identificación de pista de audio. Con este fin, se puede recuperar un registro de datos que se corresponde con la expresión de identificación de pista de audio a partir de una base de datos de plantillas de pistas de audio. Entonces, se puede realizar un análisis de por lo menos uno de un nombre de pista, un identificador de pista, un timbre, una estructura rítmica, un rango de frecuencias, una muestra de sonido y una densidad de armónicos de por lo menos una pista de audio de entre la pluralidad de pistas de audio. Se puede comparar un resultado del análisis con el registro de datos dando como resultado por lo menos una puntuación de coincidencia. Entonces se puede determinar la pista de audio objetivo sobre la base de la por lo menos una puntuación de coincidencia entre la por lo menos una pista de audio y el registro de datos.

15 El método también puede comprender una acción para identificar una sección de tiempo objetivo dentro de la pluralidad de pistas de audio, estando indicada la sección de tiempo objetivo dentro del comando de mezclado semántico por una expresión de identificación de sección de tiempo. La acción para identificar la sección de tiempo objetivo puede estar configurada para estructurar la pluralidad de pistas de audio en una pluralidad de secciones de tiempo. La identificación de sección de tiempo puede comprender realizar un análisis de la pluralidad de pistas de audio para determinar por lo menos un instante de tiempo en el que tiene lugar un cambio de una propiedad característica de una señal de audio que está representada por la pluralidad de pistas de audio, y para usar el por lo menos un instante de tiempo determinado como por lo menos un límite entre dos secciones de tiempo adyacentes.

25 De acuerdo con otro aspecto de las enseñanzas que se divulgan en el presente documento, el método además puede comprender recibir metadatos en relación con la pluralidad de pistas de audio en una interfaz de metadatos. Los metadatos pueden ser indicativos de por lo menos uno de un nombre de pista, un identificador de pista, una información de estructura de tiempo, una información de intensidad, atributos espaciales de una pista de audio o una parte de la misma, características de timbre y características rítmicas.

30 El método puede comprender además recibir el comando de mezclado semántico en un formato lingüístico en una interfaz de comando de un mezclador de audio correspondiente.

35 De acuerdo con otro aspecto de las enseñanzas que se divulgan en el presente documento, el método puede comprender además: recibir una señal de mezcla a modo de ejemplo en una interfaz de ejemplo, analizar la señal de mezcla a modo de ejemplo por medio de un analizador de señal de mezcla, y generar el comando de mezclado semántico basándose en el análisis de la señal de mezcla a modo de ejemplo.

40 La acción de obtener la pluralidad de parámetros de mezclado a partir del comando de mezclado semántico puede comprender: transformar el comando de mezclado semántico en la pluralidad de parámetros de mezclado de acuerdo con un modelo perceptual de propiedades relacionadas con la audición de la señal de mezcla.

45 De acuerdo con un aspecto de las enseñanzas que se divulgan en el presente documento, la acción de obtener la pluralidad de parámetros de mezclado puede comprender: recibir por lo menos una regla difusa que se obtiene a partir del comando de mezclado semántico mediante un intérprete de comandos semánticos, y generar la pluralidad de parámetros de mezclado sobre la base de la por lo menos una regla difusa. La percepción de la por lo menos una regla difusa y la generación de la pluralidad de parámetros de mezclado sobre la base de la por lo menos una regla difusa se puede realizar mediante un procesador de lógica difusa.

50 El método puede comprender además: recibir por lo menos dos reglas difusas concurrentes, y seleccionar de forma aleatoria una regla difusa concurrente de entre las por lo menos dos reglas difusas concurrentes.

Las enseñanzas que se divulgan en el presente documento son diferentes de la técnica previa mencionada en los siguientes puntos:

- 55
- El método propuesto por Perez-Gonzalez y col. no tiene en cuenta descripciones semánticas para controlar el procesamiento.
 - El proyecto HiFi Semántico no aborda el procesamiento de formatos de múltiples pistas. No aborda el mezclado de señales de acuerdo con descripciones semánticas. No aborda los aspectos perceptuales que son necesarios para computar una señal de mezcla que cumpla las descripciones semánticas.
 - 60 - El proyecto de "Audio "Estructurado" se refiere a la sintetización de señales de audio. En contraste, las enseñanzas que se divulgan en el presente documento (Mezclado Semántico) se refieren a la mezcla de señales de audio.

Para resumir brevemente algunos de los aspectos cruciales de las enseñanzas que se divulgan en el presente documento, el mezclado de una grabación de múltiples pistas es una tarea de creación de contenidos. El Mezclado Semántico tiene como fin desarrollar soluciones para mezclar una grabación de múltiples pistas que está guiada por descripciones semánticas. Este combina técnicas de análisis de audio semántico, de psicoacústica y de procesamiento de señales de audio. El mezclado semántico es aplicable a varias aplicaciones como producción musical, SAOC (*Spatial Object Audio Coding*, Codificación de Audio de Objeto Espacial), creación de contenidos de vídeo doméstico, realidad virtual y juegos.

El Mezclado Semántico se puede describir brevemente con los siguientes rasgos distintivos (parcialmente opcionales):

- Proporciona medios para la interacción con el usuario
- El Mezclado Semántico aborda en gran medida la componente perceptual. Esto puede incluir también la adaptación al ambiente, el sistema de reproducción y preferencias de usuario.
- Combina la parte semántica y la parte psicoacústica. Es necesario que todo procesamiento semántico tenga en cuenta aspectos perceptuales. Se centra en el procesamiento de señales de audio más que en aplicaciones tradicionales de análisis semántico (recuperación de información musical, generación de listas de reproducción). Tiene como fin nuevas maneras de interacción con el contenido.
- Se refiere al procesamiento de grabaciones de múltiples pistas

Las enseñanzas que se divulgan en el presente documento se refieren, entre otras cosas, a un método para mezclar señales de múltiples pistas de acuerdo con especificaciones del usuario. Se refiere al procesamiento de señales de audio, en particular a la tarea de mezclar una grabación de múltiples pistas de acuerdo con un conjunto de criterios definidos por el usuario. Los criterios definidos por el usuario proporcionan una descripción semántica de la señal de mezcla resultante. Las enseñanzas que se divulgan en el presente documento pueden incorporar análisis de audio, psicoacústica y procesamiento de señales de audio con el fin de obtener una señal de mezcla de forma automática basándose en la descripción semántica.

Los rasgos distintivos anteriores y otros rasgos distintivos de las enseñanzas que se divulgan en el presente documento serán evidentes a partir de la siguiente descripción, la cual se hace solo a modo de ejemplo con referencia a los dibujos esquemáticos adjuntos en los cuales:

la figura 1 muestra un diagrama de bloques esquemático de un mezclador de audio;

la figura 2 ilustra una estructura de tiempo a modo de ejemplo de una obra musical en la estructura de canción que se emplea a menudo en música popular;

la figura 3 ilustra otra estructura de tiempo a modo de ejemplo de una obra musical en forma de sonata conocida en música clásica;

la figura 4 ilustra una distribución de pistas de audio a modo de ejemplo de una grabación de música popular;

la figura 5 muestra un diagrama de bloques esquemático de un mezclador de acuerdo con las enseñanzas que se divulgan en el presente documento;

la figura 6 ilustra un diagrama de bloques esquemático de un procesador de lógica difusa;

la figura 7 ilustra una función de pertenencia a modo de ejemplo para un conjunto difuso;

la figura 8 muestra un diagrama de bloques esquemático de un mezclador de audio que comprende un procesador de lógica difusa;

la figura 9 muestra un diagrama de bloques esquemático de otra configuración de un mezclador de audio de acuerdo con las enseñanzas que se divulgan en el presente documento;

la figura 10 ilustra un comando de mezclado semántico y su descomposición de acuerdo con un aspecto de las enseñanzas que se divulgan en el presente documento;

la figura 11 ilustra otro comando de mezclado semántico y su descomposición de acuerdo con un aspecto de las enseñanzas que se divulgan en el presente documento;

la figura 12 muestra un diagrama de bloques esquemático de una configuración adicional de un mezclador de audio de acuerdo con las enseñanzas que se divulgan en el presente documento;

- la figura 13 muestra un diagrama de bloques esquemático de una configuración de un mezclador de audio de acuerdo con las enseñanzas que se divulgan en el presente documento que comprende un identificador de pista de audio;
- 5 la figura 14 muestra un diagrama de bloques esquemático de una configuración de un mezclador de audio de acuerdo con las enseñanzas que se divulgan en el presente documento que comprende un identificador de sección de tiempo;
- 10 la figura 15 muestra un diagrama de bloques esquemático de una configuración de un mezclador de audio de acuerdo con las enseñanzas que se divulgan en el presente documento que comprende una interfaz de metadatos;
- 15 la figura 16 muestra un diagrama de bloques esquemático de una configuración de un mezclador de audio de acuerdo con las enseñanzas que se divulgan en el presente documento que comprende una interfaz de ejemplo para recibir señales de mezcla a modo de ejemplo;
- 20 la figura 17 muestra un diagrama de bloques esquemático de una configuración de un mezclador de audio de acuerdo con las enseñanzas que se divulgan en el presente documento que comprende un procesador perceptual y un modelo perceptual; y
- la figura 18 muestra un diagrama de flujo esquemático de un método para mezclar una pluralidad de pistas de audio de acuerdo con las enseñanzas que se divulgan en el presente documento.

25 La figura 1 muestra un diagrama de bloques esquemático de un mezclador de audio. El mezclador de audio permite combinar una pluralidad de pistas singulares ST de tal modo que se forme una señal de mezcla MS. Para controlar la combinación de las pistas singulares ST, cada pista singular por lo general se alimenta a un procesador de señal individual. El procesador de señal individual para una pista singular puede comprender por ejemplo, un ecualizador EQ, un control de paneo PAN, un reverberador REVERB, un control de volumen VOL, y posiblemente efectos de sonido adicionales. Un papel central del mezclador de audio es ajustar el volumen de cada una de la pluralidad de pistas de audio singulares ST de tal modo que la señal de mezcla sea una superposición bien equilibrada de las señales de audio proporcionadas por la pluralidad de pistas singulares ST. La decisión de qué ajuste particular de los efectos de sonido y los volúmenes de las pistas singulares ST constituye una superposición bien equilibrada, por lo general la toma un ingeniero de mezclado. La pluralidad de procesadores de señal individuales modifica la pluralidad de señales de pista de audio. Las señales de pistas de audio modificadas se combinan entonces en un combinador de señales Σ para producir la señal de mezcla MS.

40 La figura 2 ilustra una estructura de tiempo de una canción típica que pertenece al género de la música popular. La canción que se usa como ejemplo dentro de la figura 2 empieza con una sección de introducción (INTRO), seguida por una sección de verso (VERSO 1), un coro (CORO), una segunda sección de verso (VERSO 2) substancialmente con la misma música pero diferente letra, una repetición del coro, un puente (PUENTE), otra repetición del coro, y una coda u outro (OUTRO). A pesar de que existe una multitud de variaciones de este esquema básico, por lo general la mayoría de la gente es capaz de distinguir las diversas secciones de una canción de música popular. Por ejemplo, el coro por lo general se repite en varias ubicaciones a lo largo de la canción sustancialmente con la misma letra y melodía, de tal modo que puede ser reconocido fácilmente por un oyente.

45 La figura 3 ilustra una estructura de tiempo de una obra musical compuesta de forma de sonata. La forma de sonata ha sido usada por un gran número de compositores de música clásica. Tal como sugiere el nombre, la forma de sonata se usa ampliamente en sonatas, por lo general el primer movimiento de la misma. El primer movimiento de muchas sinfonías por lo general está también en forma de sonata. Son secciones características de la forma de sonata la exposición, el desarrollo y la recapitulación, en donde se presenta básicamente el mismo material musical con diversas modificaciones, en particular con respecto a la progresión de acordes. Opcionalmente, se pueden presentar una introducción y una coda en el comienzo y en el final de la obra musical, respectivamente. A pesar de que distinguir las diversas secciones de tiempo de la forma de sonata requiere algo de experiencia, esto en general es factible para un oyente humano.

50 Un ingeniero de mezclado podría querer tratar diferentes secciones de tiempo de una obra musical de diferentes maneras. La razón puede ser el deseo de lograr un cierto efecto artístico, o hacer que la señal de mezcla MS suene de manera más uniforme mediante la compensación de imperfecciones potenciales que puedan haber tenido lugar durante la grabación de la pluralidad de pistas de audio. El conocimiento acerca de la estructura de tiempo de la obra musical o una grabación de audio general (por ejemplo, audiolibro, ponencia, etc.) puede ayudar al ingeniero de mezclado a hallar los puntos de inicio y los puntos de finalización de las diversas secciones de tiempo en la grabación.

La figura 4 ilustra una distribución de pistas de audio a modo de ejemplo de una grabación de una canción en el

género musical popular. Existen pistas de audio singulares ST para los siguientes instrumentos: guitarra solista, guitarra rítmica, parte vocal, piano y bajo. Se ha grabado una batería usando varias pistas de audio singulares para las diversas partes de la batería: platillo de remate, platillo de ritmo, platos “Charleston”, tom toms, redoblante y bombo. El uso de varias pistas de audio ST para las diferentes partes de la batería permite que el ingeniero de
5 mezclado equilibre mejor la parte de tambor que cuando se hubiera usado solo una pista de audio singular para toda la batería.

Al tener la pluralidad de pistas de audio singulares disponibles, la obra musical se puede masterizar según desee el artista y / o el ingeniero de mezclado. En particular, el carácter o “humor” de una obra musical se puede alterar de
10 una manera significativa al cambiar los parámetros de mezclado que se usan para la pluralidad de pistas de audio ST. Proporcionar la pluralidad de pistas de audio ST para que un consumidor las mezcle ofrece al consumidor un alto grado de libertad. Sin embargo, muchos usuarios carecen de la experiencia para seleccionar apropiadamente los parámetros de mezclado, en particular debido a unas interconexiones e interacciones complicadas entre los parámetros de mezclado. Para lograr un cierto efecto que parezca afectar a una pista de audio singular, puede ser
15 necesario ajustar los parámetros de mezclado de varias, o incluso de todas, las pistas de audio.

La figura 5 muestra un diagrama de bloques esquemático de un mezclador de acuerdo con las enseñanzas que se divulgan en el presente documento que tiene una primera posible configuración.

20 Por lo general, el usuario (u oyente) tiene una cierta idea de cómo debe sonar la señal de mezcla, pero no sabe cómo se deben ajustar los parámetros de mezclado para lograr esa idea.

El mezclador de audio de acuerdo con las enseñanzas que se divulgan en el presente documento establece un vínculo entre una expresión semántica que describe la idea o el deseo del usuario de forma concisa, y los
25 parámetros de mezclado reales necesarios para mezclar la pluralidad de pistas de audio singulares ST para dar la señal de mezcla MS.

Un ejemplo simple pero ilustrativo de una descripción semántica que guía un proceso de mezclado es el siguiente: “Durante el solo de guitarra, mezclar la guitarra de forma prominente y mover los teclados ligeramente a un segundo
30 plano”.

Para llevar a cabo esto, por lo general es necesario abordar por lo menos algunas de las diversas subtareas que se enumeran en lo sucesivo:

- 35 - Es necesario que se capturen las descripciones semánticas dadas por el usuario usando una interfaz de usuario apropiada.
- Es necesario que se traduzca la entrada de usuario a una forma legible por máquina.
- Es necesario que se realice un análisis semántico de la señal de audio musical (por ejemplo, identificar la pista de guitarra y la pista de teclado, hallar el comienzo y el final del solo de guitarra).
- 40 - Es necesario que se determinen los parámetros de mezclado físicos al tener en cuenta el mecanismo de audición humana (por ejemplo, determinar la ganancia de tal modo que la intensidad sonora percibida de la guitarra en la mezcla sea más alta que la de cualquier otro instrumento, pero no demasiado alta; para los teclados, determinar la ganancia, el retardo y la ganancia de la pista de envío para el efecto de reverberación para la percepción deseada de distancia)
- 45 - La mezcla se obtiene usando los parámetros de mezclado físicos computados. Los parámetros comprenden factores de ganancia y retardos de tiempo para cada combinación de una pista de audio singular ST y canal de salida. Además, los parámetros de mezclado físicos controlan procesadores de efectos de audio digital (DAFx, *Digital Audio Effects*), por ejemplo, reverberación artificial y procesamiento de rango dinámico.

50 Las descripciones semánticas pueden, por ejemplo, especificar

- posición e intensidad sonora percibidas para cada objeto de sonido SO en la señal de mezcla MS
- parámetros del DAFx para cada pista
- características para la señal de mezcla MS (por ejemplo, la cantidad de reverberación, el rango dinámico).

55 En el diagrama de bloques esquemático de la figura 5 que muestra una posible configuración de un mezclador de audio de acuerdo con las enseñanzas que se divulgan en el presente documento, las subtareas que se han mencionado en lo que antecede se llevan a cabo mediante módulos del mezclador de audio. El mezclador de audio comprende una interfaz de usuario (UI, *user interface*) 20, un intérprete de comandos (CI, *command interpreter*) 30,
60 un análisis de audio semántico (SAA, *semantic audio analysis*) 40, una unidad de asignación de descriptor (DAU, *descriptor assignment unit*) objetivo 50, un procesador perceptual 60 y un procesador de señal 70.

La interfaz de usuario 20 proporciona facilidades para capturar una entrada procedente de un usuario del mezclador de audio. Existen diferentes opciones para la implementación de la entrada de usuario, tal como se ilustra mediante

una pluralidad de submódulos que son parte de la interfaz de usuario 20. Son ejemplos:

- la selección de uno de un conjunto de ajustes predeterminados (submódulo 22);
- un conjunto de controladores n-dimensionales que se asignan a diferentes características de las pistas singulares y la señal de mezcla resultante MS (submódulo 21);
- entrada de lenguaje natural (submódulo 24);
- entrada de un ejemplo de una señal de mezcla MS o un ejemplo de una de múltiples pistas junto con una señal de mezcla MS correspondiente (submódulo 23). El ejemplo dado se analizará a continuación para obtener la descripción semántica para la señal de mezcla MS. En la descripción subsiguiente, se hará referencia a un modo de operación del mezclador de audio que se controla mediante esta clase de entrada de usuario como "mezclado mediante ejemplo".

El intérprete de comandos 30 se conecta a la interfaz de usuario 20 y traduce la entrada (la cual es legible por un ser humano o es dada por ejemplos) a comandos legibles por máquina. Estos comandos por lo general tienen un vocabulario limitado y una gramática conocida que se proporciona y / o se controla mediante un submódulo de vocabulario - gramática 31.

En la tabla 1 se dan ejemplos para la implementación del intérprete de comandos 30, para diferentes diseños de interfaz de usuario.

Entrada	Implementación del intérprete de comandos 30
Ajustes predeterminados	Seleccionar un conjunto de reglas usando una consulta en tabla
Controlador n-dimensional	Establecer una correspondencia de la función de entradas de controlador con comandos
Mezclar mediante ejemplo	Análisis de señales de audio
Lenguaje natural	Reconocimiento y comprensión del habla

Además de la entrada de usuario, el mezclador de audio también recibe unos datos 10 que comprenden las pistas de audio singulares ST como una entrada. En el caso de que los datos 10 se correspondan con pistas de audio de una obra musical, los datos 10 pueden comprender un contenedor de música 11 y un contenedor de metadatos 12 opcional. Los datos 10 se pueden proporcionar al mezclador de audio por medio de una interfaz adecuada (que no se muestra en la figura 5).

Los datos 10 se alimentan al Análisis de Audio Semántico (SAA) 40. El análisis de audio semántico 40 por lo general es un proceso automático que calcula un conjunto de metadatos para cada una de las pistas de audio ST. Además, se pueden computar metadatos que describen las múltiples pistas, es decir, la pluralidad de pistas de audio (por ejemplo, género musical). Los metadatos son descriptores semánticos que caracterizan las señales de audio.

El análisis de audio semántico 40 puede comprender:

- identificación de instrumentos
- análisis estructural (etiquetado de verso, coro y otras partes de cada señal)
- identificación de estilo de ejecución (solo, acompañamiento, melódico, armónico y entropía rítmica)
- análisis rítmico (por ejemplo, rastreo de ritmo para efectos de sonido sincrónico de ritmo)
- análisis melódico y armónico
- caracterización del timbre (por ejemplo, brillo, aspereza, agudeza)
- caracterización de las similitudes (con respecto al timbre, el estilo de ejecución, la forma) de entre las pistas de audio singulares ST
- género musical

Estos metadatos se pueden usar para asignar el procesamiento de señales apropiado, por medio de los parámetros de mezclado, a cada una de las pistas singulares ST.

El análisis de audio semántico se puede realizar cada vez que se realiza el proceso de Mezclado Semántico. Opcionalmente, el análisis de audio semántico se puede realizar una vez (durante la producción / creación de contenidos) y los metadatos resultantes se pueden almacenar y transmitir junto con el artículo multimedia.

Opcionalmente, el análisis de audio semántico 40 puede estar guiado por entradas de usuario, es decir, el usuario puede ayudar al análisis de audio semántico 40 o este puede introducir metadatos si no está satisfecho con uno o más de los resultados obtenidos de forma automática del análisis de audio semántico. Estas entradas correctivas del usuario se pueden almacenar por el análisis de audio semántico para ser tenidas en cuenta durante análisis futuros de tal modo que el análisis de audio semántico 40 puede adaptarse a las preferencias de usuario, es decir, se

entrenado con el tiempo el análisis de audio semántico 40 por medio de las entradas del usuario.

5 El análisis de audio semántico 40 puede comprender un primer submódulo 41 para computar los metadatos basándose en las señales de audio contenidas en la pluralidad de pistas de audio ST. Adicionalmente o como alternativa, el análisis de audio semántico 40 puede comprender un segundo submódulo 42 para leer metadatos que se proporcionan junto con la pluralidad de pistas de audio ST.

10 Conectada con el intérprete de comandos 30 y con el análisis de audio semántico 40 se encuentra la unidad de asignación de descriptor (DAU) objetivo 50. Dados los comandos procedentes del intérprete de comandos 30 y los metadatos que se obtienen del análisis de audio semántico 40, la unidad de asignación de descriptor objetivo 50 selecciona partes de la señal de audio (esta determina las pistas y los tiempos de comienzo y los tiempos de parada que se corresponden con objetos de sonido para los cuales existen comandos) y asigna a los mismos unos descriptores objetivo perceptuales (PTD, *perceptual target descriptor*) apropiados.

15 El descriptor objetivo perceptual puede describir:

- la intensidad percibida de un objeto de sonido (intensidad sonora)
- atributos espaciales de un objeto de sonido (ángulo lateral, altura, distancia, grado de difusión, ancho)
- características de timbre (por ejemplo, brillo, agudeza, aspereza) para un objeto de sonido
- 20 - características en relación con efectos de audio digital (DAFx)

Si los comandos se dan usando variables lingüísticas, la unidad de asignación de descriptor objetivo 50 puede usar lógica difusa para la conversión entre variables lingüísticas para dar valores nítidos.

25 Una salida de la unidad de asignación de descriptor objetivo 50 que proporciona el descriptor objetivo perceptual se conecta a una entrada del procesador perceptual (PP) 60. El procesador perceptual 60 computa los parámetros físicos (parámetros de mezclado) para un mezclado y un procesamiento de señales adicional (por ejemplo DAFx) dado el descriptor objetivo perceptual asignado. Por lo general, esta es una tarea sumamente exigente que necesita tener en cuenta la psicoacústica 61 y un conocimiento experto 62.

30 Esto se ilustra con el siguiente ejemplo: Para una señal de audio particular, por ejemplo, una pista de guitarra, el descriptor para el nivel percibido se ajusta a "alto". Una solución simple es aumentar la ganancia de la guitarra una cantidad fija, por ejemplo 6 dB. Esta solución simple puede no tener el efecto deseado en todos los casos, debido a que la percepción de intensidad sonora depende de las características espectrales y temporales de la señal de audio procesada y de la señal de mezcla MS, de una manera muy compleja.

35 Se pueden obtener mejores resultados ajustando todos los niveles de tal modo que la intensidad sonora de la guitarra, la cual es percibida por el oyente, en la mezcla se ajuste, por ejemplo, mediante el uso de un modelo perceptual de intensidad sonora y de intensidad sonora parcial. La intensidad sonora parcial es la intensidad sonora de una señal presente en una mezcla de sonido, en donde la señal de interés está parcialmente enmascarada.

40 Por lo general es necesario abordar diferentes aspectos de la audición humana y la percepción del sonido además de la percepción de la intensidad sonora. Estos son la percepción de la cantidad de reverberación, la localización del sonido y la percepción de atributos espaciales.

45 La parte psicoacústica es importante para traducir la descripción semántica (por ejemplo, "dar un poco más de intensidad sonora a esto") a un parámetro físico (por ejemplo, "reforzar 4,5 dB").

50 El procesador perceptual 60 se conecta por medio de una de sus salidas a una entrada del procesador de señal (SP) 70. El procesador de señal 70 puede comprender un módulo que manipula los parámetros de mezclado físicos 71, uno o más efectos de audio digital 72 y un módulo para dar formato 73. Con los parámetros físicos para el mezclado y el procesamiento de señales, el procesador de señal 70 computa la señal de mezcla MS.

55 En el Documento de la Convención "Sistema de Producción Musical Automático Empleando Sistemas Expertos Probabilísticos" ("*Automatic Music Production System Employing Probabilistic Expert Systems*"), de la Audio Engineering Society, presentado en la 129ª Convención, de 4 a 7 de noviembre de 2010, los autores R. Gang y col. proponen emplear un modelo gráfico probabilístico para incorporar conocimiento de ingeniería de audio profesional e inferir decisiones de producción automática basándose en información musical extraída de archivos de audio. El patrón de producción, el cual se representa como modelo gráfico probabilístico, se puede aprender a partir de los datos de operación de un ingeniero de audio humano o se puede construir de forma manual a partir del conocimiento en el área. El procesador perceptual 60 y / o el intérprete de comandos semánticos 30 pueden implementar los rasgos distintivos técnicos propuestos en este Documento de Convención.

Mezclar una grabación de múltiples pistas comprende:

- Ajuste de niveles y posiciones de paneo para cada pista singular (módulo para manipular parámetros de mezclado físicos 71)
- Ecuación (para pistas singulares ST y la señal de mezcla MS)
- 5 - Procesamiento de Rango Dinámico (DPR) (para pistas singulares ST y la señal de mezcla MS)
- Reverberación Artificial
- Aplicación de efectos de sonido (DAFx 72)

10 Cada una de estas operaciones es controlada por los parámetros físicos según lo computado por el procesador perceptual 60.

Opcionalmente se requiere dar formato 73 para encargarse de las restricciones físicas (por ejemplo, aplicar un control de ganancia automático) y la conversión de formato (codificación / decodificación de audio).

15 La siguiente sección detalla una implementación a modo de ejemplo de cada uno de los bloques de procesamiento.

La interfaz de usuario 20 se puede implementar como un conjunto de ajustes predeterminados. Cada ajuste predeterminado representa un "tipo de mezclado" con un conjunto de características. Estas características pueden estar dadas como expresiones semánticas en forma de "reglas de mezclado", y se describen en lo sucesivo en el contexto de la descripción del intérprete de comandos 30.

20 Un tipo de mezclado puede ser, por ejemplo, la "Mezcla de Baile", la "Mezcla de Ambiente", la "Mezcla de Guitarra de Rock" y otras.

25 Estos nombres dan una descripción de la señal de mezcla MS objetivo de una manera muy comprimida, pero el usuario puede interpretar estos (o un subconjunto de los mismos). La capacidad del usuario para interpretar los nombres de los ajustes predeterminados se basa en convenciones y clasificaciones estilísticas ampliamente usadas. Por ejemplo, un usuario puede asociar un estilo de ejecución y / o sonido específico con el nombre de un cierto artista.

30 Dentro del contexto del intérprete de comandos 30, se asigna un conjunto de reglas de mezclado a cada uno de los ajustes predeterminados usando una tabla de consulta. Las reglas de mezclado están representadas como implicaciones lógicas en la forma de sentencias SI - ENTONCES, como en la Lógica Difusa (J. M. Mendel, "Tutorial de Sistemas de Lógica Difusa para Ingeniería" ("*Fuzzy Logic Systems for Engineering: A Tutorial*"), Actas de IEEE, vol. 83, págs. 345 - 377, 1995), tal como se muestra en el presente documento, en donde un descriptor de objeto de sonido <SOD> es el antecedente y un descriptor de operación de mezclado <MOD> es el consecuente:

SI <SOD> ENTONCES <MOD>

40 Las sentencias SI - ENTONCES especifican

- Cómo aparecen los objetos de sonido en la señal de mezcla MS, expresado como descriptores de operación de mezclado (MOD). Los MOD se seleccionan de acuerdo con características de los objetos de sonido, dadas por los descriptores de objeto de sonido (SOD).
- Características de señal de mezcla MS que son independientes de un descriptor de operación de mezclado MOD específico, y especifican los parámetros de las operaciones para la señal de mezcla MS.

Un descriptor de objeto de sonido SOD puede estar representado como una estructura (de datos), por ejemplo:

SO.ID	Identificador de objeto de sonido, por ejemplo, nombre del ejecutante
SO.INSTR	Clase de instrumento del objeto de sonido SO
SO.BRIGHTNESS	Brillo percibido del objeto de sonido SO
SO.PERCUSSIVENESS	Cuantificador para la percusividad del SO
SO.CHARACTERISTIC	Otra característica del objeto de sonido SO

50 Los descriptores de operación de mezclado MOD describen el nivel (es decir, el volumen), la posición de paneo, la distancia y otras características de un objeto de sonido SO que se puede percibir en una señal de mezcla MS. Los descriptores de operación de mezclado MOD que se aplican a un objeto de sonido SO pueden estar designados por SO.MOD dentro de la estructura de datos. Los descriptores de operación de mezclado MOD también pueden ser aplicados a la señal de mezcla MS. Estos descriptores de operación de mezclado MOD están designados por MT.MOD. Por lo general, estos descriptores de operación de mezclado MOD controlan el procesamiento de señales que se aplica a todas las señales de audio o a la señal de mezcla, por ejemplo, reverberación o procesamiento de rango dinámico DPR (*dynamic range processing*).

Un descriptor de operación de mezclado MOD puede consistir de un atributo perceptual y un valor el cual se asigna al atributo perceptual. Los descriptors de operación de mezclado se pueden implementar como variables lingüísticas.

5 Una lista de atributos perceptuales puede contener los siguientes (aparte de otros):

<i>Atributo Perceptual</i>	<i>Descripción</i>
PLOUDLEVEL	nivel percibido
LATERALDISPLACEMENT	el ángulo lateral con el cual la posición del sonido se desvía del centro
PDISTANCE	la distancia a la cual se percibe el objeto de sonido SO
FX1LEVEL	cantidad percibida de DAFx 1
FX2LEVEL	cantidad percibida de DAFx 2
REVERBLEVEL	cantidad percibida de reverberación
BRIGHTNESS	descriptor de timbre
DIFFUSENESS	describe cómo se mezcla el sonido difuso frente al directo

Los atributos perceptuales pueden ser variables lingüísticas. Los valores asignados pueden ser uno de los siguientes: {"Muy bajo", "Bajo", "Medio", "Alto", "Muy Alto"}.

10 Los atributos perceptuales que no son ajustados por el descriptor de operación de mezclado MOD se ajustan a valores por defecto.

Entonces, una regla de mezclado puede tener el siguiente aspecto:

15 SI <SO.INSTR=valor> Y <SO.C1=valor> ... Y <SO.Cn=valor> ENTONCES <SO.MOD1=valor> Y <SO.MOD2=valor> ... Y <SO.MODn=valor>

20 Se debe hacer notar que el uso de una conjunción (es decir, "Y") es suficiente y una disyunción (es decir, "O") puede expresarse como reglas separadas.

Conjunto de reglas a modo de ejemplo: se da un conjunto de reglas de mezclado para el caso de uso en cuestión para el ejemplo de la Mezcla de Baile:

25 Estas reglas de mezclado se especifican para clases de instrumento:

1. SI <SO.INSTR="bombo">
ENTONCES <SO.PLOUDLEVEL="alto"> Y <SO.LATERALDISPLACEMENT="cero"> Y <SO.DISTANCE="cerca">
2. SI <SO.INSTR="bajo">
ENTONCES <SO.PLOUDLEVEL="alto"> Y <SO.LATERALDISPLACEMENT="cero"> Y <SO.DISTANCE="cerca">
3. SI <SO.INSTR="percusión"> Y <SO.ENTROPY="alto">
ENTONCES <SO.PLOUDLEVEL="alto"> Y <SO.FXLEVEL="alto">
4. SI <SO.INSTR="percusión"> Y <SO.ENTROPY="bajo">
ENTONCES <SO.PLOUDLEVEL="bajo">

35 Las siguientes reglas de mezclado se especifican para características independientemente de la clase de instrumento:

5. SI <SO.INSTR="*"> Y <SO.ENTROPY="bajo">
ENTONCES <SO.LATERALDISPLACEMENT="extremo izquierdo">
6. SI <SO.INSTR="*"> Y <SO.CREST="bajo">
ENTONCES <SO.PLOUDLEVEL="bajo">

45 Opcionalmente, se pueden especificar reglas de mezclado para la señal de mezcla MS. Estas no están vinculadas a características de los objetos de sonido SO. Las operaciones resultantes se aplican a todos los objetos de sonido, si no se especifica objeto de sonido alguno en la parte correspondiente a SI de la regla de mezclado.

7. SI *
ENTONCES <MS.REVERBLEVEL="bajo">
8. SI *
ENTONCES <MS.FX1LEVEL="alto">

50

Además, en la parte correspondiente a SI de las reglas, los atributos también se pueden comparar con valores relativos en lugar de con valores absolutos. Esto significa que un atributo de un objeto de sonido SO se puede comparar con el mismo atributo de todos los otros objetos de sonido SO usando operaciones como “máximo” o “mínimo”, por ejemplo

5
 9. SI <SO.INSTR=“*”> Y <SO.ENTROPY=“máximo”>
 ENTONCES <SO.FX2LEVEL=“alto”>

10 Se debe hacer notar que los atributos y las reglas que se han enumerado en lo que antecede son ejemplos y no se pretende que sean el conjunto completo para el ajuste predeterminado de mezclado particular.

15 De acuerdo con un aspecto de las enseñanzas que se divulgan en el presente documento, se puede realizar una variación del conjunto de reglas. En particular, el conjunto de reglas se puede implementar para contener reglas concurrentes (reglas con el mismo antecedente pero diferente consecuente) de las cuales una se selecciona de forma arbitraria (se aleatoriza). Esto introduce variaciones en los resultados y, de ese modo, aumenta la satisfacción del usuario. Esto también es beneficioso en situaciones en las que no se puede acordar conjunto de reglas uniforme alguno en el proceso de producir el conjunto de reglas.

20 Pasando a continuación a una implementación a modo de ejemplo del análisis de audio semántico 40, recuérdese que el análisis de audio semántico 40 se aplica para recopilar la información acerca de la pluralidad de pistas de audio ST y posiblemente las múltiples pistas que pueden ser útiles para determinar qué parámetros de mezclado se han de aplicar a cuál de la pluralidad de pistas de audio ST. El análisis de audio semántico procesa cada pista de audio ST de la pluralidad de pistas de audio por separado y adicionalmente una representación de las múltiples pistas MT. La representación de las múltiples pistas MT se puede obtener, por ejemplo, en forma de una señal de mezcla que se obtiene a partir de un mezclado descendente de todas las pistas de audio ST con ganancias unidad.

30 Los resultados se pueden representar como una matriz de estructuras (en donde cada elemento de la matriz contiene los metadatos para una pista de audio ST) y una estructura adicional conteniendo los metadatos de las múltiples pistas. Los tipos variables de los elementos de estructura pueden ser cadenas (por ejemplo, nombres de instrumento), valores escales (por ejemplo, tiempo, entropía), o matrices (por ejemplo, para tiempos de comienzo y tiempos de parada para la descripción de estilos de ejecución), o estructuras dedicadas para sí mismas (por ejemplo, una estructura para describir la forma de una pieza musical).

35 Un resultado del análisis puede estar acompañado por una medida de confianza, la cual representa el grado de fiabilidad del resultado respectivo.

Ejemplo para la representación de un resultado producido por el análisis de audio semántico 40:

40 ST(1).ID = “TR909”
 ST(1).INSTRUMENT = “bombo”
 ST(1).INSTRUMENT_CONFIDENCE = 0.93
 ST(1).ENTROPY = 0.12

45 ST(2).ID = “guitarra solista”
 ST(2).INSTRUMENT = “guitarra”
 ST(2).INSTRUMENT_CONFIDENCE = 0.68
 ST(2).SOLO = [[123.4 234.5] [567.7 789.0]]

50 ST(3).ID = “coros”
 ST(3).INSTRUMENT = “ser humano cantando”
 ST(3).INSTRUMENT_CONFIDENCE = 0.8
 ST(3).BRIGHTNESS = 0.12

55 MT.TEMPO=“126”
 MT.GENRE=“electro”
 MT.FORM=<estructura de forma>

60 El análisis de audio semántico 40 puede ser útil para normalizar el material de audio de múltiples pistas mediante la asignación de identificadores únicos a las pistas de audio ST y a las diversas secciones de tiempo de la obra musical. En particular, el material de audio de múltiples pistas por lo general no es un formato predefinido que sigue una cierta convención. Dicho de otra forma, el mezclador de audio no puede basarse en que esa pista de audio específica (por ejemplo, “pista 1”) contenga siempre un cierto instrumento (por ejemplo, “guitarra”). Los metadatos producidos por el análisis de audio semántico, sin embargo, pueden proporcionar información normalizada acerca de la organización y el contenido de la señal de múltiples pistas que ayuda a otros módulos del mezclador de audio a

llevar a cabo sus tareas respectivas. La normalización realizada por el análisis de audio semántico es útil debido a que permite relacionar el comando de mezclado proporcionado por el intérprete de comandos 30 con la situación encontrada de la señal de audio de múltiples pistas. De este modo, el intérprete de comandos 30 y el análisis de audio semántico 40 “hablan el mismo idioma”.

5 La unidad de asignación de descriptor DAU objetivo 60 procesa los metadatos proporcionados por el análisis de audio semántico 40 y las reglas de mezclado procedentes del intérprete de comandos 30 para asignar descriptores de operación de mezclado a la pluralidad de pistas de audio ST o a segmentos de las pistas de audio ST. Estos descriptores indican cómo se percibe, en la señal de mezcla MS objetivo, cada objeto de sonido SO que es
10 dominante en el segmento respectivo de la pista de audio ST.

Se supone que solo un objeto de sonido es dominante en cada instante en cada pista de audio ST. Dada esta suposición, los atributos que se obtienen a partir del análisis de audio semántico 40 (los cuales se computan para cada pista de audio ST) se procesan como atributos para el objeto de sonido SO. Como alternativa, el análisis de audio semántico puede emitir más de una estructura de atributos para cada pista de audio ST si la pista de audio ST
15 contiene múltiples objetos de sonido, en especial si los diversos objetos de sonido SO se suceden temporalmente entre sí dentro de la pista de audio ST, lo cual significa que los diversos objetos de sonido SO se pueden separar con relativa facilidad. Otra posibilidad es que un primer objeto de sonido SO1 esté presente principalmente en el canal izquierdo de una señal estéreo, al tiempo que un segundo objeto de sonido SO2 está presente principalmente en el
20 canal derecho. Otra posibilidad más sería que los diversos objetos de sonido se pudieran separar en el dominio de la frecuencia por medio de filtros de paso bajo, de paso alto y / o de paso de banda.

Se puede aplicar Lógica Difusa si las variables introducidas son valores nítidos, pero el conjunto de reglas se formula usando atributos difusos (por ejemplo, “bajo”, “alto”). Por ejemplo, el grado de variación en la ejecución de un instrumento se puede expresar como un valor escalar en el rango entre 0 y 1. Además, el análisis de audio semántico 40 puede emitir los metadatos junto con valores de confianza (por ejemplo, probabilidades) describiendo el grado de confianza con el cual se han computado los metadatos estimados.
25

La Lógica Difusa permite tareas de modelado complejas, que incorporan a menudo un conocimiento experto. Esta hace uso de Conjuntos Difusos, los cuales proporcionan un mecanismo directo para convertir, tanto en uno como en otro sentido, valores precisos y descripciones difusas.
30

En el diagrama de bloques en la figura 6 se muestra una visión de conjunto del procesamiento si se implementa como Sistema de Lógica Difusa (Mendel, 1995). El Sistema de Lógica Difusa comprende un módulo de difusificación 622, un módulo de inferencia 624, un conjunto de reglas 626 y un módulo de desdifusificación 628. El módulo de difusificación 622 recibe un conjunto de entradas nítidas, por ejemplo, a partir del análisis de audio semántico 40. Sobre la base de la entrada nítida, el módulo de difusificación 622 produce un conjunto de entrada difusa el cual se alimenta al módulo de inferencia 624. El módulo de inferencia 624 evalúa el conjunto de entrada difusa por medio de un conjunto de reglas 626 que se alimenta igualmente al módulo de inferencia 624. El conjunto de reglas 626 puede ser proporcionado por el intérprete de comandos 30. El módulo de inferencia 624 produce un conjunto de salidas difusas y lo alimenta al módulo de desdifusificación 628. En el módulo de desdifusificación 628 el conjunto de salidas difusas se traduce a salidas nítidas las cuales se pueden encontrar entonces como los parámetros de mezclado o como cantidades intermedias.
35
40

45 Pasando a continuación a la difusificación en más detalle, la asignación de descriptores de operación de mezclado MOD a las pistas de audio singulares ST se realiza sobre la base de los criterios que se describen en la parte correspondiente a SI del conjunto de reglas determinado por el intérprete de comandos 30. Si los metadatos respectivos procedentes del análisis de audio semántico 40 se dan como números reales o como cadenas junto con un valor de confianza (por ejemplo, como resultado de la clasificación de instrumento), los números reales se traducen a variables lingüísticas usando Difusificación. Los Conjuntos Difusos son conjuntos cuyos elementos tienen un grado de pertenencia. El grado de pertenencia puede ser cualquier número real en el intervalo [0, 1] (en contraste con la teoría de conjuntos clásica en la que el grado de pertenencia es o bien 0 o bien 1).
50

La Difusificación se realiza usando funciones de pertenencia para el Conjunto Difuso tal como se muestra a modo de ejemplo en la figura 7. En la Difusificación, para cada variable de entrada de valor real, se determina el Conjunto Difuso correspondiente (I. A. Zadeh, “Conjuntos Difusos, Información y Control” (“*Fuzzy Sets, Information and Control*”), vol. 8, págs. 338 - 353, 1965) y el grado de pertenencia. Por ejemplo, dado un valor de brillo de 0,25, los Conjuntos Difusos correspondientes son “muy bajo” con una pertenencia de 0,5 y “bajo” con una pertenencia de 0,5.
55

60 En la etapa o módulo de Inferencia 624, se establece una correspondencia de Conjuntos Difusos para las variables de entrada con Conjuntos Difusos para las variables de salida usando el conjunto de reglas 626. El resultado es, una vez más, un conjunto de variables lingüísticas (junto con unos grados de pertenencia de confianza correspondientes) para los atributos perceptuales.

En la siguiente etapa o módulo, la Difusificación 628, los resultados de la inferencia se convierten en valores nítidos para las variables de salida usando sus Conjuntos Difusos correspondientes. Es decir, las variables que se enumeran en la tabla anterior de atributos perceptuales tienen unos homólogos con valores nítidos.

5 Con respecto al procesador perceptual 60, las salidas del intérprete de comandos 30 y la unidad de asignación de descriptor objetivo 50 determinan cómo debería aparecer cada uno de los objetos de sonido SO en la señal de mezcla MS. Hasta ahora, esa especificación se da por medio de los valores perceptuales.

10 El procesador perceptual 60 traduce los valores perceptuales a parámetros de mezclado físicos teniendo en cuenta las características de señal y mecanismos de audición humana. Los siguientes párrafos ilustran el procesamiento de algunos valores perceptuales, en concreto, niveles de sonido, coeficientes de paneo para ángulos laterales dados, niveles de reverberación y retardos de tiempo, parámetros de DAFx, ecualización y procesamiento de rango dinámico.

15 Los niveles de sonido para los objetos de sonido SO se pueden computar usando un modelo de intensidad sonora perceptual, por ejemplo, el modelo que describió Glasberg en 2002.

20 Como alternativa, se puede usar el modelo de intensidad sonora que describió Moore en 1996 para computar la intensidad sonora de una señal de sonido dentro de mezclas de señales de sonido (B. C. J. Moore y B. R. Glasberg, "Una Revisión del Modelo de Intensidad Sonora de Zwicker" ("A Revision of Zwicker's Loudness Model"), *Acta Acustica - Acta Acustica*, vol. 82, págs. 335 - 345, 1996).

25 Se computan factores de ganancia para cada pista de audio ST de tal modo que la intensidad sonora percibida del objeto de sonido SO en la pista de audio ST (o la señal de mezcla MS) coincida con la descripción semántica tal como se expresa mediante el descriptor de operación de mezclado MOD.

30 Coeficientes de paneo para ángulos laterales dados: la percepción de posición lateral de un objeto de sonido SO está determinada por diferencias de nivel interaural (ILD, *inter-aural difference*) y diferencias de tiempo interaural (ITD, *inter-time difference*) en las entradas a los oídos (Lord Rayleigh, "Acerca de nuestra Percepción de la Dirección del Sonido" ("On our perception of Sound Direction"), *Philosophical Magazine*, vol. 6, págs. 214 - 232, 1907). Dentro del contexto del procesador perceptual 60, se determinan retardos de tiempo y diferencias de nivel para cada canal de reproducción de tal modo que se evoca la percepción de lateralización.

35 Niveles de reverberación y retardos de tiempo: se determinan los niveles para los procesadores de reverberación artificial de tal modo que la cantidad de reverberación percibida coincide con las descripciones semánticas dadas por el usuario. Se definen niveles de reverberación para cada objeto de sonido por separado y / o para la señal de mezcla MS. Se pueden ajustar los niveles de reverberación para cada objeto de sonido por separado con el fin de evocar la percepción de distancia para un objeto de sonido SO particular. La percepción de distancia se controla adicionalmente mediante el nivel, el retardo de tiempo, la curva de ecualización y la posición lateral.

40 Parámetros de DAFx: el ajuste de los parámetros para los efectos de audio digital depende del procesador de DAFx particular. El nivel de la señal procesada por DAFx se calcula usando un modelo de intensidad sonora (por ejemplo, Moore, 1996).

45 Ecualización: se ajustan los parámetros para Ecualización de tal modo que las señales procesadas coinciden con los atributos perceptuales en relación con el "brillo" del objeto de sonido o la señal de mezcla MS.

50 Procesamiento de rango dinámico: se ajustan parámetros para el procesamiento de rango dinámico DPR para que coincidan con atributos perceptuales para el rango dinámico.

55 La figura 8 muestra un diagrama de bloques esquemático de una parte de un mezclador de audio que comprende un procesador difuso 37. Una entrada del procesador difuso 37 está conectada con el análisis de audio semántico 40 y está configurada para recibir valores de análisis de pista por medio de esta conexión. Los valores de análisis de pista pueden ser o bien valores nítidos o bien variables lingüísticas. El procesador difuso 37 también tiene una entrada para recibir reglas o conjuntos de reglas a partir del intérprete de comandos semánticos 35. Tal como se ha explicado en lo que antecede, el procesador difuso 37 usa las reglas para procesar los valores de análisis de pista lo cual da como resultado unos parámetros de mezclado nítidos que se pueden proporcionar al procesador de pistas de audio 75.

60 Las reglas son creadas por el intérprete de comandos semánticos 35 sobre la base del comando de mezclado semántico proporcionado por el usuario.

Un modelo perceptual 64 proporciona parámetros de difusificación y de desdifusificación al procesador de lógica difusa 37. Los parámetros de difusificación y de desdifusificación establecen un vínculo entre los valores numéricos

y las descripciones correspondientes. Por ejemplo, los parámetros de difusificación y de desdifusificación pueden especificar rangos de intensidad sonora para señales de audio que parecen suaves, medias o fuertes al oyente.

5 Además, el modelo perceptual 64 puede especificar qué parámetros de mezclado están involucrados cuando se desea un cierto efecto. Valores correspondientes para estos parámetros de mezclado también pueden ser especificados por el modelo perceptual 64. Estas especificaciones se pueden proporcionar al intérprete de comandos semánticos 35 como directrices. El intérprete de comandos semánticos 35 puede seguir estas directrices cuando se crean las reglas difusas.

10 El mezclador de audio puede comprender un selector de reglas difusas aleatorias 38 el cual se usa cuando han sido creadas dos reglas difusas concurrentes por el intérprete de comandos semánticos 35 y solo una puede ser implementada por el procesador de lógica difusa 37. Un moderado grado de aleatoriedad puede aumentar la satisfacción del usuario, debido a que el proceso de mezclado parece ser más natural y “humano”. Después de todo, un ingeniero de mezclado humano también puede actuar en ocasiones de una forma ligeramente aleatoria, lo cual
15 puede ser percibido como “artístico” por un cliente del ingeniero de mezclado.

La figura 9 muestra un diagrama de bloques esquemático de una posible configuración básica de un mezclador de audio de acuerdo con las enseñanzas que se divulgan en el presente documento. Los datos 10 se proporcionan en la forma de una pluralidad de pistas de audio singulares ST. El mezclador de audio comprende un intérprete de comandos semánticos 35, un procesador de pistas de audio 75 y un combinador de pistas de audio (AT CMB, *audio track combiner*) 76.
20

El intérprete de comandos semánticos 35 se corresponde, en términos generales, con el intérprete de comandos 30 de la figura 5. Además, el intérprete de comandos semánticos 35 puede comprender alguna funcionalidad del módulo de asignación de descriptor objetivo 50 y el procesador perceptual 60. El intérprete de comandos semánticos 35 recibe un comando de mezclado semántico como una entrada y obtiene un parámetro de mezclado o una pluralidad de parámetros de mezclado a partir del comando de mezclado semántico. La pluralidad de parámetros de mezclado se proporcionan al procesador de pistas de audio 75 o, para ser más precisos, a unos procesadores de pistas de audio individuales ATP1, ATP2, ATP3, ATP N del procesador de pistas de audio 75. Los parámetros de mezclado están por lo general en forma de valores nítidos los cuales pueden ser fácilmente implementados por la pluralidad de procesadores de pistas de audio individuales ATP1 a ATP N.
25
30

La pluralidad de procesadores de pistas de audio individuales ATP1 a ATP N modifican las señales de audio proporcionadas por las correspondientes de las pistas de audio ST1 a ST N de acuerdo con los parámetros de mezclado.
35

Las señales de audio son combinadas por el combinador de pistas de audio 76 para obtener la señal de mezcla MS.

En la configuración que se muestra en la figura 9, el intérprete de comandos semánticos 35 es capaz de asignar un contenido semántico particular dentro del comando de mezclado semántico al parámetro de mezclado apropiado para el procesador de pistas de audio individuales ATP1 a ATP N correspondiente. Esta capacidad del intérprete de comandos semánticos 35 se puede basar en el hecho de que la pluralidad de pistas de audio singulares ST1 a ST N están organizadas de acuerdo con una norma acordada de tal modo que el intérprete de comandos semánticos 35 puede saber qué pista se corresponde con qué instrumento. En las figuras 11 a 14, se representan y se describen configuraciones alternativas del mezclador de audio en la partes correspondientes de esta descripción que son capaces de obtener información acerca de la organización de la grabación de múltiples pistas y / o una estructura de tiempo de la obra musical grabada a partir de los propios datos.
40
45

La figura 10 ilustra un comando de mezclado semántico. El comando de mezclado semántico comprende una expresión lingüística en forma de una oración en idioma castellano. Por supuesto, también se pueden usar otros idiomas. La oración dice: “Durante el solo de guitarra, mezclar la guitarra de forma prominente”. Un análisis semántico de esta oración revela que la oración se puede descomponer en tres partes. Una primera parte contiene la expresión “durante el solo de guitarra” y se puede identificar como una expresión que especifica una sección de tiempo objetivo para el comando de mezclado semántico. Una segunda parte contiene la expresión “la guitarra” y se puede identificar como una expresión que especifica una pista objetivo. Una tercera parte contiene la expresión “mezclar [...] de forma prominente” y se puede identificar como una expresión que especifica una operación de mezclado deseada.
50
55

La figura 11 ilustra un ejemplo extendido de un comando de mezclado semántico. El comando de mezclado extendido se basa en el comando de mezclado semántico de la figura 10. Además, se ha añadido una segunda operación de mezclado semántico para una segunda pista objetivo, en concreto, “[...] mover los teclados ligeramente a un segundo plano”. Se usa una conjunción para especificar la relación entre la primera operación de mezclado / primera pista objetivo y la segunda operación de mezclado / segunda pista objetivo. En el caso ilustrado, la conjunción es la palabra “y” de tal modo que tanto la primera operación de mezclado como la segunda operación de
60

mezclado se realizan de forma concurrente sobre sus pistas objetivo respectivas.

5 La figura 12 muestra un diagrama de bloques esquemático de una parte de un mezclador de audio de acuerdo con otra posible configuración. En particular, la figura 12 muestra cómo se pueden usar los datos proporcionados por la pluralidad de señales de audio ST1 a ST N y por una señal de mezcla MT (“múltiples pistas”) por defecto para obtener información útil acerca de la disposición de pistas y / o la estructura de tiempo de la obra musical. A menos que se indique lo contrario, una referencia a la pluralidad de pistas de audio incluirá una referencia a la señal de mezcla MT por defecto.

10 La pluralidad de pistas de audio ST1 a MT se proporciona al análisis de audio semántico 40. Mediante el análisis de la pluralidad de pistas de audio se puede obtener información de pistas e información de estructura de tiempo, las cuales se proporcionan a un módulo de conversión de semántico a nítido 65.

15 El comando de mezclado semántico comprende una pluralidad de expresiones, comprendiendo cada expresión especificar una sección de tiempo objetivo 26, una pista objetivo 27 y una operación de mezclado 28.

20 El módulo de conversión de semántico a nítido 65 se corresponde aproximadamente con la unidad de asignación de descriptor objetivo 50 de la figura 5. El módulo de conversión de semántico a nítido 65 también recibe información del comando de mezclado semántico como una entrada. Sobre la base de las entradas proporcionadas, el módulo de conversión de semántico a nítido 65 crea uno o más descriptores objetivo perceptuales PTD y los parámetros de mezclado correspondientes. El descriptor objetivo perceptual PTD puede contener identificadores de pista de las pistas de audio afectadas ST1 a ST N, así como una información de sección de tiempo en el caso de que solo una sección de tiempo de la pista o pistas de audio afectadas esté afectada por el comando de mezclado. Obsérvese que los parámetros de mezclado pueden ser o bien valores nítidos o bien variables lingüísticas que se van a resolver en una fase posterior.

30 El análisis de audio semántico 40 puede recibir opcionalmente la especificación de sección de tiempo objetivo 26 y / o la especificación de pista objetivo 27 como una entrada, de tal modo que el análisis de audio semántico 40 puede analizar la pluralidad de pistas de audio ST1 a MT con especial atención a las especificaciones proporcionadas.

La figura 13 muestra un diagrama de bloques esquemático de otra configuración posible del mezclador de audio de acuerdo con las enseñanzas que se divulgan en el presente documento. La configuración presenta un identificador de pista de audio 430.

35 La estructura básica de la configuración de la figura 13 es sustancialmente la misma que en la figura 9; sin embargo, algunas partes se han omitido por razones de claridad.

40 Debido a que no siempre es inmediatamente evidente qué pista de audio ST1 a ST N contiene qué instrumento o parte vocal, se puede usar el identificador de pista de audio 430 para determinar esta información. El identificador de pista de audio 430 puede ser una parte del análisis de audio semántico 40. El comando de mezclado semántico comprende una identificación de pista de audio objetivo 27 tal como se ha mencionado en relación con la figura 12. La identificación de pista de audio objetivo 27 se proporciona a una base de datos de plantillas de pistas de audio 432. La base de datos de plantillas de pistas de audio 432 proporciona uno o más registros de datos que se corresponden con la identificación de pista de audio objetivo 27 y proporciona este o estos al identificador de pista de audio 432. El registro de datos puede comprender información acerca de, por ejemplo, un instrumento en forma de valores de medición, muestras de sonido, etc. El identificador de pista de audio 430 puede comparar entonces la información contenida en el registro de datos con las señales de audio de cada una de la pluralidad de pistas de audio ST1 a ST N. Con este fin, el identificador de pista de audio puede, por ejemplo, realizar una correlación cruzada de una muestra de sonido procedente del registro de datos con una sección corta de la señal de pista de audio. Otra opción sería determinar la ubicación y magnitud de los sobretonos de la señal de pista de audio y comparar el resultado con los datos correspondientes en el registro de datos. Otra opción más es dada por el análisis y la comparación de un comportamiento de ataque - decaimiento - sostenido - relajación de la señal de pista de audio.

55 El identificador de pista de audio genera una información de identificación de pista la cual se proporciona al procesador de pistas de audio 75 de tal modo que el procesador de pistas de audio 75 pueda procesar cada pista de audio singular ST1 a ST N de acuerdo con una indicación dada, por ejemplo, por un nombre de instrumento dentro del comando de mezclado semántico.

60 La figura 14 muestra un diagrama de bloques esquemático de otra posible configuración del mezclador de audio en el que el identificador de sección de tiempo 460 extrae una información de sección de tiempo de la pluralidad de pistas de audio ST 1 a MT. El identificador de sección de tiempo 460 está conectado con la pluralidad de pistas de audio ST1 a MT y está configurado para analizar una estructura de tiempo de la obra musical que es presentada por las pistas de audio ST1 a MT. En particular, el identificador de sección de tiempo 460 puede buscar secciones similares

o substancialmente idénticas dentro de la obra musical. Si la obra musical pertenece al género de la música popular, es probablemente que estas secciones similares o substancialmente idénticas sean el coro de la canción. El identificador de sección de tiempo 460 también puede contar latidos o pulsos de la obra musical lo cual puede mejorar la precisión de la identificación de sección de tiempo.

5 La información de sección de tiempo se proporciona al intérprete de comandos semánticos 35 el cual la usa para traducir una expresión de sección de tiempo semántica usada dentro del comando de mezclado semántico, a valores de tiempo de comienzo y de finalización de sección nítidos.

10 El análisis de la estructura de tiempo de una obra musical realizado por la identificador de sección de tiempo puede emplear uno o más métodos propuestos por diversos investigadores en el pasado. En su artículo "Resumen Automático de Música Basándose en Análisis de Estructura Musical" (*"Automatic Music Summarization Based on Music Structure Analysis"*), ICASSP 2005, Xi Shao y col., sugieren un enfoque novedoso para el resumen de música basándose en un análisis de estructura musical. En particular, en primer lugar se extrae el inicio de la nota de la
15 señal de audio para obtener el tiempo de tiempos de la canción. El análisis de estructura musical se puede realizar basándose en esta información de tiempo. Después de que se haya estructurado el contenido musical en diferentes regiones tales como Introducción (Intro), Verso, Coro, Final (Outro), etc., se puede crear el resumen musical final con frases de coro y música que se incluyen antes o después del coro seleccionado para conseguir la longitud deseada del resumen final. El análisis de estructura musical distingue entre regiones de similitud basadas en
20 melodía (versos) y regiones de similitud basadas en contenido (coro).

En el documento "Detección de Coro con Uso Combinado de MFCC y Rasgos Distintivos de Cromo y Filtros de Procesamiento de Imagen" (*"Chorus Detection with Combined Use of MFCC and Chroma Features and Image Processing Filters"*), Actas de la 10ª Conferencia Internacional sobre Efectos de Audio Digital (DAFx-07), Burdeos,
25 Francia, 10 - 15 de septiembre de 2007, el autor Antti Eronen describe un método computacionalmente eficiente para detectar una sección de coro en música rock popular. El método utiliza una representación de matriz de distancia que se obtiene mediante la suma de dos matrices de distancia separadas que se calculan usando los coeficientes cepstrales en las frecuencias de Mel y rasgos distintivos de croma de altura.

30 Mark Levy y col. son los autores de un artículo "Extracción de Estructura Musical de Alto Nivel a partir de Datos de Audio y su Aplicación a la Generación de Imágenes en Miniatura" (*"Extraction of High-Level Musical Structure from Audio Data and its Application to Thumbnail Generation"*), ICASSP 2006. En el artículo se presenta un método para segmentar audio musical con un modelo jerárquico de timbre. Se presentan nuevas evidencias para mostrar que la segmentación de música se puede reformular como un agrupamiento de rasgos distintivos de timbre, y se describe
35 un nuevo algoritmo de agrupamiento.

En el documento "Un Método de Detección de Secciones de Coro para Señales de Audio Musical y Su Aplicación a una Estación de Escucha de Música" (*"A Chorus Section Detection Method for Musical Audio Signals and Its Application to a Music Listening Station"*), Interacciones sobre Procesamiento de Audio, Voz, Habla y Lenguaje de
40 IEEE (*"Transactions on Audio, Speech, and Language Processing"*), Vol. 14, n.º 5, septiembre de 2006, el autor Masataka Goto describe un método para obtener una lista de secciones de coros repetidos ("gancho") en grabaciones de discos compactos de música popular. En primer lugar, se extrae un vector de rasgos distintivos de dimensión 12 que se denomina vector de croma, el cual es robusto con respecto a cambios de acompañamientos, de cada trama de una señal de entrada y a continuación se calcula la similitud entre estos vectores. Las secciones
45 que se identifican como secciones repetidas, se enumeran y se integran. El método puede detectar incluso secciones de coro moduladas mediante la introducción de un rasgo distintivo acústico perceptualmente motivado y una similitud que posibilita la detección de una sección de coro repetida incluso después de la modulación.

Una visión de conjunto de los métodos de análisis estructural musical automático conocidos ha sido recopilada por
50 Bee Suang Ong en su tesis "Análisis Estructural y Segmentación de Señales Musicales" (*"Structural Analysis and Segmentation of Music Signals"*), Universitat Pompeu Barcelona, 2007, ISBN 978-84-691-1756-9.

La figura 15 muestra un diagrama de bloques esquemático de una posible configuración adicional del mezclador de audio en la cual una interfaz de metadatos 480 se proporciona para aprovechar unos metadatos 12 que se
55 suministran junto con la señal de múltiples pistas. Los metadatos pueden comprender información acerca de la organización de pistas de audio o información de sección de tiempo tal como se ha explicado en el contexto de las figuras 12 y 13.

Los metadatos 12, si están presentes, guardan al mezclador de audio de tener que determinar información de pistas de audio, información de sección de tiempo, u otra información útil a partir de la señal de múltiples pistas. Una determinación de este tipo puede involucrar tareas de procesamiento de datos intensivas en cuanto al cómputo, las cuales pueden llevar un tiempo relativamente largo. Además, los resultados de la determinación realizada por el propio mezclador de audio pueden ser menos fiables que los metadatos proporcionados, que son producidos y proporcionados por un elemento de origen de la señal de audio de múltiples pistas.

La interfaz de metadatos 480 está configurada para extraer los metadatos 12 a partir de los datos 12 de la grabación de múltiples pistas. En un lado de salida, la interfaz de metadatos 480 está conectada con una entrada del intérprete de comandos semánticos 35. En la configuración que se muestra en la figura 15, el intérprete de comandos semánticos 35 está configurado para usar los metadatos 12 proporcionados por la interfaz de metadatos 480 en el proceso de obtener la pluralidad de parámetros de mezclado a partir del comando de mezclado semántico.

La figura 16 muestra un diagrama de bloques esquemático de otra posible configuración del mezclador de audio en la cual se proporcionan una interfaz de ejemplo 490 y un analizador de señal de mezcla a modo de ejemplo 492 para generar el comando de mezclado semántico sobre la base de una señal de mezcla a modo de ejemplo.

La interfaz de ejemplo 490 está configurada para recibir una señal de mezcla a modo de ejemplo. La señal de mezcla a modo de ejemplo se puede, por ejemplo, almacenar en una memoria o recuperar a través de una red. El usuario puede seleccionar la señal de mezcla a modo de ejemplo de entre una recopilación de señales de mezcla a modo de ejemplo de acuerdo con sus preferencias, por ejemplo, debido a que a este le gusta cómo se ha mezclado una señal de mezcla particular. En general, se puede usar cualquier señal de audio como la señal de mezcla a modo de ejemplo, pero por lo general son de esperar mejores resultados si la señal de mezcla a modo de ejemplo tiene una estructura y un estilo que sea similar a la grabación de múltiples pistas. Por ejemplo, puede ser útil si la instrumentación de la señal de mezcla a modo de ejemplo es sustancialmente la misma que la instrumentación de la señal de múltiples pistas que va a ser mezclada por el mezclador de audio.

La interfaz de ejemplo 490 reenvía la señal de mezcla a modo de ejemplo al analizador de señal de mezcla 492. El analizador de señal de mezcla 492 puede estar configurado para identificar partes de instrumento y vocales en la señal de mezcla a modo de ejemplo. Además, el analizador de señal de mezcla 492 puede determinar niveles de intensidad sonora y / o curvas de frecuencia de las partes instrumentales identificadas, las partes vocales identificadas y / o la señal de mezcla a modo de ejemplo como un todo. También puede ser posible determinar una cantidad de un efecto de audio, tal como reverberación. Basándose en los valores determinados, el analizador de señal de mezcla 492 puede establecer un perfil de la señal de mezcla a modo de ejemplo y / o un comando de mezclado semántico. Por ejemplo, el análisis realizado por el analizador de señal de audio 492 puede revelar que una pista de tambor y una pista de bajo de la señal de mezcla a modo de ejemplo son relativamente prominentes, mientras que otras pistas son más suaves. En consecuencia, el comando de mezclado semántico puede comprender una expresión que indique que la pista de tambor y la pista de bajo deben ser prominentes a lo largo de toda la señal de mezcla MS que va a ser producida por el mezclador de audio.

La interfaz de ejemplo 490 también se puede configurar para recibir pistas de audio a modo de ejemplo junto con la señal de mezcla a modo de ejemplo. Las pistas de audio a modo de ejemplo están representadas por un romboide con línea de trazos marcado "ST a modo de ejemplo" en la figura 16. Las pistas de audio a modo de ejemplo son proporcionadas al analizador de señal de mezcla 492 por la interfaz de ejemplo 490. Las pistas de audio a modo de ejemplo se corresponden con la señal de mezcla a modo de ejemplo ya que las pistas de audio a modo de ejemplo se usaron para generar la señal de mezcla a modo de ejemplo. Al estar disponibles las pistas de audio a modo de ejemplo, el analizador de señal de mezcla 492 puede comparar la señal de mezcla a modo de ejemplo con cada una de las pistas de audio a modo de ejemplo para descubrir cómo se ha modificado una cierta señal de mezcla a modo de ejemplo antes de que se mezcle para dar la señal de mezcla a modo de ejemplo. De esta manera se pueden determinar, de una forma semántica o semisemántica, parámetros de mezclado relacionados con la pista mediante el analizador de señal de mezcla 492.

La figura 17 muestra un diagrama de bloques esquemático de otra posible configuración del mezclador de audio en el que se usan un procesador perceptual 63 y un modelo perceptual 64 en el proceso de convertir el comando de mezclado semántico en parámetros de mezclado. El procesador perceptual 63 y el modelo perceptual 64 están representados como partes del intérprete de comandos semánticos 35 en la configuración de la figura 17. Tal como se ha indicado en lo que antecede, el procesador perceptual 63 traduce los valores perceptuales a los parámetros de mezclado físicos al tener en cuenta las características de señal y mecanismos de audición humana. Los parámetros que describen los mecanismos de audición humana son proporcionados por el modelo perceptual 64. El modelo perceptual 64 se puede organizar como una base de datos o una base de conocimiento. Las entradas de la base de datos pueden comprender una descripción semántica de un fenómeno relacionado con la audición y una implementación correspondiente en forma de parámetros para efectos de audio, intensidad sonora, intensidad sonora relativa, contenido de frecuencia, etc. El fenómeno relacionado con la audición se puede describir por ejemplo, mediante expresiones tales como "distante", "cercano", "plano", "pleno", "brillante", "desplazado hacia bajas frecuencias", "desplazado hacia altas frecuencias", etc.. La correspondiente implementación puede comprender valores numéricos que indican cómo se deben elegir los parámetros de mezclado para una o más de la pluralidad de pistas de audio ST para lograr el efecto deseado. Este establecimiento de correspondencia de una descripción semántica con unos valores correspondientes de los parámetros de mezclado está basado por lo general en un conocimiento experto y en psicoacústica. El conocimiento experto y la psicoacústica se pueden haber obtenido durante elaborados ensayos y estudios científicos.

Las configuraciones que se muestran en las figuras 8 y 11 a 16 se pueden combinar entre sí en cualquier combinación. Por ejemplo, mediante la combinación de las configuraciones que se muestran en las figuras 12 y 13, se puede proporcionar un mezclador de audio que comprende un identificador de pista de audio 430 y un
 5 identificador de sección de tiempo 460.

La figura 18 muestra un diagrama de flujo esquemático de un método para mezclar una pluralidad de señales de audio. Después del inicio del método en 102, se recibe un comando de mezclado semántico, tal como se ilustra en el bloque 104. El comando de mezclado semántico puede ser introducido por un usuario en forma de texto usando un
 10 teclado, de forma oral como un comando hablado, como una selección de entre una pluralidad de ajustes predeterminados, mediante el ajuste de uno o más parámetros, como una señal de mezcla a modo de ejemplo, o de otra manera.

En una acción representada por el bloque 106, se obtiene una pluralidad de parámetros de mezclado a partir del comando de mezclado semántico. Esta acción puede involucrar un conocimiento experto y psicoacústica de tal modo que los parámetros de mezcla conduzcan a un resultado deseado por el usuario.
 15

La pluralidad de pistas de audio se procesa de acuerdo con los parámetros de mezclado en el contexto de una acción representada por el bloque 108. El procesamiento de la pluralidad de pistas de audio puede comprender ajustar niveles de intensidad sonora, posiciones de panning, efectos de audio, filtrado de frecuencias (ecualización) y otras modificaciones.
 20

En una acción representada por el bloque 110, las pistas de audio que resultan del procesamiento se combinan para formar una señal de mezcla, antes de que el método termine en un bloque 112.
 25

De acuerdo con realizaciones adicionales, como una alternativa o además, el intérprete de comandos semánticos 30; 35 puede comprender una base de datos de vocabulario 31 para identificar expresiones semánticas dentro del comando de mezclado semántico.

De acuerdo con realizaciones adicionales con respecto al identificador de sección de tiempo, además o como alternativa, el identificador de sección de tiempo 40; 460 se puede configurar para estructurar la pluralidad de pistas de audio en una pluralidad de secciones de tiempo.
 30

De acuerdo con realizaciones adicionales, además o como alternativa, el mezclador de audio puede comprender además una interfaz de comando para recibir el comando de mezclado semántico en un formato lingüístico.
 35

De acuerdo con realizaciones adicionales, además o como alternativa, el procesador de lógica difusa se puede configurar para recibir por lo menos dos reglas difusas concurrentes preparadas por el intérprete de comandos semánticos, y el mezclador de audio puede comprender además un selector aleatorio para seleccionar una regla difusa concurrente de entre las por lo menos dos reglas difusas concurrentes.
 40

A pesar de que se han descrito algunos aspectos en el contexto de un aparato, es evidente que estos aspectos también representan una descripción del método correspondiente, en donde un bloque o dispositivo se corresponde con una etapa de método o con un rasgo distintivo de una etapa de método. De forma análoga, los aspectos que se describen en el contexto de una etapa de método también representan una descripción de un bloque o componente o rasgo distintivo correspondiente de un aparato correspondiente. Algunas o todas las etapas de método pueden ser ejecutadas por (o usando) un aparato de soporte físico, como por ejemplo, un microprocesador, un ordenador programable o un circuito electrónico. En algunas realizaciones, alguna o más de las etapas de método más importantes se pueden realizar mediante un aparato de este tipo.
 45
 50

Dependiendo de ciertos requisitos de implementación, las realizaciones de la invención se pueden implementar en soporte físico o en soporte lógico. La implementación se puede realizar usando un medio de almacenamiento digital, por ejemplo, un disco flexible, un DVD, un CD, una memoria de solo lectura, una PROM, una EEPROM o una memoria FLASH, que tenga señales de control electrónicamente legibles almacenadas en el mismo, las cuales cooperan (o son capaces de cooperar) con un sistema informático programable de tal modo que se realice el método respectivo. Por lo tanto, el medio de almacenamiento digital puede ser legible por ordenador.
 55

Algunas realizaciones de acuerdo con la invención comprenden un soporte de datos que tiene señales de control electrónicamente legibles, las cuales son capaces de cooperar con un sistema informático programable, de tal modo que se realice uno de los métodos que se describen en el presente documento.
 60

En general, se pueden implementar realizaciones de la presente invención como un programa informático con un código de programa, siendo el código de programa operativo para realizar uno de los métodos cuando el producto de programa informático se ejecuta en un ordenador. El código de programa se puede almacenar, por ejemplo, en

un soporte legible por máquina.

Otras realizaciones comprenden el programa informático para realizar uno de los métodos que se describen en el presente documento, almacenado en un soporte legible por máquina.

5 Dicho de otra forma, una realización del método de la invención es, por lo tanto, un programa informático que tiene un código de programa para realizar uno de los métodos que se describen en el presente documento, cuando el programa informático se ejecuta en un ordenador.

10 Una realización adicional de los métodos de la invención es, por lo tanto, un soporte de datos (o un medio de almacenamiento digital, o un medio legible por ordenador) que comprende, grabado en el mismo, el programa informático para realizar uno de los métodos que se describen en el presente documento. El soporte de datos, el medio de almacenamiento digital o el medio de registro por lo general son tangibles y / o no transitorios.

15 Una realización adicional del método de la invención es, por lo tanto, un tren de datos o una secuencia de señales que representan el programa informático para realizar uno de los métodos que se describen en el presente documento. El tren de datos o la secuencia de señales se pueden, por ejemplo, configurar para que se transfieran por medio de una conexión de comunicación de datos, por ejemplo, por medio de Internet.

20 Una realización adicional comprende un medio de procesamiento, por ejemplo, un ordenador, o un dispositivo lógico programable, configurado para o adaptado para realizar uno de los métodos que se describen en el presente documento.

25 Una realización adicional comprende un ordenador que tiene instalado en el mismo el programa informático para realizar uno de los métodos que se describen en el presente documento.

30 Una realización adicional de acuerdo con la invención comprende un aparato o un sistema configurado para transferir (por ejemplo, de forma electrónica u óptica) un programa informático para realizar uno de los métodos que se describen en el presente documento, a un receptor. El receptor puede, por ejemplo, ser un ordenador, un dispositivo móvil, un dispositivo de memoria o similar. El aparato o sistema puede, por ejemplo, comprender un servidor de archivos para transferir el programa informático al receptor.

35 En algunas realizaciones se puede usar un dispositivo de lógica programable (por ejemplo una disposición de compuertas programable de campo) para realizar algunas o todas las funcionalidades de los métodos que se describen en el presente documento. En algunas realizaciones, la disposición de compuertas programable de campo puede cooperar con un microprocesador con el fin de realizar uno de los métodos que se describen en el presente documento. En general, los métodos se realizan preferiblemente mediante cualquier aparato de soporte físico.

40 Las realizaciones que se han descrito en lo que antecede son meramente ilustrativas para los principios de la presente invención. Se entiende que las modificaciones y variaciones de las disposiciones y de los detalles que se describen en el presente documento serán evidentes para otros expertos en la materia. Por lo tanto, se pretende estar limitado solo por el alcance de las reivindicaciones de patente inmediatamente siguientes y no por los detalles específicos que se presentan a modo de descripción y de explicación de las realizaciones en el presente documento.

REIVINDICACIONES

1. Mezclador de audio para mezclar una pluralidad de pistas de audio para dar una señal de mezcla (MS), comprendiendo el mezclador de audio:
- 5 un análisis de audio semántico (40) configurado para obtener información de pistas mediante el análisis de la pluralidad de pistas de audio;
- 10 un intérprete de comandos semánticos (30; 35) para recibir un comando de mezclado semántico y para obtener una pluralidad de parámetros de mezclado para la pluralidad de pistas de audio a partir del comando de mezclado semántico, en el que la información de pista se proporciona a un módulo de semántico a nítido (65), en el que el módulo de semántico a nítido (65) recibe información que se obtiene a partir del comando de mezclado semántico, en el que el módulo de semántico a nítido (65) crea la pluralidad de parámetros de mezclado basándose en la información de pista y la información que se obtiene a partir del comando de mezclado semántico;
- 15 un procesador de pistas de audio (70; 75) para procesar la pluralidad de pistas de audio de acuerdo con la pluralidad de parámetros de mezclado; y
- un combinador de pistas de audio (76) para combinar la pluralidad de pistas de audio procesadas por el procesador de pistas de audio para dar la señal de mezcla (MS).
- 20 2. Mezclador de audio de acuerdo con la reivindicación 1, que comprende además un identificador de pista de audio (40; 430) para identificar una pista de audio objetivo de entre la pluralidad de pistas de audio, estando indicada la pista de audio objetivo dentro del comando de mezclado semántico por una expresión de identificación de pista de audio.
- 25 3. Mezclador de audio de acuerdo con la reivindicación 2, en el que el identificador de pista de audio (40; 430) está configurado
- para recuperar un registro de datos que se corresponde con la expresión de identificación de pista de audio a partir de una base de datos de plantillas de pistas de audio (432), comprendiendo el registro de datos una información acerca de un instrumento musical correspondiente en forma de por lo menos uno de un valor de medición y una
- 30 muestra de sonido,
- para realizar un análisis de por lo menos uno de un timbre, una estructura rítmica, un rango de frecuencias, una muestra de sonido y una densidad de armónicos de por lo menos una pista de audio de entre la pluralidad de pistas de audio,
- para comparar un resultado del análisis con el registro de datos dando como resultado por lo menos una puntuación de coincidencia, y
- 35 para determinar la pista de audio objetivo sobre la base de la por lo menos una puntuación de coincidencia entre la por lo menos una pista de audio y el registro de datos.
- 40 4. Mezclador de audio de acuerdo con una cualquiera de las reivindicaciones precedentes, que además comprende un identificador de sección de tiempo (40; 460) para identificar una sección de tiempo objetivo dentro de la pluralidad de pistas de audio, estando indicada la sección de tiempo objetivo dentro del comando de mezclado semántico por una expresión de identificación de sección de tiempo.
- 45 5. Mezclador de audio de acuerdo con la reivindicación 4, en el que el identificador de sección de tiempo (40; 460) está configurado para realizar un análisis de la pluralidad de pistas de audio para determinar por lo menos un instante de tiempo en el que tiene lugar un cambio de una propiedad característica de una señal de audio que está representada por la pluralidad de pistas de audio, y para usar el por lo menos un instante de tiempo determinado como por lo menos un límite entre dos secciones de tiempo adyacentes.
- 50 6. Mezclador de audio de acuerdo con una cualquiera de las reivindicaciones precedentes, que además comprende una interfaz de metadatos(42; 480) para recibir metadatos (12) en relación con la pluralidad de pistas de audio, siendo los metadatos (12) indicativos de por lo menos uno de un nombre de pista, un identificador de pista, una información de estructura de tiempo, una información de intensidad, atributos espaciales de una pista de audio o una parte de la misma, características de timbre y características rítmicas.
- 55 7. Mezclador de audio de acuerdo con una cualquiera de las reivindicaciones precedentes, que además comprende una interfaz de ejemplo (23; 490) para recibir otra señal de mezcla como una señal de mezcla a modo de ejemplo de acuerdo con unas preferencias de usuario en relación con cómo se ha mezclado la señal de mezcla a modo de ejemplo, y
- 60 un analizador de señal de mezcla (492) para analizar la señal de mezcla a modo de ejemplo y para generar el comando de mezclado semántico basándose en el análisis de la señal de mezcla a modo de ejemplo.
8. Mezclador de audio de acuerdo con una cualquiera de las reivindicaciones precedentes, en el que el intérprete de comandos semánticos (30; 35) comprende un procesador perceptual (63) para transformar el comando de mezclado

semántico en la pluralidad de parámetros de mezclado de acuerdo con un modelo perceptual (64) de propiedades relacionadas con la audición de la señal de mezcla.

5 9. Mezclador de audio de acuerdo con una cualquiera de las reivindicaciones precedentes, en el que el intérprete de comandos semánticos (30; 35) comprende un procesador de lógica difusa para recibir por lo menos una regla difusa que se obtiene a partir del comando de mezclado semántico mediante el intérprete de comandos semánticos, y para generar la pluralidad de parámetros de mezclado sobre la base de la por lo menos una regla difusa.

10 10. Mezclador de audio de acuerdo con una cualquiera de las reivindicaciones precedentes, en el que el intérprete de comandos semánticos (30; 35) comprende:

15 una unidad de asignación de descriptor objetivo (50) para seleccionar por lo menos una parte de una señal de múltiples pistas que comprende la pluralidad de pistas de audio y asignar por lo menos un descriptor objetivo perceptual (PTD) apropiado a la por lo menos parte de la señal de múltiples pistas sobre la base del comando de mezclado semántico;

un procesador perceptual (60) para traducir valores perceptuales definidos en el por lo menos un descriptor objetivo perceptual (PTD) a los parámetros de mezclado al tener en cuenta características de señal y mecanismos de audición humana.

20 11. Método para mezclar una pluralidad de pistas de audio para dar una señal de mezcla, comprendiendo el método:

recibir un comando de mezclado semántico;

obtener información de pistas mediante el análisis de la pluralidad de pistas de audio;

25 obtener una pluralidad de parámetros de mezclado para la pluralidad de pistas de audio a partir del comando de mezclado semántico, en el que la información de pista se proporciona a un módulo de semántico a nítido (65), en el que el módulo de semántico a nítido (65) recibe información que se obtiene a partir del comando de mezclado semántico, en el que el módulo de semántico a nítido (65) crea la pluralidad de parámetros de mezclado basándose en la información de pista y la información que se obtiene a partir del comando de mezclado semántico;

30 procesar la pluralidad de pistas de audio de acuerdo con la pluralidad de parámetros de mezclado; y combinar la pluralidad de pistas de audio que resultan del procesamiento de la pluralidad de pistas de audio para formar la señal de mezcla.

35 12. Método de acuerdo con la reivindicación 12, que comprende además:

seleccionar por lo menos una parte de la señal de múltiples pistas que comprende la pluralidad de pistas de audio y asignar por lo menos un descriptor objetivo perceptual (PTD) apropiado a la por lo menos una parte sobre la base del comando de mezclado semántico;

40 traducir valores perceptuales definidos en el por lo menos un descriptor objetivo perceptual (PTD) a los parámetros de mezclado al tener en cuenta características de señal y mecanismos de audición humana.

13. Mezclador de audio para mezclar una pluralidad de pistas de audio para dar una señal de mezcla (MS), comprendiendo el mezclador de audio:

45 un intérprete de comandos semánticos (30; 35) para recibir un comando de mezclado semántico y para obtener una pluralidad de parámetros de mezclado para la pluralidad de pistas de audio a partir del comando de mezclado semántico;

un procesador de pistas de audio (70; 75) para procesar la pluralidad de pistas de audio de acuerdo con la pluralidad de parámetros de mezclado;

50 un combinador de pistas de audio (76) para combinar la pluralidad de pistas de audio procesadas por el procesador de pistas de audio para dar la señal de mezcla (MS); y

un identificador de pista de audio (40, 430) para identificar una pista de audio objetivo de entre la pluralidad de pistas de audio, estando indicada la pista de audio objetivo dentro del comando de mezclado semántico por una expresión de identificación de pista de audio, estando configurado el identificador de pista de audio para

55 recuperar un registro de datos que se corresponde con la expresión de identificación de pista de audio a partir de una base de datos de plantillas de pistas de audio (432), comprendiendo el registro de datos una información acerca de un instrumento musical correspondiente en forma de por lo menos uno de un valor de medición y una muestra de sonido, para analizar las pistas de audio y para comparar señales de audio de las pistas de audio con el registro de datos, con el fin de determinar una pista de audio o varias pistas de datos que parecen coincidir

60 con la pista de audio objetivo.

14. Método para mezclar una pluralidad de pistas de audio para dar una señal de mezcla, comprendiendo el método:

recibir un comando de mezclado semántico;

- obtener una pluralidad de parámetros de mezclado para la pluralidad de pistas de audio a partir del comando de mezclado semántico, comprendiendo la pluralidad de parámetros de mezclado un parámetro de mezclado para una pista de audio objetivo;
- 5 identificar una pista de audio objetivo que se está indicando dentro del comando de mezclado semántico por una expresión de identificación de pista de audio;
- recuperar, a partir de una base de datos de plantillas de pistas de audio, un registro de datos que se corresponde con la expresión de identificación de pista de audio, comprendiendo el registro de datos una información acerca de un instrumento musical correspondiente en forma de por lo menos uno de un valor de medición y una muestra de sonido,
- 10 identificar la pista de audio objetivo de entre la pluralidad de pistas de audio mediante el análisis de señales de audio de las pistas de audio y comparar estas con el registro de datos, para determinar una pista de audio o varias pistas de audio que parecen coincidir con la pista de audio objetivo;
- procesar la pluralidad de pistas de audio de acuerdo con la pluralidad de parámetros de mezclado; y
- 15 combinar la pluralidad de pistas de audio que resultan del procesamiento de la pluralidad de pistas de audio para formar la señal de mezcla.

15. Programa informático para dar instrucciones a un ordenador para que realice el método de la reivindicación 11, 12 o 14.

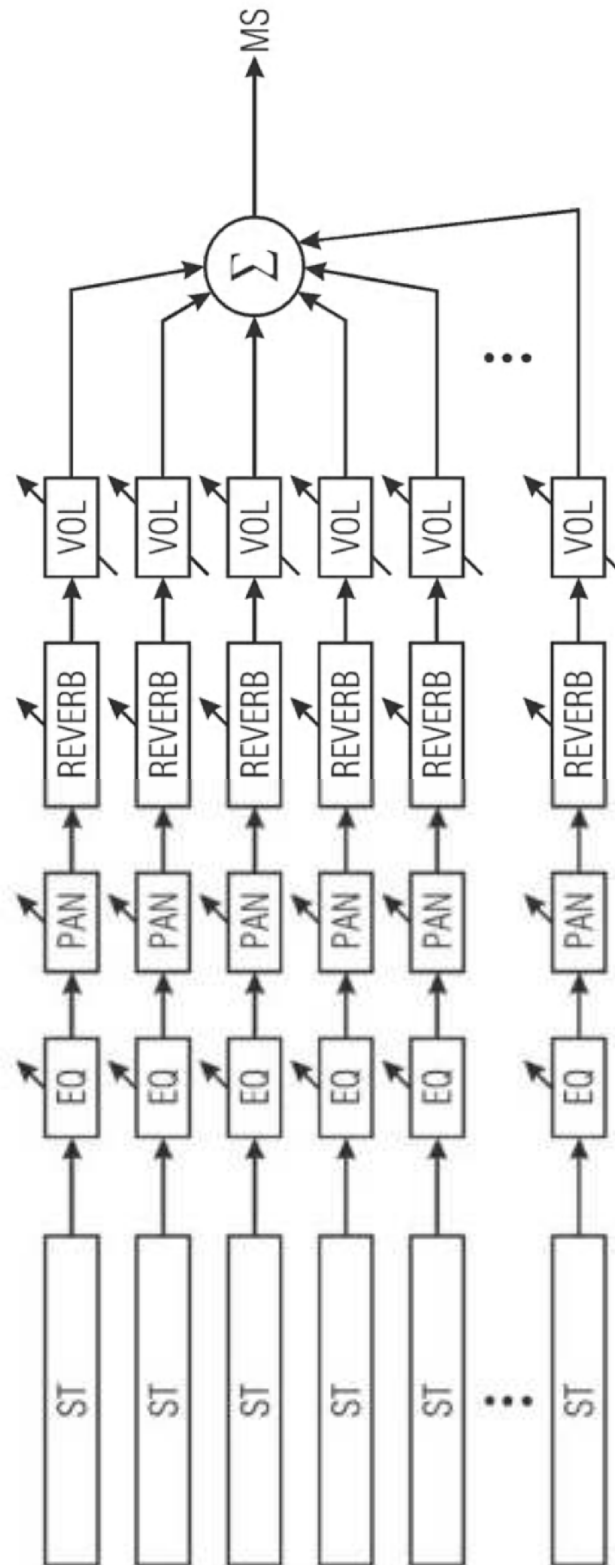


FIGURA 1

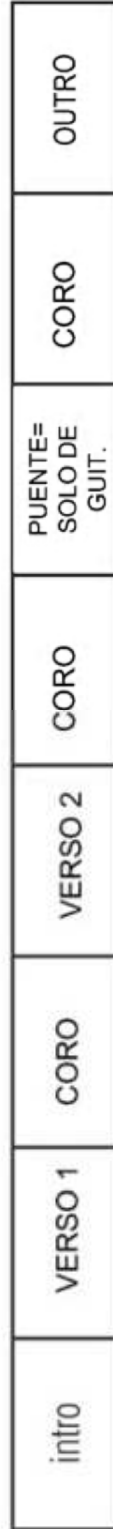


FIGURA 2



FIGURA 3

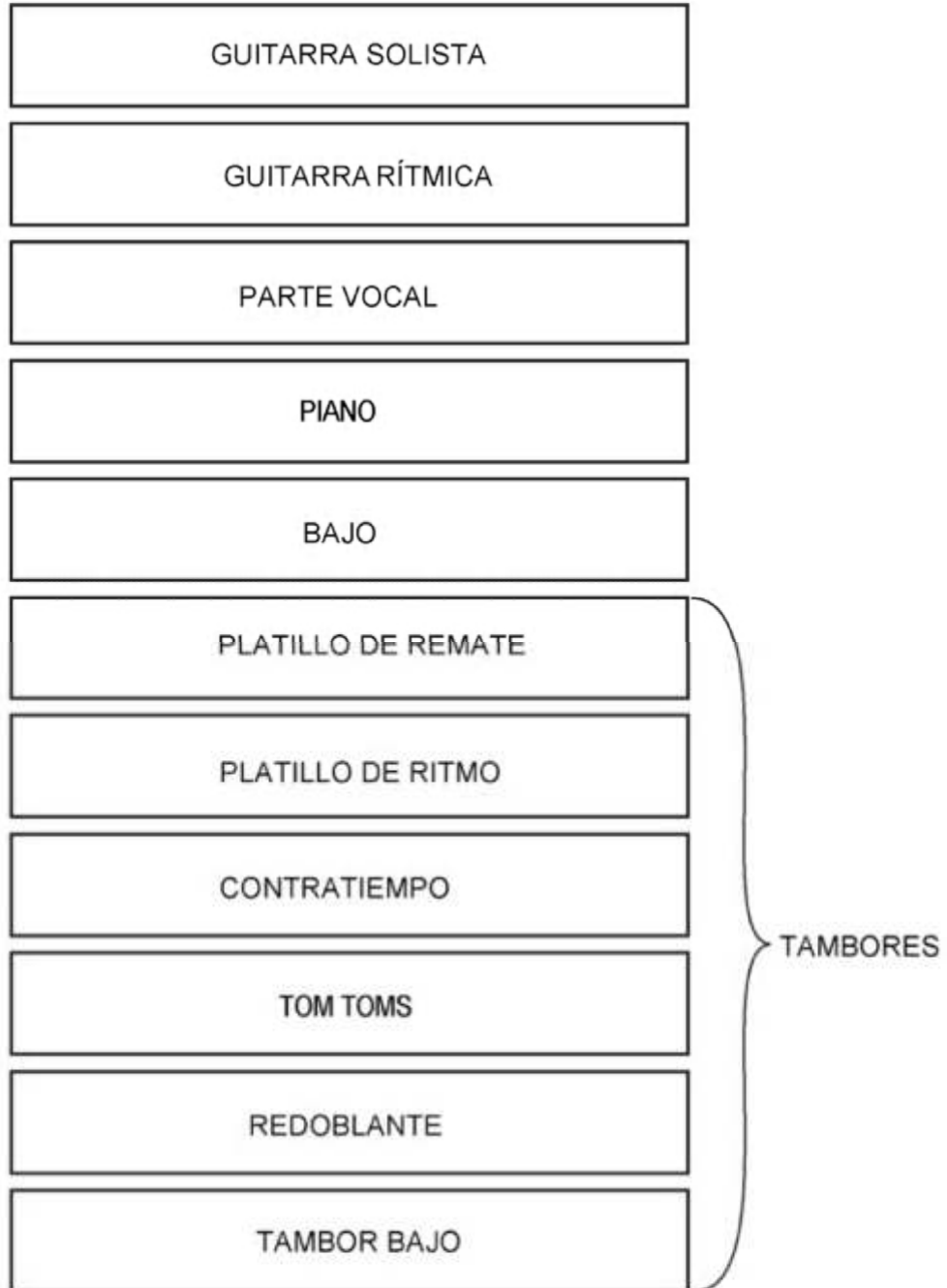


FIGURA 4

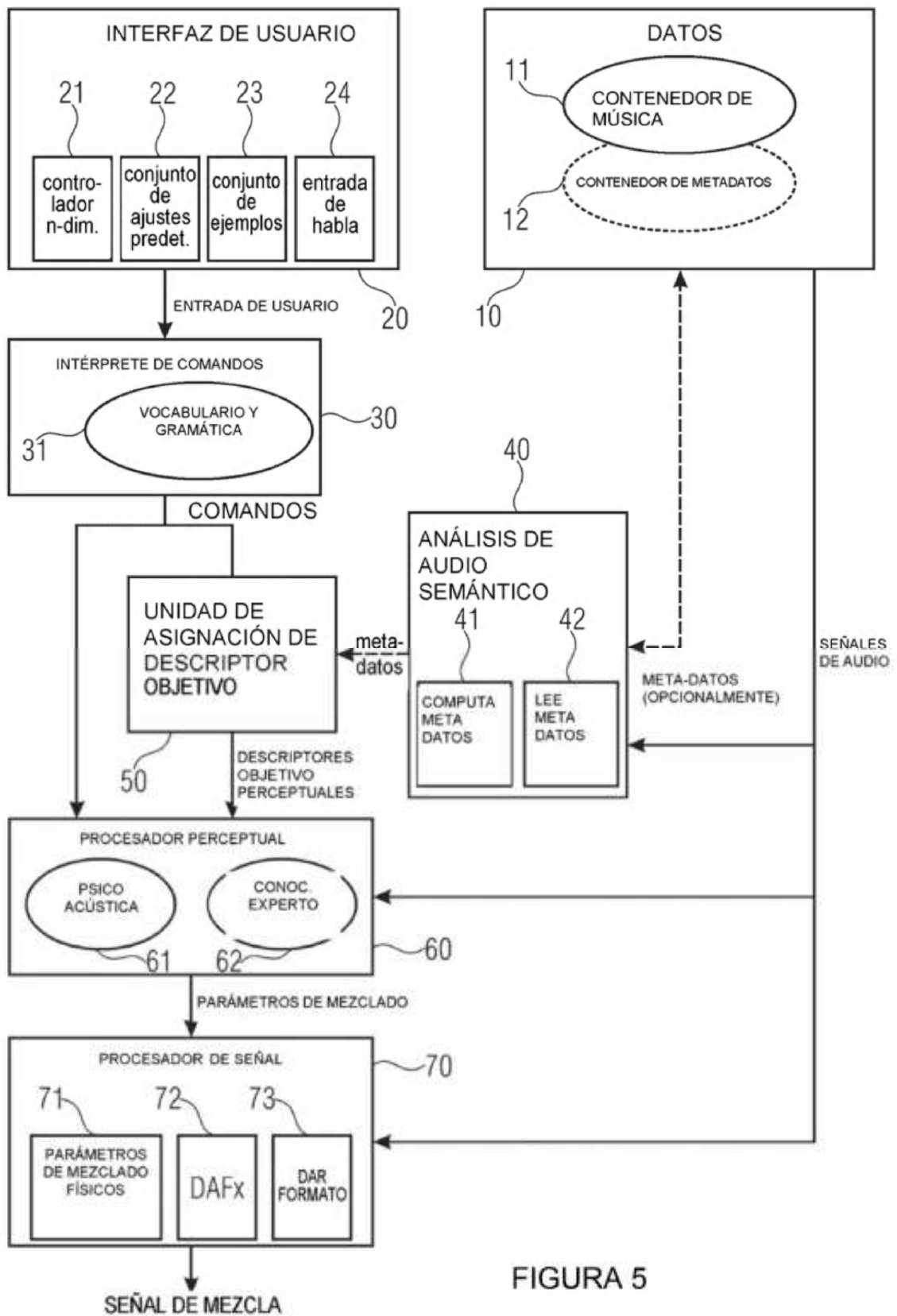


FIGURA 5

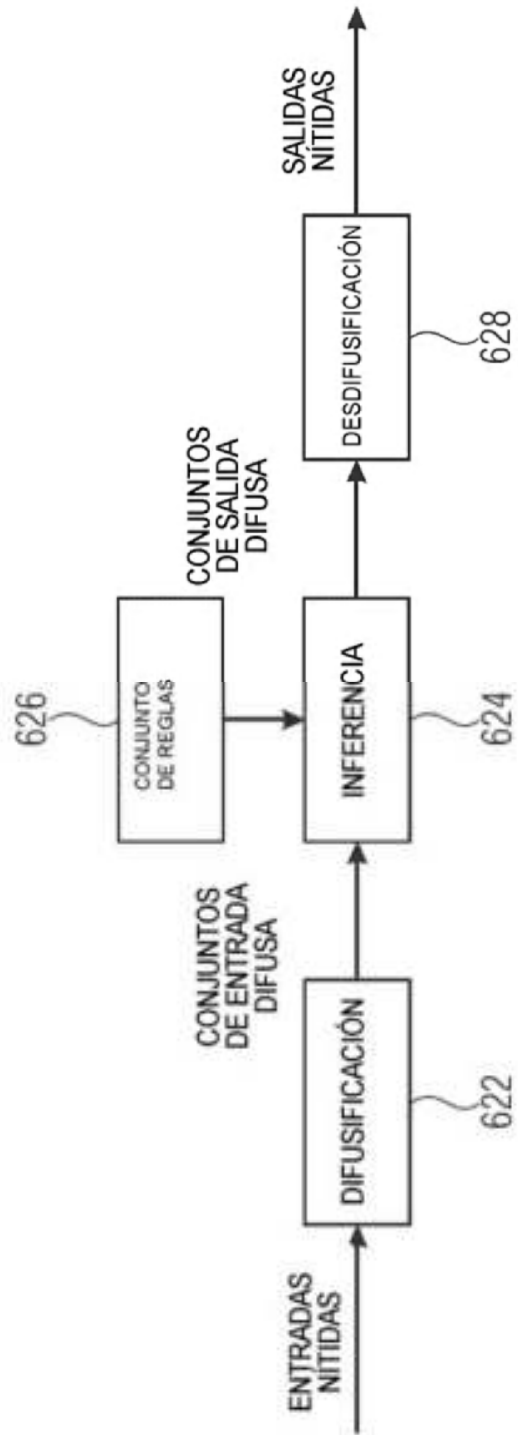


FIGURA 6

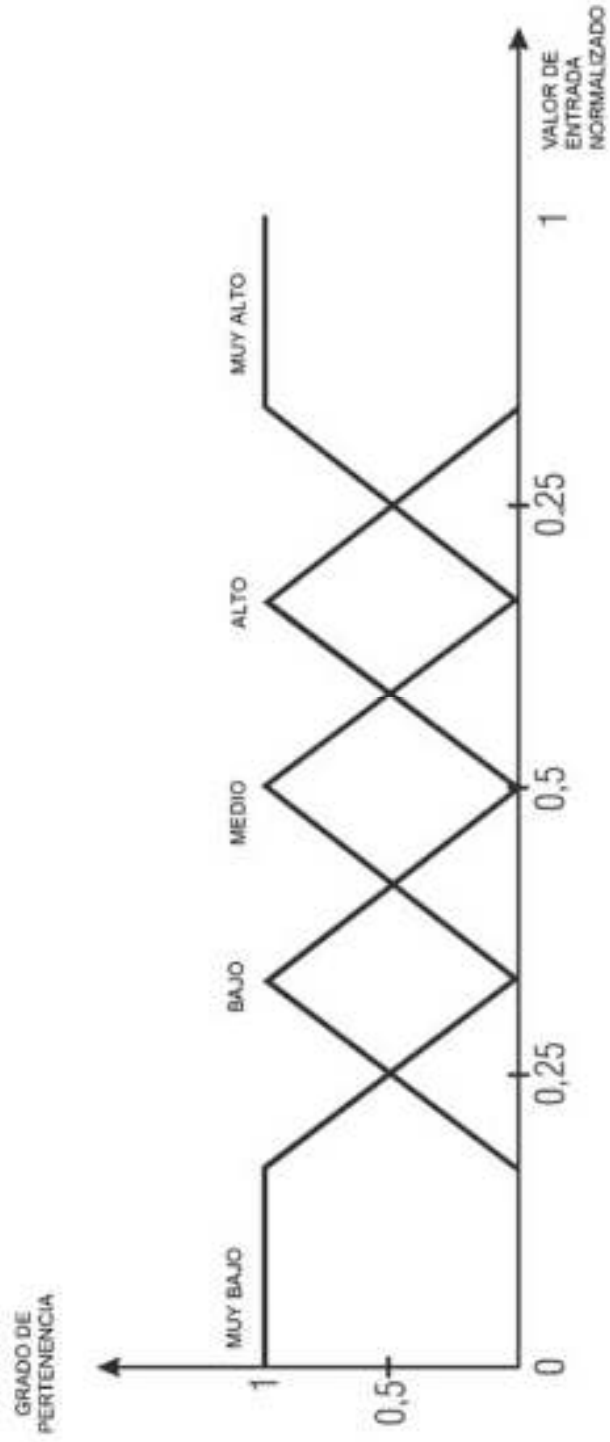


FIGURA 7

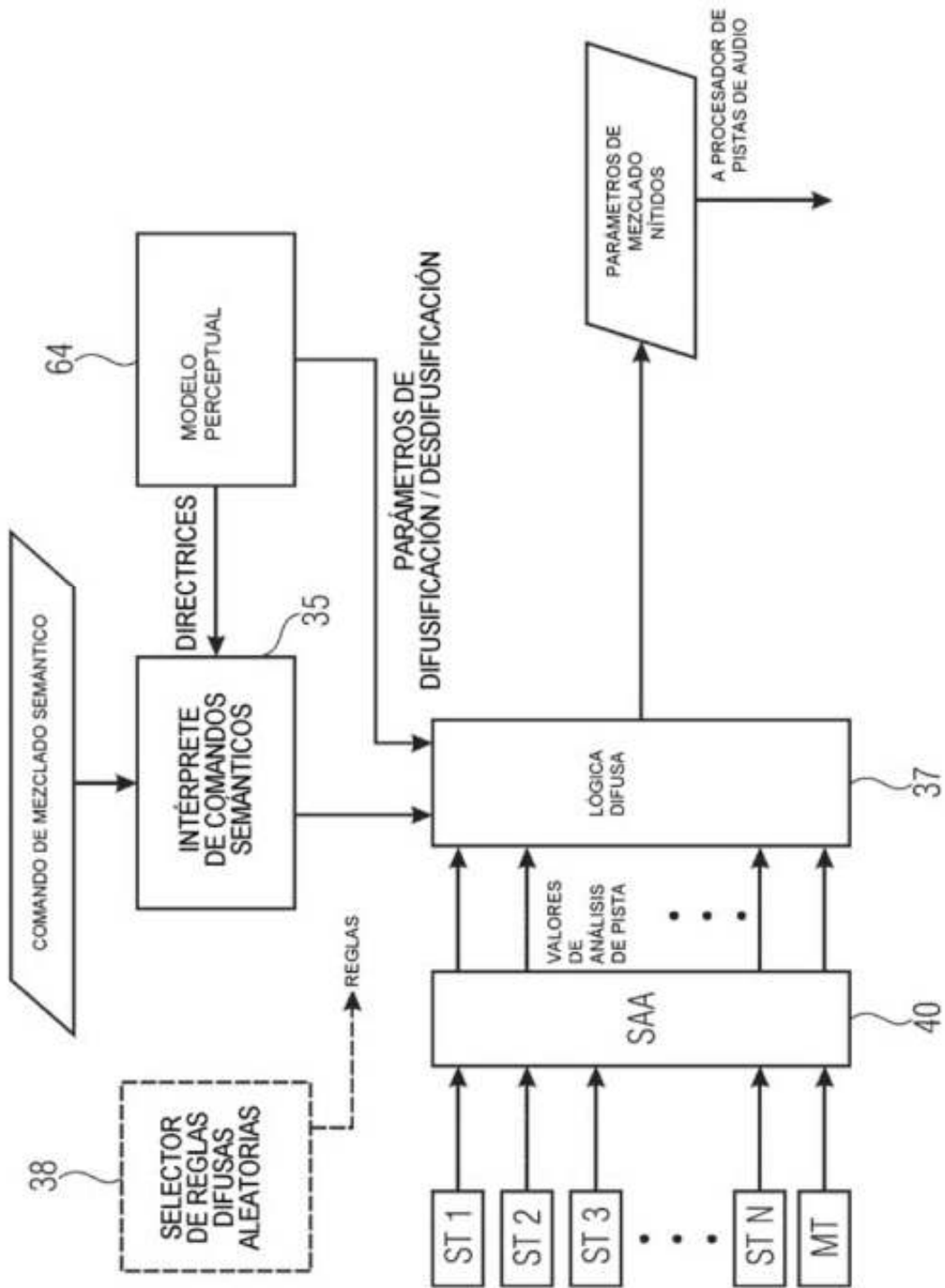


FIGURA 8

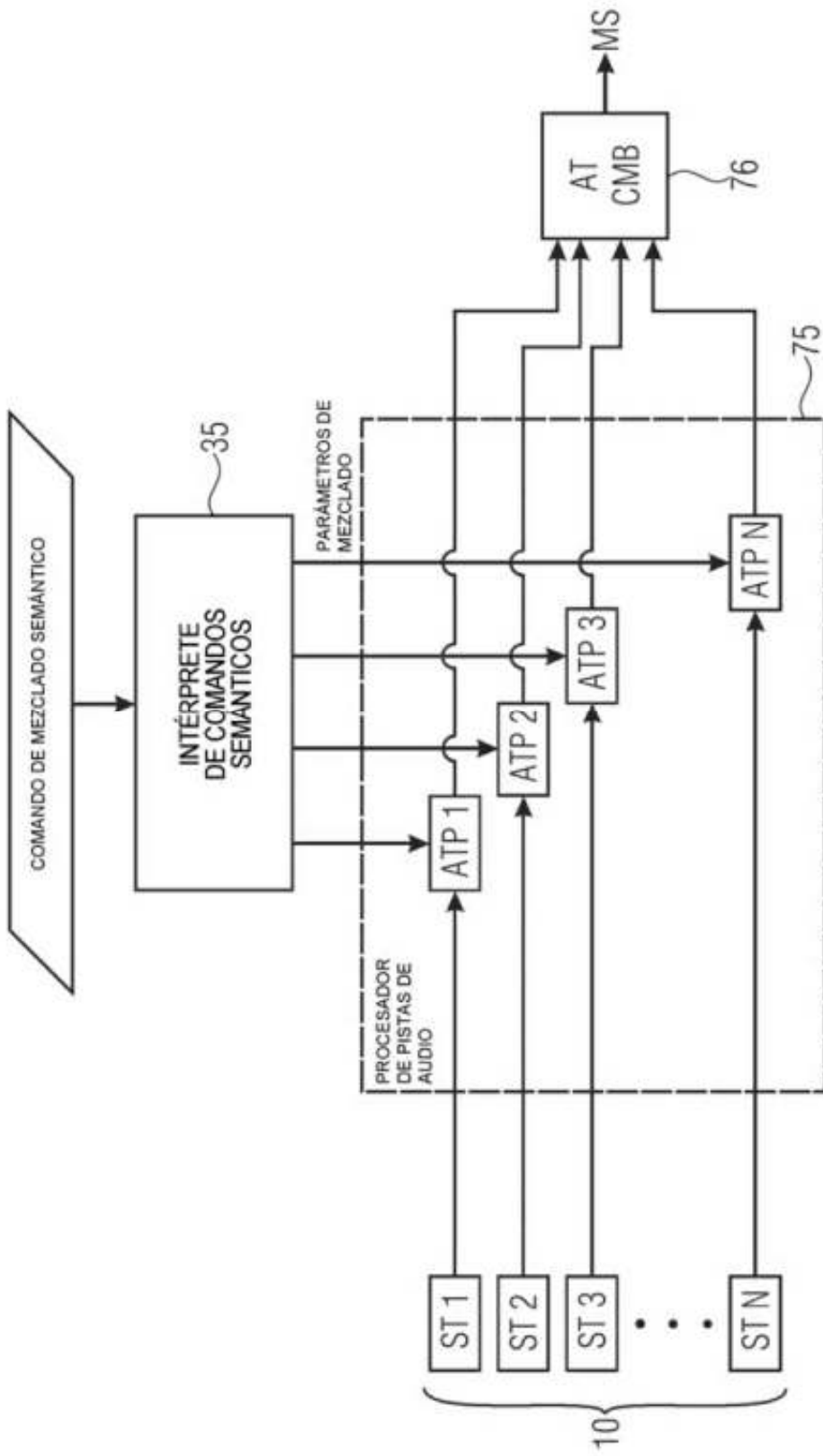


FIGURA 9

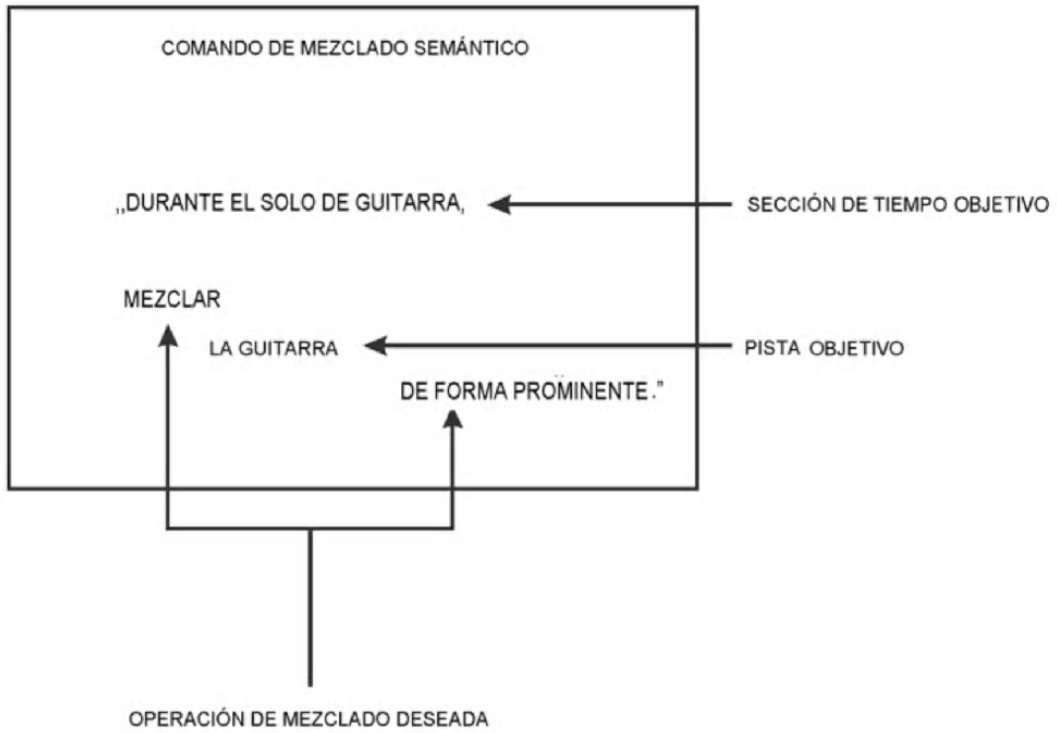


FIGURA 10



FIGURA 11

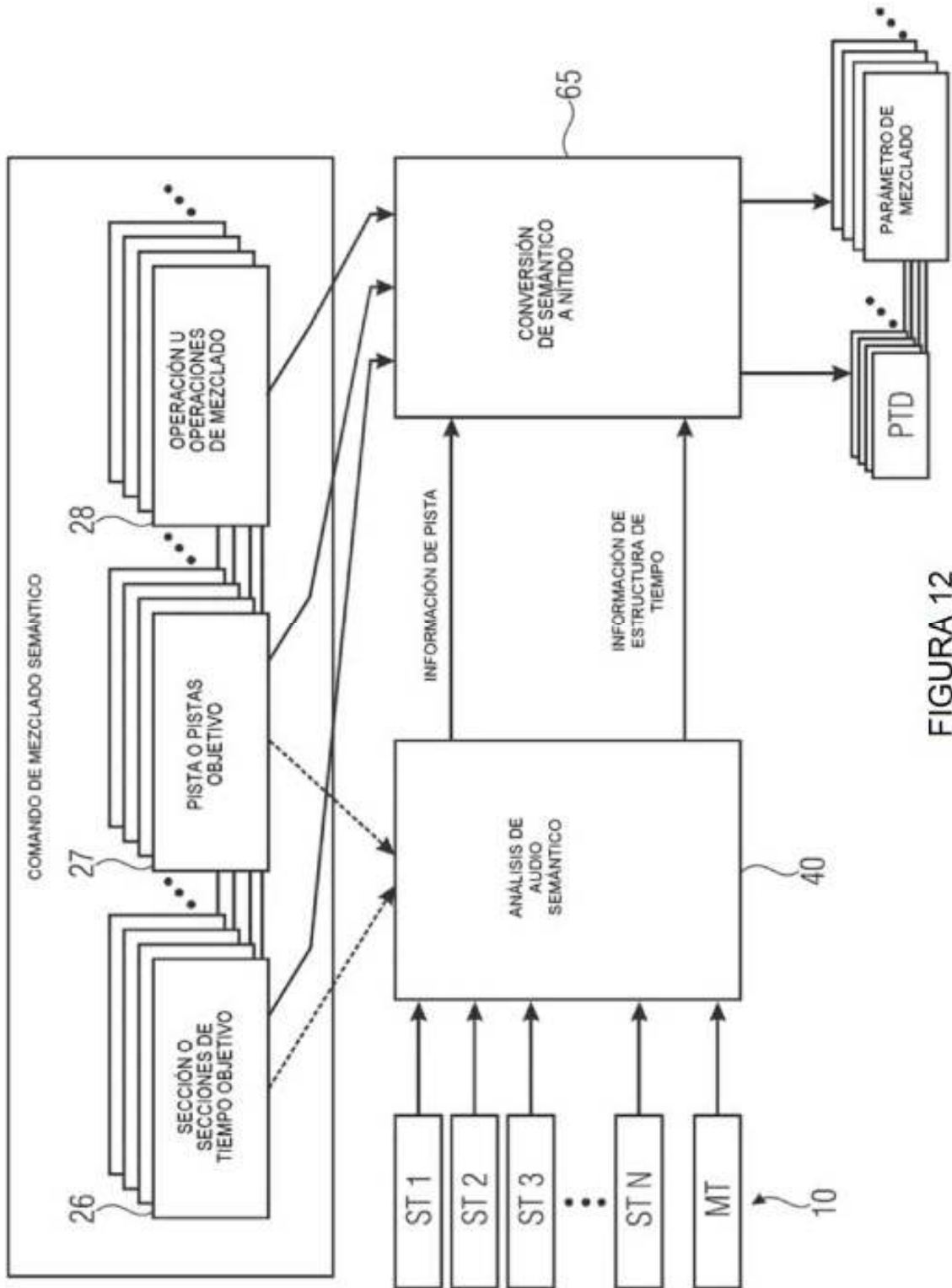


FIGURA 12

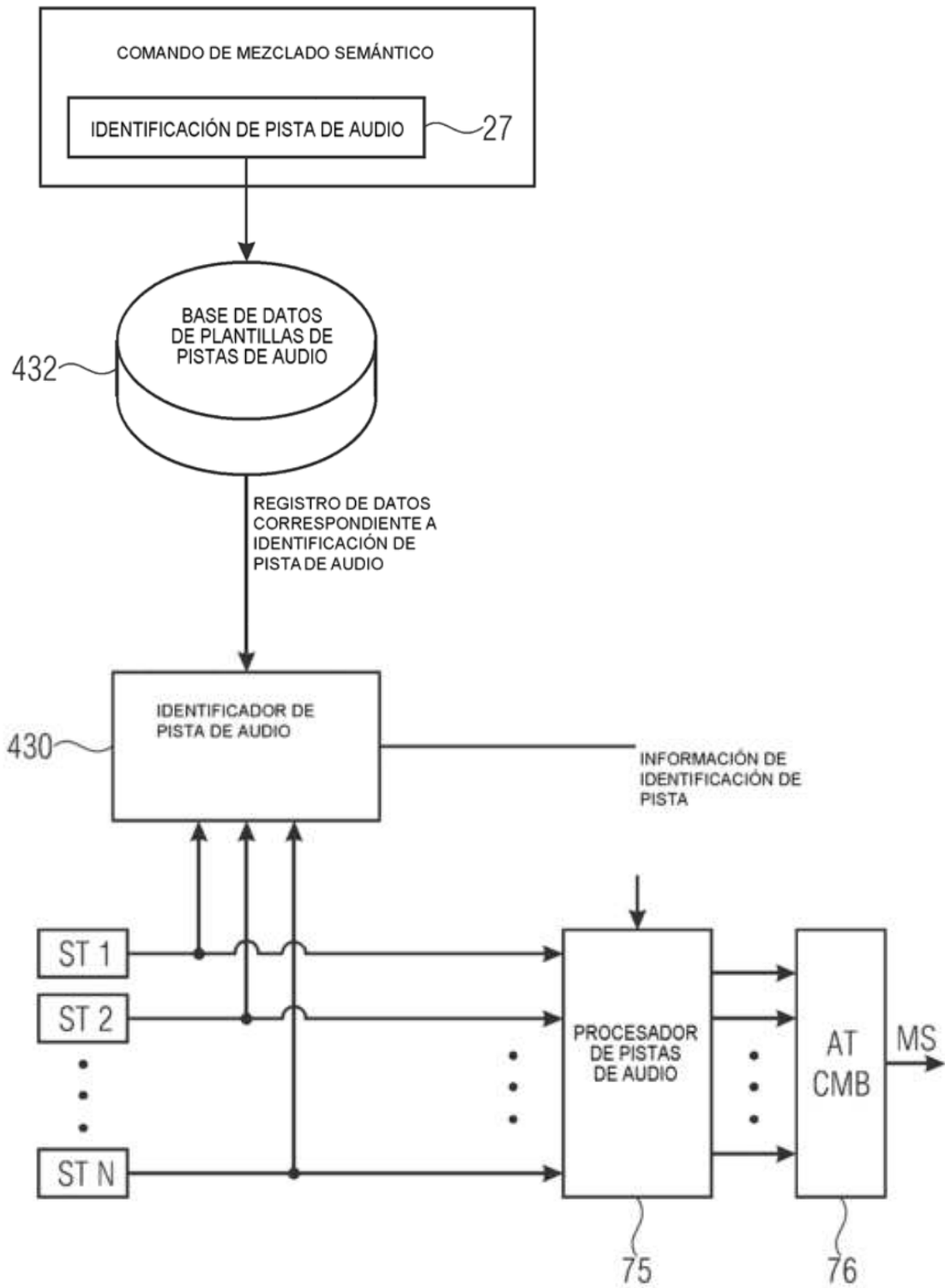


FIGURA 13

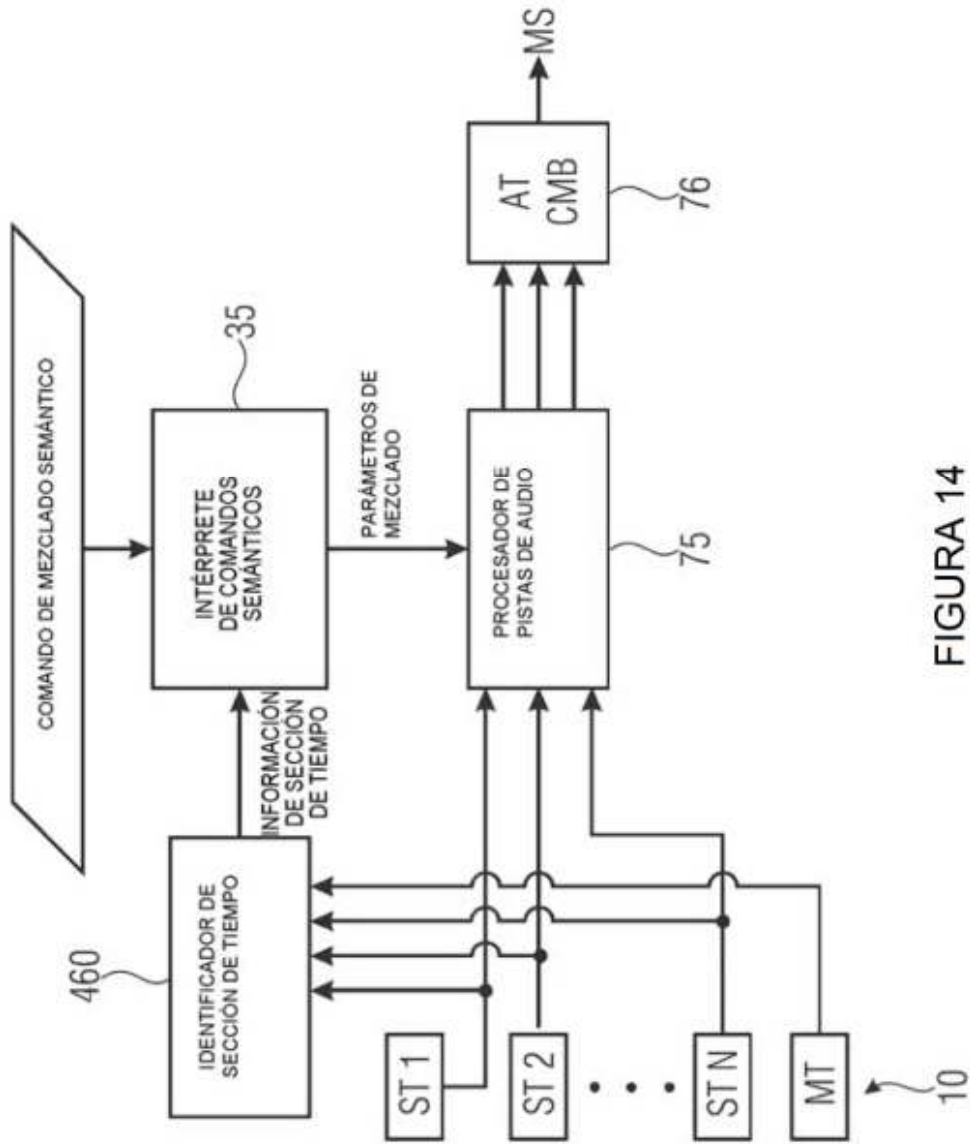


FIGURA 14

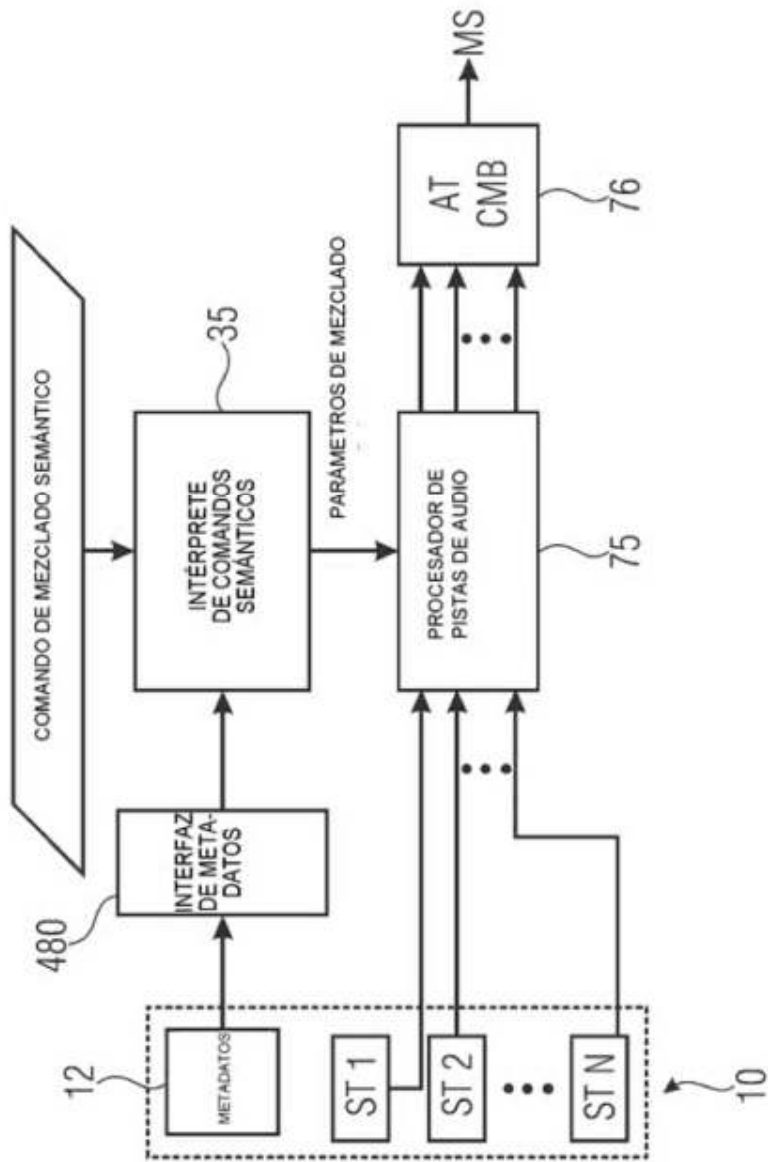


FIGURA 15

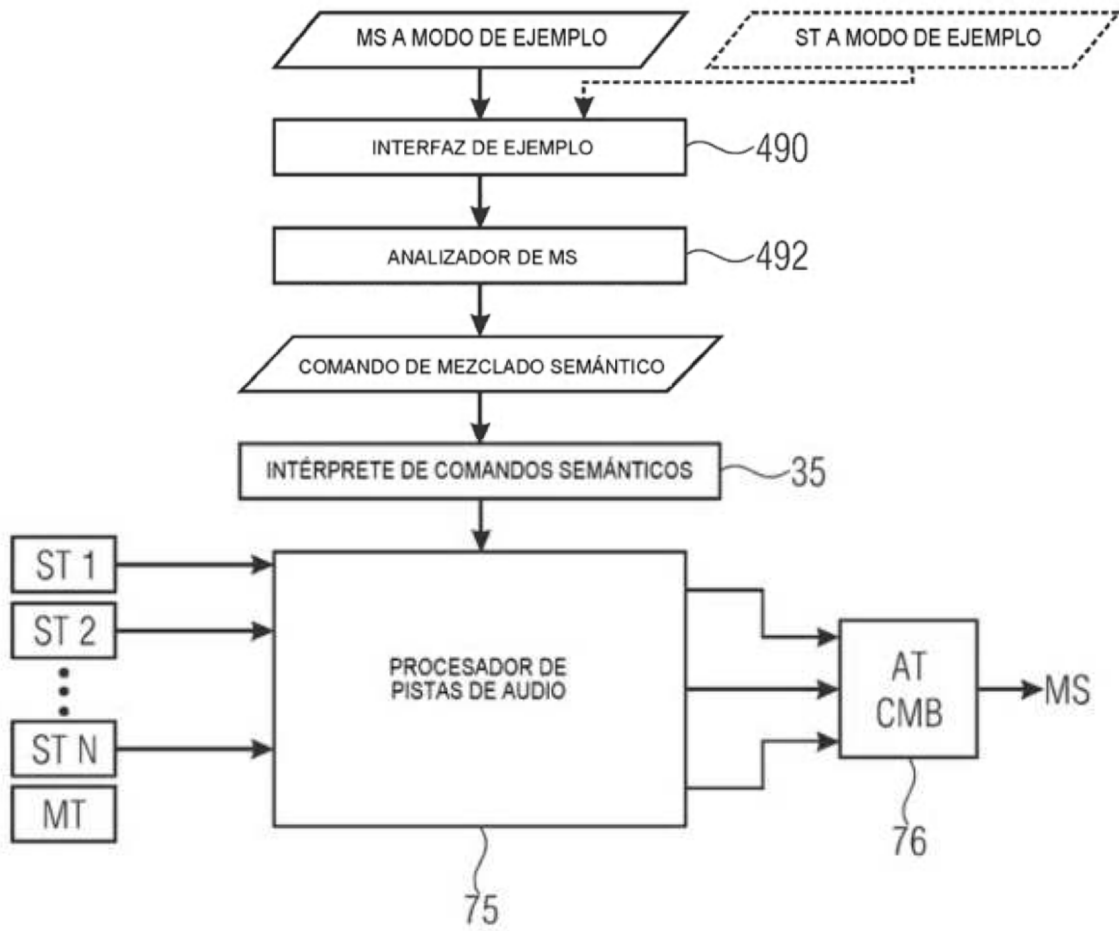


FIGURA 16

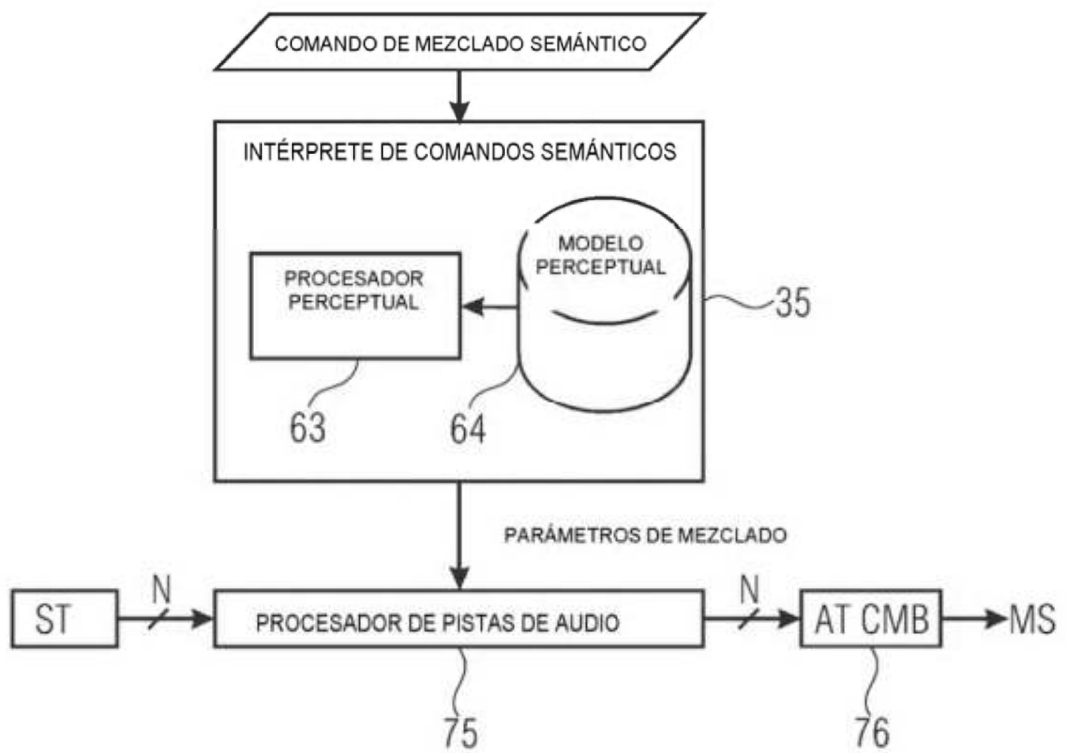


FIGURA 17

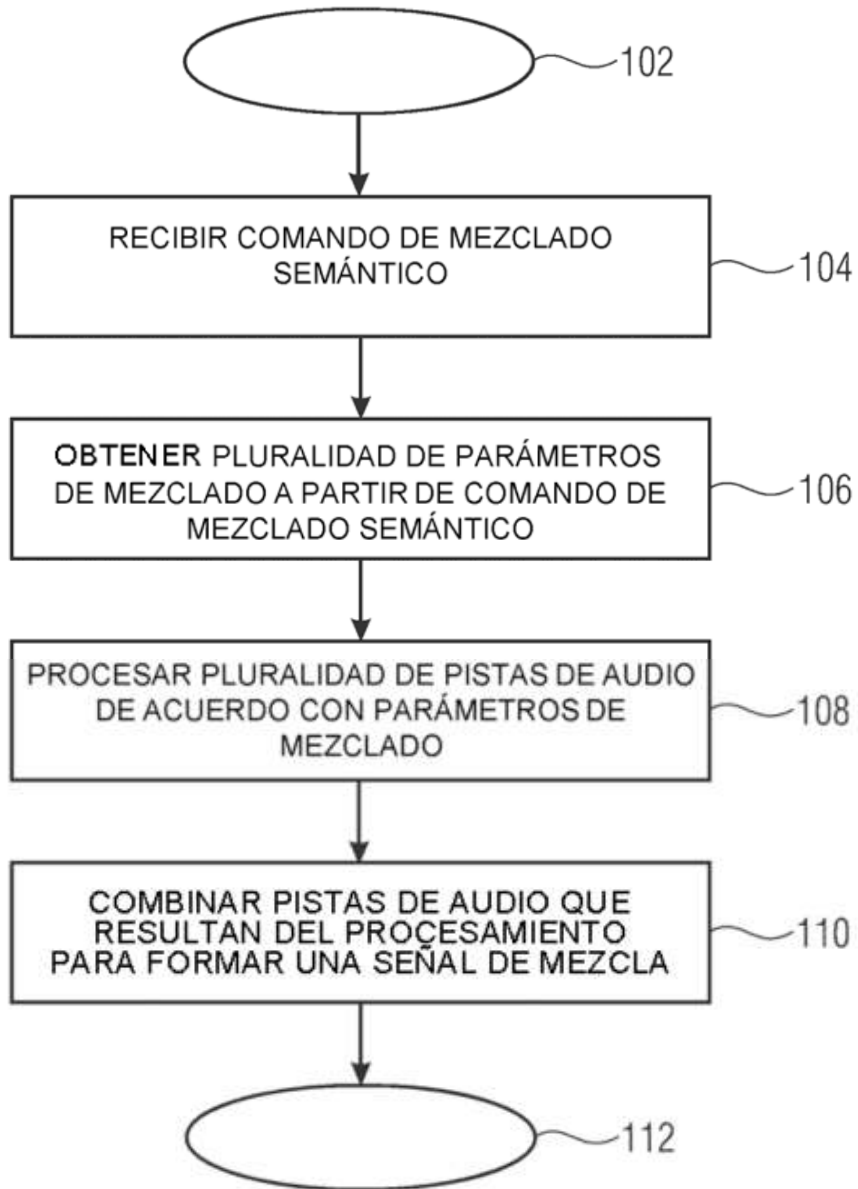


FIGURA 18