(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 453 074**

(51) Int. CI.:

H04S 3/00 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Fecha de presentación y número de la solicitud europea: **06.07.2009 E 09776987 (1)**

(97) Fecha y número de publicación de la concesión europea: **12.03.2014 EP 2297978**

(54) Título: **Aparato y procedimiento para generar señales de salida de audio mediante el uso de metadatos basados en objetos**

(30) Prioridad:

17.07.2008 EP 08012939
09.10.2008 EP 08017734

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
03.04.2014

(73) Titular/es:

FRAUNHOFER-GESELLSCHAFT ZUR FÖRDERUNG DER ANGEWANDTEN FORSCHUNG E.V. (100.0%)
Hansastrasse 27c
80686 München, DE

(72) Inventor/es:

SCHREINER, STEPHAN;
FIESEL, WOLFGANG;
NEUSINGER, MATTHIAS;
HELLMUTH, OLIVER y
SPERSCHNEIDER, RALPH

(74) Agente/Representante:

PONTI SALES, Adelaida

DESCRIPCIÓN

Aparato y procedimiento para generar señales de salida de audio mediante el uso de metadatos basados en objetos

Campo de la invención

[0001] La presente invención se refiere al procesamiento de audio y, en particular, al procesamiento de audio en el contexto de la codificación de objetos de audio tal como la codificación espacial de objetos de audio.

Antecedentes de la invención y técnica relacionada

[0002] En los modernos sistemas de difusión tales como la televisión, en algunos casos es deseable no reproducir las pistas de audio tal como los diseñó el técnico de sonido, sino más bien llevar a cabo ajustes especiales para ocuparse de las restricciones impuestas en el tiempo de renderización (En este contexto es el proceso de edición para obtener un efecto deseado). Una tecnología, bien conocida, para controlar tales ajustes post–producción, consiste en proveer metadatos adecuados juntos con dichas pistas de audio.

[0003] Los sistemas tradicionales para la reproducción del sonido, por ejemplo los sistemas antiguos de la televisión doméstica, consisten en un altavoz o en un par estéreo de altavoces. Los sistemas de reproducción multicanal más sofisticados utilizan cinco altavoces, o una cantidad mayor aún.

[0004] Si se consideran los sistemas de reproducción multicanal, los técnicos de sonido pueden ser mucho más flexibles en la colocación de fuentes individuales en un plano bidimensional y por ello también pueden utilizar un margen dinámico más elevado para sus pistas de audio globales, ya que la inteligibilidad de la voz es mucho más fácil debido al bien conocido efecto de “fiesta de cócteles (cocktail party effect” – se refiere a la capacidad del ser humano de focalizar su atención auditiva a una fuente puntual en un entorno altamente ruidoso).

[0005] Sin embargo, estos sonidos realistas, muy dinámicos, pueden ser causa de problemas en los sistemas de reproducción tradicionales. Puede haber escenarios en los que un consumidor no desea esta señal muy dinámica, sea porque ella o él está escuchando el contenido en un entorno ruidoso (por ejemplo, en un automóvil en movimiento o con un sistema de entretenimiento en vuelo o móvil), ella o él está utilizando audífonos, o ella o él no desea molestar a sus vecinos (en un vuelo nocturno, por ejemplo).

[0006] Además, las empresas de radiodifusión se enfrentan al problema de que diferentes ítems en un programa (por ejemplo, los avisos comerciales) pueden presentar diferentes niveles de sonoridad debido a diferentes factores de cresta que requieren el ajuste del nivel de ítems consecutivos.

[0007] En una cadena clásica de difusión el usuario final recibe la pista de audio ya mezclada. Cualquier manipulación ulterior en el lado del receptor puede efectuarse solamente de una manera muy limitada. En la actualidad un pequeño conjunto de aspectos de metadatos de Dolby permite al usuario modificar alguna propiedad de la señal de audio.

[0008] Usualmente, las manipulaciones basadas en los metadatos mencionados en lo que precede, se aplican sin ninguna distinción selectiva de las frecuencias, ya que los metadatos tradicionalmente adjuntos a la señal de audio no proveen suficiente información para proceder de esta manera.

[0009] Además, sólo es posible manipular el conjunto de la corriente de audio propiamente dicha. Adicionalmente, no hay manera de adoptar y separar cada objeto de audio dentro de esta corriente de audio. En especial en entornos de escucha inadecuados, esto puede ser no satisfactorio.

[0010] En el modo medianoche, es imposible para el procesador de audio real distinguir entre ambientes y diálogo, debido a la ausencia de información de guiado. Por ello, en el caso de ruidos de elevado nivel (que han ser comprimidos/limitados en cuanto a nivel sonoro), también los diálogos se manipularán en paralelo. Este podría ser perjudicial para la inteligibilidad del habla.

[0011] Aumentar el nivel del diálogo en comparación con el sonido ambiente, ayuda a mejorar la percepción del habla en especial para las personas con problemas de audición. Esta técnica funciona solamente si la señal de audio se halla realmente separada en componentes de diálogo y ambiente en el lado de receptor, además de la información de control de propiedad. Si sólo hay una señal de mezclado descendente de estéreo disponible, ya no es posible aplicar una separación ulterior adicional para diferenciar y manipular la información de habla por separado.

[0012] Las soluciones de mezclado descendente actuales permiten una sintonización dinámica del nivel de estéreo para los canales central y de “surround”. Pero para cualquier configuración variante de altavoces en lugar de estéreo, no hay una descripción real desde el transmisor acerca de cómo mezclar descendentemente la fuente final

de audio de multicanal. El contar solamente con una fórmula por defecto dentro del decodificador lleva a cabo el mezclado de las señales de una manera muy inflexible.

[0013] En todos los escenarios descritos, por lo general existen dos enfoques diferentes. El primer enfoque es que, cuando se genera la señal de audio que debe transmitirse, se mezcla descendentemente un conjunto de objetos de audio en forma de un canal mono, estéreo o multicanal. Esta señal, que de transmitirse a un usuario de esta señal mediante radiodifusión, por medio de cualquier otro protocolo de transmisión o mediante una distribución sobre un medio de almacenamiento legible por computadora, tiene normalmente una cantidad de canales que es más pequeña que la cantidad de objetos de audio que habían sido mezclados descendentemente por un técnico sonidista, por ejemplo en el entorno de un estudio. Además, puede haber metadatos adjuntos destinados a permitir varias modificaciones diferentes, pero estas modificaciones sólo pueden aplicarse a la totalidad de la señal transmitida o, si la señal transmitida tiene varios canales transmitidos diferentes, a los canales transmitidos individuales como un conjunto. Sin embargo, dado que tales canales transmitidos son siempre superposiciones de varios objetos de audio, no es posible en absoluto una manipulación individual de determinado objeto de audio, sin dejar de manipular otro objeto de audio.

[0014] El otro enfoque consiste en no realizar el mezclado descendente de objetos, sino transmitir las señales de objeto de audio, como canales transmitidos separados. Un escenario de este tipo funciona bien, cuando la cantidad de objetos de audio es pequeña. Si, por ejemplo, hay solamente cinco objetos de audio, entonces es posible transmitir estos cinco objetos de audio diferentes por separado entre sí, dentro de un escenario 5.1. Es posible asociar metadatos con estos canales que indiquen la naturaleza específica de un objeto/canal. Entonces, en el lado del receptor, los canales transmitidos pueden manipularse sobre la base de los metadatos transmitidos.

[0015] Una desventaja de este enfoque es que no es compatible inversamente y que sólo funciona bien en el contexto de una pequeña cantidad de objetos de audio. Si la cantidad de objetos de audio aumenta, la velocidad de bits requerida para transmitir todos los objetos en forma de pistas de audio explicitas aumenta rápidamente. Este incremento de la velocidad de bits no es específicamente útil en el contexto de las aplicaciones de radiodifusión.

[0016] Por ello, los enfoques actuales eficientes de velocidad de bits no permiten una manipulación individual de objetos de audio distintos. Una manipulación individual de este tipo sólo se permite si se desea transmitir cada objeto por separado. Sin embargo, este enfoque no es eficiente en materia de velocidad de bits, por lo que no es específicamente factible en los escenarios de difusión.

[0017] Un objeto de la presente invención es el de proveer una solución a estos problemas que sea eficiente en cuanto a la velocidad de bits, pero también flexible.

[0018] De acuerdo con el primer aspecto de la presente invención, este objetivo se logra mediante un aparato para generar al menos una señal de salida de audio que representa una superposición de al menos dos objetos de audio diferentes, que comprende: un procesador para procesar una señal de entrada de audio a efectos de proveer una representación de objetos de la señal de entrada de audio, en el que los al menos dos objetos de audio diferentes se separan entre si, los al menos dos objetos de audio diferentes se encuentran disponibles como señales de objeto de audio separados, y los al menos dos objetos de audio separados pueden manipularse independientemente entre sí; un manipulador de objetos para manipular la señal objeto de audio o una señal de objeto de audio mejorada de al menos un objeto de audio basado en metadatos basados en objetos de audio, relacionados con el al menos un objeto de audio para obtener una señal del objeto de audio manipulada o una señal del objeto de audio mixta manipulada para el al menos un objeto de audio; y un mezclador de objetos para mezclar la representación de objetos por medio de la combinación del objeto de audio manipulado con un objeto de audio no modificado o con un objeto de audio manipulado diferente que haya sido manipulado de una manera diferente del al menos un objeto de audio.

[0019] De acuerdo con un segundo aspecto de la presente invención, este objetivo se logra mediante este procedimiento para la generación de al menos una señal de salida de audio que representa una superposición de al menos dos objetos de audio diferentes, que comprende: procesar una señal de entrada de audio a efectos de proveer una representación de objetos de la señal de entrada de audio, en el que los al menos dos objetos de audio diferentes se separan entre si, los al menos dos objetos de audio diferentes se encuentran disponibles en forma de señales objeto de audio separados, y los al menos dos objetos de audio diferentes pueden manipularse independientemente entre sí; manipular la señal objeto de audio o una señal de objeto de audio mezclada de al menos un objeto de audio sobre la base de metadatos basados en objetos de audio que se refieren a por menos un objeto de audio, para obtener una señal de objeto de audio manipulada o una señal de objetos de audio manipulada mixta para el al menos un objeto de audio; y mezclar la representación de objetos por medio de la combinación del objeto de audio manipulado con un objeto de audio no modificado o con un objeto de audio manipulado diferente que haya sido manipulado de una manera diferente del al menos un objeto de audio.

[0020] De acuerdo con un tercer aspecto de la presente invención, este objetivo se logra mediante un aparato para generar una señal de audio codificada que representa una superposición de al menos dos objetos de audio diferentes, que comprende: un formateador de corrientes de datos para formatear una corriente de datos de manera

que la corriente de datos comprenda una señal de mezclado descendente de objetos que representa una combinación de los al menos dos objetos de audio diferentes, y, como información secundaria, metadatos que se refieren a al menos uno de los objetos de audio diferentes.

[0021] De acuerdo con un cuarto aspecto de la presente invención, este objetivo se logra mediante un procedimiento para la generación de una señal de audio codificada que representa una superposición de al menos dos objetos de audio diferentes, que comprende: formatear una corriente de datos de manera tal que la corriente de datos comprenda una señal de mezclado descendente de objetos que representa una combinación de los al menos dos objetos de audio diferentes, y, como información secundaria, metadatos que se refieren a al menos uno de los objetos de audio diferentes.

[0022] Otros aspectos de la presente invención se refieren a programas de computadora que permiten implementar los procedimientos inventivos y un medio de almacenamiento legible por computadora que tiene almacenado en él una señal de mezclado descendente de objetos, y como información secundaria, datos de parámetros de los objetos y metadatos para uno o más objetos de audio incluidos en la señal de mezclado descendente de objetos.

[0023] La presente invención se basa en la conclusión de que una manipulación individual de señales de objetos de audio separadas o de conjuntos separados de señales de objetos de audio mixtas, permite un procesamiento relacionado con los objetos, sobre la base de metadatos relacionados con objetos. De acuerdo con la presente invención, el resultado de la manipulación no se emite directamente a un altavoz, sino que se provee a un mezclador de objetos, el que genera señales egresadas para un determinado escenario de renderización, en el que las señales egresadas se generan mediante una superposición de al menos un señal de objeto manipulado o de un conjunto de señales de objeto mezcladas junto con otras señales de objeto mezcladas y/o una señal de objeto no modificada. Por supuesto, no es necesario manipular cada objeto, sino que, en algunos casos, puede ser suficiente manipular un sólo objeto y no manipular otro objeto de la pluralidad de objetos de audio. El resultado de la operación del mezclado de objetos es una señal de salida de audio o una pluralidad de señales de salida de audio, que están basadas en objetos manipulados. Estas señales de salida de audio pueden transmitirse a altavoces, o pueden almacenarse para uso ulterior, o pueden aún transmitirse a un receptor adicional, en función del escenario de aplicación específico.

[0024] Es preferible que la señal ingresada en el dispositivo de manipulación/mezclado de acuerdo con la invención sea una señal de mezclado descendente generada mediante el mezclado descendente de una pluralidad de señales de objeto de audio. La operación del mezclado descendente puede ser controlada en metadatos para cada objeto individualmente, o puede estar sin controlar; puede ser el mismo para cada objeto. En el primer caso, la manipulación del objeto de acuerdo con los metadatos es la operación de mezclado individual de objeto controlado y específico en cuanto a objeto, en la que se genera una señal de componente de altavoz representativa de este objeto. Es preferible que también se provean parámetros espaciales de objeto, que pueden utilizarse para reconstruir las señales originales mediante versiones aproximadas de las mismas que utilicen la señal de mezclado descendente de objeto transmitida. En tal caso, el procesador para procesar una señal de entrada de audio para proveer una representación de objetos de la señal de entrada de audio opera de manera de calcular versiones reconstruidas del objeto de audio original sobre la base de los datos paramétricos, pudiendo estas señales de objeto aproximadas seguidamente ser manipuladas individualmente mediante metadatos basados en objetos.

[0025] Es preferible que también se provea información sobre la renderización de objetos, y que la información sobre la renderización de objetos incluya información acerca del establecimiento previsto sobre la reproducción de audio e información acerca del posicionamiento de los objetos de audio individuales dentro del escenario de reproducción. Sin embargo, hay formas de realización específicas que también pueden funcionar sin estos datos sobre la ubicación de los datos. Tales configuraciones abarcan por ejemplo la provisión de posiciones estacionarias para los objetos, que pueden establecerse de manera fija y que pueden ser objeto de una negociación entre un transmisor y un receptor para una pista de audio completa.

Breve descripción de los dibujos

[0026] A continuación se exponen formas de realización preferidas de la presente invención en el contexto de las figuras adjuntas, en las cuales:

La figura 1 ilustra una forma de realización preferida de un aparato para generar al menos una señal de salida de audio;

La figura 2 ilustra una implementación preferida del procesador de la Figura 1;

La figura 3a ilustra una forma de realización preferida del manipulador para manipular señales de objetos;

La figura 3b ilustra una implementación preferida del mezclador de objetos en el contexto de un manipulador como el ilustrado en la Figura 3a;

La figura 4 ilustra una configuración de procesador/manipulador/objeto en una situación en la que la manipulación se lleva a cabo subsiguientemente a un mezclado descendente de objetos, pero antes de un mezclado final de los objetos;

La figura 5a ilustra una forma de realización preferida de un aparato para generar una señal de audio codificada;

La figura 5b ilustra una señal de transmisión que tiene un mezclado descendente de objetos, metadatos basados en objetos, y parámetros espaciales para los objetos;

La figura 6 ilustra un mapa que indica varios objetos de audio identificado mediante un determinado ID, que tiene un archivo de audio de objetos (object audio file), y una matriz adjunta de información de audio, **E**;

La figura 7 ilustra una explicación de una matriz de covarianza de objetos, **E** de la Figura 6:

La figura 8 ilustra una matriz de mezclado descendente y un codificador de objetos de audio controlado por la matriz de mezclado descendente **D**;

La figura 9 ilustra una matriz de renderización teórica deseada **A** normalmente provista por un usuario y un ejemplo de un escenario específico de renderización teórica deseada;

La figura 10ilustra una forma de realización preferida de un aparato para generar al menos una señal de salida de audio de acuerdo con otro aspecto de la presente invención;

La figura 11a ilustra otra forma de realización;

La figura 11b ilustra otra forma de realización más;

La figura 11c ilustra otra forma de realización más;

La figura 12a ilustra un escenario dado a título de aplicación; y

La figura 12b ilustra otro escenario, dado a título de ejemplo, para una aplicación.

Descripción detallada de las formas de realización preferidas

[0027] Para enfrentar los problemas mencionados arriba, un enfoque preferido consiste en proveer metadatos adecuados junto con dichas pistas de audio. Tales metadatos pueden consistir en información para controlar los tres factores siguientes (los tres D “clásicos”):

- *dialog normalization* (normalización del diálogo);
- *dynamic range control* (control dinámico del margen);
- *downmix* (mezclado descendente).

[0028] Dichos metadatos de audio ayudan al receptor a manipular la señal de audio recibida sobre la base de los ajustes llevados a cabo por un escucha (listener). Para diferenciar este tipo de metadatos de audio de otros (por ejemplo metadatos descriptivos tales como Author (Autor), Title (Título),…) se los denomina usualmente como “Metadatos Dolby” (por cuanto hasta ahora sólo han sido implementados por Dolby). Por lo tanto, solamente se tienen en cuenta este tipo de metadatos de audio, y se los denomina simplemente “metadatos”.

[0029] Los metadatos de audio son información de control adicional que se transporta junto con el programa de audio y que tiene información esencial acerca del audio, a un receptor. Los metadatos proveen muchas funciones importantes que incluyen el control dinámico del margen audible para entornos de escucha que no son ideales, con concordancia de nivel entre programas, información de mezclado descendente para la reproducción de audio multicanal a través de una menor cantidad de locutores y otra información.

[0030] Los metadatos proveen las herramientas necesarias para que los programas de audio se reproduzcan de manera exacta y artística en muchas situaciones de escucha diferentes, que abarcan desde teatros domésticos completamente equipados al entretenimiento en vuelo, independientemente de la cantidad de canales de locutor, calidad del equipo de reproducción, o nivel relativo del ruido ambiente.

[0031] Si bien un técnico o un productor de contenido se esfuerzan en proveer la máxima calidad de audio en su programa, ella o él no tiene control sobre el amplio conjunto de los circuitos electrónicos de los consumidores ni sobre los entornos de escucha que intentarán reproducir la pista sonora original. Los metadatos proveen al técnico o

al productor de contenido un mayor control sobre cómo se reproduce y se disfruta su trabajo en casi cualquier entorno de escucha concebible.

[0032] Los metadatos de Dolby son un formato especial para proveer información destinada a controlar los tres factores mencionados.

[0033] Las tres principales funcionalidades de los metadatos de Dolby son:

- Normalización de los diálogos, para lograr un nivel promedio de diálogo a largo plazo dentro de una presentación, que frecuentemente consiste en diferentes tipos de programa, tales como una película de largometraje, avisos comerciales, etc.

- Control Dinámico del Margen Audible, a efectos de satisfacer la mayor parte de la audiencia con una compresión de audio placentera pero de manera de al mismo tiempo permitir a cada cliente individual controlar el aspecto dinámico de la señal de audio y ajustar la compresión a su entorno de escucha personal.

- Mezclado descendente para mapear los sonidos de una señal de audio multicanal a dos o más canales en el caso en que no se disponga de un equipo de reproducción de audio multicanal.

[0034] Se utilizan metadatos de Dolby junto con Dolby Digital (AC–3) y Dolby E. El formato de los metadatos de audio de Dolby se ha descrito en [16] Dolby Digital (AC–3) y tiene por objeto la traducción de audio en el hogar mediante difusión de televisión digital (en definición estándar o alta definición), DVD u otros medios.

[0035] El Dolby Digital puede transportar cualquier cosa desde un sólo canal de audio hasta un programa de canal 5.1, que incluye metadatos. Tanto en la televisión digital como en DVD, se lo utiliza comúnmente para la transmisión de estéreo así como de programas de audio discretos de 5.1 completos.

[0036] El Dolby E está específicamente destinado a la distribución de audio multicanal dentro de los entornos de producción y distribución profesionales. En cualquier momento antes de su entrega al consumidor, el Dolby E es el procedimiento preferido para la distribución de audio multicanal/multiprograma con video. El Dolby E puede llevar hasta ocho canales de audio discretos configurados en cualquier cantidad de configuraciones de programa individuales (inclusive metadatos para cada uno de ellos) dentro de una infraestructura existente de audio digital de dos canales. A diferencia del Dolby Digital, el Dolby E puede tratar muchas generaciones de codificar/decodificar, y es síncrono con la velocidad de cuadros de video. Lo mismo que el Dolby Digital, el Dolby E transporta metadatos para cada programa de audio codificado dentro de la corriente de datos. El uso del Dolby E permite decodificar, modificar, y recodificar la corriente de datos de audio resultante, sin una degradación audible. Dado que la corriente de Dolby E es sincrónica con la velocidad de los cuadros de video, se la puede encaminar, conmutar, y editar en un entorno de difusión profesional.

[0037] Además de este medio provisto junto con MPEG AAC para llevar a cabo el control dinámico del margen audible y controlar la generación de mezclado descendente.

[0038] A efectos de manipular el material de fuente con niveles de pico, niveles medios y margen audible dinámico, variables, de una manera que minimiza las variabilidades para el consumidor, es necesario controlar el nivel reproducido de manera que, por ejemplo, el nivel del diálogo o en nivel medio de la música se ajuste a un nivel de reproducción controlado por el consumidor, independientemente de cómo se originó el programa. Adicionalmente, no todos los consumidores serán capaces de oír los programas en un buen entorno (es decir, de bajo ruido), sin restricciones de a que valor eleven el nivel del sonido. El entorno de los automóviles, por ejemplo, tiene un nivel de alto ruido y por ello puede preverse que el oyente deseará reducir el margen de niveles que de otra manera se reproducirían.

[0039] Por estas dos razones, el control dinámico del margen audible ha de estar disponible dentro de la especificación del AAC. Para lograr esto, es necesario acompañar el audio de velocidad de bits reducida con datos utilizados para ajustar y controlar el margen audible dinámico de los ítems del programa. Este control debe especificarse con respecto a un nivel de referencia y en relación con los elementos importantes del programa, por ejemplo el diálogo.

[0040] Los aspectos del control dinámico del margen audible son como sigue:

1.– El Control Dinámico del Margen es completamente opcional. Por ello, con una sintaxis correcta, no hay cambios en la complejidad para aquellos que no deseen invocar el DRC.

2.– Los datos de audio de velocidad de bits reducida se transmiten con el margen dinámico completo del material de fuente, con datos de soporte para ayudar en el control dinámico del margen audible.

3.– El control dinámico del margen audible puede enviarse cada cuadro a efectos de reducir a un mínimo la latencia en el ajuste de las ganancias de reproducción.

4.– Los datos del control dinámico del margen audible se envían utilizando el aspecto de “fill_element” del AAC.

5.– El nivel de referencia se define como escala completa.

6.– El nivel de referencia del programa se transmite a efectos de permitir la paridad de niveles entre los niveles de reproducción de las diferentes fuentes y a efectos de proveer una referencia acerca de cuál control dinámico del margen audible puede aplicarse. Es este aspecto de la señal de fuente que es el que está más relacionado con la impresión subjetiva de la sonoridad de un programa, tal como el nivel del contenido de diálogo de un programa o el nivel promedio de un programa de música.

7. El Nivel de Referencia del Programa representa aquel nivel de programa que puede reproducirse en un nivel establecido con respecto al Nivel de Referencia en el hardware del consumidor a efectos de lograr la paridad del nivel de reproducción. Con respecto a esto, las porciones más silenciosas del programa pueden incrementarse en cuanto a nivel, y es posible reducir el nivel de las porciones más sonoras del programa.

8.– El Nivel de Referencia del Programa se especifica dentro del margen o intervalo de 0 a –31.75 dB con respecto al Nivel de Referencia.

9.– El Nivel de Referencia del Programa utiliza un 7 bit de archivo con pasos de 0,25 dB.

10.– El control dinámico del margen audible se especifica dentro del intervalo ± 31,75 dB.

11.– El control dinámico del margen audible utiliza un campo de 8 bits (1 signo, 7 magnitudes) con pasos de 0,25 dB.

12.–El control dinámico del margen audible puede aplicarse a la totalidad de los coeficientes espectrales de un canal de audio o bandas de frecuencia como una entidad individual, o es posible dividir los coeficientes en diferentes bandas de factores de escala, controlándose cada uno de ellos mediante conjuntos separados de datos de control dinámico del margen audible.

13.–El control dinámico del margen audible puede aplicarse a todos los canales (de una corriente de bits estéreo o multicanal) como una única entidad, o es posible dividirla, controlándose los conjuntos de canales por separado mediante conjuntos separados de datos de control dinámico del margen audible.

14.– Si falta un conjunto previsto de datos del control dinámico del margen audible, deberían utilizarse los valores válidos más recientemente recibidos.

15.– No todos los elementos de los datos del control dinámico del margen audible se envían cada vez. Por ejemplo, el Nivel de Referencia del Programa puede enviarse solamente en promedio cada 200 ms.

16.–Cuando sea necesario, la Capa de Transporte provee detección/protección de los datos.

17.– El usuario recibirá los medios para alterar la cantidad del control dinámico del margen audible, presente en la corriente de bits, que se aplica al nivel de la señal.

[0041] Además de la posibilidad de transmitir canales de mezclado descendente mono o estéreo separados en una transmisión de canales de tipo 5.1–, el AAC también permite una generación automática de mezclado descendente a partir de la pista fuente de 5 canales. En este caso se omitirá el canal LFE.

[0042] Este procedimiento de matriz de mezclado descendente puede ser controlado por el editor de una pista de audio con un pequeño conjunto de parámetros que definen la cantidad de los canales posteriores adicionados al mezclado descendente.

[0043] El procedimiento de la matriz–mezclado descendente se aplica solamente para mezclar una configuración desde 3 altavoces de frontales / 2 altavoces posteriores y un programa de 5 canales, hasta un programa estéreo o mono. No puede aplicarse a ningún programa que no tenga la configuración 3/2.

[0044] Dentro del MPEG se proveen varios medios para controlar la renderización (rendering) de audio en el lado del receptor.

[0045] Se provee una tecnología genérica mediante un lenguaje de descripción de escena, por ejemplo BIFS y LASeR. Se utilizan ambas tecnologías para renderizar elementos audio–visuales a partir de objetos codificados separados, en una escena de reproducción.

[0046] El BIFS se normaliza en [5] y el LASeR en [6].

[0047] El MPEG–D trata principalmente con descripciones paramétricas (es decir, metadatos).

- para generar audio multicanal sobre la base de representaciones de audio mezclados descendentemente (MPEG Surround); y

- generar parámetros de MPEG Surround sobre la base de objetos de Audio (Codificación Espacial de MPEG de Objetos de Audio)

[0048] El MPEG Surround aprovecha las diferencias intercanal en cuanto a nivel, fase y coherencia equivalente a las indicaciones de ILD, ITD y IC, para capturar la imagen espacial de una señal de audio multicanal con respecto a una señal de mezclado descendente transmitida, y codifica estos indicios en una manera muy compacta de manera que los indicios y la señal transmitida puedan decodificarse para sintetizar una representación multicanal de alta calidad. El codificador MPEG Surround recibe una señal de audio multicanal, siendo N la cantidad de canales de entrada (por ejemplo, 5.1). Un aspecto clave del proceso de codificación es que una señal de mezclado descendente, xt1 y xt2, que típicamente son estéreo (pero que también podrían ser mono) se deriva desde la señal ingresada multicanal, y es esta señal de mezclado descendente que se comprime para su transmisión por medio del canal en lugar de la señal de multicanal. El codificador puede ser capaz de aprovechar el proceso del mezclado descendente, de manera de crear un equivalente real de la señal de multicanal en el mezclado descendente mono o estéreo, y también crea la mejor decodificación multicanal posible sobre la base del mezclado descendente y de los indicios espaciales codificados. Como alternativa, el mezclado descendente podría suministrarse externamente. El proceso de codificación de MPEG Surround no toma en cuenta el algoritmo de compresión utilizado para los canales transmitidos; podría ser cualquiera de entre una cantidad de algoritmos de compresión de alta performance tales como el MPEG–1 Layer (Capa) III, MPEG–4 AAC o AAC de alta eficiencia de MPEG–4, o aún podría ser el PCM.

[0049] La tecnología de MPEG surround soporta la codificación paramétrica muy eficiente de señales de audio de multicanal. La idea del MPEG SAOC es la de aplicar suposiciones básicas junto con una representación de parámetros similar para la codificación paramétrica muy eficiente de objetos de audio individuales (pistas). Adicionalmente, se incluye una funcionalidad de renderización para renderizar de manera interactiva los objetos de audio en una escena acústica de sistemas de reproducción (1.0, 2.0, 5.0, ... para altavoces o biaural para auriculares). El SAOC ha sido diseñado para transmitir una cantidad de objetos de audio en una señal de mezclado descendente mono o estéreo conjunta a efectos de permitir ulteriormente una reproducción de los objetos individuales en una escena de audio renderizada acústicamente. Para esta finalidad, el SAOC codifica OLDS (Object Level Differences, Diferencias de Nivel de los Objetos), IOCs (Inter–Object Cross Coherences, Coherencias Cruzadas Interobjetos), y DCLDs (Downmix Channel Level Differences, Diferencias de Niveles de Canales de Mezclado Descendente). El decodificador SAOC convierte la representación de los parámetros de SAOC en una representación de parámetros de MPEG Surround, que seguidamente se decodifica junto con la señal de mezclado descendente mediante un decodificador de Surround a efectos de producir la escena de audio deseada. El usuario controla interactivamente este proceso a efectos de alterar la representación de los objetos de audio en la escena de audio resultante. Entre las numerosas aplicaciones concebibles para el SAOC, a continuación se enumeran algunos escenarios típicos:

[0050] Los consumidores pueden crear nuevas mezclas interactivas personales mediante el uso de un escritorio virtual de mezclado. Es posible por ejemplo atenuar determinados instrumentos para su reproducción simultánea (como el Karaoke), es posible modificar la mezcla original de acuerdo con el gusto personal, es posible ajustar el nivel de diálogo en las películas/ difusiones para que el habla sea más inteligible, etc.

[0051] Para los juegos interactivos, el SAOC es una manera de almacenamiento y muy eficiente desde el punto de vista de la computación, para reproducir pistas sonoras. El movimiento en la escena virtual se refleja mediante una adaptación de los parámetros de renderización de los objetos. Los juegos de múltiples jugadores basados en red, se benefician de la eficiencia de la transmisión mediante el uso de una corriente de SAOC para representar todos los objetos de sonido que sean externos con respecto a la terminal de un jugador determinado.

[0052] En el contexto de esta solicitud, la expresión “objeto de audio” también comprende un “tallo o vástago” (“stems”) conocido en los escenarios de producción de sonido. En particular, los tallos son los componentes individuales de una mezcla, guardados por separado (usualmente en un disco) a los fines de su uso en una nueva mezcla. Los tallos relacionados son típicamente generados desde la misma instalación original. Los ejemplos podrían ser un tallo de tambor (incluye todos los instrumentos de tipo tambor relacionados en una mezcla), un tallo de vocales (incluye solamente las pistas de vocales) o un tallo de ritmo (incluye todos los instrumentos relacionados con la base rítmica tales como tambores, guitarra, teclado, …).

[0053] La estructura de telecomunicaciones actual es monofónica y puede extenderse en su funcionalidad. Los terminales equipados con una extensión SAOC recogen varias fuentes de sonido (objetos) y producen una señal de mezclado descendente monofónica, que se transmite de una manera compatible mediante el uso de los

codificadores (de habla) existentes. La información secundaria puede transportarse de una manera incorporada, compatible inversamente. Los terminales heredados seguirán produciendo salida monofónica mientras que los terminales habilitados para SAOC pueden renderizar una escena acústica y por lo tanto incrementar la inteligibilidad mediante la separación espacial de los parlantes diferentes ("cocktail party effect", efecto de la fiesta de cócteles).

[0054] En el capitulo siguiente se describe una revisión de las aplicaciones de metadatos de audio Dolby realmente disponibles:

Modo de medianoche (Midnight Mode)

[0055] Como se mencionó anteriormente, hay muchos escenarios en los que el escucha no desea una señal muy dinámica. Por ello ella o él puede activar el denominado "modo de medianoche" en su receptor. Seguidamente se aplica un compresor sobre la señal de audio total. Para controlar los parámetros de este compresor, se evalúan los metadatos transmitidos, y se los aplica a la señal de audio total.

Audio limpio

[0056] Otro escenario es el de las personas con problemas de audición o hipoacúsicos, que no deseen tener un elevado ruido ambiente dinámico, sino que desean una señal por demás limpia que contenga diálogos ("Clean Audio"). Este modo también puede habilitarse mediante el uso de metadatos.

[0057] En [15] – Anexo E se define una solución actualmente presentada. En este caso, se trata del equilibrio entre la señal principal estéreo y el canal de descripción de diálogo mono adicional mediante un conjunto individual de parámetros de nivel. La solución propuesta basada en una síntaxis separada, recibe en DVB la denominación de servicio complementario de audio.

Mezclado descendente

[0058] Hay parámetros de metadatos separados que gobiernan el mezclado descendente L/R. Determinados parámetros de metadatos permiten al técnico seleccionar cómo se construye el mezclado descendente de estéreo y cuál análogo de estéreo se prefiere. Aquí, el centro y el nivel de mezclado descendente de surround definen el equilibrio final de mezclado de la señal de mezclado descendente para cada decodificador.

[0059] La Figura 1 ilustra un aparato para generar al menos una señal de salida de audio que representa una superposición de al menos dos objetos de audio diferentes, de acuerdo con una forma de realización preferida de la presente invención. El aparato de la Figura 1 comprende un procesador 10 para procesar una señal de entrada de audio 11 a efectos de proveer una representación de objetos 12 de la señal de entrada de audio, en la que los al menos dos objetos de audio diferentes están separados entre si, en el que los al menos dos objetos de audio diferentes están disponibles como señales de objetos de audio separadas y en el que los al menos dos objetos de audio diferentes pueden manipularse independientemente entre sí.

[0060] La manipulación de la representación de objetos se lleva a cabo en un manipulador de objetos 13 para manipular la señal del objeto de audio o una representación mixta de la señal de objeto de audio de al menos un objeto de audio sobre la base de metadatos de objeto basados en objeto de audio, 14, que se refieren a al menos un objeto de audio. El manipulador de objetos de audio, 13, está adaptado para obtener la representación de una señal de objeto de audio manipulada o de una señal de objetos de audio manipulada mixta, 15, para el al menos un objeto de audio.

[0061] Las señales generadas por el manipulador de objetos se ingresan en un mezclador de objetos, 16, para mezclar la representación de objetos mediante la combinación del objeto de audio manipulado con un objeto de audio no modificado o con un objeto de audio diferente manipulado, en donde el objeto de audio diferente manipulado ha sido manipulado de una manera diferente del al menos un objeto de audio. El resultado del mezclador de objetos comprende una o más señales de salida de audio, 17a, 17b, 17c. Es preferible que las una o más señales de audio, 17a a 17c estén diseñadas para un establecimiento de renderización específico tal como un establecimiento de renderización mono, un establecimiento de renderización estéreo, un establecimiento de renderización multicanal que comprende tres o más canales tales como un establecimiento de surround que requiere al menos cinco o al menos siete señales de salida de audio diferentes.

[0062] La Figura 2 ilustra una implementación preferida del procesador 10 para procesar la señal de entrada de audio. Es preferible que la señal de entrada de audio 11 se implemente como un mezclado descendente de objeto 11 obtenido mediante un mezclador descendente de objetos, 101a de la Figura 5a que se describe más adelante. En esta situación, el procesador recibe adicionalmente parámetros de objeto, 18, tales como por ejemplo generados por el calculador de parámetros de objeto, 101b, en la Figura 5b como se describe más adelante. Seguidamente, el procesador 10 se halla en la posición de calcular señales de objetos de audio separadas, 12. La cantidad de señales de objeto de audio 12 puede ser superior a la cantidad de canales en el mezclado descendente de objetos, 11. El mezclado descendente de objetos, 11, puede incluir un mezclado descendente mono, un mezclado descendente

estéreo o aún un mezclado descendente que tiene más de dos canales. Sin embargo, el procesador 12 puede operar de manera de generar más señales de objeto de audio, 12, en comparación con la cantidad de señales individual en el mezclado descendente de objetos, 11. Debido al procesamiento paramétrico llevado a cabo por el procesador 10, las señales de objeto de audio no son una reproducción de los objetos de audio originales que se hallaban presentes antes de llevarse a cabo el mezclado descendente de objeto, 11, sino que las señales de objeto de audio son versiones aproximadas de los objetos de audio originales, en donde la exactitud de la aproximación depende del tipo de algoritmo de separación llevado a cabo en el procesador 10 y, por supuesto, de la exactitud de los parámetros transmitidos. Los parámetros de objeto preferidos son los parámetros conocidos de la codificación espacial de objetos de audio y un algoritmo de reconstrucción preferido para generar las señales de objetos de audio individualmente separados es el algoritmo de reconstrucción llevado a cabo de acuerdo con la norma de codificación espacial de objetos de audio. Posteriormente, en las Figuras 6 a 9 se expone una forma de realización preferida del procesador 10 y de los parámetros de objeto.

[0063] Las Figuras 3a y 3b ilustran conjuntamente una implementación en la que la manipulación del objeto se lleva a cabo ante del mezclado descendente de objeto para el establecimiento de la reproducción, mientras que la Figura 4 ilustra otra implementación en la que se lo lleva a cabo antes del manipulación, y el manipulación se lleva a cabo antes de la operación final del mezclado de objetos. El resultado del procedimiento en las Figuras 3a, 3b en comparación con la Figura 4, es el mismo, pero la manipulación del objeto se lleva a cabo a diferentes niveles en el escenario de procesamiento. Cuando la manipulación de las señales de objeto de audio es un problema en el contexto de la eficiencia y de los recursos de computación, se prefiere la forma de realización de las Figuras 3a/3b, ya que la manipulación de las señales de audio ha de llevarse a cabo únicamente sobre una única señal de audio en lugar de sobre una pluralidad de señales de audio como en la Figura 4. En una implementación diferente en la que podría haber un requerimiento de que el mezclado descendente de objeto debe ejecutarse mediante el uso de una señal de objeto no modificada, se prefiere la forma de realización preferida de la Figura 4 en la que la manipulación se lleva a cabo subsiguientemente al mezclado descendente de objetos, pero antes del mezclado descendente de objeto final para obtener las sales egresadas para, por ejemplo, el canal izquierdo L, el canal central C o el canal derecho R.

[0064] La Figura 3a ilustra la situación en la que el procesador 10 de la Figura 2 emite señales de objetos de audio separadas. Por lo menos una señal del objeto de audio tal como la señal para el objeto 1, se manipula en un manipulador 13a sobre la base de metadatos para este objeto 1. En función de la implementación, otros objetos tales como un objeto 2, también son manipulados por un manipulador 13b. Por supuesto, puede presentarse una situación en la que realmente existe un objeto tal como un objeto 3 que no experimenta una manipulación pero que sin embargo se genera mediante la separación de los objetos. El resultado del procesamiento de la Figura 3a consiste, para el caso de la Figura 3a, en dos señales de objeto manipulados y una señal no manipulada.

[0065] Estos resultados se ingresan en el mezclador de objetos, 16, que incluya una primera etapa de mezclador implementada como mezcladores descendentes de objetos, 19a, 19b, 19c, y que además comprende un segundo mezclador de objetos implementado por los dispositivos 16a, 16b, 16c.

[0066] La primera etapa del mezclador de objetos 16 incluye, para cada salida de la Figura 3a, un mezclador descendente de objetos tal como un mezclador descendente de objetos 19a para la salida 1 de la Figura 3a, un mezclador descendente de objetos 19b para la salida 2 de la Figura 3a, un mezclador descendente de objetos 19c para la salida 3 de la Figura 3a. La finalidad de los mezcladores descendente de objetos 19a a 19c es la “distribuir“ cada objeto a los canales de salida. Por ello, cada mezclador descendente de objetos 19a, 19b, 19c tiene una salida para una señal componente izquierda, L, una señal componente central, C, y una señal componente derecha, R. Por lo tanto, si por ejemplo el objeto 1 fuese el objeto individual, el mezclador descendente 19a sería un mezclador descendente directo y la salida del bloque 19a sería la misma que la salida final, L, C, R indicada en 17a, 17b, 17c. Es preferible que los mezcladores descendentes de objetos, 19a a 19c, reciban información de renderización indicada en 30, donde la información de renderización puede describir el establecimiento de la renderización, es decir, como en la forma de realización de la Figura 3e en la que solamente existen tres parlantes de salida. Estas salidas son un parlante izquierdo L, un parlante central C y un parlante derecha R. Si, por ejemplo, el establecimiento y renderización o establecimiento de reproducción comprende un escenario 5.1, entonces cada mezclador descendente de objetos tendría seis canales de salida, y existirían seis sumadores de manera tal que habría una señal de salida final para el canal izquierdo, una señal final de salida para el canal derecho, una señal final de salida para el canal central, una señal final de audio para el canal surround izquierdo, una señal final egresada para el canal surround derecho, y una señal final de salida para el canal de refuerzo de baja frecuencia.

[0067] Específicamente, los sumadores 16a, 16b, 16c están adaptados para combinar las señales componentes para el canal respectivo, que fueron generadas por los correspondientes mezcladores descendentes de objetos. Es preferible que esta combinación sea una muestra directa por simple suma, pero en función de la implementación, también es posible aplicar factores de ponderación. Por otra parte, es posible llevar a cabo las funcionalidades en las Figuras 3a, 3b en el dominio de las frecuencias o de las subbandas, por lo que los elementos 19a a 16c podrían operar en el dominio de las frecuencias y habría algún tipo de conversión de frecuencia/tiempo antes de emitirse realmente las señales a los parlantes en un establecimiento de reproducción.

[0068] La Figura 4 ilustra una implementación alternativa en la que las funcionalidades de los elementos 19a, 19b, 19c, 16a, 16b, 16c son similares a las de la forma de realización de la Figura 3b. Sin embargo, lo importante es que la manipulación que tuvo lugar en la Figura 3a antes del mezclado descendente de objetos, 19a, tiene ahora lugar subsiguientemente al mezclado descendente de objetos, 19a. Por lo tanto, la manipulación específica de los objetos, que se controla por los metadatos para el respectivo objeto, tiene lugar en el dominio del mezclado descendente, es decir, antes de la suma real de las señales componentes entonces manipuladas. Si se compara la Figura 4 con la Figura 1, se hace evidente que el mezclador descendente de objetos tal como 19a, 19b, 19c se implementará dentro del procesador 10, y que el mezclador de objetos 16 comprenderá los sumadores 16a, 16b, 16c. Si se implementa la Figura 4 y los mezcladores descendentes de objetos son parte del procesador, entonces el procesador recibirá, además de los parámetros de objeto 18 de la Figura 1, la información de renderización 30, es decir, la información sobre la posición de cada objeto de audio en el establecimiento de renderización e información adicional según el caso.

[0069] Además, la manipulación puede incluir la operación de mezclado descendente implementada por los bloques 19a, 19b, 19c. En esta forma de realización, el manipulador incluye estos bloques, y pueden tener lugar manipulaciones adicionales, pero las mismas no se requieren en ningún caso.

[0070] La Figura 5a ilustra una forma de realización del lado del codificador que puede generar una corriente de datos como se ilustra esquemáticamente en la Figura 5b. Específicamente, la Figura 5a ilustra un aparato para generar una señal de audio codificada, 50, que representa una superposición de al menos dos objetos de audio diferentes. Básicamente, el aparato de la Figura 5a ilustra una formateador de corrientes de datos, 51, para formatear la corriente de datos 50 de manera que la corriente de datos comprenda una señal de mezclado descendente de objetos, 52, que representa una combinación tal como una combinación ponderada o no ponderada, de los al menos dos objetos de audio. Además, la corriente de datos 50 comprende, como información secundaria, metadatos relacionados con objetos, 53, referidos a al menos uno de los objetos de audio diferentes. Es preferible que la corriente de datos, 50, además comprenda datos paramétricos, 54, que son selectivos en cuanto a tiempo y frecuencia, y que permitan una separación de alta calidad de la señal de mezclado descendente de objetos en varios objetos de audio; esta operación también recibe la denominación de operación de mezclado ascendente de los objetos llevado a cabo por el procesador de la Figura 1, como se expuso en lo que precede.

[0071] Es preferible que la señal de mezclado descendente de objetos, 52, sea generada por un mezclador descendente de objetos, 101a. Es preferible que los datos paramétricos 54 sean generados por un calculador de parámetros de objetos, 101b, y que los metadatos selectivos en cuanto a los objetos, 53, sean generados por un proveedor de metadatos selectivo en cuanto a objetos, 55. El proveedor de metadatos selectivo en cuanto a objetos, puede ser una entrada para recibir metadatos tal como los mismos son generados por un productor de audio en un estudio de sonido, o pueden ser datos generados mediante un análisis relacionado con objetos, que podrían llevarse a cabo subsiguientemente a la separación de los objetos. Específicamente, el proveedor de los metadatos selectivo en cuanto a objetos, podría implementarse para analizar la salida del objeto por el procesador 10 a efectos de, por ejemplo, descubrir si un objeto es un objeto de habla, un objeto de sonido o un objeto de sonido surround. Por lo tanto, podría analizarse un objeto de habla mediante alguno de los algoritmos de habla bien conocidos de la codificación del habla, y el análisis selectivo en cuanto a objetos podría implementarse para también descubrir objetos de sonido procedentes de instrumentos. Tales objetos de sonido son de una naturaleza de alta tonalidad, por lo que pueden distinguirse de los objetos de habla o de los objetos de sonido de surround. Los objetos de sonido de surround serán de una naturaleza más bien ruidosa que refleja el sonido de fondo que típicamente existe en, por ejemplo, las películas cinematográficas en las que, por ejemplo los ruidos de fondo son ruidos de tránsito callejero o cualquier otro ruido estacionario o cualquier otra señal de ruido, estacionaria o no estacionaria, que tenga un espectro de banda ancha tal como el que se genera cuando por ejemplo tiene lugar una escena de tiroteo en una sala de cine,.

[0072] Sobre la base de este análisis, se podría amplificar un objeto sonoro y atenuar los otros objetos a efectos de hacer énfasis en el habla ya que ello es útil para una mejor comprensión de la película para personas hipoacúsicas o de edad avanzada. Como se mencionó en lo que precede, otras implementaciones incluyen la provisión de los metadatos específicos para objetos tales como una identificación de objetos y los datos relacionados con el objeto por un técnico sonidista que genere la señal de mezclado descendente de objetos real sobre un CD o un DVD tal como un mezclado descendente de estéreo o un mezclado descendente de sonido surround.

[0073] La Figura 5d ilustra una corriente de datos dada a título de ejemplo, 50, que tiene, como información principal, el mezclado descendente de objetos mono, estéreo o multicanal, y que tiene como información secundaria, los parámetros de los objetos, 54, y los metadatos basados en objetos, 53, que son estacionarios en el caso de la sola identificación de objetos como habla o surround, o que son variables en el tiempo en el caso de la provisión de datos de nivel tales como metadatos basados en objetos requeridos por el modo de medianoche. Sin embargo, es preferible que los metadatos basados en objetos no se provean de una manera selectiva en frecuencias, a efectos de ahorrar velocidades de datos.

[0074] La Figura 6 ilustra una forma de realización de un mapa de objetos de audio que ilustra una cantidad de N objetos. En la explicación, dada a título de ejemplo, de la Figura 6, cada objeto tiene un ID de objeto, un

correspondiente archivo de objetos de audio y, lo que es importante, información de parámetros de objetos de audio que preferentemente se refiere a la energía del objeto de audio, y la correlación interobjetos de los parámetros de objetos de audio incluye una matriz de covarianza de objetos, **E,** para cada subbanda y para cada bloque de tiempo.

[0075] En la Figura 7 se ilustra un ejemplo para una matriz de información de objetos de audio, **E,** de este tipo. Los elementos en diagonal, e_{ii}, incluyen información de energía eléctrica o de potencia del objeto de audio i en la correspondiente subbanda y el correspondiente bloque de tiempo. A tal efecto, la señal de subbanda que representa un determinado objeto de audio, i, se ingresa en un calculador de energía o de potencia que pueden por ejemplo llevar a cabo una función de autocorrelación (acf, auto correlation function) para contener el valor e_{11} con alguna normalización, o sin ella. Como alternativa, es posible calcular la energía como la suma de los cuadrados de la señal a lo largo de una longitud determinada (es decir, el producto vectorial: ss*). De alguna manera el acf puede describir la configuración espectral de la energía, pero debido al hecho de que de cualquier manera se utiliza preferentemente una transformada T/F para la selección de frecuencias, el cálculo de la energía puede llevarse a cabo sin un acf para cada subbanda por separado. Por lo tanto, los elementos de diagonal principal de la matriz de parámetros de objetos de audio, **E**, indican una medida para la potencia de la energía de un objeto de audio en una determinada subbanda en un determinado bloque de tiempo.

[0076] Por otra parte, los elementos fuera de la diagonal, e_{ij} indican una respectiva medida de correlación entre los objetos de audio i, j en la correspondiente subbanda y el bloque de tiempo. De la Figura 7 es evidente que para las entradas realmente valuadas la matriz **E** es simétrica con respecto a la diagonal principal. En términos generales, esta matriz es una matriz de Hermite. El elemento de medida de correlación, e_{ij} puede calcularse, por ejemplo mediante una correlación cruzada de las dos señales de subbanda de los respectivos objetos de audio de manera que se obtiene una medida de correlación cruzada, que puede normalizarse o no. Pueden utilizarse otras medidas de correlación que no se calculan mediante una operación de correlación cruzada pero que se calculan de otras maneras para determinar la correlación entre dos señales. Por razones prácticas, todos los elementos de la matriz **E** se normalizan de manera que tengan magnitudes entre 0 y 1, donde 1 indica una máxima potencia o una máxima correlación y 0 indica una potencia mínima (potencia cero) y –1 indica una correlación mínima (fuera de fase).

[0077] La matriz de mezclado descendente, D, de magnitud $K \times N$ donde $K > 1$ determina la K señal de mezclado descendente de canal en la forma de una matriz con K filas mediante la multiplicación matricial

$$\mathbf{X} = \mathbf{DS} \quad (2)$$

[0078] La Figura 8 ilustra un ejemplo de una matriz de mezclado descendente **D** que tiene elementos de matriz de mezclado descendente, d_{ij}. Un elemento d_{ij} de este tipo indica si una porción o la totalidad del objeto j se halla incluido o no en la señal de mezclado descendente de objetos. Si, por ejemplo, d_{12} es igual a cero, esto significa que el objeto 2 no está incluido en la señal de mezclado descendente de objetos, 1. Por otra parte, un valor de d_{23} igual a 1 indica que el objeto 3 se halla incluido por completo en la señal de mezclado descendente de objetos, 2.

[0079] Son posibles valores de elementos de matriz de mezclado descendente entre 0 y 1. Específicamente, el valor de 0,5 indica que un determinado objeto se halla incluido en una señal de mezclado descendente, pero con solamente la mitad de su energía. Por lo tanto, cuando un objeto de audio tal como el objeto número 4 se halla distribuido igualmente a ambos canales de señales de mezclado descendente, entonces d_{24} y d_{14} serian iguales a 0,5. Esta manera de mezclado descendente es una operación de mezclado descendente que conserva energía, que para algunas situaciones es preferible. Sin embargo, como alternativa también puede utilizarse un mezclado descendente que no conserve energía, en la que la totalidad del objeto de audio se introduce en el canal de mezclado descendente izquierdo y en el canal de mezclado descendente derecha de manera que la energía de este objeto de audio se ha doblado con respecto a los otros objetos de audio dentro de la señal de mezclado descendente.

[0080] En la porción inferior de la Figura 8, se ilustra un diagrama esquemático del codificador de objetos, 101, de la Figura 1. Específicamente, el codificador de objetos, 101 incluye dos porciones diferentes 101a y 101b. La porción 101a es un mezclador descendente que preferentemente lleva a cabo una combinación lineal ponderada de objetos de audio 1, 2, …, N, y la segunda porción del codificador de objetos 101 un calculador de parámetros de objetos de audio, 101b, que calcula la información de parámetros de objetos de audio tal como la matriz **E** para cada bloque de tiempo o subbanda a efectos de proveer la energía de audio y la información de correlación que es una información paramétrica y puede, por ello, transmitirse con una baja velocidad de los datos o que puede almacenarse consumiendo una pequeña cantidad de recursos de memoria.

[0081] La matriz de renderización de objetos, controlada, $\mathbf{A}$, de magnitud $M \times N$ determina la renderización teórica de los objetos de audio en la forma de una matriz con M filas por medio de de la multiplicación matricial:

$$\mathbf{Y} = \mathbf{AS} \quad (3)$$

[0082] A lo largo de la siguiente derivación se supondrá que $M = 2$ ya que el enfoque es sobre la renderización de estéreo. Dada una matriz de renderización inicial a más de dos canales, y una regla de mezclado descendente a partir de dichos varios canales en dos canales, es evidente para las personas con pericia en la técnica derivar la correspondiente matriz de renderización $\mathbf{A}$ de magnitud $2 \times N$ para la renderización de estéreo. También se supondrá por razones de sencillez que $K = 2$, por lo que el mezclado descendente de objetos es también una señal estéreo. El caso de un mezclado descendente de objetos estéreo es además el caso especial más importante en términos de escenarios de aplicación.

[0083] La Figura 9 ilustra una explicación detallada de la matriz de renderización teórica deseada, **A**. En función de la aplicación, la matriz de renderización teórica deseada **A** puede ser provista por el usuario. El usuario tiene libertad completa para indicar dónde debería estar situado un objeto de audio de una manera virtual para un establecimiento de reproducción. La fuerza del concepto de objeto de audio es que la información de mezclado descendente y la información de los parámetros de los objetos de audio dependen por completo de una ubicación específica de los objetos de audio. Esta localización de objetos de audio se provee por un usuario en la forma de información de renderización teórica deseada. Es preferible que la información de renderización teórica deseada pueda implementarse como una matriz de renderización teórica deseada **A** que puede estar en la forma de la matriz de la Figura 9. Específicamente, la matriz de renderización **A** tiene M líneas y N columnas, siendo M igual a la cantidad de canales en la señal de salida renderizada, y siendo N igual a la cantidad de objetos de audio. M es igual a dos en el escenario preferido para la renderización de estéreo, pero si se lleva a cabo una renderización de canal M, entonces la matriz **A** tiene M líneas.

[0084] Específicamente, un elemento de matriz, a_{ij}, indica si una porción del objeto entero j ha de renderizarse o no en el canal de salida específico i. La porción inferior de la Figura 9 da un ejemplo simple para la matriz de renderización teórica deseada correspondiente a un escenario, en el que hay seis objetos de audio AO1 a AO6 en el que sólo los cinco objetos de audio deben renderizarse en posiciones específicas y en el que el sexto objeto de audio no debe ser renderizado en absoluto.

[0085] En cuanto al objeto de audio, el usuario desea que este objeto de audio sea renderizado en el lado izquierdo de un escenario de reproducción. Por ello, este objeto se coloca en la posición de un parlante izquierdo en un ambiente o sala (virtual) de reproducción, lo que tiene como resultado que la columna de la matriz de renderización **A** sea (10). En cuanto al segundo objeto de audio, a_{22} es uno y a_{12} es 0. lo que significa que el segundo objeto de audio debe renderizarse en el lado derecho.

[0086] El objeto de audio 3 debe renderizarse en el medio, entre el parlante izquierdo y el parlante derecho de manera que el 50 % del nivel o señal de este objeto de audio vaya al canal izquierdo y el 50 % del nivel o señal vaya al canal derecho, de modo que la correspondiente tercera columna de la matriz de renderización teórica deseada **A** sea (0,5 longitud 0,5).

[0087] De manera similar, cualquier colocación entre el parlante izquierdo y el parlante derecho puede indicarse mediante la matriz de renderización teórica deseada. En cuanto al objeto de audio 4, la colocación es más hacia el lado derecho, ya que el elemento de matriz a_{24} es más grande que a_{14}. De manera similar, el quinto objeto de audio A05 se renderiza de manera de orientarse en mayor grado hacia el parlante izquierdo como se indica mediante los elementos de la matriz de renderización teórica deseada a_{15} y a_{25}. La matriz de renderización teórica deseada **A** permite adicionalmente no renderizar en absoluto un determinado objeto de audio. Este se ilustra a título de ejemplo mediante la sexta columna de la matriz de renderización teórica deseada **A** que tiene elementos cero.

[0088] Subsiguientemente se resumen una forma de realización preferida de la presente invención, haciéndose referencia a la Figura 10.

[0089] Es preferible que los procedimientos conocidos del SAOC (Spatial Audio Object Coding, Codificación Espacial de Objetos de Audio) descompongan una señal de audio en diferentes partes. Estas partes pueden ser por ejemplo diferentes objetos de sonido, pero podrían no limitarse a los mismos.

[0090] Si los metadatos se transmiten para cada parte individual de la señal de audio, esto permite ajustar sólo algunos componentes de señal, mientras que otras partes permanecerán sin cambios o podrían aún modificarse con diferentes metadatos.

[0091] Esto podría hacerse para diferentes objetos de sonido, pero también para intervalos espectrales individuales.

[0092] Los parámetros para la separación de objetos son metadatos clásicos o aún nuevos (ganancia, compresión, nivel, …) para cada objeto de audio individual. Es preferible que estos datos sean transmitidos.

[0093] La caja de procesamiento del decodificador se implementa en dos etapas diferentes: en una primera etapa, los parámetros de separación de objetos se utilizan para generar (10) objetos de audio individuales. En la segunda etapa, la unidad de procesamiento 13 tiene múltiples casos, en los que cada caso es para un objeto individual. Aquí,

deberían aplicarse los metadatos específicos para objetos. En el extremo del decodificador, se combinan nuevamente todos los objetos individuales (16) en una única señal de audio. Adicionalmente, un controlador húmedo/segundo 20 puede permitir el sobredesvanecimiento suave entre señal original y manipulada, de manera de conferir al usuario (a la usuaria) final una posibilidad sencilla de encontrar su ajuste preferido.

[0094] En función de la implementación específica, la Figura 10 ilustra dos aspectos. En un aspecto básico, los metadatos relacionados con los objetos se limitan a indicar una descripción de objeto para un objeto específico. Es preferible que la descripción del objeto se refiera a un ID del objeto, como se indica en 21 en la Figura 10. Por ello, los metadatos basados en objeto para el objeto superior manipulado por el dispositivo 13a es solamente la información de que este objeto es un objeto “de habla”. Los metadatos basados en objetos para el otro objeto procesado por el ítem 13b tienen información de que este segundo objeto es un objeto de “surround”.

[0095] Estos metadatos básicos relacionados con los objetos, para ambos objetos, podrían ser suficientes para implementar un modo reforzado de audio limpio, en el que el objeto de habla se amplifica y el objeto de surround se atenúa, o, hablando en términos generales, el objeto de habla se amplifica con respecto al objeto de surround o el objeto de surround se atenúa con respecto al objeto de habla. Sin embargo, el usuario puede preferentemente implementar diferentes modos de procesamiento en el lado de receptor/decodificador, por medio de una entrada de control de modo. Estos diferentes modos pueden ser un modo de nivel de diálogo, un modo de compresión, un modo de mezclado descendente, un modo de medianoche reforzado, un modo reforzado de audio limpio, un modo dinámico de mezclado descendente, un modo para la reubicación de objetos, etc.

[0096] En función de la implementación, los diferentes modos requieren metadatos basados en diferentes objetos además de la información básica que indica el tipo o característica de un objeto tal como de habla o surround. En el modo de medianoche, en el que el margen dinámico de una señal de audio ha de ser comprimido, se prefiere que, para cada objeto tal como objeto de habla y el objeto de surround, se provean como metadatos sea el nivel real sea el nivel teórico deseado para el modo de medianoche. Si se provee el nivel real del objeto, entonces el receptor ha de calcular el nivel objetivo deseado para el modo de medianoche. Sin embargo, cuando está dado el nivel relativo teórico deseado, entonces se reduce el procesamiento del lado de decodificador/receptor.

[0097] En esta implementación, cada objeto tiene una secuencia de información sobre niveles, basada en objetos que varían con el tiempo, que se utilizan por un receptor para comprimir el margen dinámico de manera que se reduzcan las diferentes de nivel dentro de un objeto individual. Esto resulta automáticamente en una señal de audio final, en la que las diferencias de nivel se reducen cada tanto como lo requiere una implementación de modo de medianoche. Para implementaciones de audio limpio, también puede proveerse un nivel teórico deseado para el objeto de habla. En tal caso, el objeto de surround podría ajustarse en cero o casi en cero a efectos de hacer énfasis pesado en el objeto de habla dentro del sonido generado por un determinado establecimiento de altavoces. En una implementación de alta fidelidad, que es lo opuesto al modo de medianoche, el margen dinámico del objeto o el margen dinámico de la diferencia entre los objetos podrían aun reforzarse. En esta implementación, se preferiría, a efectos de proveer niveles de ganancia de objeto teóricos deseados, ya que estos niveles teóricos deseados garantizan que al final se obtiene un sonido creado por un técnico sonidista artístico dentro de la pista de un estudio de sonido y, por ello, tenga la mayor calidad en comparación con un ajuste automático o ajustado por el usuario.

[0098] En otras implementaciones, en las que los metadatos basados en objetos se refieren a mezclados descendentes avanzados, la manipulación de los objetos incluye un mezclado descendente que es diferente que para los establecimientos de renderización específicos. En tal caso, los metadatos basados en objetos se introducen en los bloques de mezclador descendente de objetos 19a a 19c en la Figura 3b o en la Figura 4. En esta implementación, el manipulador puede incluir bloques 19a a 19c, cuando se lleva a cabo un mezclado descendente de objeto individual en función del establecimiento de renderización. Específicamente, los bloques de mezclado descendente de objetos, 19a a 19c, pueden ajustarse de maneras distintas entre si. En este caso, podrían introducirse un objeto de habla en el canal central en lugar de en un canal izquierdo o en un canal derecho, en función de la configuración de los canales. Entonces, los bloques de mezclador descendente, 19a a 19c, podrían tener diferentes cantidades de salidas de señales componentes. También es posible implementar el mezclado descendente dinámicamente.

[0099] Adicionalmente, también es posible proveer información guiada de mezclado ascendente e información para la reubicación de objetos.

[0100] A continuación se da una síntesis de las maneras preferidas de proveer metadatos y la aplicación de metadatos específicos para los objetos.

[0101] Es posible que no puedan separarse los objetos de audio de manera ideal como en una aplicación típica de SOAC. Para la manipulación de audio, puede ser suficiente tener una “máscara” de los objetos, no una separación total.

[0102] Esto podría conducir a una cantidad menor de parámetros para la separación de objetos, o a parámetros menos groseros para dicha separación.

[0103] Para la aplicación denominada “modo de medianoche”, es necesario que el técnico de audio defina todos los parámetros de metadatos independientemente para cada objeto, resultando por ejemplo un volumen constante para el diálogo pero un ruido ambiental manipulado (“modo de medianoche reforzado”).

[0104] Esto también puede ser útil para personas que utilicen audífonos (“enhanced clean audio, audio limpio reforzado”).

[0105] Nuevos escenarios de mezclado descendente: diferentes objetos separados pueden tratarse de manera diferente para cada situación específica de mezclado descendente. Por ejemplo, es necesario mezclar descendentemente una señal de 5.1–canales para un sistema de televisión hogareño estéreo, y otro receptor tiene aún solamente un sistema de reproducción mono. Por ello, diferentes objetos pueden tratarse de diferentes maneras (y todo esto se controla por el técnico sonidista durante la producción gracias a los metadatos provistos por el técnico sonidista).

[0106] También se prefieren mezclados descendentes a 3.0, etc.

[0107] El mezclado descendente no será definido por un parámetro (conjunto de parámetros) global fijo, pero puede generarse a partir de parámetros dependientes de objetos variables en el tiempo.

[0108] Con nuevos metadatos basados en objetos, también es posible proveer un mezclado ascendente guiado.

[0109] Los objetos pueden colocarse en diferentes posiciones, por ejemplo para hacer que la imagen espectral sea más ancha cuando se atenúe el ambiente. Esto ayudará a una inteligibilidad del habla para las personas con discapacidad auditiva.

[0110] El procedimiento propuesto en este documento amplía el concepto existente implementado de los metadatos y utilizado principalmente en los Dolby Codecs. Es ahora posible aplicar el concepto conocido de los metadatos no solamente a la totalidad de la corriente de audio, sino también a los objetos extraídos dentro de esta corriente. Esto confiere a los técnicos y artistas de audio una flexibilidad más amplia, mayores márgenes de ajuste y por ello, un mayor goce para los oyentes.

[0111] Las Figuras 12a, 12b ilustran diferentes escenarios de aplicación del concepto inventivo. En un escenario clásico, existen deportes en televisión, en los que se tiene la atmósfera de estadio deportivo en todos los 5.1 canales, en los que el canal de locutor se mapea en el canal central. Este “mapeo” puede efectuarse mediante una suma directa del canal del locutor a un canal central existente para los 5.1 canales que llevan la atmósfera de estadio deportivo. Ahora bien, el proceso inventivo permite tener este canal central en la descripción del sonido de la atmósfera del estadio deportivo. En tal caso, la operación de la suma mezcla el canal central tomado de la atmósfera del estadio y el locutor. Mediante la generación de parámetros de los objetos para el locutor y el canal central de la atmósfera del estadio deportivo, la presente invención permite separar estos dos objetos de sonido en un lado decodificador, y permite reforzar o atenuar el locutor o el canal central con respecto a la atmósfera del estadio deportivo. Otro escenario es cuando se dispone de dos locutores. Una situación de este tipo puede presentarse cuando dos personas están comentando el mismo juego de soccer. Específicamente, cuando haya dos locutores que están hablando simultáneamente, podría ser útil considerar estos dos locutores como objetos separados y, adicionalmente, tener estos dos locutores separados con respecto a los canales de la atmósfera del estadio deportivo. En una aplicación de este tipo, los canales 5.1 y los canales de los locutores pueden procesarse como ocho objetos de audio diferentes o como siete objetos de audio diferentes, si se desdeña el canal de refuerzo de baja frecuencia (canal de subwoofer). Ya que la infraestructura de distribución directa está adaptada para una señal de sonido de 5.1 canales, es posible mezclar descendentemente los siete (u ocho) objetos en una señal de mezclado descendente de 5.1 canales, y es posible proveer los parámetros de los objetos además de los 5.1 canales de mezclado descendente de manera que, en el lado del receptor, es posible separar nuevamente los objetos, y debido al hecho de que los metadatos basados en objetos identificarán los objetos de locutor con respecto a los objetos de la atmósfera del estadio deportivo, es posible un procesamiento específico para los objetos, antes de que en el lado receptor tenga lugar un mezclado descendente final de 5.1 canales.

[0112] En este escenario, también sería posible tener un primer objeto que comprende el primer locutor, un segundo objeto que comprende el segundo locutor, y un tercer objeto que comprende la atmósfera completa del estadio deportivo.

[0113] A continuación se exponen diferentes implementaciones de escenarios de mezclado descendente basados en objetos, en el contexto de las Figuras 11a a 1c.

[0114] Si, por ejemplo, el sonido generado por el escenario de la Figura 12a o 12b ha de reproducirse en un sistema de reproducción 5.1 convencional, entonces es posible no tener en cuenta la corriente incluida de metadatos, y la corriente recibida puede reproducirse tal cual. Sin embargo, si ha de tener lugar una reproducción sobre establecimientos de estéreo de locutor, debe tener lugar un mezclado descendente de 5.1 a estéreo. Si los canales

de surround se añaden simplemente a izquierda/derecha, los moderadores pueden estar en un nivel que es demasiado pequeño. Por ello se prefiere reducir el nivel de la atmósfera antes o después del mezclado descendente, antes de que se (re)adicione el objeto del moderador.

[0115] Las personas con discapacidad auditiva pueden desear reducir el nivel de atmósfera para tener una mejor inteligibilidad sin dejar de tener ambos locutores separados en izquierdo/derecha, lo que se conoce como el "efecto de cocktail party", donde uno oye su nombre y seguidamente se concentra en la dirección en la que oyó su nombre. Desde un punto de vista sicoacústico, esta concentración específica en cuanto a la dirección, atenuará el sonido procedente de diferentes direcciones. Por ello, una ubicación nítida de un objeto específico tal como el locutor a la izquierda o a la derecha de manera que el locutor parezca estar en el medio entre izquierda y derecha, podría incrementar la inteligibilidad. A tal efecto, es preferible dividir la corriente de audio de entrada en objetos separados, debiendo los objetos tener un tipo de margen en los metadatos que diga que el objeto es importante o menos importante. Seguidamente puede ajustarse la diferencia de nivel entre los mismos de acuerdo con los metadatos, o es posible reubicar la posición del objeto para incrementar la inteligibilidad de acuerdo con los metadatos.

[0116] Para lograr este objetivo, los metadatos se aplican no sobre la señal transmitida, sino que se los aplica a objetos separable individuaoles de audio antes o después del mezclado descendente de objetos, según el caso. Ahora bien, la presente invención ya no requiere que los objetos se limiten a canales espaciales, por lo que estos canales pueden manipularse individualmente. En cambio, el concepto inventivo de los metadatos basados en objetos no requiere tener un objeto específico en un canal específico, pero los objetos pueden ser mezclados descendentemente a varios canales, y todavía pueden manipularse individualmente.

[0117] La Figura 11a ilustra otra forma de realización de una forma de realización preferida. El mezclador descendente 16 genera m canales de salida a partir de de k x n canales de entrada, siendo k la cantidad de objetos, generándose n canales para cada objeto. La Figura 11b corresponde al escenario de la Figura 3a, 3b, donde tiene lugar la manipulación 13a, 13b, 13c antes del mezclado descendente.

[0118] La Figura 11a comprende además los manipuladores de nivel 19d, 19e, 19f, que pueden implementarse sin un control de los metadatos. Sin embargo, como alternativa estos manipuladores de nivel también pueden ser controlados mediante metadatos basados en objetos, por lo que la modificación de nivel implementada por los bloques 19d a 19f también es parte del manipulador de objetos 13 de la Figura 1. Rige lo mismo para las operaciones de mezclado descendente 19a a 19b a 19c, cuando estas operaciones de mezclado descendente son controladas por los metadatos basados en objetos. Sin embargo este caso no se ilustra en la Figura 11a, pero también podría implementarse cuando los metadatos basados en objetos también se envían a los bloques de mezclado descendente 19a a 19c. En este último caso, estos bloques también serian parte del manipulador de objetos 13 de la Figura 11a, y la funcionalidad restante del mezclador de objetos 16 se implementa por la combinación de canales de salida de las señales componentes de objeto manipuladas para los correspondientes canales de salida. La Figura 11a comprende además una funcionalización de normalización de diálogos 25, que puede implementarse con metadatos convencionales, ya que esta normalización del diálogo no tiene lugar en el dominio de los objetos sino en el dominio de los canales de salida.

[0119] La Figura 11 ilustra una implementación de un mezclado descendente 5.1–estéreo basado en objetos. Aquí, el mezclado descendente se lleva a cabo antes de la manipulación y, por ello, la Figura 11b corresponde al escenario de la Figura 4. La modificación de nivel 13a, 13b se lleva a cabo mediante metadatos basados en objetos donde, por ejemplo, el ramal superior corresponde a un objeto de habla y el ramal inferior corresponde a un objeto de surround o, para el ejemplo de las Figuras 12a, 12b, el ramal superior corresponde a uno o ambos locutores y el ramal inferior corresponde a toda la información de surround. Seguidamente, los bloques del manipulador de niveles, 13a, 13b manipularían ambos objetos sobre la base de parámetros establecidos fijos, por lo que los metadatos basados en objetos serian simplemente una identificación de los objetos, pero los manipuladores de nivel 13a, 13b también podrían manipular los niveles sobre la base de niveles teóricos deseados provistos por los metadatos 14 o sobre la base de niveles reales provistos por los metadatos 14. Por ello, para generar un mezclado descendente estéreo multicanal de entrada, se aplica una fórmula de mezclado descendente para cada objeto, y los objetos se ponderan en un nivel dado antes de su remezclado en forma de una señal de salida.

[0120] Para aplicaciones de audio limpias como se ilustra en la Figura 11c, se transmite un nivel de importancia como metadatos a efectos de habilitar una reducción de componentes de señal menos importantes. Entonces, el otro ramal correspondería a los componentes de importancia, que se amplifican, mientras que el ramal inferior podría corresponder a los componentes menos importantes que pueden atenuarse. El cómo la atenuación y/o amplificación específicas de los diferentes objetos se llevan a cabo, puede ajustarse de manera fija por un receptor. También es posible controlarlo, en suma, mediante metadatos basados en objetos como se implementa mediante el control de "seco/húmedo", 14, en la Figura 11c.

[0121] En términos generales, puede efectuarse un control dinámico del margen audible en el dominio de los objetos, lo que se efectúa de manera similar a la implementación de control dinámico del margen audible de AAC como una compresión de múltiple bandas. Los metadatos basados en objetos pueden aún ser un conjunto de datos

selectivos en frecuencia de manera que se lleva a cabo una compresión selectiva en cuanto a frecuencia que es similar a una implementación de ecualizador.

[0122] Como se mencionó en lo que precede, es preferible que se lleve a cabo una normalización del diálogo subsiguientemente al mezclado descendente, es decir, en la señal de mezclado descendente. Por lo general, el mezclado descendente debería ser capaz de procesar k objetos con n canales de ingreso en m canales de salida.

[0123] No es necesariamente importante separar objetos en forma de objetos discretos. Puede ser suficiente "desenmascarar" los componentes de señal que han de ser manipulados. Esto es similar a editar máscaras en el procesamiento de imágenes. En tal caso, un "objeto" generalizado es una superposición de varios objetos originales; esta superposición incluye una cantidad de objetos que es más pequeña que la cantidad total de objetos originales. Todos los objetos se añaden nuevamente en una etapa final. Esto podría carecer de interés en objetos individuales separados, y para algunos objetos, el valor del nivel puede ajustarse en 0, que es una cifra de dB altamente negativa, cuando ha de removerse un determinado objeto por completo, por ejemplo para aplicaciones de karaoke en las que alguien podría estar interesado en remover por completo el objeto vocal de manera que el cantante de karaoke pueda introducir sus propias vocales en los objetos instrumentales restantes.

[0124] Otras formas de realización preferidas de la invención son como se indicó antes: un modo de medianoche reforzado, en la que es posible reducir el margen dinámico de objetos individuales, o un modo de alta fidelidad, en el que se expande el margen dinámico de los objetos. En este contexto, es posible comprimir la señal transmitida, y la idea es invertir esta compresión. La aplicación de una normalización del diálogo tiene preferentemente lugar para la señal total como salida para los locutores, pero se ajusta la normalización del diálogo. Además de los datos paramétricos para separar los objetos de audio diferentes de la señal de mezclado descendente de objetos, se prefiere transmitir, para cada objeto y su señal, además de los metadatos clásicos relacionados con la señal de suma, valores de nivel para el mezclado descendente, valores de importancia indicativos de un nivel de importancia para audio limpio, una identificación de objetos, niveles reales absolutos o relativos como información variable en el tiempo o niveles teóricos deseados absolutos o relativos como información variable en el tiempo, etc.

[0125] Las formas de realización descritas son meramente ilustrativas de los principios de la presente invención. Se da por entendido que las modificaciones y variaciones de las disposiciones y los detalles descritos en la presente serán evidentes para otras personas con pericia en la especialidad. Por ello nuestra intención es que la invención se limite solamente por los alcances de las reivindicaciones adjuntas y no por los detalles específicos presentados a título de descripción y explicación de las formas de realización descritas en la presente.

[0126] En función de determinados requerimientos de los procedimientos de la invención, es posible implementarlos en hardware o en software. La implementación puede efectuarse mediante un medio de almacenamiento digital, en particular, un disco, un DVD o un CD que tengan señales de control electrónicamente legibles almacenadas en ellos, que cooperan con sistemas de computadora programables de manera de llevar a cabo los procedimientos de la invención. En términos generales, la presente invención es por lo tanto un producto programa de computadora con un código de programa almacenado en un portador legible por máquina, operándose el código de programa para llevar a cabo los procedimientos de la invención cuando se ejecutan los productos programa de computadora en una computadora. En otras palabras, los procedimientos de acuerdo con la invención son por lo tanto un programa de computadora que tiene un programa de computadora para llevar a cabo al menos uno de los procedimientos de acuerdo con la invención cuando se ejecuta el programa de computadora en una computadora.

Referencias

[0127]

[1] ISO/IEC 13818–7: MPEG–2 (Generic coding of moving pictures and associated audio information, Codificación genérica para mover imágenes e información de audio asociada) – Part 7: Advanced Audio Coding (AAC)
[2] ISO/IEC 23003–1: MPEG–D (MPEG audio technologies, Tecnologías de audio MPEG) – Part 1: MPEG Surround
[3] ISO/IEC 23003–2: MPEG–D (MPEG audio technologies, Tecnologías de audio MPEG) – Part 2: Spatial Audio Object Coding (SAOC)
[4] ISO/IEC 13818–7: MPEG–2 (Generic coding of moving pictures and associated audio information, Codificación genérica para mover imágenes e información de audio asociada) – Part 7: Advanced Audio Coding (AAC)
[5] ISO/IEC 14496–11: MPEG 4 (Coding of audio–visual objects, Codificación de objetos audio–visuales) – Part 11: Scene Description and Application Engine (BIFS)
[6] ISO/IEC 14496–: MPEG 4 (Coding of audio–visual objects, Codificación de objetos audio–visuales) – Part 20: Lightweight Application Scene Representation (LASER) and Simple Aggregation Format (SAF)
[7] http:/www.dolby.com/assets/pdf/techlibrary/17. AllMetadata.pdf
[8] http:/www.dolby.com/assets/pdf/tech_library/ 18_Metadata.Guide.pdf
[9] Krauss, Kurt; Röden, Jonas; Schildbach, Wolfgang: Transcoding of Dynamic Range Control Coefficients and Other Metadata into MPEG–4 HE AA, Transcodificación de Coeficientes de Control Dinámico del Margen Audible, AES convention 123, October 2007, pp 7217

[10] Robinson, Charles Q., Gundry, Kenneth: Dynamic Range Control via Metadata (Control Dinámico del Margen Audible`) AES Convention 102, September 1999, pp 5028
[11] Dolby, “Standards and Practices for Authoring Dolby Digital and Dolby E Bitstreams (Normas y Prácticas para Autor de Corrientes de Bits de Dolby Digital y Dolby E)”, Issue 3
[14] Coding Technologies/Dolby, “Dolby E / aacPlus Metadata Transcoder Solution for aacPlus Multichannel Digital Video Broadcast (DVB)”, V1.1.0
[15] ETSI TS101154: Digital Video Broadcasting (DVB), (Difusion de Video Digital), V1.8.1
[16] SMPTE RDD 6–2008: Description and Guide to the Use of Dolby E audio Metadata Serial Bitstream (Descripción y Guía para el Uso de Corriente de Bits Seriales de Metadatos de audio Dolby E)

REIVINDICACIONES

1. Aparato para generar por lo menos una señal de audio que representa una superposición de por lo menos dos objetos de audio diferentes, que comprende:

un procesador, para procesar una señal de entrada de audio a efectos de proveer una representación de objetos de la señal de entrada de audio, en el que los por lo menos dos objetos de audio diferentes están separados entre si, los por lo menos dos objetos de audio diferentes están disponibles como señales de objetos de audio separadas, y los al menos dos objetos de audio diferentes son manipulables independientemente entre sí;
un manipulador de objetos, para manipular la señal del objeto de audio o una señal de objeto de audio mejorada de por lo menos un objeto de audio sobre la base de metadatos basados en objetos de audio que se refieren a al menos un objeto de audio para obtener una señal de objeto de audio manipulada o una señal de objetos de audio manipulada mixta para el al menos un objeto de audio; y
un mezclador de objetos, para mezclar la representación de objetos mediante la combinación del objeto de audio manipulado con un objeto de audio diferente manipulado de una manera diferente del al menos un objeto de audio.

2. Aparato de acuerdo con la reivindicación 1, que está adaptado para generar m señales egresadas, siendo m un número entero superior a 1,
en el que el procesador opera de manera de proveer una representación de objetos que tiene k objetos de audio, siendo k un número entero mayor que m,
en el que el manipulador de objetos está adaptado para manipular al menos dos objetos diferentes entre si sobre la base de los metadatos asociados con al menos un objeto de los al menos dos objetos; y
en el que el mezclador de objetos opera de manera de combinar las señales de audio manipuladas de los al menos dos diferentes objetos de manera que cada señal egresada es influida por las señales de audio manipuladas de los al menos dos objetos diferentes.

3. Aparato de acuerdo con la reivindicación 1 en el que
el procesador está adaptado para recibir la señal ingresada, siendo la señal ingresada una representación en mezclado descendente, de una pluralidad de objetos de audio originales,
el procesador está adaptado para recibir parámetros de objetos de audio para controlar un algoritmo de reconstrucción para reconstruir una representación aproximada de los objetos de audio originales, y
el procesador está adaptado para ejecutar el algoritmo de reconstrucción mediante el uso de la señal ingresada y de los parámetros de objetos de audio para obtener la representación de objetos que comprende señales de objeto de audio que son una aproximación de las señales de objeto de audio de los objetos de audio originales.

4. Aparato de acuerdo con la reivindicación 1, en el que
la señal de entrada de audio es una representación en mezclado descendente de una pluralidad de objetos de audio originales y comprende, como información secundaria, metadatos basados en objetos que tienen información acerca de uno o más objetos de audio incluidos en la representación de mezclado descendente, y
el manipulador de objetos está adaptado para extraer los metadatos basados en objetos a partir de la señal de entrada de audio.

5. Aparato de acuerdo con la reivindicación 3, en el que la señal de entrada de audio comprende, como información secundaria, los parámetros de objetos de audio, y porque el procesador está adaptado para extraer la información secundaria de la señal de entrada de audio.

6. Aparato de acuerdo con la reivindicación 1, en el que
el manipulador de objetos opera de manera de manipular la señal del objeto de audio, y
el mezclador de objetos opera de manera de aplicar una regla de mezclado descendente para cada objeto sobre la base de una posición de renderización para el objeto y un establecimiento de reproducción para obtener una señal componente de señal para cada señal de salida de audio, y
el mezclador de objetos está adaptado para añadir señales componentes de objeto a partir de diferentes objetos para el mismo canal de salida de manera de obtener la señal de salida de audio para el canal de salida.

7. Aparato de acuerdo con la reivindicación 1, en el que el manipulador de objetos opera de manera de manipular cada una de entre una pluralidad de señales componentes de objetos de la misma manera sobre la base de metadatos para obtener señales de componentes de objetos para el objeto de audio, y
el mezclador de objetos está adaptado para añadir señales de componentes de objetos tomados de diferentes objetos para el mismo canal de salida de manera de obtener la señal de salida de audio para el canal de salida.

8. Aparato de acuerdo con la reivindicación 1, que además comprende un mezclador de señales de salida para mezclar la señal de salida de audio obtenida sobre la base de una manipulación de por lo menos un objeto de audio y una correspondiente señal de salida de audio obtenida sin la manipulación del al menos un objeto de audio.

9. Aparato de acuerdo con la reivindicación 1, en el que los metadatos comprenden la información acerca de una ganancia, una compresión, un nivel, un establecimiento de mezclado descendente o una característica específica para un objeto determinado, y
el manipulador de objetos está adaptado para manipular el objeto u otros objetos sobre la base de los metadatos para implementar, de una manera específica para el objeto, un modo de medianoche, un modo de alta fidelidad, un modo de audio limpio, la normalización de diálogo, una manipulación específica de mezclado descendente, un mezclado descendente dinámico, un mezclado ascendente guiado, una reubicación de los objetos de habla o una atenuación de un objeto de ambiente,

10. Aparato de acuerdo con la reivindicación 1, en el que los parámetros de los objetos comprenden, para una pluralidad de porciones de tiempo de una señal de audio de objeto, parámetros para cada banda de una pluralidad de bandas de frecuencia en la respectiva porción de tiempo, y
los metadatos incluyen solamente información no selectiva en cuanto a frecuencia, para un objeto de audio.

11. Aparato para generar una señal de audio codificada que representa una superposición de al menos dos objetos de audio diferentes, que comprende:
un formateador de la corriente de datos, para formatear una corriente de datos de manera que la corriente de datos comprende una señal de mezclado descendente de objetos que representa una combinación de al menos dos objetos de audio diferentes, y, como información secundaria, metadatos que se refieren a al menos uno de los objetos de audio diferentes.

12. Aparato de acuerdo con la reivindicación 11, en el que el formateador de corrientes de datos opera de manera de adicionalmente introducir, como información secundaria, datos paramétricos que permiten una aproximación de los al menos dos objetos de audio diferentes, en la corriente de datos.

13. Aparato de acuerdo con la reivindicación 11, que comprende además un calculador de parámetros para calcular datos paramétricos para una aproximación de los al menos dos objetos de audio diferentes, un mezclador descendente para el mezclado descendente de los al menos dos objetos de audio diferentes para obtener la señal de mezclado descendente, y una entrada para metadatos individualmente relacionados con los al menos dos objetos de audio diferentes.

14. Procedimiento para generar al menos una señal de salida de audio que representa una superposición de al menos dos objetos de audio diferentes, que comprende:
procesar una señal de entrada de audio a efectos de proveer una representación de objetos de la señal de entrada de audio, en el que los al menos dos objetos de audio diferentes están separados entre sí, los al menos dos objetos de audio diferentes están disponibles como señales de objeto de audio separadas, y los al menos dos objetos de audio diferentes pueden manipularse independientemente entre sí;
manipular la señal del objeto de audio o una señal de objeto de audio mixta de al menos un objeto de audio basado en metadatos sobre la base de objetos de audio que se refieren a al menos un objeto de audio para obtener una señal de objeto de audio manipulada o una señal de objetos de audio mixta manipulada para el al menos un objeto de audio; y
mezclar la representación de objetos mediante la combinación del objeto de audio manipulado con un objeto de audio no manipulado o con un objeto de audio diferente manipulado que ha sido manipulado de una manera diferente del al menos un objeto de audio.

15. Procedimiento para generar una señal de audio codificada que representa una superposición de al menos dos objetos de audio diferentes, que comprende:
formatear una corriente de datos de manera que la corriente de datos comprenda una señal de mezclado descendente de objetos que representa una combinación de los al menos dos objetos de audio diferentes y, como información secundaria, metadatos que se refieren a al menos uno de los objetos de audio diferentes.

16. Programa de computadora que lleva a cabo, cuando se lo ejecuta en una computadora, un procedimiento para generar al menos una señal de salida de audio de acuerdo con la reivindicación 14 o un procedimiento para generar una señal de audio codificada de acuerdo con la reivindicación 15.

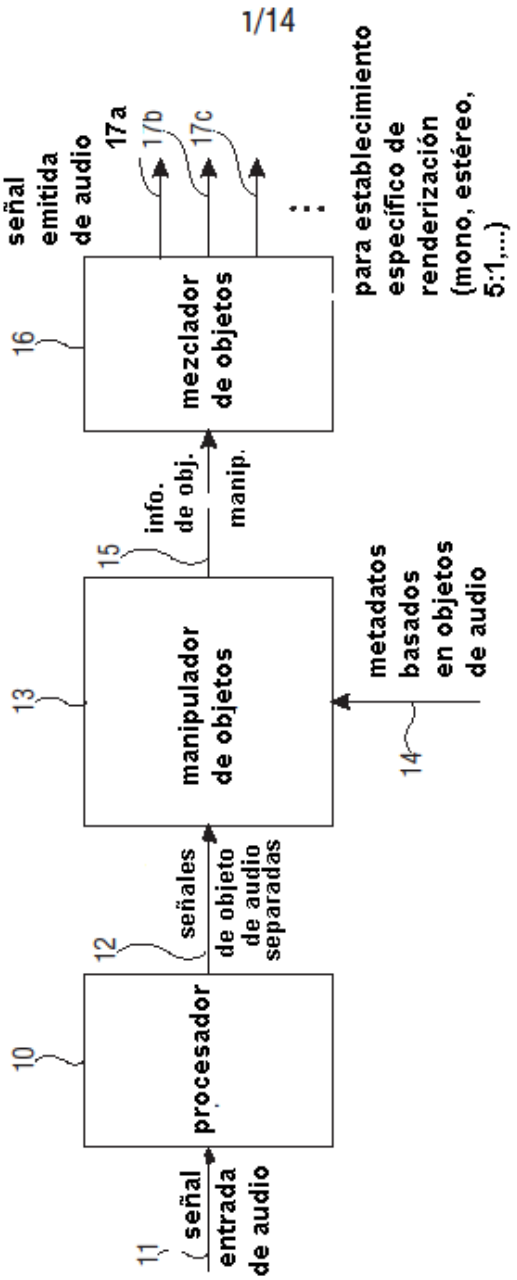

1/14
señal emitida de audio
17a
17b
17c
para establecimiento específico de renderización (mono, estéreo, 5:1,...)
16
mezclador de objetos
info. de obj.
manip.
15
13
manipulador de objetos
metadatos basados en objetos de audio
14
12
señales de objeto de audio separadas
10
procesador
11
señal entrada de audio

FIGURA 1

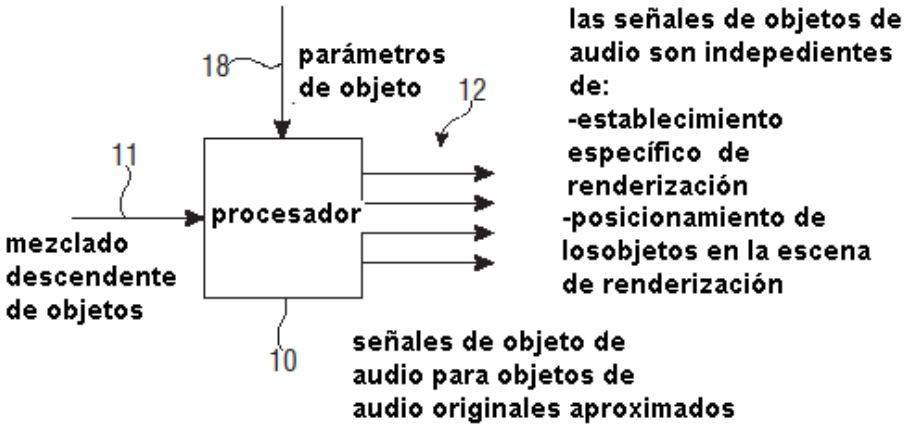


FIGURA 2

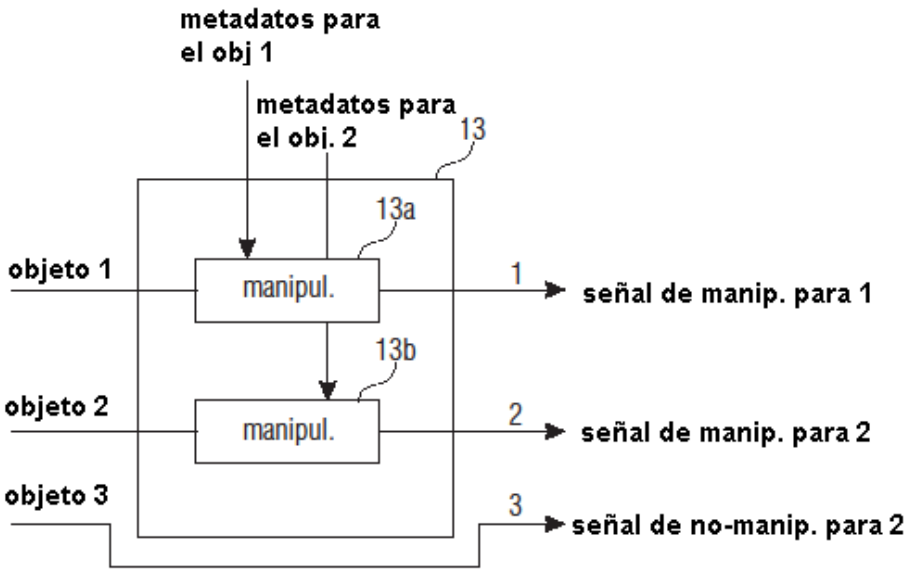


FIGURA 3A

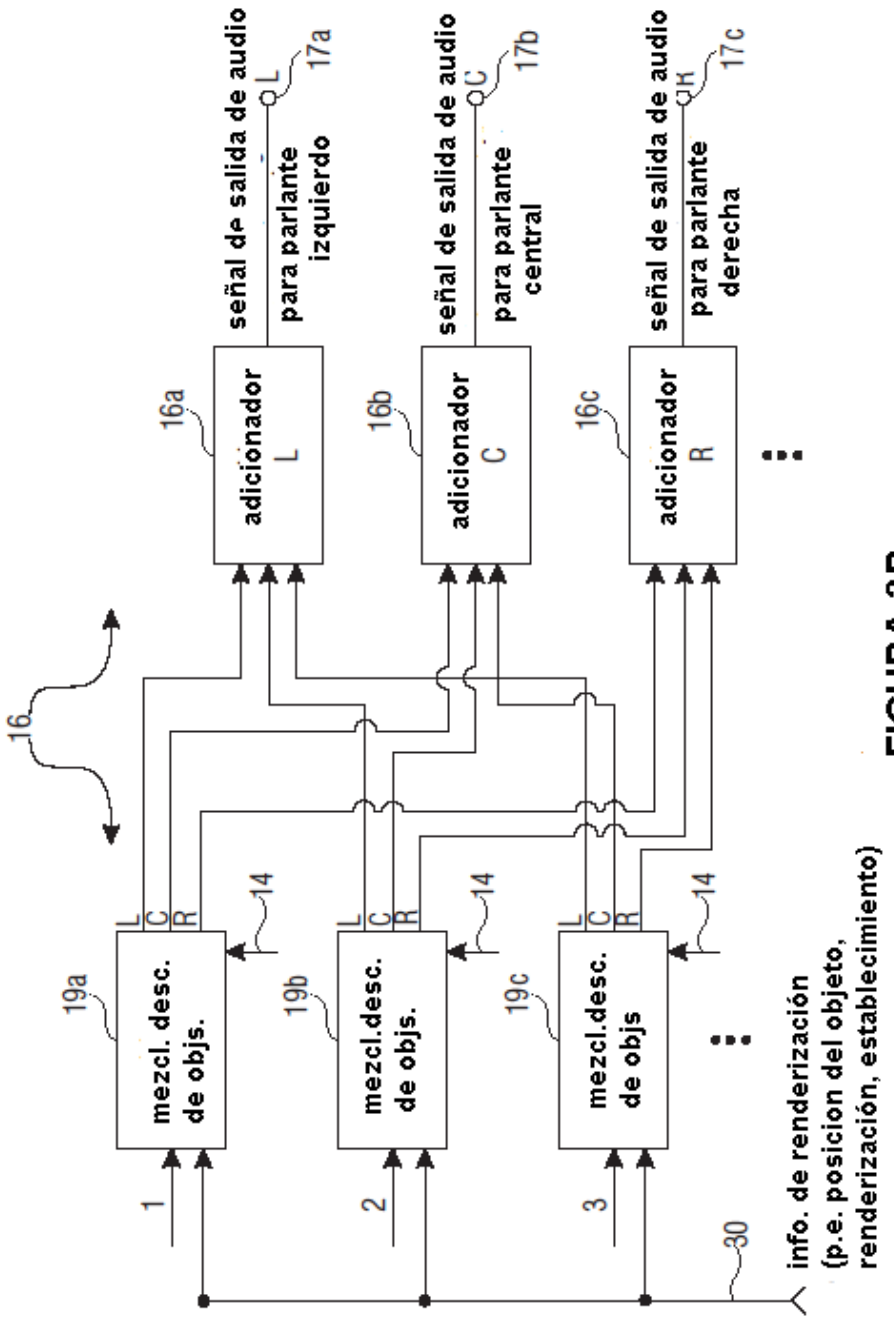

señal de salida de audio
para parlante izquierdo
17a
señal de salida de audio
para parlante central
17b
señal de salida de audio
para parlante derecha
17c
adicionador L
adicionador C
adicionador R
16a
16b
16c
16
mezcl. desc. de objs.
mezcl.desc. de objs.
mezcl.desc. de objs
19a
19b
19c
14
L
C
R
1
2
3
30
info. de renderización
(p.e. posicion del objeto,
renderización, establecimiento)

FIGURA 3B

4/14

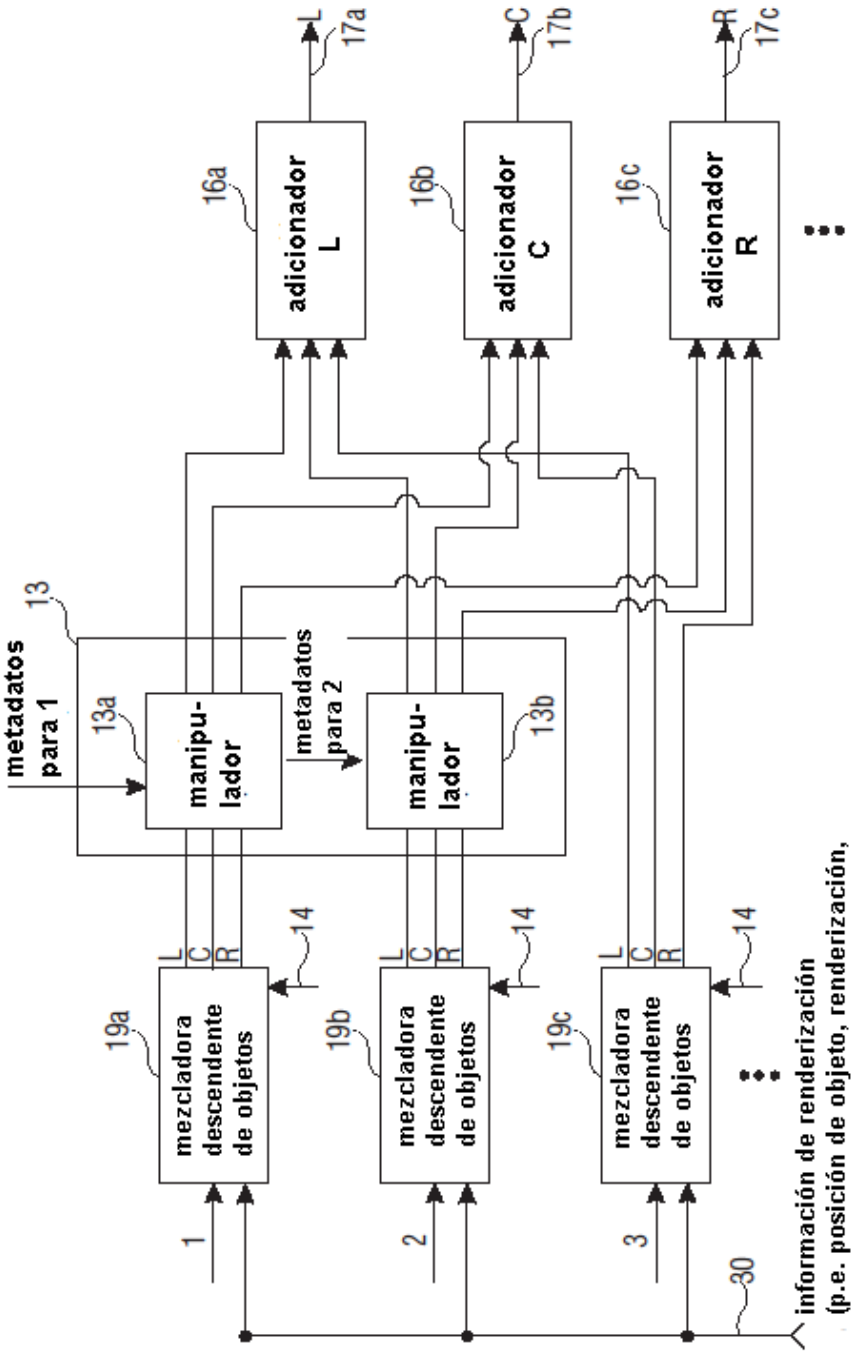

metadatos para 1
13
13a
manipu-lador
metadatos para 2
manipu-lador
13b
16a
adicionador L
16b
adicionador C
16c
adicionador R
17a
17b
17c
L
C
R
14
19a
mezcladora descendente de objetos
19b
mezcladora descendente de objetos
19c
mezcladora descendente de objetos
1
2
3
30
información de renderización (p.e. posición de objeto, renderización, establecimiento,...)

FIGURA 4

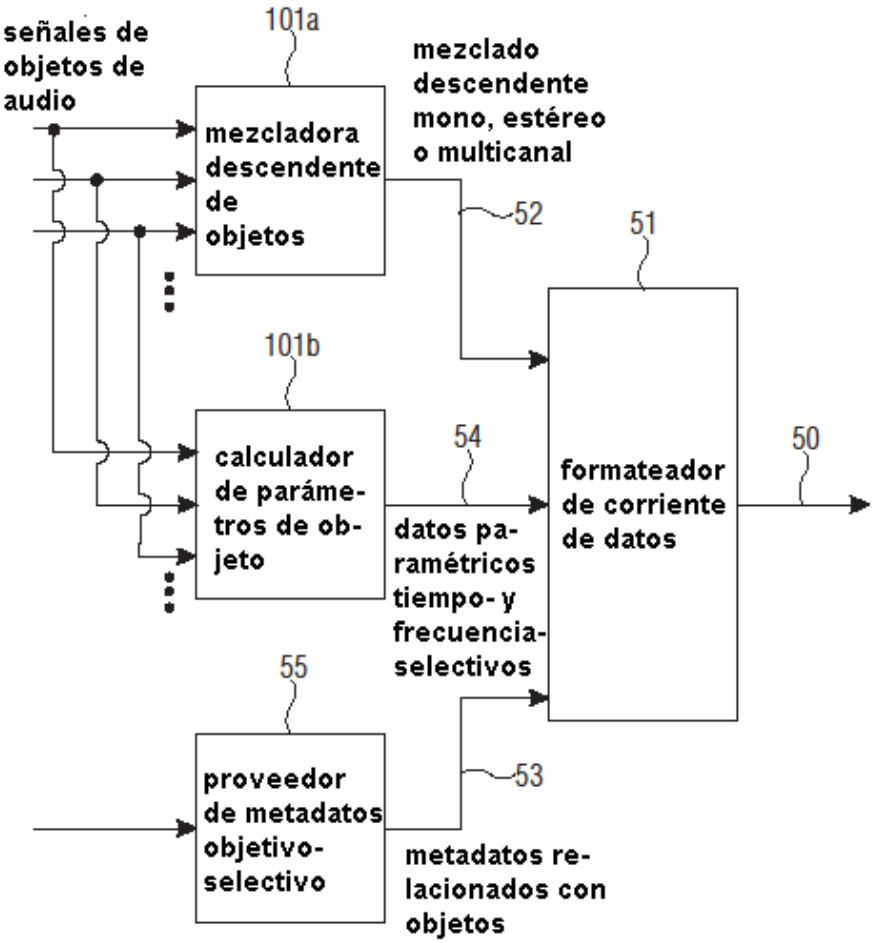


FIGURA 5A

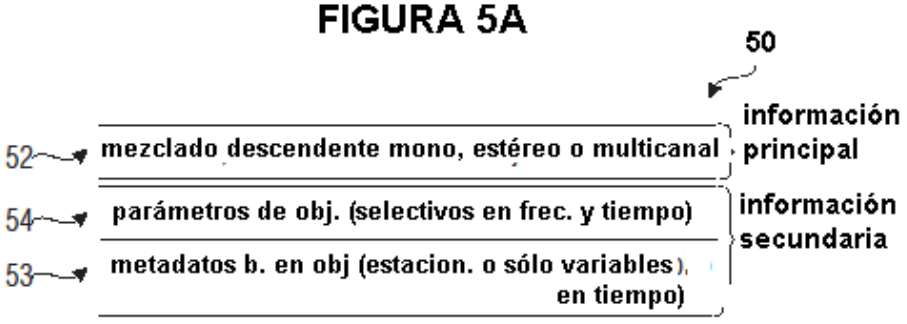


FIGURA 5B

6/14

ID del objeto	archivo del objeto de audio	energía de los parámetros de obj de audio, y correlación
1		matriz E para cada subbanda y para cada bloque de tiempo
2		
3		
...	...	
N		

FIGURA 6

matriz de covarianza de objetos para un bloque de tiempo y una subbanda

$$E = \begin{bmatrix} e_{11} & & & & & \\ e_{21} & e_{22} & & & & \\ e_{31} & e_{32} & e_{33} & & & \\ e_{41} & e_{42} & e_{43} & e_{44} & & \\ \vdots & & & & e_{55} & \\ & & & & & e_{66} \\ & & & & & \ddots \\ & & & & & e_{nn} \end{bmatrix}$$

e_{ij}: potencia de las señales de audio i en la correspondiente subbanda y bloque de tiempo

e_{ij}: medida de correlación entre objetos de audio i,j en la correspondiente subbanda y bloque de tiempo

FIGURA 7

7/14

matriz de mezclado descendente

$$D = \begin{bmatrix} d_{11} & d_{12} & d_{13} & d_{14} \cdots d_{1N} \\ d_{21} & d_{22} & d_{23} & d_{24} \cdots d_{2N} \\ \vdots & & & \\ d_{K1} & & & \cdots d_{KN} \end{bmatrix}$$

dij: indica si una porción del objeto j entero j está incluída o no en la señal de mezclado descendente del objeto

por ejemplo:

$d_{12} = 0 \Rightarrow$ **el objeto 2 NO está incluido en la señal 1 de mezclado descendente de objeto**

$d_{23} = 1 \Rightarrow$ **el objeto 3 está COMPLETAMENTE incluido en la señal de mezclado descendente de objeto**

$d_{24} = d_{14} = 0{,}5 \Rightarrow$ **el objeto 4 se halla en ambas señales de mezclado descendente de obj, pero con la 1/2 de energía en cada señal de mezc.desc.de obj.**

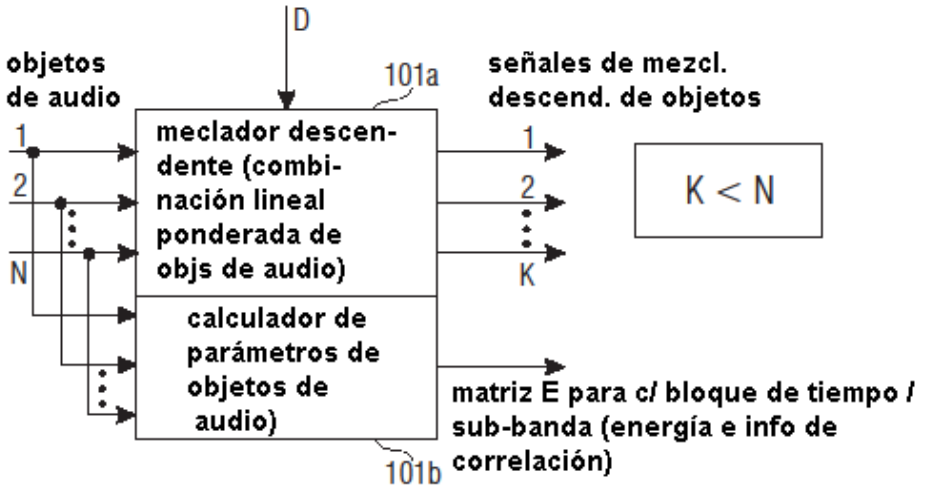


FIGURA 8

8/14

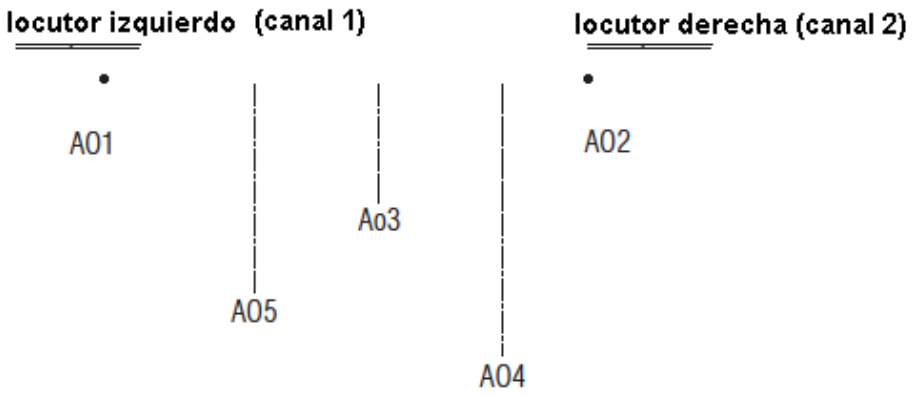
matriz A de renderización teórica deseada **normalmente provista por el usuario**

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & \cdots & a_{1N} \\ a_{21} & a_{22} & a_{23} & a_{24} & \cdots & a_{2N} \\ \vdots & & & & & \vdots \\ a_{M1} & \cdots & & & & a_{MN} \end{bmatrix}$$

M = 2 para renderización estéreo

M = M para renderización de canal M

aij indica si una porción del objeto entero debe renderizarse o no en el canal de salida



matriz ejemplo A = $\begin{bmatrix} 1 & 0 & 0.5 & 0.25 & 0.75 & 0 \\ 0 & 1 & 0.5 & 0.75 & 0.25 & 0 \end{bmatrix}$

(el objeto 6 NO debe renderizarse en absoluto)

Figura 9

9/14

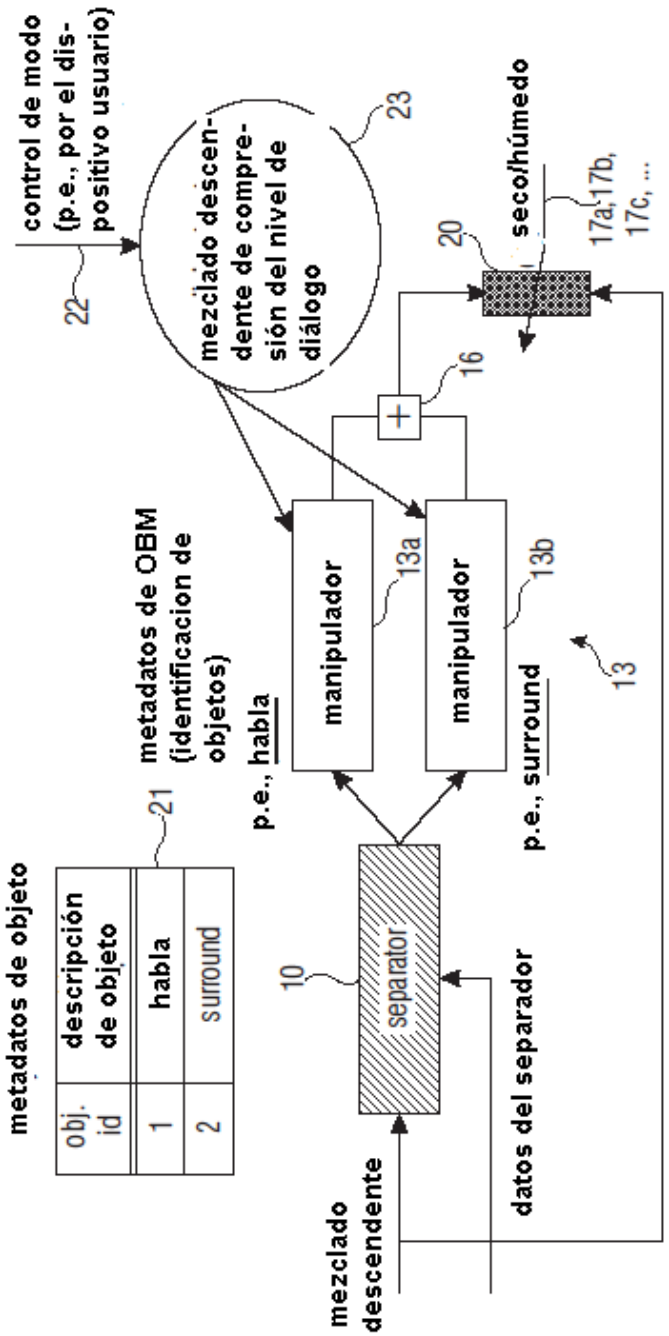


FIGURA 10

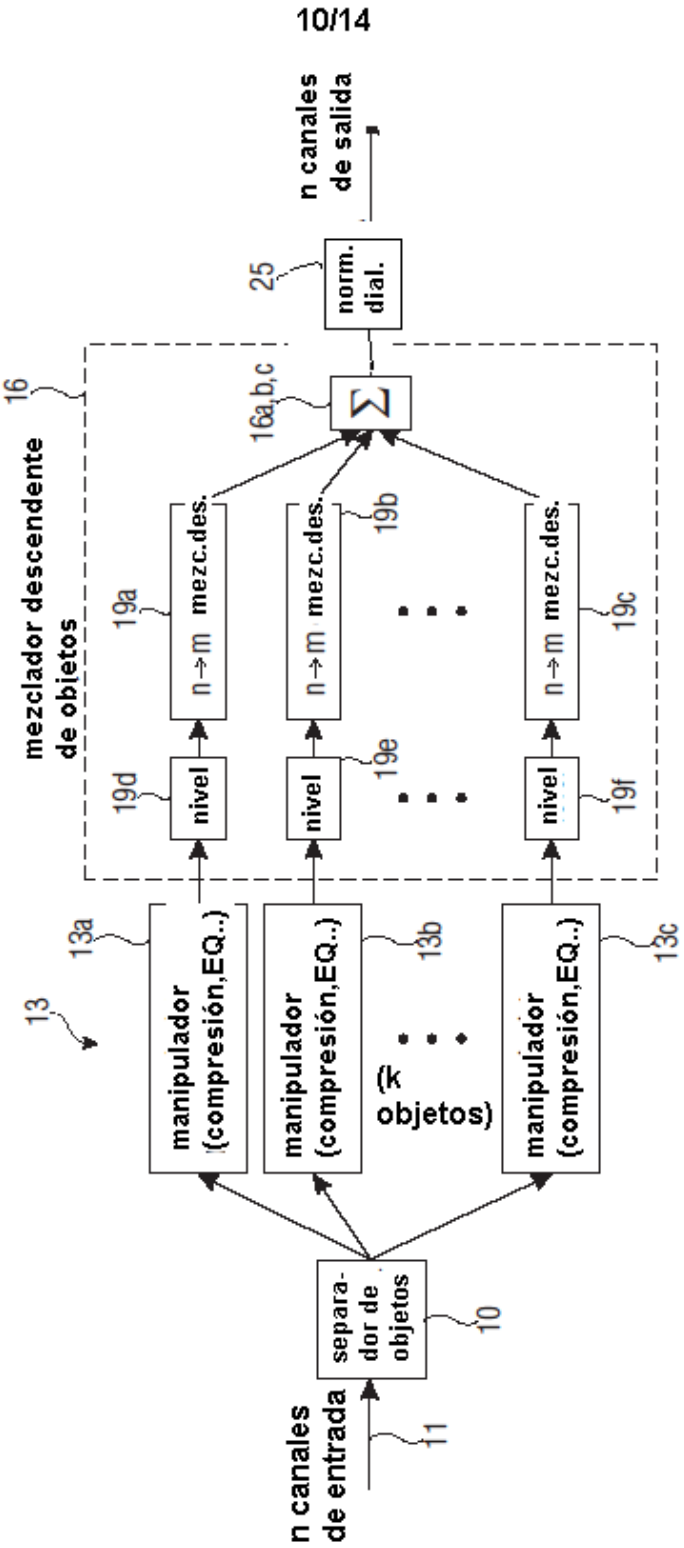

10/14
n canales de entrada
11
separa-dor de objetos
10
13
13a
manipulador (compresión,EQ..)
13b
manipulador (compresión,EQ..)
(k objetos)
13c
manipulador (compresión,EQ..)
mezclador descendente de objetos
16
19d
nivel
19a
n→m mezc.des.
19e
nivel
19b
n→m mezc.des.
19f
nivel
19c
n→m mezc.des.
16a,b,c
Σ
25
norm. dial.
n canales de salida

FIGURA 11A

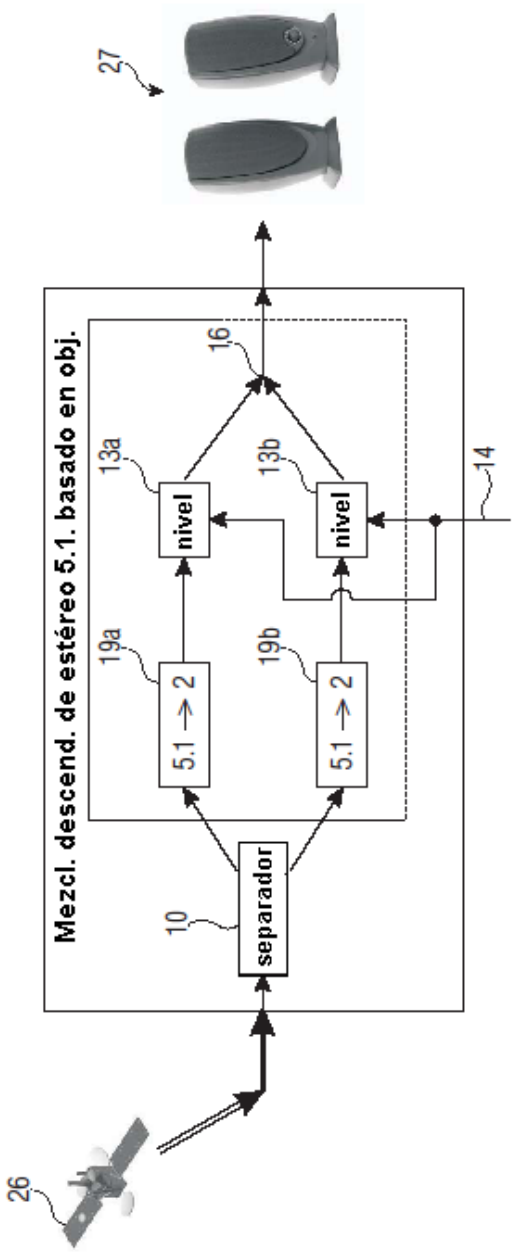

27
Mezcl. descend. de estéreo 5.1. basado en obj.
16
13a
13b
14
nivel
nivel
19a
19b
5.1 → 2
5.1 → 2
10
separador
26

FIGURA 11B

12/14

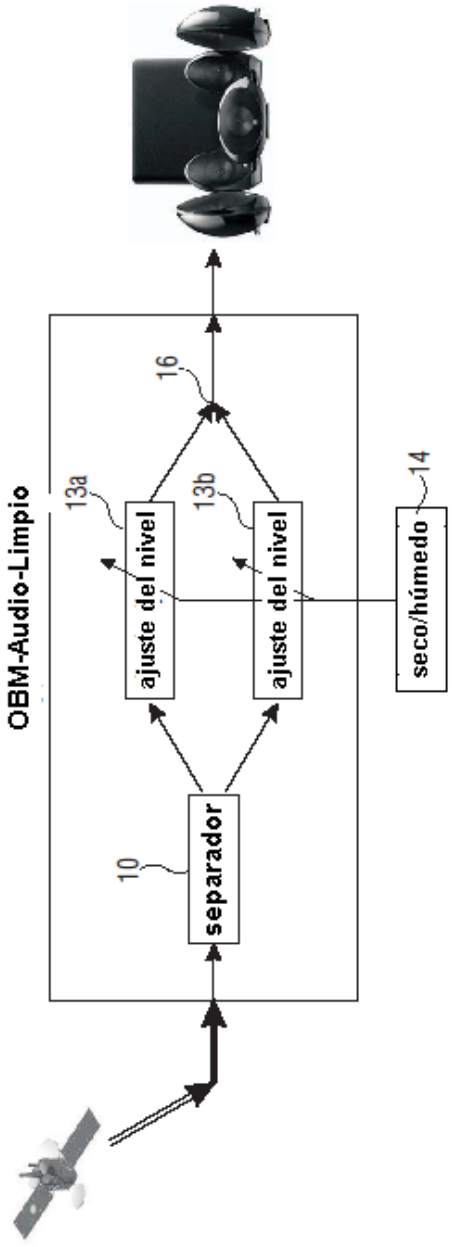


FIGURA 11C

13/14

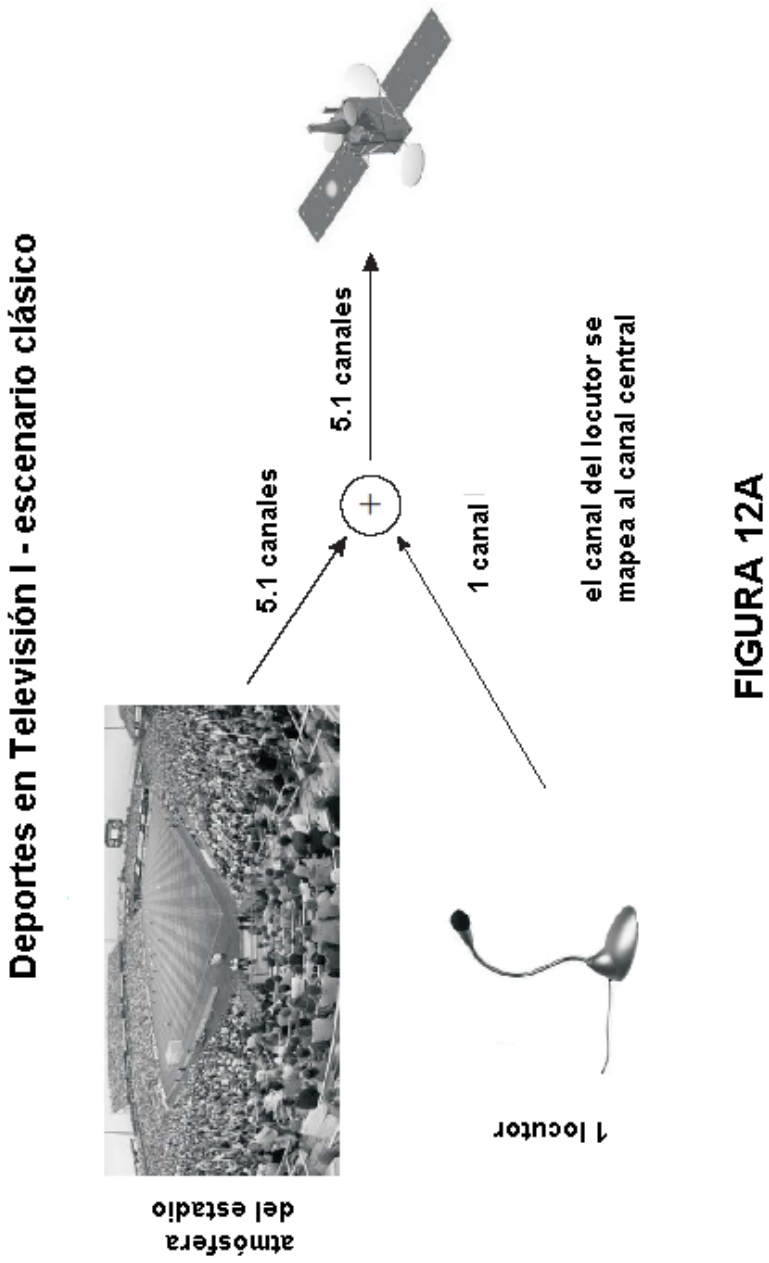


FIGURA 12A

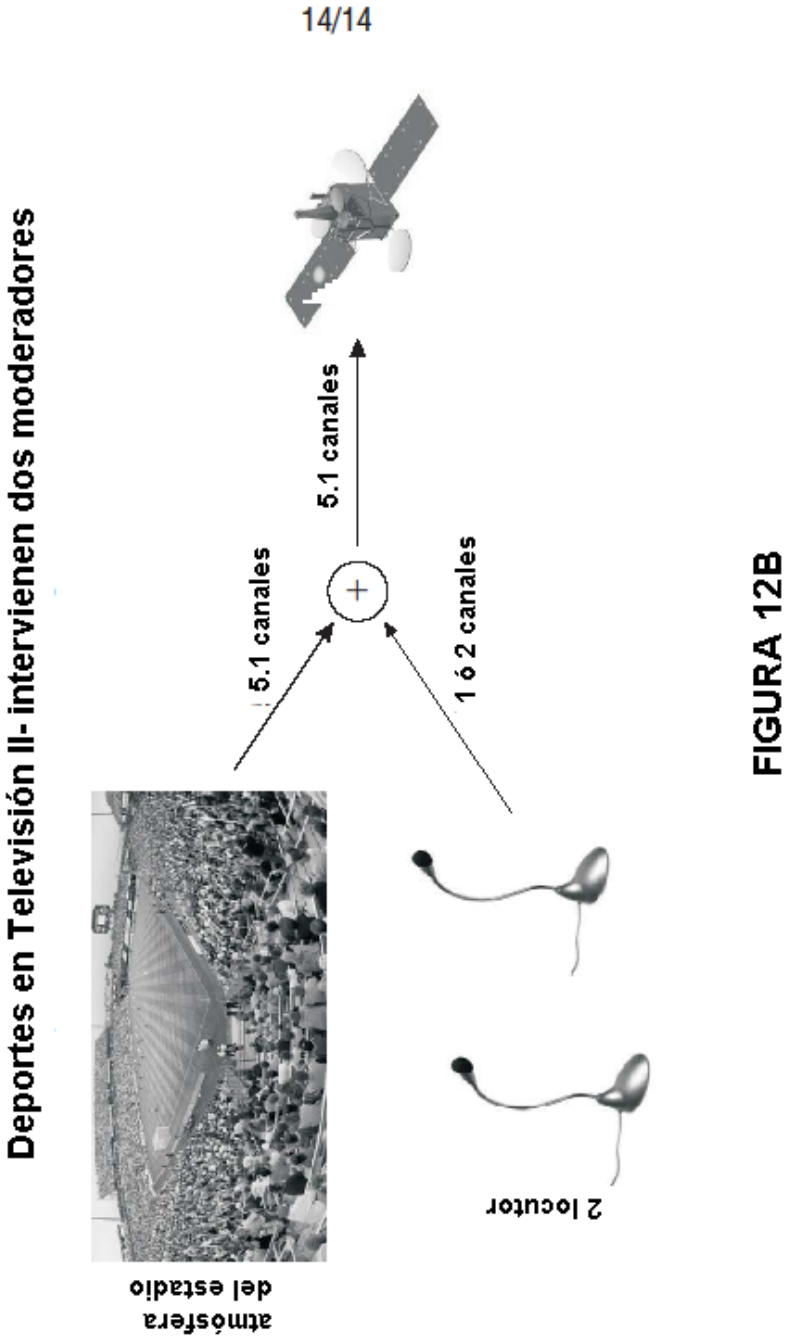

14/14
Deportes en Televisión II- intervienen dos moderadores
5.1 canales
5.1 canales
1 ó 2 canales
atmósfera del estadio
2 locutor

FIGURA 12B