

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 629 527**

51 Int. Cl.:

G10L 19/008 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **27.11.2014 PCT/EP2014/075801**

87 Fecha y número de publicación internacional: **04.06.2015 WO15078964**

96 Fecha de presentación y número de la solicitud europea: **27.11.2014 E 14802914 (3)**

97 Fecha y número de publicación de la concesión europea: **29.03.2017 EP 2941771**

54 Título: **Decodificador, codificador y procedimiento para la estimación de sonoridad informada empleando señales de objeto de audio anuladas en sistemas de codificación de audio basados en objetos**

30 Prioridad:

27.11.2013 EP 13194664

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

10.08.2017

73 Titular/es:

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

72 Inventor/es:

**PAULUS, JOUNI;
DISCH, SASCHA;
FUCHS, HARALD;
GRILL, BERNHARD;
HELLMUTH, OLIVER;
MURTAZA, ADRIAN;
RIDDERBUSCH, FALKO y
TERENTIV, LEON**

74 Agente/Representante:

SALVA FERRER, Joan

ES 2 629 527 T3

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

DESCRIPCIÓN

Decodificador, codificador y procedimiento para la estimación de sonoridad informada empleando señales de objeto de audio anuladas en sistemas de codificación de audio basados en objetos

5

[0001] La presente invención se refiere a la codificación, procesamiento y decodificación de señales de audio, y, en particular, a un decodificador, un codificador y un procedimiento para la estimación de sonoridad informada en sistemas de codificación de audio basados en objetos.

10 **[0002]** Recientemente, en el campo de la codificación de audio [BCC, JSC, SAOC, SAOC1, SAOC2] y la separación de fuentes informadas [ISS1, ISS2, ISS3, ISS4, ISS5, ISS6] se han propuesto técnicas paramétricas para la transmisión/almacenamiento eficaz de la velocidad binaria de escenas de audio que comprenden múltiples señales de objetos de audio. Estas técnicas apuntan a reconstruir una escena de audio de salida deseada o un objeto fuente de audio en base a información lateral adicional que describe la escena de audio
15 transmitida/almacenada y/o los objetos fuente en la escena de audio. Esta reconstrucción tiene lugar en el decodificador utilizando un esquema informado de separación de fuentes. Los objetos reconstruidos se pueden combinar para producir la escena de audio de salida. En función de la forma como se combinen los objetos, la sonoridad perceptual de la escena de salida puede variar.

20 **[0003]** En la emisión de televisión y radio, los niveles de volumen de las pistas de audio de varios programas pueden normalizarse en base a diversos aspectos, tales como el nivel de señal pico o el nivel de sonoridad. En función de las propiedades dinámicas de las señales, dos señales con el mismo nivel pico pueden tener un nivel muy diferente de sonoridad percibida. Actualmente al cambiar entre programas o canales las diferencias en la sonoridad de la señal son muy molestas y han sido una fuente importante de quejas de los usuarios finales con respecto a la
25 emisión.

[0004] En la técnica anterior, se ha propuesto normalizar todos los programas de todos los canales de manera similar a un nivel de referencia común utilizando una medida basada en la sonoridad de la señal perceptual. Una recomendación de este tipo en Europa es la Recomendación R128 de la UER [EBU] (más adelante denominada
30 R128).

[0005] La recomendación dice que la "sonoridad del programa", por ejemplo, la sonoridad media de un programa (o un anuncio u otro programa significativo) debe ser igual a un nivel especificado (con pequeñas desviaciones permitidas). Cuantos más radiodifusores cumplan con esta recomendación y la normalización
35 necesarias, las diferencias en la sonoridad media entre programas y canales deberían minimizarse.

[0006] La estimación de sonoridad se puede realizar de varias maneras. Existen diferentes modelos matemáticos para estimar la sonoridad perceptual de una señal de audio. La recomendación R128 de la UER se basa en el modelo presentado en la UIT-R BS.1770 (más adelante denominado BS.1770) (véase [UIT]) para la
40 estimación de sonoridad.

[0007] Como se ha indicado anteriormente, por ejemplo, según la recomendación R128 de la UER, la sonoridad del programa, por ejemplo, la sonoridad media de un programa debe ser igual a un nivel especificado con pequeñas desviaciones permitidas. Sin embargo, esto supone problemas significativos cuando se lleva a cabo la
45 renderización de audio, y que hasta ahora la técnica anterior no había solucionado. Llevar a cabo una renderización de audio en el lado del decodificador tiene un efecto significativo en la sonoridad total/general de la señal de entrada de audio recibida. Sin embargo, a pesar de que se lleva a cabo la renderización de escenas, la sonoridad total de la señal de audio recibida permanecerá igual.

50 **[0008]** Actualmente, no existe una solución específica en el lado del decodificador para este problema.

[0009] El documento EP 2 146 522 A1 ([EP]) se refiere a conceptos para generar señales de salida de audio utilizando metadatos basados en objetos. Se genera al menos una señal de salida de audio que representa una superposición de, al menos dos, señales de objetos de audio diferentes, pero no proporciona una solución para este
55 problema.

[0010] El documento WO 2008/035275 A2 ([BRE]) describe un sistema de audio que comprende un codificador que codifica objetos de audio en una unidad de codificación que genera una señal de audio de mezcla descendente y datos paramétricos que representan la pluralidad de objetos de audio. La señal de audio de mezcla

descendente y los datos paramétricos se transmiten a un decodificador que comprende una unidad de decodificación que genera réplicas aproximadas de los objetos de audio y una unidad de renderización que genera una señal de salida a partir de los objetos de audio. El decodificador contiene además un procesador para generar datos de modificación de la codificación que se envían al codificador. A continuación el codificador modifica la

5 codificación de los objetos de audio, y en particular modifica los datos paramétricos, en respuesta a los datos de modificación de la codificación. La estrategia permite que la manipulación de los objetos de audio sea controlada con el decodificador pero realizada totalmente o en parte con el codificador. Por lo tanto, la manipulación puede realizarse en los objetos de audio independientes reales en lugar de en las réplicas aproximadas proporcionando así un rendimiento mejorado.

10

[0011] El documento EP 2 146 522 A 1 describe un aparato para generar al menos una señal de salida de audio que representa una superposición de al menos dos objetos de audio diferentes comprende un procesador para procesar una señal de entrada de audio que proporciona una representación de objetos de la señal de entrada de audio, donde esta renderización de objetos puede ser generada por una aproximación guiada paramétricamente de

15 los objetos originales usando una señal de mezcla descendente de objetos. Un manipulador de objetos manipula individualmente objetos utilizando metadatos basados en objetos de audio que se refieren a los objetos de audio individuales para obtener objetos de audio manipulados. Los objetos de audio manipulados se mezclan utilizando un mezclador de objetos que obtiene en último término una señal de salida de audio, que tiene una o varias señales de canales que dependen de una configuración de renderización específica.

20

[0012] El documento 2008/046531 A1 ([ENG]) describe un codificador de objetos de audio que genera una señal de objetos codificados utilizando una pluralidad de objetos de audio que incluye un generador de información de mezcla descendente para generar una información de mezcla descendente que indica una distribución de la pluralidad de objetos de audio en, al menos dos, canales de mezcla descendente, un generador de parámetros de

25 objetos de audio que genera parámetros de objetos para los objetos de audio y una interfaz de salida que genera la señal de salida de audio importada utilizando la información de mezcla descendente y los parámetros de los objetos. Un sintetizador de audio utiliza la información de mezcla descendente para generar unos datos de salida utilizables para crear una pluralidad de canales de salida de la configuración de salida de audio predefinida.

30

[0013] Además, el documento WO2012/125855 se refiere a la creación, codificación, transmisión, decodificación y reproducción de bandas sonoras espaciales de audio. De este modo, el formato de codificación es compatible con los formatos de codificación de sonido envolvente heredados.

35

[0014] Sería deseable tener una estimación precisa de la sonoridad media de salida o del cambio en la sonoridad media sin retardo y, cuando el programa no cambia o la escena de renderización no cambia, la estimación de sonoridad media también debería permanecer estática.

40

[0015] El objetivo de la presente invención es proporcionar unos conceptos mejorados para la codificación, procesamiento y decodificación de señales de audio. El objetivo de la presente invención se resuelve mediante un decodificador según la reivindicación 1, mediante un codificador según la reivindicación 9, mediante un sistema según la reivindicación 11, mediante un procedimiento según la reivindicación 12, mediante un procedimiento según la reivindicación 13 y mediante un programa informático según la reivindicación 15.

45

[0016] Se proporciona una manera informada para estimar la sonoridad de la salida en un sistema de codificación de audio basado en objetos. Los conceptos proporcionados se basan en la información sobre la sonoridad de los objetos en la mezcla de audio que se proporcionará al decodificador. El decodificador utiliza esta información junto con la información de renderización para estimar la sonoridad de la señal de salida. Esto permite posteriormente, por ejemplo, estimar la diferencia de sonoridad entre la mezcla descendente por defecto y la salida renderizada. Entonces es posible compensar la diferencia para obtener una sonoridad aproximadamente constante

50 en la salida independientemente de la información de renderización. La estimación de sonoridad en el decodificador tiene lugar de una manera completamente paramétrica, y es computacionalmente muy ligera y preciso en comparación con los conceptos de estimación de sonoridad basados en la señal.

55

[0017] Se proporcionan conceptos para obtener información sobre la sonoridad de la escena de salida específica utilizando conceptos puramente paramétricos, lo que permite el procesamiento de sonoridad sin realizar una estimación de sonoridad explícita basada en la señal en el decodificador. Además, se describe la tecnología específica de la Codificación de Objetos de Audio Espacial (SAOC) estandarizada por MPEG [SAOC], pero los conceptos proporcionados pueden ser utilizados junto con otras tecnologías de codificación de objetos de audio.

[0018] Se proporciona un decodificador para generar una señal de salida de audio que comprende uno o más canales de salida de audio. El decodificador comprende una interfaz de recepción que recibe una señal de entrada de audio que comprende una pluralidad de señales de objetos de audio, para recibir información de sonoridad sobre las señales de objetos de audio y para recibir información de renderización que indica si una o más de las señales de objetos de audio se amplificarán o atenuarán. Además, el decodificador comprende un procesador de señal para generar uno o más canales de salida de audio de la señal de salida de audio. El procesador de señal está configurado para determinar un valor de compensación de sonoridad en función de la información de sonoridad y en función de la información de renderización. Además, el procesador de señal está configurado para generar uno o más canales de salida de audio de la señal de salida de audio a partir de la señal de entrada de audio en función de la información de renderización y en función del valor de compensación de sonoridad.

[0019] Según una realización, el procesador de señal puede estar configurado para generar uno o más canales de salida de audio de la señal de salida de audio a partir de la señal de entrada de audio en función de la información de renderización y en función del valor de compensación de sonoridad, de manera que una sonoridad de la señal de salida de audio es igual a una sonoridad de la señal de entrada de audio, o de manera que la sonoridad de la señal de salida de audio está más próxima a la sonoridad de la señal de entrada de audio que una sonoridad de una señal de audio modificada que resultaría de la modificación de la señal de entrada de audio amplificando o atenuando las señales de objetos de audio de la señal de entrada de audio según la información de renderización.

[0020] Según otra realización, cada una de las señales de objetos de audio de la señal de entrada de audio se puede asignar exactamente a un grupo de dos o más grupos, en la que cada uno de los dos o más grupos puede comprender una o más de las señales de objetos de audio de la señal de entrada de audio. En tal realización, la interfaz de recepción puede estar configurada para recibir un valor de sonoridad para cada grupo de los dos o más grupos como la información de sonoridad, en la que dicho valor de sonoridad indica una sonoridad original total de una o más señales de objetos de audio de dicho grupo. Además, la interfaz de recepción puede estar configurada para recibir la información de renderización que indica, para al menos un grupo de los dos o más grupos, si la o las señales de objetos de audio de dicho grupo se amplificarán o atenuarán indicando una sonoridad modificada total de la o las señales de objetos de audio de dicho grupo. Además, en una realización de este tipo, el procesador de señales puede estar configurado para determinar el valor de compensación de sonoridad en función de la sonoridad modificada total de cada uno de los dichos, al menos uno, grupos de los dos o más grupos y, en función de la sonoridad original total de cada uno de los dos o más grupos. Además, el procesador de señal puede estar configurado para generar uno o más canales de salida de audio de la señal de salida de audio a partir de la señal de entrada de audio en función de la sonoridad modificada total de cada uno de los dichos, al menos uno, grupos de los dos o más grupos y, en función del valor de compensación de sonoridad.

[0021] En realizaciones particulares, al menos un grupo de los dos o más grupos puede comprender dos o más de las señales de objetos de audio.

[0022] Además, se proporciona un codificador. El codificador comprende una unidad codificadora basada en objetos que codifica una pluralidad de señales de objetos de audio para obtener una señal de audio codificada que comprende la pluralidad de señales de objetos de audio. Además, el codificador comprende una unidad codificadora de sonoridad de objetos para codificar información de sonoridad sobre las señales de objetos de audio. La información de sonoridad comprende uno o más valores de sonoridad, en el que cada uno, o más, de los valores de sonoridad depende de una, o más, señales de objetos de audio.

[0023] Según una realización, cada una de las señales de objetos de audio de la señal de audio codificada puede ser asignada exactamente a un grupo de dos o más grupos, en el que cada uno de los dos o más grupos comprende una, o más, señales de objetos de audio de la señal de audio codificada. La unidad codificadora de sonoridad de objetos puede configurarse para determinar uno o más valores de sonoridad de la información de sonoridad mediante la determinación de un valor de sonoridad para cada grupo de los dos o más grupos, en el que dicho valor de sonoridad de dicho grupo indica una sonoridad original total de la o las señales de objetos de audio de dicho grupo.

[0024] Además, se proporciona un sistema. El sistema comprende un codificador según una de las realizaciones descritas anteriormente que codifica una pluralidad de señales de objetos de audio para obtener una señal de audio codificada que comprende la pluralidad de señales de objetos de audio y que codifica información de sonoridad sobre las señales de objetos de audio. Además, el sistema comprende un decodificador según una de las realizaciones descritas anteriormente para generar una señal de salida de audio que comprende uno o más canales

de salida de audio. El decodificador está configurado para recibir la señal de audio codificada como una señal de entrada de audio y la información de sonoridad. Además, el decodificador está configurado para recibir más información de renderización. Por otro lado, el decodificador está configurado para determinar un valor de compensación de sonoridad en función de la información de sonoridad y en función de la información de renderización. Además, el decodificador está configurado para generar uno o más canales de salida de audio de la señal de salida de audio a partir de la señal de entrada de audio en función de la información de renderización y en función del valor de compensación de sonoridad.

[0025] Además, se proporciona un procedimiento para generar una señal de salida de audio que comprende uno o más canales de salida de audio. El procedimiento comprende:

- Recepción de una señal de entrada de audio que comprende una pluralidad de señales de objetos de audio.
- Recepción de la información de sonoridad en las señales de objetos de audio.
- Recepción de información de renderización que indica si una o más de las señales de objetos de audio serán amplificadas o atenuadas.
- Determinación de un valor de compensación de sonoridad en función de la información de sonoridad y en función de la información de renderización. Y:
- Generación de uno o más canales de salida de audio de la señal de salida de audio a partir de la señal de entrada de audio en función de la información de renderización y en función del valor de compensación de sonoridad.

[0026] Por otro lado, se proporciona un procedimiento para la codificación. El procedimiento comprende:

- Codificación de una señal de entrada de audio que comprende una pluralidad de señales de objetos de audio. Y:
- Codificación de información de sonoridad en las señales de objetos de audio, en la que la información de sonoridad comprende uno o más valores de sonoridad, en la que cada uno, o más, de los valores de sonoridad depende de una, o más, de las señales de objetos de audio.

[0027] Además, se proporciona un programa informático que implementa el procedimiento anteriormente descrito cuando se ejecuta en un ordenador o procesador de señales.

[0028] Las realizaciones preferentes se definen en las reivindicaciones dependientes.

[0029] En los párrafos siguientes se describen con mayor detalle realizaciones de la presente invención en relación con las figuras en las que:

La Fig. 1 ilustra un decodificador para generar una señal de salida de audio que comprende uno o más canales de salida de audio según una realización,

La Fig. 2 ilustra un sistema según una realización,

La Fig. 3 ilustra un codificador según una realización,

La Fig. 4 ilustra un sistema de Codificación de Objetos de Audio Espacial que comprende un codificador SAOC y un decodificador SAOC,

La Fig. 5 ilustra un decodificador SAOC que comprende un decodificador de información lateral, un separador de objetos y un renderizador,

La Fig. 6 ilustra un comportamiento de las estimaciones de sonoridad de la señal de salida en un cambio de sonoridad,

La Fig. 7 representa la estimación de sonoridad informada según una realización, que ilustra los componentes de un codificador y un decodificador según una realización,

La Fig. 8 ilustra un codificador según otra realización,

La Fig. 9 ilustra un codificador y un decodificador según una realización relacionada con el SAOC-Dialogue Enhancement, que comprende canales anulados,

La Fig. 10 representa una primera ilustración de un cambio medido de sonoridad y el resultado de utilizar los conceptos proporcionados para estimar el cambio de sonoridad de una manera paramétrica,

La Fig. 11 representa una segunda ilustración de un cambio medido de sonoridad y el resultado de utilizar los 5 conceptos proporcionados para estimar el cambio de sonoridad de una manera paramétrica, y

La Fig. 12 ilustra otra realización para llevar a cabo una compensación de sonoridad.

10 **[0030]** Antes de describir con detalle las realizaciones preferidas, se describen la estimación de sonoridad, la Codificación de Objetos de Audio Espacial (SAOC) y la Mejora de Diálogo (DE).

[0031] En primer lugar, se describe la estimación de sonoridad.

15 **[0032]** Como ya se ha indicado anteriormente, la recomendación R128 de la UER se basa en el modelo presentado en la Rec. UIT-R BS.1770 para la estimación de sonoridad. Esta medida se utilizará como ejemplo, pero los conceptos descritos a continuación pueden aplicarse también para otras medidas de sonoridad.

[0033] El funcionamiento de la estimación de sonoridad según BS.1770 es relativamente simple y se basa en 20 las siguientes etapas principales [UIT]:

- La señal de entrada x_i (o señales en el caso de una señal multicanal) se filtra con un filtro K (una combinación de un shelving y un filtro de etapa alto) para obtener la(s) señal(es) y_i
- Se calcula la energía cuadrática media z_i de la señal y_i .

25 - En el caso de la señal multicanal, se aplica la ponderación de canales G_i y se suman las señales ponderadas. Entonces la sonoridad de la señal se define como

$$L = c + 10 \log_{10} \sum_i G_i z_i ,$$

30 con el valor constante $c = -0,691$. La salida se expresa en las unidades de "LKFS" (sonoridad, K-ponderado, relativo a escala máxima) que escala de forma similar a la escala de decibelios.

[0034] En la fórmula anterior, G_i puede, por ejemplo, ser igual a 1 para algunos de los canales, mientras que G_i puede, por ejemplo, ser 1,41 para otros canales. Por ejemplo, si se considera un canal izquierdo, un canal 35 derecho, un canal central, un canal envolvente izquierdo y un canal envolvente derecho, los pesos G_i respectivos pueden ser, por ejemplo, 1 para los canales izquierdo, derecho y central y puede, por ejemplo, ser 1,41 para el canal envolvente izquierdo y el canal envolvente derecho, consulte [UIT].

[0035] Se puede ver, que el valor de sonoridad L está estrechamente relacionado con el logaritmo de la 40 energía de la señal.

[0036] A continuación, se describe la Codificación de Objetos de Audio Espacial.

[0037] Los conceptos de codificación de audio basados en objetos permiten una gran flexibilidad en el lado 45 del decodificador de la cadena. Un ejemplo de un concepto de codificación de audio basado en objetos es la Codificación de Objetos de Audio Espacial (SAOC).

[0038] La Fig. 4 ilustra un sistema de Codificación de Objetos de Audio Espacial (SAOC) que comprende un 50 codificador SAOC 410 y un decodificador SAOC 420,

[0039] El codificador SAOC 410 recibe N señales de objetos de audio S_1, \dots, S_N como entrada. Además, el codificador SAOC 410 recibe instrucciones "Información de mezcla **D**" cómo estos objetos deben combinarse para obtener una señal de mezcla descendente que comprende M canales de mezcla descendente X_1, \dots, X_M . El codificador SAOC 410 extrae cierta información lateral a partir de los objetos y a partir de proceso de mezcla 55 descendente, y esta información lateral se transmite y/o se almacena junto con las señales de mezcla descendente.

[0040] Una propiedad importante de un sistema SAOC es que la señal de mezcla descendente ue

comprende los canales de mezcla descendente X_1, \dots, X_M forma una señal semánticamente significativa. En otras palabras, es posible escuchar la señal de mezcla descendente. Si, por ejemplo, el receptor no tiene la funcionalidad del decodificador SAOC, el receptor puede, a pesar de ello, proporcionar siempre la señal de mezcla descendente como la salida.

5

[0041] La Fig. 5 ilustra un decodificador SAOC que comprende un decodificador de información lateral 510, un separador de objetos 520 y un renderizador 530. El decodificador SAOC ilustrado por la Fig. 5 recibe, por ejemplo, desde un codificador SAOC, la señal de mezcla descendente y la información lateral. La señal de mezcla descendente puede considerarse como una señal de entrada de audio que comprende las señales de objeto de audio, ya que las señales de objeto de audio se mezclan dentro de la señal de mezcla descendente (las señales de objeto de audio se mezclan dentro de uno o más canales de mezcla descendente de la señal de mezcla descendente).

10

[0042] El decodificador SAOC puede, por ejemplo, intentar reconstruir (virtualmente) los objetos originales, por ejemplo, empleando el separador de objetos 520, por ejemplo, utilizando la información lateral decodificada. Estas reconstrucciones de objetos (virtuales) $\hat{S}_1, \dots, \hat{S}_N$, por ejemplo, las señales de objetos de audio reconstruidas, se combinan a continuación en base a la información de renderización, por ejemplo, una matriz de renderización \mathbf{R} , que produce K canales de salida de audio Y_1, \dots, Y_K de una señal de salida de audio Y .

15

[0043] En SAOC, a menudo las señales de objetos de audio son, por ejemplo, reconstruidas, por ejemplo, empleando información de covarianza, por ejemplo, una matriz de covarianza de la señal \mathbf{E} , que se transmite desde el codificador SAOC al decodificador SAOC.

20

[0044] Por ejemplo, se puede emplear la siguiente fórmula para reconstruir las señales de objetos de audio en el lado del decodificador:

25

$$\mathbf{S} = \mathbf{G}\mathbf{X} \text{ con } \mathbf{G} \approx \mathbf{E}\mathbf{D}^H (\mathbf{D}\mathbf{E}\mathbf{D}^H)^{-1}$$

en la que

30

N número de señales de objetos de audio,
 $N_{samples}$ número de muestras consideradas de una señal de objeto de audio
 M número de canales de mezcla descendente,
 \mathbf{X} señal de audio de mezcla descendente, tamaño $M \times N_{samples}$

35

\mathbf{D} matriz de mezcla descendente, tamaño $M \times N$
 \mathbf{E} matriz de covarianza de la señal, tamaño $N \times N$ definido como $\mathbf{E} = \mathbf{X}\mathbf{X}^H$
 \mathbf{S} reconstruido paramétricamente N señales de objeto de audio, tamaño $N \times N_{samples}$
 $(\cdot)^H$ operador autoadjunto (hermitiano) que representa el traspuesto conjugado de (\cdot)

40

[0045] A continuación, una matriz de renderización \mathbf{R} puede aplicarse sobre las señales de objetos de audio reconstruidas \mathbf{S} para obtener los canales de salida de audio de la señal de salida de audio \mathbf{Y} , por ejemplo, según la fórmula:

$$\mathbf{Y} = \mathbf{R}\mathbf{S}$$

45

en la que

50

K número de canales de salida de audio Y_1, \dots, Y_K de la señal de salida de audio \mathbf{Y} .
 \mathbf{R} matriz de renderización de tamaño $K \times N$
 \mathbf{Y} señal de salida de audio que comprende los K canales de salida de audio, tamaño $K \times N_{samples}$

[0046] En la Fig. 5, el proceso de reconstrucción de objetos, por ejemplo, llevado a cabo por el separador de objetos 520, se refiere con la noción "virtual" u "opcional", ya que no necesariamente tiene que producirse, pero la funcionalidad deseada puede obtenerse combinando la reconstrucción y las etapas de renderización en el dominio paramétrico (es decir, combinando las ecuaciones).

55

[0047] En otras palabras, en lugar de reconstruir las señales de objetos de audio utilizando la información de mezcla \mathbf{D} y la información de covarianza \mathbf{E} primero y, a continuación, aplicar la información de renderización \mathbf{R} en las señales de objetos de audio reconstruidas para obtener los canales de salida de audio Y_1, \dots, Y_K ambas etapas pueden realizarse en una sola etapa, de modo que los canales de salida de audio Y_1, \dots, Y_K se generan directamente a partir de los canales de mezcla descendente.

[0048] Por ejemplo, se puede emplear la siguiente fórmula:

$$\mathbf{Y} = \mathbf{R}\mathbf{G}\mathbf{X} \text{ con } \mathbf{G} \approx \mathbf{E}\mathbf{D}^H (\mathbf{D}\mathbf{E}\mathbf{D}^H)^{-1} .$$

10

[0049] En principio, la información de renderización \mathbf{R} puede solicitar cualquier combinación de las señales de objetos de audio originales. En la práctica, sin embargo, las reconstrucciones de objetos pueden comprender errores de reconstrucción y la escena de salida solicitada puede no ser necesariamente alcanzada. Como una regla general aproximada que cubre muchos casos prácticos, cuanto más se diferencia la escena de salida solicitada de la señal de mezcla descendente, más errores de reconstrucción audibles habrá.

[0050] A continuación, se describe la mejora de diálogo (DE). La tecnología SAOC puede, por ejemplo, emplearse para producir el escenario. Cabe señalar que, aunque el nombre "mejora de diálogo" sugiere centrarse en las señales orientadas al diálogo, el mismo principio se puede utilizar también con otros tipos de señales.

[0051] En el escenario DE, los grados de libertad en el sistema están limitados a partir del caso general.

[0052] Por ejemplo, las señales de objetos de audio $S_1, \dots, S_N = \mathbf{S}$ se agrupan (y posiblemente se mezclan) en dos metaobjetos de un objeto en primer plano (FGO) S_{FGO} y un objeto de fondo (BGO) S_{BGO} .

[0053] Además, la escena de salida $Y_1, \dots, Y_K = \mathbf{Y}$ se parece a la señal de mezcla descendente $X_1, \dots, X_M = \mathbf{X}$. Más específicamente, ambas señales tienen las mismas dimensionalidades, es decir, $K = M$, y el usuario final solo puede controlar los niveles de mezcla relativos de los dos metaobjetos FGO y BGO. Para ser más exactos, la señal de mezcla descendente se obtiene mezclando el FGO y BGO con unos pesos escalares

$$\mathbf{X} = h_{FGO}\mathbf{S}_{FGO} + h_{BGO}\mathbf{S}_{BGO} ,$$

y la escena de salida se obtiene de manera similar con una ponderación escalar del FGO y del BGO:

35

$$\mathbf{Y} = g_{FGO}\mathbf{S}_{FGO} + g_{BGO}\mathbf{S}_{BGO} .$$

[0054] En función de los valores relativos de los pesos de mezcla, el equilibrio entre el FGO y el BGO puede cambiar. Por ejemplo, con el ajuste

40

$$\begin{cases} g_{FGO} > h_{FGO} \\ g_{BGO} = h_{BGO} \end{cases}$$

[0055] es posible aumentar el nivel relativo del FGO en la mezcla. Si el FGO es el diálogo, este ajuste proporciona funcionalidad en la mejora de diálogo.

45

[0056] Como ejemplo de caso de uso, el BGO puede ser el ruido del estadio y otros sonidos de fondo durante un evento deportivo y el FGO es la voz del comentarista. La funcionalidad DE permite que el usuario final pueda amplificar o atenuar el nivel del comentarista en relación con el fondo.

[0057] Las realizaciones se basan en el hallazgo de que la utilización de la tecnología SAOC (o similar) en un escenario de radiodifusión permite proporcionar al usuario final una funcionalidad extendida de manipulación de la señal. Se proporciona más funcionalidad que solo cambiar el canal y ajustar el volumen de reproducción.

5

[0058] Una posibilidad de emplear la tecnología DE se describe brevemente más arriba. Si la señal de radiodifusión, que es la señal de mezcla descendente para SAOC, está normalizada en nivel, por ejemplo, según R128, los diferentes programas tienen una sonoridad media similar si no se aplica ningún procesamiento (SAOC-) (o la descripción de renderización es la misma que la descripción de mezcla descendente). Sin embargo, si se aplica algún procesamiento (SAOC-), la señal de salida se diferencia de la señal de mezcla descendente por defecto y la sonoridad de la señal de salida puede ser diferente de la sonoridad de la señal de mezcla descendente por defecto. Desde el punto de vista del usuario final, esto puede conducir a una situación en la que la sonoridad de la señal de salida entre canales o programas tenga de nuevo saltos o diferencias no deseables. En otras palabras, los beneficios de la normalización aplicada por el radiodifusor se pierden parcialmente.

10
15

[0059] Este problema no es específico solamente de SAOC o del escenario DE, sino que también puede ocurrir con otros conceptos de codificación de audio que permiten que el usuario final pueda interactuar con el contenido. Sin embargo, en muchos casos, si la señal de salida tiene una sonoridad diferente a la de la mezcla descendente por defecto, no causa ningún daño.

20

[0060] Como se ha indicado antes, la sonoridad total de un programa de señal de entrada de audio debe ser igual a un nivel determinado con pequeñas desviaciones permitidas. Sin embargo, como ya se ha indicado, esto conduce a problemas significativos cuando se realiza la renderización de audio, ya que la renderización puede tener un efecto significativo sobre la sonoridad total/general de la señal de entrada de audio recibida. Sin embargo, a pesar de que se lleva a cabo la renderización de escenas, la sonoridad total de la señal de audio recibida permanecerá igual.

25

[0061] Un enfoque sería estimar la sonoridad de una señal mientras se está reproduciendo y, con un concepto de integración temporal adecuado, la estimación puede converger a la sonoridad media verdadera pasado un tiempo. Sin embargo, el tiempo necesario para la convergencia es problemático desde el punto de vista del usuario final. Cuando la estimación de sonoridad cambia incluso cuando no se aplican cambios a la señal, la compensación de cambio de sonoridad también debe reaccionar y cambiar su comportamiento. Esto conduciría a una señal de salida con una sonoridad media variable en el tiempo, que puede percibirse como algo molesto.

30

[0062] La Fig. 6 ilustra un comportamiento de las estimaciones de sonoridad de la señal de salida sobre un cambio de sonoridad. Entre otras cosas, se representa una estimación de sonoridad de señal de salida basada en la señal, que ilustra el efecto de una solución como se acaba de describir. La estimación se acerca a la estimación correcta muy lentamente. En lugar de una estimación de sonoridad de la señal de salida basada en la señal, sería preferible una estimación informada de sonoridad de señal de salida, que determina inmediatamente la sonoridad de la señal de salida correctamente.

35
40

[0063] En particular, en la Fig. 6, la entrada del usuario, por ejemplo, el nivel del objeto de diálogo, cambia en el instante de tiempo T aumentando de valor. El verdadero nivel de señal de salida y, consecuentemente la sonoridad, cambia al mismo instante. Cuando la estimación de sonoridad de la señal de salida se realiza a partir de la señal de salida transcurrido cierto tiempo de integración temporal, la estimación cambiará gradualmente y alcanzará el valor correcto después de un cierto retardo. Durante este retardo, los valores de estimación están cambiando y no se pueden utilizar de forma fiable para un procesamiento posterior de la señal de salida, por ejemplo, para la corrección del nivel de sonoridad.

45

[0064] Como ya se ha indicado, sería deseable tener una estimación precisa de la sonoridad media de salida o del cambio en la sonoridad media sin retardo y, si el programa no cambia o la escena de renderización no cambia, la estimación de sonoridad media también debe permanecer estática. En otras palabras, cuando se aplica alguna compensación de cambio de sonoridad, el parámetro de compensación solo debe cambiar cuando cambia el programa o si existe alguna interacción con el usuario.

50
55

[0065] El comportamiento deseado se ilustra en la ilustración más baja de la Fig. 6 (estimación informada de sonoridad de señal de salida). La estimación de sonoridad de señal de salida cambiará inmediatamente cuando cambie la entrada del usuario.

[0066] La Fig. 2 ilustra un codificador según una realización.

[0067] El codificador comprende una unidad codificadora basada en objetos 210 que codifica una pluralidad de señales de objetos de audio para obtener una señal de audio codificada que comprende la pluralidad de señales de objetos de audio.

[0068] Por otro lado, el codificador comprende una unidad codificadora de sonoridad de objetos 220 que codifica información de sonoridad sobre las señales de objetos de audio. La información de sonoridad comprende uno o más valores de sonoridad, en la que cada uno, o más, de los valores de sonoridad depende de una, o más, de las señales de objetos de audio.

[0069] Según una realización, cada una de las señales de objetos de audio de la señal de audio codificada se asigna exactamente a un grupo de dos o más grupos, en el que cada uno de los dos o más grupos comprende una, o más, de las señales de objetos de audio de la señal de audio codificada. La unidad codificadora de sonoridad de objetos 220 está configurada para determinar uno, o más, valores de sonoridad de la información de sonoridad determinando un valor de sonoridad para cada grupo de los dos, o más, grupos, en el que dicho valor de sonoridad de dicho grupo indica una sonoridad original total de la, o las, señales de objetos de audio de dicho grupo.

[0070] La Fig. 1 ilustra un decodificador que genera una señal de salida de audio que comprende uno o más canales de salida de audio según una realización.

[0071] El decodificador comprende una interfaz de recepción 110 que recibe una señal de entrada de audio que comprende una pluralidad de señales de objetos de audio, que recibe información de sonoridad sobre las señales de objetos de audio y que recibe información de renderización que indica si una, o más, de las señales de objetos de audio se amplificará o se atenuará.

[0072] Además, el decodificador comprende un procesador de señales 120 que genera uno, o más, canales de salida de audio de la señal de salida de audio. El procesador de señales 120 está configurado para determinar un valor de compensación de sonoridad en función de la información de sonoridad y en función de la información de renderización. Por otro lado, el procesador de señales 120 está configurado para generar uno o más canales de salida de audio de la señal de salida de audio a partir de la señal de entrada de audio en función de la información de renderización y en función del valor de compensación de sonoridad.

[0073] Según una realización, el procesador de señales 110 está configurado para generar uno o más canales de salida de audio de la señal de salida de audio a partir de la señal de entrada de audio en función de la información de renderización y en función del valor de compensación de sonoridad, de tal manera que una sonoridad de la señal de salida de audio es igual a una sonoridad de la señal de entrada de audio, o tal que la sonoridad de la señal de salida de audio está más próxima a la sonoridad de la señal de entrada de audio que una sonoridad de una señal de audio modificada que resultaría de modificar la señal de entrada de audio amplificando o atenuando las señales de objetos de audio de la señal de entrada de audio según la información de renderización.

[0074] Según otra realización, cada una de las señales de objetos de audio de la señal de entrada de audio se asigna exactamente a un grupo de dos o más grupos, en la que cada uno de los dos, o más, grupos comprende una, o más, de las señales de objetos de audio de la señal de entrada de audio.

[0075] En una realización de este tipo, la interfaz de recepción 110 está configurada para recibir un valor de sonoridad respecto a cada grupo de los dos, o más, grupos como la información de sonoridad, en el que dicho valor de sonoridad indica una sonoridad original total de una, o más, señales de audio de dicho grupo. Por otro lado, la interfaz de recepción 110 está configurada para recibir la información de renderización que indica, para al menos un grupo de los dos o más grupos, si la o las señales de objetos de audio de dicho grupo se amplificarán o atenuarán indicando una sonoridad modificada total de la o las señales de objetos de audio de dicho grupo. Además, en una realización de este tipo, el procesador de señales 120 está configurado para determinar el valor de compensación de sonoridad en función de la sonoridad modificada total de cada uno de los dichos, al menos uno, grupos de los dos o más grupos y, en función de la sonoridad original total de cada uno de los dos o más grupos. Además, el procesador de señales 120 está configurado para generar uno o más canales de salida de audio de la señal de salida de audio a partir de la señal de entrada de audio en función de la sonoridad modificada total de cada uno de los dichos, al menos uno, grupos de los dos o más grupos y, en función del valor de compensación de sonoridad.

[0076] En realizaciones particulares, al menos un grupo de los dos o más grupos comprende dos o más de

las señales de objetos de audio.

[0077] Existe una relación directa entre la energía e_i de una señal de objeto de audio i y la sonoridad L_i de la señal de objeto de audio i según las fórmulas:

5

$$L_i = c + 10 \log_{10} e_i, \quad e_i = 10^{(L_i - c)/10}$$

en las que c es un valor constante.

10 **[0078]** Las realizaciones se basan en las siguientes conclusiones: Diferentes señales de objetos de audio de la señal de entrada de audio pueden tener una sonoridad diferente y, por lo tanto, una energía diferente. Si, por ejemplo, un usuario desea aumentar la sonoridad de una de las señales de objetos de audio, la información de renderización puede ajustarse de forma correspondiente y el aumento de la sonoridad de esta señal de objeto de audio aumenta la energía de este objeto de audio. Esto conduciría a una mayor sonoridad de la señal de salida de audio. Para mantener la sonoridad total constante, debe realizarse una compensación de sonoridad. En otras palabras, la señal de audio modificada que resultaría de la aplicación de la información de renderización en la señal de entrada de audio tendría que ser ajustada. Sin embargo, el efecto exacto de la amplificación de una de las señales de objetos de audio en la sonoridad total de la señal de audio modificada depende de la sonoridad original de la señal de objeto de audio amplificado, por ejemplo, de la señal de objeto de audio. Si la sonoridad original de este objeto corresponde a una energía que era bastante baja, el efecto sobre la sonoridad total de la señal de entrada de audio será menor. Si, no obstante, la sonoridad original de este objeto corresponde a una energía que era bastante alta, el efecto sobre la sonoridad total de la señal de entrada de audio será significativo.

25 **[0079]** Se pueden considerar dos ejemplos. En ambos ejemplos, una señal de entrada de audio comprende dos señales de audio y, en ambos ejemplos, aplicando la información de renderización, la energía de una primera señal de objeto de audio se incrementa en un 50%.

30 **[0080]** En el primer ejemplo, la primera señal de objeto de audio contribuye con el 20% y la segunda señal de objeto de audio contribuye con el 80% de la energía total de la señal de entrada de audio. Sin embargo, en el segundo ejemplo, el primer objeto de audio, la primera señal de objeto de audio contribuye con el 40% y la segunda señal de objeto de audio contribuye con el 60% de la energía total de la señal de entrada de audio. En ambos ejemplos estas contribuciones son derivables de la información de sonoridad sobre las señales de objetos de audio, ya que existe una relación directa entre sonoridad y energía.

35 **[0081]** En el primer ejemplo, un aumento del 50% de la energía del primer objeto de audio da como resultado que, una señal de audio modificada que se genera aplicando la información de renderización sobre la señal de entrada de audio, tiene una energía total $1,5 \times 20\% + 80\%$ 110% de la energía de la señal de entrada de audio.

40 **[0082]** En el segundo ejemplo, un aumento del 50% de la energía del primer objeto de audio da como resultado que, la señal de audio modificada que se genera aplicando la información de renderización en la señal de entrada de audio, tiene una energía total $1,5 \times 40\% + 60\%$ 120% de la energía de la señal de entrada de audio.

45 **[0083]** Por lo tanto, después de aplicar la información de renderización sobre la señal de entrada de audio, en el primer ejemplo, la energía total de la señal de audio modificada debe reducirse solo en un 9% (10/110) para obtener la misma energía tanto en la señal de entrada de audio como en la señal de salida de audio, mientras que en el segundo ejemplo, la energía total de la señal de audio modificada tiene que ser reducida en un 17% (20/120). Para este propósito, se puede calcular un valor de compensación de sonoridad.

50 **[0084]** Por ejemplo, el valor de compensación de sonoridad puede ser un escalar que se aplica en todos los canales de salida de audio de la señal de salida de audio.

55 **[0085]** Según una realización, el procesador de señales está configurado para generar la señal de audio modificada modificando la señal de entrada de audio amplificando o atenuando las señales de objetos de audio de la señal de entrada de audio según la información de renderización. Además, el procesador de señales está configurado para generar la señal de salida de audio aplicando el valor de compensación de sonoridad en la señal de audio modificada, de manera que la sonoridad de la señal de salida de audio sea igual a la sonoridad de la señal de entrada de audio, o de manera que la sonoridad de la señal de salida de audio está más cerca de la sonoridad de

la señal de entrada de audio que la sonoridad de la señal de audio modificada.

[0086] Por ejemplo, en el primer ejemplo anterior, el valor de compensación de sonoridad /cv puede, por ejemplo, ajustarse a un valor /cv = 10/11, y un factor de multiplicación de 10/11 puede aplicarse en todos los canales 5 que resulten de renderizar los canales de entrada de audio según la información de renderización.

[0087] Por consiguiente, por ejemplo, en el segundo ejemplo anterior, el valor de compensación de sonoridad /cv puede, por ejemplo, ajustarse a un valor /cv = 10/12 = 5/6, y un factor de multiplicación de 5/6 puede ser aplicado 10 en todos los canales que resultan de renderizar los canales de entrada de audio según la información de renderización.

[0088] En otras realizaciones, cada una de las señales de objetos de audio se puede asignar a uno entre una pluralidad de grupos y se puede transmitir un valor de sonoridad para cada uno de los grupos que indica un valor de sonoridad total de las señales de objetos de audio de dicho grupo. Si la información de renderización especifica que 15 la energía de uno de los grupos es atenuada o amplificada, por ejemplo, amplificada en un 50% como se ha indicado anteriormente, se puede calcular un incremento total de energía y se puede determinar un valor de compensación de sonoridad como se ha descrito anteriormente.

[0089] Por ejemplo, según una realización, cada una de las señales de objetos de audio de la señal de 20 entrada de audio está asignada exactamente a un grupo de exactamente dos grupos como los dos o más grupos. Cada una de las señales de audio de la señal de entrada de audio se asigna a un grupo de objetos en primer plano de los exactamente dos grupos o a un grupo de objetos de fondo de los exactamente dos grupos. La interfaz de recepción 110 está configurada para recibir la sonoridad original total de una o más señales de objetos de audio del grupo de objetos en primer plano. Además, la interfaz de recepción 110 está configurada para recibir la sonoridad 25 original total de una o más señales de objetos de audio del grupo de objetos de fondo. Además, la interfaz de recepción 110 está configurada para recibir la información de renderización que indica, para al menos un grupo de los exactamente dos grupos, si una o más señales de objetos de audio de cada uno del dicho al menos un grupo, se amplificarán o atenuarán indicando una sonoridad total modificada de la o las señales de objetos de audio de dicho grupo.

[0090] En una realización de este tipo, el procesador de señales 120 está configurado para determinar el 30 valor de compensación de sonoridad en función de la sonoridad modificada total de cada uno de los dichos, al menos un grupo, en función de la sonoridad original total de una o más señales de objetos de audio del objeto en primer plano, y en función de la sonoridad original total de una o más señales de objetos de audio del grupo de 35 objetos de fondo. Además, el procesador de señales 120 está configurado para generar uno o más canales de salida de audio de la señal de salida de audio de la señal de entrada de audio en función de la sonoridad modificada total de cada uno de los dichos, al menos uno, grupos en función del valor de compensación de sonoridad.

[0091] De acuerdo con algunas realizaciones, cada una de las señales de objetos de audio está asignada a 40 uno de tres o más grupos y la interfaz de recepción puede configurarse para recibir un valor de sonoridad para cada uno de los tres o más grupos que indican la sonoridad total de las señales de objetos de audio de dicho grupo.

[0092] Según una realización, para determinar el valor total de sonoridad de dos o más señales de objetos de 45 audio, por ejemplo, se determina el valor de energía correspondiente al valor de sonoridad para cada señal de objeto de audio, se suman los valores de energía de todos los valores de sonoridad para obtener una suma de energías y el valor de sonoridad correspondiente a la suma de energías se determina como el valor de sonoridad total de las dos o más señales de objetos de audio. Por ejemplo, pueden emplearse las fórmulas

$$L_i = c + 10 \log_{10} e_i, \quad e_i = 10^{(L_i - c)/10}$$

50 **[0093]** En algunas realizaciones, se transmiten valores de sonoridad para cada una de las señales de objetos de audio, o cada una de las señales de objetos de audio se asigna a uno o dos o más grupos, en las que para cada uno de los grupos se transmite un valor de sonoridad.

55 **[0094]** Sin embargo, en algunas realizaciones, para una o más señales de objetos de audio o para uno o más de los grupos que comprenden señales de objetos de audio, no se transmite valor de sonoridad. En su lugar, el decodificador puede asumir, por ejemplo, que estas señales de objetos de audio o grupos de señales de objetos de

audio, para las que no se transmite un valor de sonoridad, tienen un valor de sonoridad predefinido. El decodificador puede, por ejemplo, basar todas las determinaciones adicionales en este valor de sonoridad predefinido.

[0095] Según una realización, la interfaz de recepción 110 está configurada para recibir una señal de mezcla descendente que comprende uno o más canales de mezcla descendente como la señal de entrada de audio, en la que uno o más canales de mezcla descendente comprenden las señales de objetos de audio y en la que el número de las señales de objetos de audio es menor que el número de uno o más canales de mezcla descendente. La interfaz de recepción 110 está configurada para recibir información de mezcla descendente que indica cómo se mezclan las señales de objetos de audio dentro de uno o más canales de mezcla descendente. Además, el procesador de señales 120 está configurado para generar uno o más canales de salida de audio de la señal de salida de audio a partir de la señal de entrada de audio en función de la información de mezcla descendente, en función de la información de renderización y en función del valor de compensación de sonoridad. En una realización particular, el procesador de señales 120 puede, por ejemplo, estar configurado para calcular el valor de compensación de sonoridad en función de la información de mezcla descendente.

[0096] Por ejemplo, la información de mezcla descendente puede ser una matriz de mezcla descendente. En realizaciones, el decodificador puede ser un decodificador SAOC. En tales realizaciones, la interfaz de recepción 110 puede, por ejemplo, estar configurada adicionalmente para recibir una información de covarianza, por ejemplo, una matriz de covarianza como se ha descrito anteriormente.

[0097] Con respecto a la información de renderización que indica si una o más de las señales de objetos de audio se amplificará o atenuará, debe observarse que, por ejemplo, la información que indica cómo se deben amplificar o atenuar una o más de las señales de objetos de audio, es la información de renderización. Por ejemplo, una matriz de renderización R , por ejemplo, una matriz de renderización de SAOC, es la información de renderización.

[0098] La Fig. 3 ilustra un sistema según una realización,

[0099] El sistema comprende un codificador 310 según una de las realizaciones descritas anteriormente que codifica una pluralidad de señales de objetos de audio para obtener una señal de audio codificada que comprende la pluralidad de señales de objetos de audio.

[0100] Además, el sistema comprende un decodificador 320 según una de las realizaciones descritas anteriormente para generar una señal de salida de audio que comprende uno o más canales de salida de audio. El decodificador está configurado para recibir la señal de audio codificada como una señal de entrada de audio y la información de sonoridad. Además, el decodificador 320 está configurado para recibir información de renderización adicional. Por otro lado, el decodificador 320 está configurado para determinar un valor de compensación de sonoridad en función de la información de sonoridad y en función de la información de renderización. Además, el decodificador 320 está configurado para generar uno o más canales de salida de audio de la señal de salida de audio a partir de la señal de entrada de audio en función de la información de renderización y en función del valor de compensación de sonoridad.

[0101] La Fig. 7 ilustra la estimación de sonoridad informada según una realización. A la izquierda del flujo de transporte 730, se ilustran componentes de un codificador de codificación de audio basada en objetos. En particular, se ilustra una unidad codificadora basada en objetos 710 ("codificador de audio basado en objetos") y una unidad codificadora de sonoridad de objetos 720 ("estimación de sonoridad de objetos").

[0102] El flujo de transporte 730 comprende por sí mismo la información de sonoridad L , la información de mezcla D y la salida del codificador de audio basado en objetos 710 B .

[0103] A la derecha del flujo de transporte 730, se ilustran componentes de un procesador de señales de codificación de audio basada en objetos. La interfaz de recepción del decodificador no se ilustra. Se representa un estimador de sonoridad de salida 740 y una unidad de decodificación de audio basada en objetos 750. El estimador de sonoridad de salida 740 puede configurarse para determinar el valor de compensación de sonoridad. La unidad decodificadora de audio basada en objetos 750 puede configurarse para determinar una señal de audio modificada a partir de una señal de audio, que es introducida en el decodificador, aplicando la información de renderización R . La aplicación del valor de compensación de sonoridad en la señal de audio modificada para compensar un cambio de sonoridad total provocado por la renderización no se muestra en la Fig. 7.

[0104] La entrada al codificador consiste en los objetos de entrada **S** en el mínimo. El sistema calcula la sonoridad de cada objeto (o alguna otra información relacionada con la sonoridad, tal como las energías de los objetos), por ejemplo, con la unidad codificadora de sonoridad de objetos 720, y esta información **L** se transmite y/o se almacena. (También es posible que la sonoridad de los objetos se proporcione como una entrada al sistema y se puede omitir la etapa de estimación dentro del sistema).

[0105] En la realización de la Fig. 7, el decodificador recibe, al menos la información de sonoridad del objeto, y, por ejemplo, la información de renderización **R** describiendo la mezcla de los objetos en la señal de salida. En base a éstos, por ejemplo, el estimador de sonoridad de salida 740, estima la sonoridad de la señal de salida y proporciona esta información como salida.

[0106] La información de mezcla descendente **D** puede proporcionarse como la información de renderización, en cuyo caso la estimación de sonoridad proporciona una estimación de sonoridad de la señal de mezcla descendente. También es posible proporcionar la información de mezcla descendente como una entrada a la estimación de sonoridad de objetos, y transmitirla y/o almacenarla a lo largo de la información de sonoridad de objetos. La estimación de sonoridad de salida puede entonces estimar simultáneamente la sonoridad de la señal de mezcla descendente y la salida renderizada, y proporcionar estos dos valores o su diferencia como la información de sonoridad de salida. El valor de diferencia (o su inversa) describe la compensación necesaria que debe aplicarse a la señal de salida renderizada para hacer que su sonoridad sea similar a la sonoridad de la señal de mezcla descendente. La información de sonoridad de objetos puede incluir adicionalmente información con respecto a los coeficientes de correlación entre diversos objetos y esta información de correlación puede utilizarse en la estimación de sonoridad de salida para obtener una estimación más precisa.

[0107] A continuación, se describe una realización preferente para la aplicación de mejora de diálogo.

[0108] En la aplicación de mejora de diálogo, como se ha descrito anteriormente, las señales de objetos de audio de entrada están agrupadas y parcialmente mezcladas para formar dos metaobjetos, FGO y BGO, que pueden sumarse trivialmente para obtener la señal final de mezcla descendente.

[0109] Siguiendo la descripción de SAOC [SAOC], N señales de objeto de entrada se representan como una matriz **S** del tamaño $N \times N_{samples}$, y la información de mezcla descendente como una matriz **D** del tamaño $M \times N$. Las señales de mezcla descendente pueden obtenerse como $\mathbf{X} = \mathbf{D}\mathbf{S}$.

[0110] En este punto, la información de mezcla descendente **D** puede dividirse en dos partes

$$\mathbf{D} = \mathbf{D}_{FGO} + \mathbf{D}_{BGO}$$

para los metaobjetos.

[0111] Como cada columna de la matriz **D** se corresponde con una señal de objeto de audio original, las matrices de mezcla descendente de dos componentes se pueden obtener estableciendo las columnas, que se corresponden con el otro metaobjeto en cero (suponiendo que ningún objeto original puede estar presente en ambos metaobjetos). En otras palabras, las columnas correspondientes al metaobjeto BGO se ponen a cero en \mathbf{D}_{FGO} , y viceversa.

[0112] Estas nuevas matrices de mezcla descendente describen la forma en que se pueden obtener los dos metaobjetos a partir de los objetos de entrada, concretamente:

$$\mathbf{S}_{FGO} = \mathbf{D}_{FGO}\mathbf{S} \quad \text{y} \quad \mathbf{S}_{BGO} = \mathbf{D}_{BGO}\mathbf{S},$$

y la mezcla descendente real se simplifica para

$$\mathbf{X} = \mathbf{S}_{FGO} + \mathbf{S}_{BGO} .$$

[0113] También se puede considerar que el decodificador de objeto (por ejemplo, SAOC) intenta reconstruir los metaobjetos:

5

$$\tilde{\mathbf{S}}_{FGO} \approx \mathbf{S}_{FGO} \quad \text{y} \quad \tilde{\mathbf{S}}_{BGO} \approx \mathbf{S}_{BGO} ,$$

y la renderización específica de DE se puede escribir como una combinación de estas dos reconstrucciones de metaobjetos:

10

$$\mathbf{Y} = \mathbf{g}_{FGO} \mathbf{S}_{FGO} + \mathbf{g}_{BGO} \mathbf{S}_{BGO} \approx \mathbf{g}_{FGO} \tilde{\mathbf{S}}_{FGO} + \mathbf{g}_{BGO} \tilde{\mathbf{S}}_{BGO} .$$

[0114] La estimación de sonoridad de objetos recibe los dos metaobjetos \mathbf{S}_{FGO} y \mathbf{S}_{BGO} como la entrada y estima la sonoridad de cada uno de ellos: Siendo L_{BGO} la sonoridad (total/general) de \mathbf{S}_{FGO} y L_{FGO} la sonoridad (total/general) de \mathbf{S}_{BGO} . Estos valores de sonoridad se transmiten y/o se almacenan.

15

[0115] Como alternativa, utilizando uno de los metaobjetos, por ejemplo, el FGO, como referencia, es posible calcular la diferencia de sonoridad de estos dos objetos, por ejemplo, como

$$\Delta L_{FGO} = L_{BGO} - L_{FGO} .$$

20

[0116] Este único valor se transmite y/o se almacena.

[0117] La Fig. 8 ilustra un codificador según otra realización. El codificador de la Fig. 8 comprende un mezclador descendente de objetos 811 y un estimador de información lateral de objetos 812. Por otro lado, el codificador de la Fig. 8 comprende además una unidad codificadora de sonoridad de objetos 820. Además, el codificador de la Fig. 8 comprende un mezclador de metaobjetos de audio 805.

25

[0118] El codificador de la Fig. 8 utiliza metaobjetos de audio intermedios como una entrada para la estimación de sonoridad de objetos. En realizaciones, el codificador de la Fig. 8 puede configurarse para generar dos metaobjetos de audio. En otras realizaciones, el codificador de la Fig. 8 puede configurarse para generar tres o más metaobjetos de audio.

30

[0119] Entre otras cosas, los conceptos proporcionados proporcionan la nueva característica de que el codificador puede, por ejemplo, estimar la sonoridad media de todos los objetos de entrada. Los objetos pueden, por ejemplo, mezclarse en una señal de mezcla descendente que se transmite. Los conceptos proporcionados proporcionan además la nueva característica de que la sonoridad de objetos y la información de mezcla descendente pueden, por ejemplo, ser incluidas en la información lateral de codificación de objetos que se transmite.

35

[0120] El decodificador puede, por ejemplo, utilizar la información lateral de codificación de objetos para la separación (virtual) de los objetos y recombina los objetos utilizando la información de renderización.

40

[0121] Además, los conceptos proporcionados proporcionan la nueva característica de que la información de mezcla descendente puede utilizarse para estimar la sonoridad de la señal de mezcla descendente por defecto, la información de renderización y la sonoridad de objetos recibidos pueden utilizarse para estimar la sonoridad media de la señal de salida y/o el cambio de sonoridad se puede estimar a partir de estos dos valores. O bien, la información de mezcla descendente y de renderización puede utilizarse para estimar el cambio de sonoridad a partir de la mezcla descendente por defecto, otra nueva característica de los conceptos proporcionados.

45

[0122] Además, los conceptos proporcionados proporcionan la nueva característica de que la salida del decodificador puede modificarse para compensar el cambio en la sonoridad de modo que la sonoridad media de la

50

señal modificada coincida con la sonoridad media de la mezcla descendente por defecto.

[0123] En la Fig. 9 se ilustra una realización específica relacionada con SAOC-DE. El sistema recibe las señales de objetos de audio de entrada, la información de mezcla descendente y la información de la agrupación de los objetos en metaobjetos. En base a estos, el mezclador de metaobjetos de audio 905 forma los dos metaobjetos S_{FGO} y S_{BGO} . Es posible que la parte de la señal que se procesa con SAOC no constituya la señal entera. Por ejemplo, en una configuración de canal 5.1, SAOC puede desplegarse en un subconjunto de canales, como en el canal frontal (izquierdo, derecho y central), mientras que los otros canales (envolvente izquierdo, envolvente derecho y efectos de frecuencia baja) se desvían, pasan por alto la SAOC y se proporcionan como tales. Estos canales no procesados por SAOC se indican con X_{BYPASS} . Los posibles canales anulados deben proporcionarse para que el codificador pueda realizar una estimación más precisa de la información de sonoridad.

[0124] Los canales anulados pueden manejarse de varias maneras.

15 **[0125]** Por ejemplo, los canales anulados pueden, por ejemplo, formar un metaobjeto independiente. Esto permite definir la renderización para que los tres metaobjetos se escalen de forma independiente.

[0126] O bien, por ejemplo, los canales anulados pueden, por ejemplo, combinarse con uno de los otros dos metaobjetos. Las configuraciones de renderización de ese metaobjeto también controlan la parte de canales anulados. Por ejemplo, en el escenario de mejora de diálogo, puede ser significativo combinar los canales anulados con el metaobjeto de fondo: $X_{BGO} = S_{BGO} + X_{BYPASS}$

[0127] O bien, por ejemplo, los canales anulados pueden, por ejemplo, ser ignorados.

25 **[0128]** Según las realizaciones, la unidad codificadora basada en objetos 210 del codificador está configurada para recibir las señales de objetos de audio, en la que cada una de las señales de objetos de audio está asignada exactamente a uno de exactamente de dos grupos, en la que cada uno de los exactamente dos grupos comprende una o más de las señales de objetos de audio. Además, la unidad codificadora basada en objetos 210 está configurada para mezclar de forma descendente las señales de objetos de audio, comprendidas en los exactamente dos grupos, para obtener una señal de mezcla descendente que comprende uno o más canales de audio de mezcla descendente como la señal de audio codificada, en la que el número de uno o más canales de mezcla descendente es menor que el número de las señales de objetos de audio que están comprendidas exactamente por los dos grupos. La unidad codificadora de sonoridad de objetos 220 está asignada para recibir una o más señales de objetos de audio anuladas adicionales, en la que cada una de la o las señales de objetos de audio de anuladas adicionales está asignada a un tercer grupo, en la que cada una de la o las señales de objetos de audio anuladas no está comprendida en el primer grupo y no está comprendida en el segundo grupo, en la que la unidad codificadora basada en objetos 210 está configurada para no mezclar la o las señales de objeto de audio anuladas dentro de la señal de mezcla descendente.

40 **[0129]** En una realización, la unidad codificadora de sonoridad de objetos 220 está configurada para determinar un primer valor de sonoridad, un segundo valor de sonoridad y un tercer valor de sonoridad de la información de sonoridad, el primer valor de sonoridad que indica una sonoridad total de una o más señales de objetos de audio del primer grupo, el segundo valor de sonoridad que indica una sonoridad total de una o más señales de objetos de audio del segundo grupo y el tercer valor de sonoridad que indica una sonoridad total de una o más señales de objetos de audio anuladas adicionales del tercer grupo. En otra realización, la unidad codificadora de sonoridad de objetos 220 está configurada para determinar un primer valor de sonoridad y un segundo valor de sonoridad de la información de sonoridad, el primer valor de sonoridad que indica una sonoridad total de una o más señales de objeto de audio del primer grupo, y el segundo valor de sonoridad que indica una sonoridad total de una o más señales de objetos de audio del segundo grupo y de una o más señales de objetos de audio anuladas adicionales del tercer grupo.

[0130] Según una realización, la interfaz de recepción 110 del decodificador está configurada para recibir la señal de mezcla descendente. Además, la interfaz de recepción 110 está configurada para recibir una o más señales de objetos de audio anuladas adicionales, en la que la o las señales de objetos de audio anuladas adicionales no se mezclan dentro de la señal de mezcla descendente. Además, la interfaz de recepción 110 está configurada para recibir la información de sonoridad que indica información sobre la sonoridad de las señales de objetos de audio que se mezclan dentro de la señal de mezcla descendente e información que indica la sonoridad de una o más señales de objetos de audio anuladas adicionales que no se han mezclado dentro de la señal de mezcla descendente. Además, el procesador de señal 120 está configurado para determinar el valor de compensación de sonoridad en

función de la información sobre la sonoridad de las señales de objetos de audio que se mezclan dentro de la señal de mezcla descendente y en función de la información sobre la sonoridad de la o las señales de objetos de audio anuladas adicionales que no se mezclan dentro de la señal de mezcla descendente.

5 **[0131]** La Fig. 9 ilustra un codificador y un decodificador según una realización relacionada con el SAOC-DE, que comprende canales anulados. Entre otras cosas, el codificador de la Fig. 9 comprende un codificador SAOC 902.

10 **[0132]** En la realización de la Fig. 9, la posible combinación de los canales anulados con los otros metaobjetos tiene lugar en los dos bloques de "inclusión de anuladas" 913, 914, que producen los metaobjetos X_{FGO} y X_{BGO} con las partes definidas a partir de los canales anulados incluidos.

15 **[0133]** La sonoridad perceptual L_{BYPASS} , L_{FGO} , y L_{BGO} de ambos metaobjetos se calculan en las unidades de estimación de sonoridad 921, 922, 923. Esta información de sonoridad se transforma entonces en una codificación adecuada de un estimador de información de sonoridad de metaobjetos 925 que después se transmite y/o se almacena.

20 **[0134]** El codificador y decodificador SAOC real funciona como se espera extrayendo la información lateral de objetos a partir de los objetos, creando la señal de mezcla descendente X y transmitiendo y/o almacenando la información en el decodificador. Los posibles canales anulados se transmiten y/ o se almacenan junto con la otra información en el decodificador.

25 **[0135]** El decodificador SAOC-DE 945 recibe un valor de ganancia "Ganancia de diálogo" como una entrada del usuario. En base a esta entrada y en la información de mezcla descendente recibida, el decodificador SAOC 945 determina la información de renderización. A continuación, el decodificador SAOC 945 produce la escena de salida renderizada como la señal Y . Además de eso, produce un factor de ganancia (y un valor de retardo) que debe aplicarse sobre las posibles señales anuladas X_{BYPASS} .

30 **[0136]** La unidad de "inclusión de anuladas" 955 recibe esta información junto con la escena de salida renderizada y las señales anuladas y crea la señal de escena de salida completa. El decodificador SAOC 945 produce también un conjunto de valores de ganancia de metaobjeto, cuya cantidad depende del agrupamiento de metaobjetos y de la forma de información de sonoridad deseada.

35 **[0137]** Los valores de ganancia se proporcionan al estimador de sonoridad de la mezcla 960 que también recibe la información de sonoridad de metaobjetos desde el codificador.

40 **[0138]** A continuación, el estimador de sonoridad de la mezcla 960 puede determinar la información de sonoridad deseada, que puede incluir, entre otros, la sonoridad de la señal de mezcla descendente, la sonoridad de la escena de salida renderizada y/o la diferencia en la sonoridad entre la señal de mezcla descendente y la escena de salida renderizada.

45 **[0139]** En algunas realizaciones, la información de sonoridad en sí misma es suficiente, mientras que en otras realizaciones es deseable procesar la salida completa en función de la información de sonoridad determinada. Este procesamiento puede ser, por ejemplo, una compensación de cualquier diferencia posible en la sonoridad entre la señal de mezcla descendente y la escena de salida renderizada. Dicho procesamiento, por ejemplo, con una unidad de procesamiento de sonoridad 970, tendría sentido en el escenario de radiodifusión, ya que reduciría los cambios en la sonoridad de la señal percibida independientemente de la interacción del usuario (ajuste de la entrada "ganancia de diálogo").

50 **[0140]** El procesamiento relacionado con la sonoridad en esta realización específica comprende la pluralidad de nuevas características. Entre otras cosas, el FGO, el BGO y los posibles canales anulados se mezclan previamente en la configuración de canal final de modo que la mezcla descendente se puede hacer simplemente añadiendo las dos señales premezcladas juntas (por ejemplo, coeficientes de matriz de mezcla descendente de 1), lo que constituye una nueva característica. Además, como una nueva característica adicional, se calcula la
55 sonoridad media del FGO y del BGO y se calcula la diferencia. Por otro lado, los objetos se mezclan en una señal de mezcla descendente que se transmite. Además, como otra característica nueva, la información de diferencia de sonoridad se incluye en la información lateral que se transmite. (nuevo) Por otro lado, el decodificador utiliza la información lateral para la separación (virtual) de los objetos y vuelve a combinar los objetos utilizando la información de renderización que se basa en la información de mezcla descendente y la ganancia de modificación como entrada

del usuario. Además, como otra característica nueva, el decodificador utiliza la ganancia de modificación y la información de sonoridad transmitida para estimar el cambio en la sonoridad media de la salida del sistema en comparación con la mezcla descendente por defecto.

5 **[0141]** A continuación, se proporciona una descripción formal de las realizaciones.

[0142] Suponiendo que los valores de sonoridad del objeto se comportan de forma similar al logaritmo de los valores energéticos cuando se suman los objetos, es decir, los valores de sonoridad deben transformarse en dominio lineal, se añaden allí y finalmente se transforman de nuevo en el dominio logarítmico. Incentivando esto a
10 través de la definición de medida de sonoridad BS.1770, en este punto se presentará (por sencillez, el número de canales se establece en uno, pero el mismo principio se puede aplicar en señales multicanal con la suma adecuada a través de los canales).

[0143] La sonoridad de la señal i K-filtrada z_i con la energía cuadrada media e_i se define como
15

$$L_i = c + 10 \log_{10} e_i,$$

en la que c es una constante de desplazamiento. Por ejemplo, c puede ser $-0,691$. De esto se deduce que la energía de la señal puede determinarse a partir de la sonoridad con
20

$$e_i = 10^{(L_i - c)/10}.$$

$$z_{SUM} = \sum_{i=1}^N z_i$$

[0144] La energía de la suma de N señales no correlacionadas

$$e_{SUM} = \sum_{i=1}^N e_i = \sum_{i=1}^N 10^{(L_i - c)/10},$$

25 y la sonoridad de esta suma de señales es

$$L_{SUM} = c + 10 \log_{10} e_{SUM} = c + 10 \log_{10} \sum_{i=1}^N 10^{(L_i - c)/10}.$$

30 **[0145]** Si las señales no están sin correlación, los coeficientes de correlación C_{ij} deben tenerse en cuenta al aproximar la energía de la suma de señales como

$$e_{SUM} = \sum_{i=1}^N \sum_{j=1}^N e_{i,j},$$

35 en la que la energía cruzada e_{ij} entre los objetos i -ésimo y j -ésimo se define como

$$\begin{aligned}
 e_{i,j} &= C_{i,j} \sqrt{e_i e_j} \\
 &= C_{i,j} \sqrt{10^{(L_i-c)/10} 10^{(L_j-c)/10}}, \\
 &= C_{i,j} \sqrt{10^{(L_i+L_j-2c)/10}}
 \end{aligned}$$

en la que $-1 \leq C_{i,j} \leq 1$ es el coeficiente de correlación entre los dos objetos i y j . Cuando dos objetos no están correlacionados, el coeficiente de correlación es igual a 0, y cuando los dos objetos son idénticos, el coeficiente de correlación es igual a 1.

[0146] Si se amplía adicionalmente el modelo con los pesos de mezcla que hay que aplicar sobre las señales

$$z_{SUM} = \sum_{i=1}^N g_i z_i$$

en el proceso de mezcla, es decir, , la energía de la suma de señales será

$$e_{SUM} = \sum_{i=1}^N \sum_{j=1}^N g_i g_j e_{i,j},$$

10

y la sonoridad de la señal de mezcla se puede obtener a partir de esto, como anteriormente, con

$$L_{SUM} = c + 10 \log_{10} e_{SUM}.$$

15

[0147] La diferencia entre la sonoridad de las dos señales se puede estimar como

$$\Delta L(i, j) = L_i - L_j.$$

20 **[0148]**

Si la definición de sonoridad se utiliza ahora como antes, puede escribirse como

$$\begin{aligned}
 \Delta L(i, j) &= L_i - L_j \\
 &= (c + 10 \log_{10} e_i) - (c + 10 \log_{10} e_j), \\
 &= 10 \log_{10} \frac{e_i}{e_j}
 \end{aligned}$$

lo que se puede observar como una función de las energías de las señales. Si ahora se desea estimar la diferencia de sonoridad entre las dos mezclas

25

$$z_A = \sum_{i=1}^N g_i z_i \quad \text{y} \quad z_B = \sum_{i=1}^N h_i z_i$$

con pesos de mezcla posiblemente diferentes g_i y h_i , se puede estimar con

$$\begin{aligned} \Delta L(A, B) &= 10 \log_{10} \frac{e_A}{e_B} \\ &= 10 \log_{10} \frac{\sum_{i=1}^N \sum_{j=1}^N g_i g_j e_{i,j}}{\sum_{i=1}^N \sum_{j=1}^N h_i h_j e_{i,j}} \\ &= 10 \log_{10} \frac{\sum_{i=1}^N \sum_{j=1}^N g_i g_j C_{i,j} \sqrt{10^{(L_i+L_j-2c)/10}}}{\sum_{i=1}^N \sum_{j=1}^N h_i h_j C_{i,j} \sqrt{10^{(L_i+L_j-2c)/10}}} \\ &= 10 \log_{10} \frac{\sum_{i=1}^N \sum_{j=1}^N g_i g_j C_{i,j} \sqrt{10^{(L_i+L_j)/10}}}{\sum_{i=1}^N \sum_{j=1}^N h_i h_j C_{i,j} \sqrt{10^{(L_i+L_j)/10}}} \end{aligned}$$

5

[0149] En el caso de que los objetos no estén correlacionados ($C_{ij} = 0, \forall i \neq j$ y $C_{ij} = 1, \forall i = j$), la estimación de la diferencia se convierte en

$$\begin{aligned} \Delta L(A, B) &= 10 \log_{10} \frac{\sum_{i=1}^N g_i^2 10^{(L_i-c)/10}}{\sum_{i=1}^N h_i^2 10^{(L_i-c)/10}} \\ &= 10 \log_{10} \frac{\sum_{i=1}^N g_i^2 10^{L_i/10}}{\sum_{i=1}^N h_i^2 10^{L_i/10}} \end{aligned}$$

10

[0150] A continuación, se considera la codificación diferencial.

[0151] Es posible codificar los valores de sonoridad por objeto como diferencias con respecto a la sonoridad de un objeto de referencia seleccionado:

15

$$K_i = L_i - L_{REF}$$

en la que L_{REF} es la sonoridad del objeto de referencia. Esta codificación es favorable si no se necesitan valores absolutos de sonoridad como resultado, porque ahora es necesario transmitir un valor menos y la estimación de la diferencia de sonoridad puede escribirse como

20

$$\Delta L(A, B) = 10 \log_{10} \frac{\sum_{i=1}^N \sum_{j=1}^N g_i g_j C_{i,j} \sqrt{10^{(k_i+k_j)/10}}}{\sum_{i=1}^N \sum_{j=1}^N h_i h_j C_{i,j} \sqrt{10^{(k_i+k_j)/10}}},$$

o bien, en el caso de objetos no correlacionados

$$\Delta L(A, B) = 10 \log_{10} \frac{\sum_{i=1}^N g_i^2 10^{k_i/10}}{\sum_{i=1}^N h_i^2 10^{k_i/10}}.$$

5

[0152] A continuación, se considera un escenario de mejora de diálogo.

[0153] Considerando nuevamente el escenario de aplicación de la mejora de diálogo. La libertad de definir la información de renderización en el decodificador se limita solamente a cambiar los niveles de los dos metaobjetos. Supongamos además que los dos metaobjetos no están correlacionados, es decir, $C_{FGO, BGO} = 0$. Si los pesos de mezcla descendente de los metaobjetos son h_{FGO} y h_{BGO} , y se renderizan con las ganancias f_{FGO} y f_{BGO} , la sonoridad de la salida en relación con la mezcla por defecto es

$$\begin{aligned} \Delta L(A, B) &= 10 \log_{10} \frac{f_{FGO}^2 10^{(L_{FGO}-c)/10} + f_{BGO}^2 10^{(L_{BGO}-c)/10}}{h_{FGO}^2 10^{(L_{FGO}-c)/10} + h_{BGO}^2 10^{(L_{BGO}-c)/10}} \\ &= 10 \log_{10} \frac{f_{FGO}^2 10^{L_{FGO}/10} + f_{BGO}^2 10^{L_{BGO}/10}}{h_{FGO}^2 10^{L_{FGO}/10} + h_{BGO}^2 10^{L_{BGO}/10}}. \end{aligned}$$

15

[0154] Esto es también la compensación necesaria si se desea tener la misma sonoridad en la salida que en la mezcla descendente por defecto.

20 **[0155]** $\Delta L(A, B)$ puede ser considerado como un valor de compensación de sonoridad, que puede ser transmitido con el procesador de señales 120 del decodificador. $\Delta L(A, B)$ también se puede denominar como un valor de cambio de sonoridad y, por lo tanto, el valor de compensación real puede ser un valor inverso. Por lo tanto, el valor de compensación de sonoridad lc_v mencionado anteriormente en este documento correspondería al valor g_{Δ} a continuación.

25

[0156] Por ejemplo, $g_{\Delta} = 10^{-\Delta L(A,B)/20}$ o $g_{\Delta} = 1/\Delta L(A,B)$ puede aplicarse como un factor de multiplicación en cada canal de una señal de audio modificada que resulta de aplicar la información de renderización en la señal de entrada de audio. Esta ecuación para g_{Δ} funciona en el dominio lineal. En el dominio logarítmico, la ecuación sería diferente, como $1 / \Delta L(A, B)$ y aplicarse en consecuencia.

30

[0157] Si el proceso de mezcla descendente se simplifica de tal manera que los dos metaobjetos pueden mezclarse con pesos unitarios para obtener la señal de mezcla descendente, es decir, $h_{FGO} = h_{BGO} = 1$, y ahora las ganancias de renderización para estos dos objetos se indican con g_{FGO} y g_{BGO} , esto simplifica la ecuación del cambio de sonoridad en

35

$$\begin{aligned}\Delta L(A, B) &= 10 \log_{10} \frac{g_{FGO}^2 10^{(L_{FGO}-c)/10} + g_{BGO}^2 10^{(L_{BGO}-c)/10}}{10^{(L_{FGO}-c)/10} + 10^{(L_{BGO}-c)/10}} \\ &= 10 \log_{10} \frac{g_{FGO}^2 10^{L_{FGO}/10} + g_{BGO}^2 10^{L_{BGO}/10}}{10^{L_{FGO}/10} + 10^{L_{BGO}/10}}\end{aligned}$$

[0158] De nuevo $\Delta L(A, B)$ puede considerarse como un valor de compensación de sonoridad, que se determina con el procesador de señales 120.

5

[0159] En general, g_{FGO} se puede considerar como una ganancia de renderización para el objeto en primer plano FGO (grupo de objetos en primer plano) y g_{BGO} puede considerarse como una ganancia de renderización para el objeto de fondo BGO (grupo de objetos de fondo).

10

[0160] Como se mencionó anteriormente, es posible transmitir diferencias de sonoridad en lugar de sonoridad absoluta. Definamos la sonoridad de referencia como la sonoridad del metaobjeto FGO $L_{REF} = L_{FGO}$, es decir, $K_{FGO} = L_{FGO} - L_{REF} = 0$ y $K_{BGO} = L_{BGO} - L_{REF} = L_{BGO} - L_{FGO}$. Ahora, el cambio de sonoridad es

$$\Delta L(A, B) = 10 \log_{10} \frac{g_{FGO}^2 + g_{BGO}^2 10^{K_{BGO}/10}}{1 + 10^{K_{BGO}/10}}$$

15

[0161] También puede ser, como es el caso en el SAOC-DE, que dos metaobjetos no tengan factores individuales de escala, pero uno de los objetos se deja sin modificar, mientras que el otro se atenúa para obtener la proporción de mezcla correcta entre los objetos. En este ajuste de renderización, la salida será menor en sonoridad que la mezcla por defecto, y el cambio en la sonoridad es

20

$$\Delta L(A, B) = 10 \log_{10} \frac{\hat{g}_{FGO}^2 + \hat{g}_{BGO}^2 10^{K_{BGO}/10}}{1 + 10^{K_{BGO}/10}},$$

con

25

$$\hat{g}_{FGO} = \begin{cases} 1 & , \text{ if } g_{FGO} \geq g_{BGO} \\ \frac{g_{FGO}}{g_{BGO}} & , \text{ if } g_{FGO} < g_{BGO} \end{cases}, \text{ y } \hat{g}_{BGO} = \begin{cases} \frac{g_{BGO}}{g_{FGO}} & , \text{ if } g_{BGO} < g_{FGO} \\ 1 & , \text{ if } g_{BGO} \geq g_{FGO} \end{cases}.$$

[0162] Esta forma ya es bastante sencilla y es bastante agnóstica con respecto a la medida de sonoridad utilizada. El único requisito real es que los valores de sonoridad se deben sumar en el dominio exponencial. Es posible transmitir/almacenar valores de energías de señales en lugar de valores de sonoridad, ya que los dos tienen una conexión estrecha.

30

[0163] En cada una de las fórmulas anteriores, $\Delta L(A, B)$ puede ser considerado como un valor de compensación de sonoridad, que puede ser transmitido con el procesador de señales 120 del decodificador.

35

[0164] A continuación, se consideran ejemplos de casos. La exactitud de los conceptos proporcionados se ilustra a través de dos señales de ejemplo. Ambas señales tienen una mezcla descendente 5.1 con los canales envolvente y LFE anulados desde el procesamiento SAOC.

[0165] Se utilizan dos estrategias principales: una ("3-términos") con tres metaobjetos: FGO, BGO y canales anulados, por ejemplo,

$$\mathbf{X} = \mathbf{X}_{FGO} + \mathbf{X}_{BGO} + \mathbf{X}_{BYPASS} ,$$

5

[0166] Y otra ("2-términos") con dos metaobjetos, por ejemplo:

$$\mathbf{X} = \mathbf{X}_{FGO} + \mathbf{X}_{BGO} .$$

10 **[0167]** En la estrategia de 2 términos, los canales anulados pueden, por ejemplo, mezclarse junto con el BGO para la estimación de sonoridad de metaobjetos. Se estiman la sonoridad de ambos (o de los tres) objetos, así como la sonoridad de la señal de mezcla descendente y se almacenan los valores.

15 **[0168]** Las instrucciones de renderización son de forma

$$\mathbf{Y} = \hat{g}_{FGO} \mathbf{X}_{FGO} + \hat{g}_{BGO} \mathbf{X}_{BGO} + \hat{g}_{BGO} \mathbf{X}_{BYPASS}$$

y

20

$$\mathbf{Y} = \hat{g}_{FGO} \mathbf{X}_{FGO} + \hat{g}_{BGO} \mathbf{X}_{BGO}$$

para las dos estrategias respectivamente.

25 **[0169]** Los valores de ganancia son, por ejemplo, determinados según:

$$\hat{g}_{FGO} = \begin{cases} 1 & , \text{ si } g_{FGO} > 1 \\ g_{FGO} & , \text{ en caso contrario} \end{cases} \text{ y } \hat{g}_{BGO} = \begin{cases} 1/g_{FGO} & , \text{ si } g_{FGO} > 1 \\ 1 & , \text{ en caso contrario} \end{cases} ,$$

en el que la ganancia FGO g_{FGO} varía entre -24 y +24 dB.

30

[0170] Se renderiza el escenario de salida, se mide la sonoridad y se calcula la atenuación de la sonoridad de la señal de mezcla descendente.

[0171] Este resultado se muestra en la Fig. 10 y la Fig. 11 con la línea azul de marcadores de círculo. La Fig. 10 representa una primera ilustración y la Fig. 11 representa una segunda ilustración de un cambio medido de sonoridad y el resultado de utilizar los conceptos proporcionados para estimar el cambio de sonoridad de una manera puramente paramétrica,

[0172] A continuación, la atenuación por la mezcla descendente se calcula paraméricamente empleando los valores de sonoridad de metaobjetos almacenados y la información de renderización y mezcla descendente. La estimación utilizando la sonoridad de tres metaobjetos se ilustra con la línea verde de marcadores cuadrados y la estimación utilizando la sonoridad de dos metaobjetos se ilustra con la línea roja de marcadores de estrellas.

[0173] Se puede ver a partir de las figuras, que las estrategias de 2 y 3 términos proporcionan resultados prácticamente idénticos y ambos se aproximan bastante bien al valor medido.

[0174] Los conceptos proporcionados presentan una pluralidad de ventajas. Por ejemplo, los conceptos proporcionados permiten estimar la sonoridad de una señal de mezcla a partir de la sonoridad de las señales

componentes que forman la mezcla. La ventaja de esto es que la sonoridad de la señal del componente se puede estimar una vez y la estimación de sonoridad de la señal de la mezcla se puede obtener paramétricamente para cualquier mezcla sin la necesidad de una estimación de sonoridad real basada en la señal. Esto proporciona una mejora considerable en la eficacia computacional del sistema global en el que se necesita la estimación de sonoridad de varias mezclas. Por ejemplo, cuando el usuario final cambia los ajustes de renderización, la estimación de sonoridad de la salida está inmediatamente disponible.

[0175] En algunas aplicaciones, como por ejemplo cuando está conforme a la recomendación UER R128, es importante que la sonoridad media de todo el programa sea importante. Si la estimación de sonoridad en el receptor, por ejemplo, en un escenario de radiodifusión, se hace en base a la señal recibida, la estimación converge en la sonoridad media solo después de que se haya recibido todo el programa. Debido a esto, cualquier compensación de la sonoridad tendrá errores o presentará variaciones temporales. Cuando se estima la sonoridad de los objetos componentes como se propone y se transmite la información de sonoridad, es posible estimar la sonoridad media de la mezcla en el receptor sin retardo.

[0176] Si se desea que la sonoridad media de la señal de salida permanezca (aproximadamente) constante, independientemente de los cambios en la información de renderización, los conceptos proporcionados permiten determinar un factor de compensación por este motivo. Los cálculos necesarios para esto en el decodificador son, desde su complejidad computacional insignificantes, y, de este modo, es posible añadir la funcionalidad a cualquier decodificador.

[0177] Hay casos en los que el nivel de sonoridad absoluta de la salida no es importante, pero la importancia reside en determinar el cambio en la sonoridad a partir de una escena de referencia. En dichos casos los niveles absolutos de los objetos no son importantes, pero sus niveles relativos sí lo son. Esto permite definir uno de los objetos como el objeto de referencia y representar la sonoridad de los otros objetos en relación con la sonoridad de este objeto de referencia. Esto tiene algunas ventajas considerando el transporte y/o almacenamiento de la información de sonoridad.

[0178] En primer lugar, no es necesario transportar el nivel de sonoridad de referencia. En el caso de aplicación de dos metaobjetos, esto reduce a la mitad la cantidad de datos a transmitir. La segunda ventaja se refiere a la posible cuantización y representación de los valores de sonoridad. Puesto que los niveles absolutos de los objetos pueden ser casi cualquier cosa, los valores absolutos de sonoridad también pueden ser casi cualquier cosa. Por otro lado, se supone que los valores relativos de sonoridad tienen una media de 0 y una distribución bastante bien formada alrededor de la media. La diferencia entre las representaciones permite definir la cuadrícula de cuantización de la representación relativa de una manera que tiene una mayor exactitud posible con el mismo número de bits utilizados para la representación cuantizada.

[0179] La Fig. 12 ilustra otra realización para llevar a cabo una compensación de sonoridad. En la Fig. 12, se puede realizar la compensación de la sonoridad, por ejemplo, para compensar la pérdida de sonoridad. Para este propósito, por ejemplo, se pueden utilizar los valores **DE_loudness_diff_dialogue** (= K_{FGO}) y **DE_loudness_diff_background** (= K_{BGO}) desde **DE_control_info**. En este punto, **DE_control_info** puede concretar la información de control de "Mejora de diálogo" (DE) con Advanced Clean Audio

[0180] La compensación de sonoridad se consigue aplicando un valor de ganancia "g" en la señal de salida SAOC-DE y en los canales anulados (en el caso de una señal multicanal).

[0181] En la realización de la Fig. 12, esto se hace de la siguiente manera:

Se utiliza un valor de ganancia de modificación de diálogo limitado m_G para determinar las ganancias eficaces para el objeto en primer plano (FGO, por ejemplo, diálogo) y para el objeto de fondo (BGO, por ejemplo, ambiente). Esto se hace mediante el bloque "mapeo de ganancia" 1220 que produce los valores de ganancia m_{FGO} y m_{BGO} .

[0182] El bloque "estimador de sonoridad de salida" 1230 utiliza la información de sonoridad K_{FGO} y K_{BGO} , Y los valores de ganancia eficaces m_{FGO} y m_{BGO} para estimar este posible cambio en la sonoridad en comparación con el caso de mezcla descendente por defecto. El cambio se mapea a continuación en el "factor de compensación de sonoridad" que se aplica en los canales de salida para producir las "señales de salida" finales.

[0183] Se aplican las siguientes etapas para la compensación de sonoridad:

- Recibir del valor de ganancia limitada m_G desde el decodificador SAOC-DE (como se define en la cláusula 12.8 "Modificación del intervalo de control para SAOC-DE" [DE]) y determinar las ganancias FGO/BGO aplicadas:

$$\begin{cases} m_{FGO} = m_G, \text{ y } m_{BGO} = 1 & \text{if } m_G \leq 1 \\ m_{FGO} = 1, \text{ y } m_{BGO} = m_G^{-1} & \text{if } m_G > 1 \end{cases}$$

5

- Obtener la información de sonoridad de metaobjetos K_{FGO} y K_{BGO} .
 - Calcular el cambio en la sonoridad de salida en comparación con la mezcla descendente por defecto con

$$\Delta L = 10 \log_{10} \frac{m_{FGO}^2 10^{K_{FGO}/10} + m_{BGO}^2 10^{K_{BGO}/10}}{10^{K_{FGO}/10} + 10^{K_{BGO}/10}}$$

10

- Calcular la ganancia de compensación de sonoridad $g_\Delta = 10^{0,05\Delta L}$.

$$\mathbf{g} = \begin{bmatrix} g_1 \\ \vdots \\ g_N \end{bmatrix},$$

- Calcular los factores de escala en los que

$$g_i = \begin{cases} g_\Delta & \text{si el canal } i \text{ pertenece a la salida de SAOC-DE} \\ m_{BGO} g_\Delta & \text{si el canal } i \text{ es un canal de anuladas} \end{cases}$$

y N es el número total de canales de

15 salida.

En la Fig. 12, el ajuste de ganancia se divide en dos etapas: la ganancia de los posibles "canales anulados" se ajusta con m_{BGO} antes de combinarlos con los "canales de salida SAOC-DE" y luego posteriormente una ganancia común g_Δ se aplica posteriormente en todos los canales combinados. Esto es solo una posible reordenación de las operaciones de ajuste de ganancia, mientras que g combina ambas etapas de ajuste de ganancia en un ajuste de ganancia.

20

- Aplicar los valores de escala g en los canales de audio \mathbf{Y}_{FULL} que consiste en los "canales de salida SAOC-DE" \mathbf{Y}_{SAOC} y los posibles "canales anulados" alineados en el tiempo \mathbf{Y}_{BYPASS} : $\mathbf{Y}_{FULL} = \mathbf{Y}_{SAOC} \cup \mathbf{Y}_{BYPASS}$

25

[0184] La aplicación de los valores de escala g en los canales de audio \mathbf{Y}_{FULL} se lleva a cabo con la unidad de ajuste de ganancia 1240.

[0185] ΔL , tal como se ha calculado anteriormente, puede considerarse como un valor de compensación de sonoridad. En general, m_{FGO} indica una ganancia de renderización para el objeto en primer plano FGO (grupo de objetos en primer plano) y m_{BGO} indica una ganancia de renderización para el objeto de fondo BGO (grupo de objetos de fondo).

30

[0186] Aunque se han descrito algunos aspectos en el contexto de un aparato, es evidente que estos

aspectos también representan una descripción del procedimiento correspondiente, donde un bloque o dispositivo se corresponde con una etapa del procedimiento o una característica de una etapa del procedimiento. De forma análoga los aspectos que se describen en el contexto de una etapa del procedimiento también representan una descripción de un bloque correspondiente o un elemento o característica del aparato correspondiente.

5

[0187] La señal descompuesta de la invención se puede almacenar en un medio de almacenamiento digital o se puede transmitir con un medio de transmisión tal como un medio de transmisión inalámbrico o un medio de transmisión por cable como Internet.

10 **[0188]** En función de ciertos requisitos de implementación, las realizaciones de la invención se pueden implementar en hardware o en software. La implementación puede realizarse utilizando un medio de almacenamiento digital, por ejemplo un disquete, un DVD, un CD, una ROM, una PROM, una EPROM, una EEPROM o una memoria FLASH, que tiene señales de control legibles electrónicamente y almacenadas en el mismo, que coopera (o es capaz de cooperar) con un sistema informático programable de manera que se lleve a
15 cabo el procedimiento respectivo.

[0189] De acuerdo con la invención algunas realizaciones comprenden un soporte de datos no transitorio que tiene señales de control legibles electrónicamente y que son capaces de cooperar con un sistema informático programable, de manera que se lleve a cabo uno de los procedimientos descritos en este documento.

20

[0190] En general las realizaciones de la presente invención se pueden implementar como un producto de programa informático con un código de programa, siendo el código de programa operativo para llevar a cabo uno de los procedimientos cuando el producto de programa informático se ejecuta en un ordenador. El código de programa se puede almacenar, por ejemplo, en un soporte legible por máquina.

25

[0191] Otras realizaciones comprenden el programa de ordenador para llevar a cabo uno de los procedimientos descritos en este documento almacenado en un soporte legible por máquina.

30 **[0192]** En otras palabras, una realización del procedimiento de la invención es, por lo tanto, un programa informático que tiene un código de programa para realizar uno de los procedimientos descritos en este documento cuando el programa informático se ejecuta en un ordenador.

35 **[0193]** Una realización adicional de los procedimientos de la invención es, por lo tanto, un soporte de datos (o un medio de almacenamiento digital o un medio legible por ordenador) que comprende, grabado en el mismo, el programa informático para llevar a cabo uno de los procedimientos descritos en este documento.

40 **[0194]** Una realización adicional del procedimiento de la invención es, por lo tanto, un flujo de datos o una secuencia de señales que representan el programa informático para llevar a cabo uno de los procedimientos descritos en este documento. El flujo de datos o la secuencia de señales pueden, por ejemplo, estar configurados para ser transferidos a través de una conexión de comunicación de datos, por ejemplo, a través de Internet.

45 **[0195]** Una realización adicional comprende además un medio de procesamiento, por ejemplo, un ordenador o un dispositivo lógico programable, configurado o adaptado para llevar a cabo uno de los procedimientos descritos en este documento.

[0196] Una realización adicional comprende un ordenador que tiene instalado en el mismo el programa informático para llevar a cabo uno de los procedimientos descritos en este documento.

50 **[0197]** En algunas realizaciones se puede utilizar un dispositivo lógico programable (por ejemplo, un campo de matriz de puertas programables) para llevar a cabo algunas o todas las funcionalidades de los procedimientos descritos en el este documento. En algunas realizaciones una matriz de puertas programables por campo podrá cooperar con un microprocesador a fin de llevar a cabo uno de los procedimientos descritos en este documento. En general los procedimientos se llevan a cabo, preferentemente, mediante cualquier aparato de hardware.

55

[0198] Las realizaciones anteriormente descritas son meramente ilustrativas de los principios de la presente invención. Se entiende que las modificaciones y variaciones de las disposiciones y los detalles descritos en este documento serán evidentes para otros expertos en la técnica. Es la intención, por lo tanto, de estar limitado sólo por el alcance de las reivindicaciones de patentes inminentes y no por los detalles específicos presentados a modo de

descripción y las explicaciones de las realizaciones de este documento.

Referencias

5 **[0199]**

[BCC] C. Faller y F. Baumgarte, "Binaural Cue Coding - Part II: Schemes and applications," IEEE Trans. on Speech and Audio Proc., Vol. 11, no. 6 de noviembre de 2003.

10 [UER] Recomendación R128 de la UER "Normalización de la sonoridad y nivel máximo permitido de las señales de audio", Ginebra, 2011.

[JSC] C. Faller, "Parametric Joint-Coding of Audio Sources", 120ª Convención AES, París, 2006.

15 [ISS1] M. Parvaix y L. Girin: "Informed Source Separation of underdetermined instantaneous Stereo Mixtures using Source Index Embedding", IEEE ICASSP, 2010.

[ISS2] M. Parvaix, L. Girin, J.-M. Brossier: "A watermarking-based method for informed source separation of audio signals with a single sensor", IEEE Transactions on Audio, Speech and Language Processing, 2010.

20

[ISS3] A. Liutkus y J. Pinel y R. Badeau y L. Girin y G. Richard: "Informed source separation through spectrogram coding and data embedding", Signal Processing Journal, 2011.

25

[ISS4] A. Ozerov, A. Liutkus, R. Badeau, G. Richard: "Informed source separation: source coding meets source separation", IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2011.

[ISS5] S. Zhang y L. Girin: "An Informed Source Separation System for Speech Signals", INTERSPEECH, 2011.

30 [ISS6] L. Girin y J. Pinel: "Informed Audio Source Separation from Compressed Linear Stereo Mixtures", AES 42nd International Conference: Semantic Audio, 2011.

[UIT] Unión Internacional de Telecomunicaciones: "Recommendation ITU-R BS.1770-3 - Algorithms to measure audio programme loudness and true-peak audio level", Ginebra, 2012.

35

[SAOC1] J. Herre, S. Disch, J. Hilpert, O. Hellmuth: "From SAC To SAOC - Recent Developments in Parametric Coding of Spatial Audio", 22nd Regional UK AES Conference, Cambridge, Reino Unido, Abril de 2007.

40 [SAOC2] J. Engdegard, B. Resch, C. Falch, O. Hellmuth, J. Hilpert, A. Holzer, L. Terentiev, J. Breebaart, J. Koppens, E. Schuijers y W. Oomen: "Spatial Audio Object Coding (SAOC) - The Upcoming MPEG Standard on Parametric Object Based Audio Coding", 124th AES Convention, Amsterdam 2008.

[SAOC] ISO/IEC, "MPEG audio technologies - Part 2: Spatial Audio Object Coding (SAOC)," ISO/IEC JTC1/SC29/WG11 (MPEG) International Standard 23003-2.

45

[EP] EP 2146522 A1: S. Schreiner, W. Fiesel, M. Neusinger, O. Helimuth, R. Sperschneider, "Apparatus and method for generating audio output signals using object based metadata", 2010.

50 [DE] ISO/IEC, "MPEG audio technologies - Part 2: Spatial Audio Object Coding (SAOC) - Amendment 3, Dialogue Enhancement," ISO/IEC 23003-2:2010/DAM 3, Dialogue Enhancement.

[BRE] WO 2008/035275 A2. [SCH] EP 2 146 522 A1. [ES] WO 2008/046531 A1.

REIVINDICACIONES

1. Un decodificador para generar una señal de salida de audio que comprende uno o más canales de salida de audio, en el que el decodificador comprende:

5 una interfaz de recepción (110) que recibe una señal de entrada de audio que comprende una pluralidad de señales de objetos de audio, que recibe información de sonoridad sobre las señales de objetos de audio y que recibe información de renderización que indica si una o más de las señales de objetos de audio serán amplificadas o atenuadas, y

10 procesador de señales (120) que genera uno, o más, canales de salida de audio de la señal de salida de audio.

en el que la interfaz de recepción (110) está configurada para recibir una señal de mezcla descendente que comprende uno o más canales de mezcla descendente como la señal de entrada de audio, en el que el uno o más canales de mezcla descendente comprenden las señales de objetos de audio y en el que el número del uno o más canales de mezcla descendente es menor que el número de las señales de objetos de audio,

en el que la interfaz de recepción (110) está configurada para recibir información de mezcla descendente que indica cómo se mezclan las señales de objetos de audio dentro del uno o más canales de mezcla descendentes, en el que la interfaz de recepción (110) está configurada para recibir uno o más señales de objetos de audio anuladas adicionales, en el que la o las señales de objetos de audio anuladas adicionales no se mezclan dentro de la señal de mezcla descendente,

en el que la interfaz de recepción (110) está configurada para recibir la información de sonoridad que indica información sobre la sonoridad de las señales de objetos de audio que se mezclan dentro de la señal de mezcla descendente e indica información sobre la sonoridad de una o más señales de objetos de audio anuladas adicionalmente que no se mezclan dentro de la señal de mezcla descendente,

en el que el procesador de señales (120) está configurado para determinar un valor de compensación de sonoridad en función de la información sobre la sonoridad de las señales de objetos de audio que se mezclan dentro de la señal de mezcla descendente y en función de la información sobre la sonoridad de una o más señales de objetos de audio anuladas adicionalmente que no se mezclan dentro de la señal de mezcla descendente, y

en el que el procesador de señales (120) está configurado para generar uno o más canales de salida de audio de la señal de salida de audio a partir de la señal de entrada de audio en función de la información de renderización y en función del valor de compensación de sonoridad,

2. Un decodificador según la reivindicación 1, en el que el procesador de señales (120) está configurado para generar uno o más canales de salida de audio de la señal de salida de audio a partir de la señal de entrada de audio en función de la información de renderización y en función del valor de compensación de sonoridad, de manera que una sonoridad de la señal de salida de audio es igual a una sonoridad de la señal de entrada de audio, o de manera que la sonoridad de la señal de salida de audio está más próxima a la sonoridad de la señal de entrada de audio que una sonoridad de una señal de audio modificada que resultaría de modificar la señal de entrada de audio amplificando o atenuando las señales de objetos de audio de la señal de entrada de audio según la información de renderización.

3. Un decodificador de audio según la reivindicación 2 y

en el que el procesador de señales (120) está configurado para generar la señal de audio modificada modificando la señal de entrada de audio amplificando o atenuando las señales de objetos de audio de la señal de entrada de audio según la información de renderización, y

en el que el procesador de señales (120) está configurado para generar la señal de salida de audio aplicando el valor de compensación de sonoridad en la señal de audio modificada, de manera que la sonoridad de la señal de salida de audio sea igual a la sonoridad de la señal de entrada de audio, o de manera que la sonoridad de la señal de salida de audio está más próxima a la sonoridad de la señal de entrada de audio que la sonoridad de la señal de audio modificada.

4. Un decodificador según cualquiera de las reivindicaciones anteriores,

en el que cada una de las señales de objetos de audio de la señal de entrada de audio está asignada exactamente a un grupo de dos o más grupos, en el que cada uno de los dos o más grupos comprende una o más de las señales de objetos de audio de la señal de entrada de audio,

5 en el que la interfaz de recepción (110) está configurada para recibir un valor de sonoridad para cada grupo de los dos o más grupos como la información de sonoridad,

en el que el procesador de señales (120) está configurado para determinar el valor de compensación de sonoridad en función del valor de sonoridad de cada uno de los dos o más grupos, y

10

en el que el procesador de señales (120) está configurado para generar uno o más canales de salida de audio de la señal de salida de audio a partir de la señal de entrada de audio en función del valor de compensación de sonoridad.

5. Un decodificador según cualquiera de las reivindicaciones precedentes, en el que al menos un grupo
15 de los dos o más grupos comprende dos o más de las señales de objetos de audio.

6. Un decodificador según de las reivindicaciones anteriores,

en el que cada una de las señales de objetos de audio de la señal de entrada de audio está asignada exactamente a
20 un grupo de exactamente dos grupos como los dos o más grupos,

en el que cada una de las señales de objetos de audio de la señal de entrada de audio se asigna a un grupo de objetos en primer plano de exactamente dos grupos o a un grupo de objetos de fondo de los exactamente dos
25 grupos,

25

en el que la interfaz de recepción (110) está configurada para recibir el valor de sonoridad del grupo de objetos en primer plano,

en el que la interfaz de recepción (110) está configurada para recibir el valor de sonoridad del grupo de objetos de
30 fondo,

en el que el procesador de señales (120) está configurado para determinar el valor de compensación de sonoridad en función del valor de sonoridad del grupo de objetos en primer plano, y en función del valor de sonoridad del grupo de objetos de fondo, y
35

en el que el procesador de señales (120) está configurado para generar uno o más canales de salida de audio de la señal de salida de audio a partir de la señal de entrada de audio en función del valor de compensación de sonoridad.

7. Un decodificador según la reivindicación 6
40

en el que el procesador de señales (120) está configurado para determinar un valor de cambio de sonoridad ΔL de acuerdo con la fórmula

$$\Delta L = 10 \log_{10} \frac{m_{FGO}^2 10^{K_{FGO}/10} + m_{BGO}^2 10^{K_{BGO}/10}}{10^{K_{FGO}/10} + 10^{K_{BGO}/10}}$$

45

en el que K_{FGO} indica el valor de sonoridad del grupo de objetos en primer plano,

en el que K_{BGO} indica el valor de sonoridad del grupo de objetos de fondo,

en el que m_{FGO} indica una ganancia de renderización del grupo de objetos en primer plano, y

en el que m_{BGO} indica una ganancia de renderización del grupo de objetos de fondo.

50

8. Un decodificador según la reivindicación 6 y

en el que el procesador de señales (120) está configurado para determinar un valor de compensación de sonoridad ΔL según la fórmula

$$\Delta L(A, B) = 10 \log_{10} \frac{g_{FGO}^2 10^{L_{FGO}/10} + g_{BGO}^2 10^{L_{BGO}/10}}{10^{L_{FGO}/10} + 10^{L_{BGO}/10}}$$

- en el que L_{FGO} indica el valor de sonoridad del grupo de objetos en primer plano,
 5 en el que L_{BGO} indica el valor de sonoridad del grupo de objetos de fondo,
 en el que g_{FGO} indica una ganancia de renderización del grupo de objetos en primer plano, y
 en el que g_{BGO} indica una ganancia de renderización del grupo de objetos de fondo.
9. Un codificador que comprende:
- 10 una unidad codificadora basada en objetos (210; 710) que codifica una pluralidad de señales de objetos de audio para obtener una señal de audio codificada que comprende la pluralidad de señales de objetos de audio y
- una unidad codificadora de sonoridad de objetos (220; 720; 820) que codifica información de sonoridad en las
 15 señales de objetos de audio, en la que la información de sonoridad comprende uno o más valores de sonoridad, en la que cada uno, o más, de los valores de sonoridad depende de una, o más, de las señales de objetos de audio,
- en la que la unidad codificadora basada en objetos (210; 710) está configurada para recibir las señales de objetos de audio, en la que cada una de las señales de objetos de audio está asignada exactamente a uno de dos o más
 20 grupos, en la que cada uno de los dos o más grupos comprende una o más de las señales de objetos de audio,
- en la que la unidad codificadora basada en objetos (210; 710) está configurada para mezclar de forma descendente las señales de objetos de audio, comprendidas en los dos o más grupos, para obtener una señal de mezcla descendente que comprende uno o más canales de audio de mezcla descendente como la señal de audio
 25 codificada, en la que el número de uno o más canales de mezcla descendente es menor que el número de las señales de objetos de audio que están comprendidas en los dos o más grupos,
- en la que la unidad codificadora de sonoridad de objetos (220; 720; 820) está asignada para recibir una o más señales de objetos de audio anuladas adicionales, en la que cada una de una o más señales de objetos de audio
 30 anuladas adicionales está asignada a un tercer grupo, en la que cada una de la o las señales de objetos de audio anuladas adicionales no está comprendida en el primer grupo y no está comprendida en el segundo grupo, en la que la unidad codificadora basada en objetos (210; 710) está configurada para no mezclar de forma descendente la o las señales de objeto anuladas adicionales dentro de la señal de mezcla descendente, y
- 35 en la que la unidad codificadora de sonoridad de objetos (220; 720; 820) está configurada para determinar un primer valor de sonoridad, un segundo valor de sonoridad y un tercer valor de sonoridad de la información de sonoridad, el primer valor de sonoridad que indica una sonoridad total de una o más señales de objetos de audio del primer grupo, el segundo valor de sonoridad que indica una sonoridad total de una o más señales de objetos de audio del segundo grupo, y el tercer valor de sonoridad que indica una sonoridad total de una o más señales de objetos de audio
 40 anuladas adicionales del tercer grupo, o bien está configurada para determinar un primer valor de sonoridad y un segundo valor de sonoridad de la información de sonoridad, el primer valor de sonoridad que indica una sonoridad total de una o más señales de objetos de audio del primer grupo y el segundo valor de sonoridad que indica una sonoridad total de una o más señales de objetos de audio del segundo grupo y de una o más señales de objetos de audio anuladas adicionales del tercer grupo.
- 45
10. Un codificador según la reivindicación 9,
- en el que los dos o más grupos son exactamente dos grupos,
- 50 en el que cada una de las señales de objetos de audio está asignada exactamente a uno de los exactamente dos grupos, en el que cada uno de los exactamente dos grupos comprende una o más de las señales de objetos de audio,
- en el que la unidad codificadora basada en objetos (210; 710) está configurada para mezclar de forma descendente
 55 las señales de objetos de audio, comprendidas en los exactamente dos grupos, para obtener una señal de mezcla

descendente que comprende uno o más canales de audio de mezcla descendente como la señal de audio codificada, en el que el número de uno o más canales de mezcla descendente es menor que el número de las señales de objetos de audio que están comprendidas en los exactamente dos grupos.

5 11. Un sistema que comprende:

un codificador (310) según la reivindicación 9 o 10 que codifica una pluralidad de señales de objetos de audio para obtener una señal de audio codificada que comprende la pluralidad de señales de objetos de audio, y

10 un decodificador (320) según una de las reivindicaciones 1 a 8 que genera una señal de salida de audio que comprende uno o más canales de salida de audio,

en el que el decodificador (320) está configurado para recibir la señal de audio codificada como una señal de entrada de audio y para recibir la información de sonoridad

15 en el que el decodificador (320) está configurado para recibir información de renderización adicional,

en el que el decodificador (320) está configurado para determinar un valor de compensación de sonoridad en función de la información de sonoridad y en función de la información de renderización, y

20 en el que el decodificador (320) está configurado para generar uno o más canales de salida de audio de la señal de salida de audio a partir de la señal de entrada de audio en función de la información de renderización y en función del valor de compensación de sonoridad.

25 12. Un procedimiento para generar una señal de salida de audio que comprende uno o más canales de salida de audio, en el que el procedimiento comprende:

recibir una señal de entrada de audio que comprende una pluralidad de señales de objetos de audio,

30 recibir información de sonoridad que indica información sobre la sonoridad de las señales de objetos de audio que se mezclan dentro de la señal de mezcla descendente e información que indica la sonoridad de una o más señales de objetos de audio anuladas adicionales que no se mezclan dentro de la señal de mezcla descendente, y

35 recibir información de renderización que indica si una o más de las señales de objetos de audio se amplificarán o atenuarán,

40 recibir una señal de mezcla descendente que comprende uno o más canales de mezcla descendente como la señal de entrada de audio, en el que uno o más canales de mezcla descendente comprenden las señales de objetos de audio y en el que el número de uno o más canales de mezcla descendente es menor que el número de las señales de objetos de audio ,

recibir información de mezcla descendente que indica cómo se mezclan las señales de objetos de audio dentro del uno o más canales de mezcla descendente,

45 recibir una o más señales de objetos de audio anuladas adicionales, en el que la o las señales de objetos de audio anuladas adicionales no se mezclan dentro de la señal de mezcla descendente,

50 determinar un valor de compensación de sonoridad en función de la información sobre la sonoridad de las señales de objetos de audio que se mezclan dentro de la señal de mezcla descendente y en función de la información sobre la sonoridad de la o las señales de objetos de audio anuladas que no se mezclan dentro de la señal de mezcla descendente y generar uno o más canales de salida de audio de la señal de salida de audio a partir de la señal de entrada de audio en función de la información de mezcla descendente, en función de la información de renderización y en función del valor de compensación de sonoridad.

55 13. Un procedimiento para codificar que comprende:

codificar una señal de entrada de audio que comprende una pluralidad de señales de objetos de audio, y

codificar la información de sonoridad en las señales de objetos de audio, en el que la información de sonoridad

comprende uno o más valores de sonoridad, en el que cada uno, o más, de los valores de sonoridad depende de una, o más, de las señales de objetos de audio,

5 en el que cada una de las señales de objetos de audio está asignada exactamente a uno de dos o más grupos, en el que cada uno de los dos o más grupos comprende una o más de las señales de objetos de audio,

en el que la codificación de la información de sonoridad sobre las señales de objetos de audio se lleva a cabo mezclando de forma descendente las señales de objetos de audio, comprendidas en los dos o más grupos, para obtener una señal de mezcla descendente que comprende uno o más canales de audio mixtos como la señal de audio codificada, en el que el número de uno o más canales de mezcla descendente

10 es menor que el número de las señales de objetos de audio que están comprendidas en los dos o más grupos,

en el que cada una de una o más señales de objetos de audio anuladas adicionales está asignada a un tercer grupo, 15 en el que cada una de la o las señales de objetos de audio anuladas adicionales no está comprendida en el primer grupo y no está comprendida en el segundo grupo,

en el que la codificación de la información de sonoridad sobre las señales de objetos de audio se lleva a cabo no mezclando de forma descendente la o las señales de objetos de audio anuladas adicionales dentro de la señal de 20 mezcla descendente, y

en el que la codificación de la información de sonoridad sobre las señales de objetos de audio se realiza determinando un primer valor de sonoridad, un segundo valor de sonoridad y un tercer valor de sonoridad de la información de sonoridad, el primer valor de sonoridad que indica una sonoridad total de la o las señales de objetos de audio del primer grupo, el segundo valor de sonoridad que indica una sonoridad total de la o las señales de objetos de audio del segundo grupo, y el tercer valor de sonoridad que indica una sonoridad total de la o las señales de objetos de audio anuladas adicionales del tercer grupo, o bien está configurada para determinar un primer valor de sonoridad y un segundo valor de sonoridad de la información de sonoridad, el primer valor de sonoridad que indica una sonoridad total de la o las señales de objetos de audio del primer grupo y el segundo valor de sonoridad que indica una sonoridad total de una o más señales de objetos de audio del segundo grupo y de la una o más 30 señales de objetos de audio anuladas adicionales del tercer grupo.

14. Un procedimiento según la reivindicación 13,

35 en el que los dos o más grupos son exactamente dos grupos,

en el que cada una de las señales de objetos de audio está asignada exactamente a uno de los exactamente dos grupos, en el que cada uno de los exactamente dos grupos comprende una o más de las señales de objetos de audio,

40 en el que la codificación de la información de sonoridad sobre las señales de objetos de audio se lleva a cabo mezclando de forma descendente las señales de objetos de audio, comprendidas en los exactamente dos grupos, para obtener una señal de mezcla descendente que comprende uno o más canales de audio de mezcla descendente como la señal de audio codificada, en el que el uno o más canales de mezcla descendente es menor que el número de las señales de objetos de audio que están comprendidas exactamente en los dos grupos.

15. Un programa informático para implementar el procedimiento de las reivindicaciones 12 a 14 cuando se ejecuta en un ordenador o procesador de señales.

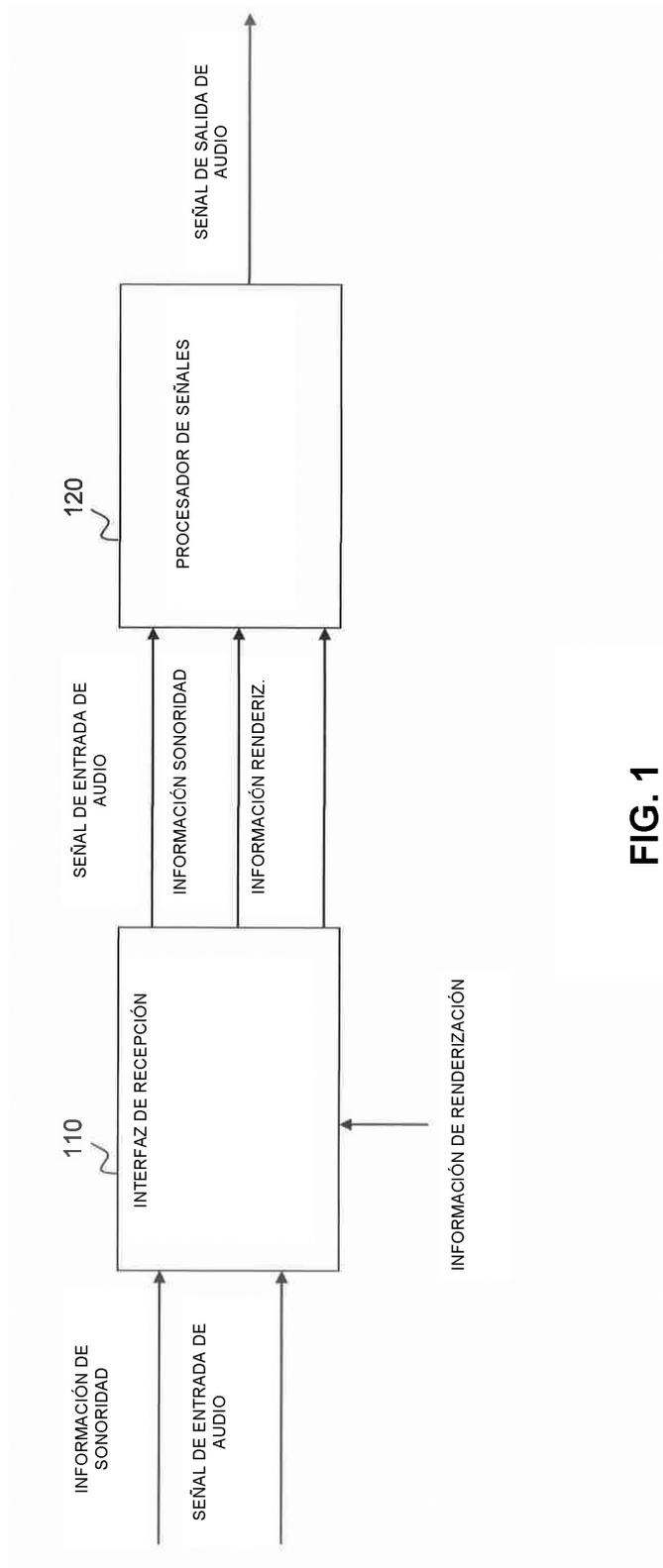
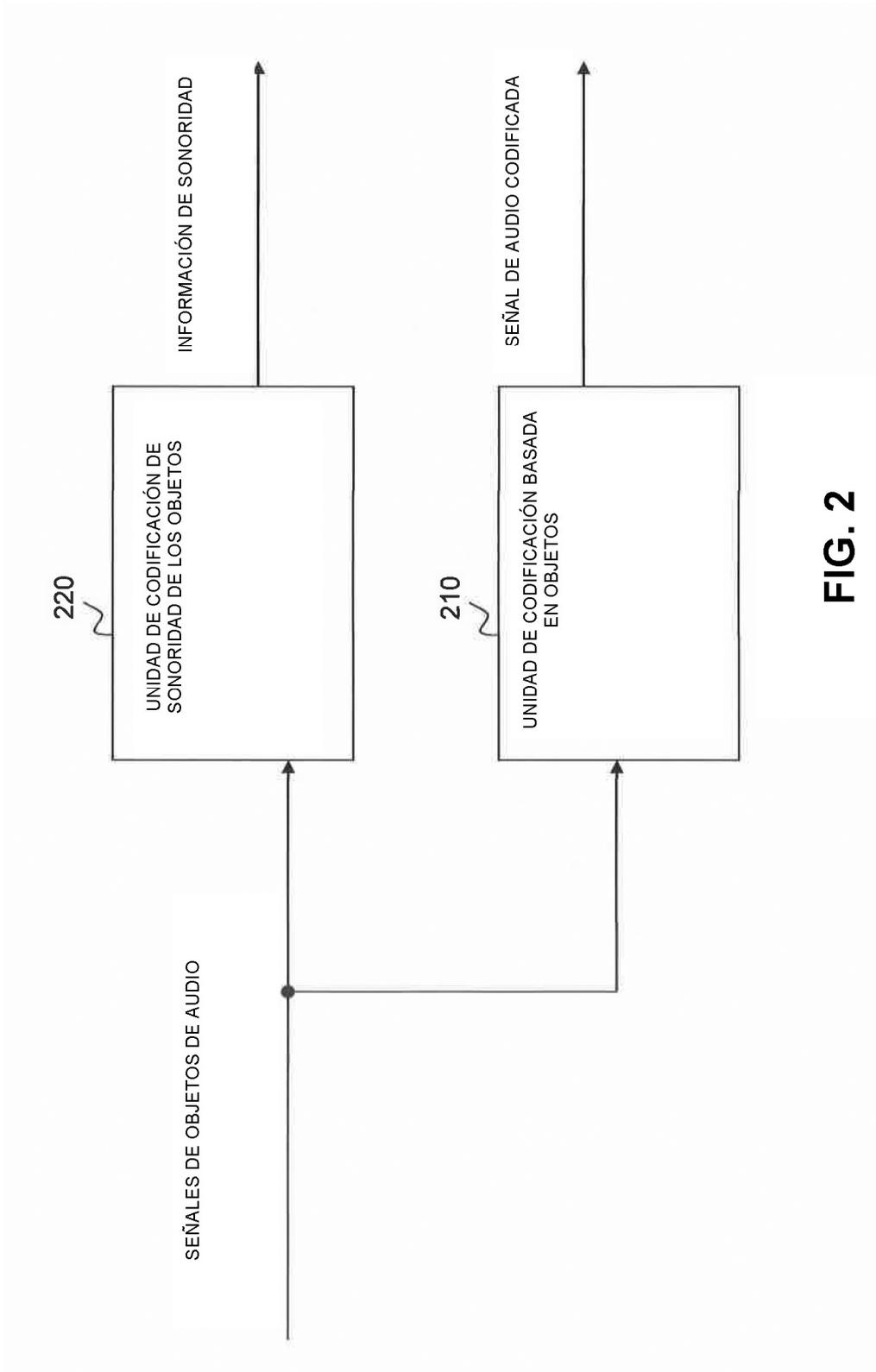


FIG. 1



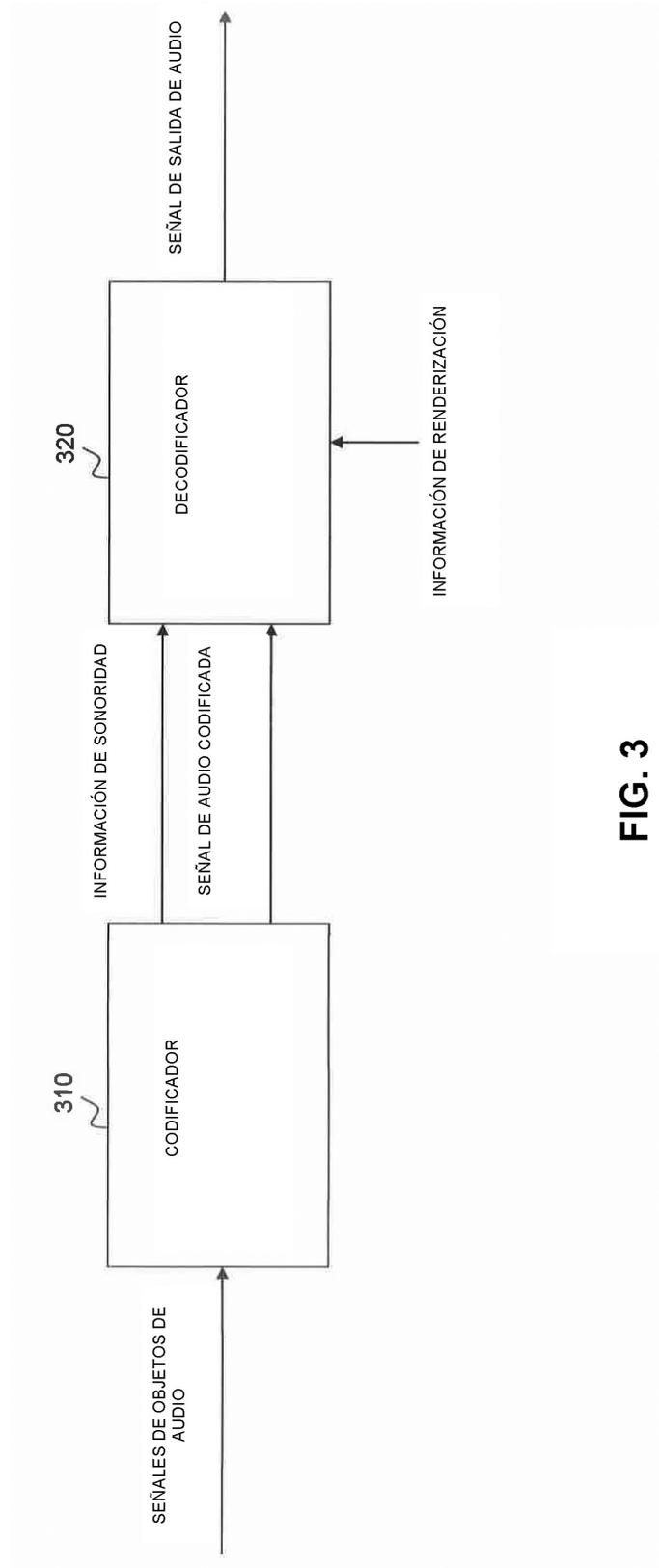


FIG. 3

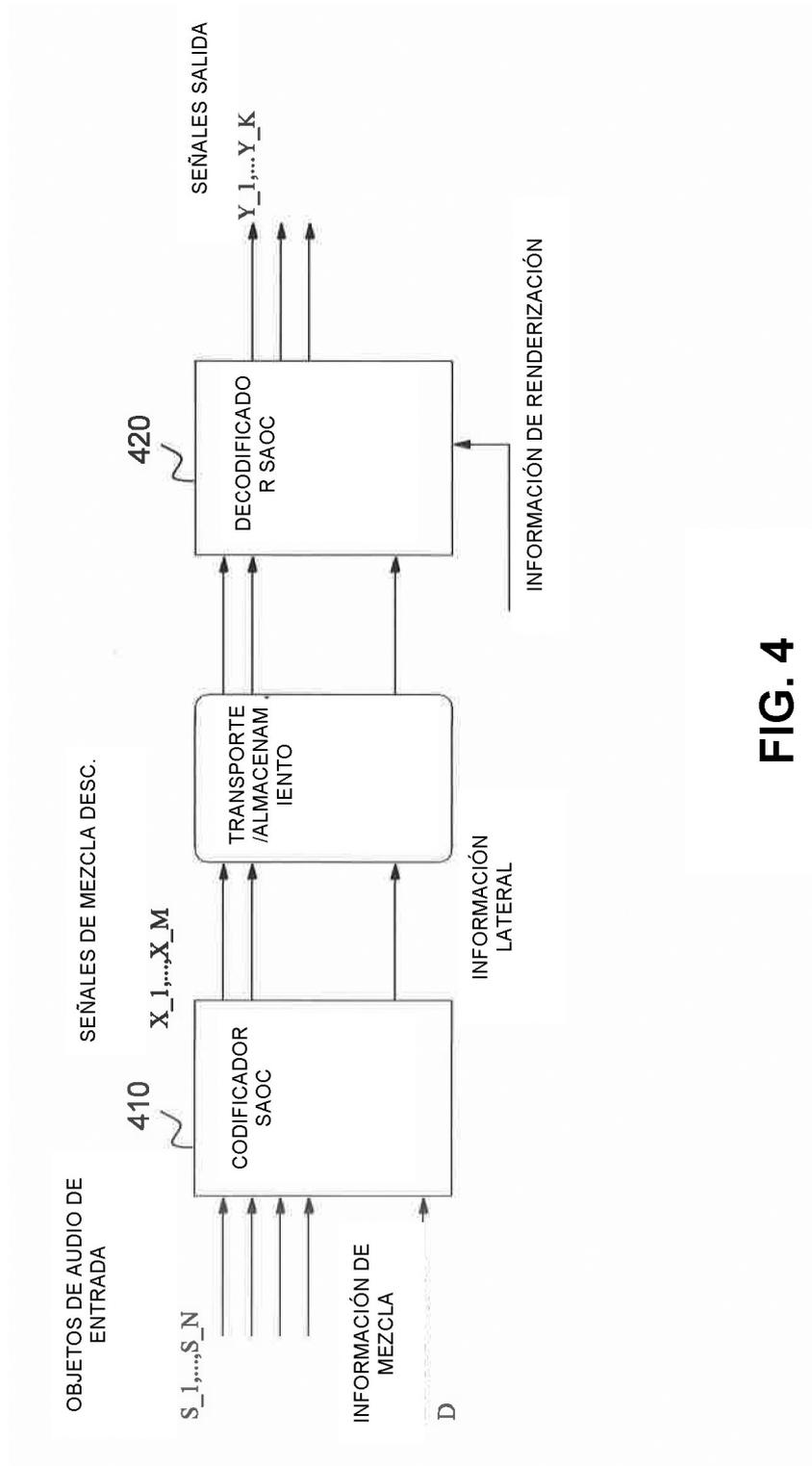


FIG. 4

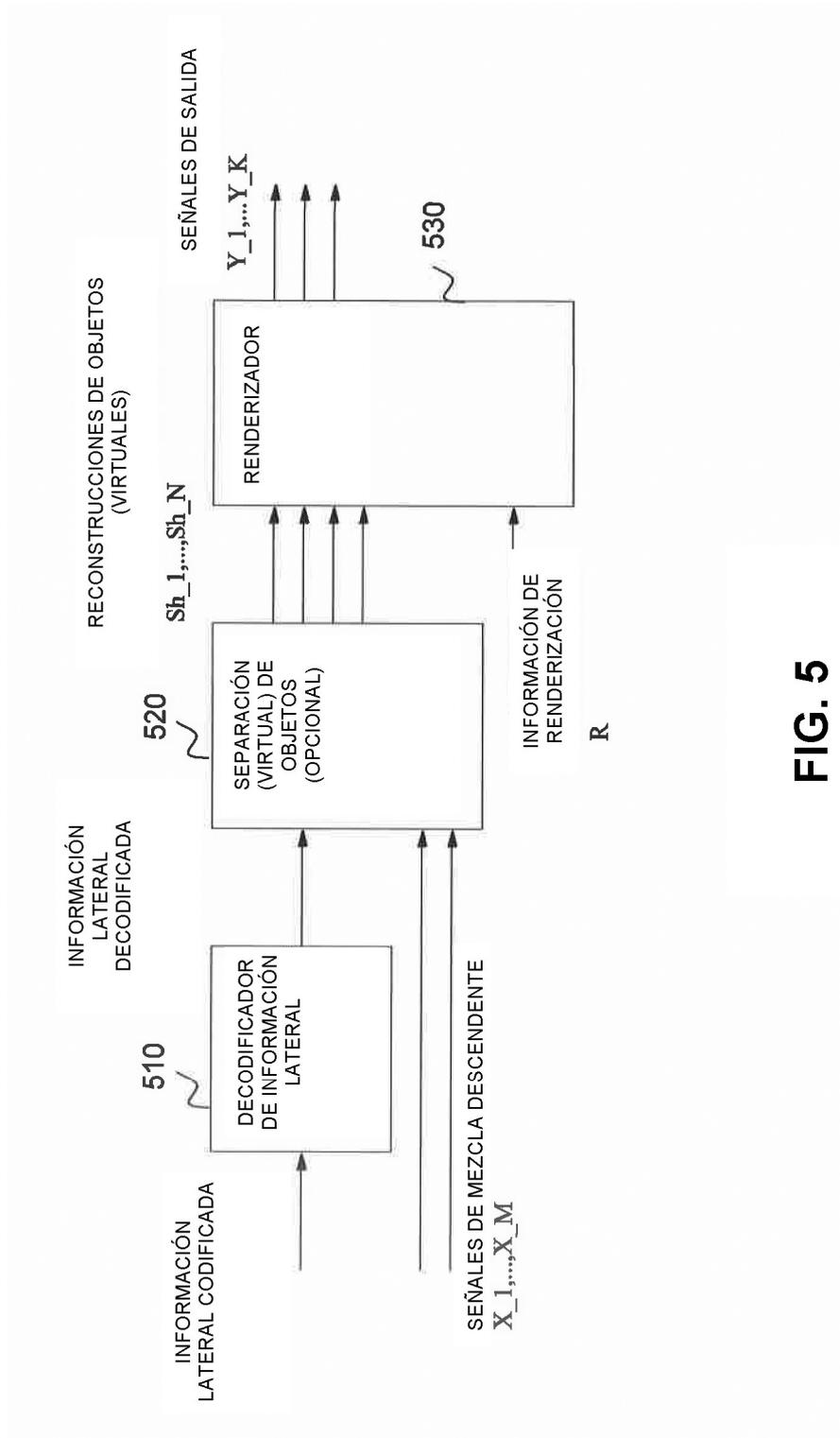


FIG. 5

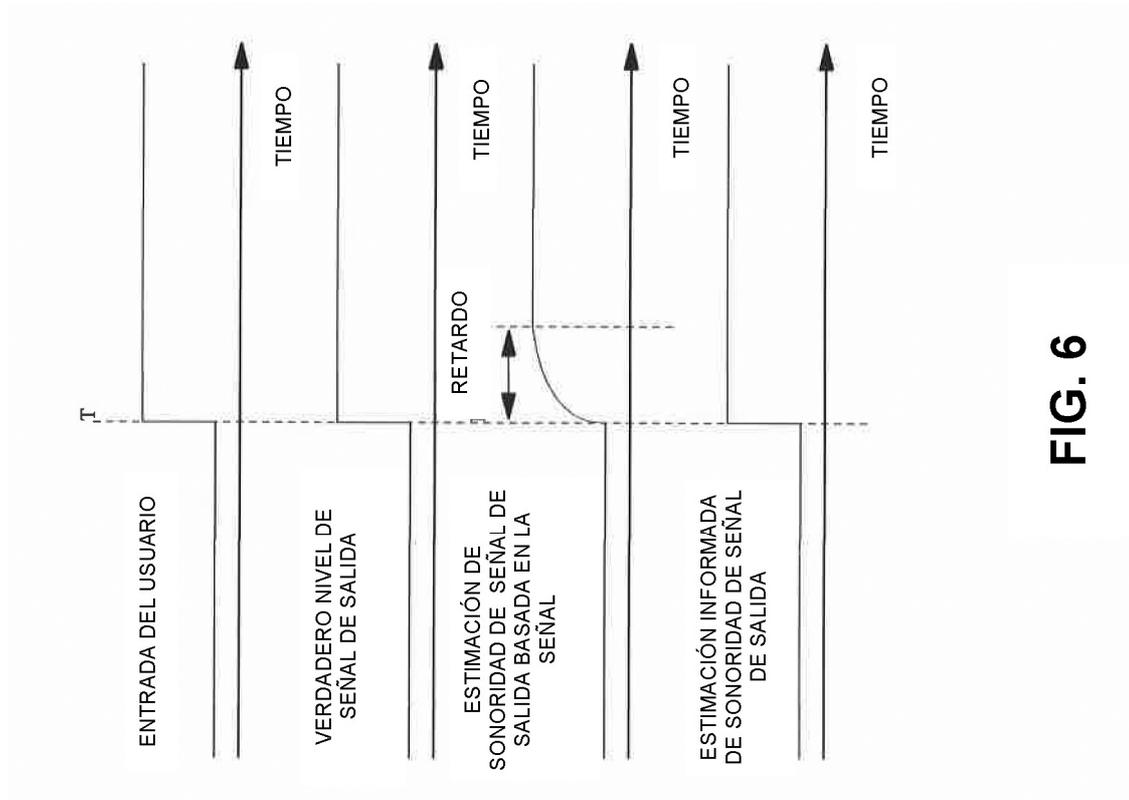


FIG. 6

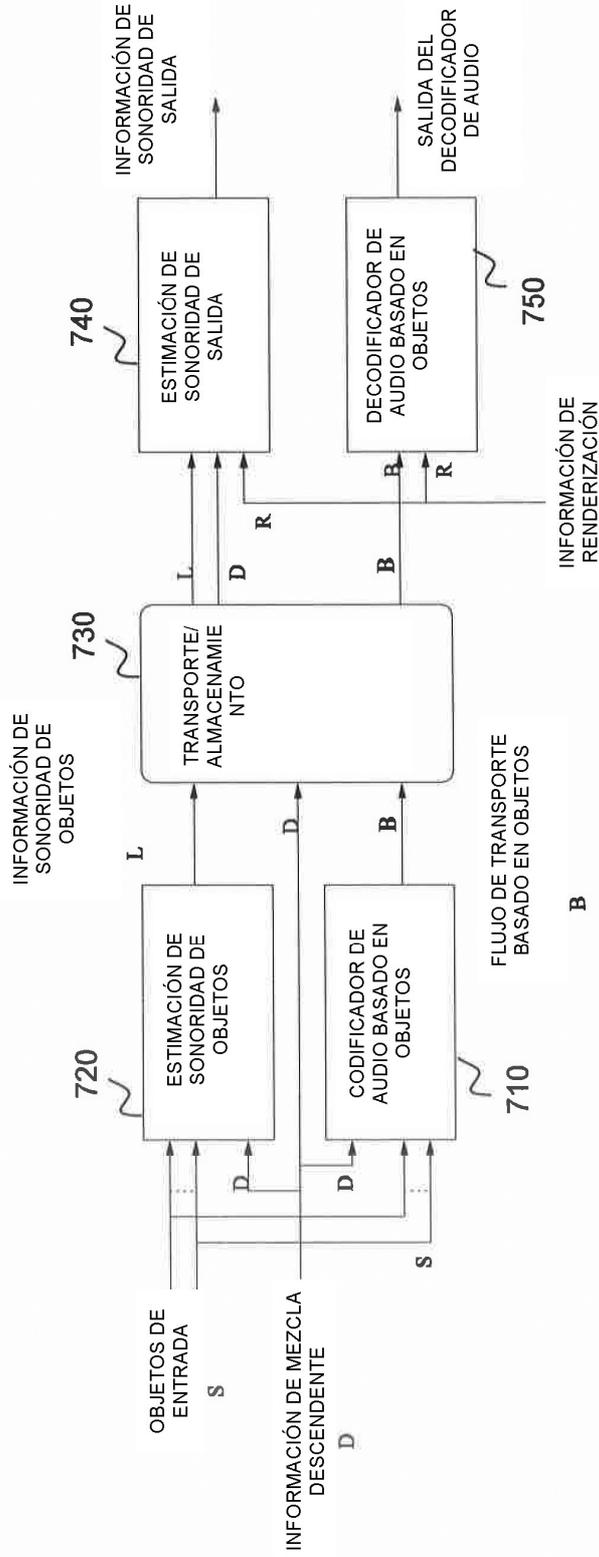


FIG. 7

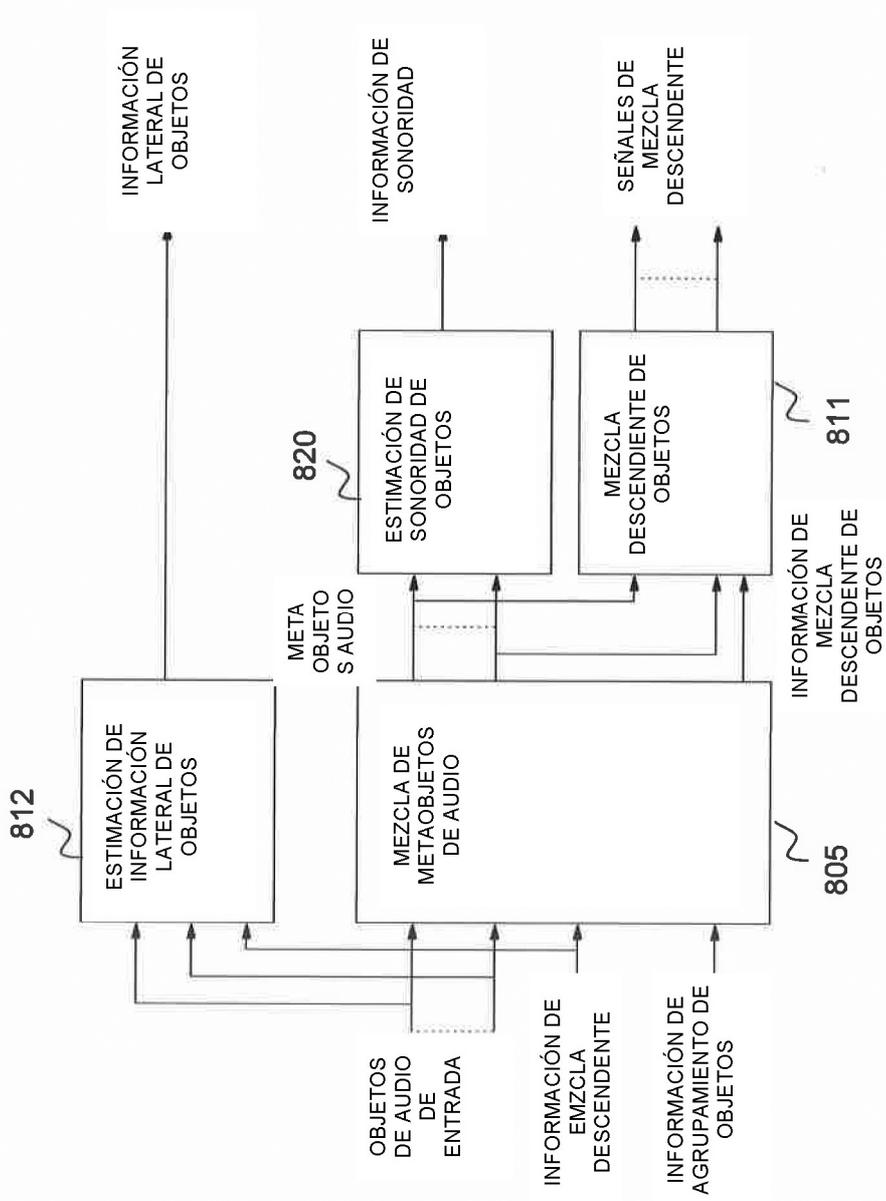


FIG. 8

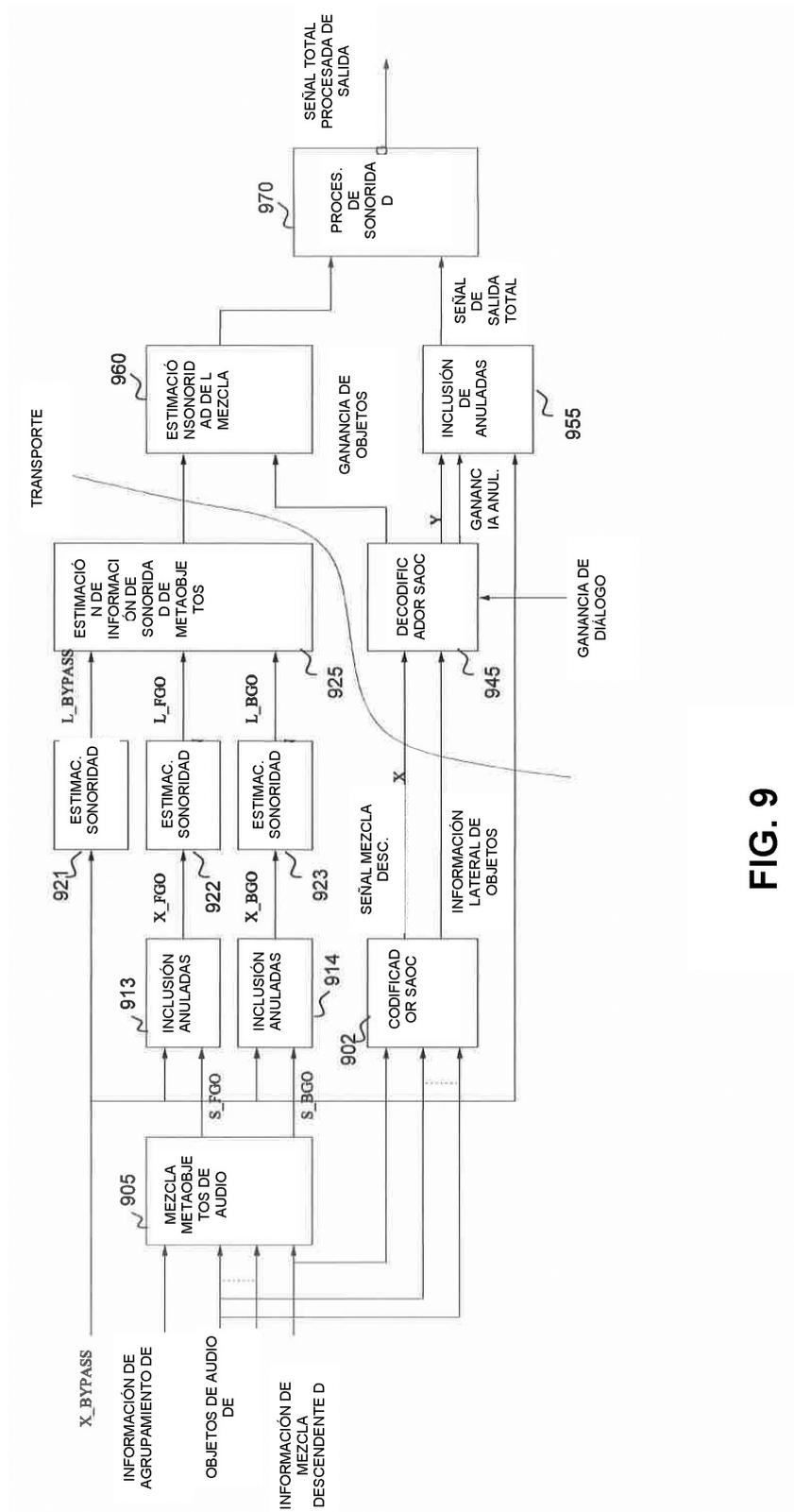


FIG. 9

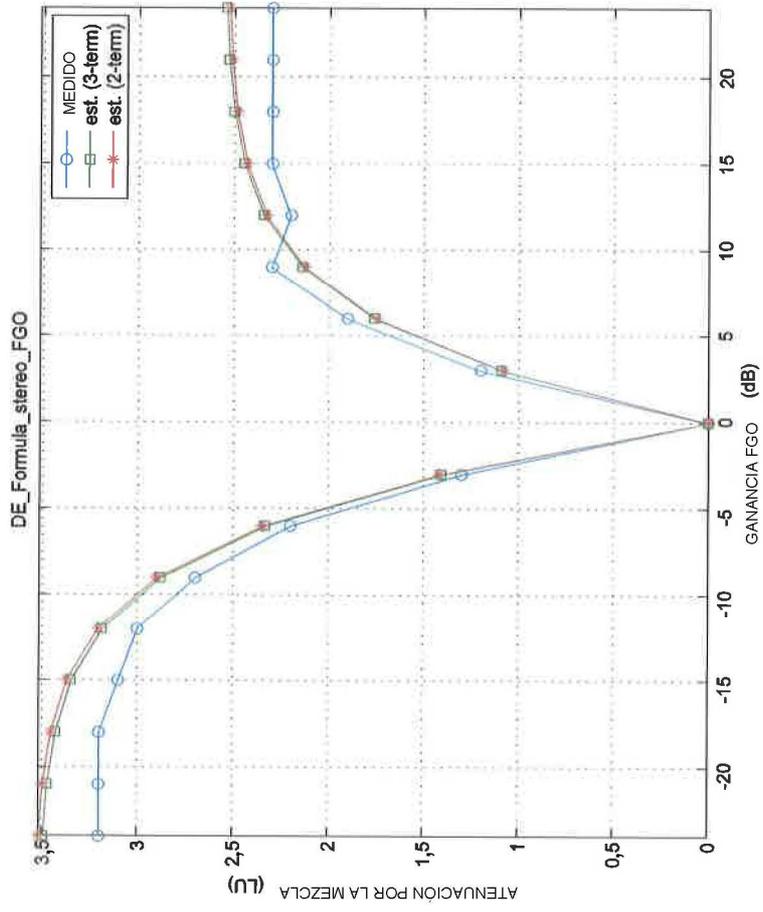


FIG. 10

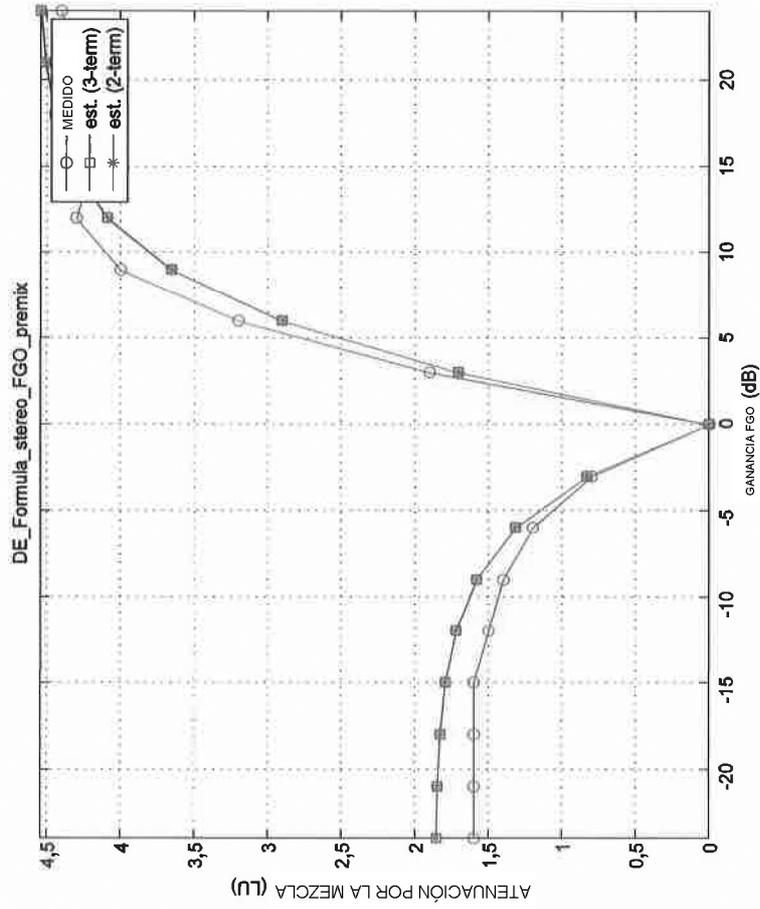


FIG. 11

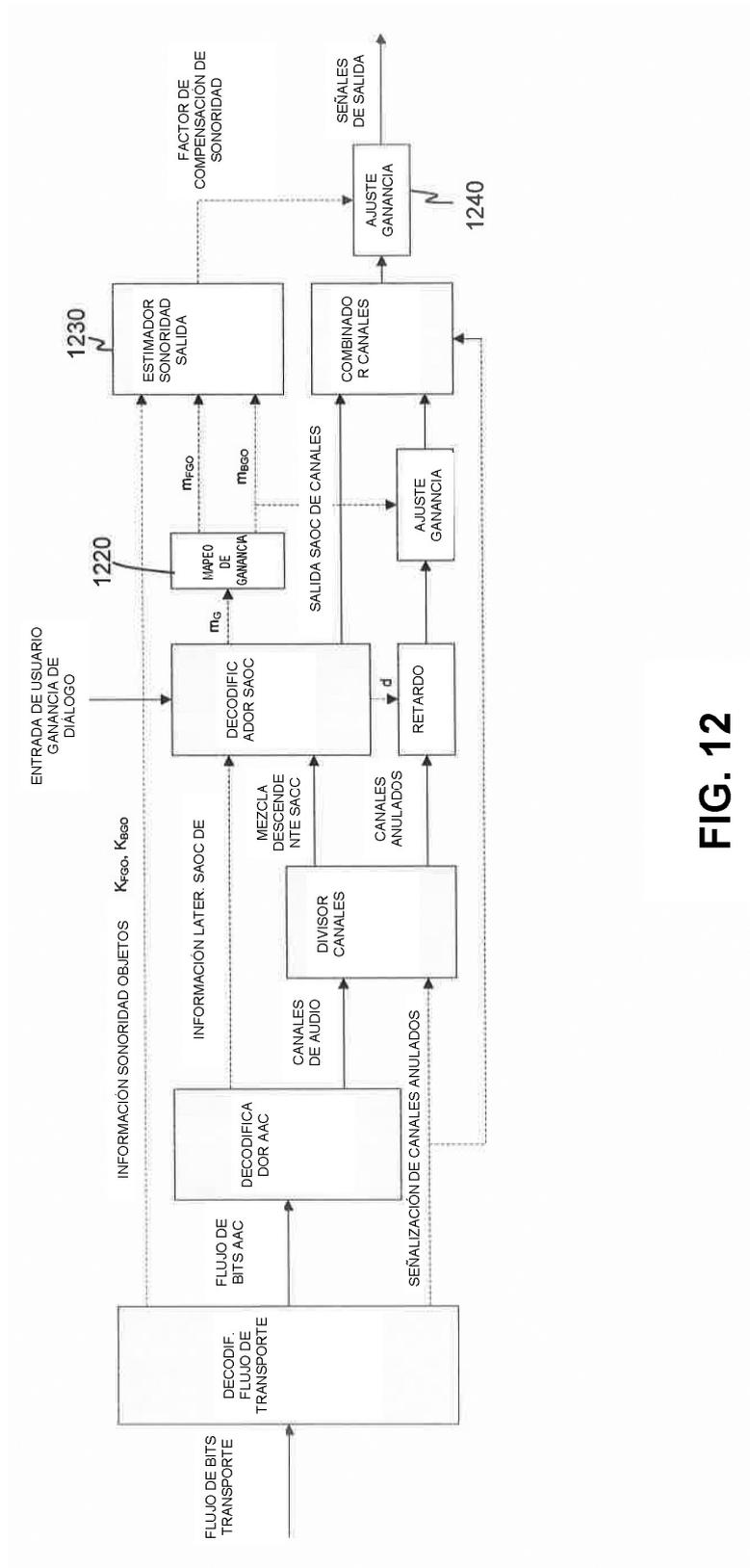


FIG. 12