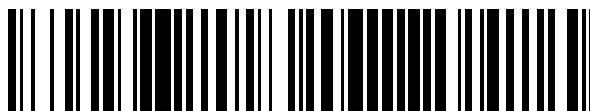


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 666 127**

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/EP2014/075787**

87 Fecha y número de publicación internacional: **04.06.2015 WO15078956**

96 Fecha de presentación y número de la solicitud europea: **27.11.2014 E 14805849 (8)**

97 Fecha y número de publicación de la concesión europea: **21.02.2018 EP 3074971**

54 Título: **Decodificador, codificador y método para la estimación informada de sonoridad en sistemas de codificación de audio basada en objetos**

30 Prioridad:

27.11.2013 EP 13194664

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

03.05.2018

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:

ARIZTI ACHA, Monica

Observaciones:

Véase nota informativa (Remarks, Remarques o Bemerkungen) en el folleto original publicado por la Oficina Europea de Patentes

ES 2 666 127 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 método para la estimación informada de sonoridad en sistemas de codificación de audio basada en objetos

5 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 método para la estimación informada de sonoridad en sistemas de codificación de audio basada en objetos.

10 Recientemente, se han propuesto técnicas paramétricas para la transmisión/almacenamiento a tasas de bits eficientes de escenas de audio que comprenden múltiples señales de objetos en el campo de la codificación de audio [BCC, JSC, SAOC, SAOC1, SAOC2] y la separación informada de fuentes [ISS1, ISS2, ISS3, ISS4, ISS5, ISS6]. Estas técnicas tienen como objetivo la reconstrucción de una escena de audio o un objeto fuente de audio de salida deseado basándose en información secundaria adicional que describe la escena de audio transmitida/almacenada y/o los objetos fuente de la escena de audio. Esta reconstrucción tiene lugar en el
15 decodificador que utiliza un esquema de separación informada de fuentes. Los objetos reconstruidos se pueden combinar para producir la escena de audio de salida. Dependiendo de la manera en que se combinan los objetos, la sonoridad perceptual de la escena de salida puede variar.

20 En la transmisión de TV y radio, se pueden normalizar los niveles de volumen de las pistas de audio de diversos programas basándose en diversos aspectos, tales como el nivel pico de la señal o el nivel de sonoridad. Dependiendo de las propiedades dinámicas de las señales, dos señales con el mismo nivel pico puede tener un nivel de sonoridad percibida con grandes variaciones. Ahora al conmutar entre programas o canales son muy molestas las diferencias de sonoridad de la señal y han sido una fuente principal de quejas de los usuarios finales en la difusión.

25 En la técnica anterior, se ha propuesto la normalización de todos los programas en todos los canales de manera similar en un nivel de referencia común utilizando una medida basada en la sonoridad perceptual de la señal. Una de esas recomendaciones en Europa es la Recomendación de EBU R128 [EBU] (a la que a continuación se hace referencia como R128).

30 La recomendación dice que la "sonoridad del programa", por ejemplo, la sonoridad promedio con respecto a un programa (o un anuncio u otra entidad de programación significativa) debe ser igual a un nivel especificado (con pequeñas desviaciones permitidas). Cuando cada vez más difusores cumplan con esta recomendación y la normalización requerida, las diferencias en la sonoridad promedio entre programas y canales se deberían minimizar.

35 La estimación de la sonoridad se puede realizar de varias maneras. Existen varios modelos matemáticos para estimar la sonoridad perceptual de una señal de audio. La recomendación EBU R128 se basa en el modelo presentado en ITU-R BS.1770 (que a continuación se denomina BS.1770) (véase [ITU]) para la estimación de sonoridad.

40 Como se ha indicado anteriormente, por ejemplo, de acuerdo con la Recomendación EBU R128, la sonoridad de los programas, por ejemplo, la sonoridad promedio a través de un programa debe ser igual a un nivel especificado con pequeñas desviaciones permitidas. Sin embargo, esto conduce a considerables problemas al llevar a cabo la representación de audio, hasta ahora sin solución en la técnica anterior. La realización de la representación de audio
45 en el lado del decodificador tiene un efecto significativo sobre la sonoridad general/total de la señal de entrada de audio recibida. Sin embargo, pese a que se lleva a cabo la representación de escenas, la sonoridad total de la señal de audio recibida debe mantenerse constante.

50 En la actualidad no existe ninguna solución específica en el lado del decodificador para este problema.

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 problema.

55 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 con 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 representación que genera una señal de salida a partir
60 de los objetos de audio. El decodificador contiene además un procesador para generar datos de modificación de codificación que se envían al codificador. El codificador modifica a continuación la codificación de los objetos de audio y, en particular, modifica los datos paramétricos en respuesta a los datos de modificación de la codificación. El enfoque permite la manipulación de los objetos de audio para que se controlen por el decodificador pero que se

realiza total o parcialmente por el codificador. De esa manera, la manipulación se puede realizar sobre los objetos de audio independientes en lugar de hacerlo sobre réplicas aproximadas, proporcionando de esta manera un rendimiento mejorado.

5 El documento EP 2 146 522 A1 ([SCH]) 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 que comprende un procesador para procesar una señal de entrada de audio para proporcionar una representación de objetos de la señal de entrada de audio, en el que esta representación de objetos se puede generar mediante una aproximación paramétricamente guiada de los objetos originales utilizando una señal de mezcla descendente de objeto. Un manipulador de objetos
10 manipula individualmente los objetos empleando metadatos basados en objetos de audio que hacen referencia a los objetos de audio individuales para obtener objetos de audio manipulados. Los objetos de audio manipulados se mezclan utilizando un mezclador de objetos para obtener, en última instancia, una señal de salida de audio con una o varias señales de canales, dependiendo de la configuración específica para la representación.

15 El documento WO 2008/046531 A1 ([ENG]) describe un codificador de objeto de audio para generar una señal de objeto codificada utilizando una pluralidad de objetos de audio que incluye un generador de información de mezcla descendente para generar 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 objeto de audio para generar parámetros de objeto para los objetos de audio, y una interfaz de salida para generar la señal de salida
20 de audio importada utilizando la información de mezcla descendente y los parámetros de objetos. Un sintetizador de audio utiliza la información de mezcla descendente para generar datos de salida que se pueden emplear para crear una pluralidad de canales de salida de la configuración de salida de audio predefinida.

25 El documento WO 2008/035275 A2 se refiere a 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 representación que genera una señal de salida desde los objetos de audio.
30

El documento EP 2 146 522 A1 se refiere a un aparato para generar al menos una señal de salida de audio que representa una superposición de al menos dos objetos de audio diferentes. Un manipulador de objetos manipula invidamente objetos usando metadatos basados en objetos de audio que hacen referencia a objetos de audio individuales para obtener objetos de audio manipulados. Los objetos de audio manipulados se mezclan usando un
35 mezclador de objetos para obtener finalmente una señal de salida de audio que tiene una o varias señales de canal que dependen de una configuración de representación específica.

40 El documento WO 2008/046531 A1 se refiere a un codificador de objetos de audio para generar una señal de objeto codificada usando una pluralidad de objetos de audio que incluye un generador de información de mezcla descendente para generar 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 objeto de audio para generar parámetros de objeto para los objetos de audio, y una interfaz de salida para generar la señal de salida de audio importada usando la información de mezcla descendente y los parámetros de objeto. Un sintetizador de audio usa la información de mezcla descendente para generar datos de salida que pueden usarse para crear una pluralidad de canales de salida de la configuración de salida de audio predefinida.
45

El documento "A Guide to Dolby Metadata", 2005, páginas 1 - 28, "http://www.dolby.com/uploadedFiles/Assets/US/Doc/Professional/18_Metadata.Guide.pdf", se refiere a metadatos que proporcionan capacidad para que los productores de contenido entreguen el audio a los consumidores en una gama de entornos de escucha. También proporciona elecciones que permiten a los consumidores ajustar sus
50 ajustes para adecuar sus entornos de escucha.

Sería deseable contar con una estimación precisa de la sonoridad promedio de salida o del cambio de la sonoridad promedio sin retardo y cuando el programa no cambia o la escena de representación no cambia, la estimación de sonoridad promedia también se mantenga estática.

55 El objetivo de la presente invención es presentar conceptos mejorados sobre la codificación, procesamiento y decodificación de señales de audio. El objetivo de la presente invención se resuelve mediante un decodificador de acuerdo con la reivindicación 1, mediante un codificador de acuerdo con la reivindicación 15, mediante un sistema de acuerdo con la reivindicación 17, mediante un método de acuerdo con la reivindicación 18, mediante un método de acuerdo con la reivindicación 19 y mediante un programa informático de acuerdo con la reivindicación 22. Se proporciona una manera informada de estimar la sonoridad de la salida en un sistema de codificación de audio
60 basada en objetos. Los conceptos proporcionados se basan en la información sobre la sonoridad de los objetivos en la mezcla de audio que se ha de proporcionar al decodificador. El decodificador utiliza esta información junto con la información de representación para estimar la sonoridad de la señal de salida. Esto permite entonces, por ejemplo, estimar la diferencia de sonoridad entre la mezcla descendente por defecto y la salida representada. Es posible así

compensar la diferencia para obtener una sonoridad aproximadamente constante en la salida independientemente de la información de representación. La estimación de sonoridad en el decodificador tiene lugar de manera totalmente paramétrica y es muy ligera, en términos informáticos, y precisa en comparación con los conceptos de estimación de la sonoridad basada en la señal.

5 Se proporcionan conceptos para obtener información sobre la sonoridad de la escena de salida específica utilizando conceptos puramente paramétricos, lo que a continuación da lugar al procesamiento de la sonoridad sin estimación de la sonoridad basada en la señal explícita en el decodificador. Más aun, se describe la tecnología específica de la Codificación De Objetos de Audio Espacial (SAOC) normalizada por MPEG [SAOC], aunque los conceptos
10 proporcionados también pueden utilizarse en conjunto con otras tecnologías de codificación de objetos de audio.

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 para recibir una señal de entrada de audio que comprende una pluralidad de señales de objetos de audio, para recibir información de sonoridad sobre las
15 señales de objetos de audio, y para recibir información de representación que indica si una o más de las señales de objetos de audio deberán amplificarse o atenuarse. Más aun, el decodificador comprende un procesador de señales para generar dichos uno o más canales de salida de audio de la señal de salida de audio. El procesador de señales está configurado para determinar un valor de compensación de sonoridad dependiendo de la información de sonoridad y dependiendo de la información de representación. Además, el procesador de señales está configurado
20 para generar dichos uno o más canales de salida de audio de la señal de salida de audio a partir de la señal de entrada de audio dependiendo de la información de representación y dependiendo del valor de compensación de sonoridad.

De acuerdo con una realización, el procesador de señales puede estar configurado para generar dichos uno o más
25 canales de salida de audio de la señal de salida de audio a partir de la señal de entrada de audio dependiendo de la información de representación y dependiendo del valor de compensación de sonoridad, de tal manera que una sonoridad de la señal de salida de audio sea igual a una sonoridad de la señal de entrada de audio, o de tal manera que la sonoridad de la señal de salida de audio se aproxime más a la sonoridad de la señal de entrada de audio que una sonoridad de una señal de audio modificada que se produciría como resultado de la modificación de la señal de
30 entrada de audio mediante la amplificación o atenuación de las señales de objetos de audio de la señal de entrada de audio de acuerdo con la información de representación.

De acuerdo con otra realización, cada una de las señales de objetos de audio de la señal de entrada de audio puede asignarse exactamente a un grupo de dos o más grupos, en el 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 una realización de este tipo, la interfaz de recepción puede estar configurada para recibir un valor de sonoridad por cada grupo de los dos o más grupos como información de sonoridad, en el que dicho valor de sonoridad indica una sonoridad total original de dichas una o más señales de objetos de audio de dicho grupo. Además, la interfaz de recepción puede estar configurada para recibir la información de representación que indica, para al menos un grupo de los dos o más
40 grupos, si dichas una o más señales de objetos de audio de dicho grupo deberán amplificarse o atenuarse mediante la indicación de una sonoridad total modificada de dichas una o más señales de objetos de audio de dicho grupo. Más aun, en una realización de este tipo, el procesador de señales puede estar configurado para determinar el valor de compensación de sonoridad dependiendo de la sonoridad total modificada de cada uno de dicho al menos un grupo de dichos dos o más grupos y dependiendo de la sonoridad total original de cada uno de dichos dos o más grupos. Además, el procesador de señales puede estar configurado para generar dichos uno o más canales de salida de audio de la señal de salida de audio a partir de la señal de entrada de audio dependiendo de la sonoridad total modificada de cada uno de dicho al menos un grupo de dichos dos o más grupos y dependiendo del valor de compensación de sonoridad.

50 En realizaciones particulares, al menos un grupo de dichos dos o más grupos puede comprender dos o más de las señales de objetos de audio.

Más aun, se proporciona un codificador. El codificador comprende una unidad de codificación basada en objetos para codificar una pluralidad de señales de objetos de audio para obtener una señal de audio codificada que
55 comprende la pluralidad de señales de objetos de audio. Además, el codificador comprende una unidad de codificación 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 de dichos uno o más valores de sonoridad depende de una o más de las señales de objetos de audio.

60 De acuerdo con una realización, cada una de las señales de objetos de audio de la señal de audio codificada puede asignarse 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 de codificación de sonoridad de objetos puede estar configurada para determinar dicho uno o más valores de sonoridad de la información de sonoridad mediante la determinación de un valor de sonoridad por cada grupo de los dos o más

grupos, en el que dicho valor de sonoridad de dicho grupo indica una sonoridad total original de dichas una o más señales de objetos de audio de dicho grupo.

Además, se proporciona un sistema. El sistema comprende un codificador de acuerdo con una de las realizaciones anteriormente descritas para codificar 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 para codificar información de sonoridad sobre las señales de objetos de audio. Más aun, el sistema comprende un decodificador de acuerdo con una de las realizaciones anteriormente descritas 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. Más aun, el decodificador está configurado para recibir adicionalmente información de representación. Además, el decodificador está configurado para determinar un valor de compensación de sonoridad dependiendo de la información de sonoridad y dependiendo de la información de representación. Más aun, el decodificador está configurado para generar dichos uno o más canales de salida de audio de la señal de salida de audio a partir de la señal de entrada de audio dependiendo de la información de representación y dependiendo del valor de compensación de sonoridad.

Se proporciona además un método para generar una señal de salida de audio que comprende uno o más canales de salida de audio. El método comprende:

- 20 - Recibir una señal de entrada de audio que comprende una pluralidad de señales de objetos de audio.
- Recibir información de sonoridad sobre las señales de objetos de audio.
- 25 - Recibir información de representación que indica si una o más de las señales de objetos de audio deberá amplificarse o atenuarse.
- Determinar un valor de compensación de sonoridad dependiendo de la información de sonoridad y dependiendo de la información de representación. Y:
- 30 - Generar dichos uno o más canales de salida de audio de la señal de salida de audio a partir de la señal de entrada de audio dependiendo de la información de representación y dependiendo del valor de compensación de sonoridad.

Además, se proporciona un método de codificación. El método comprende:

- 35 - Codificar una señal de entrada de audio que comprende una pluralidad de señales de objetos de audio. Y:
- Codificar información de sonoridad sobre 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 de dichos uno o más valores de sonoridad depende de una o más de las señales de objetos de audio.
- 40

Más aun, se proporciona un programa informático para implementar el método antes descrito al ejecutarse en una computadora o un procesador de señales.

45 Se proporcionan realizaciones preferidas en las reivindicaciones dependientes.

A continuación se describen las realizaciones de la presente invención en forma más detallada con referencia a las figuras, en las cuales:

- 50 Fig. 1 ilustra un decodificador para generar una señal de salida de audio que comprende uno o más canales de salida de audio de acuerdo con una realización,
- Fig. 2 ilustra un codificador de acuerdo con una realización,
- 55 Fig. 3 ilustra un sistema de acuerdo con una realización,
- Fig. 4 ilustra un sistema de Codificación de Objetos de Audio Espacial que comprende un codificador SAOC y un decodificador SAOC,
- 60 Fig. 5 ilustra un decodificador SAOC que comprende un decodificador de información secundaria, un separador de objetos y un representador,
- Fig. 6 ilustra el comportamiento de las estimaciones de sonoridad de la señal de salida en un cambio de sonoridad,

- Fig. 7 ilustra la estimación de sonoridad informada de acuerdo con una realización, ilustrando los componentes de un codificador y un decodificador de acuerdo con una realización,
- 5 Fig. 8 ilustra un codificador de acuerdo con otra realización,
- Fig. 9 ilustra un codificador y un decodificador de acuerdo con una realización relacionada con la Mejora de SAOC-Diálogo, que comprende canales de desviación,
- 10 Fig. 10 presenta una primera ilustración de un cambio de sonoridad medido y el resultado del uso de los conceptos proporcionados para estimar el cambio de sonoridad de manera paramétrica,
- Fig. 11 presenta una segunda ilustración de un cambio de sonoridad medido y el resultado del uso de los conceptos proporcionados para estimar el cambio de sonoridad de manera paramétrica, y
- 15 Fig. 12 ilustra otra realización de la realización de la compensación de sonoridad.

Antes de describir en detalle las realizaciones preferidas, se describe la estimación de sonoridad, la Codificación de Objetos de Audio Espacial (SAOC) y la Mejora de Diálogos (DE).

20 En primer lugar se describe la estimación de sonoridad.

Como ya se ha indicado anteriormente, la recomendación de EBU R128 se basa en el modelo presentado en ITU-R BS.1770 para la estimación de sonoridad. Esta medida se utilizará como ejemplo, aunque los conceptos descritos a continuación también se pueden aplicar a otras medidas de la sonoridad.

La operación de estimación de sonoridad de acuerdo con BS.1770 es relativamente sencilla y se basa en los siguientes pasos principales [ITU]:

- 30 - Se filtra la señal de entrada x_i (o señales, en el caso de la señal multicanal) con un filtro K (una combinación de filtros “shelving” o y de paso alto) para obtener la señal (o señales) y_i .
- Se calcula la energía cuadrática media z_i de la señal y_i .
- 35 - En el caso de la señal multicanal, se aplica la ponderación de canales G_i y se suman las señales ponderadas. A continuación se define la sonoridad de la señal de la siguiente manera

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

40 con el valor constante $c=-0,691$. A continuación la salida se expresa en las unidades “LKFS” (Sonoridad, ponderada con K, relativa a Escala Completa) que se escala de manera similar a la escala de decibelios.

En la fórmula anterior, G_i puede ser, por ejemplo, igual a 1 en algunos de los canales, mientras que G_j puede ser, por ejemplo, 1,41 en el caso de algunos otros canales. Por ejemplo, si se contempla un canal izquierdo, un canal derecho, un canal central, un canal de envolvente izquierdo y un canal de envolvente derecho, las respectivas ponderaciones G_i pueden ser, por ejemplo, 1 en el caso del canal izquierdo, derecho y central y pueden ser, por ejemplo, 1,41 en el caso del canal de envolvente izquierdo y el canal de envolvente derecho, véase [ITU].

Se puede ver que el valor de sonoridad L está estrechamente relacionado con el logaritmo de la energía de la señal.

50 A continuación se describe la Codificación de Objetos de Audio Espacial.

Los conceptos de la codificación de audio basada en objetos permiten una gran flexibilidad en el lado del decodificador de la cadena. Un ejemplo de concepto de codificación de audio basada en objetos es la Codificación de Objetos de Audio Espacial (SAOC).

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

60 El codificador SAOC 410 recibe N señales de objetos de audio S_1, \dots, S_N como entrada. Más aun, el codificador SAOC 410 recibe asimismo instrucciones “Mezclar Información D” sobre cómo se deben combinar estos objetos 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 secundaria de los objetos y del proceso de mezcla descendente, y

esta información secundaria se transmite y/o almacena junto con las señales de mezcla descendente.

Una propiedad muy importante del sistema SAOC es que la señal de mezcla descendente \mathbf{X} que comprende los canales de mezcla descendente X_1, \dots, X_M forma una señal semánticamente representativa. En otras palabras, es posible escuchar la señal de mezcla descendente. Si, por ejemplo, el receptor no posee la funcionalidad de decodificador SAOC, de todas maneras el receptor puede proporcionar siempre la señal de mezcla descendente como salida.

La Fig. 5 ilustra un decodificador SAOC que comprende un decodificador de información secundaria 510, un separador de objetos 520 y un representador 530. El decodificador SAOC ilustrado por la Fig. 5 recibe, por ejemplo, de un codificador SAOC, la señal de mezcla descendente y la información secundaria. La señal de mezcla descendente se puede considerar una señal de entrada de audio que comprende las señales de objetos de audio, ya que las señales de objetos de audio se mezclan dentro de la señal de mezcla descendente (las señales de objetos de audio se mezclan dentro de dichos uno o más canales de mezcla descendente de la señal de mezcla descendente).

Seguidamente, el decodificador SAOC puede intentar, por ejemplo, reconstruir (virtualmente) los objetos originales, por ejemplo, empleando el separador de objetos 520, por ejemplo, utilizando la información secundaria decodificada.

A continuación se combinan estas reconstrucciones (virtuales) de objetos $\hat{S}_1, \dots, \hat{S}_N$, por ejemplo, las señales de objetos de audio reconstruidas, basándose en la información de representación, por ejemplo, una matriz de representación \mathbf{R} , para producir K canales de salida de audio Y_1, \dots, Y_K de una señal de salida de audio \mathbf{Y} .

En la SAOC, con frecuencia se reconstruyen señales de objetos de audio, por ejemplo, empleando información de covarianza, por ejemplo, una matriz de covarianza de señal \mathbf{E} , que se transmite desde el codificador SAOC al decodificador SAOC.

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

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

en el que

- N número de señales de objetos de audio,
- $N_{muestras}$ número de muestras consideradas de una señal de objetos de audio
- M número de canales de mezcla descendente,
- \mathbf{X} señal de audio con mezcla descendente, tamaño $M \times N_{muestras}$,
- \mathbf{D} matriz de mezcla descendente, tamaño $M \times N$
- \mathbf{E} matriz de covarianza de señal, tamaño $N \times N$ definido según $\mathbf{E} = \mathbf{X} \mathbf{X}^H$
- \mathbf{S} N señales de objetos de audio paramétricamente reconstruidas, tamaño $N \times N_{muestras}$
- $(\cdot)^H$ operador auto-adjunto (Hermitiano) que representa la transpuesta conjugada de (\cdot)

Seguidamente se puede aplicar una matriz de representación \mathbf{R} a las señales de objetos de audio reconstruidas para obtener los canales de salida de audio de la señal de salida de audio \mathbf{Y} , por ejemplo, de acuerdo con la fórmula:

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

en el que

K número de los canales de salida de audio Y_1, \dots, Y_K de la señal de salida de audio \mathbf{Y} .

\mathbf{R} matriz de representación de un tamaño $K \times N$

\mathbf{Y} señal de salida de audio que comprende los K canales de salida de audio, tamaño $K \times N_{muestras}$

En la Fig. 5, se hace referencia al proceso de reconstrucción de objetos, por ejemplo, el que lleva a cabo el separador de objetos 520, con la noción "virtual" u "opcional", ya que puede no ser indispensable que se produzca necesariamente, sino que se puede obtener la funcionalidad deseada combinando las etapas de reconstrucción y representación en el dominio paramétrico (es decir, combinando las ecuaciones).

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} en primer lugar, y a continuación aplicando la información de representación \mathbf{R} a las señales de objetos de audio reconstruidas para obtener los canales de salida de audio Y_1, \dots, Y_K , se pueden realizar

ambas etapas en una sola etapa, de manera que los canales de salida de audio Y_1, \dots, Y_K se generen directamente de los canales de mezcla descendente.

Por ejemplo, se puede emplear la siguiente fórmula:

5

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

En principio, la información de representación \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 puede no lograrse necesariamente la escena de salida solicitada. Como regla general aproximada que cubre muchos casos prácticos, cuánto más difiere la escena de salida solicitada de la señal de mezcla descendente, más errores de reconstrucción audibles hay.

10

A continuación se describe la Mejora de Diálogos (DE). Se puede emplear, por ejemplo, la tecnología SAOC para lograr la situación. Cabe señalar que aunque el nombre "Mejora de Diálogos" sugiere un centrarse en las señales orientadas al diálogo, se puede usar también el mismo principio con otros tipos de señales.

15

En el escenario de DE, los grados de libertad del sistema están limitados con respecto al caso general.

20

Por ejemplo, las señales de objetos de audio $S_1, \dots, S_N = \mathbf{S}$ se agrupan (y posiblemente se mezclan) formando dos meta-objetos de un objeto en primer plano (FGO) \mathbf{S}_{FGO} y un objeto de segundo plano (BGO) \mathbf{S}_{BGO} .

Más aun, la escena de salida $Y_1, \dots, Y_K = \mathbf{Y}$ se asemeja a la señal de mezcla descendente $X_1, \dots, X_K = \mathbf{X}$. 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 meta-objetos FGO y BGO. Para ser más precisos, la señal de mezcla descendente se obtiene mezclando el FGO y el BGO con algunas ponderaciones escalares

25

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

30

y la escena de salida se obtiene de manera similar con alguna ponderación de escalado del FGO y el BGO:

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

35

Dependiendo de los valores relativos de las ponderaciones de mezcla, el balance entre el FGO y el BGO puede cambiar. Por ejemplo, con la configuración

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

40

es posible incrementar el nivel relativo del FGO en la mezcla. Si el FGO es el diálogo, esta configuración proporciona una funcionalidad de Mejora de Diálogos.

Como ejemplo de caso práctico, el BGO puede consistir en ruidos de un estadio y otro sonido de segundo plano durante un evento deportivo y el FGO es la voz del comentarista. La funcionalidad DE permite que el usuario final amplifique o atenúe el nivel del comentarista en relación con el segundo plano.

45

Las realizaciones se basan en el hallazgo de que el uso de la tecnología SAOC (o similar) en un escenario de difusión permite proporcionar al usuario final una funcionalidad de manipulación de señales incrementada. Se proporciona mayor funcionalidad que solo el cambio de canal y el ajuste del volumen de reproducción.

50

Se ha descrito anteriormente de manera breve una posibilidad del empleo de la tecnología DE. Si se normaliza el nivel de la señal de difusión, que es la señal de mezcla descendente para SAOC, por ejemplo, de acuerdo con R128, los diferentes programas tienen una sonoridad promedio similar cuando no se aplica el procesamiento (SAOC) (o la descripción de representación es igual a la descripción de la mezcla descendente). Sin embargo, cuando se aplica cierto grado de procesamiento (SAOC), la señal de salida difiere 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 llevar a una situación en la cual la sonoridad de la señal de salida entre canales o programas puede tener, una vez más, los saltos o diferencias molestas. En otras palabras, se pierden parcialmente los beneficios de la normalización aplicada por el difusor.

55

60

Este problema no es específico del escenario de SAOC o DE solamente, sino que también puede aparecer con otros

conceptos de codificación de audio que permitan al usuario final interactuar con el contenido. Sin embargo, en muchos casos no causa perjuicio alguno si la señal de salida tiene una sonoridad diferente de la de la mezcla descendente por defecto.

5 Como se ha indicado anteriormente, la sonoridad total de un programa de señal de entrada de audio debe ser igual a un nivel especificado con pequeñas desviaciones permitidas. Sin embargo, de acuerdo con lo mencionado anteriormente, esto lleva a problemas significativos al llevar a cabo la representación del audio, ya que la representación puede tener un efecto significativo sobre la sonoridad general/total de la señal de entrada de audio recibida. Sin embargo, a pesar de llevarse a cabo la representación de la escena, la sonoridad total de la señal de
10 audio recibida se mantiene igual.

Un enfoque sería estimar la sonoridad de una señal mientras se está reproduciendo, y con un concepto de integración temporal apropiado, la estimación puede converger en la verdadera sonoridad promedio después de un tiempo. Sin embargo, el tiempo requerido para la convergencia es problemático desde el punto de vista del usuario
15 final. Cuando la estimación de la sonoridad cambia aunque no se apliquen cambios a la señal, la compensación por cambio de sonoridad también debería reaccionar y cambiar su comportamiento. Esto traería aparejado una señal de salida con sonoridad promedio temporalmente variable, lo que puede percibirse como algo ciertamente molesto.

La Fig. 6 ilustra el comportamiento de las estimaciones de sonoridad de la señal de salida ante un cambio de sonoridad. Entre otras cosas, se ilustra una estimación de la sonoridad de la señal de salida basada en la señal, que demuestra el efecto de una solución como la que se acaba de describir. La estimación aborda la estimación correcta con bastante lentitud. En lugar de una estimación de la sonoridad de la señal de salida basada en la señal, sería preferible una estimación informada de la sonoridad de la señal de salida, que determina correctamente de inmediato la sonoridad de la señal de salida.
20

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 por el incremento de valor. El verdadero nivel de la señal de salida y, por consiguiente la sonoridad, cambia en el mismo instante de tiempo. Cuando se realiza la estimación de sonoridad de la señal de salida a partir de la señal de salida con cierto tiempo de integración temporal, la estimación cambia gradualmente y llega al valor correcto después de un cierto retardo. Durante este retardo, los valores de estimación son cambiantes y no se pueden utilizar con seguridad para el procesamiento posterior de la señal de salida, por ejemplo, para la corrección del nivel de sonoridad.
25

Como ya se ha mencionado, sería deseable tener una estimación precisa de la sonoridad promedio de salida o el cambio de sonoridad promedio sin retardo y cuando el programa no cambia o no cambia la escena de la representación, la estimación de sonoridad debería permanecer estática también. En otras palabras, cuando se aplica cierta compensación de cambio de sonoridad, el parámetro de compensación debe cambiar solo cuando cambia el programa o cuando hay alguna interacción con el usuario.
30

El comportamiento ventajoso está ilustrado en el esquema inferior de la Fig. 6 (estimación informada de la sonoridad de la señal de salida). La estimación de la sonoridad de la señal de salida debe cambiar inmediatamente cuando cambia la entrada del usuario.
35

La Fig. 2 ilustra un codificador de acuerdo con una realización.
40

El codificador comprende una unidad de codificación basada en objetos 210 para codificar 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.
45

Además, el codificador comprende una unidad de codificación de sonoridad de objetos 220 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 de dichos uno o más valores de sonoridad dependen de una o más de las señales de objetos de audio.
50

De acuerdo con 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 de codificación de sonoridad de objetos 220 está configurada para determinar dicho uno o más valores de sonoridad de la información de sonoridad mediante la determinación de un valor de sonoridad por cada grupo de los dos o más grupos, en el que dicho valor de sonoridad de dicho grupo indica una sonoridad total original de dichas una o más señales de objetos de audio de dicho grupo.
55
60

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

El decodificador comprende una interfaz de recepción 110 para recibir 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 representación que indica si una o más de las señales de objetos de audio deberá amplificarse o atenuarse.

Más aun, el decodificador comprende un procesador de señales 120 para generar dichos 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 dependiendo de la información de sonoridad y dependiendo de la información de representación. Además, el procesador de señales 120 está configurado para generar dichos uno o más canales de salida de audio de la señal de salida de audio a partir de la señal de entrada de audio dependiendo de la información de representación y dependiendo del valor de compensación de sonoridad.

De acuerdo con una realización, el procesador de señales 110 está configurado para generar dichos uno o más canales de salida de audio de la señal de salida de audio a partir de la señal de entrada de audio dependiendo de la información de representación y dependiendo del valor de compensación de sonoridad, de tal 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 tal manera que la sonoridad de la señal de salida de audio se aproxime más a la sonoridad de la señal de entrada de audio que la sonoridad de una señal de audio modificada que se produciría como resultado de la modificación de la señal de entrada de audio mediante la amplificación o atenuación de las señales de objetos de audio de la señal de entrada de audio de acuerdo con la información de representación.

De acuerdo con 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 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.

En una realización de este tipo, la interfaz de recepción 110 está configurada para recibir un valor de sonoridad por cada grupo de los dos o más grupos como información de sonoridad, en el que dicho valor de sonoridad indica una sonoridad total original de dichas una o más señales de objetos de audio de dicho grupo. Además, la interfaz de recepción 110 está configurada para recibir la información de representación que indica, con respecto a al menos un grupo de los dos o más grupos, si dichas una o más señales de objetos de audio de dicho grupo deberá amplificarse o atenuarse mediante la indicación de una sonoridad total modificada de dichas una o más señales de objetos de audio de dicho grupo. Más aun, en una realización de este tipo, el procesador de señales 120 está configurado para determinar el valor de compensación de sonoridad dependiendo de la sonoridad total modificada de cada uno de dicho al menos un grupo de dichos dos o más grupos y dependiendo de la sonoridad total original de cada uno de dichos dos o más grupos. Además, el procesador de señales 120 está configurado para generar dichos uno o más canales de salida de audio de la señal de salida de audio a partir de la señal de entrada de audio dependiendo de la sonoridad total modificada de cada uno de dicho al menos un grupo de dichos dos o más grupos y dependiendo del valor de compensación de sonoridad.

En realizaciones particulares, al menos un grupo de dichos dos o más grupos comprende dos o más de las señales de objetos de audio.

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 de acuerdo con las fórmulas:

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

en la que c es un valor constante.

Las realizaciones se basan en los siguientes hallazgos: diferentes señales de objetos de audio de la señal de entrada de audio puede tener diferente sonoridad y, por consiguiente, diferente energía. Si, por ejemplo, un usuario desea aumentar la sonoridad de una de las señales de objetos de audio, se puede ajustar de manera correspondiente la información de representación, y el aumento de la sonoridad de esta señal de objeto de audio aumenta la energía de este objeto de audio. Esto daría lugar a una sonoridad incrementada de la señal de salida de audio. Para mantener constante la sonoridad total, se tiene que llevar a cabo una compensación de la sonoridad. En otras palabras, la señal de audio modificada que se generaría como resultado de la aplicación de la información de representación sobre la señal de entrada de audio debería ajustarse. Sin embargo, el efecto exacto de la amplificación de una de las señales de objetos de audio sobre la sonoridad total de la señal de audio modificada depende de la sonoridad original de la señal de objeto de audio amplificada, por ejemplo, de la señal de objeto de audio, la sonoridad de la cual se incrementa. 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 ha de ser menor. Si, por el contrario, 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.

Se pueden considerar dos ejemplos. En ambos ejemplos, una señal de entrada de audio comprende dos señales de objetos de audio, y en ambos ejemplos, aplicando la información de representación se incrementa el 50 % la energía de la primera de las señales de objetos de audio.

5 En el primer ejemplo, la primera señal de objeto de audio contribuye el 20 % y la segunda señal de objeto de audio contribuye el 80 % a 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 el 40 % y la segunda señal de objeto de audio contribuye el 60 % a la energía total de la señal de entrada de audio. En ambos ejemplos estas contribuciones se derivan de la información de sonoridad sobre las señales de objetos de audio, ya que existe una relación directa
10 entre sonoridad y energía.

En el primer ejemplo, se produce un aumento del 50 % de la energía del primer objeto de audio dando como resultado que la señal de audio modificada se genera mediante la aplicación de la información de representación a la señal de entrada de audio tiene una energía total de $1,5 \times 20 \% + 80 \% = 110 \%$ de la energía de la señal de
15 entrada de audio.

En el segundo ejemplo, se produce un aumento del 50 % de la energía del primer objeto de audio dando como resultado que la señal de audio modificada que se genera mediante la aplicación de la información de representación a la señal de entrada de audio tiene una energía total de $1,5 \times 40 \% + 60 \% = 120 \%$ de la energía de la señal de entrada de audio.
20

Por consiguiente, después de aplicar la información de representación sobre la señal de entrada de audio, en el primer ejemplo, solo se tiene que reducir la energía total de la señal de audio modificada el 9 % ($10/110$) para obtener igual 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 reducirse el 17 % ($20/120$). Para este
25 fin, se puede calcular un valor de compensación de sonoridad.

Por ejemplo, el valor de compensación de sonoridad puede ser un valor escalar que se aplica a todos los canales de salida de audio de la señal de salida de audio.

30 De acuerdo con una realización, el procesador de señales está configurado para generar la señal de audio modificada mediante la modificación de la señal de entrada de audio mediante la amplificación o atenuación de las señales de objetos de audio de la señal de entrada de audio de acuerdo con la información de representación. Más aun, el procesador de señales está configurado para generar la señal de salida de audio mediante la aplicación del valor de compensación de sonoridad a la señal de audio modificada, de tal manera que la sonoridad de la señal de
35 salida de audio sea igual a la sonoridad de la señal de entrada de audio, o de tal manera que la sonoridad de la señal de salida de audio se aproxime más a la sonoridad de la señal de entrada de audio que la sonoridad de la señal de audio modificada.

Por ejemplo, en el primer ejemplo anterior, se puede fijar el valor de compensación de sonoridad l_{cv} , por ejemplo, en un valor $l_{cv} = 10/11$, y se puede aplicar un factor de multiplicación de $10/11$ a todos los canales que se producen como resultado de la representación de los canales de entrada de audio de acuerdo con la información de representación.
40

En consecuencia, por ejemplo, en el segundo ejemplo anterior, se puede fijar el valor de compensación de sonoridad l_{cv} , por ejemplo, en un valor $l_{cv} = 10/12 = 5/6$, y se puede aplicar un factor de multiplicación de $5/6$ en todos los canales que se producen como resultado de la representación de los canales de entrada de audio de acuerdo con la información de representación.
45

En otras realizaciones, cada una de las señales de objetos de audio puede asignarse a una de una pluralidad de grupos, y se puede transmitir un valor de sonoridad para cada uno de los grupos que indican un valor de sonoridad total de las señales de objetos de audio de dicho grupo. Si la información de representación especifica que la energía de uno de los grupos se atenúa o amplifica, por ejemplo, se amplifica el 50 % como en el caso anterior, se puede calcular un aumento de la energía total y se puede determinar un valor de compensación de sonoridad de la manera anteriormente descrita.
50

Por ejemplo, de acuerdo con una realización, cada una de las señales de objetos de audio de la señal de entrada de audio se asigna exactamente a un grupo de exactamente dos grupos como dicho dos o más grupos. Cada una de las señales de objetos de audio de la señal de entrada de audio se asigna a un grupo de objetos de primer plano de los exactamente dos grupos o a un grupo de objetos de segundo plano de los exactamente dos grupos. La interfaz de recepción 110 está configurada para recibir la sonoridad total original de dichas una o más señales de objetos de
60 audio del grupo de objetos de primer plano. Más aun, la interfaz de recepción 110 está configurada para recibir la sonoridad total original de dichas una o más señales de objetos de audio del grupo de objetos de segundo plano. Además, la interfaz de recepción 110 está configurada para recibir la información de representación que indica, en el caso de al menos un grupo de los exactamente dos grupos, si dichas una o más señales de objetos de audio de

cada uno de dicho al menos un grupo deberán amplificarse o atenuarse mediante la indicación de una sonoridad total modificada de dichas una o más señales de objetos de audio de dicho grupo.

5 En una realización de este tipo, el procesador de señales 120 está configurado para determinar el valor de compensación de sonoridad dependiendo de la sonoridad total modificada de cada uno de dicho al menos un grupo, dependiendo de la sonoridad total original de dichas una o más señales de objetos de audio del grupo de objetos de primer plano, y dependiendo de la sonoridad total original de dichas una o más señales de objetos de audio del grupo de objetos de segundo plano. Más aun, el procesador de señales 120 está configurado para generar dichos uno o más canales de salida de audio de la señal de salida de audio a partir de la señal de entrada de audio
10 dependiendo de la sonoridad total modificada de cada uno de dicho al menos un grupo y dependiendo del valor de compensación de sonoridad.

De acuerdo con algunas realizaciones, cada una de las señales de objetos de audio se asigna a uno de tres o más grupos, y la interfaz de recepción puede estar configurada para recibir un valor de sonoridad para cada uno de los
15 tres o más grupos, que indica la sonoridad total de la señales de objetos de audio de dicho grupo.

De acuerdo con una realización, para determinar el valor total de sonoridad de dos o más señales de objetos de audio, por ejemplo, se determina el valor de energía correspondiente al valor de sonoridad por cada señal de objeto de audio, se suman los valores de energía de todos los valores de sonoridad para obtener una suma de energía y se
20 determina el valor de sonoridad que corresponde a la suma de energías como valor de sonoridad total de las dos o más señales de objetos de audio. Por ejemplo, se pueden emplear las siguientes fórmulas

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

25 En algunas realizaciones, se transmiten los 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 el que por cada uno de los grupos, se transmite un valor de sonoridad.

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 ningún valor de sonoridad. Por el contrario, el decodificador puede presumir, por ejemplo, que estas señales de objetos de audio o grupos de señales de objetos de audio, para los cuales no se transmite ningún valor de sonoridad, tienen un valor de sonoridad predefinido. El decodificador puede basar todas las determinaciones posteriores, por ejemplo, en este valor de sonoridad predefinido.
30

De acuerdo con 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 señal de entrada de audio, en el que dichos uno o más canales de mezcla descendente comprenden las señales de objetos de audio, y en el que el número de señales de objetos de audio es menor que el número de dichos uno o más canales de mezcla descendente. La interfaz de recepción 110 está configurada para recibir información de mezcla descendente que indica de cómo se mezclan las señales de objetos de audio dentro de dichos uno o más canales de mezcla descendente. Más aun, el procesador de señales 120 está configurado para generar dichos uno o más canales de salida de audio de la señal de salida de audio a partir de la señal de entrada de audio dependiendo de la información de mezcla descendente, dependiendo de la información de representación y dependiendo del valor de compensación de sonoridad. En una realización particular, el procesador de señales 120 puede estar configurado, por ejemplo, para calcular el valor de compensación de sonoridad dependiendo de la información de mezcla descendente.
35

Por ejemplo, la información de mezcla descendente puede ser una matriz de mezcla descendente. En algunas realizaciones, el decodificador puede ser un decodificador SAOC. En realizaciones de este tipo, la interfaz de recepción 110 puede estar configurada además, por ejemplo, para recibir información de covarianza, por ejemplo, una matriz de covarianza como la anteriormente descrita.
40

Con respecto a la información de representación que indica si una o más de las señales de objetos de audio deberá amplificarse o atenuarse, se debe tener en cuenta que, por ejemplo, la información que indica cómo una o más de las señales de objetos de audio deberá amplificarse o atenuarse, es información de representación. Por ejemplo, una matriz de representación R, por ejemplo, una matriz de representación de SAOC, es información de representación.
45

60 La Fig. 3 ilustra un sistema de acuerdo con una realización.

El sistema comprende un codificador 310 de acuerdo con una de las realizaciones anteriormente descritas para codificar 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.

Más aun, el sistema comprende un decodificador 320 de acuerdo con una de las realizaciones anteriormente descritas 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 señal de entrada de audio y la información de sonoridad. Más aun, el decodificador 320 está configurado para recibir asimismo información de representación. Además, el decodificador 320 está configurado para determinar un valor de compensación de sonoridad dependiendo de la información de sonoridad y dependiendo de la información de representación. Asimismo, el decodificador 320 está configurado para generar dichos uno o más canales de salida de audio de la señal de salida de audio a partir de la señal de entrada de audio dependiendo de la información de representación y dependiendo del valor de compensación de sonoridad.

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

El flujo de transporte 730 en sí comprende información de sonoridad **L**, información de mezcla descendente **D** y la salida del codificador de audio basado en objetos 710 **B**.

A la derecha del flujo de transporte 730, se ilustran los componentes de un procesador de señales de un decodificador para la codificación de audio basada en objetos. No se ilustra la interfaz de recepción del decodificador. Se ilustra 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 estar configurado para determinar el valor de compensación de sonoridad. La unidad de decodificación de audio basada en objetos 750 puede estar configurada para determinar una señal de audio modificada a partir de una señal de audio, que se envía al decodificador, mediante la aplicación de la información de representación **R**. No se muestra la aplicación del valor de compensación de sonoridad a la señal de audio modificada para compensar un cambio de sonoridad total causado por la representación en la Fig. 7.

La entrada enviada al codificador consiste en los objetos de entrada **S** en el mínimo. El sistema estima la sonoridad de cada objeto (o alguna otra información relacionada con la sonoridad, tal como por ejemplo las energías de los objetos), por ejemplo, por la unidad de codificación de sonoridad de objetos 720, y esta información **L** se transmite y/o almacena. (También es posible que la sonoridad de los objetos se envíe como entrada al sistema, y puede omitirse el paso de estimación dentro del sistema).

En la realización de la Fig. 7, el decodificador recibe al menos la información de la sonoridad de los objetos y, por ejemplo, la información de representación **R** que describe la mezcla de los objetos en la señal de salida. Basándose en estas, por ejemplo, el estimador de sonoridad de salida 740, estima la sonoridad de la señal de salida y proporciona esta información como su salida.

La información de mezcla descendente **D** puede proporcionarse como la información de representación, en cuyo caso la estimación de sonoridad proporciona una estimación de la sonoridad de la señal de mezcla descendente. También es posible proporcionar la información de mezcla descendente como entrada a la estimación de sonoridad de los objetos y transmitir y/o almacenarla junto con la información de la sonoridad de los objetos. A continuación la estimación de sonoridad de salida puede estimar simultáneamente la sonoridad de la señal de mezcla descendente y la salida representada y enviar estos dos valores o su diferencia como información de la sonoridad de salida. El valor de diferencia (o su inversa) describe la compensación requerida que debería aplicarse a la señal de salida representada para hacer que su sonoridad sea similar a la sonoridad de la señal de mezcla descendente. La información de la sonoridad de los objetos puede incluir además información con respecto a los coeficientes de correlación entre diversos objetos y esta correlación puede utilizarse en la estimación de sonoridad de salida para una estimación más precisa.

A continuación, se describe una realización preferida para aplicación de mejora de diálogos.

En la aplicación de mejora de diálogos, como se ha descrito anteriormente, se agrupan las señales de objetos de audio de entrada y se realiza la mezcla descendente parcial para formar dos meta-objetos, FGO y BGO, que a continuación pueden ser sumados fácilmente para determinar la señal de mezcla descendente final.

Después de la descripción de SAOC [SAOC], se representan N señales de objetos de entrada como una matriz **S** del tamaño $N \times N_{\text{Muestras}}$, y la información de mezcla descendente como una matriz **D** del tamaño $M \times N$. Seguidamente se pueden obtener las señales de mezcla descendente como $\mathbf{X} = \mathbf{DS}$.

Ahora se puede dividir la información de mezcla descendente **D** en dos partes

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

para los meta-objetos.

- 5 Como cada columna de la matriz \mathbf{D} corresponde a una señal de objeto de audio original, se pueden obtener las dos matrices de mezcla descendente componentes estableciendo las columnas, que corresponden al otro meta-objeto en cero (suponiendo que no hay ningún objeto original en ambos meta-objetos). En otras palabras, las columnas correspondientes al meta-objeto BGO son fijados en cero en \mathbf{D}_{FGO} y viceversa.
- 10 Estas nuevas matrices de mezcla descendente describen la manera en que se pueden obtener dos meta-objetos a partir de los objetos de entrada, es decir:

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

- 15 y la mezcla descendente real se simplifica a

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

- 20 Se puede considerar asimismo que el decodificador de objetos (por ejemplo, SAOC) intenta reconstruir los meta-objetos:

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

- 25 y la representación específica de DE se puede escribir como una combinación de estas dos reconstrucciones de meta-objetos:

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

- 30 La estimación de sonoridad de objetos recibe los dos meta-objetos \mathbf{S}_{FGO} y \mathbf{S}_{BGO} como entrada y estima la sonoridad de cada uno de ellos: L_{FGO} que es la sonoridad (total/general) de \mathbf{S}_{FGO} , y L_{BGO} que es la sonoridad (total/general) de \mathbf{S}_{BGO} . Estos valores de sonoridad se transmiten y/o almacenan.

- 35 Como alternativa, utilizando uno de los meta-objetos, por ejemplo, el FGO, como referencia, es posible calcular la diferencia de sonoridad de estos dos objetos, por ejemplo, como

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

A continuación, este valor único se transmite y/o almacena.

- 40 La Fig. 8 ilustra un codificador de acuerdo con otra realización. El codificador de la Fig. 8 comprende un mezclador descendente de objetos 811 y un estimador de información secundaria de objetos 812. Además, el codificador de la Fig. 8 comprende asimismo una unidad de codificación de sonoridad de objetos 820. Más aun, el codificador de la Fig. 8 comprende un mezclador de meta-objetos de audio 805.

- 45 El codificador de la Fig. 8 utiliza meta-objetos de audio intermedios como entrada a la estimación de sonoridad de objetos. En ciertas realizaciones, el codificador de la Fig. 8 puede estar configurado para generar dos meta-objetos de audio. En otras realizaciones, el codificador de la Fig. 8 puede estar configurado para generar tres o más meta-objetos de audio.

- 50 Entre otras cosas, los conceptos proporcionados proporcionan una nueva característica, y es que el codificador puede, por ejemplo, estimar la sonoridad promedio de todos los objetos de entrada. Los objetos se pueden mezclar, por ejemplo, para obtener una señal de mezcla descendente que se transmite. Los conceptos proporcionados proporcionan además la nueva característica de que se puede incluir la información de sonoridad de objeto y de mezcla descendente, por ejemplo, en la información secundaria de codificación de objetos que se transmite.

- 55 El decodificador puede utilizar, por ejemplo, la información secundaria de codificación de objetos para la separación (virtual) de los objetos y recombinar los objetos usando la información de representación.

- 60 Además, los conceptos proporcionados presentan la nueva característica de que se puede usar la información de mezcla descendente para estimar la sonoridad de la señal de mezcla descendente por defecto, la información de representación y la sonoridad de objetos recibida para para estimar la sonoridad promedio de la señal de salida, y/o se puede estimar el cambio de sonoridad tomando estos dos valores. O bien, se puede utilizar la información de

mezcla descendente y de representación para estimar el cambio de sonoridad a partir de la mezcla descendente por defecto, otra nueva característica de los conceptos proporcionados.

5 Además, los conceptos proporcionados proporcionan la nueva característica de que se puede modificar la salida del decodificador para compensar el cambio de sonoridad de manera que la sonoridad promedio de la señal modificada se equipare a la sonoridad promedio de la mezcla descendente por defecto.

10 En la Fig. 9 se ilustra una realización particular relacionada con SAOC-DE. El sistema recibe las señales de objetos de audio de entrada, la información de mezcla descendente, y la información sobre el agrupamiento de los objetos en meta-objetos. Basándose en estas, el mezclador de meta-objetos de audio 905 forma los dos meta-objetos S_{FGO} y S_{BGO} . Es posible que la porción de la señal que se procesa con SAOC, no constituya la totalidad de la señal. Por ejemplo, en una configuración de 5.1 canales, se puede implementar SAOC en un subconjunto de canales, como en el canal delantero (izquierdo, derecho y central), mientras que los demás canales (izquierdo de envolvente, derecho de envolvente y efectos de baja frecuencia) se encaminan a rodear (desviando) la SAOC y se entregan como tal.

15 Estos canales no procesados por SAOC se indican como X_{DESVIO} . Es necesario incluir los posibles canales de desviación al codificador para una estimación más precisa de la información de sonoridad.

Los canales de desviación pueden tratarse de diversas maneras.

20 Por ejemplo, los canales de desviación pueden formar, por ejemplo, un meta-objeto independiente. Esto permite definir la representación de manera que los tres meta-objetos se escalan de manera independiente.

O bien, por ejemplo, se puede combinar los canales de desviación, por ejemplo, con uno de los otros dos meta-objetos. Las configuraciones de representación de ese meta-objeto también controlan la porción de canales de desviación. Por ejemplo, en el escenario de mejora de diálogos, puede ser ventajoso combinar los canales de desviación con el meta-objeto de segundo plano: $X_{BCO} = S_{BGO} + X_{DESVIO}$.

25

O, por ejemplo, se pueden ignorar los canales de desviación, por ejemplo.

30 De acuerdo con algunas realizaciones, la unidad de codificación basada en objetos 210 del codificador está configurada para recibir las señales de objetos de audio, en el que cada una de las señales de objetos de audio se asigna exactamente a uno de exactamente dos grupos, en el que cada uno de los exactamente dos grupos comprende una o más de las señales de objetos de audio. Más aun, la unidad de codificación basada en objetos 210 está configurada para efectuar la mezcla descendente de las señales de objetos de audio, que están comprendidas por dichos 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 dichos uno o más canales de mezcla descendente es menor que el número de señales de objetos de audio que están comprendidas por dichos exactamente dos grupos. La unidad de codificación de sonoridad de objetos 220 está asignada para recibir una o más señales de objetos de audio de desvío adicionales, en el que cada una de dichas una o más señales de objetos de audio de desvío adicionales se asigna a un tercer grupo, en el que cada una de dichas una o más señales de objetos de audio de desvío adicionales no está comprendida en el primer grupo y no está comprendida en el segundo grupo, en el que la unidad de codificación basada en objetos 210 está configurada para no efectuar la mezcla descendente de dichas una o más señales de objetos de audio de desvío adicionales dentro de la señal de mezcla descendente.

35

40

45 En una realización, la unidad de codificación 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, indicando el primer valor de sonoridad una sonoridad total de dichas una o más señales de objetos de audio del primer grupo, indicando el segundo valor de sonoridad una sonoridad total de dichas una o más señales de objetos de audio del segundo grupo e indicando el tercer valor de sonoridad una sonoridad total de dichas una o más señales de objetos de audio de desvío adicionales del tercer grupo. En otra realización, la unidad de codificación 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, en el que el primer valor de sonoridad indica una sonoridad total de dichas una o más señales de objetos de audio del primer grupo, y en el que el segundo valor de sonoridad indica una sonoridad total de dichas una o más señales de objetos de audio del segundo grupo y de dichas una o más señales de objetos de audio de desvío adicionales del tercer grupo.

50

55

60 De acuerdo con una realización, la interfaz de recepción 110 del decodificador está configurada para recibir la señal de mezcla descendente. Más aun, la interfaz de recepción 110 está configurada para recibir una o más señales de objetos de audio de desvío adicionales, en el que dichas una o más señales de objetos de audio de desvío 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 contiene información sobre la sonoridad de la señales de objetos de audio que se mezclan en la señal de mezcla descendente y que indica información sobre la sonoridad de dichas una o más señales de objetos de audio de desvío adicionales que no se mezclan en la señal de mezcla

descendente. Más aun, el procesador de señales 120 está configurado para determinar el valor de compensación de sonoridad dependiendo de la información sobre la sonoridad de las señales de objetos de audio que se mezclan en la señal de mezcla descendente, y dependiendo de la información sobre la sonoridad de dichas una o más señales de objetos de audio de desvío adicionales que no se mezclan en la señal de mezcla descendente.

5 La Fig. 9 ilustra un codificador y un decodificador de acuerdo con una realización relacionada con el SAOC-DE, que comprende canales de desviación. Entre otras cosas, el codificador de la Fig. 9 comprende un codificador SAOC 902.

10 En la realización de la Fig. 9, la posible combinación de los canales de desviación con los demás meta-objetos tiene lugar en los dos bloques de "inclusión de desvío" 913, 914, que producen los meta-objetos X_{FGO} y X_{BGO} en el que se incluyen las partes definidas de los canales de desviación.

15 La sonoridad perceptual $L_{DESVÍO}$, L_{FGO} y L_{BGO} de estos dos meta-objetos se estima en las unidades de estimación de sonoridad 921, 922, 923. A continuación se transforma esta información de sonoridad en una codificación apropiada en un estimador de información de sonoridad de meta-objetos 925 y a continuación se transmite y/o almacena.

20 El codificador y el decodificador SAOC reales operan como se espera extrayendo la información secundaria de objetos de los objetos, creando la señal de mezcla descendente X , y almacenando y/o transmitiendo la información al decodificador. Los posibles canales de desviación se transmiten y/o almacenan junto con la demás información al decodificador.

25 El decodificador SAOC-DE 945 recibe un valor de ganancia "Ganancia de diálogo" como entrada del usuario. Basándose en esta entrada y la información de mezcla descendente recibida, el decodificador SAOC 945 determina la información de representación. El decodificador SAOC 945 produce a continuación la escena representada de salida como la señal Y . Además de eso, produce un factor de ganancia (y un valor de retardo) que se aplicaría a las posibles señales de desvío $X_{DESVÍO}$.

30 La unidad de "inclusión de desvío" 955 recibe esta información junto con la escena de salida representada y las señales de desvío y crea la señal de escena de salida total. El decodificador SAOC 945 produce asimismo un conjunto de valores de ganancia de meta-objetos, dependiendo la cantidad de estos del agrupamiento de meta-objetos y de la forma de información de sonoridad pretendida.

35 Los valores de ganancia se envían al estimador de sonoridad de la mezcla 960 que también recibe información de sonoridad de meta-objetos procedente del codificador.

40 A continuación el estimador de sonoridad de la mezcla 960 puede determinar la información de sonoridad deseada, que puede incluir, pero sin limitación, la sonoridad de la señal de mezcla descendente, la sonoridad de la escena de salida representada, y/o la diferencia de sonoridad entre la señal de mezcla descendente y la escena de salida representada.

45 En algunas realizaciones, la información de sonoridad en sí es suficiente, mientras que, en otras realizaciones, es conveniente procesar la salida completa dependiendo de la información de sonoridad determinada. Este procesamiento puede ser, por ejemplo, la compensación de cualquier diferencia posible de sonoridad entre la señal de mezcla descendente y la escena de salida representada. Un procesamiento de este tipo, por ejemplo, mediante una unidad de procesamiento de sonoridad 970, tendría sentido en el escenario de difusión, ya que reduciría los cambios en la sonoridad percibida de la señal independientemente de la interacción con el usuario (configuración de la "ganancia de diálogo" de entrada).

50 El procesamiento relacionado con la sonoridad efectuado en esta realización particular comprende una pluralidad de características nuevas. Entre otras cosas, el FGO, BGO, y los posibles canales de desviación se pre-mezclan en la configuración final de canales para que se pueda realizar la mezcla descendente simplemente agregando las dos señales pre-mezcladas juntas (por ejemplo, coeficientes de matriz de mezcla descendente de 1), que constituye una nueva característica. Más aun, como nueva característica adicional, se estima la sonoridad promedio del FGO y BGO y se calcula la diferencia. Además, se mezclan los objetos para obtener una señal de mezcla descendente que se transmite. Más aun, como nueva característica adicional, se incluye la información de diferencia de sonoridad en la información secundaria que se transmite. (nueva) Además, el decodificador utiliza la información secundaria para la separación (virtual) de los objetos y recombina los objetos usando la información de representación que se basa en la información de mezcla descendente y la ganancia de modificación de entrada de usuario. Más aun, como nueva característica adicional, el decodificador utiliza la ganancia de modificación y la información de sonoridad transmitida para estimar el cambio en la sonoridad promedio de la salida del sistema en comparación con la mezcla descendente por defecto.

A continuación, se presenta una descripción formal de las realizaciones.

Suponiendo que los valores de sonoridad de los objetos se comportan de manera similar al logaritmo de los valores de energía al sumar los objetos, es decir, que los valores de sonoridad se deben transformar al dominio lineal, sumarse ahí y por último transformarse nuevamente al dominio logarítmico. Se presentará ahora la motivación de esto por medio de la definición de la medida de sonoridad BS.1770 (para simplificar, se establece un número de canales de uno, aunque se puede aplicar el mismo principio a señales multicanal con la suma apropiada de los canales adicionales).

La sonoridad de la señal de orden i filtrada con $K z_i$ en la que la energía cuadrática media e_i se define como

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

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

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

La energía de la suma de N señales no correlacionadas
$$z_{SUM} = \sum_{i=1}^N z_i$$
 es entonces

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

y la sonoridad de esta señal de suma es entonces

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

Si las señales no están correlacionadas, se deben tener en cuenta los coeficientes de correlación C_{ij} al aproximar la energía de la señal sumada de la siguiente manera

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

en el que la energía cruzada e_{ij} entre los objetos de orden i y j se define de la siguiente manera

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

en el que $-1 \leq C_{ij} \leq 1$ 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.

Extendiendo aún más el modelo con las ponderaciones de mezcla g_i que se han de aplicar a las señales en el

proceso de mezcla, es decir,
$$z_{SUM} = \sum_{i=1}^N g_i z_i,$$
 la energía de la señal sumada será

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

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

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

La diferencia entre la sonoridad de las dos señales se puede estimar de la siguiente manera

5
$$\Delta L(i, j) = L_i - L_j .$$

Si ahora se utiliza la definición de sonoridad como anteriormente, esto se puede escribir de la siguiente manera

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

10 que puede observarse en función de las energías de señal. Si ahora se desea estimar la diferencia de sonoridad entre dos mezclas

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

15 posiblemente con ponderaciones de mezcla diferentes g_i y h_i , esto se puede estimar calculando:

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

20 En 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

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

25 A continuación se considera la codificación diferencial.

Es posible codificar los valores de sonoridad por objeto como diferencias con respecto a la sonoridad de un objeto de referencia elegido:

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

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

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

10 o en el caso de objetos no correlacionados

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

15 A continuación, se considera un enfoque de mejora de diálogos.

Consideremos, una vez más, el escenario de aplicación de mejora de diálogos. La libertad para definir la información de representación en el decodificador se limita solo a los cambios de nivel de los dos meta-objetos. Supongamos, además, que los dos meta-objetos no están correlacionados, es decir, $C_{FGO, BGO} = 0$. Si las ponderaciones de mezcla descendente de los meta-objetos son h_{FGO} y h_{BGO} , y se representarían con las ganancias f_{FGO} y f_{BGO} , la sonoridad de la salida con respecto a la mezcla descendente por defecto es

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

25 Esta es entonces también la compensación requerida si se desea tener la misma sonoridad en la salida que en la mezcla descendente por defecto.

$\Delta L(A, B)$ se puede considerar como un valor de compensación de sonoridad, que puede transmitirse por el procesador de señales 120 del decodificador. $\Delta L(A, B)$ se puede nombrar asimismo como valor de cambio de sonoridad y, por consiguiente, el valor de compensación real puede ser un valor inverso. ¿O también es aceptable usar el nombre "factor de compensación de sonoridad" para el mismo? De esa manera, el valor de compensación de sonoridad lc_v mencionado anteriormente en este documento correspondería al siguiente valor g_{Delta} .

35 Por ejemplo, se puede aplicar $g_{\Delta} = 10^{-\Delta L(A, B)/20} 1 / \Delta L(A, B)$ como un factor de multiplicación a cada canal de una señal de audio modificada que surja como resultado de la aplicación de la información de representación a la señal de entrada de audio. Esta ecuación g_{Delta} actúa en el dominio lineal. En el dominio logarítmico, la ecuación sería diferente, tal como $1 / \Delta L(A, B)$ y se aplicaría de manera correspondiente.

40 Si el proceso de mezcla descendente se simplifica de tal manera que los dos meta-objetos se puedan mezclar con ponderaciones unitarias para obtener la señal de mezcla descendente, es decir, $h_{FGO} = h_{BGO} = 1$, y ahora las ganancias de representación correspondientes a estos dos objetos están indicadas con g_{FGO} y g_{BGO} . Esto simplifica la ecuación de cambio de sonoridad en

$$\begin{aligned} \Delta L(A, B) &= 10 \log_{10} \frac{g_{FGO}^2 10^{(L_{FGO}-\epsilon)^{n_0}} + g_{BGO}^2 10^{(L_{BGO}-\epsilon)^{n_0}}}{10^{(L_{FGO}-\epsilon)^{n_0}} + 10^{(L_{BGO}-\epsilon)^{n_0}}} \\ &= 10 \log_{10} \frac{g_{FGO}^2 10^{L_{FGO}^{n_0}} + g_{BGO}^2 10^{L_{BGO}^{n_0}}}{10^{L_{FGO}^{n_0}} + 10^{L_{BGO}^{n_0}}} \end{aligned}$$

Una vez más, se puede considerar $\Delta L(A, B)$ como un valor de compensación de sonoridad determinado por el procesador de señales 120.

En general, se puede considerar que g_{FGO} es una ganancia de representación correspondiente al objeto en primer plano FGO (grupo de objetos de primer plano), y se puede considerar que g_{BGO} es una ganancia de representación correspondiente al objeto de segundo plano BGO (grupo de objetos de segundo plano).

Como se ha mencionado anteriormente, es posible transmitir diferencias de sonoridad en lugar de la sonoridad absoluta. Definamos la sonoridad de referencia como sonoridad del meta-objeto 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

$$\Delta L(A, B) = 10 \log_{10} \frac{g_{FGO}^2 + g_{BGO}^2 10^{K_{BGO}/n_0}}{1 + 10^{K_{BGO}/n_0}}$$

También puede ocurrir, como en el caso de la SAOC-DE, que dos meta-objetos no tengan factores de escala individuales, pero que uno de los objetos quede sin modificar, mientras que el otro se atenúa para obtener la relación de mezcla correcta entre los objetos. En esta configuración de representación, la salida ha de ser más baja en sonoridad que la mezcla por defecto, y el cambio en la sonoridad es

$$\Delta L(A, B) = 10 \log_{10} \frac{\hat{g}_{FGO}^2 + \hat{g}_{BGO}^2 10^{K_{BGO}/n_0}}{1 + 10^{K_{BGO}/n_0}},$$

con

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

Esta forma ya es bastante sencilla, y bastante independiente con respecto a la medida de sonoridad empleada. El único requisito real es que los valores de sonoridad deben sumarse en el dominio exponencial. Es posible transmitir/almacenar los valores de las energías de señal en lugar de los valores de sonoridad, ya que entre los dos hay una estrecha conexión.

En cada una de las fórmulas anteriores, $\Delta L(A, B)$ se puede considerar un valor de compensación de sonoridad, que puede transmitirse por el procesador de señales 120 del decodificador.

A continuación, se considera un ejemplo. Se ilustra la precisión de los conceptos proporcionados por medio de dos señales ejemplificativas. Ambas señales tienen una mezcla descendente 5.1 con los canales de envolvente y LFE desviados del procesamiento de SAOC.

Se utilizan dos enfoques principales: uno ("de 3 términos") con tres meta-objetos: FGO, BGO y canales de desviación, por ejemplo,

$$\mathbf{X} = \mathbf{X}_{FGO} + \mathbf{X}_{BGO} + \mathbf{X}_{DES\text{V}IO}.$$

Y otro ("de 2 términos") con dos meta-objetos, por ejemplo:

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

En el enfoque de 2 términos, los canales de desviación se pueden mezclar, por ejemplo, junto con el BGO para la estimación de la sonoridad de los meta-objetos. Se estima la sonoridad de ambos (o de los tres) objetos, así como la sonoridad de la señal de mezcla descendente, y se guardan los valores.

5 Las instrucciones de representación están en la forma

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

y

10

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

para los dos enfoques, respectivamente.

15 Los valores de ganancia se determinan, por ejemplo, de acuerdo con:

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

en el que la ganancia del FGO g_{FGO} se varía de -24 a +24 dB.

20

Se representa el escenario de salida, se mide la sonoridad y se calcula la atenuación de la sonoridad de la señal de mezcla descendente.

25

Este resultado se visualiza en la Fig. 10 y en la Fig. 11 con la línea azul con marcadores circulares. La Fig. 10 presenta una primera ilustración y la Fig. 11 presenta una segunda ilustración de un cambio de sonoridad medido y el resultado del uso de los conceptos proporcionados para estimar el cambio de sonoridad de manera puramente paramétrica.

30

A continuación se estima la atenuación de la mezcla descendente paramétricamente empleando los valores de sonoridad de los meta-objetos almacenados y la información de mezcla descendente y representación. Se ilustra la estimación utilizando la sonoridad de tres meta-objetos con la línea verde con marcadores cuadrados y se ilustra la estimación que utiliza la sonoridad de dos meta-objetos con la línea roja con marcadores en forma de estrella.

35

Se puede observar por las figuras que los enfoques de 2 y 3 términos producen resultados prácticamente idénticos, y que ambos aproximan el valor medido con bastante eficacia.

40

Los conceptos proporcionados muestran 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 integran la mezcla. El beneficio de esto es que la sonoridad de las señales componentes se puede estimar una vez y que se puede obtener la estimación de la sonoridad de la señal de mezcla paramétricamente respecto de cualquier mezcla sin necesidad de estimación de sonoridad basada en la señal real. Esto constituye una mejora considerable en la eficiencia informática del sistema en su conjunto en el cual es necesaria la estimación de la sonoridad de diversas mezclas. Por ejemplo, al cambiar el usuario final la configuración de representación, de inmediato está disponible la estimación de la sonoridad de la salida.

45

En algunas aplicaciones, como en el caso de adaptarse a la recomendación EBU R128, es importante la sonoridad promedio de la totalidad del programa. Si la estimación de sonoridad en el receptor, por ejemplo, en un escenario de transmisión, se realiza basándose en la señal recibida, la estimación converge hacia la sonoridad promedio solo una vez que se ha recibido la totalidad del programa. Debido a esto, cualquier compensación de la sonoridad ha de contener errores o mostrar variaciones temporales. Cuando se estima la sonoridad de los objetos componentes según lo propuesto y se transmite la información de sonoridad, es posible estimar la sonoridad promedio de la mezcla en el receptor sin retardo.

50

55

Si se desea que la sonoridad promedio de la señal de salida se mantenga (aproximadamente) constante independientemente de los cambios en la información de representación, los conceptos proporcionados permiten determinar un factor de compensación por esta razón. Los cálculos necesarios para esto en el decodificador son insignificantes desde el punto de su complejidad computacional y por ello es posible agregar la funcionalidad a cualquier decodificador.

Hay casos en que el nivel de sonoridad absoluto de la salida no es importante, sino que la importancia reside en determinar el cambio de sonoridad con respecto a una escena de referencia. En tales casos, los niveles absolutos de los objetos no son importantes, aunque sus niveles relativos sí lo son. Esto permite definir uno de los objetos como objeto de referencia y representar la sonoridad de los demás objetos en relación con la sonoridad de este objeto de referencia. Esto ofrece algunos beneficios, teniendo en cuenta el transporte y/o almacenamiento de la información de sonoridad.

En primer lugar, no es necesario transportar el nivel de sonoridad de referencia. En el caso de aplicación de los dos meta-objetos, esto divide por la mitad la cantidad de datos a transmitir. El segundo beneficio está relacionado con la posible cuantificación y representación de los valores de sonoridad. Dado que los niveles absolutos de los objetos pueden ser casi cualquiera, los valores de sonoridad absoluta también pueden ser casi cualquiera. Se supone que los valores relativos de sonoridad, por otra parte, tienen una media 0 y una distribución bien formada alrededor de la media. La diferencia entre las representaciones permite definir la cuadrícula de cuantificación de la representación relativa de manera que tenga una precisión potencialmente mayor con el mismo número de bits utilizado para la representación cuantificada.

La Fig. 12 ilustra otra realización para realizar la compensación de sonoridad. En la Fig. 12, se puede llevar a cabo la compensación de sonoridad, por ejemplo, para compensar la pérdida de sonoridad. Para este fin, por ejemplo, se pueden utilizar los valores **DE_loudness_diff_dialogue** (= K_{FGO}) y **DE_loudness_diff_background** (= K_{BGO}) de **DE_control_info**. En este caso, **DE_control_info** puede especificar la información de control de "Mejora de Diálogos" (DE) de Audio Limpio Avanzado.

La compensación de sonoridad se obtiene aplicando un valor de ganancia "g" a la señal de salida de SAOC-DE y a los canales de desviación (en el caso de una señal multicanal).

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

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

El bloque "Estimador de sonoridad de salida" 1230 utiliza la información de sonoridad K_{FGO} y K_{BGO} , y los valores de ganancia efectiva m_{FGO} y m_{BGO} para estimar este posible cambio de la sonoridad en comparación con el caso de la mezcla descendente por defecto. A continuación se mapea el cambio con el "Factor de compensación de sonoridad" que se aplica a los canales de salida para producir las "Señales de salida" finales.

Se aplican las siguientes etapas para la compensación de sonoridad:

- Recibir el valor de ganancia limitada m_G del decodificador SAOC-DE (según lo definido en el artículo 12.8 "Modification range control for SAOC-DE" [DE] (Control del rango de Modificación para SAOC-DE)), y determinar las ganancias de FGO/BGO aplicadas:

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

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

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

- Calcular la ganancia por 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 \mathbf{g} , en los que
- $$g_i = \begin{cases} g_\Delta & \text{si canal } i \text{ pertenece a salida de SAOC-DE} \\ m_{BGO} g_\Delta & \text{si canal } i \text{ es un canal de desvío} \end{cases}$$
- , y N es el número total de canales de

5 salida. En la Fig. 12, el ajuste de ganancia se divide en dos etapas: se ajusta la ganancia de los posibles “canales de desviación” con m_{BGO} antes de combinarlos con los “canales de salida de SAOC-DE”, y a continuación se aplica una ganancia común g_Δ a todos los canales combinados. Esto solo es posible reordenando las operaciones de ajuste de ganancia, mientras que \mathbf{g} en este caso combina ambas etapas de ajuste de ganancia en un solo ajuste de ganancia.

- 10 - Aplicar los valores de escala \mathbf{g} a los canales de audio \mathbf{Y}_{TOTAL} que consisten en los “canales de salida de SAOC-DE” \mathbf{Y}_{SAOC} y los posibles “canales de desviación” alineados en el tiempo $\mathbf{Y}_{DESUDIO}$: $\mathbf{Y}_{TOTAL} = \mathbf{Y}_{SAOC} \cup \mathbf{Y}_{DESUDIO}$.

La aplicación de los valores de escala \mathbf{g} a los canales de audio \mathbf{Y}_{TOTAL} se realiza por la unidad de ajuste de ganancia 1240.

15 Se puede considerar a ΔL como se ha calculado anteriormente como un valor de compensación de sonoridad. En general, m_{FGO} indica una ganancia de representación para el objeto en primer plano FGO (grupo de objetos de primer plano) y m_{BGO} indica una ganancia de representación para el objeto de segundo plano BGO (grupo de objetos de segundo plano).

20 Si bien se han descrito algunos aspectos en el contexto de un aparato, es obvio que estos aspectos también representan una descripción del método correspondiente, en el cual un bloque o dispositivo corresponde a una etapa del método o a una característica de una etapa del método. De manera análoga, los aspectos descritos en el contexto de una etapa del método también representan una descripción de un bloque o elemento correspondiente o de una característica de un aparato correspondiente.

25 La señal descompuesta de la presente invención puede almacenarse en un medio digital o se puede transmitir por un medio de transmisión tal como un medio de transmisión inalámbrico o un medio de transmisión por cables tal como Internet.

30 Dependiendo de ciertos requisitos de implementación, las realizaciones de la invención pueden implementarse en hardware o en software. La implementación se puede realizar empleando un medio de almacenamiento digital, por ejemplo un disco flexible, un DVD, un CD, una ROM, una PROM, una EPROM, una EEPROM o una memoria FLASH, que tiene almacenadas en la misma señales control legibles electrónicamente, que cooperan (o tienen capacidad para cooperar) con un sistema informático programable de tal manera que se ejecute el método respectivo.

Algunas realizaciones de acuerdo con la invención comprenden un portador de datos no transitorio que tiene señales de control legibles electrónicamente, con capacidad para cooperar con un sistema informático programable de tal manera que se ejecute uno de los métodos descritos en el presente documento.

40 En general, las realizaciones de la presente invención pueden implementarse en forma de producto de programa informático con un código de programa, en el que el código de programa cumple la función de ejecutar uno de los métodos al ejecutarse el programa informático en una computadora. El código de programa puede almacenarse, por ejemplo, en un portador legible por una máquina.

45 Otras realizaciones comprenden el programa informático para realizar uno de los métodos descritos en el presente documento, almacenado en un portador legible por una máquina.

50 En otras palabras, una realización del método de la invención consiste, por lo tanto, en un programa informático que consta de un código de programa para realizar uno de los métodos descritos en el presente documento al ejecutarse el programa informático en una computadora.

55 Otra realización de los métodos de la invención consiste, por lo tanto, en un portador de datos (o medio de almacenamiento digital, o medio legible por computadora) que comprende, grabado en el mismo, el programa informático para ejecutar uno de los métodos descritos en el presente documento.

Otra realización del método de la invención es, por lo tanto, un flujo de datos o una secuencia de señales que representa el programa informático para ejecutar uno de los métodos descritos en el presente documento. El flujo de

datos o la secuencia de señales pueden estar configurados, por ejemplo, para transferirse a través de una conexión de comunicación de datos, por ejemplo por la Internet.

5 Otra realización comprende un medio de procesamiento, por ejemplo una computadora, o un dispositivo lógico programable, configurado o adaptado para ejecutar uno de los métodos descritos en el presente documento.

Otra realización comprende una computadora en la que se ha instalado el programa informático para ejecutar uno de los métodos descritos en el presente documento.

10 En algunas realizaciones, se puede utilizar un dispositivo lógico programable (por ejemplo un campo de matrices de puertas programables) para ejecutar algunas o todas las funcionalidades de los métodos descritos en el presente documento. En algunas realizaciones, un campo de matrices de puertas programables puede cooperar con un microprocesador para ejecutar uno de los métodos descritos en el presente documento. Por lo general, los métodos se ejecutan preferentemente por cualquier aparato de hardware.

15 Referencias

[BCC] C. Faller y F. Baumgarte, "Binaural Cue Coding - Part II: Schemes and applications," IEEE Trans. on Speech and Audio Proc., vol. 11, n.º 6, noviembre de 2003.

20 [EBU] EBU Recommendation R 128 "Loudness normalization and permitted maximum level of audio signals", Ginebra, 2011.

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

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

30 [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.

[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.

35 [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.

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

[ITU] International Telecommunication Union: "Recommendation ITU-R BS.1770-3 - Algorithms to measure audio programme loudness and true-peak audio level", Ginebra, 2012.

45 [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, RU, abril de 2007.

50 [SAOC2] J. Engdegård, B. Resch, C. Falch, O. Hellmuth, J. Hilpert, A. Hölzer, 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) Norma internacional 23003-2.

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

60 [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] Documento WO 2008/035275 A2.

[SCH] Documento EP 2 146 522 A1.

[ENG] Documento 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) para recibir 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 representación que indica si una o más de las señales de objetos de audio deberá amplificarse o atenuarse, y
 10 un procesador de señales (120) para generar el uno o más canales de salida de audio de la señal de salida de audio, en el que el procesador de señales (120) está configurado para determinar un valor de compensación de sonoridad dependiendo de la información de sonoridad y dependiendo de la información de representación, y en el que el procesador de señales (120) está configurado para generar el uno o más canales de salida de audio de la señal de salida de audio a partir de la señal de entrada de audio dependiendo de la información de representación y dependiendo del valor de compensación de sonoridad,
 15 en el que el procesador de señales (120) está configurado para generar el uno o más canales de salida de audio de la señal de salida de audio a partir de la señal de entrada de audio dependiendo de la información de representación y dependiendo del valor de compensación de sonoridad, de tal manera que una sonoridad de la señal de salida de audio sea igual a una sonoridad de la señal de entrada de audio, o de tal manera que la sonoridad de la señal de salida de audio se aproxime más a la sonoridad de la señal de entrada de audio que una sonoridad de una señal de audio modificada que se produciría como resultado de la modificación de la señal de entrada de audio mediante la amplificación o atenuación de las señales de objetos de audio de la señal de entrada de audio de acuerdo con la información de representación.

25 2. Un decodificador de acuerdo con la reivindicación 1, en el que el procesador de señales (120) está configurado para generar la señal de audio modificada mediante la modificación de la señal de entrada de audio por la amplificación o atenuación de las señales de objetos de audio de la señal de entrada de audio de acuerdo con la información de representación, y en el que el procesador de señales (120) está configurado para generar la señal de salida de audio mediante la aplicación de valor de compensación de sonoridad a la señal de audio modificada, de tal 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 tal manera que la sonoridad de la señal de salida de audio se aproxime más a la sonoridad de la señal de entrada de audio que la sonoridad de la señal de audio modificada.

35 3. Un decodificador de acuerdo con la reivindicación 1 o 2, en el que 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 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, en el que la interfaz de recepción (110) está configurada para recibir un valor de sonoridad por cada grupo de los dos o más grupos como la información de sonoridad,
 40 en el que el procesador de señales (120) está configurado para determinar el valor de compensación de sonoridad dependiendo del valor de sonoridad de cada uno de los dos o más grupos, y en el que el procesador de señales (120) está configurado para generar el uno o más canales de salida de audio de la señal de salida de audio a partir de la señal de entrada de audio dependiendo del valor de compensación de sonoridad.

4. Un decodificador de acuerdo con la reivindicación 3, en el que al menos un grupo de los dos o más grupos comprende dos o más de las señales de objetos de audio.

50 5. Un decodificador de acuerdo con las reivindicaciones 1 o 2, en el que cada una de las señales de objetos de audio de la señal de entrada de audio se asigna a exactamente un grupo de más de dos grupos, en el que cada uno de los más de dos grupos comprende una o más de las señales de objetos de audio de la señal de entrada de audio, en el que la interfaz de recepción (110) está configurada para recibir un valor de sonoridad por cada grupo de los más de dos grupos como la información de sonoridad,
 55 en el que el procesador de señales (120) está configurado para determinar el valor de compensación de sonoridad dependiendo del valor de sonoridad de cada uno de los más de dos grupos, y en el que el procesador de señales (120) está configurado para generar el uno o más canales de salida de audio de la señal de salida de audio a partir de la señal de entrada de audio dependiendo del valor de compensación de sonoridad.

60 6. Un decodificador de acuerdo con la reivindicación 5, en el que al menos un grupo de los más de dos grupos comprende dos o más de las señales de objetos de audio.

7. Un decodificador de acuerdo con una de las reivindicaciones 3 a 6, en el que el procesador de señales (120) está configurado para determinar el valor de compensación de sonoridad de acuerdo con la fórmula

$$\Delta L = 10 \log_{10} \frac{\sum_{i=1}^N g_i^2 10^{(L_i - c)/10}}{\sum_{i=1}^N h_i^2 10^{(L_i - c)/10}}$$

5

o de acuerdo con la fórmula

$$\Delta L = 10 \log_{10} \frac{\sum_{i=1}^N g_i^2 10^{L_i/10}}{\sum_{i=1}^N h_i^2 10^{L_i/10}}$$

10

en la que ΔL es el valor de compensación de sonoridad,
 en la que i indica una señal de objeto de audio de orden i de las señales de objetos de audio,
 en la que L_i es la sonoridad de la señal de objeto de audio de orden i ,
 en la que g_i es una primera ponderación de mezcla para la señal de objeto de audio de orden i ,
 en la que h_i es una segunda ponderación de mezcla para la señal de objeto de audio de orden i ,
 en la que c es un valor constante, y
 en la que N es un número.

15

8. Un decodificador de acuerdo con una de las reivindicaciones 3 a 6, en el que el procesador de señales (120) está configurado para determinar el valor de compensación de sonoridad de acuerdo con la fórmula

$$\Delta L = 10 \log_{10} \frac{\sum_{i=1}^N g_i^2 10^{K_i/10}}{\sum_{i=1}^N h_i^2 10^{K_i/10}}$$

25

en la que ΔL es el valor de compensación de sonoridad,
 en la que i indica una señal de objeto de audio de orden i de las señales de objetos de audio,
 en la que g_i es una primera ponderación de mezcla para la señal de objeto de audio de orden i ,
 en la que h_i es una segunda ponderación de mezcla para la señal de objeto de audio de orden i ,
 en la que N es un número, y
 en la que K_i se define de acuerdo con

30

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

35

en la que L_i es una sonoridad de la señal de objeto de audio de orden i , y
 en la que L_{REF} es la sonoridad de un objeto de referencia.

9. Un decodificador de acuerdo con la reivindicación 3 o 4, en el que cada una de las señales de objetos de audio de la señal de entrada de audio se asigna a exactamente un grupo de exactamente dos grupos como los dos o más grupos,

40

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 de primer plano de los exactamente dos grupos o a un grupo de objetos de segundo plano de los exactamente dos grupos,
 en el que la interfaz de recepción (110) está configurada para recibir el valor de sonoridad del grupo de objetos de primer plano,

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

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

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

10. Un decodificador de acuerdo con la reivindicación 9, en el que el procesador de señales (120) está configurado para determinar el valor de compensación de sonoridad de acuerdo con la fórmula

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

en la que ΔL es el valor de compensación de sonoridad,

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

en la que K_{BGO} indica el valor de sonoridad del grupo de objetos de segundo plano,

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

en la que m_{BGO} indica una ganancia de representación del grupo de objetos de segundo plano.

11. Un decodificador de acuerdo con la reivindicación 9, en el que el procesador de señales (120) está configurado para determinar el valor de compensación de sonoridad de acuerdo con 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 la que ΔL es el valor de compensación de sonoridad,

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

en la que L_{BGO} indica el valor de sonoridad del grupo de objetos de segundo plano,

en la que g_{FGO} indica una ganancia de representación del grupo de objetos de primer plano, y

en la que g_{BGO} indica una ganancia de representación del grupo de objetos de segundo plano.

12. Un decodificador de acuerdo con una de las reivindicaciones anteriores,

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 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 descendente, y

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

13. Un decodificador de acuerdo con la reivindicación 12,

en el que la interfaz de recepción (110) está configurada para recibir una o más señales de objetos de audio de desvío adicionales, en el que la una o más señales de objetos de audio de desvío 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 en la señal de mezcla descendente y que indica información sobre la sonoridad de la una o más señales de desvío de objetos de audio adicionales que no se mezclan en la señal de mezcla descendente, y

en el que el procesador de señales (120) está configurado para determinar el valor de compensación de sonoridad dependiendo de la información sobre la sonoridad de las señales de objetos de audio que se mezclan en la señal de mezcla descendente, y dependiendo de la información sobre la sonoridad de la una o más señales de desvío de objetos de audio adicionales que no se mezclan en la señal de mezcla descendente.

14. 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) para recibir 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 representación que indica si una o más de las señales de objetos de audio deberá amplificarse o atenuarse, y
 10 un procesador de señales (120) para generar el uno o más canales de salida de audio de la señal de salida de audio, en el que el procesador de señales (120) está configurado para determinar un valor de compensación de sonoridad dependiendo de la información de sonoridad y dependiendo de la información de representación, y en el que el procesador de señales (120) está configurado para generar el uno o más canales de salida de audio de la señal de salida de audio a partir de la señal de entrada de audio dependiendo de la información de representación y dependiendo del valor de compensación de sonoridad, 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 señales de objetos de audio,
 20 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 descendente, y en el que el procesador de señales (120) está configurado para generar el uno o más canales de salida de audio de la señal de salida de audio a partir de la señal de entrada de audio dependiendo de la información de mezcla descendente, dependiendo de la información de representación y dependiendo del valor de compensación de sonoridad,
 25 en el que la interfaz de recepción (110) está configurada para recibir una o más señales de objetos de audio de desvío adicionales, en el que la una o más señales de objetos de audio de desvío adicionales no se mezclan dentro de la señal de mezcla descendente,
 30 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 en la señal de mezcla descendente y que indica información sobre la sonoridad de la una o más señales de desvío de objetos de audio adicionales que no se mezclan en la señal de mezcla descendente, y en el que el procesador de señales (120) está configurado para determinar el valor de compensación de sonoridad dependiendo de la información sobre la sonoridad de las señales de objetos de audio que se mezclan en la señal de mezcla descendente, y dependiendo de la información sobre la sonoridad de la una o más señales de desvío de objetos de audio adicionales que no se mezclan en la señal de mezcla descendente.

15. Un codificador que comprende,

40 una unidad de codificación basada en objetos (210; 710) para codificar 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 de codificación de sonoridad de objetos (220; 720; 820) para codificar información de sonoridad sobre 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 del uno o
 45 más valores de sonoridad depende de una o más de las señales de objetos de audio, en el que 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, en el que al menos un grupo de los dos o más grupos comprende dos o más de las señales de objetos de audio,
 50 en el que la unidad de codificación de sonoridad de objetos (220; 720; 820) está configurada para determinar el uno o más valores de sonoridad de la información de sonoridad mediante la determinación de un valor de sonoridad por cada grupo de los dos o más grupos, en el que dicho valor de sonoridad de dicho grupo indica una sonoridad total de la una o más señales de objetos de audio de dicho grupo.

55 16. Un codificador, que comprende:

una unidad de codificación basada en objetos (210; 710) para codificar 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 de codificación de sonoridad de objetos (220; 720; 820) para codificar información de sonoridad sobre las señales de objetos de audio,
 60 en el que la información de sonoridad comprende uno o más valores de sonoridad, en el que cada uno del uno o más valores de sonoridad depende de una o más de las señales de objetos de audio, en el que la unidad de codificación basada en objetos (210; 710) está configurada para recibir las señales de objetos de audio, en el que cada una de las señales de objetos de audio se asigna exactamente a uno de

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 al menos un grupo de los exactamente dos grupos comprende dos o más de las señales de objetos de audio,

en el que la unidad de codificación basada en objetos (210; 710) está configurada para efectuar la mezcla descendente de las señales de objetos de audio, que están comprendidas por 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 del uno o más canales de mezcla descendente es menor que el número de las señales de objetos de audio que están comprendidas por los exactamente dos grupos,

en el que la unidad de codificación de sonoridad de objetos (220; 720; 820) está configurada para recibir una o más señales de desvío de objetos de audio adicionales, en el que cada una de la una o más señales de desvío de objetos de audio adicionales se asigna a un tercer grupo, en el que cada una de la una o más señales de desvío de objetos de audio adicionales no está comprendida en el primer grupo y no está comprendida en el segundo grupo, en el que la unidad de codificación basada en objetos (210; 710) está configurada para no efectuar la mezcla descendente de la una o más señales de desvío de objetos de audio adicionales dentro de la señal de mezcla descendente, y

en el que la unidad de codificación 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, indicando el primer valor de sonoridad una sonoridad total de la una o más señales de objetos de audio del primer grupo, indicando el segundo valor de sonoridad una sonoridad total de la una o más señales de objetos de audio del segundo grupo, e indicando el tercer valor de sonoridad una sonoridad total de la una o más señales de desvío de objetos de audio adicionales del tercer grupo, o está configurada para determinar un primer valor de sonoridad y un segundo valor de sonoridad de la información de sonoridad, indicando el primer valor de sonoridad una sonoridad total de la una o más señales de objetos de audio del primer grupo, e indicando el segundo valor de sonoridad una sonoridad total de la una o más señales de objetos de audio del segundo grupo y de la una o más señales de desvío de objetos de audio adicionales del tercer grupo.

17. Un sistema que comprende:

un codificador (310) que comprende:

una unidad de codificación basada en objetos (210; 710) para codificar 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 de codificación de sonoridad de objetos (220; 720; 820) para codificar información de sonoridad sobre 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 del uno o más valores de sonoridad depende de una o más de las señales de objetos de audio,

un decodificador (320) de acuerdo con una de las reivindicaciones 1 a 14, para generar 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

en el que el decodificador (320) está configurado para recibir adicionalmente información de representación,

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

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

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

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

recibir información de sonoridad sobre las señales de objetos de audio,

recibir información de representación que indica cómo una o más de las señales de objetos de audio deberá amplificarse o atenuarse,

determinar un valor de compensación de sonoridad dependiendo de la información de sonoridad y dependiendo de la información de representación, y

generar el uno o más canales de salida de audio de la señal de salida de audio a partir de la señal de entrada de audio dependiendo de la información de representación y dependiendo del valor de compensación de sonoridad,

en el que generar el uno o más canales de salida de audio de la señal de salida de audio a partir de la señal de entrada de audio se realiza dependiendo de la información de representación y dependiendo 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 de tal manera que la sonoridad de la señal de salida de audio se

aproxime más a la sonoridad de la señal de entrada de audio que una sonoridad de una señal de audio modificada que se produciría como resultado de la modificación de la señal de entrada de audio mediante la amplificación o atenuación de las señales de objetos de audio de la señal de entrada de audio de acuerdo con la información de representación.

5 19. Un método para generar una señal de salida de audio que comprende uno o más canales de salida de audio, en el que el método comprende:

10 recibir una señal de entrada de audio que comprende una pluralidad de señales de objetos de audio, en el que recibir la señal de audio de entrada se realiza recibiendo 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 señales de objetos de audio,

15 recibir información de representación que indica si una o más de las señales de objetos de audio deberá amplificarse o atenuarse, 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, y recibir una o más señales de objetos de audio de desvío adicionales, en el que la una o más señales de objetos de audio de desvío adicionales no se mezclan dentro de la señal de mezcla descendente,

20 recibir información de sonoridad sobre las señales de objetos de audio, en el que la información de sonoridad indica información sobre la sonoridad de las señales de objetos de audio que se mezclan en la señal de mezcla descendente e indica información sobre la sonoridad de la una o más señales de objetos de audio de desvío adicionales que no se mezclan en la señal de mezcla descendente, y determinar un valor de compensación de sonoridad dependiendo de la información sobre la sonoridad y dependiendo de la información sobre la representación, en el que la determinación del valor de compensación de sonoridad se realiza dependiendo de la información sobre la sonoridad de las señales de objetos de audio que se mezclan en la señal de mezcla descendente, y dependiendo de la información sobre la sonoridad de la una o más señales de objetos de audio de desvío adicionales que no se mezclan en la señal de mezcla descendente, y generar el uno o más canales de salida de audio de la señal de salida de audio a partir de la señal de entrada de audio dependiendo de la información de mezcla descendente, dependiendo de la información de representación y dependiendo del valor de compensación de sonoridad.

20. Un método para codificar, que comprende:

35 codificar 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 determinar información de sonoridad sobre 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 del uno o más valores de sonoridad depende de una o más de las señales de objetos de audio

40 en el que determinar el uno o más valores de sonoridad de la información de sonoridad se realiza 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 total de la una o más señales de objetos de audio de dicho grupo codificar la información de sonoridad en las señales de objetos de audio,

45 en el que 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, en el que al menos un grupo de los dos o más grupos comprende dos o más de las señales de objetos de audio.

21. Un método para codificar, que comprende:

50 recibir las señales de objetos de audio, en el que cada una de las señales de objetos de audio se asigna exactamente a uno de 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 al menos un grupo de los exactamente dos grupos comprende dos o más de las señales de objetos de audio,

55 codificar la pluralidad de señales de objetos de audio para obtener una señal de audio codificada que comprende la pluralidad de señales de objetos de audio efectuando la mezcla descendente de las señales de objetos de audio, que están comprendidas por 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 del uno o más canales de mezcla descendente es menor que el número de las señales de objetos de audio que están comprendidas por los exactamente dos grupos,

60 determinar información de sonoridad sobre 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 del uno o más valores de sonoridad depende de una o más de las señales de objetos de audio,

- 5 determinando un primer valor de sonoridad, un segundo valor de sonoridad y un tercer valor de sonoridad de la información de sonoridad, indicando el primer valor de sonoridad una sonoridad total de la una o más señales de objetos de audio del primer grupo, indicando el segundo valor de sonoridad una sonoridad total de la una o más señales de objetos de audio del segundo grupo, e indicando el tercer valor de sonoridad una sonoridad total de la una o más señales de desvío de objetos de audio adicionales del tercer grupo, o
- 10 determinando un primer valor de sonoridad y un segundo valor de sonoridad de la información de sonoridad, indicando el primer valor de sonoridad una sonoridad total de la una o más señales de objetos de audio del primer grupo, e indicando el segundo valor de sonoridad una sonoridad total de la una o más señales de objetos de audio del segundo grupo y de la una o más señales de desvío de objetos de audio adicionales del tercer grupo,
- 15 codificar la información de sonoridad en las señales de objetos de audio, recibir una o más señales de desvío de objetos de audio adicionales, en el que cada una de la una o más señales de desvío de objetos de audio adicionales se asigna a un tercer grupo, en el que cada una de la una o más señales de desvío de objetos de audio adicionales no está comprendida por el primer grupo y no está comprendida por el segundo grupo, y no efectuar la mezcla descendente de la una o más señales de desvío de objetos de audio adicionales en la señal de mezcla descendente.
- 20 22. Un programa informático para implementar el método de acuerdo con la reivindicación 18 o 21 cuando se ejecuta en una computadora o un procesador de señales.

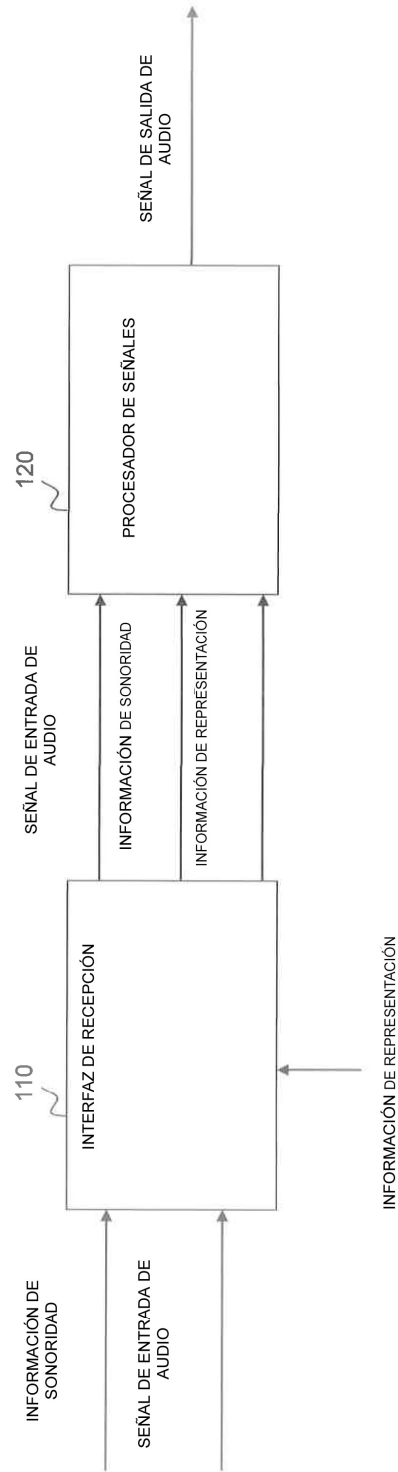


FIG 1

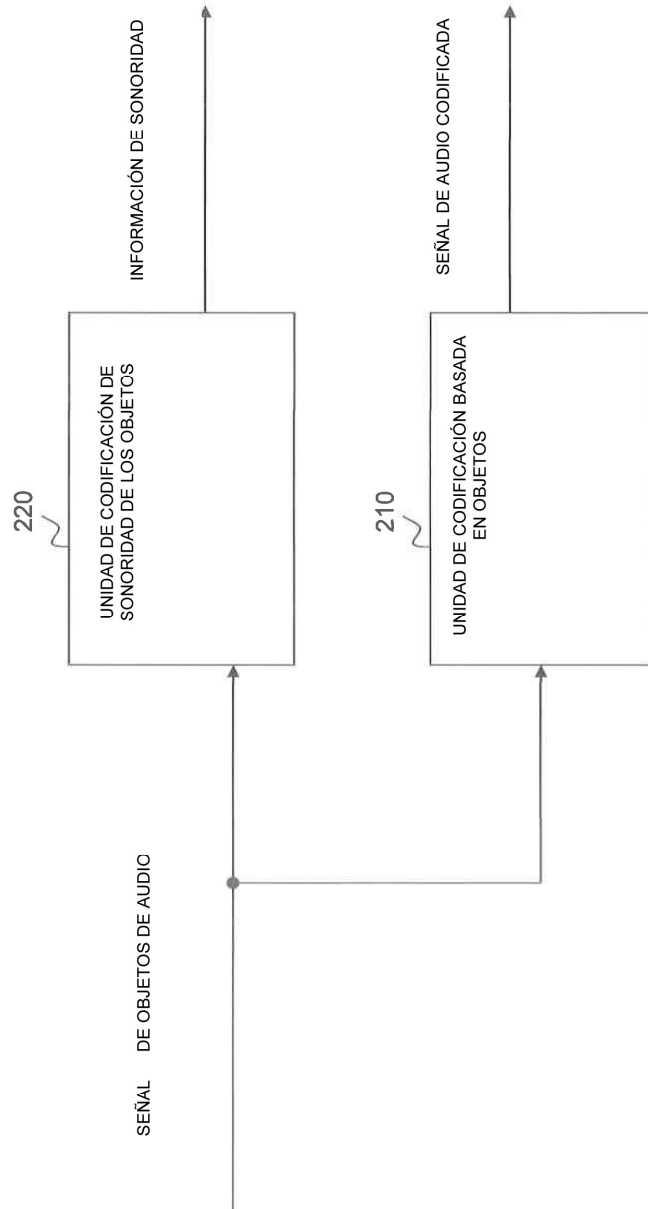


FIG 2

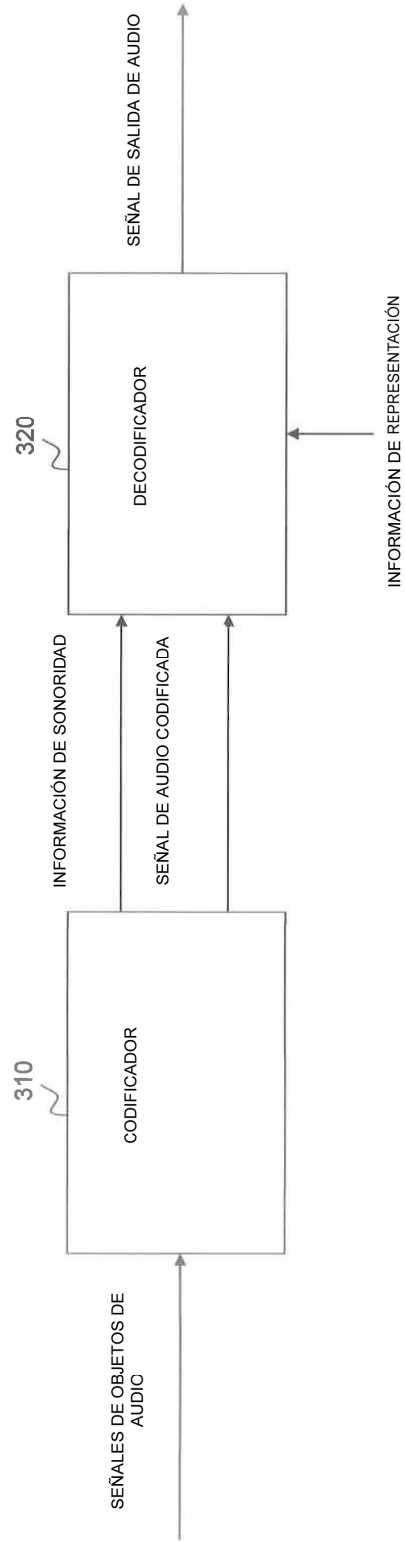


FIG 3

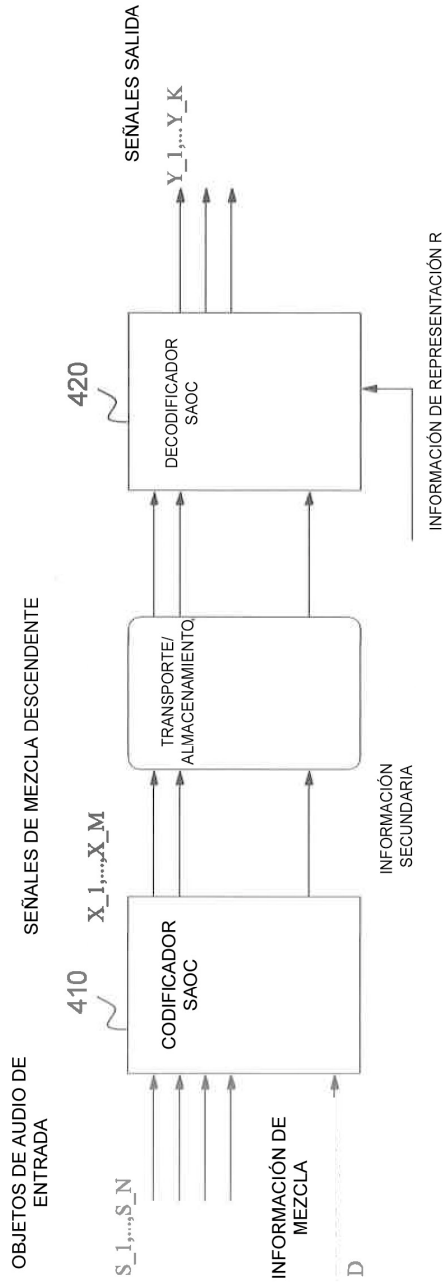


FIG 4

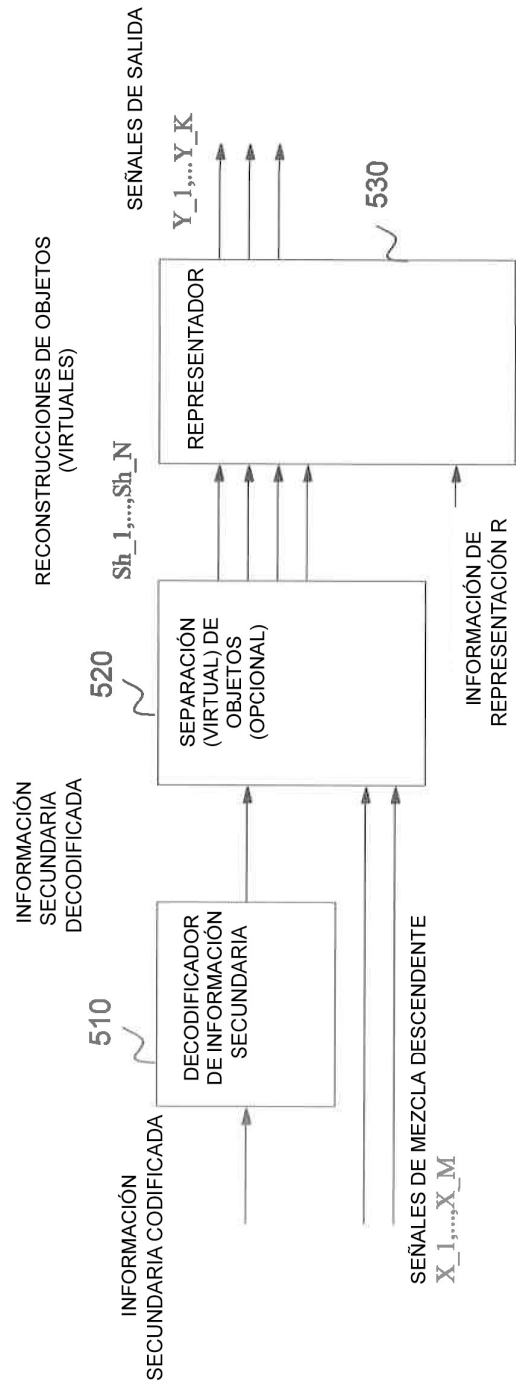


FIG 5

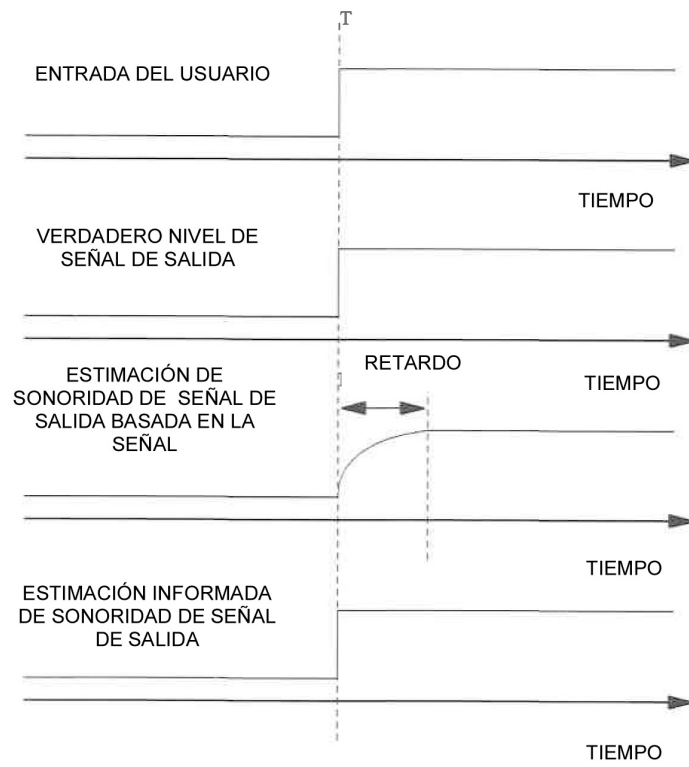


FIG 6

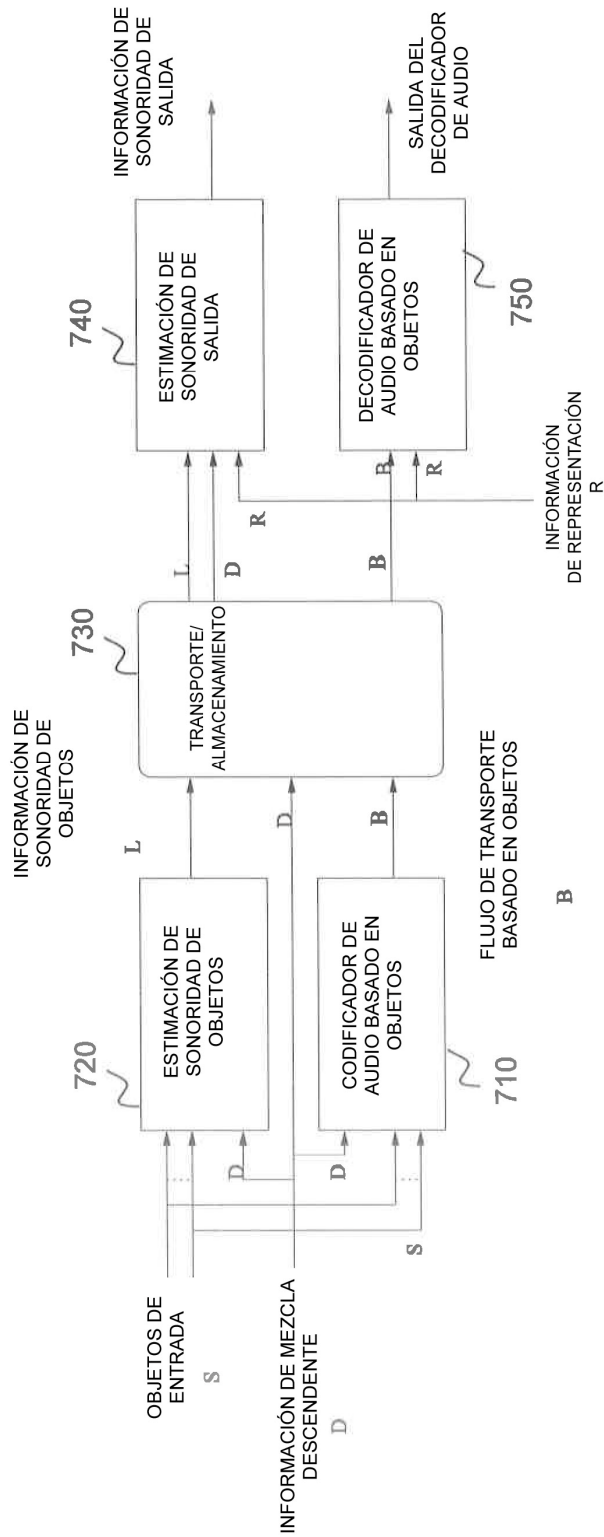


FIG 7

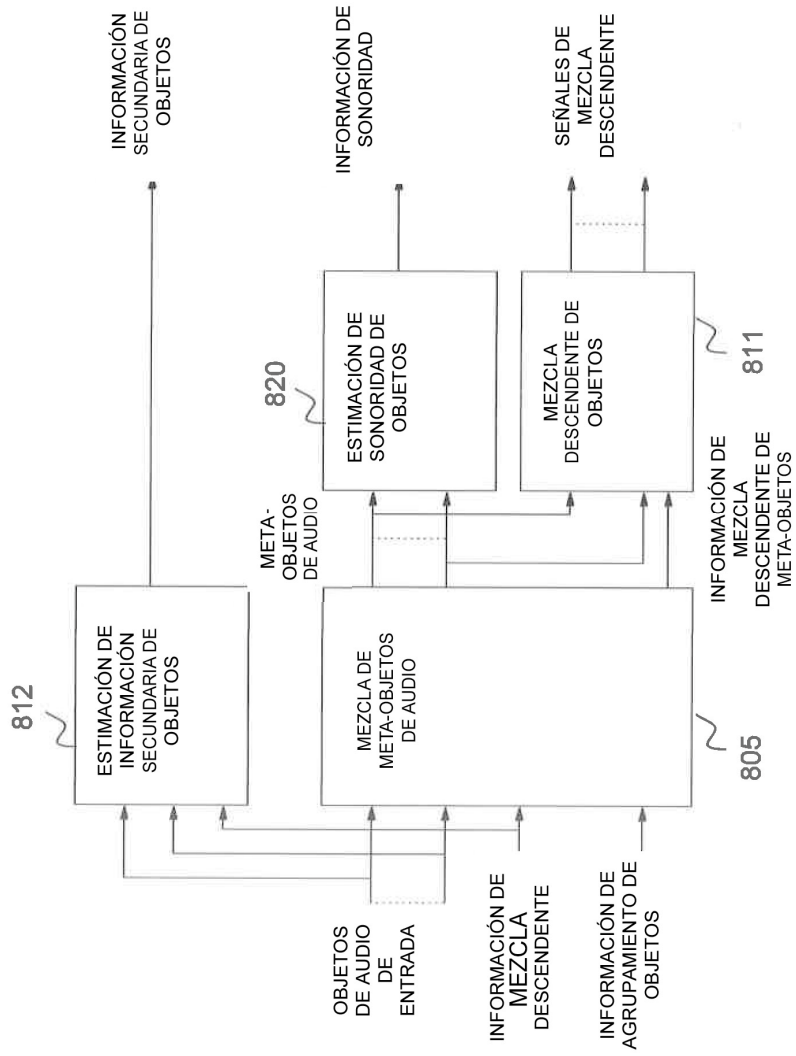


FIG 8

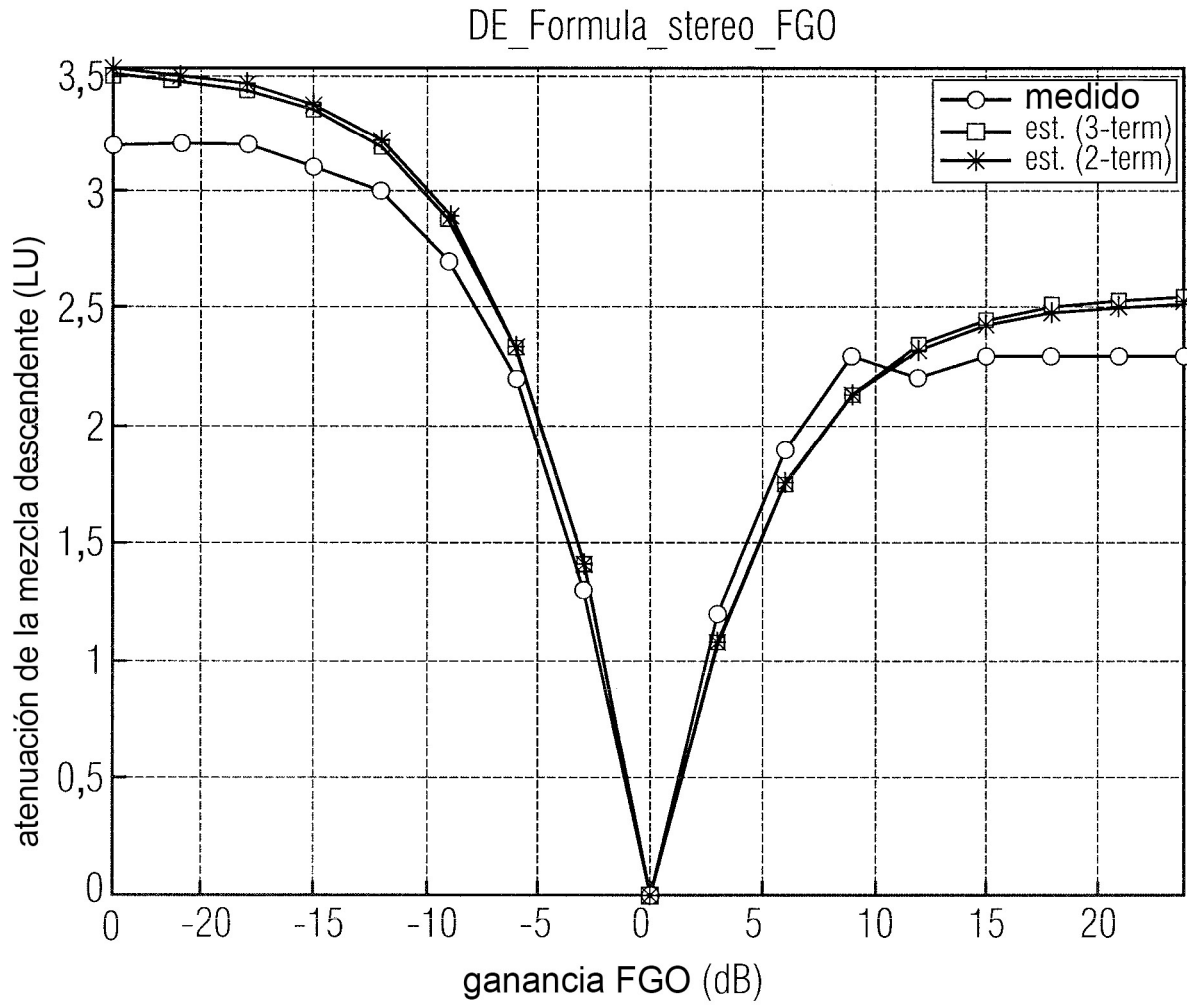


FIG 10

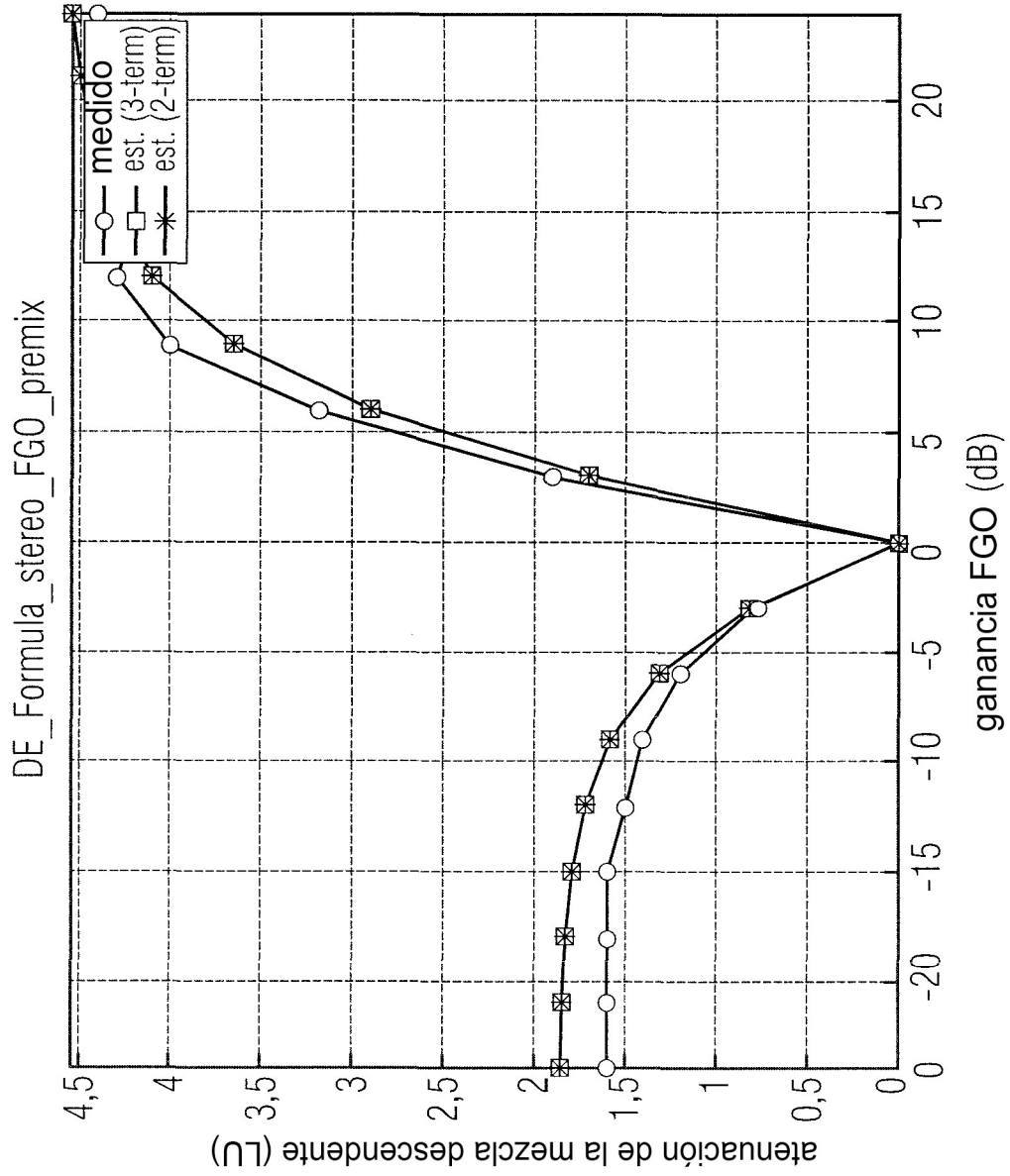


FIG 11

