

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 969 138**

51 Int. Cl.:

G10L 19/008 (2013.01)

H04S 3/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **06.12.2019** **PCT/EP2019/084055**

87 Fecha y número de publicación internacional: **11.06.2020** **WO20115310**

96 Fecha de presentación y número de la solicitud europea: **06.12.2019** **E 19813055 (1)**

97 Fecha y número de publicación de la concesión europea: **29.11.2023** **EP 3891735**

54 Título: **Aparato, método y programa informático para codificación, decodificación, procesamiento de escenas y otros procedimientos relacionados con codificación de audio espacial basada en dirac que utiliza compensación directa de componentes**

30 Prioridad:

07.12.2018 EP 18211064

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
16.05.2024

73 Titular/es:

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

72 Inventor/es:

**FUCHS, GUILLAUME;
THIERGART, OLIVER;
KORSE, SRIKANTH;
DÖHLA, STEFAN;
MULTRUS, MARKUS;
KÜCH, FABIAN;
BOUTHÉON, ALEXANDRE;
EICHENSEER, ANDREA y
BAYER, STEFAN**

74 Agente/Representante:

ARIZTI ACHA, Monica

ES 2 969 138 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

Aparato, método y programa informático para codificación, decodificación, procesamiento de escenas y otros procedimientos relacionados con codificación de audio espacial basada en dirac que utiliza compensación directa de componentes

La presente invención se refiere a codificación de audio y, en particular, a la generación de una descripción de campo de sonido a partir de una señal de entrada usando uno o más generadores de componentes de sonido.

La técnica de codificación de audio direccional (DirAC) [1] es un enfoque eficiente para el análisis y la reproducción del sonido espacial. La DirAC utiliza una representación perceptualmente motivada del campo de sonido con base en la dirección de llegada (DOA) y difusión medida por banda de frecuencia. Se basa en la suposición de que en un instante de tiempo y en una banda crítica, la resolución espacial del sistema auditivo se limita a decodificar una pista para dirección y otra para coherencia interaural. El sonido espacial se representa entonces en el dominio de frecuencia mediante fundido cruzado de dos flujos: un flujo difuso no direccional y un flujo no difuso direccional.

DirAC se propuso originalmente para sonido grabado en formato B, pero también se puede extender para señales de micrófono que coinciden con una configuración de altavoz específica como 5.1 [2] o cualquier configuración de arreglos de micrófonos [5]. En el último caso, se puede lograr más flexibilidad al grabar las señales no para una configuración de altavoz específica, sino en cambio grabando las señales de un formato intermedio.

Este formato intermedio, que está bien establecido en la práctica, se representa por Ambisonics (de mayor orden) [3]. A partir de una señal Ambisonics, se pueden generar las señales de cada configuración de altavoz deseada, que incluyen las señales binaurales para reproducción de auriculares. Esto requiere un renderizador específico que se aplica a la señal Ambisonics, utilizando un renderizador Ambisonics lineal [3] o un renderizador paramétrico como la codificación de audio direccional (DirAC).

Una señal Ambisonics se puede representar como una señal multicanal donde cada canal (denominado componente Ambisonics) es equivalente al coeficiente de una llamada función de base espacial. Con una suma ponderada de estas funciones de base espacial (con los factores de ponderación correspondientes a los coeficientes) se puede recrear el campo de sonido original en la ubicación de grabación [3]. Por lo tanto, los coeficientes de función de base espacial (es decir, los componentes Ambisonics) representan una descripción compacta del campo de sonido en la ubicación de grabación. Existen diferentes tipos de funciones de base espacial, por ejemplo, armónicos esféricos (SH) [3] o armónicos cilíndricos (CH) [3]. Los CH se pueden usar cuando se describe el campo de sonido en el espacio 2D (por ejemplo, para la reproducción de sonido 2D), mientras que los SH se pueden usar para describir el campo de sonido en el espacio 2D y 3D (por ejemplo, para la reproducción de sonido 2D y 3D).

Como un ejemplo, una señal de audio $f(t)$ que llega de una cierta dirección (φ, θ) da como resultado una señal de audio espacial $f(\varphi, \theta, t)$ que se puede representar en formato Ambisonics expandiendo los armónicos esféricos hasta un orden de truncamiento H:

$$f(\varphi, \theta, t) = \sum_{l=0}^H \sum_{m=-l}^{+l} Y_l^m(\varphi, \theta) \phi_{lm}(t)$$

por lo cual $Y_l^m(\varphi, \theta)$ son los armónicos esféricos de orden l y modo m , y $\phi_{lm}(t)$ los coeficientes de expansión. Con un orden de truncamiento creciente H, la expansión da como resultado una representación espacial más precisa. Los armónicos esféricos hasta el orden H = 4 con el índice de numeración de canal Ambisonics (ACN) se ilustran en la figura 1a para el orden n y el modo m .

DirAC ya se extendió para distribuir señales Ambisonics de mayor orden desde una señal Ambisonics de primer orden (FOA como el llamado formato B) o desde diferentes arreglos de micrófonos [5]. Este documento se centra en una forma más eficiente de sintetizar señales Ambisonics de mayor orden a partir de parámetros DirAC y una señal de referencia. En este documento, la señal de referencia, también denominada señal de mezcla descendente, se considera un subconjunto de una señal Ambisonics de mayor orden o una combinación lineal de un subconjunto de los componentes Ambisonics.

Además, la presente invención considera el caso en el que se utiliza la DirAC para la transmisión en forma paramétrica de la escena de audio. En este caso, la señal de mezcla descendente se codifica mediante un codificador de núcleo de audio convencional mientras que los parámetros DirAC se transmiten de manera comprimida como información secundaria. La ventaja del presente método es tener en cuenta el error de cuantificación que se produce durante la

codificación de audio.

A continuación, se presenta una descripción general de un sistema de codificación de audio espacial basado en DirAC diseñado para servicios de voz y audio inmersivos (IVAS). Esto representa uno de los diferentes contextos, como una
 5 visión general del sistema de un codificador de audio espacial DirAC. El objetivo de este sistema es poder manejar diferentes formatos de audio espacial que representan la escena de audio y codificarlos a bajas velocidades de bits y reproducir la escena de audio original lo más fielmente posible después de la transmisión.

El sistema puede aceptar como entrada diferentes representaciones de escenas de audio. La escena de audio de
 10 entrada se puede capturar mediante señales multicanal dirigidas a reproducirse en las diferentes posiciones de los altavoces, objetos auditivos junto con metadatos que describen las posiciones de los objetos con respecto al tiempo, o un formato Ambisonics de primer orden o de mayor orden que representa el campo de sonido en la posición del oyente o de referencia.

Preferiblemente, el sistema se basa en Servicios de Voz Mejorados (EVS) 3GPP, ya que se espera que la solución
 15 funcione con baja latencia para permitir servicios de conversación en redes móviles.

Como se muestra en la figura 1b, el codificador (codificador IVAS) es capaz de soportar diferentes formatos de audio
 20 presentados al sistema por separado o al mismo tiempo. Las señales de audio pueden ser de naturaleza acústica, captadas por micrófonos, o de naturaleza eléctrica, que se supone que se transmiten a los altavoces. Los formatos de audio admitidos pueden ser señales multicanal, componentes Ambisonics de primer orden y de mayor orden y objetos de audio. Una escena de audio compleja también se puede describir combinando diferentes formatos de entrada. Luego, todos los formatos de audio se transmiten al análisis DirAC, que extrae una representación paramétrica de la
 25 escena de audio completa. Una dirección de llegada y una difusión medida por unidad de tiempo-frecuencia forman los parámetros. El análisis DirAC es seguido por un codificador de metadatos espaciales, que cuantifica y codifica los parámetros DirAC para obtener una representación paramétrica de baja velocidad de bits.

Junto con los parámetros, se codifica una señal de mezcla descendente derivada de las diferentes fuentes o señales
 30 de entrada de audio para su transmisión mediante un codificador de núcleo de audio convencional. En este caso, se adopta un codificador de audio basado en EVS para codificar la señal de mezcla descendente. La señal de mezcla descendente consta de diferentes canales, llamados canales de transporte: la señal puede ser, por ejemplo, las cuatro señales de coeficiente que componen una señal de formato B, un par estéreo o una mezcla descendente monofónica dependiendo de la velocidad de bits objetivo. Los parámetros espaciales codificados y el flujo de bits de audio
 35 codificado se multiplexan antes de transmitirse a través del canal de comunicación.

El lado de codificador de la codificación de audio espacial basada en DirAC que soporta diferentes formatos de audio
 se ilustra en la figura 1b. Una entrada acústica/eléctrica 1000 se introduce en una interfaz de codificador 1010, donde la interfaz de codificador tiene una funcionalidad específica para Ambisonics de primer orden (FOA) o Ambisonics de
 40 alto orden (HOA) ilustrada en 1013. Además, la interfaz de codificador tiene una funcionalidad para datos multicanal (MC) tal como datos estéreo, datos 5.1 o datos que tienen más de dos o cinco canales. Además, la interfaz de codificador 1010 tiene una funcionalidad para codificación de objetos como, por ejemplo, SAOC (codificación de objetos de audio espacial) ilustrada 1011. El codificador IVAS comprende una etapa DirAC 1020 que tiene un bloque de análisis DirAC 1021 y un bloque de mezcla descendente (DMX) 1022. La salida de señal por el bloque 1022 se
 45 codifica por un codificador de núcleo IVAS 1040 tal como codificador AAC o EVS, y los metadatos generados por el bloque 1021 se codifican usando un codificador de metadatos DirAC 1030.

En el decodificador, que se muestra en la figura 2, los canales de transporte se decodifican por el decodificador de
 núcleo, mientras que los metadatos DirAC se decodifican primero antes de que se porten con los canales de transporte
 50 decodificados a la síntesis DirAC. En este momento, se pueden considerar diferentes opciones. Se puede solicitar reproducir la escena de audio directamente en cualquier altavoz o configuración de auriculares como suele ser posible en un sistema DirAC convencional (MC en la figura 2).

El decodificador también puede entregar los objetos individuales tal como se presentaron en el lado de codificador
 55 (objetos en la figura 2).

Alternativamente, también se puede solicitar que la escena se represente en formato Ambisonics para otras
 manipulaciones adicionales, como rotación, reflexión o movimiento de la escena (FOA/HOA en la figura 2) o para usar
 un procesador externo no definido en el sistema original.

El decodificador de la codificación de audio espacial DirAC que distribuye diferentes formatos de audio se ilustra en la
 60 figura 2 y comprende un decodificador IVAS 1045 y la interfaz de decodificador conectada posteriormente 1046. El decodificador IVAS 1045 comprende un decodificador de núcleo IVAS 1060 que se configura para realizar una operación de decodificación de contenido codificado por el codificador de núcleo IVAS 1040 de la figura 1b. Además,

se proporciona un decodificador de metadatos DirAC 1050 que proporciona la funcionalidad de decodificación para decodificar el contenido codificado por el codificador de metadatos DirAC 1030. Un sintetizador DirAC 1070 recibe datos del bloque 1050 y 1060 y usando algo de interactividad de usuario o no, la salida se ingresa en una interfaz de decodificador 1046 que genera datos FOA/HOA ilustrados en 1083, datos multicanal (datos MC) como se ilustra en el bloque 1082, o datos de objeto como se ilustra en el bloque 1080.

En la figura 3 se ilustra una síntesis HOA convencional utilizando el paradigma DirAC. Una señal de entrada llamada señal de mezcla descendente es la frecuencia de tiempo analizada por un banco de filtros de frecuencia. El banco de filtros de frecuencia 2000 puede ser un banco de filtros de valor complejo como QMF de valor complejo o una transformada de bloque como STFT. La síntesis HOA genera en la salida una señal Ambisonics de orden H que contiene $(H + 1)^2$ componentes. Opcionalmente, también puede emitir la señal Ambisonics representada en un diseño de altavoz específico. A continuación, se detalla cómo obtener los $(H + 1)^2$ componentes de la señal de mezcla descendente acompañados en algunos casos por parámetros espaciales de entrada.

La señal de mezcla descendente puede ser las señales de micrófono originales o una mezcla de las señales originales que representan la escena de audio original. Por ejemplo, si la escena de audio se captura mediante un micrófono de campo de sonido, la señal de mezcla descendente puede ser el componente omnidireccional de la escena (W), una mezcla descendente estéreo (L/R) o la señal Ambisonics de primer orden (FOA).

Para cada mosaico de tiempo-frecuencia, una dirección de sonido, también llamada Dirección de llegada (DOA), y un factor de difusión se estiman por el estimador de dirección 2020 y por el estimador de difusión 2010, respectivamente, si la señal de mezcla descendente contiene información suficiente para determinar estos parámetros DirAC. Es el caso, por ejemplo, si la señal de mezcla descendente es una señal Ambisonics de primer orden (FOA). Alternativamente o si la señal de mezcla descendente no es suficiente para determinar estos parámetros, los parámetros se pueden transmitir directamente a la síntesis DirAC a través de un flujo de bits de entrada que contiene los parámetros espaciales. El flujo de bits podría consistir, por ejemplo, de parámetros cuantificados y codificados recibidos como información secundaria en el caso de aplicaciones de transmisión de audio. En este caso, los parámetros se derivan fuera del módulo de síntesis DirAC de las señales de micrófono originales o los formatos de audio de entrada proporcionados al módulo de análisis DirAC en el lado de codificador como se ilustra mediante el interruptor 2030 o 2040.

Las direcciones de sonido se utilizan por un evaluador de ganancias direccionales 2050 para evaluar, para cada mosaico de tiempo-frecuencia de la pluralidad de mosaicos de tiempo-frecuencia, uno o más conjuntos de $(H + 1)^2$ ganancias direccionales $G_l^m(k, n)$, donde H es el orden de la señal Ambisonics sintetizada.

Las ganancias direccionales se pueden obtener mediante la evaluación de la función de base espacial para cada dirección de sonido estimada en el orden deseado (nivel) l y modo m de la señal Ambisonics para sintetizar. La dirección de sonido se puede expresar, por ejemplo, en términos de un vector normal unitario $n(k, n)$ o en términos de un ángulo de acimut $\varphi(k, n)$ y/o ángulo de elevación $\theta(k, n)$, que se relacionan, por ejemplo, como:

$$n(k, n) = \begin{bmatrix} \cos \varphi(k, n) \cos \theta(k, n) \\ \sin \varphi(k, n) \cos \theta(k, n) \\ \sin \theta(k, n) \end{bmatrix}$$

Después de estimar u obtener la dirección del sonido, una respuesta de una función de base espacial del orden deseado (nivel) l y modo se pueden determinar, por ejemplo, considerando armónicos esféricos de valor real con normalización SN3D como función de base espacial:

$$Y_l^m(\varphi, \theta) = N_l^{|m|} P_l^{|m|} \sin \theta \begin{cases} \sin(|m|\varphi) & \text{si } m < 0 \\ \cos(|m|\varphi) & \text{si } m \geq 0 \end{cases}$$

con los intervalos $0 \leq l \leq H$, y $-l \leq m \leq l$. $P_l^{|m|}$ son las funciones de Legendre y $N_l^{|m|}$ es un término de normalización tanto para las funciones de Legendre como para las funciones trigonométricas que toma la siguiente forma para SN3D:

$$N_l^{|m|} = \sqrt{\frac{2 - \delta_m (l - |m|)!}{4\pi (l + |m|)!}}$$

donde la delta Kronecker δ_m es uno para $m = 0$ y cero en caso contrario. Las ganancias direccionales se deducen directamente para cada mosaico de tiempo-frecuencia de índices (k, n) como:

5

$$G_l^m(k, n) = Y_l^m(\varphi(k, n), \theta(k, n))$$

Los componentes Ambisonics de sonido directo $P_{s,l}^m$ se calculan derivando una señal de referencia P_{ref} de la señal de mezcla descendente y se multiplican por las ganancias direccionales y una función de factor de la difusión $\Psi(k, n)$:

10

$$P_{s,l}^m(k, n) = P_{ref}(k, n) \sqrt{1 - \Psi(k, n)} G_l^m(k, n)$$

Por ejemplo, la señal de referencia P_{ref} puede ser el componente omnidireccional de la señal de mezcla descendente o una combinación lineal de los K canales de la señal de mezcla descendente.

15

El componente Ambisonics de sonido difuso se puede modelar mediante el uso de una respuesta de una función de base espacial para los sonidos que llegan de todas las direcciones posibles. Un ejemplo es definir la respuesta promedio D_l^m al considerar la integral de la magnitud al cuadrado de la función de base espacial D_l^m a través de todos los ángulos posibles φ y θ :

$$D_l^m = \int_0^{2\pi} \int_0^\pi |Y_l^m(\varphi, \theta)|^2 \sin \theta d\theta d\varphi$$

20

Los componentes Ambisonics de sonido difuso $P_{d,l}^m$ se calculan a partir de una señal P_{diff} multiplicada por la respuesta promedio y una función de factor de la difusión $\Psi(k, n)$:

25

$$P_{d,l}^m(k, n) = P_{diff,l}^m(k, n) \sqrt{\Psi(k, n)} \sqrt{D_l^m}$$

La señal $P_{diff,l}^m$ se puede obtener mediante el uso de diferentes descorrelacionadores aplicados a la señal de referencia P_{ref} .

30

Finalmente, el componente Ambisonics de sonido directo y el componente Ambisonics de sonido difuso se combinan 2060, por ejemplo, mediante la operación de suma, para obtener el componente Ambisonics final P_l^m del orden deseado (nivel) l y modo m para el mosaico de tiempo-frecuencia (k, n) , es decir,

$$P_l^m(k, n) = P_{s,l}^m(k, n) + P_{diff,l}^m(k, n)$$

35

Los componentes Ambisonics obtenidos se pueden transformar de nuevo en el dominio de tiempo utilizando un banco de filtros inversos 2080 o una STFT inversa, almacenados, transmitidos o utilizados, por ejemplo, para aplicaciones de reproducción de sonido espacial. Alternativamente, se puede aplicar un renderizador Ambisonics lineal 2070 para cada banda de frecuencia para obtener señales que se van a reproducir en un diseño de altavoz específico o a través de auriculares antes de transformar las señales de altavoces o las señales binaurales en el dominio de tiempo.

40

Se debe señalar que [5] también enseñó la posibilidad de que los componentes de sonido difuso $P_{diff,l}^m$ solo se podían sintetizar hasta un orden L , donde $L < H$. Esto reduce la complejidad computacional mientras se evitan artefactos sintéticos debido al uso intensivo de descorrelacionadores.

45

El objeto de la presente invención es proporcionar un concepto mejorado para generar una descripción de campo de sonido a partir de una señal de entrada. Otros ejemplos para generar una representación de un campo sonoro pueden encontrarse en [6] y [7].

Este objeto se logra mediante un aparato para generar una descripción de campo de sonido de la reivindicación 1, un método para generar una descripción de campo de sonido de la reivindicación 21 o un programa informático de la reivindicación 22.

- 5 La presente invención de acuerdo con un primer aspecto se basa en el hallazgo de que no es necesario realizar una síntesis de componentes de campo de sonido que incluye un cálculo de porción difusa para todos los componentes generados. Es suficiente realizar una síntesis de componentes difusos solo hasta un cierto orden. Sin embargo, con el fin de no tener fluctuaciones de energía o errores de energía, se realiza una compensación de energía al generar los componentes de campo de sonido de un primer grupo de componentes de campo de sonido que tienen un
- 10 componente difuso y directo, donde esta compensación de energía depende de los datos de difusión, y al menos uno de un número de componentes de campo de sonido en el segundo grupo, un orden máximo de componentes de campo de sonido del primer grupo y un orden máximo de los componentes de campo de sonido del segundo grupo. Particularmente, de acuerdo con el primer aspecto, un aparato para generar una descripción de campo de sonido a partir de una señal de entrada que comprende uno o más canales comprende un analizador de señal de entrada para obtener datos de difusión de la señal de entrada y un generador de componentes de sonido para generar, a partir de la señal de entrada, uno o más componentes de campo de sonido de un primer grupo de componentes de campo de sonido que tienen para cada componente de campo de sonido un componente directo y un componente difuso, y para generar, a partir de la señal de entrada, el segundo grupo de componentes de campo de sonido que tienen solo el componente directo. Particularmente, el generador de componentes de sonido realiza una compensación de energía
- 15 cuando se genera el primer grupo de componentes de campo de sonido, la compensación de energía dependiendo de los datos de difusión y al menos uno de un número de componentes de campo de sonido en el segundo grupo, un número de componentes difusos en el primer grupo, un orden máximo de componentes de campo de sonido del primer grupo y un orden máximo de componentes de campo de sonido del segundo grupo.
- 25 El primer grupo de componentes de campo de sonido puede comprender componentes de campo de sonido de bajo orden y componentes de campo de sonido de orden medio, y el segundo grupo comprende componentes de campo de sonido de alto orden.
- Un aparato para generar una descripción de campo de sonido a partir de una señal de entrada que comprende al menos dos canales de acuerdo con un segundo aspecto correspondiente a la presente invención comprende un analizador de señal de entrada para obtener datos de dirección y datos de difusión de la señal de entrada. El aparato comprende además un estimador para estimar una primera medición relacionada con energía o amplitud para un componente omnidireccional derivado de la señal de entrada y para estimar una segunda medición relacionada con energía o amplitud para un componente direccional derivado de la señal de entrada. Además, el aparato comprende
- 30 un generador de componentes de sonido para generar componentes de campo de sonido del campo de sonido, donde el generador de componentes de sonido se configura para realizar una compensación de energía del componente direccional usando la primera medición relacionada con energía o amplitud, la segunda medición relacionada con energía o amplitud, los datos de dirección y los datos de difusión.
- 40 Particularmente, el segundo aspecto correspondiente a la presente invención se basa en el hallazgo de que en una situación, donde el aparato recibe un componente direccional para generar una descripción de campo de sonido y, al mismo tiempo, también se reciben datos de dirección y datos de difusión, los datos de dirección y difusión se pueden utilizar para compensar cualquier error probablemente introducido debido a una cuantificación o cualquier otro procesamiento del componente direccional u omnidireccional dentro del codificador. Por lo tanto, los datos de dirección y difusión no se aplican simplemente con el propósito de generación de descripción de campo de sonido tal como son, sino estos datos se utilizan una "segunda vez" para corregir el componente direccional con el fin de deshacer o al menos parcialmente deshacer y, por lo tanto, compensar una pérdida de energía del componente direccional.
- 45
- Preferentemente, esta compensación de energía se realiza a componentes de bajo orden que se reciben en una interfaz de decodificador o que se generan a partir de datos recibidos de un codificador de audio que genera la señal de entrada.
- 50
- De acuerdo con un tercer aspecto, un aparato para generar una descripción de campo de sonido usando una señal de entrada que comprende una señal mono o una señal multicanal comprende un analizador de señal de entrada, un generador de componentes de audio bajo, un generador de componentes de orden medio y un generador de componentes de alto orden. Particularmente, los diferentes "sub"-generadores se configuran para generar componentes de campo de sonido en el orden respectivo con base en un procedimiento de procesamiento específico que es diferente para cada uno de los generadores de componentes de bajo, medio o alto orden. Esto asegura que se mantenga un equilibrio óptimo entre los requisitos de procesamiento por un lado, los requisitos de calidad de audio por otro lado y los procedimientos de practicidad por otro lado. Mediante este procedimiento, el uso de descorrelacionadores, por ejemplo, se limita solo a la generación de componentes de orden medio, pero se evitan los descorrelacionadores propensos a artefactos para la generación de componentes de bajo orden y la generación de componentes de alto orden. Por otro lado, preferentemente se realiza una compensación de energía para la pérdida
- 55
- 60

de energía de componentes difusos y esta compensación de energía se realiza dentro de los componentes de campo de sonido de bajo orden solo o dentro de los componentes de campo de sonido de orden medio solo o tanto en los componentes de campo de sonido de bajo orden como en los componentes de campo de sonido de orden medio. Preferentemente, una compensación de energía para el componente direccional formado en el generador de componentes de bajo orden también se realiza usando datos de difusión direccional transmitidos.

Las realizaciones preferidas se refieren a un aparato, un método o un programa informático para sintetizar una señal Ambisonics (de mayor orden) usando un paradigma de Codificación de Audio Direccional (DirAC), una técnica perceptualmente motivada para el procesamiento de audio espacial.

Las realizaciones se refieren a un método eficiente para sintetizar una representación Ambisonics de una escena de audio a partir de parámetros espaciales y una señal de mezcla descendente. En una aplicación del método, pero no se limita a, la escena de audio se transmite y, por lo tanto, se codifica para reducir la cantidad de datos transmitidos. La señal de mezcla descendente se limita fuertemente en número de canales y calidad por la velocidad de bits disponible para la transmisión. Las realizaciones se refieren a una forma eficaz de aprovechar la información contenida en la señal de mezcla descendente transmitida para reducir la complejidad de la síntesis al tiempo que aumenta la calidad.

Otra realización de la invención se refiere al componente difuso del campo de sonido que se puede limitar a modelarse únicamente hasta un orden predeterminado de los componentes sintetizados para evitar la síntesis de artefactos. La realización proporciona una manera de compensar la pérdida de energía resultante mediante la amplificación de la señal de mezcla descendente.

Otra realización se refiere al componente direccional del campo de sonido cuyas características se pueden alterar dentro de la señal de mezcla descendente. La señal de mezcla descendente se puede normalizar en energía además para preservar la relación de energía dictada por un parámetro de dirección transmitido pero interrumpido durante la transmisión por cuantificación inyectada u otros errores.

Posteriormente, las realizaciones preferidas de la presente invención se describen con respecto a las figuras anexas, en las cuales:

La figura 1a ilustra armónicos esféricos con numeración de canales/componentes Ambisonics;

La figura 1b ilustra un lado de codificador de un procesador de codificación de audio espacial basado en DirAC;

La figura 2 ilustra un decodificador del procesador de codificación de audio espacial basado en DirAC;

La figura 3 ilustra un procesador de síntesis Ambisonics de alto orden conocido en la técnica;

La figura 4 ilustra una realización preferida de la presente invención que aplica el primer aspecto, el segundo aspecto y el tercer aspecto;

La figura 5 ilustra un procesamiento de resumen de compensación de energía;

La figura 6 ilustra un aparato para generar una descripción de campo de sonido de acuerdo con un primer;

La figura 7 ilustra un aparato para generar una descripción de campo de sonido de acuerdo con un segundo aspecto correspondiente a la presente invención;

La figura 8 ilustra un aparato para generar una descripción de campo de sonido de acuerdo con un tercer aspecto de la presente invención;

La figura 9 ilustra una implementación preferida del generador de componentes de bajo orden de la figura 8;

La figura 10 ilustra una implementación preferida del generador de componentes de orden medio de la figura 8;

La figura 11 ilustra una implementación preferida del generador de componentes de alto orden de la figura 8;

La figura 12a ilustra una implementación preferida del cálculo de ganancia de compensación de acuerdo con el primer aspecto;

La figura 12b ilustra una implementación del cálculo de compensación de energía de acuerdo con la invención; y

La figura 12c ilustra una implementación preferida de la compensación de energía que combina el primer aspecto y la invención.

La figura 6 ilustra un aparato para generar una descripción de campo de sonido de acuerdo con el primer. El aparato comprende un analizador de señal de entrada 600 para obtener datos de difusión de la señal de entrada ilustrada a la izquierda en la figura 6. Además, el aparato comprende un generador de componentes de sonido 650 para generar, a partir de la señal de entrada, uno o más componentes de campo de sonido de un primer grupo de componentes de campo de sonido que tienen para cada componente de campo de sonido un componente directo y un componente difuso. Además, el generador de componentes de sonido genera, a partir de la señal de entrada, un segundo grupo de componentes de campo de sonido que tienen solo un componente directo.

Particularmente, el generador de componentes de sonido 650 se configura para realizar una compensación de energía cuando se genera el primer grupo de componentes de campo de sonido. La compensación de energía depende de los datos de difusión y el número de componentes de campo de sonido en el segundo grupo o en un orden máximo de los componentes de campo de sonido del primer grupo o un orden máximo de los componentes de campo de sonido del segundo grupo. Particularmente, de acuerdo con el primer aspecto, se realiza una compensación de energía para compensar una pérdida de energía debido al hecho de que, para el segundo grupo de componentes de campo de sonido, solo se generan componentes directos y no se generan componentes difusos.

Contrariamente a esto, en el primer grupo de componentes de campo de sonido, las porciones directa y difusa se incluyen en los componentes de campo de sonido. Por lo tanto, el generador de componentes de sonido 650 genera, como se ilustra por el arreglo superior, componentes de campo de sonido que solo tienen una porción directa y no una porción difusa como se ilustra, en otras figuras, por el número de referencia 830 y el generador de componentes de sonido genera componentes de campo de sonido que tienen una porción directa y una porción difusa como se ilustra por los números de referencia 810, 820 que se explican más adelante con respecto a otras figuras.

La figura 7 ilustra un aparato para generar una descripción de campo de sonido a partir de una señal de entrada que comprende al menos dos canales de acuerdo con el segundo aspecto correspondiente a la invención. El aparato comprende un analizador de señal de entrada 600 para obtener datos de dirección y datos de difusión de la señal de entrada. Además, se proporciona un estimador 720 para estimar una primera medición relacionada con energía o amplitud para un componente omnidireccional derivado de la señal de entrada y para estimar una segunda medición relacionada con energía o amplitud para un componente direccional derivado de la señal de entrada.

Además, el aparato para generar la descripción de campo de sonido comprende un generador de componentes de sonido 750 para generar componentes de campo de sonido del campo de sonido, donde el generador de componentes de sonido 750 se configura para realizar una compensación de energía del componente direccional usando la primera medición de amplitud, la segunda medición relacionada con energía o amplitud, los datos de dirección y los datos de difusión. Por lo tanto, el generador de componentes de sonido genera, de acuerdo con el segundo aspecto correspondiente a la presente invención, componentes direccionales (directos) corregidos/compensados y, si se implementa de manera correspondiente, otros componentes del mismo orden que la señal de entrada, tales como componentes omnidireccionales que preferentemente no se compensan en energía o solo se compensan en energía con el propósito de compensación de energía difusa como se analiza en el contexto de la figura 6. Se debe señalar que la medición relacionada con amplitud también puede ser la normal o magnitud o valor absoluto del componente direccional u omnidireccional tal como B_0 y B_1 . Preferentemente, la potencia o energía derivada por la potencia de 2 se prefiere como se describe en la ecuación, pero también se pueden utilizar otras potencias aplicadas a la normal o magnitud o valor absoluto para obtener la medición relacionada con energía o amplitud.

En una implementación, el aparato para generar una descripción de campo de sonido de acuerdo con el segundo aspecto correspondiente a la invención realiza una compensación de energía del componente de señal direccional incluido en la señal de entrada que comprende al menos dos canales de modo que un componente direccional se incluye en la señal de entrada o se puede calcular a partir de la señal de entrada tal como mediante el cálculo de una diferencia entre los dos canales. Este aparato solo puede realizar una corrección sin generar ningún dato de mayor orden o similar. Sin embargo, en otras realizaciones, el generador de componentes de sonido se configura para generar también otros componentes de campo de sonido de otros órdenes como se ilustra por los números de referencia 820, 830 descritos más adelante, pero para estos componentes de sonido (o de mayor orden), para los cuales no se incluyeron contrapartes en la señal de entrada, no se realiza necesariamente alguna compensación de energía de componente direccional.

La figura 8 ilustra una implementación preferida del aparato para generar una descripción de campo de sonido usando una señal de entrada que comprende una señal mono o una señal multicanal de acuerdo con el tercer aspecto. El aparato comprende un analizador de señal de entrada 600 para analizar la señal de entrada para derivar datos de dirección y datos de difusión. Además, el aparato comprende un generador de componentes de bajo orden 810 para generar una descripción de campo de sonido de bajo orden de la señal de entrada hasta un orden predeterminado y

un modo predeterminado, en el que el generador de componentes de bajo orden 810 se configura para derivar la descripción de campo de sonido de bajo orden al copiar o tomar la señal de entrada o una parte de la señal de entrada tal como es o mediante la realización de una combinación ponderada de los canales de la señal de entrada cuando la señal de entrada es una señal multicanal. Además, el aparato comprende un generador de componentes de orden medio 820 para generar una descripción de campo de sonido de orden medio por encima del orden predeterminado o en el orden predeterminado y por encima del modo predeterminado y por debajo o en un primer orden de truncamiento usando una síntesis de al menos una porción directa y de al menos una porción difusa usando los datos de dirección y los datos de difusión de modo que la descripción de campo de sonido de orden medio comprenda una contribución directa y una contribución difusa.

El aparato para generar la descripción de campo de sonido comprende además un generador de componentes de alto orden 830 para generar una descripción de campo de sonido de alto orden que tiene un componente por encima del primer orden de truncamiento usando una síntesis de al menos una porción directa, en el que la descripción de campo de sonido de alto orden comprende una contribución directa solamente. Por lo tanto, en una realización, la síntesis de la por lo menos una porción directa se realiza sin ninguna síntesis de componente difuso, de modo que la descripción de campo de sonido de alto orden comprende una contribución directa solamente.

Por lo tanto, el generador de componentes de bajo orden 810 genera la descripción de campo de sonido de bajo orden, el generador de componentes de orden medio 820 genera la descripción de campo de sonido de orden medio y el generador de componentes de alto orden genera la descripción de campo de sonido de alto orden. La descripción de campo de sonido de bajo orden se extiende hasta un cierto orden y modo como, por ejemplo, en el contexto de componentes esféricos Ambisonics de alto orden como se ilustra en la figura 1. Sin embargo, cualquier otra descripción de campo de sonido tal como una descripción de campo de sonido con funciones cilíndricas o una descripción de campo de sonido con cualquier otro componente diferente de cualquier representación Ambisonics se puede generar también de acuerdo con el primer aspecto de la invención y/o el tercer aspecto.

El generador de componentes de orden medio 820 genera componentes de campo de sonido por encima del orden o modo predeterminado y hasta un cierto orden de truncamiento que también se indica con L en la siguiente descripción. Finalmente, el generador de componentes de alto orden 830 se configura para aplicar la generación de componentes de campo de sonido desde el orden de truncamiento L hasta un orden máximo indicado como H en la siguiente descripción.

Dependiendo de la implementación, la compensación de energía proporcionada por el generador de componentes de sonido 650 de la figura 6 no se puede aplicar dentro del generador de componentes de bajo orden 810 o el generador de componentes de orden medio 820 como se ilustra por los números de referencia correspondientes en la figura 6 para el componente de sonido directo/difuso. Además, el segundo grupo de componentes de campo de sonido generado por el componente de campo de sonido generado por el generador de componentes de campo de sonido 650 corresponde a la salida del generador de componentes de alto orden 830 de la figura 8 ilustrado por el número de referencia 830 por debajo de la notación directa/no difusa en la figura 6.

Con respecto a la figura 7, se indica que la compensación de energía de componente direccional se realiza preferentemente dentro del generador de componentes de bajo orden 810 ilustrado en la figura 8, es decir, se realiza a algunos o todos los componentes de campo de sonido hasta el orden predeterminado y el modo predeterminado como se ilustra por el número de referencia 810 por encima de la flecha superior que sale del bloque 750. La generación de los componentes de orden medio y los componentes de alto orden se ilustra con respecto a la flecha sombreada superior que sale del bloque 750 en la figura 7 como se ilustra por los números de referencia 820, 830 indicados debajo de la flecha superior. Por lo tanto, el generador de componentes de bajo orden 810 de la figura 8 puede aplicar la compensación de energía difusa de acuerdo con el primer aspecto y la compensación de señal direccional (directa) de acuerdo con la invención, mientras que el generador de componentes de orden medio 820 puede realizar la compensación de componentes difusos solamente, ya que el generador de componentes de orden medio genera datos de salida que tienen porciones difusas que se pueden mejorar con respecto a su energía para tener una mayor distribución de energía de componente difuso en la señal de salida.

Posteriormente, se hace referencia a la figura 4 que ilustra una implementación del primer aspecto, la invención y el tercer aspecto de la invención dentro de un aparato para generar una descripción de campo de sonido.

La figura 4 ilustra el analizador de entrada 600. El analizador de entrada 600 comprende un estimador de dirección 610, un estimador de difusión 620 y conmutadores 630, 640. El analizador de señal de entrada 600 se configura para analizar la señal de entrada, típicamente posterior a un banco de filtros de análisis 400, con el fin de encontrar, para cada información de dirección de periodo de tiempo/frecuencia indicada como información DOA y/o de difusión. La información de dirección DOA y/o la información de difusión también pueden provenir de un flujo de bits. Por lo tanto, en situaciones en las que estos datos no se pueden recuperar de la señal de entrada, es decir, cuando la señal de entrada solo tiene un componente omnidireccional W, entonces el analizador de señal de entrada recupera datos de

dirección y/o datos de difusión del flujo de bits. Cuando, por ejemplo, la señal de entrada es una señal de dos canales que tiene un canal izquierdo L y un canal derecho R, entonces se puede realizar un análisis para obtener datos de dirección y/o difusión. Cuando la señal de entrada es una señal Ambisonics de primer orden (FOA) o, cualquier otra señal con más de dos canales tal como una señal de formato A o una señal de formato B, entonces se puede realizar un análisis de señal real realizado por el bloque 610 o 620. Sin embargo, cuando se analiza el flujo de bits para recuperar, del flujo de bits, los datos de dirección y/o los datos de difusión, esto también representa un análisis realizado por el analizador de señal de entrada 600, pero sin un análisis de señal real como en el otro caso. En este último caso, el análisis se realiza en el flujo de bits, y la señal de entrada consiste tanto de la señal de mezcla descendente como de los datos del flujo de bits.

Además, el aparato para generar una descripción de campo de sonido ilustrada en la figura 4 comprende un bloque de cálculo de ganancias direccionales 410, un divisor 420, un combinador 430, un decodificador 440 y un banco de filtros de síntesis 450. El banco de filtros de síntesis 450 recibe datos para una representación Ambisonics de alto orden o una señal que se reproducirá mediante auriculares, es decir, una señal binaural, o una señal que se va a reproducir mediante altavoces arreglados en una determinada configuración de altavoz que representa una señal multicanal adaptada a la configuración de altavoz específica de la descripción de campo de sonido que es típicamente independiente de la configuración de altavoz específica.

Además, el aparato para generar la descripción de campo de sonido comprende un generador de componentes de sonido que generalmente consiste en el generador de componentes de bajo orden 810 que comprende el bloque de "generar componentes de bajo orden" y el bloque de "mezclar componentes de bajo orden". Además, se proporciona el generador de componentes de orden medio 820 que consiste en el bloque de señal de referencia generado 821, descorrelacionadores 823, 824 y el bloque de componentes de orden medio de mezcla 825. Y, el generador de componentes de alto orden 830 también se proporciona en la figura 4 que comprende el bloque de componentes de alto orden de mezcla 822. Además, se proporciona un bloque de cálculo de ganancias de compensación (difusa) ilustrado en los números de referencia 910, 920, 930, 940. Los números de referencia 910 a 940 se explican adicionalmente con referencia a las figuras 12a a 12c.

Aunque no se ilustra en la figura 4, al menos la compensación de energía de señal difusa no solo se realiza en el generador de componentes de sonido para el bajo orden como se ilustra explícitamente en la figura 4, sino que esta compensación de energía también se puede realizar en el mezclador de componentes de orden medio 825.

Además, la figura 4 ilustra la situación, en la que todo el procesamiento se lleva a cabo en mosaicos de tiempo/frecuencia individuales generados por el banco de filtros de análisis 400. Por lo tanto, para cada mosaico de tiempo/frecuencia, se realiza un cierto valor DOA, un cierto valor de difusión y un cierto procesamiento para aplicar estos valores y también para aplicar las diferentes compensaciones. Además, los componentes de campo de sonido también se generan/sintetizan para los mosaicos de tiempo/frecuencia individuales y la combinación realizada por el combinador 430 también tiene lugar dentro del dominio de tiempo/frecuencia para cada mosaico de tiempo/frecuencia individual, y, además, el procedimiento del decodificador HOA 440 se realiza en el dominio de tiempo/frecuencia y, la síntesis de banco de filtros 450 luego genera las señales de dominio de tiempo para la banda de frecuencia completa con componentes HOA de ancho de banda completo, con señales binaurales de ancho de banda completo para los auriculares o con señales de altavoz de ancho de banda completo para altavoces de una determinada configuración de altavoz.

Realizaciones de la presente invención aprovechan dos principios principales:

- Los componentes Ambisonics de sonido difuso $P_{diff,l}^m$ se pueden restringir para que se sintetizan solo para los componentes de bajo orden de la señal Ambisonics sintetizada hasta el orden $L < H$.

- A partir de, la señal de mezcla descendente, K componentes Ambisonics de bajo orden generalmente se pueden extraer, para lo cual no se requiere una síntesis completa.

- En caso de mezcla descendente mono, la mezcla descendente generalmente representa el componente omnidireccional W de la señal Ambisonics.

- En caso de mezcla estéreo descendente, los canales izquierdo (L) y derecho (R) se pueden transformar fácilmente en componentes Ambisonics W e Y .

$$\begin{cases} W = L + R \\ Y = L - R \end{cases}$$

- En caso de una mezcla descendente FOA, los componentes Ambisonics de orden 1 ya están disponibles.

Alternativamente, el FOA se puede recuperar de una combinación lineal de una señal DMX de mezcla descendente de 4 canales que está, por ejemplo, en formato A:

$$\begin{bmatrix} W \\ Y \\ Z \\ X \end{bmatrix} = T^{-1} \begin{bmatrix} DMX_0 \\ DMX_1 \\ DMX_2 \\ DMX_3 \end{bmatrix}$$

5 Con

$$T = 0.5 \begin{bmatrix} 1 & \sin \theta & 0 & \cos \theta \\ 1 & -\sin \theta & 0 & \cos \theta \\ 1 & 0 & \sin \theta & -\cos \theta \\ 1 & 0 & -\sin \theta & -\cos \theta \end{bmatrix}$$

10 y

$$\theta = \cos^{-1} \frac{1}{\sqrt{3}}$$

A través de estos dos principios, también se pueden aplicar dos mejoras:

- La pérdida de energía al no modelar los componentes Ambisonics de sonido difuso hasta el orden H se puede compensar amplificando los componentes Ambisonics de bajo orden extraídos de la señal de mezcla descendente.
- En las aplicaciones de transmisión donde la señal de mezcla descendente se codifica con pérdida, la señal de mezcla descendente transmitida se corrompe por errores de cuantificación que se pueden mitigar restringiendo la relación de energía de los componentes Ambisonics de bajo orden K extraídos de la señal de mezcla descendente.

La figura 4 ilustra una realización del nuevo método. Una diferencia con respecto al estado de la ilustración en la figura 3 es la diferenciación del proceso de mezcla que difiere de acuerdo con el orden del componente Ambisonics que se va a sintetizar. Los componentes de los bajos órdenes se determinan principalmente a partir de los componentes de bajo orden extraídos directamente de la señal de mezcla descendente. La mezcla de los componentes de bajo orden puede ser tan simple como copiar directamente los componentes extraídos a la salida.

Sin embargo, en la realización preferida, los componentes extraídos se procesan adicionalmente mediante la aplicación de una compensación de energía, función de la difusión y los órdenes de truncamiento L y H , o mediante la aplicación de una normalización de energía, función de la difusión y las direcciones de sonido, o mediante la aplicación de ambos.

La mezcla de los componentes de orden medio es en realidad similar al método del estado de la técnica (aparte de una compensación de difusión opcional), y genera y combina tanto componentes Ambisonics de sonido directos como difusos hasta el orden de truncamiento L pero ignorando los componentes de bajo orden K ya sintetizados por la mezcla de componentes de bajo orden. La mezcla de los componentes de alto orden consiste en generar los componentes Ambisonics restantes $(H - L + 1)^2$ hasta el orden de truncamiento H , pero solo para el sonido directo e ignorando el sonido difuso. A continuación se detalla la mezcla o generación de los componentes de bajo orden.

El primer aspecto se refiere a la compensación de energía generalmente ilustrada en la figura 6 que proporciona una descripción general de procesamiento en el primer aspecto. El principio se explica para el caso específico para $K = (L + 1)^2$ sin pérdida de generalidad.

La figura 5 muestra una descripción general del procesamiento. El vector de entrada $\vec{b}_L$ es una señal Ambisonics físicamente correcta de orden de truncamiento L . Contiene $(L + 1)^2$ coeficientes denotados por $B_{m,l}$, donde $0 \leq l \leq L$ es el orden del coeficiente y $-l \leq m \leq l$ es el modo. Típicamente, la señal Ambisonics $\vec{b}_L$ se representa en el dominio de tiempo-frecuencia.

En el bloque de síntesis HAO 820, 830, los coeficientes Ambisonics se sintetizan desde $\vec{b}_L$ hasta un orden máximo H , donde $H > L$. El vector resultante $\vec{y}_H$ contiene los coeficientes sintetizados de orden $L < l \leq H$, denotados por $Y_{m,l}$. La síntesis HOA normalmente depende de la difusión Ψ (o una medición similar), que describe cuán difuso es el campo de sonido para el punto de tiempo-frecuencia actual. Normalmente, los coeficientes en $\vec{y}_H$ solo se sintetizan si el campo de sonido se vuelve no difuso, mientras que en situaciones difusas, los coeficientes se vuelven cero. Esto evita artefactos en situaciones difusas, pero también da como resultado una pérdida de energía. Más adelante se explican los detalles de la síntesis HOA.

Para compensar la pérdida de energía en situaciones difusas mencionadas anteriormente, aplicamos una compensación de energía $\vec{b}_L$ en el bloque de compensación de energía 650, 750. La señal resultante se denota por $\vec{x}_L$ y tiene el mismo orden máximo L que $\vec{b}_L$. La compensación de energía depende de la difusión (o medición similar) e incrementa la energía de los coeficientes en situaciones difusas de modo que se compensa la pérdida de energía de los coeficientes en $\vec{y}_H$. Más adelante se explican los detalles.

En el bloque de combinación, los coeficientes compensados en energía en $\vec{x}_L$ se combinan 430 con los coeficientes sintetizados en $\vec{y}_H$ para obtener la señal de salida Ambisonics $\vec{z}_H$ que contiene todos los $(H+1)^2$ coeficientes, es decir,

$$\vec{z}_H = \begin{bmatrix} \vec{x}_L \\ \vec{y}_H \end{bmatrix}.$$

Posteriormente, una síntesis HOA se explica como una realización. Existen varios enfoques del estado de la técnica para sintetizar los coeficientes HOA en $\vec{y}_H$, por ejemplo, una renderización basada en covarianza o una renderización directa utilizando codificación de audio direccional (DirAC). En el caso más simple, los coeficientes en $\vec{y}_H$ se sintetizan a partir del componente omnidireccional B_0^0 en $\vec{b}_L$ usando

$$Y_l^m = B_0^0 \sqrt{1 - \Psi} G_l^m(\varphi, \theta).$$

Aquí, (φ, θ) es la dirección de llegada (DOA) del sonido y $G_l^m(\varphi, \theta)$ es la ganancia correspondiente del coeficiente Ambisonics de orden l y modo m . Normalmente, $G_l^m(\varphi, \theta)$ corresponde al patrón de directividad de valor real de la bien conocida función de armónico esférico de orden l y modo m , evaluada en la DOA (φ, θ) . La difusión Ψ se vuelve 0 si el campo de sonido no es difuso, y 1 si el campo de sonido es difuso. En consecuencia, los coeficientes Y_l^m calculados por encima del orden L se vuelven cero en situaciones de registro difuso. Se señala que los parámetros φ , θ y Ψ se pueden estimar a partir de una señal Ambisonics de primer orden $\vec{b}_1$ basada en el vector de intensidad de sonido activo como se explica en los documentos de DirAC originales.

Posteriormente se analiza la compensación de energía de los componentes de sonido difuso. Para derivar la compensación de energía, consideramos un modelo de campo de sonido típico donde el campo de sonido está compuesto por un componente de sonido directo y un componente de sonido difuso, es decir, la señal omnidireccional se puede escribir como

$$B_0^0 = P_s + P_d,$$

donde P_s es el sonido directo (por ejemplo, onda plana) y P_d es el sonido difuso. Asumiendo este modelo de campo de sonido y una normalización SN3D de los coeficientes Ambisonics, la potencia esperada de los coeficientes físicamente correctos $B_{m,l}$ está dada por

$$E\{|B_l^m|^2\} = E\{|G_l^m(\varphi, \theta)|^2\} \Phi_s + Q_l \Phi_d.$$

Aquí, $\Phi_s = E\{|P_s|^2\}$ es la potencia del sonido directo y $\Phi_d = E\{|P_d|^2\}$ es la potencia del sonido difuso. Además, Q_l es el factor de directividad de los coeficientes del l -ésimo orden, que está dado por $Q_l = 1/N$, donde $N = 2l + 1$ es el número de coeficientes por orden l . Para calcular la compensación de energía, podemos considerar la DOA (φ, θ) (compensación de energía más precisa) o se asume que (φ, θ) es una variable aleatoria distribuida

uniformemente (enfoque más práctico). En este último caso, la potencia esperada de B_l^m es

$$E\{|B_l^m|^2\} = Q_l \Phi_s + Q_l \Phi_d.$$

- 5 En lo siguiente, permítase que $\vec{b}_H$ denote una señal Ambisonics físicamente correcta de orden máximo H . Usando las ecuaciones anteriores, la potencia total esperada de $\vec{b}_H$ está dada por

$$\sum_{l=0}^H \sum_{m=-l}^l E\{|B_l^m|^2\} = (H+1)\Phi_s + (H+1)\Phi_d.$$

- 10 De manera similar, cuando se utiliza la definición de difusión común $\Psi = \frac{\Phi_d}{\Phi_s + \Phi_d}$, la potencia total esperada de la señal Ambisonics sintetizada $\vec{y}_H$ está dada por

$$\sum_{l=L+1}^H \sum_{m=-l}^l E\{|Y_l^m|^2\} = (H-L)\Phi_s.$$

- 15 La compensación energética se lleva a cabo multiplicando un factor g por $\vec{b}_L$, es decir,

$$\vec{x}_L = g\vec{b}_L.$$

La potencia total esperada de la señal Ambisonics de salida $\vec{z}_H$ ahora está dada por

20

$$\sum_{l=0}^H \sum_{m=-l}^l E\{|Z_l^m|^2\} = \underbrace{g^2(L+1)\Phi_s + g^2(L+1)\Phi_d}_{\text{potencia total } \vec{x}_L} + \underbrace{(H-L)\Phi_s}_{\text{potencia total } \vec{y}_H}.$$

La potencia total esperada de $\vec{z}_H$ debe coincidir con la potencia total esperada de $\vec{b}_H$. Por lo tanto, el factor de compensación al cuadrado se calcula como

25

$$g^2 = \frac{(L+1)\Phi_s + (H+1)\Phi_d}{(L+1)(\Phi_s + \Phi_d)}$$

Esto se puede simplificar para

30

$$g = \sqrt{1 + \Psi \left(\frac{H+1}{L+1} - 1 \right)},$$

donde Ψ es la difusión, L es el orden máximo de la señal Ambisonics de entrada y H es el orden máximo de la señal Ambisonics de salida.

- 35 Es posible adoptar el mismo principio $K < (L+1)^2$ donde los $(L+1)^2 - K$ componentes Ambisonics de sonido difuso se sintetizan utilizando descorrelacionadores y una respuesta difusa promedio.

En ciertos casos, se sintetizan $K < (L+1)^2$ y componentes de sonido no difuso. Es especialmente cierto para altas frecuencias donde las fases absolutas son inaudibles y el uso de descorrelacionadores es irrelevante. Los componentes de sonido difuso luego se pueden modelar mediante la compensación de energía mediante el cálculo del orden Lk y el número de modos mk correspondientes a los K componentes de bajo orden, donde K representa un número de componentes difusos en el primer grupo:

40

$$\begin{cases} Lk = \lfloor \sqrt{K} - 1 \rfloor \\ mk = K - (Lk + 1)^2, \\ N = 2(Lk + 1) + 1 \end{cases}$$

La ganancia de compensación se convierte entonces en:

$$g = \sqrt{1 + \Psi \left(\frac{H + 1}{Lk + 1 + \frac{mk}{N}} - 1 \right)}$$

Posteriormente, se ilustran realizaciones de la normalización de energía de componentes de sonido directo correspondientes al segundo aspecto ilustrado generalmente en la figura 7. En lo anterior, se asumió que el vector de entrada $\vec{b}_L$ era una señal Ambisonics físicamente correcta de orden máximo L . Sin embargo, la señal de entrada de mezcla descendente se puede ver afectada por errores de cuantificación, que pueden romper la relación de energía. Esta relación se puede restaurar normalizando la señal de entrada de mezcla descendente:

$$\vec{x}_L = g_s \vec{b}_L.$$

Dada la dirección del sonido y los parámetros de difusión, los componentes directos y difusos se pueden expresar como:

$$P_{s,l}^m = B_0^0 \sqrt{1 - \Psi} G_l^m(\varphi, \theta)$$

$$P_{d,l}^m = \sqrt{\Psi} B_l^m.$$

La potencia esperada de acuerdo con el modelo se puede expresar entonces para cada componente de $\vec{x}_L$ como:

$$E\{|X_l^m|^2\} = g_s^2 E\{|B_l^m|^2\} = E\{|B_0^0|^2\} (1 - \Psi)(G_l^m(\varphi, \theta))^2 + \Psi Q_l E\{|B_0^0|^2\}$$

La ganancia de compensación se convierte entonces en:

$$g_s = \sqrt{\frac{E\{|B_0^0|^2\}}{E\{|B_l^m|^2\}} (Q_l \Psi + (1 - \Psi)(G_l^m(\varphi, \theta))^2)},$$

donde $0 \leq l \leq L$ y $-l \leq m \leq l$

Alternativamente, la potencia esperada de acuerdo con el modelo se puede expresar entonces para cada componente de $\vec{x}_L$ como:

$$E\{|X_l^m|^2\} = g_s^2 E\{|B_l^m|^2\} = E\{|B_0^0|^2\} (1 - \Psi)(G_l^m(\varphi, \theta))^2 + \Psi E\{|B_l^m|^2\}$$

La ganancia de compensación se convierte entonces en:

$$g_s = \sqrt{\Psi + \frac{E\{|B_0^0|^2\}}{E\{|B_l^m|^2\}} (1 - \Psi)(G_l^m(\varphi, \theta))^2},$$

donde $0 \leq l \leq L$ y $-l \leq m \leq l$

B_0^0 y B_l^m son valores complejos y para el cálculo de g_s , la normal o magnitud o valor absoluto o la representación de

coordenadas polares del valor complejo se toma y se eleva al cuadrado para obtener la potencia o energía esperada como la medición relacionada con energía o amplitud.

- 5 La compensación de energía de los componentes de sonido difuso y la normalización de energía de los componentes de sonido directo se pueden lograr conjuntamente mediante la aplicación de una ganancia de la forma:

$$g_{s,d} = g \cdot g_s$$

- 10 En una implementación real, la ganancia de normalización obtenida, la ganancia de compensación o la combinación de las dos se pueden limitar para evitar grandes factores de ganancia que dan como resultado una ecualización severa que podría conducir a artefactos de audio. Por ejemplo, las ganancias se pueden limitar a estar entre -6 y +6 dB. Además, las ganancias se pueden suavizar con respecto al tiempo y/o la frecuencia (por un promedio móvil o un promedio recursivo) para evitar cambios abruptos y para luego el proceso de estabilización.

- 15 Posteriormente, se resumirán algunos de los beneficios y ventajas de realizaciones preferidas sobre el estado de la técnica.

- Síntesis HOA simplificada (menos compleja) dentro de DirAC.

- 20 ○ Síntesis más directa sin una síntesis completa de todos los componentes Ambisonics.

- Reducción del número de descorrelacionadores requeridos y su impacto en la calidad final.

- 25 • Reducción de los artefactos de codificación introducidos en la señal de mezcla descendente durante la transmisión.

- Separación del procesamiento de tres pedidos diferentes para tener una compensación óptima entre calidad y eficiencia de procesamiento.

- 30 Posteriormente, se resumen varios aspectos inventivos incluidos parcial o totalmente en la descripción anterior que se pueden usar independientemente entre sí o en combinación entre sí o solo en una determinada combinación que combina solo dos aspectos seleccionados arbitrariamente de los tres aspectos.

Primer aspecto: Compensación de energía para los componentes de sonido difuso

- 35 La presente invención parte del hecho de que cuando se genera una descripción de campo de sonido a partir de una señal de entrada que comprende uno o más componentes de señal, la señal de entrada se puede analizar para obtener al menos datos de difusión para el campo de sonido representado por la señal de entrada. El análisis de señal de entrada puede ser una extracción de datos de difusión asociados como metadatos al uno o más componentes de señal o el análisis de señal de entrada puede ser un análisis de señal real, cuando, por ejemplo, la señal de entrada
- 40 tiene dos, tres o incluso más componentes de señal tal como una representación de primer orden completa tal como una representación de formato B o una representación de formato A.

- Ahora, hay un generador de componentes de sonido que genera uno o más componentes de campo de sonido de un primer grupo que tienen un componente directo y un componente difuso. Y, además, se generan uno o más
- 45 componentes de campo de sonido de un segundo grupo, donde, para este segundo grupo, el componente de campo de sonido solo tiene componentes directos.

- A diferencia de una generación de campo de sonido completo, esto dará lugar a un error de energía siempre que el valor de difusión para el cuadro actual o el intervalo de tiempo/frecuencia actual considerado tengan un valor diferente
- 50 de cero.

- Con el fin de compensar este error de energía, se realiza una compensación de energía al generar el primer grupo de componentes de campo de sonido. Esta compensación de energía depende de los datos de difusión y un número de componentes de campo de sonido en el segundo grupo que representan la pérdida de energía debido a la no síntesis
- 55 de componentes difusos para el segundo grupo.

- En una realización, el generador de componentes de sonido para el primer grupo puede ser la ramificación de bajo orden de la figura 4 que extrae los componentes de campo de sonido del primer grupo mediante la copia o realización de una suma ponderada, es decir, sin realizar una evaluación de función de base espacial compleja. Por lo tanto, el
- 60 componente de campo de sonido del primer grupo no está disponible por separado como una porción directa y una porción difusa. Sin embargo, aumentar todo el componente de campo de sonido del primer grupo con respecto a su energía aumenta automáticamente la energía de la porción difusa.

Alternativamente, el generador de componentes de sonido para uno o más componentes de campo de sonido del primer grupo también puede ser la ramificación de orden medio en la figura 4 que se basa en una síntesis de porción directa separada y síntesis de porción difusa. Aquí, tenemos la porción difusa disponible por separado y, en una realización, la porción difusa del componente de campo de sonido se incrementa pero no la porción directa para compensar la pérdida de energía debido al segundo grupo. Alternativamente, sin embargo, se podría, en este caso, aumentar la energía del componente de campo de sonido resultante después de haber combinado la porción directa y la porción difusa.

Alternativamente, el generador de componentes de sonido para uno o más componentes de campo de sonido del primer grupo también puede ser las ramificaciones de componentes de bajo y medio orden en la figura 4. La compensación de energía se puede aplicar entonces solo a los componentes de bajo orden o a los componentes de bajo y medio orden.

Segundo aspecto: Normalización Energética de Componentes de Sonido Directo

En esta invención, se parte de la suposición de que la generación de la señal de entrada que tiene dos o más componentes de sonido estuvo acompañada por algún tipo de cuantificación. Típicamente, cuando se consideran dos o más componentes de sonido, un componente de sonido de la señal de entrada puede ser una señal omnidireccional, tal como señales de micrófono omnidireccionales W en una representación de formato B, y los otros componentes de sonido pueden ser señales direccionales individuales, tal como las señales de micrófono X, Y, Z en una representación de formato B, es decir, una representación Ambisonics de primer orden.

Cuando un codificador de señal entra en una situación en la que los requisitos de velocidad de bits son demasiado altos para una operación de codificación perfecta, entonces un procedimiento típico es que el codificador codifica la señal omnidireccional lo más exacta posible, pero el codificador solo gasta un número menor de bits para los componentes direccionales que incluso pueden ser tan bajos que uno o más componentes direccionales se reducen a cero completamente. Esto representa un desajuste y pérdida de energía en la información direccional.

Ahora, sin embargo, se tiene el requisito que, por ejemplo, se obtiene al tener información secundaria paramétrica explícita que dice que un cierto cuadro o intervalo de tiempo/frecuencia tiene una cierta difusión que es menor que uno y una dirección de sonido. Por lo tanto, puede surgir la situación de que uno tiene, de acuerdo con los datos paramétricos, un cierto componente no difuso con una cierta dirección mientras que, en el otro lado, la señal omnidireccional transmitida y las señales direccionales no reflejan esta dirección. Por ejemplo, la señal omnidireccional se podría haber transmitido sin ninguna pérdida significativa de información mientras que la señal direccional, Y, responsable de la dirección izquierda y derecha se podría haber establecido a cero por falta de razón de bits. En este escenario, incluso si en la escena de audio original un componente de sonido directo proviene de la izquierda, las señales transmitidas reflejarán una escena de audio sin ninguna característica direccional izquierda-derecha.

Por lo tanto, de acuerdo con la segunda invención, se realiza una normalización de energía para los componentes de sonido directo con el fin de compensar la ruptura de la relación de energía con la ayuda de datos de dirección/difusión que se incluyen explícitamente en la señal de entrada o se derivan de la señal de entrada en sí.

Esta normalización de energía se puede aplicar en el contexto de todas las ramificaciones de procesamiento individuales de la figura 4 ya sea en conjunto o solo por separado.

La presente invención permite utilizar los datos paramétricos adicionales recibidos de la señal de entrada o derivados de porciones no comprometidas de la señal de entrada y, por lo tanto, los errores de codificación que se incluyen en la señal de entrada por alguna razón se pueden reducir utilizando los datos de dirección y difusión adicionales derivados de la señal de entrada.

En esta invención, se estima una medición relacionada con energía o amplitud para un componente omnidireccional derivado de la señal de entrada y una medición adicional relacionada con energía o amplitud para el componente direccional derivado de la señal de entrada y se utiliza para la compensación de energía junto con los datos de dirección y los datos de difusión. Esta medición relacionada con energía o amplitud puede ser la amplitud en sí misma o la potencia, es decir, las amplitudes al cuadrado y sumadas o puede ser la energía tal como potencia multiplicada por un período de tiempo determinado o puede ser cualquier otra medición derivada de la amplitud con un exponente para una amplitud que es diferente de una y una suma posterior. Por lo tanto, una medición adicional relacionada con energía o amplitud también podría ser una intensidad con un exponente de tres en comparación con la potencia que tiene un exponente de dos.

Tercer aspecto: Implementación de sistema con diferentes procedimientos de procesamiento para los diferentes órdenes.

En la tercera invención, que se ilustra en la figura 4, se genera un campo de sonido usando una señal de entrada que comprende una señal mono o una señal de múltiples componentes que tiene dos o más componentes de señal. Un analizador de señal obtiene datos de dirección y difusión de la señal de entrada ya sea mediante un análisis de señal explícito en el caso de que la señal de entrada tenga dos o más componentes de señal o mediante el análisis de la

Un generador de componentes de bajo orden genera la descripción de sonido de bajo orden de la señal de entrada hasta un orden predeterminado y realiza esta tarea para los modos disponibles que se pueden extraer de la señal de entrada al copiar un componente de señal de la señal de entrada o mediante la realización de una combinación ponderada de componentes en la señal de entrada.

El generador de componentes de orden medio genera una descripción de sonido de orden medio que tiene componentes de órdenes por encima del orden predeterminado o en el orden predeterminado y por encima del modo predeterminado e inferior o igual a un primer orden de truncamiento usando una síntesis de al menos un componente directo y una síntesis de al menos un componente difuso usando los datos de dirección y los datos de difusión obtenidos del analizador de modo que la descripción de sonido de orden medio comprenda una contribución directa y una contribución difusa.

Además, un generador de componentes de alto orden genera una descripción de sonido de alto orden que tiene componentes de órdenes por encima del primer orden truncado e inferior o igual a un segundo orden de truncamiento usando una síntesis de al menos un componente directo sin ninguna síntesis de componente difuso de modo que la descripción de sonido de alto orden tiene una contribución directa solamente.

Esta invención de sistema tiene ventajas significativas en que se realiza una generación de campo de sonido de bajo orden lo más exacta posible mediante la utilización de la información incluida en la señal de entrada lo mejor posible mientras que, al mismo tiempo, las operaciones de procesamiento para realizar la descripción de sonido de bajo orden requieren bajos esfuerzos debido al hecho de que solo se requieren operaciones de copia u operaciones de combinación ponderadas tales como sumas ponderadas. Por lo tanto, se realiza una descripción de sonido de bajo orden de alta calidad con una cantidad mínima de potencia de procesamiento requerida.

La descripción de sonido de orden medio requiere más potencia de procesamiento, pero permite generar una descripción de sonido de orden medio muy precisa que tiene contribuciones directas y difusas usando los datos de dirección analizados y datos de difusión típicamente hasta un orden, es decir, el alto orden, por debajo del cual todavía se requiere una contribución difusa en una descripción de campo de sonido desde un punto de vista perceptivo.

Finalmente, el generador de componentes de alto orden genera una descripción de sonido de alto orden solo mediante la realización de una síntesis directa sin realizar una síntesis difusa. Esto, una vez más, reduce la cantidad de potencia de procesamiento requerida debido al hecho de que solo se generan los componentes directos mientras que, al mismo tiempo, la omisión de la síntesis difusa no es tan problemática desde un punto de vista perceptivo.

Naturalmente, la tercera invención se puede combinar con la primera invención y/o la segunda invención, pero incluso cuando, por algunas razones, no se aplica la compensación por no realizar la síntesis difusa con el generador de componentes de alto orden, el procedimiento sin embargo da por resultado un compromiso óptimo entre la potencia de procesamiento por un lado y la calidad de audio por otro lado. Lo mismo es cierto para la realización de la normalización de energía de bajo orden que compensa la codificación utilizada para generar la señal de entrada. En una realización, esta compensación se realiza adicionalmente, pero incluso sin esta compensación, se obtienen ventajas no triviales significativas.

La figura 4 ilustra, como una ilustración simbólica de una transmisión paralela, el número de componentes procesados por cada generador de componentes. El generador de componentes de bajo orden 810 ilustrado en la figura 4 genera una descripción de campo de sonido de bajo orden de la señal de entrada hasta un orden predeterminado y un modo predeterminado, donde el generador de componentes de bajo orden 810 se configura para derivar la descripción de campo de sonido de bajo orden al copiar o tomar la señal de entrada tal como es o la realización de una combinación ponderada de los canales de la señal de entrada. Como se ilustra entre el bloque de componentes de bajo orden de generador y el bloque de componentes de bajo orden de mezcla, se procesan K componentes individuales por este generador de componentes de bajo orden 810. El generador de componentes de orden medio 820 genera la señal de referencia y, como una situación de ejemplo, se describe que se utiliza la señal omnidireccional incluida en la señal de mezcla descendente en la entrada o la salida del banco de filtros 400. Sin embargo, cuando la señal de entrada tiene el canal izquierdo y el canal derecho, entonces la señal mono obtenida mediante la adición del canal izquierdo y derecho se calcula por el generador de señal de referencia 821. Además, el número de $(L + 1)^2 - K$ componentes se genera por el generador de componentes de orden medio. Además, el generador de componentes de alto orden genera un número de $(H + 1)^2 - (L + 1)^2$ componentes de modo que, al final, a la salida del combinador, $(H + 1)^2$ componentes están allí desde el único o varios (número pequeño) componentes en la entrada en el banco de filtros

400. El divisor se configura para proporcionar los datos direccionales/de difusión individuales a los generadores de componentes correspondientes 810, 820, 830. Por lo tanto, el generador de componentes de bajo orden recibe los K elementos de datos. Esto se indica mediante la línea que recopila el divisor 420 y el bloque de componentes de bajo orden de mezcla.

Además, el bloque de componentes de orden de mezcla 825 recibe $(L + 1)^2 - K$ elementos de datos, y el bloque de componentes de alto orden de mezcla recibe $(H + 1)^2 - (L + 1)^2$ elementos de datos. De manera correspondiente, los bloques de componentes de mezcla individuales proporcionan un cierto número de componentes de campo de sonido al combinador 430.

Posteriormente, se ilustra una implementación preferida del generador de componentes de bajo orden 810 de la figura 4 con respecto a la figura 9. La señal de entrada se ingresa en un investigador de señales de entrada 811, y el investigador de señales de entrada 811 proporciona la información adquirida a un selector de modo de procesamiento 812. El selector de modo de procesamiento 812 se configura para seleccionar una pluralidad de diferentes modos de procesamiento que se ilustran esquemáticamente como un bloque de copia 813 indicado por el número 1, un bloque de toma (tal como es) 814 indicado por el número 2, un bloque de combinación lineal (primer modo) indicado por el número 3 y por el número de referencia 815, y un bloque de combinación lineal (segundo modo) 816 indicado por el número 4. Por ejemplo, cuando el investigador de señales de entrada 811 determina un cierto tipo de señal de entrada, entonces el selector de modo de procesamiento 812 selecciona uno de la pluralidad de diferentes modos de procesamiento como se muestra en la tabla de la figura 9. Por ejemplo, cuando la señal de entrada es una señal omnidireccional W o una señal mono entonces se selecciona copiar 813 o tomar 814. Sin embargo, cuando la señal de entrada es una señal estéreo con un canal izquierdo o un canal derecho o una señal multicanal con 5.1 o 7.1 canales, entonces se selecciona el bloque de combinación lineal 815 para derivar, de la señal de entrada, la señal omnidireccional W al sumar izquierda y derecha y al calcular un componente direccional mediante el cálculo de la diferencia entre izquierda y derecha.

Sin embargo, cuando la señal de entrada es una señal estéreo conjunta, es decir, una representación media/lateral, entonces se selecciona el bloque 813 o el bloque 814 ya que la señal media ya representa la señal omnidireccional y la señal lateral ya representa el componente direccional.

De manera similar, cuando se determina que la señal de entrada es una señal Ambisonics de primer orden (FOA), el selector de modo de procesamiento 812 selecciona el bloque 813 o el bloque 814. Sin embargo, cuando se determina que la señal de entrada es una señal de formato A, entonces se selecciona el bloque de combinación lineal (segundo modo) 816 para realizar una transformación lineal en la señal de formato A para obtener la señal Ambisonics de primer orden que tiene el componente omnidireccional y componentes tridireccionales que representan los K bloques de componentes de bajo orden generados por el bloque 810 de la figura 8 o figura 6. Además, la figura 9 ilustra un compensador de energía 900 que se configura para realizar una compensación de energía a la salida de uno de los bloques 813 a 816 con el fin de realizar la compensación difusa y/o la compensación directa con los valores de ganancia correspondientes g y g_s .

Por lo tanto, la implementación del compensador de energía 900 corresponde al procedimiento del generador de componentes de sonido 650 o el generador de componentes de sonido 750 de la figura 6 y la figura 7, respectivamente.

La figura 10 ilustra una implementación preferida del generador de componentes de orden medio 820 de la figura 8 o una parte del generador de componentes de sonido 650 para la flecha inferior directa/difusa del bloque 650 relacionada con el primer grupo. En particular, el generador de componentes de orden medio 820 comprende el generador de señal de referencia 821 que recibe la señal de entrada y genera la señal de referencia al copiar o tomar tal como es cuando la señal de entrada es una señal mono o al derivar de la señal de referencia de la señal de entrada al calcular como se analizó antes o como se ilustra en WO 2017/157803 A1 incorporado en el presente documento mediante referencia con toda su enseñanza.

Además, la figura 10 ilustra la calculadora de ganancia direccional 410 que se configura para calcular, a partir de determinada información DOA (Φ , θ) y a partir de un determinado número de modo m y un determinado número de orden l la ganancia direccional G_l^m . En la realización preferida, donde el procesamiento se realiza en el dominio de tiempo/frecuencia para cada mosaico individual referenciado por k , n , la ganancia direccional se calcula para cada mosaico de tiempo/frecuencia. El ponderador 820 recibe la señal de referencia y los datos de difusión para el mosaico de tiempo/frecuencia determinado y el resultado del ponderador 820 es la porción directa. La porción difusa se genera mediante el procesamiento realizado por el filtro de descorrelación 823 y el ponderador posterior 824 que recibe el valor de difusión Ψ para el periodo de tiempo determinado y el intervalo de frecuencia y, en particular, que recibe la respuesta promedio a un determinado modo m y orden l indicado por D_l generado por el proveedor de respuesta promedio 826 que recibe, como entrada, el modo requerido m y el orden requerido l .

El resultado del ponderador 824 es la porción difusa y la porción difusa se suma a la porción directa por el sumador

825 con el fin de obtener un determinado componente de campo de sonido de orden medio para un determinado modo m y un determinado orden l . Se prefiere aplicar la ganancia de compensación difusa analizada con respecto a la figura 6 solo a la porción difusa generada por el bloque 823. Esto se puede hacer ventajosamente dentro del procedimiento realizado por el ponderador (difuso). Por lo tanto, solo se mejora la porción difusa en la señal para compensar la pérdida de energía difusa incurrida por componentes superiores que no reciben una síntesis completa como se ilustra en la figura 10.

En la figura 11 se ilustra una generación de solo porción directa para el generador de componentes de alto orden. Básicamente, el generador de componentes de alto orden se implementa de la misma manera que el generador de componentes de orden medio con respecto a la ramificación directa, pero no comprende los bloques 823, 824, 825 y 826. Por lo tanto, el generador de componentes de alto orden solo comprende el ponderador (directo) 822 que recibe datos de entrada de la calculadora de ganancia direccional 410 y que recibe una señal de referencia del generador de señal de referencia 821. Preferentemente, solo se genera una única señal de referencia para el generador de componentes de alto orden y el generador de componentes de orden medio. Sin embargo, ambos bloques también pueden tener generadores de señales de referencia individuales como sea el caso. Sin embargo, se prefiere tener solo un único generador de señales de referencia. Por lo tanto, el procesamiento realizado por el generador de componentes de alto orden es extremadamente eficiente, ya que solo se debe realizar una única dirección de ponderación con una cierta ganancia direccional G^m con una cierta información de difusión Ψ para el mosaico de tiempo/frecuencia. Por lo tanto, los componentes de campo de sonido de alto orden se pueden generar de manera extremadamente eficiente y rápida y cualquier error debido a la no generación de componentes difusos o el no uso de componentes difusos en la señal de salida se compensa fácilmente mediante la mejora de los componentes de campo de sonido de bajo orden o la única porción preferentemente difusa de los componentes de campo de sonido de orden medio.

Típicamente, la porción difusa no estará disponible por separado dentro de los componentes de campo de sonido de bajo orden generados mediante copia o mediante la realización de una combinación lineal (ponderada). Sin embargo, mejorar la energía de dichos componentes mejora automáticamente la energía de la porción difusa. La mejora concurrente de la energía de la porción directa no es problemática como se ha descubierto por los inventores.

Posteriormente, se hace referencia a las figuras 12A a 12C para ilustrar aún más el cálculo de las ganancias de compensación individual.

La figura 12a ilustra una implementación preferida del generador de componentes de sonido 650 de la figura 6. La ganancia de compensación (difusa) se calcula, en una realización, utilizando el valor de difusión, el orden máximo H y el orden de truncamiento L . En la otra realización, la ganancia de compensación difusa se calcula utilizando el parámetro L_k derivado de la cantidad de componentes en la ramificación de procesamiento de bajo orden 810. Además, el parámetro m_k se utiliza dependiendo del parámetro l_k y del número K de componentes realmente generados por el generador de componentes de bajo orden. Además, también se utiliza el valor N que depende de L_k . Ambos valores H , L en la primera realización o H , L_k , m_k generalmente representan el número de componentes de campo de sonido en el segundo grupo (relacionado con el número de componentes de sonido en el primer grupo). Por lo tanto, cuantos más componentes haya para los cuales no se sintetice ningún componente difuso, mayor será la ganancia de compensación de energía. Por otro lado, cuanto mayor sea el número de componentes de campo de sonido de bajo orden que se pueden compensar, es decir, multiplicarse por el factor de ganancia, menor puede ser el factor de ganancia. Generalmente, el factor de ganancia g siempre será mayor que 1.

La figura 12a ilustra el cálculo del factor de ganancia g mediante la calculadora de ganancia de compensación (difusa) 910 y la aplicación posterior de este factor de ganancia al componente (de bajo orden) que se "corrige" como se hace mediante el aplicador de ganancia de compensación 900. En caso de números lineales, el aplicador de ganancia de compensación será un multiplicador, y en caso de números logarítmicos, el aplicador de ganancia de compensación será un sumador. Sin embargo, otras implementaciones de la aplicación de ganancia de compensación se pueden implementar dependiendo de la naturaleza específica y la forma de calcular la ganancia de compensación por el bloque 910. Por lo tanto, la ganancia no necesariamente tiene que ser una ganancia multiplicativa, sino que también puede ser cualquier otra ganancia.

La figura 12b ilustra una tercera implementación para el procesamiento de ganancia de compensación (directa). Una calculadora de ganancia de compensación (directa) 920 recibe, como entrada, la medición relacionada con energía o amplitud para el componente omnidireccional indicado como "potencia omnidireccional" en la figura 12b. Además, la segunda medición relacionada con energía o amplitud para el componente direccional también se ingresa en el bloque 920 como "direccional de potencia". Además, la calculadora de ganancia de compensación directa 920 recibe además la información Q_L o, alternativamente, la información N . N es igual a $(2l + 1)$ que es el número de coeficientes por orden l , y Q_l es igual a $1/N$. Además, la ganancia direccional G^m para el mosaico de tiempo/frecuencia determinado (k, n) también se requiere para el cálculo de la ganancia de compensación (directa). La ganancia direccional es los mismos datos que se derivan de la calculadora de ganancia direccional 410 de la figura 4, a modo de ejemplo. La

ganancia de compensación (directa) g se reenvía desde el bloque 920 al aplicador de ganancia de compensación 900 que se puede implementar de manera similar al bloque 900, es decir, recibe el componente o los componentes que se van a "corregir" y produce el componente corregido o los componentes corregidos.

5 La figura 12c ilustra una implementación preferida de la combinación de la compensación de energía de los componentes de sonido difusos y la normalización de energía de la compensación de componentes de sonido directos que se van a realizar conjuntamente. Con este fin, la ganancia de compensación (difusa) g y la ganancia de compensación (directa) g_s se introducen en un combinador de ganancia 930. El resultado del combinador de ganancia (es decir, la ganancia combinada) se ingresa en un manipulador de ganancia 940 que se implementa como un
10 posprocesador y realiza una limitación a un valor mínimo o máximo o que aplica una función de compresión con el fin de realizar algún tipo de limitación más suave o realiza un suavizado entre los mosaicos de tiempo o frecuencia. La ganancia manipulada que se limita se comprime o suaviza o procesa de otras maneras de posprocesamiento y la ganancia posprocesada se aplica luego por el aplicador de ganancia a uno o más componentes de bajo orden para obtener componentes de bajo orden corregidos.

15 En caso de ganancias lineales g , g_s , el combinador de ganancias 930 se implementa como un multiplicador. En caso de ganancias logarítmicas, el combinador de ganancias se implementa como un sumador. Además, con respecto a la implementación del estimador de la figura 7 indicado en el número de referencia 620, se describe que el estimador 620 puede proporcionar cualquier medición relacionada con energía o amplitud para el componente omnidireccional y direccional siempre que la torre aplicada a la amplitud sea mayor que 1. En el caso de una potencia como la medición
20 relacionada con energía o amplitud, el exponente es igual a 2. Sin embargo, también son útiles exponentes entre 1,5 y 2,5. Además, incluso exponentes o potencias más altos son útiles, tal como una potencia de 3 aplicada a la amplitud correspondiente a un valor de sonoridad en lugar de un valor de potencia. Por lo tanto, en general, se prefieren potencias de 2 o 3 para proporcionar las mediciones relacionadas con energía o amplitud, pero también se prefieren generalmente potencias entre 1,5 y 4.

Una señal de audio codificada de forma inventiva se puede almacenar en un medio de almacenamiento digital o un medio de almacenamiento no transitorio o se puede transmitir en un medio de transmisión tal como un medio de
30 transmisión inalámbrico o un medio de transmisión alámbrico tal como Internet.

Aunque que algunos aspectos se han descrito en el contexto de un aparato, es evidente que estos aspectos también representan una descripción del método correspondiente, donde un bloque o aparato corresponde a una etapa de procedimiento o una característica de una etapa de procedimiento. Análogamente, los aspectos descritos en el contexto de un paso de método también representan una descripción de un bloque o elemento o característica
35 correspondiente de un aparato correspondiente.

Dependiendo de ciertos requisitos de implementación, las realizaciones se pueden implementar en hardware o en software. La implementación se puede realizar utilizando un medio de almacenamiento digital, por ejemplo, un disquete, un DVD, un CD, una ROM, una PROM, una EPROM, una EEPROM o una memoria FLASH, que tenga
40 señales de control electrónicamente legibles almacenadas en el mismo, que cooperan (o sean capaces de cooperar) con un sistema informático programable de manera que se realiza el método respectivo.

Algunas realizaciones comprenden un portador de datos que tiene señales de control electrónicamente legibles, que son capaces de cooperar con un sistema informático programable, de modo que se lleva a cabo uno de los métodos
45 descritos en el presente documento.

En general, realizaciones se pueden implementar como un producto de programa informático con un código de programa, el código de programa que es operativo para realizar uno de los métodos, cuando el producto de programa informático se ejecuta en un ordenador. El código de programa se puede por ejemplo almacenar en un portador legible
50 por máquina.

Otras realizaciones comprenden el programa informático para realizar uno de los métodos descritos en el presente documento, almacenado en un portador legible por máquina o un medio de almacenamiento no transitorio.

55 En otras palabras, una realización del método inventivo es, por lo tanto, un programa informático que tiene un código de programa para realizar uno de los métodos descritos en el presente documento, cuando el programa informático se ejecuta en un ordenador.

Una realización adicional es un portador de datos (o un medio de almacenamiento digital, o un medio legible por ordenador) que comprende, registrado en el mismo, el programa informático para realizar uno de los métodos descritos en el presente documento.
60

Una realización adicional es un flujo de datos o una secuencia de señales que representan el programa informático

para realizar uno de los métodos descritos en el presente documento. El flujo de datos o la secuencia de señales se pueden por ejemplo configurar para transferirse a través de una conexión de comunicación de datos, por ejemplo, a través de Internet.

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

Una realización comprende además un ordenador que tiene instalado en la misma el programa informático para realizar uno de los métodos descritos en el presente documento.

- 10 En algunas realizaciones, un dispositivo lógico programable (por ejemplo, una matriz de puertas programable en el campo) se puede usar para realizar algunas o todas las funcionalidades de los métodos descritos en el presente documento. En algunas realizaciones, una matriz de puertas programable en el campo puede cooperar con un microprocesador con el fin de realizar uno de los métodos descritos en el presente documento. En general, los métodos se llevan a cabo preferentemente mediante cualquier aparato de hardware.

- 15 Las realizaciones anteriormente descritas son solo ilustrativas de los principios de la presente invención. Se entiende que las modificaciones y variaciones de los arreglos y los detalles descritos en el presente documento serán evidentes para otros expertos en la técnica. Por lo tanto, se pretende que se limite solo por el alcance de las reivindicaciones de patente inminentes y no por los detalles específicos presentados a modo de descripción y explicación de las realizaciones de la presente.

Referencias

- 25 [1] V. Pulkki, M-V Laitinen, J Vilkkamo, J Ahonen, T Lokki y T Pihlajamäki, "Directional audio coding - perception-based reproduction of spatial sound", International Workshop on the Principles and Application on Spatial Hearing, noviembre, 2009, Zao; Miyagi, Japón.
- 30 [2] M. V. Laitinen y V. Pulkki, "Converting 5.1 audio recordings to B-format for directional audio coding reproduction", 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Praga, 2011, pp. 61-64.
- [3] R. K. Furness, "Ambisonics —An overview", en AES 8th International Conference, abril, 1990, págs. 181—189.
- 35 [4] C. Nachbar, F. Zotter, E. Deleflie y A. Sontacchi, "AMBIX – A Suggested Ambisonics Format", Proceedings of the Ambisonics Symposium 2011
- [5] "APPARATUS, METHOD OR COMPUTER PROGRAM FOR GENERATING A SOUND FIELD DESCRIPTION" (que corresponde a WO 2017/157803 A1).
- 40 [6] Documento US 8 891 797 B2
- [7] Documento US 2018/315432 A1

45

REIVINDICACIONES

1. Un aparato para generar una descripción de campo de sonido a partir de una señal de entrada que comprende al menos dos canales, comprendiendo el aparato:
 - 5 un analizador de señal de entrada (600) para obtener datos de dirección y datos de difusión a partir de la señal de entrada;
 - 10 un estimador (620) para estimar una primera medición relacionada con energía o amplitud para un componente omnidireccional derivado de la señal de entrada y para estimar una segunda medición relacionada con energía o amplitud para un componente direccional derivado de la señal de entrada; y
 - 15 un generador de componentes de sonido (750) para generar componentes de campo de sonido del campo de sonido, en donde el generador de componentes de sonido se configura para realizar una compensación de energía del componente direccional usando la primera medición de amplitud, la segunda medición relacionada con energía o amplitud, los datos de dirección y los datos de difusión.
2. Aparato según la reivindicación 1, en el que la señal de entrada comprende al menos dos canales, en el que el estimador (620) está configurado para calcular el componente omnidireccional usando una suma de al menos dos canales, y para calcular el componente direccional usando una resta de los al menos dos canales (815).
3. Aparato según la reivindicación 1, en el que la señal de entrada comprende el componente omnidireccional y uno o más componentes direccionales, y en el que el estimador (620) está configurado para calcular la primera medida relacionada con la amplitud para el componente omnidireccional usando la señal de entrada y para calcular la segunda medición relacionada con la amplitud o energía para cada uno de los uno o más componentes direccionales de la señal de entrada.
4. Aparato según la reivindicación 1, en el que la señal de entrada comprende una representación en formato A o formato B con al menos dos canales, y en el que el estimador (620) está configurado para derivar (816) el componente omnidireccional y los componentes direccionales usando una combinación lineal ponderada de al menos dos canales.
5. Aparato según una de las reivindicaciones anteriores, en el que el analizador de señal de entrada (600) está configurado para extraer los datos de difusión de los metadatos asociados con la señal de entrada o para extraer los datos de difusión de la señal de entrada mediante un análisis de señal (610, 620) de la señal de entrada que tiene al menos dos canales o componentes.
6. Aparato según una de las reivindicaciones anteriores, en el que el estimador (620) está configurado para calcular la primera medida relacionada con la energía o la amplitud o la segunda medida relacionada con la energía o la amplitud a partir de un valor absoluto de una amplitud compleja o una magnitud elevada a una potencia mayor que 1 e inferior a 5 o igual a 2 o 3.
7. Aparato según una de las reivindicaciones anteriores, en el que el generador de componentes de sonido (750) comprende un compensador de energía (900-940) para realizar la compensación de energía, comprendiendo el compensador de energía una calculadora de ganancia de compensación (910, 920, 930, 940) para calcular una ganancia de compensación usando la primera medida relacionada con la energía o la amplitud, la segunda medida relacionada con la energía o la amplitud, los datos de dirección y los datos de difusión.
8. Aparato según una de las reivindicaciones anteriores, en el que el generador de componentes de sonido (750) está configurado para calcular (410), a partir de los datos de dirección, una ganancia direccional y para combinar (920) la ganancia direccional y los datos de difusión para realizar la compensación de energía.
9. Aparato según una de las reivindicaciones anteriores, en el que el estimador (620) está configurado para estimar la segunda medida relacionada con la energía o la amplitud para un primer componente direccional y una tercera medida relacionada con la energía o la amplitud para un segundo componente direccional, para calcular una primera ganancia de compensación para el primer componente direccional usando la primera y la segunda medidas relacionadas con la energía o la amplitud, y para calcular una segunda ganancia de compensación para el segundo componente direccional usando la primera y la tercera medidas relacionadas con la energía o la amplitud.
10. Aparato según la reivindicación 9, en el que el calculador de ganancia de compensación (910, 920) está

configurado para calcular (910) un primer factor de ganancia dependiendo de los datos de difusión y al menos uno del número de componentes de campo sonoro en el segundo grupo, el orden máximo de componentes de campo sonoro del primer grupo y el número máximo de componentes de campo sonoro del segundo grupo, para calcular (920) un segundo factor de ganancia dependiendo de la primera medida relacionada con la energía o la amplitud para el componente omnidireccional, la segunda medida relacionada con la energía o la amplitud para el componente omnidireccional, o medida relacionada con la amplitud para el componente direccional, los datos de dirección y los datos de difusión, y para calcular (930) la ganancia de compensación usando el primer factor de ganancia y el segundo factor de ganancia, en donde el generador de componente de sonido (750) está configurado para usar los mismos datos de dirección y datos de difusión para calcular la primera ganancia de compensación y la segunda ganancia de compensación.

11. Aparato según una de las reivindicaciones 7 a 10, en el que el calculador de ganancia de compensación (910) está configurado para calcular un factor de ganancia basándose en la siguiente ecuación

$$g_s = \sqrt{\frac{E\{|B_0^0|^2\}}{E\{|B_l^m|^2\}} (Q_l \Psi + (1 - \Psi)(G_l^m(\varphi, \theta))^2)},$$

$$g_s = \sqrt{\Psi + \frac{E\{|B_0^0|^2\}}{E\{|B_l^m|^2\}} (1 - \Psi)(G_l^m(\varphi, \theta))^2},$$

en donde Ψ representa los datos de difusión, $E\{|B_0^0|^2\}$ representa la primera medición relacionada con la energía o la amplitud, $E\{|B_l^m|^2\}$ representa la segunda medición relacionada con la energía o la amplitud, B_0^0 es el componente omnidireccional, B_l^m es el componente direccional, $(G_l^m(\varphi, \theta))$ es una ganancia de dirección derivada de los datos de dirección φ , θ , φ es un ángulo de azimut, θ es un ángulo de elevación. Q_l es un factor de directividad de orden l , y g_s es el factor de ganancia.

12. Aparato según una de las reivindicaciones 7 a 11, en el que la calculadora de ganancia de compensación (910) está configurada

para aumentar la ganancia de compensación con una primera medición creciente relacionada con la energía o la amplitud, o

para disminuir la ganancia de compensación con una segunda medición creciente relacionada con la energía o la amplitud, o

para aumentar la ganancia de compensación usando una ganancia de dirección creciente, o

para aumentar la ganancia de compensación con un número decreciente de componentes direccionales.

13. Aparato según una de las reivindicaciones 7 a 12, en el que un generador de componentes de sonido (650) para generar, a partir de la señal de entrada, uno o más componentes de campo de sonido de un primer grupo de componentes de campo de sonido que tiene para cada componente de campo de sonido un

componente directo y un componente difuso, y para generar, a partir de la señal de entrada, un segundo grupo de componentes de campo sonoro que tienen solo un componente directo,

en el que el calculador de ganancia de compensación (910) está configurado para calcular la ganancia de compensación usando los datos de difusividad y al menos uno de los número de componentes de campo sonoro en el segundo grupo, el número de componentes difusos en el primer grupo, un orden máximo de componentes de campo sonoro del primer grupo, y un orden máximo de componentes de campo sonoro del segundo grupo.

14. Aparato según una de las reivindicaciones 7 a 13, en el que el calculador de ganancia de compensación (910, 920, 930, 940) está configurado para realizar una manipulación del factor de ganancia (940) usando una limitación con un umbral máximo fijo o un umbral mínimo fijo o usando una función de compresión para comprimir factores de ganancia bajos o altos hacia factores de ganancia medios para obtener la ganancia de compensación.

15. Aparato según una de las reivindicaciones anteriores, en el que el generador de componentes de sonido (750) está configurado para generar otros componentes de campo de sonido de otros órdenes, en el que un combinador (430) está configurado para combinar los componentes de campo de sonido del campo de sonido y otros componentes de campo de sonido de otros órdenes para obtener una descripción de campo sonoro que tiene un orden superior al orden de la señal de entrada.

16. Aparato según una de las reivindicaciones 7 a 15, en el que el compensador de energía (910, 920, 930, 940) comprende un aplicador de ganancia de compensación (900) para aplicar la ganancia de compensación a al menos un componente del campo sonoro.

17. Aparato según una de las reivindicaciones anteriores, en el que el generador de componentes de sonido (750) comprende un generador de componentes de orden bajo (810) para generar una descripción de campo de sonido de orden bajo desde la señal de entrada hasta un orden predeterminado y el modo predeterminado, en el que el generador de componentes de orden bajo (810) está configurado para derivar la descripción del campo de sonido de orden bajo copiando o tomando la señal de entrada o formando una combinación ponderada de los canales de la señal de entrada,

en donde la descripción del campo de sonido de orden bajo comprende el omnidireccional componente y el componente direccional generado por la copia o la toma o la combinación lineal.

18. Aparato según la reivindicación 17, en el que el generador de componentes de sonido (750) comprende además:

un generador de componentes de orden medio (820) para generar una descripción de campo sonoro de orden medio por encima del orden predeterminado o en el orden predeterminado y por encima del modo predeterminado y por debajo o en un primer orden de truncamiento usando una síntesis de al menos una porción directa y de al menos una porción difusa usando los datos de dirección y los datos de difusión de modo que la descripción del campo sonoro de orden medio comprenda una contribución directa y una contribución difusa; y

un generador de componentes de alto orden (830) para generar una descripción de campo de sonido de alto orden que tiene un componente por encima del primer orden de truncamiento usando una síntesis de al menos una porción directa, en donde la descripción de campo de sonido de alto orden comprende sólo una contribución directa.

19. Aparato según una de las reivindicaciones anteriores, en el que un primer grupo de componentes del campo sonoro hasta un orden I de coeficientes y un segundo grupo de componentes del campo sonoro por encima del orden I de coeficientes son ortogonales entre sí, o en el que los componentes del campo sonoro son al menos uno de coeficientes de funciones de base ortogonales, coeficientes de funciones de base espaciales, coeficientes de armónicos esféricos o circulares y coeficientes Ambisonics.

20. Aparato según una de las reivindicaciones anteriores, un banco de filtros de análisis (400) para generar uno o más componentes de campo sonoro para una pluralidad de mosaicos de tiempo y frecuencia diferentes, en el que el analizador de señal de entrada (600) está configurado para obtener un elemento de datos de difusión para cada mosaico de tiempo-frecuencia, y en el que el generador de componentes de sonido (750) está configurado para realizar la compensación de energía por separado para cada mosaico de tiempo-frecuencia.

21. Método para generar una descripción de campo sonoro a partir de una señal de entrada que comprende al

menos dos canales, que comprende:

obtener datos de dirección y datos de difusión a partir de la señal de entrada;

- 5 estimar una primera medición relacionada con la energía o la amplitud para un componente omnidireccional derivado de la señal de entrada y para estimar una segunda medida relacionada con la energía o la amplitud para un componente direccional derivado de la señal de entrada, y
- 10 generar componentes del campo sonoro del campo sonoro, en el que el generador de componentes de sonido está configurado para realizar una compensación de energía del componente direccional usando la primera medida relacionada con la energía o la amplitud, la segunda medida relacionada con la energía o la amplitud, los datos de dirección y los datos de difusión.
- 15 22. Programa informático para realizar, cuando se ejecuta en un ordenador o en un procesador, el método según la reivindicación 21.

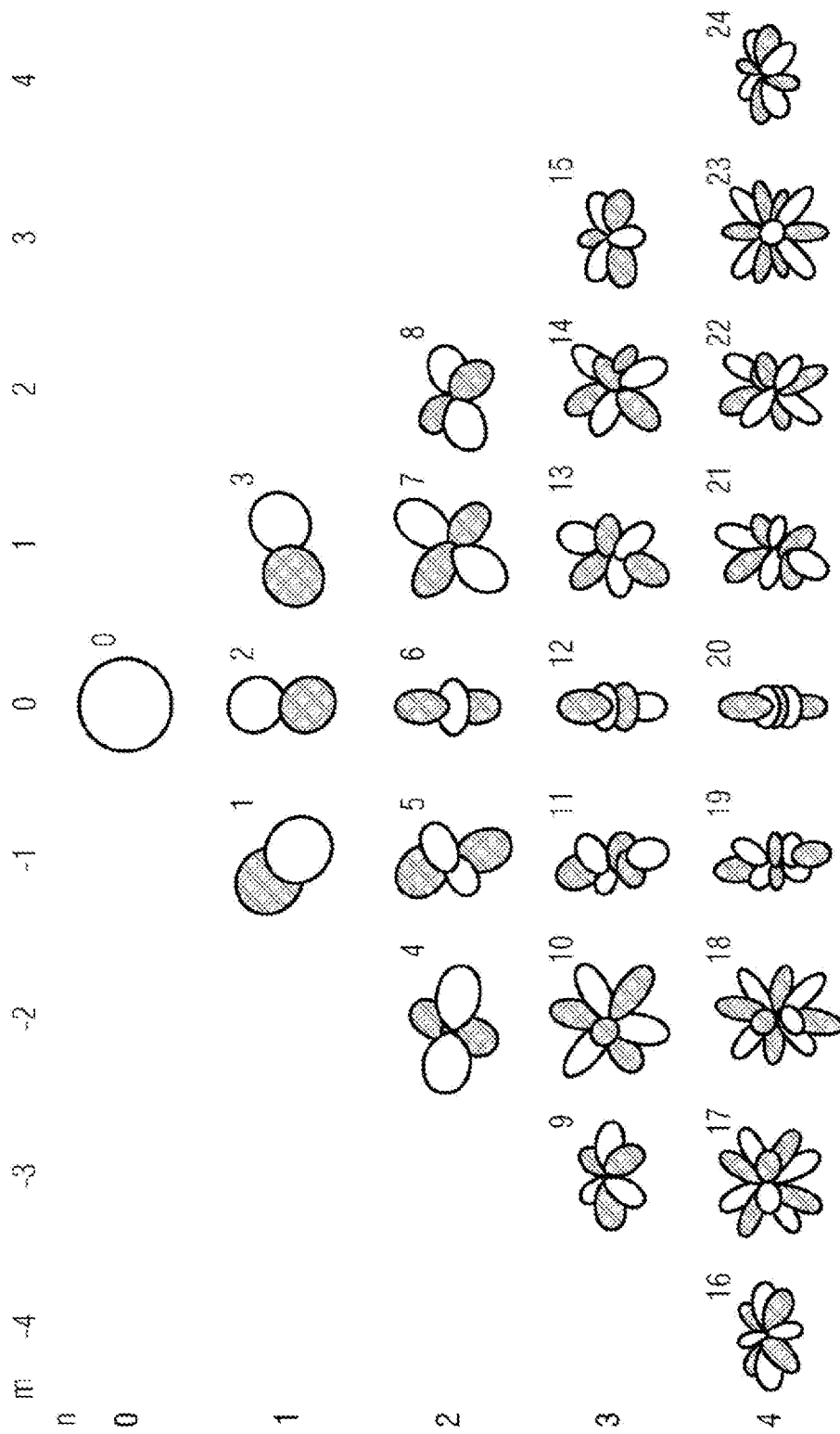


Fig. 1a

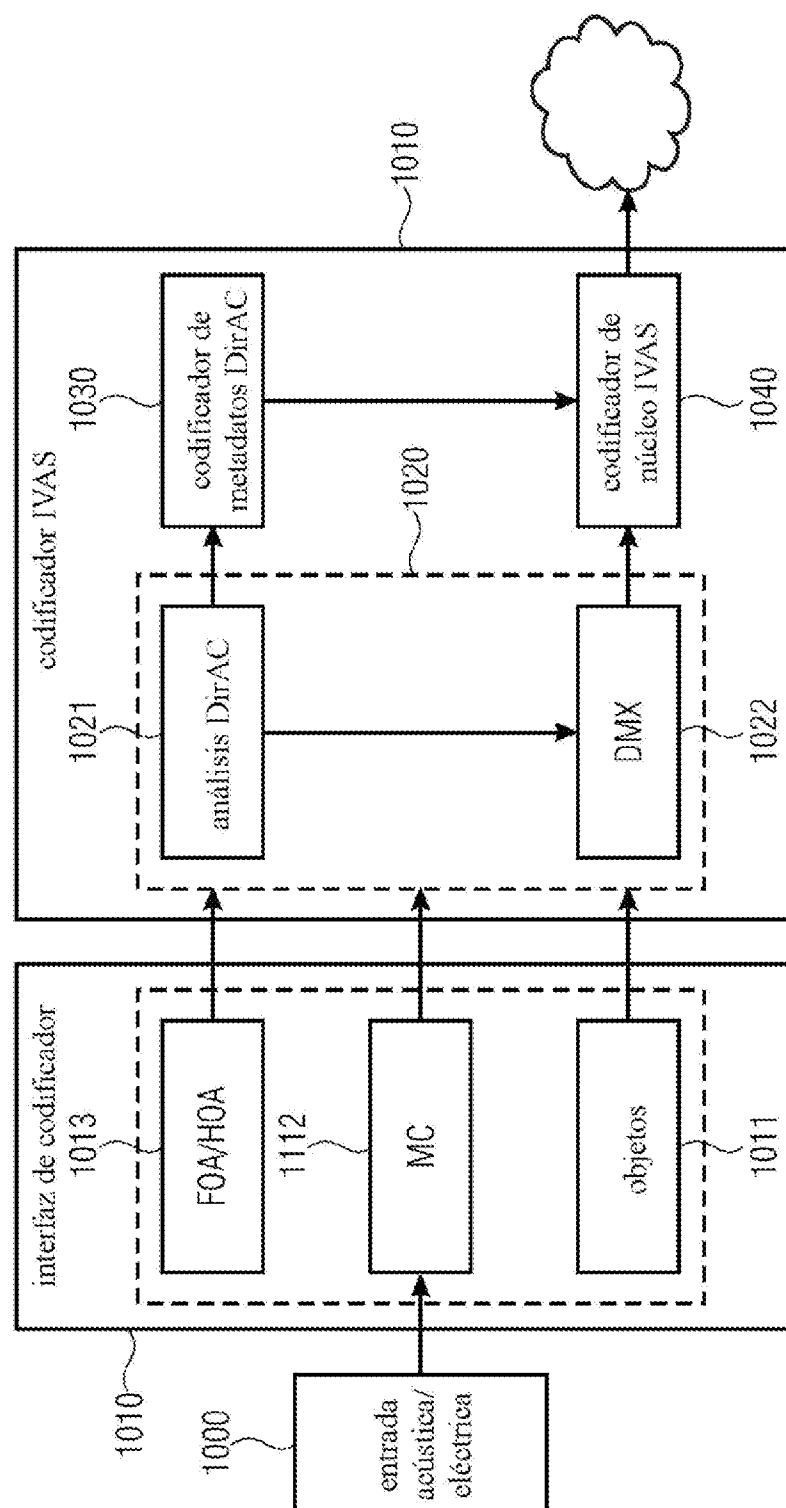


Fig. 1b

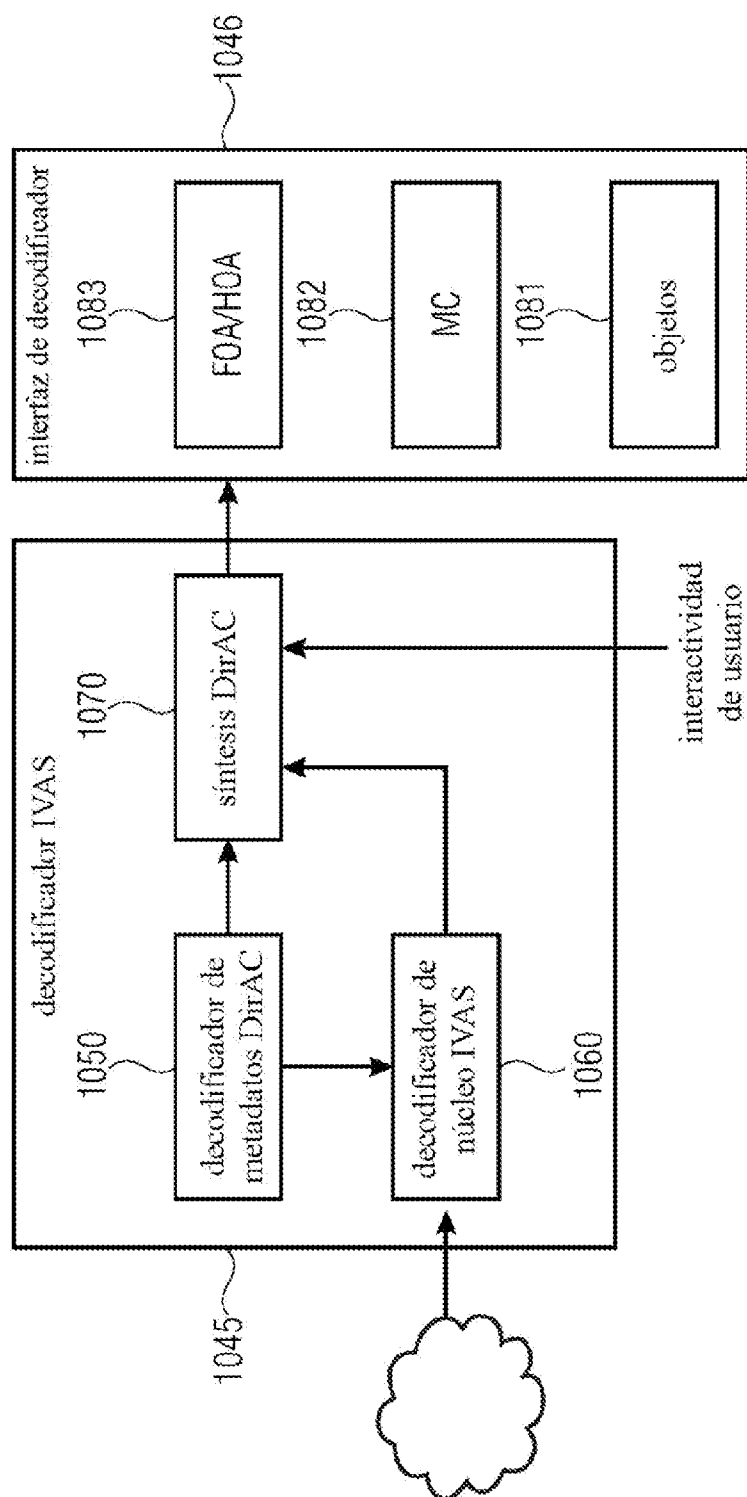


Fig. 2

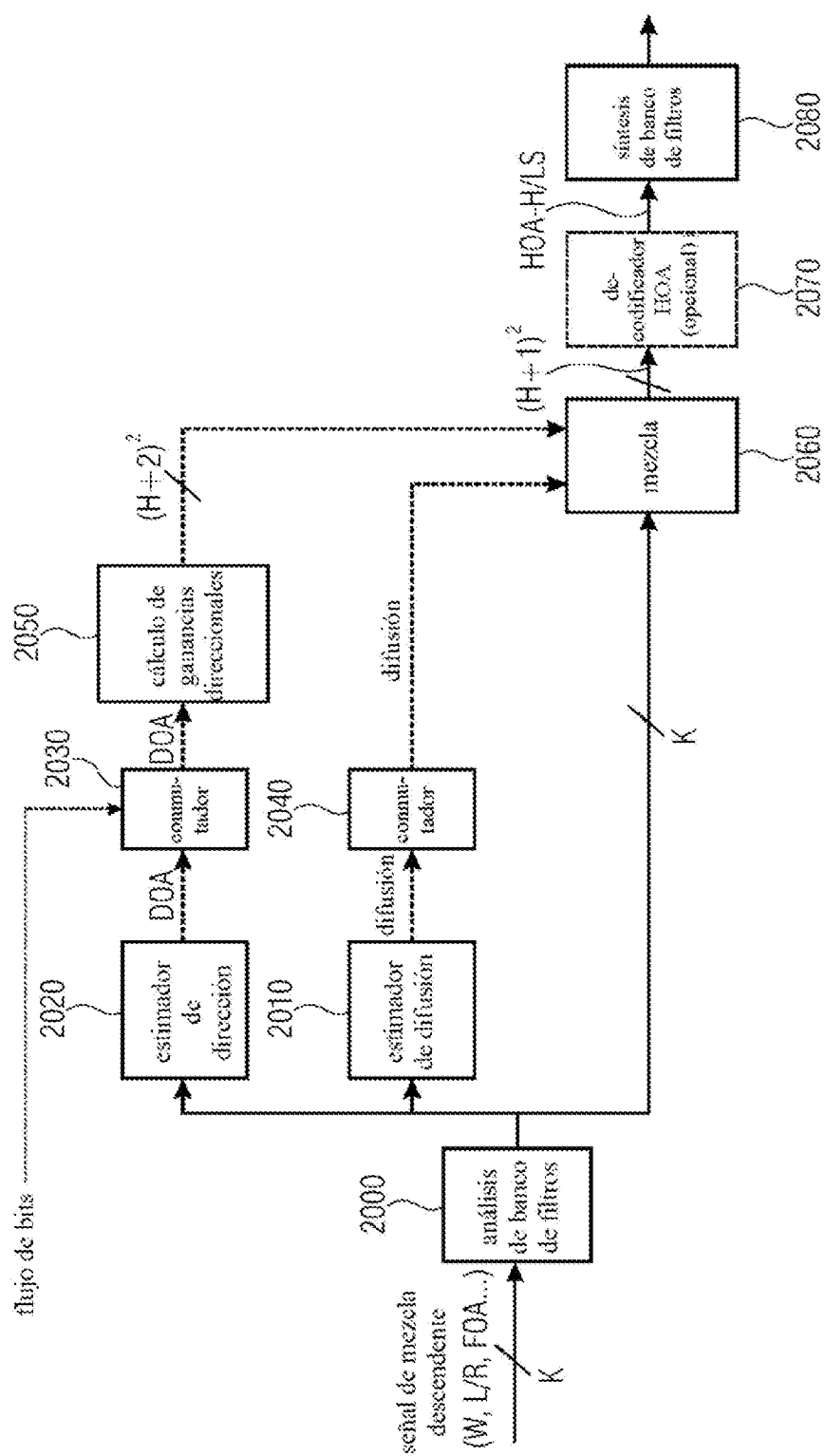


Fig. 3

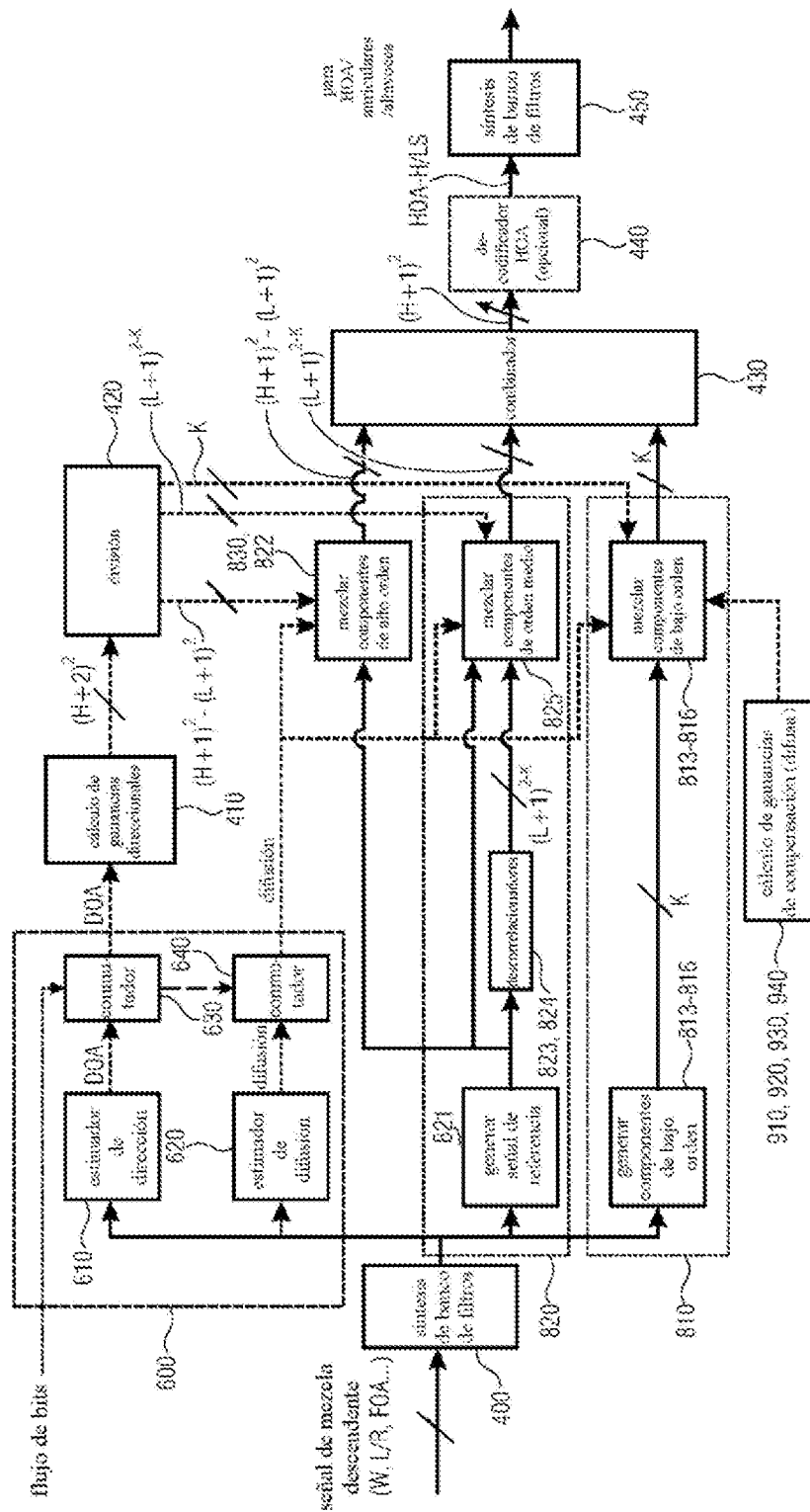


Fig. 4

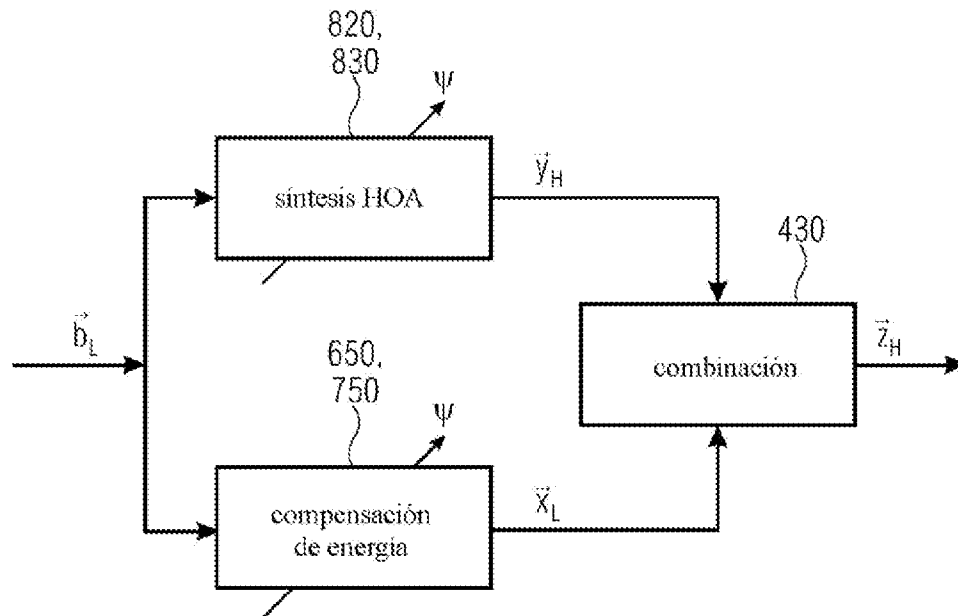


Fig. 5

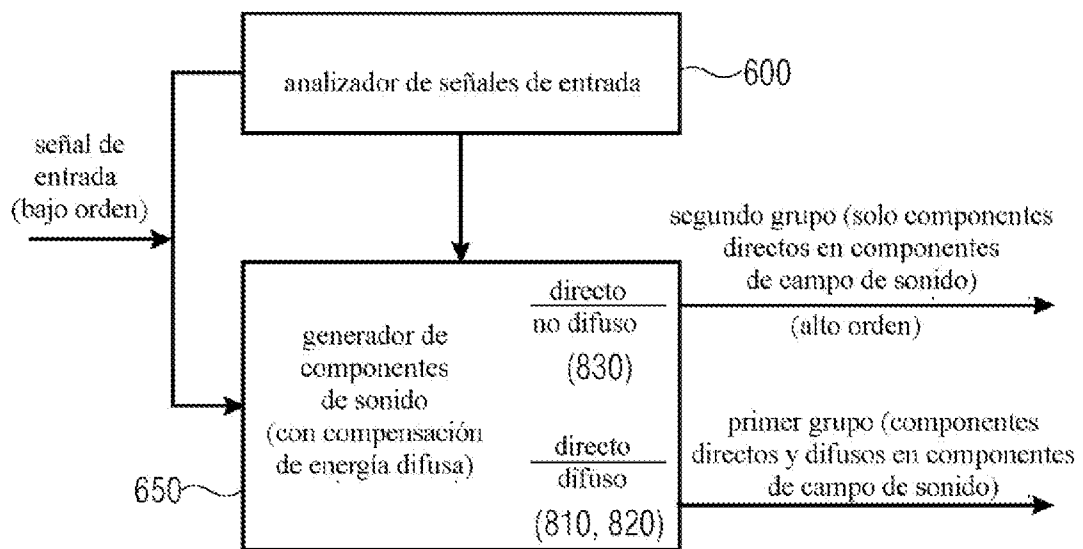


Fig. 6

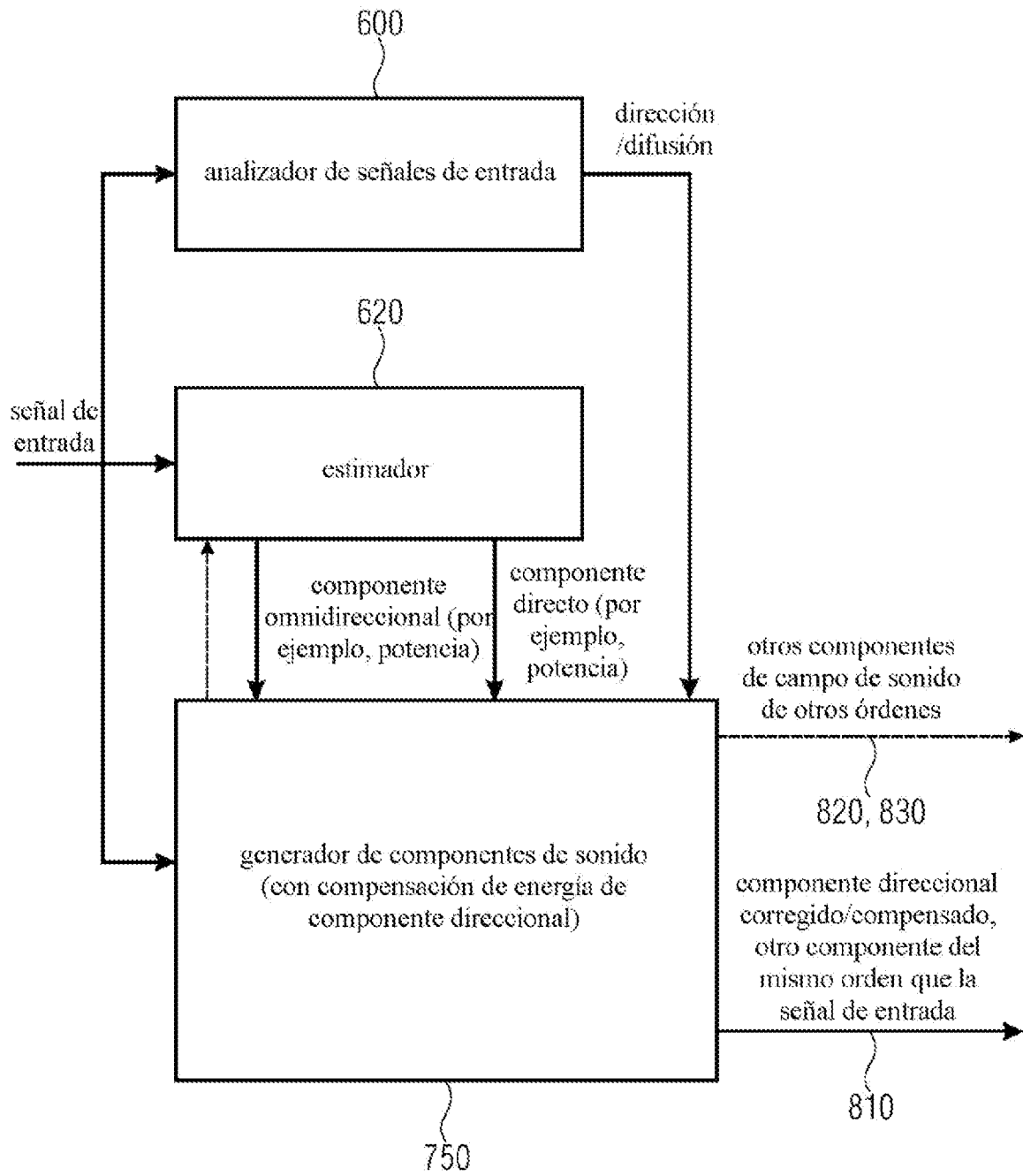


Fig. 7

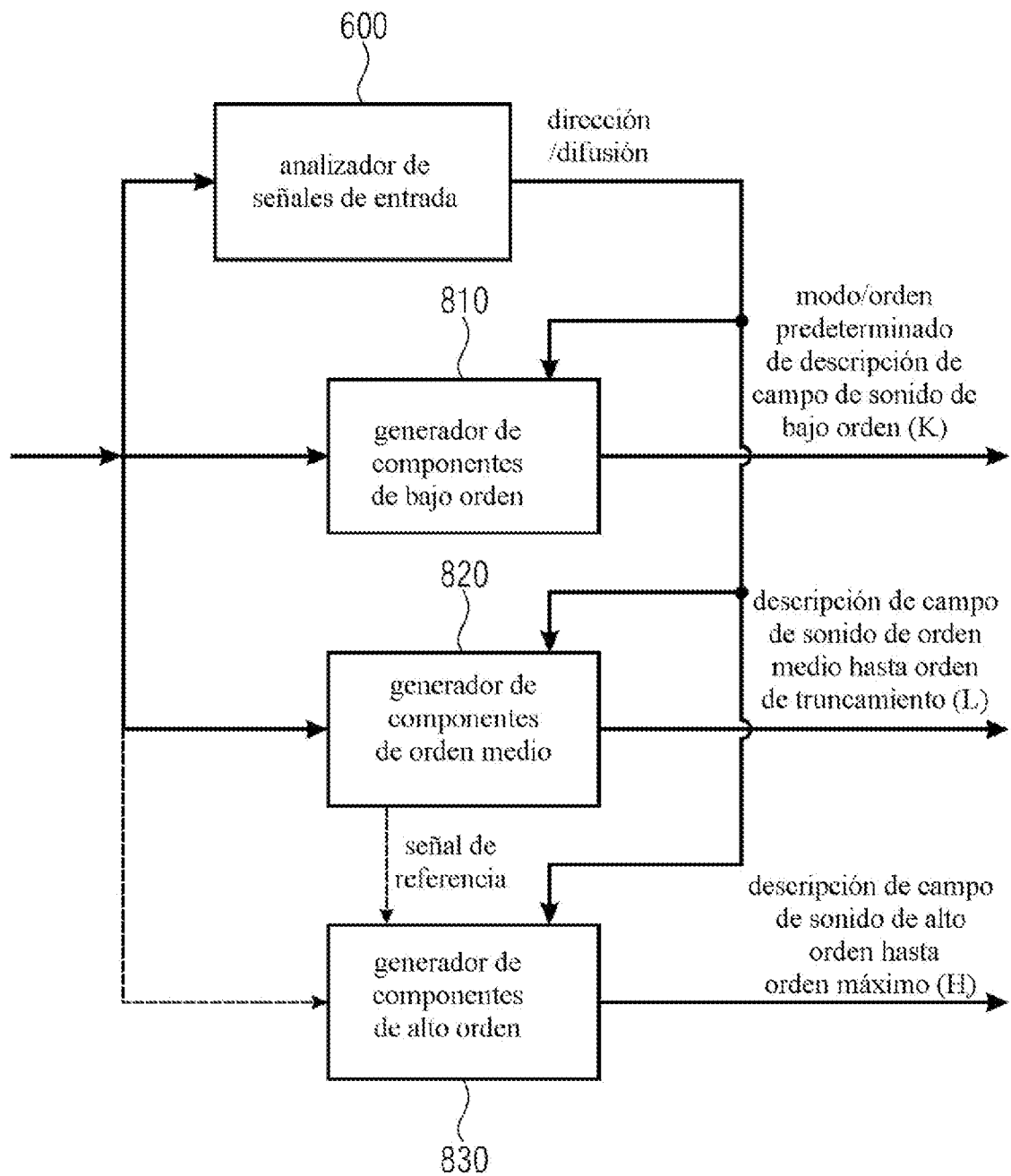


Fig. 8

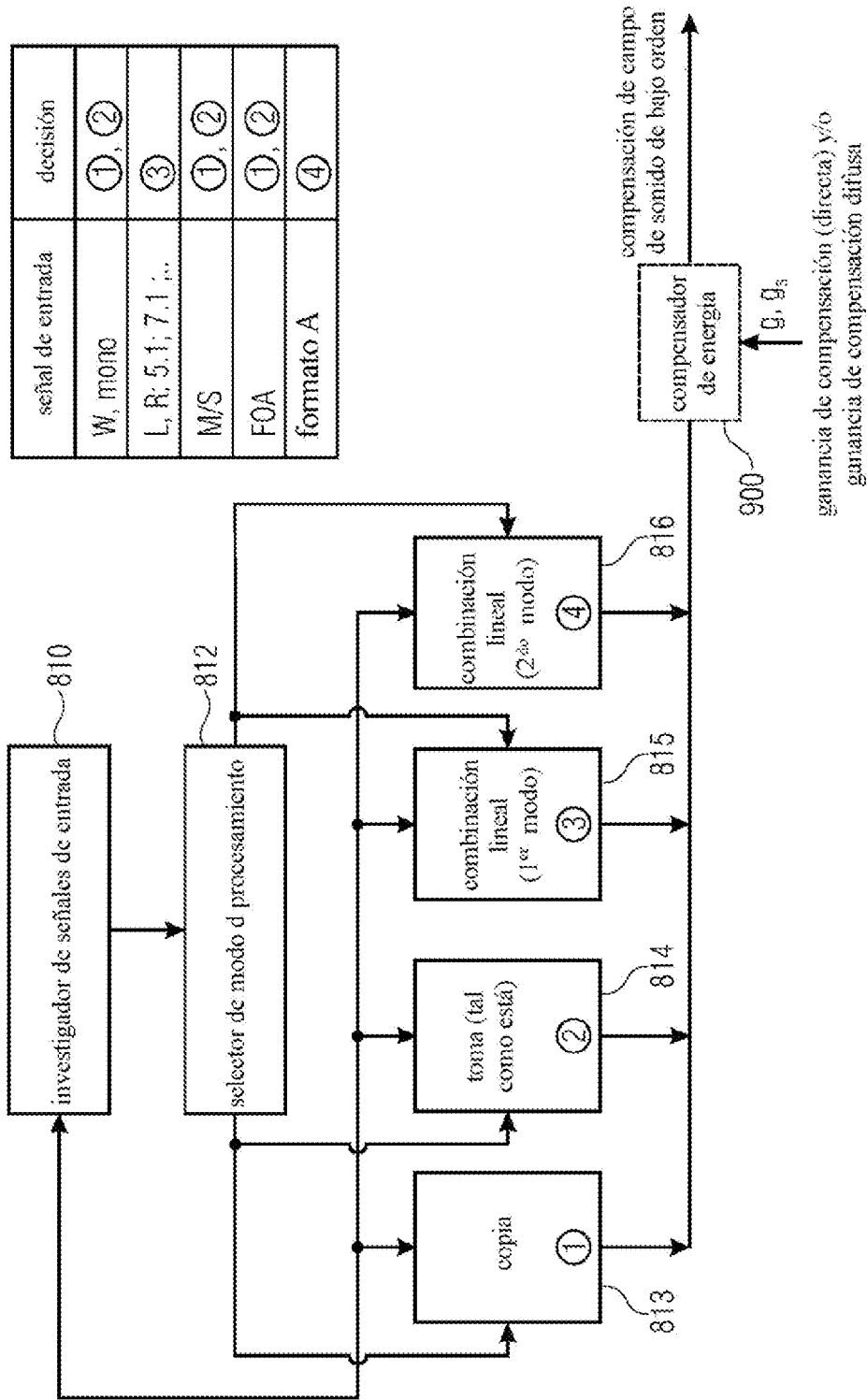


Fig. 9

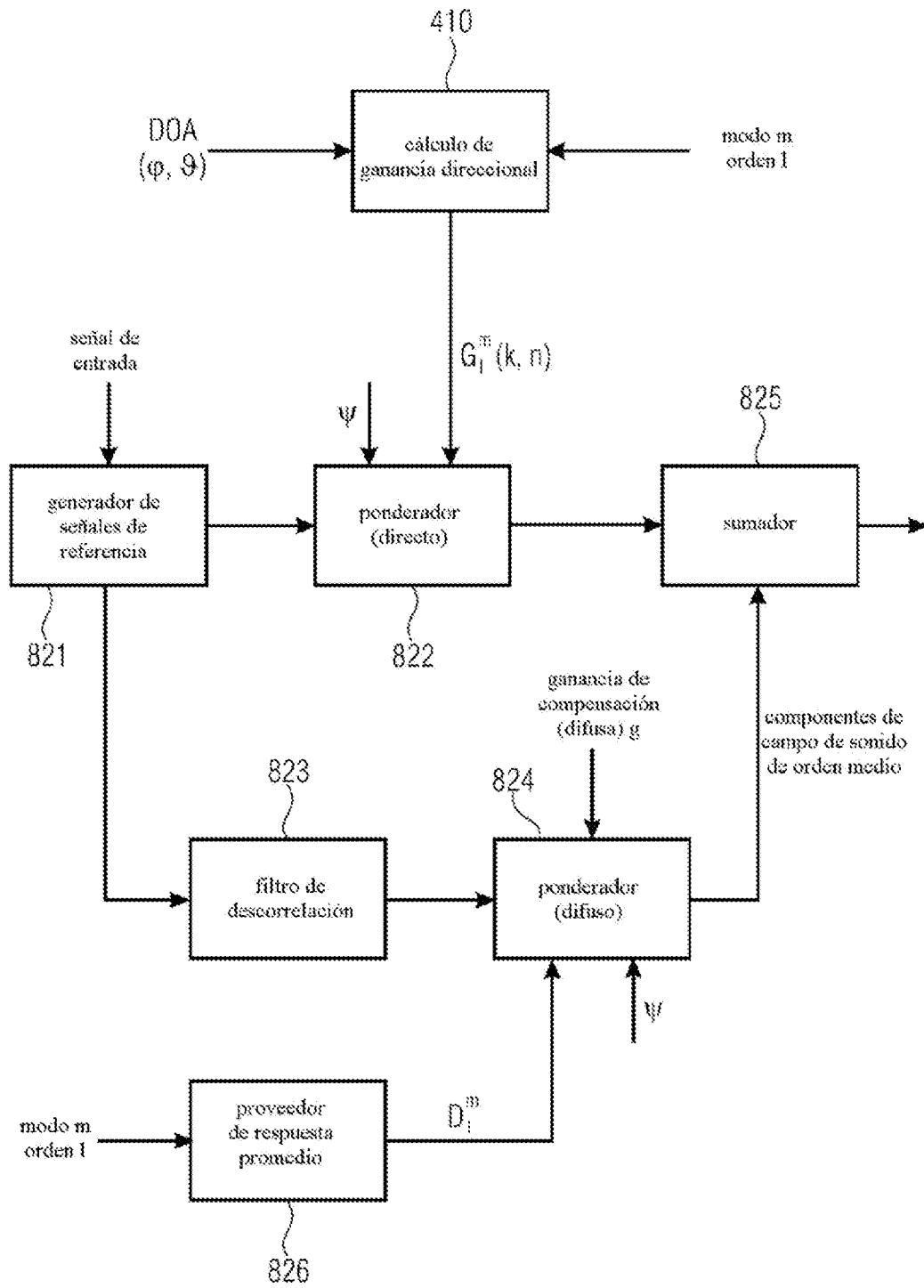


Fig. 10

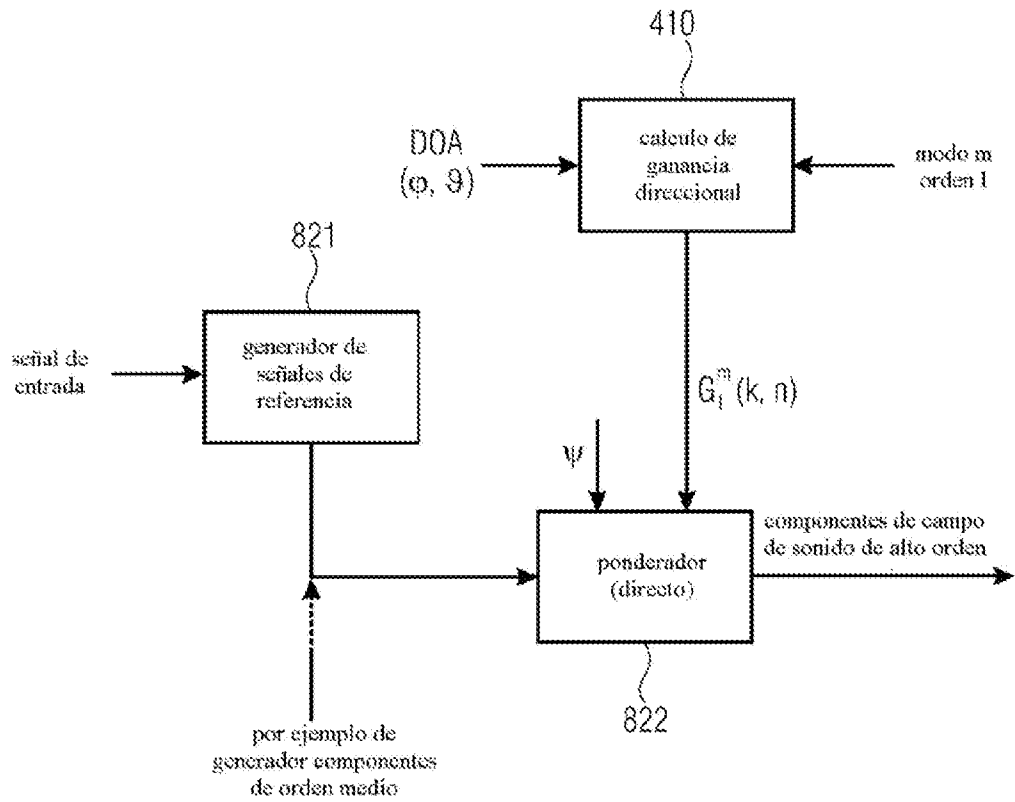


Fig. 11

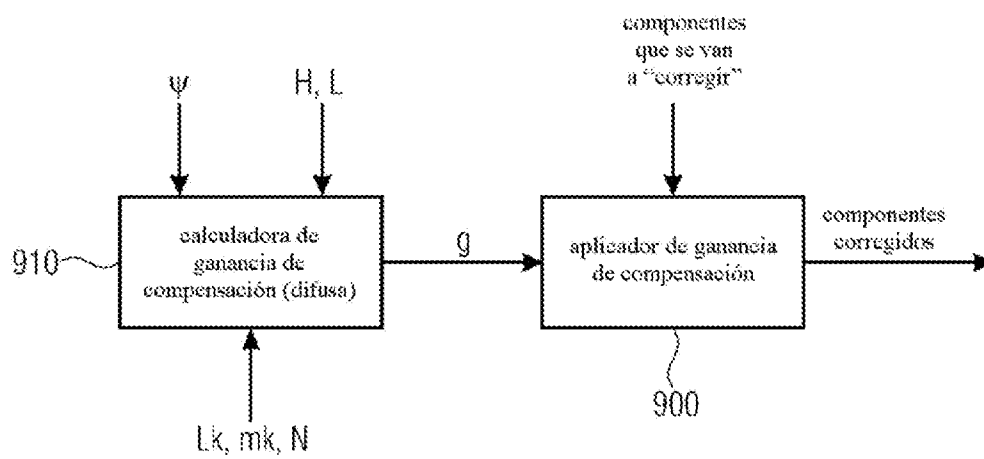


Fig. 12a

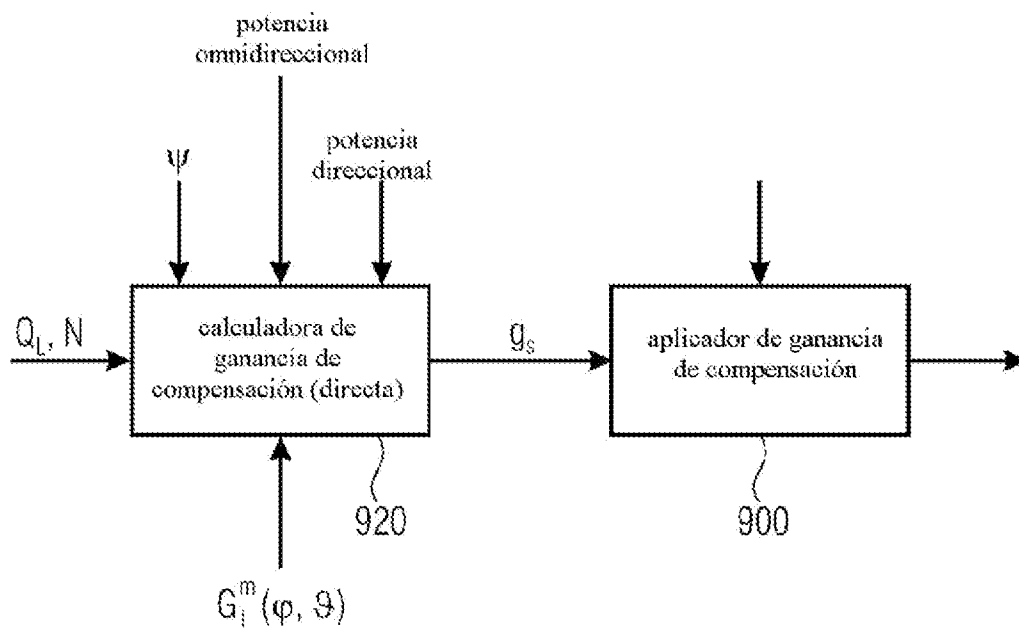


Fig. 12b

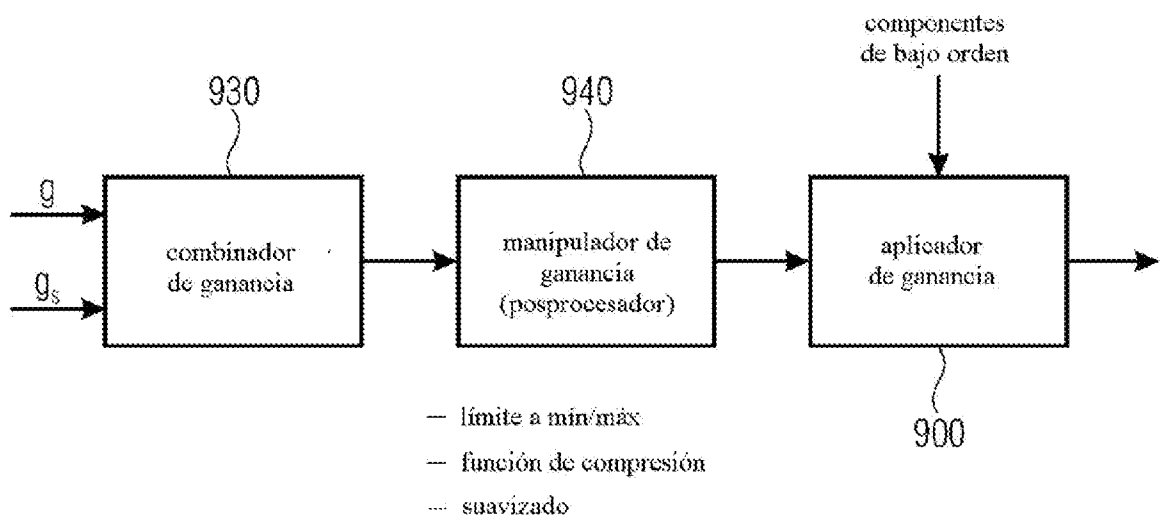


Fig. 12c