

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 965 084**

51 Int. Cl.:

G10L 19/008 (2013.01)

G10L 19/26 (2013.01)

H04S 7/00 (2006.01)

G10L 25/03 (2013.01)

H04S 5/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **24.09.2020 PCT/FR2020/051668**

87 Fecha y número de publicación internacional: **08.04.2021 WO21064311**

96 Fecha de presentación y número de la solicitud europea: **24.09.2020 E 20792467 (1)**

97 Fecha y número de publicación de la concesión europea: **06.09.2023 EP 4042418**

54 Título: **Determinación de correcciones a aplicar a una señal de audio multicanal, codificación y decodificación asociadas**

30 Prioridad:

02.10.2019 FR 1910907

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

10.04.2024

73 Titular/es:

**ORANGE (100.0%)
111, quai du Président Roosevelt
92130 Issy-les-Moulineaux, FR**

72 Inventor/es:

**MAHE, PIERRE CLÉMENT;
RAGOT, STÉPHANE y
DANIEL, JEROME**

74 Agente/Representante:

LEHMANN NOVO, María Isabel

ES 2 965 084 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

Determinación de correcciones a aplicar a una señal de audio multicanal, codificación y decodificación asociadas

5 La presente invención se refiere a la codificación/decodificación de datos de sonido espacializado, en particular en un contexto ambiofónico (en lo sucesivo también denominado "ambisónico").

10 Los codificadores/decodificadores (en lo sucesivo denominados "codificadores") utilizados actualmente en telefonía móvil son mono (un único canal de señal para la reproducción en un único altavoz). El codificador 3GPP EVS (*Enhanced Voice Services*) permite ofrecer una calidad "Super-HD" (también denominada voz de "alta definición plus" o HD+) con una banda de audio en banda superancha (SWB, "*super-wideband*" en inglés) para señales muestreadas a 32 o 48 kHz, o en banda completa (FB, "*Fullband*" en inglés) para señales muestreadas a 48 kHz; el ancho de banda de audio es de 14,4 a 16 kHz en modo SWB (de 9,6 a 128 kbit/s) y de 20 kHz en modo FB (de 16,4 a 128 kbit/s).

15 El próximo avance en la calidad de los servicios conversacionales ofrecidos por los operadores debería estar constituido por los servicios inmersivos, utilizando terminales tales como teléfonos inteligentes equipados con varios micrófonos o equipos de audioconferencia espacializados o videoconferencia de tipo telepresencia o vídeo de 360°, o incluso equipos para compartir contenidos de audio "en directo", con un renderizado de sonido espacializado en 3D mucho más inmersiva que una simple reproducción estéreo 2D. Debido al uso cada vez más extendido de la escucha por teléfono móvil con auriculares y la aparición de equipos de audio avanzados (accesorios tales como un micrófono 3D, asistentes de voz con antenas acústicas, auriculares de realidad virtual, etc.), la captura y el renderizado de escenas sonoras espacializadas son ya lo suficientemente comunes como para ofrecer una experiencia de comunicación inmersiva.

25 En este sentido, la futura norma 3GPP "IVAS" (*Immersive Voice And Audio Services*) propone ampliar la codificación EVS al audio inmersivo aceptando como formato de entrada de codificación al menos los formatos de sonido espacializado que se enumeran a continuación (y sus combinaciones):

- 30 - Formato multicanal ("*channel-based*" en inglés) de tipo estéreo o 5.1 donde cada canal alimenta un altavoz (por ejemplo L y R en estéreo o L, R, Ls, Rs y C en 5.1);
- Formato objeto ("*object-based*" en inglés) donde los objetos sonoros se describen como una señal de audio (generalmente mono) asociada a metadatos que describen los atributos de este objeto (posición en el espacio, anchura espacial de la fuente, etc.),
- 35 - Formato ambisónico ("*scene-based*" en inglés) que describe el campo sonoro en un punto dado, generalmente captado por un micrófono esférico o sintetizado en el dominio de los armónicos esféricos.

40 Lo que sigue se refiere típicamente a la codificación de sonido en formato ambisónico, a modo de ejemplo (al menos algunos aspectos presentados a continuación en relación con la invención pueden aplicarse también a formatos distintos del ambisónico).

45 La ambisonía es un método de grabación ("codificación" en el sentido acústico) de sonido espacializado y un sistema de reproducción ("decodificación" en el sentido acústico). Un micrófono ambisónico (de orden 1) comprende al menos cuatro cápsulas (típicamente de tipo cardioide o subcardioide) dispuestas en una rejilla esférica, por ejemplo los vértices de un tetraedro regular. Los canales de audio asociados a estas cápsulas se denominan "formato A". Este formato se convierte en un "formato B", en el que el campo sonoro se descompone en cuatro componentes (armónicos esféricos) denominadas W, X, Y, Z, que corresponden a cuatro micrófonos virtuales coincidentes. La componente W corresponde a una captación omnidireccional del campo sonoro, mientras que las componentes X, Y y Z, más direccionales, pueden asemejarse a micrófonos con gradientes de presión orientados a lo largo de los tres ejes ortogonales del espacio. Un sistema ambisónico es un sistema flexible en el sentido de que la grabación y la reproducción están separadas y desacopladas. Permite la decodificación (en el sentido acústico) en cualquier configuración de altavoces (por ejemplo, binaural, de sonido "envolvente" de tipo 5.1 o periférica (con elevación) de tipo 7.1.4). El enfoque ambisónico se puede generalizar a más de cuatro canales en formato B y esta representación generalizada se conoce comúnmente como HOA (*Higher-Order Ambisonics*). Descomponer el sonido en más armónicos esféricos mejora la precisión espacial de reproducción cuando se renderiza en altavoces.

55 Una señal ambisónica de orden M comprende $K=(M+1)^2$ componentes y, en el orden 1 (si $M= 1$), se encuentran las cuatro componentes W, X, Y y Z, lo que se denomina comúnmente FOA (*First-Order Ambisonics*). También hay una variante denominada "planar" de la ambisonía (W, X, Y) que descompone el sonido definido en un plano que es, en general, el plano horizontal. En este caso, el número de componentes es $K = 2M + 1$ canales. La ambisonía de orden 1 (4 canales: W, X, Y, Z), la ambisonía de orden 1 planar (3 canales: W, X, Y) y la ambisonía de orden superior se denominan en adelante indistintamente "ambisonía" para facilitar la lectura; los tratamientos presentados pueden aplicarse independientemente del tipo planar o no y del número de componentes ambisónicas.

En lo sucesivo, se denominará "señal ambisónica" a una señal en formato B de un orden predeterminado con un determinado número de componentes ambisónicas. Esto comprende también los casos híbridos, en los que, por ejemplo, en el orden 2 solo hay 8 canales (en lugar de 9); más concretamente, en el orden 2 están los 4 canales de orden 1 (W, X, Y, Z) a los que normalmente se añaden 5 canales (típicamente denotados como R, S, T, U, V), y se puede ignorar, por ejemplo, uno de los canales de orden superior (por ejemplo, R).

Las señales a procesar por el codificador/decodificador se presentan como sucesiones de bloques de muestras sonoras denominadas en lo sucesivo "tramas" o "subtramas".

Además, las notaciones matemáticas que aparecen a continuación siguen la siguiente convención:

- Escalar: s o N (minúscula para las variables o mayúscula para las constantes)
- El operador $\text{Re}(\cdot)$ designa la parte real de un número complejo
- Vector: **u** (minúscula, en negrita)
- Matriz: **A** (mayúscula, en negrita)

Las notaciones A^T y A^H indican respectivamente la transposición y la transposición hermitiana (transpuesta y conjugada) de A .

- Una señal unidimensional en tiempo discreto, $s(i)$, definida en un intervalo de tiempo $i = 0, \dots, L-1$ de longitud L se representa mediante un vector lineal

$$\mathbf{s} = [s(0), \dots, s(L-1)]$$

También se puede escribir: $\mathbf{s} = [s_0, \dots, s_{L-1}]$ para evitar el uso de paréntesis.

- Una señal multidimensional en tiempo discreto, $\mathbf{b}(i)$, definida en un intervalo de tiempo $i = 0, \dots, L-1$ de longitud L y en K dimensiones se representa mediante una matriz de tamaño $L \times K$:

$$\mathbf{B} = \begin{bmatrix} b_0(0) & \dots & b_0(L-1) \\ \vdots & \dots & \vdots \\ b_{K-1}(0) & \dots & b_{K-1}(L-1) \end{bmatrix}$$

También cabe destacar que: $\mathbf{B} = [B_{ij}]$, $i=0, \dots, K-1$, $j=0, \dots, L-1$, para evitar el uso de paréntesis.

- Un punto 3D de coordenadas cartesianas (x, y, z) se puede convertir en coordenadas esféricas (r, Θ, ϕ) , donde r es la distancia al origen, Θ es el acimut y ϕ la elevación. En el presente documento se utiliza, sin pérdida de generalidad, la convención matemática en la que la elevación se define en relación con el plano horizontal (Oxy) ; la invención puede adaptarse fácilmente a otras definiciones, incluida la convención utilizada en física en la que el acimut se define en relación con el eje Oz .

Por otro lado, no se describirán en el presente documento las convenciones conocidas del estado de la técnica en ambisonía acerca del orden de las componentes ambisónicas (que incluyen ACN (*Ambisonic Channel Number*), SID (*Single Index Designation*), FuMA (Furse-Malham)) y la normalización de las componentes ambisónicas (SN3D, N3D, maxN). Se pueden encontrar más detalles, por ejemplo, en el recurso disponible en línea: https://en.wikipedia.org/wiki/Ambisonic_data_exchange_formats. Por convención, la primera componente de una señal ambisónica corresponde, en general, a la componente omnidireccional W.

El enfoque más simple para codificar una señal ambisónica consiste en utilizar un codificador mono y aplicarlo en paralelo a todos los canales, posiblemente con una asignación de bits diferente según los canales. Este enfoque se denomina en el presente documento "multimono". El enfoque multimono se puede extender a la codificación multiestéreo (donde pares de canales se codifican por separado mediante un codificador estéreo) o, más en general, al uso de varias instancias paralelas de un mismo códec principal.

Una realización de este tipo se muestra en la figura 1. La señal de entrada se divide en canales (un canal mono o varios canales) mediante el bloque 100. Estos canales se codifican por separado mediante los bloques 120 a 122 en función de una distribución y una asignación binaria predeterminadas. Su secuencia binaria se multiplexa (bloque 130) y después de la transmisión y/o almacenamiento, se desmultiplexa (bloque 140) para aplicar una decodificación para reconstruir los canales decodificados (bloques 150 a 152) que se recombinan (bloque 160).

La calidad asociada varía según la codificación y la decodificación principales utilizadas (bloques 120 a 122 y 150 a 152) y, en general, solo es satisfactoria a muy alta velocidad. Por ejemplo, en el caso multimonero, la codificación EVS puede considerarse casi transparente (desde un punto de vista perceptivo) a una velocidad de al menos 48 kbit/s por canal (mono); así, para una señal ambisónica de orden 1 se obtiene una velocidad mínima de $4 \times 48 = 192$ kbit/s. El enfoque de codificación multimonero no tiene en cuenta la correlación entre canales, por lo que produce deformaciones espaciales con la adición de diferentes artefactos, como la aparición de fuentes sonoras fantasmas, ruidos difusos o desplazamientos de trayectorias de fuentes sonoras. Por lo tanto, la codificación de una señal ambisónica según este enfoque degrada la espacialización.

Un enfoque alternativo a la codificación separada de todos los canales es la codificación paramétrica para una señal estéreo o multicanal. Para este tipo de codificación, la señal multicanal de entrada se reduce a un menor número de canales después de un procesamiento llamado mezcla descendente (*downmix*); estos canales se codifican y transmiten, y también se codifica información de espacialización adicional. La decodificación paramétrica consiste en aumentar el número de canales después de la decodificación de los canales transmitidos, utilizando un procesamiento denominado mezcla ascendente (*upmix*) (típicamente implementado mediante decodificación) y una síntesis espacial en función de la información de espacialización adicional decodificada. El codificador 3GPP e-AAC+ ofrece un ejemplo de codificación paramétrica estéreo. Cabe señalar que la operación de mezcla descendente también degrada la espacialización; en este caso, la imagen espacial se modifica.

2. En el estado de la técnica también se conocen los siguientes documentos: La solicitud de patente EP271 7261 A1 divulga una codificación espacial orientada a objetos utilizando valores de corrección estimados a partir de parámetros OLD de diferencia de objeto divididos por un valor de espectro de potencia reconstruido para un objeto. La solicitud WO2017/153697A1 divulga una codificación-decodificación paramétrica espacial utilizando un valor de corrección opcional para el parámetro ITDmaxest.

La invención mejora el estado de la técnica.

Para ello, se propone un método de determinación de un conjunto de correcciones a realizar en una señal sonora multicanal, en donde el conjunto de correcciones se determina a partir de una información representativa de una imagen espacial de una señal multicanal original y de una información representativa de una imagen espacial de la señal multicanal original codificada y luego decodificada.

De este modo, el conjunto de correcciones determinado, a aplicar a la señal multicanal decodificada, permite limitar las degradaciones espaciales debidas a la codificación y, opcionalmente, a operaciones de reducción/aumento de canales. La implementación de la corrección permite encontrar una imagen espacial de la señal multicanal decodificada más cercana a la imagen espacial de la señal multicanal original.

En una forma de realización particular, la determinación del conjunto de correcciones se realiza en el dominio de tiempo en banda completa (una banda de frecuencia). En variantes, se realiza en el dominio de tiempo por subbanda de frecuencia. Esto permite adaptar las correcciones según las bandas de frecuencia. En otras variantes, se realiza en un dominio de transformada real o compleja (típicamente en frecuencia) de tipo transformada rápida discreta de Fourier (STFT), transformada discreta de coseno modificada (MDCT) u otro. La invención también se refiere a un método de decodificación de una señal sonora multicanal, que comprende las siguientes etapas:

- recibir un flujo binario que comprende una señal de audio codificada procedente de una señal multicanal original y una información representativa de una imagen espacial de la señal multicanal original;
- decodificar la señal de audio codificada recibida y obtener una señal multicanal decodificada;
- decodificar la información representativa de una imagen espacial de la señal multicanal original;
- determinar una información representativa de una imagen espacial de la señal multicanal decodificada;
- determinar un conjunto de correcciones a realizar en la señal decodificada de acuerdo con el método de determinación descrito anteriormente;
- corregir la señal multicanal decodificada mediante el conjunto de correcciones determinado.

Por tanto, en esta forma de realización, el decodificador es capaz de determinar las correcciones a realizar en la señal multicanal decodificada a partir de una información representativa de la imagen espacial de la señal multicanal original, recibida desde el codificador. Esto limita la información recibida desde el codificador. Es el decodificador el que se encarga tanto de la determinación como de la aplicación de las correcciones.

La invención también se refiere a un método de codificación de una señal sonora multicanal, que comprende las siguientes etapas:

- 5 - codificar una señal de audio procedente de una señal multicanal original;
- determinar una información representativa de una imagen espacial de la señal multicanal original;
- 10 - decodificar de manera local la señal de audio codificada y obtener una señal multicanal decodificada;
- determinar una información representativa de una imagen espacial de la señal multicanal decodificada;
- 15 - determinar un conjunto de correcciones a realizar en la señal multicanal decodificada de acuerdo con el método de determinación descrito anteriormente;
- codificar el conjunto de correcciones determinado.

En este modo de realización, es el codificador el que determina el conjunto de correcciones a realizar en la señal multicanal decodificada y el que la transmite al decodificador. Por lo tanto, es el codificador el que inicia esta determinación de correcciones.

25 En una primera forma de realización particular del método de decodificación tal como el descrito anteriormente o del método de codificación tal como el descrito anteriormente, la información representativa de una imagen espacial es una matriz de covarianza y la determinación del conjunto de correcciones comprende además las etapas siguientes:

- 30 - obtener una matriz de ponderación que comprende vectores de ponderación asociados a un conjunto de altavoces virtuales;
- determinar una imagen espacial de la señal multicanal original a partir de la matriz de ponderación obtenida y a partir de la matriz de covarianza de la señal multicanal original recibida;
- 35 - determinar una imagen espacial de la señal multicanal decodificada a partir de la matriz de ponderación obtenida y a partir de la matriz de covarianza de la señal multicanal decodificada determinada;
- 40 - calcular una relación entre la imagen espacial de la señal multicanal original y la imagen espacial de la señal multicanal decodificada en las direcciones de los altavoces del conjunto de altavoces virtuales, para obtener un conjunto de ganancias.

45 De acuerdo con esta forma de realización, este método que utiliza el renderizado en altavoces solo permite transmitir una cantidad limitada de datos del codificador al decodificador. De hecho, para un orden dado M , pueden bastar $K = (M+1)^2$ coeficientes a transmitir (asociados a otros tantos altavoces virtuales), pero para una corrección más estable se puede recomendar el uso de más altavoces virtuales y, por lo tanto, transmitir más puntos. Además, la corrección se puede interpretar fácilmente en términos de ganancias asociadas a los altavoces virtuales.

50 En otra variante de realización, en caso de que el codificador determine directamente la energía de la señal de acuerdo con diferentes direcciones y transmita esta imagen espacial de la señal multicanal original al decodificador, la determinación del conjunto de correcciones del proceso de decodificación comprende además las siguientes etapas:

- 55 - obtener una matriz de ponderación que comprende vectores de ponderación asociados a un conjunto de altavoces virtuales;
- determinar una imagen espacial de la señal multicanal decodificada a partir de la matriz de ponderación obtenida y a partir de la información representativa de una imagen espacial de la señal multicanal decodificada determinada;
- 60 - calcular una relación entre la imagen espacial de la señal multicanal original y la imagen espacial de la señal multicanal decodificada en las direcciones de los altavoces del conjunto de altavoces virtuales, para obtener un conjunto de ganancias.

65 Para garantizar un valor de corrección que no sea demasiado brusco, el proceso de decodificación o el proceso de codificación comprende una etapa de limitación de los valores de ganancia obtenidos según al menos un umbral.

Este conjunto de ganancias constituye el conjunto de correcciones y puede estar, por ejemplo, en forma de una matriz de corrección que comprenda el conjunto de las ganancias determinadas de esta manera.

5 En una segunda forma de realización particular del método de decodificación o del método de codificación, la información representativa de una imagen espacial es una matriz de covarianza y la determinación del conjunto de correcciones comprende una etapa de determinación de una matriz de transformación por descomposición matricial de las dos matrices de covarianza, constituyendo la matriz de transformación el conjunto de correcciones.

10 Esta forma de realización tiene la ventaja de aportar las correcciones directamente en el dominio ambisónico en el caso de una señal multicanal ambisónica. Se evitan así las etapas de transformar las señales reproducidas en los altavoces al dominio ambisónico. Esta forma de realización permite además optimizar la corrección para que sea óptima matemáticamente incluso si requiere la transmisión de un mayor número de coeficientes en comparación con el método con renderizado en altavoces. De hecho, para un orden M y, por lo tanto, un número de componentes $K = (M+1)^2$, el número de coeficientes a transmitir es $K \times (K+1)/2$. Para evitar la amplificación excesiva de determinadas zonas de frecuencia, se determina un factor de normalización y se aplica a la matriz de transformación. En caso de que el conjunto de correcciones esté representado por una matriz de transformación o una matriz de corrección como se describió anteriormente, la corrección de la señal multicanal decodificada por el conjunto de correcciones determinado se realiza mediante la aplicación del conjunto de correcciones a la señal multicanal decodificada, es decir, directamente en el dominio ambisónico en el caso de una señal ambisónica. En el modo de realización por renderizado en altavoces implementado por el decodificador, la corrección de la señal multicanal decodificada por el conjunto de correcciones determinado se realiza de acuerdo con las siguientes etapas:

- 25 - decodificar de manera acústica la señal multicanal decodificada en el conjunto definido de altavoces virtuales;
- aplicar el conjunto de ganancias obtenido a las señales procedentes de la decodificación acústica;
- 30 - codificar de manera acústica las señales procedentes de la decodificación acústica y corregidas para obtener componentes de la señal multicanal;
- sumar las componentes de la señal multicanal obtenidas de esta manera para obtener una señal multicanal corregida.

35 En una variante de realización, las anteriores etapas de decodificación, aplicación de ganancias y codificación/suma se agrupan en una operación de corrección directa mediante una matriz de corrección. Esta matriz de corrección se puede aplicar directamente a la señal multicanal decodificada, lo que tiene la ventaja, como se describió anteriormente, de realizar las correcciones directamente en el dominio ambisónico.

40 En una segunda forma de realización, donde el método de codificación implementa el método de determinación del conjunto de correcciones, el método de decodificación comprende las siguientes etapas:

- 45 - recibir un flujo binario que comprende una señal de audio codificada procedente de una señal multicanal original y un conjunto codificado de correcciones a realizar en la señal multicanal decodificada, habiéndose codificado el conjunto de correcciones de acuerdo con un método de codificación descrito anteriormente;
- 50 - decodificar la señal de audio codificada recibida y obtener una señal multicanal decodificada;
- decodificar el conjunto codificado de correcciones;
- 55 - corregir la señal multicanal decodificada mediante la aplicación del conjunto de correcciones decodificado a la señal multicanal decodificada.

60 En esta forma de realización, es el codificador el que determina las correcciones a realizar en la señal multicanal decodificada, directamente en el dominio ambisónico y es el decodificador el que implementa la aplicación de estas correcciones en la señal multicanal decodificada, directamente en el dominio ambisónico.

En este caso, el conjunto de correcciones puede ser una matriz de transformación o una matriz de corrección que comprende un conjunto de ganancias.

65 En una variante de realización del método de decodificación por renderizado en altavoces, el método de decodificación comprende las siguientes etapas:

- 5 - recibir un flujo binario que comprende una señal de audio codificada procedente de una señal multicanal original y un conjunto codificado de correcciones a realizar en la señal multicanal decodificada, habiéndose codificado el conjunto de correcciones de acuerdo con un método de codificación tal como el descrito anteriormente;
- decodificar la señal de audio codificada recibida y obtener una señal multicanal decodificada;
- 10 - decodificar el conjunto codificado de correcciones;
- corregir la señal multicanal decodificada mediante el conjunto de correcciones decodificadas de acuerdo con las siguientes etapas:
 - 15 . decodificar de manera acústica la señal multicanal decodificada en el conjunto definido de altavoces virtuales;
 - . aplicar el conjunto de ganancias obtenido a las señales procedentes de la decodificación acústica;
 - 20 . codificar de manera acústica las señales procedentes de la decodificación acústica y corregidas para obtener componentes de la señal multicanal;
 - . sumar las componentes de la señal multicanal obtenidas de esta manera para obtener una señal multicanal corregida.

25 En esta forma de realización, es el codificador el que determina las correcciones a realizar en las señales resultantes de la decodificación acústica en un conjunto de altavoces virtuales y es el decodificador el que implementa la aplicación de estas correcciones en las señales resultantes de la decodificación acústica y el que, a continuación, transforma estas señales de vuelta al dominio ambisónico en el caso de una señal multicanal ambisónica.

30 En una variante de realización, las anteriores etapas de decodificación, aplicación de ganancias y codificación/suma se agrupan en una operación de corrección directa mediante una matriz de corrección. La corrección se realiza directamente mediante la aplicación de una matriz de corrección en la señal multicanal decodificada, por ejemplo la señal ambisónica. Como se describe anteriormente, esto tiene la ventaja de hacer las correcciones directamente en el dominio ambisónico.

35 La invención también se refiere a un dispositivo de decodificación que comprende un circuito de procesamiento para la implementación de los métodos de decodificación descritos anteriormente.

40 La invención también se refiere a un dispositivo de decodificación que comprende un circuito de procesamiento para la implementación de los métodos de codificación descritos anteriormente.

 La invención se refiere a un programa informático que comprende instrucciones que implementan los métodos de decodificación o de codificación descritos anteriormente cuando son ejecutadas por un procesador.

45 Por último, la invención se refiere a un soporte de almacenamiento, legible por un procesador, que almacena un programa informático que comprende instrucciones para la ejecución de los métodos de decodificación o de codificación descritos anteriormente.

50 Otras características y ventajas de la invención aparecerán más claramente tras la lectura de la siguiente descripción de formas de realización particulares, dadas a título de simples ejemplos ilustrativos y no limitativos, y de los dibujos adjuntos, en los que:

55 [Fig. 1] La figura 1 ilustra una codificación multimono de acuerdo con el estado de la técnica y tal como se ha descrito anteriormente;

 [Fig. 2] La figura 2 ilustra, en forma de organigrama, las etapas de un método de determinación de un conjunto de correcciones de acuerdo con una forma de realización de la invención;

60 [Fig. 3] La figura 3 ilustra una primera forma de realización de un codificador y un decodificador, de un método de codificación y de un método de decodificación de acuerdo con la invención;

 [Fig. 4] La figura 4 ilustra una primera forma de realización detallada del bloque de determinación del conjunto de correcciones;

65 [Fig. 5] La figura 5 ilustra una segunda forma de realización detallada del bloque de determinación del conjunto de correcciones;

[Fig. 6] La figura 6 ilustra una segunda forma de realización de un codificador y un decodificador, de un método de codificación y de un método de decodificación de acuerdo con la invención; y

5 [Fig. 7] La figura 7 ilustra ejemplos de realización estructural de un codificador y un decodificador de acuerdo con una forma de realización de la invención.

10 El método descrito a continuación se basa en la corrección de las degradaciones espaciales, en particular para garantizar que la imagen espacial de la señal decodificada sea lo más parecida posible a la señal original. A diferencia de los enfoques de codificación paramétrica conocidos para las señales estéreo o multicanal, donde se codifican atributos perceptivos ("*perceptual cues*" en inglés), la invención no se basa en una interpretación perceptual de la información de la imagen espacial ya que el dominio ambisónico no puede "escucharse" directamente.

15 La **figura 2** representa las etapas principales implementadas para determinar un conjunto de correcciones a aplicar a la señal multicanal codificada y, a continuación, decodificada.

20 La señal multicanal original **B** de dimensión $K \times L$ (es decir, K componentes de L muestras de tiempo o frecuencia) es la entrada del método de determinación. En la etapa S1 se extrae una información representativa de una imagen espacial de la señal multicanal original.

25 El presente documento se centra en una señal multicanal con representación ambisónica, tal y como se ha descrito anteriormente. La invención también puede aplicarse a otros tipos de señal multicanal, como una señal en formato B con modificaciones, como por ejemplo la supresión de determinadas componentes (por ejemplo, la supresión de la componente R de orden 2 para mantener 8 canales solamente) o el matrizado del formato B para pasar a un dominio equivalente (denominado dominio espacial equivalente (*Equivalent Spatial Domain*)) como se describe en la especificación 3GPP TS 26.260; otro ejemplo de matrizado viene dado por el mapeo de canales 3 (*channel mapping 3*) del codificador IETF Opus y en la especificación 3GPP TS 26.918 (cláusula 6.1.6.3).

30 En el presente documento se denomina "imagen espacial" a la distribución de la energía sonora de la escena sonora ambisónica en diferentes direcciones del espacio; en variantes, esta imagen espacial que describe la escena sonora corresponde, en general, a magnitudes positivas evaluadas en diferentes direcciones predeterminadas del espacio, por ejemplo en forma de un pseudoespectro de tipo MUSIC (*MUltiple Signal Classification*) muestreado en estas direcciones o de un histograma de direcciones de llegada (donde las direcciones de llegada se deducen de acuerdo con la discretización dada por las direcciones predeterminadas); estas magnitudes positivas pueden interpretarse como energías y se consideran como tales posteriormente para simplificar la descripción de la invención.

35 Por lo tanto, una imagen espacial asociada a una escena sonora ambisónica representa la energía sonora (o más generalmente una magnitud positiva) relativa en función de diferentes direcciones en el espacio. En la invención, una información representativa de una imagen espacial puede ser, por ejemplo, una matriz de covarianza calculada entre los canales de la señal multicanal o bien una información de energía asociada a direcciones de procedencia del sonido (asociadas a direcciones de altavoces virtuales distribuidos en una esfera unitaria).

40 El conjunto de correcciones a aplicar a una señal multicanal es una información que puede definirse por un conjunto de ganancias asociadas a direcciones de procedencia del sonido que pueden estar en forma de una matriz de correcciones que comprende este conjunto de ganancias o una matriz de transformación.

45 Por ejemplo, en la etapa S1 se obtiene una matriz de covarianza de una señal multicanal **B**. Como se describe más adelante con referencia a las figuras 3 y 6, esta matriz se calcula, por ejemplo, de la siguiente manera:

50 $\mathbf{C} = \mathbf{B} \cdot \mathbf{B}^T$ con un factor de normalización aproximado (en el caso real)

o

55 $\mathbf{C} = \text{Re}(\mathbf{B} \cdot \mathbf{B}^H)$ con un factor de normalización aproximado (en el caso complejo)

En variantes, se podrán utilizar operaciones de suavizado temporal de la matriz de covarianza. En el caso de una señal multicanal en el dominio de tiempo, la covarianza se puede estimar recursivamente (muestra por muestra) bajo la forma:

60
$$C_{ij}(n) = n/(n+1) C_{ij}(n-1) + 1/(n+1) b_i(n) b_j(n).$$

65 En una variante de realización, se obtiene información de energía de acuerdo con diferentes direcciones (asociadas a direcciones de altavoces virtuales distribuidos en una esfera unitaria). Para ello, se podrá aplicar, por ejemplo, un método de tipo SRP (potencia de respuesta dirigida, "*Steered-Response Power*" en inglés) descrito más adelante con referencia a las figuras 3 y 4. En variantes, se pueden utilizar otros métodos de cálculo de imágenes espaciales (pseudoespectro MUSIC, histograma de direcciones de llegada).

Existen varias formas de realización posibles de codificar la señal multicanal original, las cuales se describen en el presente documento.

5 En una primera forma de realización, los diferentes canales \mathbf{b}_k , $k=0, \dots, K-1$, de \mathbf{B} se codifican, en la etapa S2, mediante una codificación multimonero, donde cada canal \mathbf{b}_k se codifica por separado. En variantes de realización, también es posible una codificación multiestéreo donde los canales \mathbf{b}_k se codifican en pares separados. Un ejemplo clásico para una señal de entrada 5.1 consiste en utilizar dos codificaciones estéreo separadas de L/R y de Ls/Rs con codificaciones mono LFE (solo bajas frecuencias); para el caso ambisónico, la codificación multiestéreo se puede aplicar a las componentes ambisónicas (formato B) o a una señal multicanal equivalente obtenida tras el matizado de los canales del formato B; por ejemplo, en el orden 1, los canales W, X, Y, Z se pueden convertir en cuatro canales transformados y dos pares de canales se codifican por separado y se vuelven a convertir al formato B durante la decodificación. Se proporciona un ejemplo en las versiones recientes del código Opus (mapeo de canales 3) y en la especificación 3GPP TR 26.918 (cláusula 6.1.6.3).

15 En otras variantes, también se puede utilizar en la etapa S2 una codificación multicanal conjunta, como por ejemplo la codificación MPEGH 3D Audio para el formato ambisónico (basado en escenas); en este caso, la codificación realiza una codificación de los canales de entrada de forma conjunta. En el ejemplo MPEG-H, esta codificación conjunta se descompone para una señal ambisónica en varias etapas como la extracción y la codificación de fuentes mono predominantes, la extracción de un ambiente (típicamente reducida a una señal ambisónica de orden 1), la codificación de todos los canales extraídos (denominados canales de transporte) y de metadatos que describen los vectores de formación de haces acústicos ("*beamforming*" en inglés) para la extracción de canales predominantes. La codificación multicanal conjunta permite explotar las relaciones entre todos los canales para, por ejemplo, extraer fuentes de audio predominantes y un ambiente o realizar una asignación global de bits que tenga en cuenta todo el contenido de audio.

20 En la forma de realización preferente, se toma como ejemplo de realización de la etapa S2 una codificación multimonero que se realiza utilizando la codificación 3GPP EVS tal como se ha descrito anteriormente. Sin embargo, el método de acuerdo con la invención se puede utilizar independientemente de la codificación principal (multimonero, multiestéreo, codificación conjunta) utilizada para representar los canales a codificar.

30 La señal codificada de esta manera en forma de flujo binario ("*bitstream*" en inglés) puede codificarse en la etapa S3 ya sea mediante un decodificador local del codificador o mediante un decodificador después de la transmisión. Esta señal se decodifica para encontrar los canales de la señal multicanal $\hat{\mathbf{B}}$ (por ejemplo, por varias instancias de decodificador EVS de acuerdo con una decodificación multimonero).

35 Las etapas S2a, S2b, S3a, S3b representan una variante de realización de la codificación y decodificación de la señal multicanal \mathbf{B} . La diferencia con la codificación de la etapa S2 descrita anteriormente radica en el uso de tratamientos adicionales de reducción del número de canales ("*downmix*" en inglés) en la etapa S2a y de aumento del número de canales ("*upmix*" en inglés) en la etapa S3b. Estas etapas de codificación y decodificación (S2b y S3a) son similares a las etapas S2 y S3, excepto que el número de canales respectivos de entrada y salida es menor en las etapas S2b y S3a.

45 Un ejemplo de mezcla descendente (*downmix*) para una señal de entrada ambisónica de orden 1 consiste en mantener solo el canal W; para una señal de entrada ambisónica de orden > 1 , las 4 primeras componentes W, X, Y, Z pueden tomarse como mezcla descendente (truncando así la señal a orden 1). En variantes, se podrá tomar como mezcla descendente un subconjunto de las componentes ambisónicas (por ejemplo, 8 canales de orden 2 sin la componente R) y también considerar los casos de matizado como, por ejemplo, una mezcla descendente estéreo obtenida bajo la forma: $L = W - Y + 0.3 * X$, $R = W + Y + 0.3 * X$ (solo se utilizan canales FOA).

50 Un ejemplo de mezcla ascendente de una señal mono consiste en aplicar diferentes respuestas de impulsos espaciales de sala (SRI R, de "*Spatial Room Impulse Response*") o diferentes filtros de descorrelación (de tipo pasatodo) en el dominio de tiempo o frecuencia. Por ejemplo, en el documento 3GPP S4-180975, "*pCR to 26.118 on Dolby VRStream audio profile candidate*" (cláusula X.6.2.3.5) se ofrece un ejemplo de realización de descorrelación en un dominio de frecuencia.

55 La señal \mathbf{B}' resultante de este procesamiento de mezcla descendente se codifica en la etapa S2b mediante una codificación principal (multimonero, multiestéreo, codificación conjunta), por ejemplo mediante un enfoque mono o multimonero con la codificación 3GPP EVS. La señal de audio de entrada de la etapa de codificación S2b y de salida de la etapa de decodificación S3a tiene un número de canales inferior a la señal de audio multicanal original. En este caso, la imagen espacial representada por el codificador principal ya se ha degradado sensiblemente incluso antes de la codificación. En un caso extremo, el número de canales se reduce a un solo canal mono, codificándose solo el canal W; la señal de entrada se limita a un solo canal de audio y, por lo tanto, se pierde la imagen espacial. El método según la invención permite describir y reconstruir esta imagen espacial lo más parecido posible a la de la señal multicanal original. A la salida de la etapa de mezcla ascendente en S3b de esta variante de realización se encuentra una señal multicanal decodificada $\hat{\mathbf{B}}$.

A partir de la señal multicanal decodificada $\hat{\mathbf{B}}$ según las dos variantes (S2-S3 o S2a-S2b-S3a-S3b), se extrae, en la etapa S4, una información representativa de la imagen espacial de la señal multicanal decodificada. En cuanto a la imagen original, esta información puede ser una matriz de covarianza calculada en la señal multicanal decodificada o bien una información de energía asociada a direcciones de procedencia del sonido (o de manera equivalente, a puntos virtuales en una esfera unitaria).

Esta información representativa de la señal multicanal original y de la señal multicanal decodificada se utiliza en la etapa S5 para determinar un conjunto de correcciones a realizar en la señal multicanal decodificada con el fin de limitar las degradaciones espaciales. A continuación se detallarán dos formas de realización en referencia a las figuras 4 y 5 para ilustrar esta etapa.

El procedimiento descrito en la figura 2 se puede implementar en el dominio de tiempo, en banda completa de frecuencia (con una sola banda) o en subbandas de frecuencia (con varias bandas); esto no cambia el funcionamiento del método, donde cada subbanda se trata por separado. Si el método se realiza por subbanda, el conjunto de correcciones se determina entonces por subbanda, lo que supone un coste adicional de cálculo y datos a transmitir al decodificador en comparación con el caso de una sola banda. La división en subbandas puede ser uniforme o no uniforme. Por ejemplo, podemos dividir el espectro de una señal muestreada a 32 kHz de acuerdo con diferentes variantes:

- 4 bandas con anchuras respectivas de 1, 3, 4 y 8 kHz o 2, 2, 4, 8 kHz
- 24 bandas de Bark (con una anchura de 100 Hz en bajas frecuencias de 3,5-4 kHz para la última subbanda)
- las 24 bandas de Bark se pueden agrupar opcionalmente en bloques de 4 o 6 bandas sucesivas para formar un conjunto de, respectivamente, 6 o 4 bandas "aglomeradas".

Otras divisiones son posibles (por ejemplo, bandas ERB (ancho de banda rectangular equivalente, "*equivalent rectangular bandwidth*" en inglés) o de 1/3 de octava), incluso para el caso de una frecuencia de muestreo diferente (por ejemplo, 16 o 48 kHz).

En variantes, la invención también se puede implementar en el dominio de transformada, por ejemplo en el dominio de la transformada rápida discreta de Fourier (STFT) o en el dominio de la transformada discreta de coseno modificada (MDCT).

A continuación se describen varias formas de realización para determinar este conjunto de correcciones y aplicar este conjunto de correcciones a la señal decodificada.

Cabe recordar en el presente documento la conocida técnica de codificación de una fuente sonora en formato ambisónico. Una fuente sonora mono puede espacializarse artificialmente multiplicando su señal por los valores de los armónicos esféricos asociados a su dirección de procedencia (suponiendo que la señal se transporta por una onda plana) para obtener otras tantas componentes ambisónicas. Para ello, se calculan los coeficientes para cada armónico esférico para una posición determinada en acimut Θ y en elevación ϕ del orden deseado:

$$\mathbf{B} = \mathbf{Y}(\Theta, \phi) \cdot \mathbf{s}$$

donde \mathbf{s} es la señal mono que se espacializará e $\mathbf{Y}(\Theta, \phi)$ es el vector de codificación que define los coeficientes de los armónicos esféricos asociados a la dirección (Θ, ϕ) para el orden M . A continuación se proporciona un ejemplo de vector de codificación de orden 1 con la convención SN3D y el orden de los canales SID o FuMa:

$$\mathbf{Y}(\theta, \varphi) = \begin{bmatrix} 1 \\ \cos \theta \cos \varphi \\ \sin \theta \cos \varphi \\ \sin \varphi \end{bmatrix}$$

En variantes puede utilizarse otros convenios de normalización (por ejemplo, maxN, N3D) y orden de los canales (por ejemplo: ACN) y las diferentes formas de realización se adaptarán en función de la convención utilizada para el orden o la normalización de las componentes ambisónicas (FOA u HOA). Esto equivale a cambiar el orden de las líneas $\mathbf{Y}(\Theta, \phi)$ o multiplicar estas líneas por constantes predefinidas.

Para los órdenes superiores, los coeficientes $\mathbf{Y}(\Theta, \phi)$ de los armónicos esféricos se pueden encontrar en el libro de B. Rafaely, "*Fundamentals of Spherical Array Processing*", Springer, 2015. En términos generales, para un orden M , el número de señales ambisónicas es $K = (M+1)^2$.

5 Del mismo modo, cabe recordar en el presente documento algunas nociones sobre el renderizado o reproducción ambisónica mediante altavoces. Un sonido ambisónico no está hecho para ser escuchado tal cual; para una escucha inmersiva en altavoces o auriculares, se debe realizar una etapa de "decodificación" en el sentido acústico, también denominada renderizado ("renderer" en inglés). Se considera el caso de N altavoces (virtuales o físicos) distribuidos en una esfera, típicamente de radio unitario, y cuyas direcciones (Θ_n, ϕ_n) , $n=0, \dots, N-1$, en términos de acimut y elevación se conocen. La decodificación, como se considera en este caso, es una operación lineal que consiste en aplicar una matriz **D** a las señales ambisónicas **B** para obtener las señales s_n de los altavoces, que se pueden agrupar en una matriz $\mathbf{S} = [s_0, \dots, s_{N-1}]$, $\mathbf{S}=\mathbf{D}.\mathbf{B}$ donde

$$10 \quad \mathbf{S} = \begin{bmatrix} s_0 \\ \vdots \\ s_{N-1} \end{bmatrix} .$$

La matriz **D** se puede descomponer en vectores de líneas d_n , es decir

$$15 \quad \mathbf{D} = \begin{bmatrix} d_0 \\ \vdots \\ d_{N-1} \end{bmatrix}$$

d_n se puede ver como un vector de ponderación para el n -ésimo altavoz, utilizado para recombinar las componentes de la señal ambisónica y calcular la señal reproducida en el n -ésimo altavoz: $s_n=d_n.\mathbf{B}$.

20 Existen múltiples métodos de "decodificación" en el sentido acústico. El método conocido como "decodificación básica", también conocido como "adaptación modal", se basa en la matriz de codificación **E** asociada a todas las direcciones de los altavoces virtuales:

$$25 \quad \mathbf{E} = [\mathbf{Y}(\theta_0, \varphi_0) \quad \dots \quad \mathbf{Y}(\theta_{N-1}, \varphi_{N-1})]$$

Según este método, la matriz **D** se define típicamente como la pseudoinversa de **E**: $\mathbf{D}=\text{pinv}(\mathbf{E})= \mathbf{D}^T(\mathbf{D}.\mathbf{D}^T)^{-1}$

De forma alternativa, el método que puede denominarse de "de proyección" da resultados similares para algunas distribuciones regulares de direcciones, y se describe por la ecuación:

$$30 \quad \mathbf{D} = \frac{1}{N} \mathbf{E}^T$$

En este último caso, se observa que para cada dirección de índice n ,

$$35 \quad d_n = \frac{1}{N} \mathbf{Y}(\theta_n, \varphi_n)^T$$

En el contexto de esta invención, tales matrices servirán como matriz de formación de haces directivos ("beamforming" en inglés) que describen cómo obtener señales características de direcciones del espacio con el fin de realizar un análisis y/o transformaciones espaciales.

40 En el contexto de la presente invención, es útil describir la conversión recíproca para pasar del dominio de los altavoces al dominio ambisónico. Conviene que la aplicación sucesiva de dos conversiones reproduzca de manera exacta las señales ambisónicas originales si no se aplica ninguna modificación intermedia en el dominio de los altavoces. Por lo tanto, la conversión recíproca se define como aquella en la que interviene la pseudoinversa de **D**:

$$45 \quad \text{pinv}(\mathbf{D}).\mathbf{S}=\mathbf{D}^T(\mathbf{D}.\mathbf{D}^T)^{-1}.\mathbf{S}$$

Cuando $K = (M+1)^2$, la matriz **D** de tamaño $K \times K$ puede invertirse bajo ciertas condiciones y en este caso: $\mathbf{B}=\mathbf{D}^{-1}.\mathbf{S}$

50 En el caso del método de "adaptación modal", resulta que $\text{pinv}(\mathbf{D})=\mathbf{E}$. En variantes, se podrán utilizar otros métodos de decodificación mediante **D**, con la conversión inversa **E** correspondiente; la única condición a verificar es que la combinación de la decodificación mediante **D** y la conversión inversa mediante **E** debe dar una reconstrucción perfecta (cuando no se realiza ningún procesamiento intermedio entre la decodificación acústica y la codificación acústica).

55 Dichas variantes vienen dadas, por ejemplo, mediante:

- la decodificación "adaptación modal" con un término de regulación bajo la forma $\mathbf{D}^T(\mathbf{D} \cdot \mathbf{D}^T + \epsilon \mathbf{I})^{-1}$, donde ϵ es un valor bajo (por ejemplo 0,01),
- las decodificaciones "en fase" o "max-rE" conocidas del estado de la técnica
- o variantes donde la distribución de las direcciones de los altavoces no es regular en la esfera.

5
10 La figura 3 representa una primera forma de realización de un dispositivo de codificación y de un dispositivo de decodificación para la implementación de un método de codificación y decodificación que incluye un método de determinación de un conjunto de correcciones tal como se describe en referencia a la figura 2.

15 En esta forma de realización, el codificador calcula la información representativa de la imagen espacial de la señal multicanal original y la transmite al decodificador para permitirle corregir la degradación espacial generada por la codificación. Esto permite, durante la decodificación, atenuar los artefactos espaciales en la señal ambisónica decodificada. Por lo tanto, el codificador recibe una señal de entrada multicanal, por ejemplo de representación ambisónica FOA u HOA, o una representación híbrida con un subconjunto de componentes ambisónicas de hasta un orden ambisónico parcial dado; este último caso se incluye de manera equivalente en el caso de FOA u HOA, donde las componentes ambisónicas que faltan son nulas y el orden ambisónico viene dado por el orden mínimo requerido para incluir todas las componentes definidas. Por lo tanto, sin pérdida de generalidad, en el resto de la descripción se consideran los casos FOA u HOA.

20 En la forma de realización así descrita, la señal de entrada se muestrea a 32 kHz. El codificador funciona con tramas que tienen preferentemente una longitud de 20 ms, es decir $L = 640$ muestras por trama a 32 kHz. En variantes, otras longitudes de trama y frecuencias de muestreo son posibles (por ejemplo $L=480$ muestras por trama de 10 ms a 48 kHz).

25 En una forma de realización preferente, la codificación se realiza en el dominio de tiempo (en una o más bandas); sin embargo, en variantes, la invención puede implementarse en un dominio de transformada, por ejemplo después de la transformada rápida discreta de Fourier (STFT) o la transformación discreta de coseno modificada (MDCT).

30 De acuerdo con la forma de realización de la codificación utilizada, como se explica en referencia a la figura 2, se puede implementar un bloque 310 de reducción del número de canales (DMX); la entrada del bloque 311 es la señal \mathbf{B}' en la salida del bloque 310 cuando se implementa la mezcla descendente o la señal \mathbf{B} en caso contrario. En una forma de realización, si se aplica la mezcla descendente consiste, por ejemplo, para una señal de entrada ambisónica de orden 1, en mantener solo el canal W y, para una señal de entrada ambisónica de orden > 1 , en mantener solo las 4 primeras componentes ambisónicas W, X, Y, Z (truncando así la señal a orden 1). Otros tipos de mezcla descendente (como los descritos anteriormente con una selección de un subconjunto de canales y/o una matriz) se pueden implementar sin que esto modifique el método de acuerdo con la invención.

35 El bloque 311 codifica la señal de audio \mathbf{b}'_k de \mathbf{B}' a la salida del bloque 310 en caso de que se realice la etapa de mezcla descendente o la señal de audio \mathbf{b}_k de la señal multicanal original \mathbf{B} . Esta señal corresponde a las componentes ambisónicas de la señal multicanal original si no se ha aplicado ningún procesamiento de reducción del número de canales.

40 En una forma de realización preferente, el bloque 311 utiliza una codificación multimonó (COD) con una asignación fija o variable, donde la codificación principal es la codificación normalizada 3GPP EVS. En este enfoque multimonó, cada canal \mathbf{b}_k o \mathbf{b}'_k se codifica por separado mediante una instancia del codificador; sin embargo, en variantes son posibles otros métodos de codificación, por ejemplo una codificación multiestéreo o una codificación multicanal conjunta. Por lo tanto, a la salida de este bloque de codificación 311 se obtiene una señal de audio codificada procedente de la señal multicanal original, en forma de secuencia binaria que se envía al multiplexor 340.

45 Opcionalmente, el bloque 320 realiza una división en subbandas. En variantes, esta división en subbandas podrá reutilizar procesamientos equivalentes realizados en los bloques 310 o 311; la separación del bloque 320 es funcional en este caso. En una forma de realización privilegiada, los canales de la señal de audio multicanal original se dividen en 4 subbandas de frecuencia con una anchura respectiva de 1 kHz, 3 kHz, 4 kHz, 8 kHz (lo que equivale a una división de frecuencias de acuerdo con 0-1000, 1000-4000, 4000-8000 y 8000-16000 Hz). Esta división se puede implementar a través de una transformada rápida discreta de Fourier (STFT), un filtrado de paso banda en el dominio de Fourier (mediante la aplicación de una máscara de frecuencia) y una transformada inversa con adición de superposición. En este caso, las subbandas permanecen muestreadas a la misma frecuencia original, y el tratamiento de acuerdo con la invención se aplica en el dominio de tiempo; en variantes, se puede usar un banco de filtros con muestreo crítico. Cabe señalar que la operación de división en subbandas generalmente implica un retardo en el procesamiento, que depende del tipo de banco de filtros implementado; de acuerdo con la invención, se puede aplicar una alineación de tiempo antes o después de la codificación-decodificación y/o antes de la extracción de información de imagen espacial, de modo que la información de imagen espacial esté bien sincronizada en el tiempo con la señal corregida.

En variantes, se puede realizar un procesamiento de banda completa o la división en subbandas puede ser diferente, como se explicó anteriormente.

5 En otras variantes, la señal resultante de una transformada de la señal de audio multicanal original se utiliza directamente y la invención se aplica en el dominio de transformada con una división en subbandas en el dominio de tiempo.

10 En el resto de la descripción, se describen las diferentes etapas de codificación y decodificación como si fuera un procesamiento en el dominio de tiempo o frecuencia (real o compleja) con una sola banda de frecuencia para simplificar la descripción.

15 También se podrá implementar, de manera opcional, en cada subbanda, un filtrado de paso alto (de frecuencia de corte típicamente a 20 o 50 Hz), por ejemplo en forma de un filtro IIR elíptico de orden 2 cuya frecuencia de corte está fijada de manera preferente en 20 o 50 Hz (50 Hz en variantes). Este preprocesamiento evita un sesgo potencial para la estimación posterior de la covarianza durante la codificación; sin este preprocesamiento, la corrección implementada en el bloque 390 descrito posteriormente tenderá a amplificar las bajas frecuencias durante un procesamiento de banda completa.

20 El bloque 321 determina (Inf. B) una información representativa de una imagen espacial de la señal multicanal original.

En una forma de realización, esta información es una información de energía asociada a direcciones de origen del sonido (asociadas a direcciones de altavoces virtuales distribuidos en una esfera unitaria).

25 Para ello, se define una esfera 3D virtual de radio unitario, donde esta esfera 3D es discretizada por N puntos (altavoces virtuales "puntuales") cuya posición está definida en coordenadas esféricas por las direcciones (Θ_n, ϕ_n) para el enésimo altavoz. Los altavoces se colocan típicamente de manera (casi)uniforme en la esfera. El número N de altavoces virtuales se determina como una discretización que tiene al menos $N=K$ puntos, donde M es el orden ambisónico de la señal y $K=(M+1)^2$, es decir, $N \geq K$. Por ejemplo, se puede utilizar un método de cuadratura de tipo "Lebedev" para realizar esta discretización, de acuerdo con las referencias de V.I. Lebedev y D.N. Laikov, "A quadrature formula for the sphere of the 131st algebraic order of accuracy", Doklady Mathematics, vol. 59, n.º 3, 1999, páginas 477-481 o de Pierre Lecomte, Philippe-Aubert Gauthier, Christophe Langrenne, Alexandre Garcia y Alain Berry, "On the use of a Lebedev grid for Ambisonics", AES Convention 139, Nueva York, 2015.

35 En variantes se podrán utilizar otras discretizaciones, como por ejemplo una discretización de Fliege con al menos $N=K$ puntos ($N \geq K$), como se describe en la referencia de J. Fliege y U. Maier, "A two-stage approach for computing cubature formulae for the sphere", Technical Report, Dortmund University, 1999, o bien una discretización tomando los puntos de un "diseño esférico de tipo t" como se describe en el artículo de R. H. Hardin y N. J. A. Sloane, "McLaren's Improved Snub Cube and Other New Spherical Designs in Three Dimensions", Discret and Computational Geometry, 15 (1996), páginas 429-441.

45 A partir de esta discretización, se puede determinar la imagen espacial de la señal multicanal. Un método posible es, por ejemplo, el método SRP (potencia de respuesta dirigida, "Steered-Response Power" en inglés). De hecho, este método consiste en calcular la energía a corto plazo procedente de diferentes direcciones definidas en términos de acimut y elevación. Para ello, como se explicó anteriormente, de manera similar al renderizado en N altavoces, se calcula una matriz de ponderación de componentes ambisónicas; a continuación, esta matriz se aplica a la señal multicanal para sumar la contribución de las componentes y realizar un conjunto de N haces acústicos (o "beamformers" en inglés).

50 La señal resultante del haz acústico para la dirección (Θ_n, ϕ_n) del enésimo altavoz viene dada por: $\mathbf{s}_n = \mathbf{d}_n \cdot \mathbf{B}$

donde \mathbf{d}_n es el vector (línea) de ponderación que proporciona los coeficientes de formación del haz acústico para la dirección dada y \mathbf{B} es una matriz de tamaño $K \times L$ que representa la señal ambisónica (formato B) con K componentes, en un intervalo de tiempo de longitud L.

55 El conjunto de las señales procedentes de los N haces acústicos da lugar a la ecuación:

$$\mathbf{S} = \mathbf{D} \cdot \mathbf{B}$$

60 donde

$$\mathbf{D} = \begin{bmatrix} \mathbf{d}_0 \\ \vdots \\ \mathbf{d}_{N-1} \end{bmatrix}$$

y **S** es una matriz de tamaño NxL que representa las señales de N altavoces virtuales en un intervalo de tiempo de longitud L.

La energía a corto plazo en el segmento de tiempo de longitud L para cada dirección (Θ_n, ϕ_n) es:

$$\sigma_n^2 = \mathbf{s}_n \cdot \mathbf{s}_n^T = (\mathbf{d}_n \cdot \mathbf{B}) \cdot (\mathbf{d}_n \cdot \mathbf{B})^T = \mathbf{d}_n \cdot \mathbf{B} \cdot \mathbf{B}^T \cdot \mathbf{d}_n^T = \mathbf{d}_n \cdot \mathbf{C} \cdot \mathbf{d}_n^T$$

donde $\mathbf{C} = \mathbf{B} \cdot \mathbf{B}^T$ (caso real) o $\text{Re}(\mathbf{B} \cdot \mathbf{B}^H)$ (caso complejo) es la matriz de covarianza de **B**. Cada término $\sigma_n^2 = \mathbf{s}_n \cdot \mathbf{s}_n^T$ se puede calcular de este modo para el conjunto de direcciones (Θ_n, ϕ_n) que corresponden a una discretización de la esfera 3D por parte de los altavoces virtuales.

La imagen espacial Σ viene dada entonces por:

$$\Sigma = [\sigma_0^2, \dots, \sigma_{N-1}^2]$$

Se pueden utilizar otras variantes de cálculo de una imagen espacial Σ distinta del método SRP.

- Los valores \mathbf{d}_n pueden variar según el tipo de formación de haz acústico utilizado (retardo-suma, MVDR, LCMV...). La invención también se aplica a estas variantes de cálculo de la matriz **D** y la imagen espacial

$$\Sigma = [\sigma_0^2, \dots, \sigma_{N-1}^2]$$

- El método MUSIC (*MUltiple Signal Classification*) también proporciona otra forma de calcular una imagen espacial, con un enfoque de subespacios.

La invención también se aplica en esta variante de cálculo de la imagen espacial

$$\Sigma = [\sigma_0^2, \dots, \sigma_{N-1}^2]$$

que corresponde al pseudoespectro MUSIC calculado diagonalizando la matriz de covarianza y evaluado para las direcciones (Θ_n, ϕ_n).

- La imagen espacial se puede calcular a partir de un histograma del vector intensidad (de orden 1) como, por ejemplo, en el artículo de S. Tervo, "Direction estimation based on sound intensity vectors", Proc. EUSI POO, 2009, o su generalización en vector de pseudointensidad. En este caso, el histograma (cuyos valores son el número de ocurrencias de los valores de las direcciones de llegada según las direcciones predeterminadas (Θ_n, ϕ_n)) se interpreta como un conjunto de energías según las direcciones predeterminadas.

El bloque 330 realiza entonces una cuantificación de la imagen espacial determinada de esta manera, por ejemplo con una cuantificación escalar de 16 bits por coeficiente (utilizando directamente la representación en coma flotante trunca en 16 bits). En variantes, son posibles otros métodos de cuantificación escalar o vectorial. En otra forma de realización, la información representativa de la imagen espacial de la señal multicanal original es una matriz de covarianza (de las subbandas) de los canales de entrada **B**. Esta matriz se calcula como:

$\mathbf{C} = \mathbf{B} \cdot \mathbf{B}^T$ con un factor de normalización aproximado (en el caso real).

Si la invención se implementa en un dominio de transformada de valores complejos, esta covarianza se calcula como:

$$\mathbf{C} = \text{Re}(\mathbf{B} \cdot \mathbf{B}^H)$$

con un factor de normalización aproximado.

En variantes, se podrán utilizar operaciones de suavizado temporal de la matriz de covarianza. En el caso de una señal multicanal en el dominio de tiempo, la covarianza se puede estimar recursivamente (muestra por muestra).

Dado que la matriz de covarianza **C** (de tamaño KxK) es, por definición, simétrica, solo uno de los triángulos inferior o superior se transmite al bloque de cuantificación 330 que codifica (Q) $K(K+1)/2$ coeficientes, siendo K el número de componentes ambisónicas. Este bloque 330 realiza una cuantificación de estos coeficientes, por ejemplo con una cuantificación escalar de 16 bits por coeficiente (utilizando directamente la representación de coma flotante trunca en 16 bits). En variantes, se pueden implementar otros métodos de cuantificación escalar o vectorial de la matriz de

covarianza. Por ejemplo, se puede calcular el valor máximo (varianza máxima) de la matriz de covarianza y, a continuación, codificar mediante cuantificación escalar con un paso logarítmico, con un número de bits menor (por ejemplo 8 bits), los valores del triángulo superior (o inferior) de la matriz de covarianza normalizada por su valor máximo.

5 En variantes, la matriz de covarianza \mathbf{C} podrá regularizarse antes de la cuantificación en forma de $\mathbf{C} + \epsilon \mathbf{I}$.

Los valores cuantificados se envían al multiplexor 340.

10 En esta forma de realización, el decodificador recibe en el bloque demultiplexor 350 un flujo binario que comprende una señal de audio codificada procedente de la señal multicanal original e información representativa de una imagen espacial de la señal multicanal original.

15 El bloque 360 decodifica (\mathbf{Q}^{-1}) la matriz de covarianza u otra información representativa de la imagen espacial de la señal original. El bloque 370 decodifica (DEC) la señal de audio representada por el flujo binario.

En una forma de realización de la codificación y la decodificación, sin implementar las etapas de mezcla descendente y mezcla ascendente, la señal multicanal decodificada $\hat{\mathbf{B}}$ se obtiene a la salida del bloque de decodificación 370.

20 En la forma de realización donde la etapa de mezcla descendente se ha utilizado para la codificación, la decodificación implementada en el bloque 370 permite obtener una señal de audio decodificada

$\hat{\mathbf{B}}'$

25 que se envía como entrada desde el bloque 371 de mezcla ascendente.

Por lo tanto, el bloque 371 implementa una etapa opcional (mezcla ascendente) para aumentar el número de canales. En una forma de realización de esta etapa, para el canal de una señal mono

$\hat{\mathbf{B}}'$,

30 consiste en convolucionar la señal

$\hat{\mathbf{B}}'$

35 mediante diferentes respuestas de impulsos espaciales de sala (SRIR, "*Spatial Room Impulse Response*"); estas SRIR se definen en el orden ambisónico original de \mathbf{B} . Otros métodos de descorrelación son posibles, por ejemplo la aplicación de filtros de descorrelación de paso total a los diferentes canales de la señal

$\hat{\mathbf{B}}'$.

40 El bloque 372 implementa una etapa opcional (SB) de división en subbandas para obtener subbandas en el dominio de tiempo o en un dominio de transformada. Una etapa inversa, en el bloque 391, agrupa las subbandas para encontrar una señal multicanal de salida.

45 El bloque 375 determina ($\text{Inf } \hat{\mathbf{B}}$) una información representativa de una imagen espacial de la señal multicanal decodificada de manera similar a lo que se ha descrito para el bloque 321 (para la señal multicanal original), aplicado esta vez a la señal multicanal decodificada $\hat{\mathbf{B}}$ obtenida a la salida del bloque 371 o del bloque 370 de acuerdo con las formas de realización de decodificación.

50 De la misma manera que se describió para el bloque 321, en una forma de realización, esta información es una información de energía asociada a direcciones de origen del sonido (asociadas a direcciones de altavoces virtuales distribuidos en una esfera unitaria). Como se explicó anteriormente, se puede utilizar un método de tipo SRP (u otro) para determinar la imagen espacial de la señal multicanal decodificada. En otra forma de realización, esta información es una matriz de covarianza de los canales de la señal multicanal decodificada.

55 Esta matriz de covarianza se obtiene de la siguiente manera:

$$\hat{\mathbf{C}} = \hat{\mathbf{B}} \cdot \hat{\mathbf{B}}^T$$

60 (caso real) o

$\hat{\mathbf{C}} = \text{Re}(\hat{\mathbf{B}} \cdot \hat{\mathbf{B}}^H)$ (caso complejo) con un factor de normalización aproximado.

En variantes, se podrán utilizar operaciones de suavizado temporal de la matriz de covarianza. En el caso de una señal multicanal en el dominio de tiempo, la covarianza se puede estimar recursivamente (muestra por muestra).

A partir de la información representativa de las imágenes espaciales respectivamente de la señal multicanal original (Inf. \mathbf{B}) y de la señal multicanal decodificada (Inf. $\hat{\mathbf{B}}$), por ejemplo, las matrices de covarianza \mathbf{C} y $\hat{\mathbf{C}}$, el bloque 380 implementa el método de determinación (Det.Corr) de un conjunto de correcciones como se describe en referencia a la figura 2.

En referencia a las figuras 4 y 5 se describen dos formas de realización particulares de esta determinación.

En la forma de realización de la figura 4 se utiliza un método que usa el renderizado (explícito o no) en altavoces virtuales, y en la forma de realización de la figura 5 se utiliza un método implementado basado en una factorización de tipo Cholesky.

El bloque 390 de la figura 3 implementa una corrección (CORR) de la señal multicanal decodificada por el conjunto de correcciones determinado por el bloque 380 para obtener una señal multicanal decodificada corregida.

Por lo tanto, la **figura 4** representa una forma de realización de la etapa de determinación de un conjunto de correcciones. Esta forma de realización se realiza mediante el uso de renderizado en altavoces virtuales.

En esta forma de realización, se considera en un primer momento que la información representativa de la imagen espacial de la señal multicanal original y de la señal multicanal decodificada son las matrices de covarianza respectivas \mathbf{C} y $\hat{\mathbf{C}}$.

En este caso, los bloques 420 y 421 determinan, respectivamente, las imágenes espaciales de la señal multicanal original y de la señal multicanal decodificada.

Para ello, como se describió anteriormente, se discretiza una esfera 3D virtual de radio unitario por N puntos (altavoces virtuales "puntuales") cuya dirección está definida en coordenadas esféricas por las direcciones (Θ_n, ϕ_n) para el n -ésimo altavoz.

En lo que antecede se han definido varios métodos de discretización.

A partir de esta discretización, se puede determinar la imagen espacial de la señal multicanal. Como se describió anteriormente, un método posible es el método SRP (u otro), que consiste en calcular la energía a corto plazo procedente de diferentes direcciones definidas en términos de acimut y elevación.

Este método u otros tipos de métodos, tales como los enumerados anteriormente, se pueden utilizar para determinar las imágenes espaciales Σ y $\hat{\Sigma}$ (IS \mathbf{B} y IS $\hat{\mathbf{B}}$), respectivamente, de la señal multicanal original, en 420 (IMG \mathbf{B}) y de la señal multicanal decodificada en 421 (IMG $\hat{\mathbf{B}}$).

En caso de que la información representativa de la imagen espacial de la señal de origen (Inf \mathbf{B}) recibida y decodificada en 360 por el decodificador sea la imagen espacial en sí, es decir, una información de energía (o una magnitud positiva) asociada a direcciones de procedencia del sonido (asociadas a direcciones de altavoces virtuales distribuidos en una esfera unitaria), ya no es necesario calcularla en 420. Esta imagen espacial es utilizada directamente por el bloque 430 descrito más adelante.

Del mismo modo, si la determinación en 375 de la información representativa de la imagen espacial de la señal multicanal decodificada (Inf $\hat{\mathbf{B}}$) es la imagen espacial en sí de la señal multicanal decodificada, entonces ya no es necesario calcularla en 421. Esta imagen espacial es utilizada directamente por el bloque 430 descrito más adelante.

A partir de las imágenes espaciales Σ y $\hat{\Sigma}$, el bloque 430 calcula (relación) para cada punto dado por (Θ_n, ϕ_n) , la relación de energía entre la energía $\sigma_n^2 = \Sigma_n$ de la señal original y la energía

$$\hat{\sigma}_n^2 = \hat{\Sigma}_n$$

de la señal decodificada. Se obtiene así un conjunto de ganancias g_n de acuerdo con la siguiente ecuación:

$$g_n = \frac{\sigma_n^2}{\sqrt{\sigma_n^2 + \epsilon}}$$

La relación de energía, dependiendo de la dirección (Θ_n, ϕ_n) y la banda de frecuencia, puede ser muy importante. El bloque 440 permite, de manera opcional, limitar (Limit g_n) el valor máximo que puede tomar una ganancia g_n . Cabe recordar que las magnitudes positivas dadas σ_n^2 y

5 $\hat{\sigma}_n^2$

pueden corresponder, más generalmente, a cantidades de un pseudoespectro MUSIC o a valores de un histograma de direcciones de llegada según las direcciones discretizadas (Θ_n, ϕ_n) .

10 En una posible realización, se aplica un umbral al valor de g_n . Cualquier valor por encima de este umbral está obligado a ser igual a este valor umbral. El umbral puede fijarse, por ejemplo, en 6 dB, de modo que un valor de ganancia fuera del intervalo de ± 6 dB se satura a ± 6 dB.

15 Por lo tanto, este conjunto de ganancias g_n constituye el conjunto de correcciones a realizar en la señal multicanal decodificada.

Este conjunto de ganancias se recibe como entrada del bloque 390 de corrección de la figura 3. Se puede definir una matriz de corrección directamente aplicable a la señal multicanal decodificada, por ejemplo en la forma $\mathbf{G} = \mathbf{E} \cdot \text{diag}([g_0 \dots g_{N-1}]) \cdot \mathbf{D}$ donde \mathbf{D} y \mathbf{E} son las matrices de decodificación y codificación acústicas definidas anteriormente. Esta matriz \mathbf{G} se aplica a la señal multicanal decodificada $\hat{\mathbf{B}}$ para obtener la señal ambisónica de salida corregida ($\hat{\mathbf{B}}_{\text{corr}}$).

20 A continuación se describe un desglose de las etapas implementadas para la corrección. El bloque 390 aplica para cada altavoz virtual la ganancia g_n correspondiente, determinada anteriormente. La aplicación de esta ganancia permite obtener, en este altavoz, la misma energía que la señal original.

25 De este modo, en cada altavoz se corrige el renderizado de las señales decodificadas.

Por tanto, se implementa una etapa de codificación acústica, por ejemplo una codificación ambisónica mediante la matriz \mathbf{E} , para obtener componentes de la señal multicanal, por ejemplo componentes ambisónicas. Finalmente, estas componentes ambisónicas se suman para obtener la señal multicanal de salida corregida ($\hat{\mathbf{B}}_{\text{Corr}}$). Por lo tanto, se pueden calcular explícitamente los canales asociados a los altavoces virtuales, aplicarles una ganancia y , a continuación, recombinar los canales procesados o, de manera equivalente, aplicar la matriz \mathbf{G} a la señal a corregir.

35 En variantes, a partir de la matriz de covarianza $\hat{\mathbf{C}}$ de la señal multicanal codificada, y posteriormente decodificada, y a partir de la matriz de corrección \mathbf{G} se puede calcular en el bloque 390 la matriz de covarianza de la señal corregida como:

$$\mathbf{R} = \mathbf{G} \cdot \hat{\mathbf{C}} \cdot \mathbf{G}^T$$

40 Solo se conserva el valor del primer coeficiente R_{00} de la matriz \mathbf{R} , correspondiente a la componente omnidireccional (canal W), para aplicarse como factor de normalización a \mathbf{R} y evitar un aumento de la ganancia global debido a la matriz de corrección \mathbf{G} :

45
$$\hat{\mathbf{B}}_{\text{corr}} = \mathbf{G}_{\text{norm}} \cdot \hat{\mathbf{B}}$$

$$\mathbf{G}_{\text{norm}} = g_{\text{norm}} \cdot \mathbf{G}$$

con

50
$$g_{\text{norm}} = \sqrt{\hat{C}_{00} / R_{00}}$$

donde \hat{C}_{00} corresponde al primer coeficiente de la matriz de covarianza de la señal multicanal decodificada.

55 En variantes, el factor de normalización g_{norm} se puede determinar sin calcular toda la matriz \mathbf{R} , ya que basta con calcular un subconjunto de elementos matriciales para determinar R_{00} (y, por lo tanto, g_{norm}).

La matriz \mathbf{G} o \mathbf{G}_{norm} así obtenida corresponde al conjunto de correcciones a realizar en la señal multicanal decodificada.

60 A continuación, la **figura 5** representa otra forma de realización del método de determinación del conjunto de correcciones implementado en el bloque 380 de la figura 3.

En esta forma de realización, se considera que la información representativa de la imagen espacial de la señal multicanal original y de la señal multicanal decodificada son las matrices de covarianza respectivas \mathbf{C} y $\hat{\mathbf{C}}$.

5 En esta forma de realización, no se busca hacer un renderizado en altavoces virtuales para corregir la imagen espacial de una señal multicanal. De manera particular, para una señal ambisónica, se busca calcular la corrección de la imagen espacial directamente en el dominio ambisónico.

10 Para ello, se determina una matriz de transformación \mathbf{T} a aplicar a la señal decodificada, de modo que la imagen espacial modificada después de aplicar la matriz de transformación \mathbf{T} a la señal decodificada $\hat{\mathbf{B}}$ sea la misma que la de la señal original \mathbf{B} . Por lo tanto, se busca una matriz \mathbf{T} que verifique la siguiente ecuación: $\mathbf{T} \cdot \hat{\mathbf{C}} \cdot \mathbf{T}^T = \mathbf{C}$ donde $\mathbf{C} = \mathbf{B} \cdot \mathbf{B}^T$ es la matriz de covarianza de \mathbf{B} y $\hat{\mathbf{C}} = \hat{\mathbf{B}} \cdot \hat{\mathbf{B}}^T$ es la matriz de covarianza de $\hat{\mathbf{B}}$, en la trama actual.

15 En esta forma de realización, se utiliza una factorización denominada factorización de Cholesky para resolver esta ecuación.

Dada una matriz \mathbf{A} de tamaño $n \times n$, la factorización de Cholesky consiste en determinar una matriz \mathbf{L} triangular (inferior o superior) de modo que $\mathbf{A} = \mathbf{L}\mathbf{L}^T$ (caso real) y $\mathbf{A} = \mathbf{L}\mathbf{L}^H$ (caso complejo). Para que la descomposición sea posible, la matriz \mathbf{A} debe ser una matriz simétrica definida como positiva (caso real) o hermitiana definida como positiva (caso complejo); en el caso real, los coeficientes diagonales de \mathbf{L} son estrictamente positivos.

20 En el caso real, se dice que una matriz \mathbf{M} de tamaño $n \times n$ es simétrica definida como positiva si es simétrica ($\mathbf{M}^T = \mathbf{M}$) y definida como positiva ($\mathbf{x}^T \mathbf{M} \mathbf{x} > 0$ para todo $\mathbf{x} \in \mathbb{R}^n \setminus \{0\}$).

25 Para una matriz simétrica \mathbf{M} , es posible verificar que la matriz está definida como positiva si todos sus valores propios son estrictamente positivos ($\lambda_i > 0$). Si los valores propios son positivos ($\lambda_i \geq 0$), se dice que la matriz está semidefinida como positiva.

30 Se dice que una matriz \mathbf{M} de tamaño $n \times n$ es hermitiana simétrica definida como positiva si es hermitiana ($\mathbf{M}^H = \mathbf{M}$) y definida como positiva ($\mathbf{z}^H \mathbf{M} \mathbf{z}$ es un real > 0 para todo $\mathbf{z} \in \mathbb{C}^n \setminus \{0\}$).

Por ejemplo, la factorización de Cholesky se utiliza para encontrar una solución a un sistema de ecuaciones lineales del tipo $\mathbf{A}\mathbf{x} = \mathbf{b}$. Por ejemplo, en el caso complejo, es posible transformar \mathbf{A} en $\mathbf{L}\mathbf{L}^H$ mediante la factorización de Cholesky, resolver $\mathbf{L}\mathbf{y} = \mathbf{b}$ y, a continuación, resolver $\mathbf{L}^H \mathbf{x} = \mathbf{y}$.

35 De manera equivalente, la factorización de Cholesky se puede escribir como $\mathbf{A} = \mathbf{U}^T \mathbf{U}$ (caso real) y $\mathbf{A} = \mathbf{U}^H \mathbf{U}$ (caso complejo), donde \mathbf{U} es una matriz triangular superior.

40 En la forma de realización descrita aquí, sin pérdida de generalidad, solo se trata el caso de una factorización de Cholesky por una matriz triangular \mathbf{L} .

Por tanto, la factorización de Cholesky permite descomponer una matriz $\mathbf{C} = \mathbf{L}\mathbf{L}^T$ en dos matrices triangulares a condición de que la matriz \mathbf{C} sea simétrica definida como positiva. Esto da como resultado la siguiente ecuación:

45
$$\mathbf{T} \cdot \hat{\mathbf{L}} \cdot \hat{\mathbf{L}}^T \cdot \mathbf{T}^T = \mathbf{L} \cdot \mathbf{L}^T$$

Por identificación, se obtiene:

$$\mathbf{T} \cdot \hat{\mathbf{L}} = \mathbf{L}$$

50 Es decir:

$$\mathbf{T} = \mathbf{L} \cdot \hat{\mathbf{L}}^{-1}$$

55 Dado que las matrices de covarianza \mathbf{C} y $\hat{\mathbf{C}}$ son, en general, matrices semidefinidas como positivas, la factorización de Cholesky no puede utilizarse tal cual.

Se observa aquí que cuando las matrices \mathbf{L} y $\hat{\mathbf{L}}$ son triangulares inferiores (respectivamente, superiores), la matriz de transformación \mathbf{T} es también triangular inferior (respectivamente, superior).

60 Por lo tanto, el bloque 510 obliga a la matriz de covarianza \mathbf{C} a definirse como positiva. Para ello, se añade un valor ϵ (Fact. \mathbf{C} para la factorización de \mathbf{C}) a los coeficientes de la diagonal de la matriz para garantizar que la matriz esté bien definida como positiva: $\mathbf{C} = \mathbf{C} + \epsilon \mathbf{I}$, donde ϵ es un valor bajo fijado, por ejemplo, en 10^{-9} e \mathbf{I} es la matriz identidad. De manera similar, el bloque 520 fuerza a la matriz de covarianza $\hat{\mathbf{C}}$ a ser definida como positiva, modificando esta matriz bajo la forma $\hat{\mathbf{C}} = \hat{\mathbf{C}} + \epsilon \mathbf{I}$, donde ϵ es un valor bajo fijado, por ejemplo, en 10^{-9} e \mathbf{I} es la matriz identidad.

Una vez que las dos matrices de covarianza \mathbf{C} y $\hat{\mathbf{C}}$ están acondicionadas para definirse como positivas, el bloque 530 calcula las factorizaciones de Cholesky asociadas y obtiene (Det.T) la matriz de transformación óptima \mathbf{T} bajo la forma

5
$$\mathbf{T} = \mathbf{L}\mathbf{L}^{-1}$$

En variantes, se puede hacer una resolución alternativa con una descomposición en valores propios.

10 La descomposición en valores propios ("*eigen decomposition*" en inglés) consiste en factorizar una matriz real o compleja \mathbf{A} de tamaño $n \times n$ bajo la forma:

$$\mathbf{A} = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^{-1}$$

15 donde \mathbf{A} es una matriz diagonal que contiene los valores propios λ_i y \mathbf{Q} es la matriz de vectores propios.

Si la matriz es real:

$$\mathbf{A} = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^T$$

20 En el caso complejo, la descomposición se escribe: $\mathbf{A} = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^H$

En este caso, se busca una matriz \mathbf{T} tal que: $\mathbf{T} \cdot \hat{\mathbf{C}} \cdot \mathbf{T}^T = \mathbf{C}$ donde

25
$$\mathbf{C} = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^t$$

y

$$\hat{\mathbf{C}} = \hat{\mathbf{Q}} \mathbf{\Lambda} \hat{\mathbf{Q}}^t$$

30 es decir:

$$\mathbf{T} \cdot \hat{\mathbf{Q}} \cdot \mathbf{\Lambda} \cdot \hat{\mathbf{Q}}^t \cdot \mathbf{T}^t = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^t$$

35 Por identificación, se obtiene:

$$\mathbf{T} \cdot \hat{\mathbf{Q}} \cdot \sqrt{\mathbf{\Lambda}} = \mathbf{Q} \sqrt{\mathbf{\Lambda}}$$

Es decir:

40
$$\mathbf{T} = \mathbf{Q} \cdot \sqrt{\mathbf{\Lambda}} \cdot \sqrt{\mathbf{\Lambda}}^{-1} \cdot \hat{\mathbf{Q}}^{-1}$$

La estabilidad de la solución de una trama a otra es típicamente peor que con un enfoque de factorización de Cholesky. Esta inestabilidad se ve agravada por aproximaciones computacionales potencialmente mayores durante la descomposición en valores propios.

45 En variantes, el cálculo de la matriz diagonal

$$\sqrt{\mathbf{\Lambda}} \cdot \sqrt{\mathbf{\Lambda}}^{-1}$$

50 donde

$$\mathbf{\Lambda} = (\lambda_0, \dots, \lambda_{K-1}) \text{ et } \hat{\mathbf{\Lambda}} = (\hat{\lambda}_0, \dots, \hat{\lambda}_{K-1})$$

55 se puede realizar elemento por elemento bajo la forma

$$\text{sgn}(\lambda_i, \hat{\lambda}_i) \sqrt{|\lambda_i| / (|\hat{\lambda}_i| + \varepsilon)}$$

donde $\text{sgn}(\cdot)$ es una función de signo (+ 1 si es positivo, -1 en caso contrario) y ε es un término de regularización (por ejemplo $\varepsilon = 10^{-9}$) para evitar las divisiones por cero.

5 En esta forma de realización, es posible que la diferencia relativa de energía entre la señal ambisónica decodificada y la señal ambisónica corregida sea muy importante, especialmente a nivel de altas frecuencias que pueden verse muy deterioradas por los codificadores, tal como la codificación EVS multimonos. Para evitar la amplificación excesiva de determinadas zonas de frecuencia, se puede añadir un término de regularización. El bloque 640 se encarga, de forma opcional, de normalizar (Norm. T) esta corrección.

Por lo tanto, en la forma de realización preferente, se calcula un factor de normalización para no amplificar las zonas de frecuencia.

15 A partir de la matriz de covarianza $\hat{\mathbf{C}}$ de la señal multicanal codificada, y posteriormente decodificada, y a partir de la matriz de transformación \mathbf{T} se puede calcular la matriz de covarianza de la señal corregida como:

$$\mathbf{R} = \mathbf{T} \cdot \hat{\mathbf{C}} \cdot \mathbf{T}^T$$

20 Solo se conserva el valor del primer coeficiente R_{00} de la matriz \mathbf{R} , correspondiente a la componente omnidireccional (canal W), para aplicarse como factor de normalización a \mathbf{T} y evitar un aumento de la ganancia global debido a la matriz de corrección \mathbf{T} :

$$\hat{\mathbf{B}}_{corr} = \mathbf{T}_{norm} \cdot \hat{\mathbf{B}}$$

$$\mathbf{T}_{norm} = g_{norm} \cdot \mathbf{T}$$

con

$$30 \quad g_{norm} = \sqrt{\hat{\mathbf{C}}_{00} / R_{00}}$$

donde $\hat{\mathbf{C}}_{00}$ corresponde al primer coeficiente de la matriz de covarianza de la señal multicanal decodificada.

35 En variantes, el factor de normalización g_{norm} se puede determinar sin calcular toda la matriz \mathbf{R} , ya que basta con calcular un subconjunto de elementos matriciales para determinar R_{00} (y, por lo tanto, g_{norm}).

La matriz \mathbf{T} o \mathbf{T}_{norm} así obtenida corresponde al conjunto de correcciones a realizar en la señal multicanal decodificada.

40 Con esta forma de realización, el bloque 390 de la figura 3 realiza la etapa de corrección de la señal multicanal decodificada mediante la aplicación de la matriz de transformación \mathbf{T} o \mathbf{T}_{norm} directamente a la señal multicanal decodificada, en el dominio ambisónico, para obtener la señal ambisónica de salida corregida ($\hat{\mathbf{B}}_{corr}$).

45 Se describe ahora una segunda forma de realización de un codificador/decodificador de acuerdo con la invención en donde el método de determinación del conjunto de correcciones se implementa en el codificador. La **figura 6** describe esta forma de realización. Por tanto, esta figura representa una segunda forma de realización de un dispositivo de codificación y de un dispositivo de decodificación para la implementación de un método de codificación y decodificación que incluye un método de determinación de un conjunto de correcciones tal como se describe en referencia a la figura 2.

50 En esta forma de realización, el método de determinación del conjunto de correcciones (por ejemplo, ganancias asociadas a direcciones) se realiza en el codificador que, a continuación, transmite al decodificador este conjunto de correcciones. El decodificador decodifica este conjunto de correcciones para aplicarlo a la señal multicanal decodificada. Esta forma de realización implica, por lo tanto, implementar una decodificación local al codificador; esta decodificación local está representada por los bloques 612 a 613.

55 Los bloques 610, 611, 620 y 621 son idénticos, respectivamente, a los bloques 310, 311, 320 y 321 descritos en referencia a la figura 3.

60 De este modo, a la salida del bloque 621 se obtiene una información representativa de la imagen espacial de la señal multicanal original (Inf. \mathbf{B}).

El bloque 612 implementa una decodificación local (DEC_loc) en relación con la codificación realizada por el bloque 611.

5 Esta decodificación local puede consistir en una decodificación completa a partir de la secuencia binaria del bloque 611 o, preferentemente, puede integrarse en el bloque 611.

En una forma de realización de la codificación y la decodificación, sin implementar las etapas de mezcla descendente y mezcla ascendente, la señal multicanal decodificada $\hat{\mathbf{B}}$ se obtiene a la salida del bloque de decodificación local 612.

10 En la forma de realización donde la etapa de mezcla descendente en 610 se ha utilizado para la codificación, la decodificación local implementada en el bloque 612 permite obtener una señal de audio decodificada

$\hat{\mathbf{B}}$

15 que se envía como entrada desde el bloque 613 de mezcla ascendente.

Por lo tanto, el bloque 613 implementa una etapa opcional (mezcla ascendente) para aumentar el número de canales. En una forma de realización de esta etapa, para el canal de una señal mono

20 $\hat{\mathbf{B}}'$,

consiste en convolucionar la señal

$\hat{\mathbf{B}}$

25 mediante diferentes respuestas de impulsos espaciales de sala (SRIR, "*Spatial Room Impulse Response*"); estas SRI R se definen en el orden ambisónico original de \mathbf{B} . Otros métodos de descorrelación son posibles, por ejemplo la aplicación de filtros de descorrelación de paso total a los diferentes canales de la señal

30 $\hat{\mathbf{B}}'$.

El bloque 614 implementa una etapa opcional (SB) de división en subbandas para obtener subbandas en el dominio de tiempo o en un dominio de transformada.

35 El bloque 615 determina (Inf $\hat{\mathbf{B}}$) una información representativa de una imagen espacial de la señal multicanal decodificada de manera similar a lo que se ha descrito para los bloques 621 y 321 (para la señal multicanal original), aplicado esta vez a la señal multicanal decodificada $\hat{\mathbf{B}}$ obtenida a la salida del bloque 612 o del bloque 613 de acuerdo con las formas de realización de decodificación local. Este bloque 615 es equivalente al bloque 375 de la figura 3.

40 De la misma manera que para los bloques 621 y 321, en una forma de realización, esta información es una información de energía asociada a direcciones de origen del sonido (asociadas a direcciones de altavoces virtuales distribuidos en una esfera unitaria). Como se explicó anteriormente, se puede utilizar un método de tipo SRP u otro (como las variantes descritas anteriormente) para determinar la imagen espacial de la señal multicanal decodificada.

45 En otra forma de realización, esta información es una matriz de covarianza de los canales de la señal multicanal decodificada.

Esta matriz de covarianza se obtiene de la siguiente manera:

50 $\hat{\mathbf{C}} = \hat{\mathbf{B}} \cdot \hat{\mathbf{B}}^T$ con un factor de normalización aproximado (en el caso real) o

$$\hat{\mathbf{C}} = \text{Re}(\hat{\mathbf{B}} \cdot \hat{\mathbf{B}}^H)$$

55 con un factor de normalización aproximado (en el caso complejo)

A partir de la información representativa de las imágenes espaciales respectivamente de la señal multicanal original (Inf. \mathbf{B}) y de la señal multicanal decodificada (Inf. $\hat{\mathbf{B}}$), por ejemplo, las matrices de covarianza \mathbf{C} y $\hat{\mathbf{C}}$, el bloque 680 implementa el método de determinación (Det.Corr) de un conjunto de correcciones como se describe en referencia a la figura 2.

Son posibles dos formas de realización particulares de esta determinación y se han descrito en referencia a las figuras 4 y 5.

5 En la forma de realización de la figura 4 se utiliza un método que usa el renderizado en altavoz, y en la forma de realización de la figura 5 se utiliza un método implementado directamente en el dominio ambisónico basado en una factorización de tipo Cholesky o por descomposición en valores propios.

10 Por lo tanto, si la forma de realización de la figura 4 se aplica en 630, el conjunto de correcciones determinado es un conjunto de ganancias g_n para un conjunto de direcciones (Θ_n, Φ_n) definidas por un conjunto de altavoces virtuales. Este conjunto de ganancias puede determinarse en forma de una matriz de corrección \mathbf{G} como se describe con referencia a la Figura 4.

15 Este conjunto de ganancias (Corr.) se codifica entonces en 640. La codificación de este conjunto de ganancias puede consistir en la codificación de la matriz de corrección \mathbf{G} o \mathbf{G}_{norm} .

20 Se observa que la matriz \mathbf{G} de tamaño $K \times K$ es simétrica, por lo que, de acuerdo con la invención, solo es posible codificar el triángulo inferior o superior de \mathbf{G} o \mathbf{G}_{norm} , es decir, $K \times (K+1)/2$ valores. En general, los valores en la diagonal son positivos. En una forma de realización, la codificación de la matriz \mathbf{G} o \mathbf{G}_{norm} se realiza mediante cuantificación escalar (con o sin bit de signo) en función de si los valores están fuera de diagonal o no. En las variantes en las que se utiliza \mathbf{G}_{norm} , se puede omitir la codificación y transmisión del primer valor de la diagonal (correspondiente a la componente omnidireccional) de \mathbf{G}_{norm} ya que siempre es 1; por ejemplo, en el caso ambisónico de orden 1 a $K=4$ canales, esto equivale a transmitir solo 9 valores en lugar de $K \times (K+1)/2 = 10$ valores. En variantes, se pueden utilizar otros métodos de cuantificación escalar o vectorial (con o sin predicción).

25 Si la forma de realización de la figura 5 se aplica en 630, el conjunto de correcciones determinado es una matriz de transformación \mathbf{T} o \mathbf{T}_{norm} que, a continuación, se codifica en 640.

30 Se observa que la matriz \mathbf{T} de tamaño $K \times K$ es triangular en la variante que utiliza la factorización de Cholesky y simétrica en la variante que utiliza la descomposición en valores propios; por lo tanto, de acuerdo con la invención, solo es posible codificar el triángulo inferior o superior de \mathbf{T} o \mathbf{T}_{norm} , es decir, $K \times (K+1)/2$ valores.

35 En general, los valores en la diagonal son positivos. En una forma de realización, la codificación de la matriz \mathbf{T} o \mathbf{T}_{norm} se realiza mediante cuantificación escalar (con o sin bit de signo) en función de si los valores están fuera de diagonal o no. En variantes, se pueden utilizar otros métodos de cuantificación escalar o vectorial (con o sin predicción). En las variantes en las que se utiliza \mathbf{T}_{norm} , se puede omitir la codificación y transmisión del primer valor de la diagonal (correspondiente a la componente omnidireccional) de \mathbf{T}_{norm} ya que siempre es 1; por ejemplo, en el caso ambisónico de orden 1 a $K=4$ canales, esto equivale a transmitir solo 9 valores en lugar de $K \times (K+1)/2 = 10$ valores.

40 El bloque 640 codifica así el conjunto de correcciones determinado y envía el conjunto codificado de correcciones al multiplexor 650.

45 El decodificador recibe en el bloque demultiplexor 660 un flujo binario que comprende una señal de audio codificada procedente de la señal multicanal original y el conjunto codificado de correcciones a aplicar a la señal multicanal decodificada.

El bloque 670 decodifica (Q^{-1}) el conjunto codificado de correcciones. El bloque 680 decodifica (DEC) la señal de audio codificada recibida en el flujo.

50 En una forma de realización de la codificación y la decodificación, sin implementar las etapas de mezcla descendente y mezcla ascendente, la señal multicanal decodificada $\hat{\mathbf{B}}$ se obtiene a la salida del bloque de decodificación 680.

55 En la forma de realización donde la etapa de mezcla descendente se ha utilizado para la codificación, la decodificación implementada en el bloque 680 permite obtener una señal de audio decodificada

$$\hat{\mathbf{B}}'$$

que se envía como entrada desde el bloque 681 de mezcla ascendente.

60 Por lo tanto, el bloque 681 implementa una etapa opcional (mezcla ascendente) para aumentar el número de canales. En una forma de realización de esta etapa, para el canal de una señal mono

$$\hat{\mathbf{B}}'$$

consiste en convolucionar la señal

$\hat{\mathbf{B}}$

5 mediante diferentes respuestas de impulsos espaciales de sala (SRI R, "Spatial Room Impulse Response"); estas SRI R se definen en el orden ambisónico original de \mathbf{B} . Otros métodos de descorrelación son posibles, por ejemplo la aplicación de filtros de descorrelación de paso total a los diferentes canales de la señal

$\hat{\mathbf{B}}$

10 El bloque 682 implementa una etapa opcional (SB) de división en subbandas para obtener subbandas en el dominio de tiempo o en un dominio de transformada, y el bloque 691 agrupa las subbandas para encontrar la señal multicanal de salida.

15 El bloque 690 implementa una corrección (CORR) de la señal multicanal decodificada por el conjunto de correcciones decodificado en el bloque 670 para obtener una señal multicanal decodificada corregida ($\hat{\mathbf{B}}$ Corr).

20 En una forma de realización donde el conjunto de correcciones es un conjunto de ganancias como se describe en referencia a la figura 4, este conjunto de ganancias se recibe en la entrada del bloque 690 de corrección.

25 Si el conjunto de ganancias está en forma de una matriz de corrección directamente aplicable a la señal multicanal decodificada, definida, por ejemplo, bajo la forma $\mathbf{G} = \mathbf{E} \cdot \text{diag}([g_0 \dots g_{N-1}]) \cdot \mathbf{D}$ o $\mathbf{G}_{\text{norm}} = g_{\text{norm}} \cdot \mathbf{G}$, esta matriz \mathbf{G} o \mathbf{G}_{norm} se aplica a la señal multicanal decodificada $\hat{\mathbf{B}}$ para obtener la señal ambisónica de salida corregida ($\hat{\mathbf{B}}$ Corr).

Si el bloque 690 recibe un conjunto de ganancias g_n , el bloque 690 aplica para cada altavoz virtual la ganancia g_n correspondiente. La aplicación de esta ganancia permite obtener, en este altavoz, la misma energía que la señal original.

30 De este modo, en cada altavoz se corrige el renderizado de las señales decodificadas.

35 Por tanto, se implementa una etapa de codificación acústica, por ejemplo una codificación ambisónica, para obtener componentes de la señal multicanal, por ejemplo componentes ambisónicas. A continuación, estas componentes ambisónicas se suman para obtener la señal multicanal de salida corregida ($\hat{\mathbf{B}}$ Corr).

En una forma de realización donde el conjunto de correcciones es una matriz de transformación como se describe en referencia a la figura 5, la matriz de transformación \mathbf{T} decodificada en 670 se recibe como entrada en el bloque 690 de corrección.

40 Con esta forma de realización, el bloque 690 realiza la etapa de corrección de la señal multicanal decodificada mediante la aplicación de la matriz de transformación \mathbf{T} o \mathbf{T}_{norm} directamente a la señal multicanal decodificada, en el dominio ambisónico, para obtener la señal ambisónica de salida corregida ($\hat{\mathbf{B}}$ Corr).

45 Incluso si la invención se aplica al caso ambisónico, en variantes se pueden convertir otros formatos (multicanal, objeto...) en ambisónico para aplicar los métodos implementados de acuerdo con las diferentes formas de realización descritas. Un ejemplo de realización de tal conversión de un formato multicanal u objeto a un formato ambisónico se describe en la figura 2 de la especificación 3GPP TS 26.259 (v15.0.0).

50 En la **figura 7** se ilustra un dispositivo de codificación DCOD y un dispositivo de decodificación DDEC, en el sentido de la invención; estos dispositivos son duales entre sí (en el sentido de "reversibles") y están conectados entre sí por una red de comunicación RES.

El dispositivo de codificación DOOD comprende un circuito de procesamiento que incluye típicamente:

- 55 - una memoria MEM1 para almacenar datos de instrucciones de un programa informático en el sentido de la invención (estas instrucciones se pueden repartir entre el codificador DCOD y el decodificador DDEC);
- 60 - una interfaz INT1 de recepción de una señal multicanal original \mathbf{B} , por ejemplo una señal ambisónica distribuida en diferentes canales (por ejemplo, cuatro canales W, Y, Z, X de orden 1) para su codificación en compresión en el sentido de la invención;

- un procesador PROC1 para recibir esta señal y procesarla ejecutando las instrucciones del programa informático almacenado en la memoria MEM1, para su codificación; y
- una interfaz de comunicación COM1 para transmitir las señales codificadas a través de la red.

5 El dispositivo de decodificación DDEC comprende un circuito de procesamiento propio, que incluye típicamente:

- una memoria MEM2 para almacenar datos de instrucciones de un programa informático en el sentido de la invención (estas instrucciones se pueden repartir entre el codificador DOOD y el decodificador DDEC como se ha indicado previamente);
- una interfaz COM2 para recibir de la red RES las señales codificadas para su decodificación en compresión en el sentido de la invención;
- un procesador PROC2 para procesar estas señales ejecutando las instrucciones del programa informático almacenado la memoria MEM2, para su decodificación; y
- una interfaz de salida INT2 para entregar las señales decodificadas corregidas (**B** Corr), por ejemplo en forma de canales ambisónicos W...X, para su reproducción.

10
15
20 Naturalmente, esta figura 7 ilustra un ejemplo de una realización estructural de un codificador (codificador o decodificador) en el sentido de la invención. Las figuras 3 a 6 analizadas anteriormente describen en detalle las realizaciones más funcionales de estos codificadores.

25

REIVINDICACIONES

- 5 1. Método de determinación de un conjunto de correcciones (Corr.) a realizar en una señal sonora multicanal, **caracterizado por** la determinación de dicho conjunto de correcciones a partir de una información representativa de una imagen espacial de una señal multicanal original (Inf. **B**) y de una información representativa de una imagen espacial de la señal multicanal original codificada y posteriormente decodificada (Inf. **B̂**).
- 10 2. Método de acuerdo con la reivindicación 1, en donde la determinación del conjunto de correcciones se realiza mediante subbanda de frecuencia.
- 15 3. Método de decodificación de una señal sonora multicanal, que comprende las siguientes etapas:
- recibir (350) un flujo binario que comprende una señal de audio codificada procedente de una señal multicanal original y una información representativa de una imagen espacial de la señal multicanal original;
 - decodificar (370) la señal de audio codificada recibida y obtener una señal multicanal decodificada;
 - decodificar (360) la información representativa de una imagen espacial de la señal multicanal original;
 - determinar (375) una información representativa de una imagen espacial de la señal multicanal decodificada;
 - determinar (380) un conjunto de correcciones a realizar en la señal decodificada de acuerdo con el método de determinación conforme a una de las reivindicaciones 1 a 2;
 - corregir (390) la señal multicanal decodificada mediante el conjunto de correcciones determinado.
- 20 4. Método de codificación de una señal sonora multicanal, que comprende las siguientes etapas:
- codificar (611) una señal de audio procedente de una señal multicanal original;
 - determinar (621) una información representativa de una imagen espacial de la señal multicanal original;
 - decodificar de manera local (612) la señal de audio codificada y obtener una señal multicanal decodificada;
 - determinar (615) una información representativa de una imagen espacial de la señal multicanal decodificada;
 - determinar (630) un conjunto de correcciones a realizar en la señal multicanal decodificada de acuerdo con el método de determinación conforme a una de las reivindicaciones 1 a 2;
 - codificar (640) el conjunto de correcciones determinado.
- 25 5. Método de decodificación de acuerdo con la reivindicación 3 o método de codificación de acuerdo con la reivindicación 4, en donde la información representativa de una imagen espacial es una matriz de covarianza y la determinación del conjunto de correcciones comprende además las etapas siguientes:
- obtener una matriz de ponderación que comprende vectores de ponderación asociados a un conjunto de altavoces virtuales;
 - determinar una imagen espacial de la señal multicanal original a partir de la matriz de ponderación obtenida y a partir de la matriz de covarianza de la señal multicanal original;
 - determinar una imagen espacial de la señal multicanal decodificada a partir de la matriz de ponderación obtenida y a partir de la matriz de covarianza de la señal multicanal decodificada determinada;
 - calcular una relación entre la imagen espacial de la señal multicanal original y la imagen espacial de la señal multicanal decodificada en las direcciones de los altavoces del conjunto de altavoces virtuales, para obtener un conjunto de ganancias.
- 30 6. Método de decodificación de acuerdo con la reivindicación 3, en donde la información representativa de una imagen espacial de la señal multicanal original recibida es la imagen espacial de la señal multicanal original y la determinación del conjunto de correcciones comprende además las siguientes etapas:
- obtener una matriz de ponderación que comprende vectores de ponderación asociados a un conjunto de altavoces virtuales;
 - determinar una imagen espacial de la señal multicanal decodificada a partir de la matriz de ponderación obtenida y a partir de la información representativa de una imagen espacial de la señal multicanal decodificada determinada;
 - calcular una relación entre la imagen espacial de la señal multicanal original y la imagen espacial de la señal multicanal decodificada en las direcciones de los altavoces del conjunto de altavoces virtuales, para obtener un conjunto de ganancias.
- 35 7. Método de decodificación de acuerdo con la reivindicación 3 o método de codificación de acuerdo con la reivindicación 4, en donde la información representativa de una imagen espacial es una matriz de covarianza y la determinación del conjunto de correcciones comprende una etapa de determinación de una matriz de transformación por descomposición matricial de las dos matrices de covarianza, constituyendo la matriz de transformación el conjunto de correcciones.
- 40 8. Método de decodificación de acuerdo con una de las reivindicaciones 5 a 7, en donde la corrección de la señal multicanal decodificada por el conjunto de correcciones determinado se realiza mediante la aplicación del conjunto de correcciones a la señal multicanal decodificada.
- 45
- 50
- 55
- 60
- 65

9. Método de decodificación de acuerdo con una de las reivindicaciones 5 a 6, en donde la corrección de la señal multicanal decodificada por el conjunto de correcciones determinado se realiza de acuerdo con las siguientes etapas:

- 5 - decodificar de manera acústica la señal multicanal decodificada en el conjunto definido de altavoces virtuales;
- aplicar el conjunto de ganancias obtenido a las señales procedentes de la decodificación acústica;
- codificar de manera acústica las señales procedentes de la decodificación acústica y corregidas para obtener componentes de la señal multicanal;
- sumar las componentes de la señal multicanal obtenidas de esta manera para obtener una señal multicanal corregida.

10

10. Método de decodificación de una señal sonora multicanal, que comprende las siguientes etapas:

- recibir un flujo binario que comprende una señal de audio codificada procedente de una señal multicanal original y un conjunto codificado de correcciones a realizar en la señal multicanal decodificada, habiéndose codificado el conjunto de correcciones de acuerdo con un método de codificación conforme a una de las reivindicaciones 4, 5 o 7;
- 15 - decodificar la señal de audio codificada recibida y obtener una señal multicanal decodificada;
- decodificar el conjunto codificado de correcciones;
- corregir la señal multicanal decodificada mediante la aplicación del conjunto de correcciones decodificado a la señal multicanal decodificada.

20

11. Método de decodificación de una señal sonora multicanal, que comprende las siguientes etapas:

- recibir un flujo binario que comprende una señal de audio codificada procedente de una señal multicanal original y un conjunto codificado de correcciones a realizar en la señal multicanal decodificada, habiéndose codificado el conjunto de correcciones de acuerdo con un método de codificación conforme a la reivindicación 5;
- 25 - decodificar la señal de audio codificada recibida y obtener una señal multicanal decodificada;
- decodificar el conjunto codificado de correcciones;
- corregir la señal multicanal decodificada mediante el conjunto de correcciones decodificadas de acuerdo con las siguientes etapas:

30

- . decodificar de manera acústica la señal multicanal decodificada en el conjunto altavoces virtuales;
- . aplicar el conjunto de ganancias obtenido a las señales procedentes de la decodificación acústica;
- . codificar de manera acústica las señales procedentes de la decodificación acústica y corregidas para obtener componentes de la señal multicanal;
- 35 . sumar las componentes de la señal multicanal obtenidas de esta manera para obtener una señal multicanal corregida.

12. Dispositivo de decodificación que comprende un circuito de procesamiento para la implementación del método de decodificación de acuerdo con una de las reivindicaciones 3 o 5 a 11.

40

13. Dispositivo de codificación que comprende un circuito de procesamiento para la implementación del método de codificación de acuerdo con una de las reivindicaciones 4, 5 o 7.

45

14. Soporte de almacenamiento, legible por un procesador, que almacena un programa informático que comprende instrucciones para la ejecución del método de decodificación de acuerdo con una de las reivindicaciones 3 o 5 a 11 o del método de codificación de acuerdo con una de las reivindicaciones 4, 5 o 7.

[Fig 1]

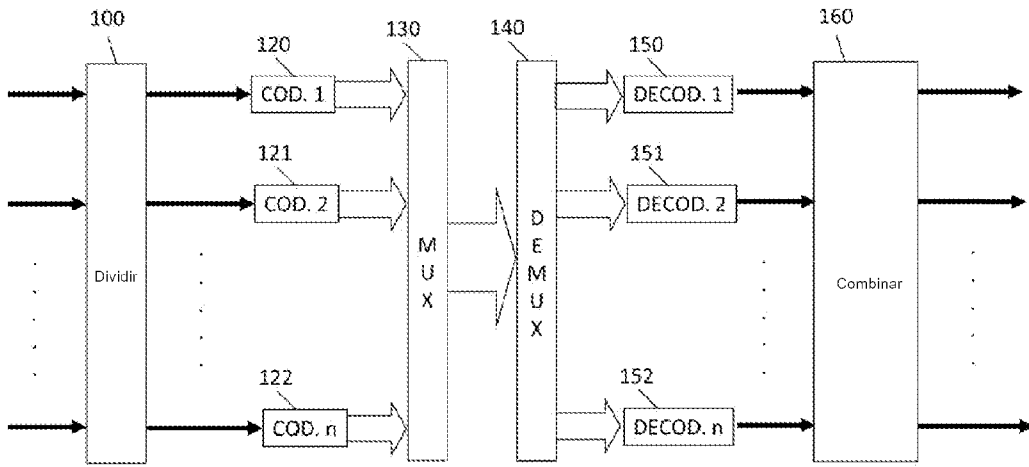


FIG.1 (Estado de la técnica)

[Fig 2]

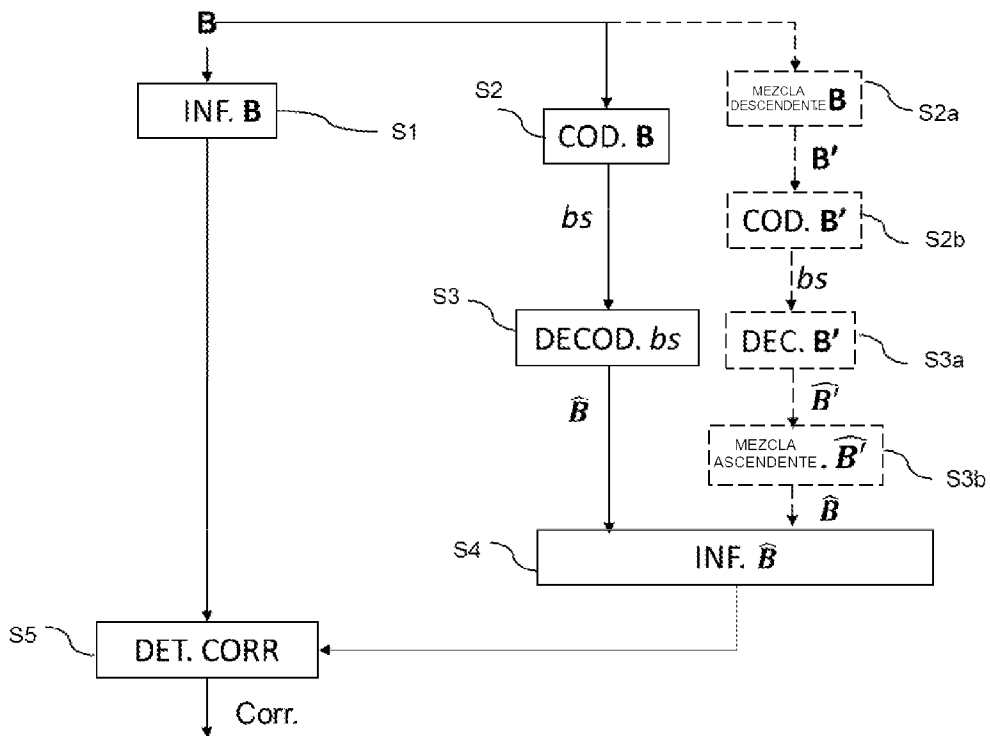


FIG.2

[Fig 3]

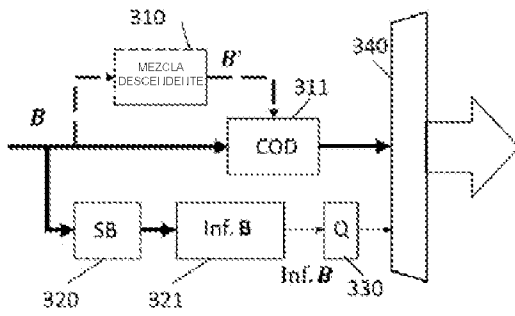
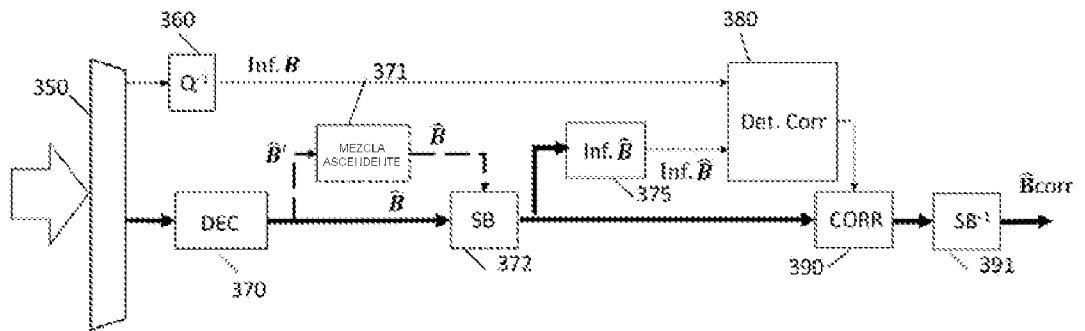


FIG.3



[Fig 4]

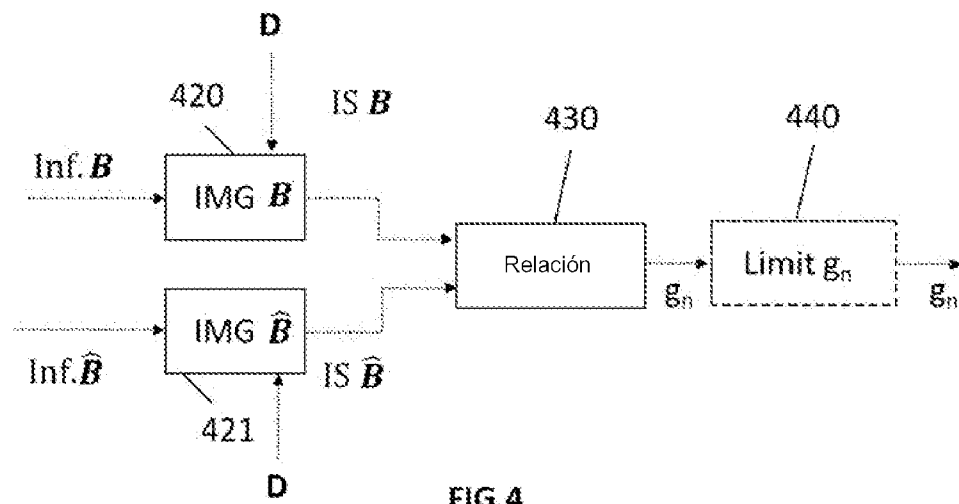


FIG.4

[Fig 5]

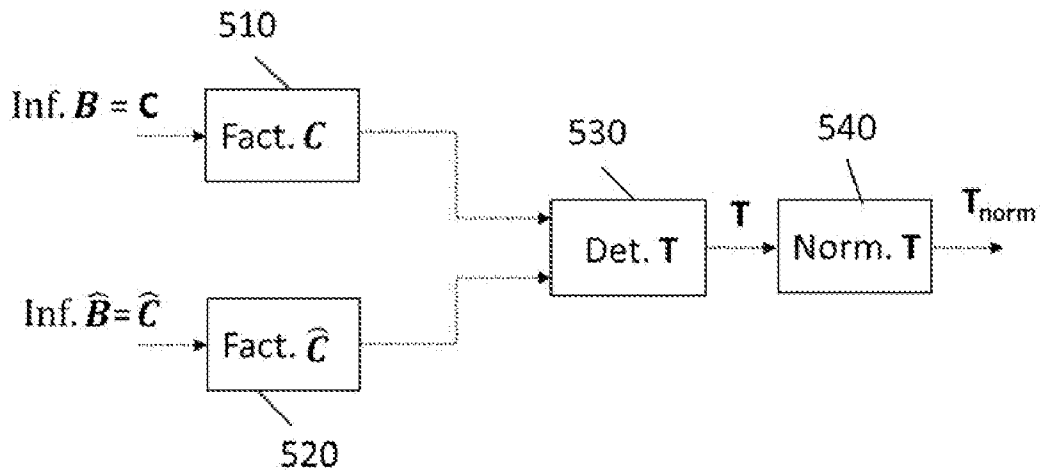


FIG.5

[Fig 6]

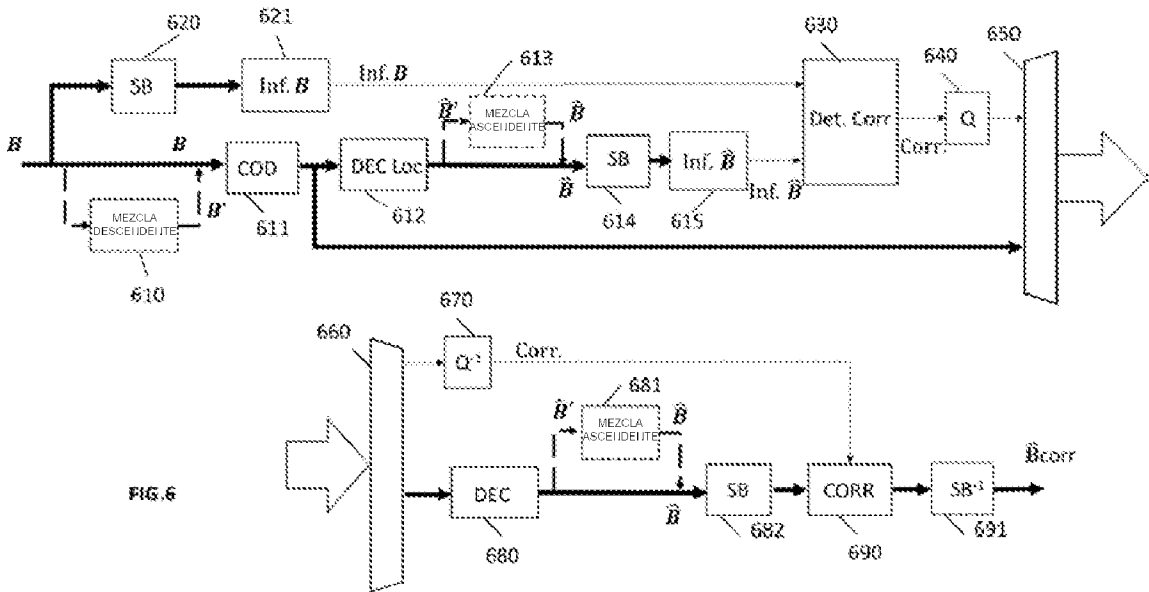


FIG. 6

[Fig 7]

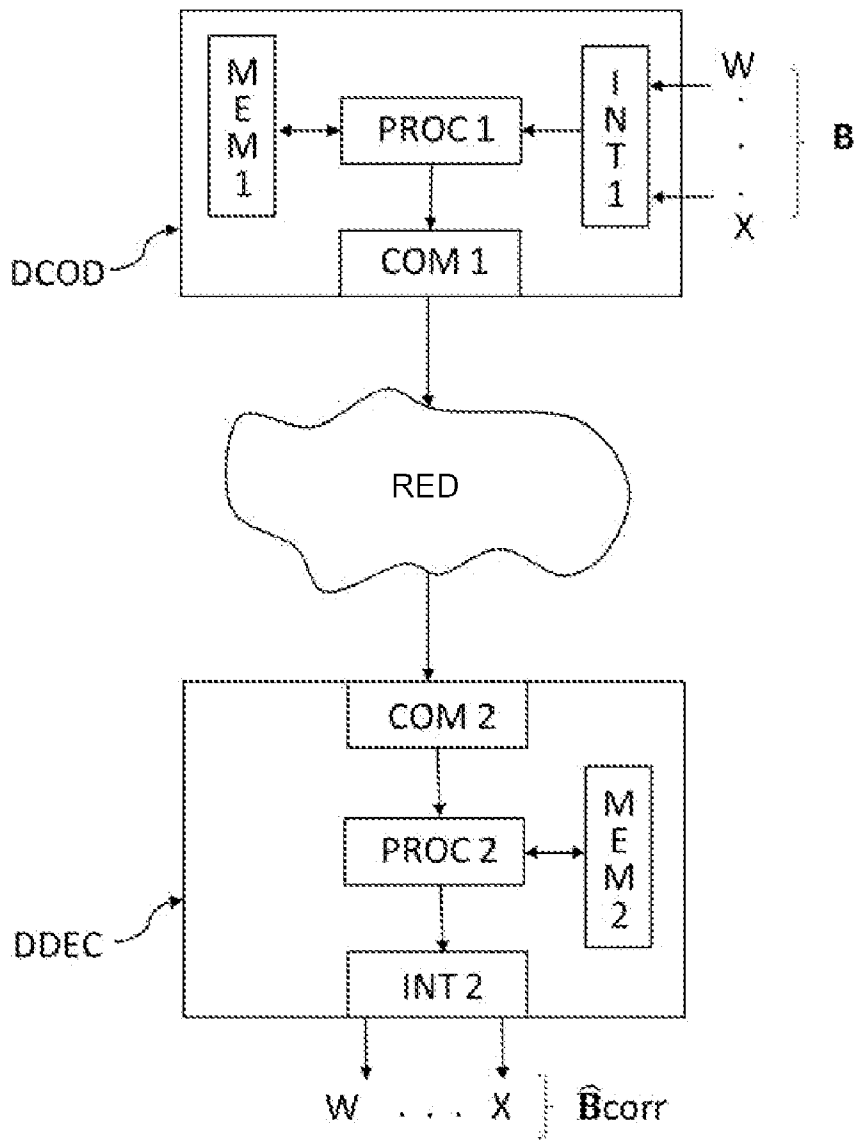


FIG.7