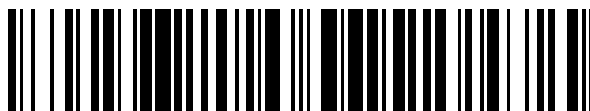


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 696 930**

51 Int. Cl.:

**H04S 7/00** (2006.01)

**G10L 19/008** (2013.01)

12

## TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **29.05.2015 PCT/US2015/033273**

87 Fecha y número de publicación internacional: **03.12.2015 WO15184316**

96 Fecha de presentación y número de la solicitud europea: **29.05.2015 E 15727844 (1)**

97 Fecha y número de publicación de la concesión europea: **15.08.2018 EP 3149972**

54 Título: **Obtención de información de simetría para renderizadores de audio ambisónicos de orden superior**

30 Prioridad:

**30.05.2014 US 201462005829 P**

**11.07.2014 US 201462023662 P**

**28.05.2015 US 201514724615**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**18.01.2019**

73 Titular/es:

**QUALCOMM INCORPORATED (100.0%)**

**5775 Morehouse Drive**

**San Diego, CA 92121-1714, US**

72 Inventor/es:

**PETERS, NILS GÜNTHER;**

**SEN, DIPANJAN y**

**MORRELL, MARTIN JAMES**

74 Agente/Representante:

**FORTEA LAGUNA, Juan José**

ES 2 696 930 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

Obtención de información de simetría para renderizadores de audio ambisónicos de orden superior

## 5 CAMPO TÉCNICO

[0001] Esta divulgación se refiere a renderización de información y, más específicamente, información de renderización para datos de audio ambisónicos de orden superior (HOA).

## 10 ANTECEDENTES

[0002] Durante la producción de contenido de audio, el ingeniero de sonido puede renderizar el contenido de audio usando un renderizador específico en un intento de personalizar el contenido de audio para las configuraciones objetivo de los altavoces utilizados para reproducir el contenido de audio. En otras palabras, el ingeniero de sonido puede renderizar el contenido de audio y reproducir el contenido de audio renderizado utilizando altavoces dispuestos en la configuración objetivo. El ingeniero de sonido puede a continuación remezclar varios aspectos del contenido de audio, renderizar el contenido de audio remezclado y reproducir de nuevo el contenido de audio remezclado y renderizado utilizando los altavoces dispuestos en la configuración objetivo. El ingeniero de sonido puede iterar de esta manera hasta que el contenido de audio proporcione una cierta intención artística. De esta forma, el ingeniero de sonido puede producir contenido de audio que proporcione un cierto propósito artístico o que proporcione un cierto campo de sonido durante la reproducción (por ejemplo, para acompañar el contenido de vídeo reproducido junto con el contenido de audio). El "Audio Engineering Society Convention Paper 8426" ["Documento de la Convención de la Sociedad de Ingeniería de Audio 8426"], con fecha del 13 al 16 de mayo de 2011, describe una disposición de reproducción de sonido espacial tridimensional que utiliza ambisónica de orden superior. Un documento titulado "Reducing the Bandwidth of Sparse Symmetric Matrices" ["Reducción del ancho de banda de matrices simétricas dispersas"], con fecha 1 de enero de 1969, describe varias disposiciones de procesamiento de matrices. Un documento titulado "Symmetric Eigenvalue Problem: Tridiagonal Reduction" ["Problema de valor propio simétrico: Reducción tridiagonal"], con fecha del 18 de mayo de 2009, describe problemas y soluciones de Eigenvalue utilizando una matriz simétrica.

## 30 SUMARIO

[0003] En general, las técnicas se describen para especificar información de renderización de audio en un flujo de bits representativo de datos de audio. En otras palabras, las técnicas pueden proporcionar una forma de señalar la información de renderización de audio utilizada durante la producción de contenido de audio a un dispositivo de reproducción, que luego puede usar la información de renderización de audio para renderizar el contenido de audio. Proporcionar la información de renderización de esta manera permite que el dispositivo de reproducción renderice el contenido de audio de la manera prevista por el ingeniero de sonido, y de ese modo potencialmente asegurar la reproducción apropiada del contenido de audio de forma tal que la intención artística sea potencialmente comprendida por un oyente. En otras palabras, la información de renderización utilizada durante la renderización por el ingeniero de sonido se proporciona de acuerdo con las técnicas descritas en esta divulgación para que el dispositivo de reproducción de audio pueda utilizar la información de renderización para renderizar el contenido de audio de la manera prevista por el ingeniero de sonido, garantizando así una experiencia más consistente durante la producción y la reproducción del contenido de audio en comparación con los sistemas que no proporcionan esta información de renderización de audio.

[0004] Los detalles de uno o más aspectos de las técnicas se exponen en los dibujos adjuntos y la descripción a continuación. Otras características, objetivos y ventajas de las técnicas resultarán evidentes a partir de la descripción y de los dibujos y a partir de las reivindicaciones.

## 50 BREVE DESCRIPCIÓN DE LOS DIBUJOS

## [0005]

55 La FIG. 1 es un diagrama que ilustra funciones de base armónicas esféricas de varios órdenes y subórdenes.

La FIG. 2 es un diagrama que ilustra un sistema que puede realizar diversos aspectos de las técnicas descritas en esta divulgación.

60 La FIG. 3 es un diagrama de bloques que ilustra, con mayor detalle, otro ejemplo del dispositivo de codificación de audio mostrado en el ejemplo de la FIG. 2 que puede realizar diversos aspectos de las técnicas descritas en esta divulgación.

65 La FIG. 4 es un diagrama de bloques que ilustra el dispositivo de descodificación de audio de la FIG. 2 con mayor detalle.

La FIG. 5 es un diagrama de flujo que ilustra el funcionamiento a modo de ejemplo de un dispositivo de codificación de audio en la realización de diversos aspectos de las técnicas de síntesis basadas en vectores descritas en esta divulgación.

La FIG. 6 es un diagrama de flujo que ilustra el funcionamiento a modo de ejemplo de un dispositivo de descodificación de audio en la realización de varios aspectos de las técnicas descritas en esta divulgación.

La FIG. 7 es un diagrama de flujo que ilustra el funcionamiento de ejemplo de un sistema, tal como uno del sistema mostrado en el ejemplo de las FIGs. 2, en la realización de diversos aspectos de las técnicas descritas en esta divulgación.

Las FIGs. 8A-8D son un diagrama que ilustra los flujos de bits formados de acuerdo con las técnicas descritas en esta divulgación.

Las FIGs. 8E-8G son diagramas que ilustran partes de la información del flujo de bits o del canal lateral, que puede especificar los componentes espaciales comprimidos con mayor detalle.

La FIG. 9 es un diagrama que ilustra un ejemplo de ganancias mínimas y máximas dependientes de orden de ambisónicos de orden superior (HOA) dentro de una matriz de renderización de HOA.

La FIG. 10 es un diagrama que ilustra una matriz de renderización HOA de sexto orden parcialmente dispersa para 22 altavoces.

La FIG. 11 es un diagrama de flujo que ilustra la señalización de las propiedades de simetría.

## DESCRIPCIÓN DETALLADA

**[0006]** La evolución del sonido envolvente ha dejado disponibles muchos formatos de salida para el entretenimiento de hoy en día. Los ejemplos de tales formatos de sonido envolvente de consumidor son en su mayoría "canales" basados en que, implícitamente, especifican las señales a los altavoces en ciertas coordenadas geométricas. Los formatos de sonido envolvente de consumidor incluyen el popular formato 5.1 (que incluye los siguientes seis canales: delantero izquierdo (FL), delantero derecho (FR), central o delantero central, posterior izquierdo o envolvente izquierdo, posterior derecho o envolvente derecho, y los efectos de baja frecuencia (LFE)), el creciente formato 7.1, varios formatos que incluyen altavoces de altura tales como el formato 7.1.4 y el formato 22.2 (por ejemplo, para usar con el estándar de Televisión de Ultra Alta Definición). Los formatos que no son de consumidor pueden abarcar cualquier número de altavoces (en geometrías simétricas y no simétricas) a menudo denominados "formaciones envolventes". Un ejemplo de una configuración de este tipo incluye 32 altavoces colocados en coordenadas en las esquinas de un icosaedro truncado.

**[0007]** La entrada a un futuro codificador de MPEG es, opcionalmente, uno de tres formatos posibles: (i) audio tradicional basado en canales (como se ha expuesto anteriormente), que está destinado a ser reproducido mediante altavoces en posiciones preestablecidas; (ii) audio basado en objetos, que implica datos discretos de modulación por código de pulso (PCM) para objetos de audio individuales con metadatos asociados que contienen sus coordenadas de ubicación (entre otra información); y (iii) audio basado en la escena, que implica la renderización del campo sonoro utilizando coeficientes de funciones de base armónicas esféricas (también llamados "coeficientes armónicos esféricos" o SHC, "Ambisónica de Orden Superior" o HOA y "coeficientes HOA"). El futuro codificador de MPEG se puede describir con mayor detalle en un documento titulado "Call for Proposals for 3D Audio" ["Convocatoria de propuestas para audio tridimensional"], de la Organización Internacional para la Estandarización/Comisión Electrotécnica Internacional (ISO)/(IEC) JTC1/SC29/WG11/N13411, publicado en enero de 2013 en Ginebra, Suiza, y disponible en <http://mpeg.chiariglione.org/sites/default/files/files/standards/parts/docs/w13411.zip>.

**[0008]** Hay varios formatos basados en canales de "sonido envolvente" en el mercado. Varían, por ejemplo, desde el sistema de cine en casa 5.1 (que es el que ha tenido el mayor éxito en cuanto a irrumpir en las salas de estar, más allá del estéreo) hasta el sistema 22.2 desarrollado por NHK (Nippon Hoso Kyokai o Corporación Difusora de Japón). Los creadores de contenido (por ejemplo, los estudios de Hollywood) desearían producir la banda sonora de una película una vez y no gastar esfuerzos en remezclarla para cada configuración de altavoces. Recientemente, las Organizaciones de Desarrollo de Normas han estado considerando maneras de proporcionar una codificación en un flujo de bits estandarizado y una descodificación subsiguiente que sea adaptable e indiferente con respecto a la geometría (y número) del altavoz y a las condiciones acústicas en la ubicación de la reproducción (que implica un reproductor).

**[0009]** Para proporcionar dicha flexibilidad a los creadores de contenido, se puede usar un conjunto jerárquico de elementos para representar un campo sonoro. El conjunto jerárquico de elementos puede referirse a un conjunto de elementos en el que los elementos están ordenados de tal manera que un conjunto básico de elementos de orden inferior proporciona una representación completa del campo sonoro modelado. A medida que el conjunto se extiende para incluir elementos de orden superior, la representación se torna más detallada, aumentando la resolución.

**[0010]** Un ejemplo de un conjunto jerárquico de elementos es un conjunto de coeficientes armónicos esféricos (SHC). La siguiente expresión demuestra una descripción o representación de un campo sonoro usando SHC:

$$p_i(t, r_r, \theta_r, \varphi_r) = \sum_{\omega=0}^{\infty} \left[ 4\pi \sum_{n=0}^{\infty} j_n(kr_r) \sum_{m=-n}^n A_n^m(k) Y_n^m(\theta_r, \varphi_r) \right] e^{j\omega t},$$

**[0011]** La expresión muestra que la presión  $p_i$  en cualquier punto  $\{r_r, \theta_r, \varphi_r\}$  del campo de sonido, en el momento  $t$ , puede ser representada de forma única por los SHC.  $A_n^m(k)$ . Aquí,  $k = \frac{\omega}{c}$ ,  $c$  es la velocidad del sonido (~343 m/s),  $\{r_r, \theta_r, \varphi_r\}$  es un punto de referencia (o punto de observación),  $j_n(\cdot)$  es la función esférica de Bessel de orden  $n$  y  $Y_n^m(\theta_r, \varphi_r)$  son las funciones de base armónicas esféricas de orden  $n$  y sub-orden  $m$ . Puede reconocerse que el término en corchetes es una representación en el dominio de la frecuencia de la señal (es decir,  $S(\omega, r_r, \theta_r, \varphi_r)$ ) que se puede aproximar mediante diversas transformaciones de tiempo-frecuencia, tales como la transformada discreta de Fourier (DFT), la transformada discreta de coseno (DCT) o una transformada de ondículas. Otros ejemplos de conjuntos jerárquicos incluyen conjuntos de coeficientes de transformada de ondículas y otros conjuntos de coeficientes de funciones de base de multi-resolución.

**[0012]** La FIG. 1 es un diagrama que ilustra las funciones de base armónicas esféricas desde el orden cero ( $n = 0$ ) hasta el cuarto orden ( $n = 4$ ). Como puede verse, para cada orden, hay una expansión de los subórdenes  $m$  que se muestran pero no se señalan explícitamente en el ejemplo de la FIG. 1, con fines de facilidad de ilustración.

**[0013]** Los SHC  $A_n^m(k)$  pueden ser físicamente adquiridos (por ejemplo, registrados) con varias configuraciones de formaciones de micrófonos o, de forma alternativa, se pueden obtener a partir de descripciones basadas en canales o basadas en objetos del campo sonoro. Los SHC representan el audio basado en la escena, donde los SHC pueden ser introducidos en un codificador de audio para obtener SHC codificados que pueden fomentar una transmisión o un almacenamiento más eficaz. Por ejemplo, puede utilizarse una representación de cuarto orden que implique coeficientes  $(1 + 4)^2$  (25 y, por lo tanto, de cuarto orden).

**[0014]** Como se ha señalado anteriormente, los SHC pueden obtenerse de una grabación con micrófono utilizando una formación de micrófonos. En el artículo de Poletti, M., "Three-Dimensional Surround Sound Systems Based on Spherical Harmonics" ["Sistemas tridimensionales de sonido envolvente basados en armónicos esféricos"], J. Audio Eng. Soc., Vol. 53, n.º 11, noviembre de 2005, págs. 1004 a 1025, se describen varios ejemplos de cómo se pueden obtener los SHC a partir de formaciones de micrófonos.

**[0015]** Para ilustrar cómo los SHC se pueden obtener a partir de una descripción basada en objetos, considérese la siguiente ecuación. Los coeficientes  $A_n^m(k)$  para el campo sonoro correspondiente a un objeto de audio individual pueden expresarse como:

$$A_n^m(k) = g(\omega)(-4\pi i k) h_n^{(2)}(kr_s) Y_n^{m*}(\theta_s, \varphi_s),$$

donde  $i$  es  $\sqrt{-1}$ ,  $h_n^{(2)}(\cdot)$  es la función esférica de Hankel (del segundo tipo) de orden  $n$  y  $\{r_s, \theta_s, \varphi_s\}$  es la ubicación del objeto. Conocer la energía fuente del objeto  $g(\omega)$  como función de la frecuencia (por ejemplo, utilizando técnicas de análisis de tiempo-frecuencia, tales como realizar una transformada rápida de Fourier en el flujo de PCM) nos permite convertir cada objeto de PCM y la ubicación correspondiente en el SHC.  $A_n^m(k)$ . Además, se puede mostrar (dado que lo anterior es una descomposición lineal y ortogonal) que los coeficientes  $A_n^m(k)$  para cada objeto son aditivos. De esta manera, una multitud de objetos de PCM puede ser representada por los coeficientes  $A_n^m(k)$  (por ejemplo, como una suma de los vectores de coeficientes para los objetos individuales). Esencialmente, los coeficientes contienen información sobre el campo sonoro (la presión como una función de las coordenadas tridimensionales), y lo anterior representa la transformación de los objetos individuales a una representación del campo sonoro global en las proximidades del punto de observación  $\{r_r, \theta_r, \varphi_r\}$ . Las cifras restantes se describen a continuación en el contexto de la codificación de audio basada en objetos y basada en SHC.

**[0016]** La FIG. 2 es un diagrama que ilustra un sistema 10 que puede realizar diversos aspectos de las técnicas descritas en esta divulgación. Como se muestra en el ejemplo de la FIG. 2, el sistema 10 incluye un dispositivo creador de contenido 12 y un dispositivo consumidor de contenido 14. Aunque se describen en el contexto del dispositivo creador de contenido 12 y del dispositivo consumidor de contenido 14, las técnicas pueden implementarse en cualquier contexto en el que los SHC (que también pueden denominarse coeficientes HOA), o cualquier otra representación jerárquica de un campo sonoro, se codifican para formar un flujo de bits representativo de los datos de audio. Además, el dispositivo creador de contenido 12 puede representar cualquier forma de dispositivo informático capaz de



implementar las técnicas descritas en esta divulgación, incluido un equipo manual (o teléfono celular), un ordenador tipo tablet, un teléfono inteligente o un ordenador de sobremesa, para proporcionar unos pocos ejemplos. Del mismo modo, el dispositivo consumidor de contenido 14 puede representar cualquier forma de dispositivo informático capaz de implementar las técnicas descritas en esta divulgación, incluido un equipo manual (o teléfono celular), un ordenador tipo tablet, un teléfono inteligente, un descodificador o un ordenador de sobremesa, para proporcionar unos pocos ejemplos.

**[0017]** El dispositivo creador de contenido 12 puede ser operado por un estudio de cine u otra entidad que pueda generar contenido de audio multicanal para el consumo por parte de los operadores de los dispositivos consumidores de contenido, tales como el dispositivo consumidor de contenido 14. En algunos ejemplos, el dispositivo creador de contenido 12 puede ser operado por un usuario individual que desea comprimir los coeficientes HOA 11. A menudo, este creador de contenido genera contenido de audio junto con el contenido de vídeo. El dispositivo consumidor de contenido 14 puede ser operado por un individuo. El dispositivo consumidor de contenido 14 puede incluir un sistema de reproducción de audio 16, que puede referirse a cualquier forma de sistema de reproducción de audio capaz de representar los SHC para su reproducción como contenido de audio multicanal.

**[0018]** El dispositivo creador de contenido 12 incluye un sistema de edición de audio 18. El dispositivo creador de contenido 12 obtiene grabaciones en vivo 7 en diversos formatos (incluso directamente como coeficientes HOA) y objetos de audio 9, que el dispositivo creador de contenido 12 puede editar utilizando el sistema de edición de audio 18. Un micrófono 5 puede capturar las grabaciones en vivo 7. El creador de contenido puede, durante el proceso de edición, renderizar los coeficientes HOA 11 de los objetos de audio 9, escuchando las alimentaciones de altavoces renderizadas en un intento de identificar varios aspectos del campo sonoro que requieran una edición adicional. El dispositivo creador de contenido 12 puede entonces editar los coeficientes HOA 11 (potencialmente, de manera indirecta mediante la manipulación de diferentes objetos entre los objetos de audio 9 a partir de los cuales pueden obtenerse los coeficientes HOA originales, de la manera descrita anteriormente). El dispositivo creador de contenido 12 puede emplear el sistema de edición de audio 18 para generar los coeficientes HOA 11. El sistema de edición de audio 18 representa cualquier sistema capaz de editar datos de audio y emitir los datos de audio como uno o más coeficientes armónicos esféricos originales.

**[0019]** Cuando se completa el proceso de edición, el dispositivo creador de contenido 12 puede generar un flujo de bits 21 basándose en los coeficientes HOA 11. Es decir, el dispositivo creador de contenido 12 incluye un dispositivo de codificación de audio 20 que representa un dispositivo configurado para codificar o comprimir de otro modo los coeficientes HOA 11 de acuerdo con diversos aspectos de las técnicas descritas en esta divulgación para generar el flujo de bits 21. El dispositivo de codificación de audio 20 puede generar el flujo de bits 21 para su transmisión, por ejemplo, a través de un canal de transmisión, que puede ser un canal cableado o inalámbrico, un dispositivo de almacenamiento de datos o algo similar. El flujo de bits 21 puede representar una versión codificada de los coeficientes HOA 11 y puede incluir un flujo de bits primario y otro flujo de bits lateral, que se puede denominar información de canal lateral.

**[0020]** Mientras se muestra en la FIG. 2 como transmitido directamente al dispositivo consumidor de contenido 14, el dispositivo creador de contenido 12 puede emitir el flujo de bits 21 a un dispositivo intermedio situado entre el dispositivo creador de contenido 12 y el dispositivo consumidor de contenido 14. Este dispositivo intermedio puede almacenar el flujo de bits 21 para su posterior suministro al dispositivo consumidor de contenido 14, que puede solicitar el flujo de bits. El dispositivo intermedio puede comprender un servidor de ficheros, un servidor web, un ordenador de sobremesa, un ordenador portátil, un ordenador tipo tablet, un teléfono móvil, un teléfono inteligente o cualquier otro dispositivo capaz de almacenar el flujo de bits 21 para su posterior recuperación mediante un descodificador de audio. Este dispositivo intermedio puede residir en una red de suministro de contenido, capaz de transmitir en flujo continuo el flujo de bits 21 (y posiblemente junto con la transmisión de un correspondiente flujo de bits de datos de vídeo) a los abonados, tales como el dispositivo consumidor de contenido 14 que solicita el flujo de bits 21.

**[0021]** De forma alternativa, el dispositivo creador de contenido 12 puede almacenar el flujo de bits 21 en un medio de almacenamiento, tal como un disco compacto, un disco de vídeo digital, un disco de vídeo de alta definición u otros medios de almacenamiento, la mayoría de los cuales pueden ser leídos por un ordenador y, por lo tanto, pueden denominarse medios de almacenamiento legibles por ordenador o medios de almacenamiento no transitorio legibles por ordenador. En este contexto, el canal de transmisión puede referirse a aquellos canales por los cuales se transmiten contenidos almacenados en los medios (y pueden incluir tiendas minoristas y otros mecanismos de suministro basados en almacenes). En cualquier caso, las técnicas de esta divulgación no deberían, por lo tanto, limitarse, a este respecto, al ejemplo de la FIG. 2.

**[0022]** Como se muestra adicionalmente en el ejemplo de la FIG. 2, el dispositivo consumidor de contenido 14 incluye el sistema de reproducción de audio 16. El sistema de reproducción de audio 16 puede representar cualquier sistema de reproducción de audio capaz de reproducir datos de audio multicanal. El sistema de reproducción de audio 16 puede incluir un número de diferentes reproductores 22. Los reproductores 22 pueden proporcionar, cada uno, una forma diferente de renderización, donde las diferentes formas de renderización pueden incluir una o más de las diversas formas de realizar la panoramización de amplitud de base vectorial (VBAP) y/o una o más de las diversas formas de realizar la síntesis del campo sonoro. Como se usa en el presente documento, "A y/o B" significa "A o B" o

ambos "A y B".

**[0023]** El sistema de reproducción de audio 16 puede incluir además un dispositivo de descodificación de audio 24. El dispositivo de descodificación de audio 24 puede representar un dispositivo configurado para descodificar los coeficientes HOA 11' desde el flujo de bits 21, donde los coeficientes HOA 11' pueden ser similares a los coeficientes HOA 11, pero difieren debido a las operaciones con pérdidas (por ejemplo, cuantificación) y/o a la transmisión mediante el canal de transmisión. El sistema de reproducción de audio 16 puede, después de descodificar el flujo de bits 21, obtener los coeficientes HOA 11' y renderizar los coeficientes HOA 11' en las señales de salida de altavoces 25. Las señales de los altavoces 25 pueden accionar uno o más altavoces (que no se muestran en el ejemplo de la FIG. 2, con fines de facilidad de ilustración).

**[0024]** Para seleccionar el renderizador apropiado o, en algunos casos, generar un reproductor apropiado, el sistema de reproducción de audio 16 puede obtener información de los altavoces 13 indicativa de un número de altavoces y/o una geometría espacial de los altavoces. En algunos casos, el sistema de reproducción de audio 16 puede obtener la información de los altavoces 13 usando un micrófono de referencia y accionando los altavoces de tal manera como para determinar dinámicamente la información de altavoces 13. En otros casos, o conjuntamente con la determinación dinámica de la información de altavoces 13, el sistema de reproducción de audio 16 puede inducir a un usuario a interactuar con el sistema de reproducción de audio 16 e introducir la información de altavoces 13.

**[0025]** El sistema de reproducción de audio 16 puede entonces seleccionar uno de los renderizadores de audio 22 basándose en la información de altavoces 13. En algunos casos, el sistema de reproducción de audio 16 puede, cuando ninguno de los renderizadores de audio 22 está dentro de alguna medición de similitud de umbral (en cuanto a geometría de altavoces) con la geometría de altavoces especificada en la información de altavoces 13, generar un renderizador entre los renderizadores de audio 22 basándose en la información de altavoces 13. El sistema de reproducción de audio 16 puede, en algunos casos, generar uno de los renderizadores de audio 22 basándose en la información de altavoces 13 sin intentar primero seleccionar un renderizador existente entre los renderizadores de audio 22. Uno o más altavoces 3 pueden a continuación reproducir las alimentaciones de altavoz renderizadas 25.

**[0026]** En algunos casos, el sistema de reproducción de audio 16 puede seleccionar cualquiera de los renderizadores de audio 22 y puede configurarse para seleccionar uno o más de los renderizadores de audio 22 dependiendo de la fuente desde la que se recibe el flujo de bits 21 (tal como un reproductor de DVD, un reproductor de Blu-ray, un teléfono inteligente, un ordenador tipo tablet, un sistema de juegos y un televisor, para dar algunos ejemplos). Si bien se puede seleccionar cualquiera de los renderizadores de audio 22, a menudo el renderizador de audio utilizado al crear el contenido proporciona una mejor (y posiblemente la mejor) forma de renderización debido al hecho de que el contenido fue creado por el creador del contenido 12 usando este uno de los renderizadores de audio, es decir, el renderizador de audio 5 en el ejemplo de la FIG. 3. La selección de uno de los renderizadores de audio 22 que es el mismo o al menos cercano (en términos de la forma de renderización) puede proporcionar una mejor representación del campo de sonido y puede dar como resultado una mejor experiencia de sonido envolvente para el consumidor de contenido 14.

**[0027]** De acuerdo con las técnicas descritas en esta divulgación, el dispositivo de codificación de audio 20 puede generar el flujo de bits 21 para incluir la información de renderización de audio 2 ("información de renderizador 2"). La información de renderización de audio 2 puede incluir un valor de la señal que identifica un renderizador de audio utilizado cuando se genera el contenido de audio multicanal, es decir, el renderizador de audio 1 en el ejemplo de la FIG. 3. En algunos casos, el valor de la señal incluye una matriz utilizada para renderizar los coeficientes armónicos esféricos en una pluralidad de alimentaciones de altavoces.

**[0028]** En algunos casos, el valor de la señal incluye dos o más bits que definen un índice que indica que el flujo de bits incluye una matriz utilizada para renderizar coeficientes armónicos esféricos en una pluralidad de alimentaciones de altavoces. En algunos casos, cuando se usa un índice, el valor de la señal incluye además dos o más bits que definen un número de filas de la matriz incluida en el flujo de bits y dos o más bits que definen un número de columnas de la matriz incluidas en el flujo de bits. Usando esta información y dado que cada coeficiente de la matriz bidimensional se define típicamente por un número de punto flotante de 32 bits, el tamaño en términos de bits de la matriz se puede calcular como una función del número de filas, el número de columnas, y el tamaño de los números de punto flotante que definen cada coeficiente de la matriz, es decir, 32 bits en este ejemplo.

**[0029]** En algunos casos, el valor de la señal especifica un algoritmo de renderización utilizado para renderizar coeficientes armónicos esféricos en una pluralidad de alimentaciones de altavoces. El algoritmo de renderización puede incluir una matriz que es conocida tanto por el dispositivo de codificación de audio 20 como por el dispositivo de descodificación 24. Es decir, el algoritmo de renderización puede incluir la aplicación de una matriz además de otros pasos de renderización, tales como la panoramización (por ejemplo, VBAP, DBAP o panoramización simple) o el filtrado NFC. En algunos casos, el valor de la señal incluye dos o más bits que definen un índice asociado con una de una pluralidad de matrices usadas para renderizar coeficientes armónicos esféricos en una pluralidad de alimentaciones de altavoces. De nuevo, tanto el dispositivo de codificación de audio 20 como el dispositivo de descodificación 24 pueden configurarse con información que indica la pluralidad de matrices y el orden de la pluralidad de matrices de modo que el índice puede identificar de manera única a una particular de la pluralidad de matrices. De

forma alternativa, el dispositivo de codificación de audio 20 puede especificar datos en el flujo de bits 21 que define la pluralidad de matrices y/o el orden de la pluralidad de matrices de modo que el índice pueda identificar de manera única a una particular de la pluralidad de matrices.

**[0030]** En algunos casos, el valor de la señal incluye dos o más bits que definen un índice asociado con uno de una pluralidad de algoritmos de renderización usados para renderizar coeficientes armónicos esféricos en una pluralidad de alimentaciones de altavoces. De nuevo, tanto el dispositivo de codificación de audio 20 como el dispositivo de descodificación 24 pueden configurarse con información que indica la pluralidad de algoritmos de renderización y el orden de la pluralidad de algoritmos de renderización de manera que el índice puede identificar de manera única a una particular de la pluralidad de matrices. De forma alternativa, el dispositivo de codificación de audio 20 puede especificar datos en el flujo de bits 21 que define la pluralidad de matrices y/o el orden de la pluralidad de matrices de modo que el índice pueda identificar de manera única a una particular de la pluralidad de matrices.

**[0031]** En algunos casos, el dispositivo de codificación de audio 20 especifica información de renderización de audio 2 basándose en cada trama de audio en el flujo de bits. En otros casos, el dispositivo de codificación de audio 20 especifica la información de renderización de audio 2 una sola vez en el flujo de bits.

**[0032]** El dispositivo de descodificación 24 puede entonces determinar la información de renderización de audio 2 especificada en el flujo de bits. Basándose en el valor de la señal incluido en la información de renderización de audio 2, el sistema de reproducción de audio 16 puede renderizar una pluralidad de alimentaciones de altavoces 25 basándose en la información de renderización de audio 2. Como se indicó anteriormente, el valor de la señal puede, en algunos casos, incluir una matriz utilizada para renderizar coeficientes armónicos esféricos en una pluralidad de alimentaciones de altavoces. En este caso, el sistema de reproducción de audio 16 puede configurar uno de los renderizadores de audio 22 con la matriz, usando este uno de los renderizadores de audio 22 para renderizar las alimentaciones de altavoces 25 basándose en la matriz.

**[0033]** En algunos casos, el valor de la señal incluye dos o más bits que definen un índice que indica que el flujo de bits incluye una matriz utilizada para renderizar los coeficientes HOA 11' en las alimentaciones de altavoces 25. El dispositivo de descodificación 24 puede analizar la matriz desde el flujo de bits en respuesta al índice, después de lo cual el sistema de reproducción de audio 16 puede configurar uno de los renderizadores de audio 22 con la matriz analizada e invocar a este uno de los renderizadores 22 para renderizar las alimentaciones de altavoces 25. Cuando el valor de la señal incluye dos o más bits que definen un número de filas de la matriz incluida en el flujo de bits y dos o más bits que definen un número de columnas de la matriz incluidas en el flujo de bits, el dispositivo de descodificación 24 puede analizar la matriz del flujo de bits en respuesta al índice y basado en los dos o más bits que definen un número de filas y los dos o más bits que definen el número de columnas de la manera descrita anteriormente.

**[0034]** En algunos casos, el valor de la señal especifica un algoritmo de renderización utilizado para renderizar los coeficientes HOA 11' en las alimentaciones de altavoces 25. En estos casos, algunos o todos los renderizadores de audio 22 pueden realizar estos algoritmos de renderización. El dispositivo de reproducción de audio 16 puede utilizar entonces el algoritmo de renderización especificado, por ejemplo, uno de los renderizadores de audio 22, para renderizar las alimentaciones de altavoces 25 desde los coeficientes HOA 11'.

**[0035]** Cuando el valor de la señal incluye dos o más bits que definen un índice asociado con una de una pluralidad de matrices usadas para renderizar los coeficientes HOA 11' en las alimentaciones de altavoces 25, algunos o todos los renderizadores de audio 22 pueden representar esta pluralidad de matrices. Por lo tanto, el sistema de reproducción de audio 16 puede renderizar las alimentaciones de altavoz 25 desde los coeficientes HOA 11' usando uno de los renderizadores de audio 22 asociados con el índice.

**[0036]** Cuando el valor de la señal incluye dos o más bits que definen un índice asociado con uno de una pluralidad de algoritmos de renderización utilizados para renderizar los coeficientes HOA 11' en las alimentaciones de altavoces 25, algunos o todos los renderizadores de audio 34 pueden representar estos algoritmos de renderización. Por lo tanto, el sistema de reproducción de audio 16 puede renderizar las alimentaciones de altavoces 25 desde los coeficientes armónicos esféricos 11' usando uno de los renderizadores de audio 22 asociados con el índice.

**[0037]** Dependiendo de la frecuencia con la que se especifique esta información de renderización de audio en el flujo de bits, el dispositivo de descodificación 24 puede determinar la información de renderización de audio 2 basándose en cada trama de audio o a una sola vez.

**[0038]** Al especificar la información de renderización de audio 3 de esta manera, las técnicas pueden dar como resultado una mejor reproducción del contenido de audio multicanal y de acuerdo con la manera en que el creador de contenido 12 pretenda que se reproduzca el contenido de audio multicanal. Como resultado, las técnicas pueden proporcionar un sonido envolvente más inmersivo o una experiencia de audio multicanal.

**[0039]** En otras palabras, como se ha descrito anteriormente, los ambisónicos de orden superior (HOA) pueden representar una manera mediante la cual describir la información direccional de un campo sonoro basándose en una transformada espacial de Fourier. Típicamente, cuanto mayor es el orden de ambisónicos N, mayor es la resolución

espacial, mayor es el número de coeficientes armónicos esféricos (SH)  $(N+1)^2$ , y mayor es el ancho de banda necesario para transmitir y almacenar los datos.

**[0040]** Una posible ventaja de esta descripción es la posibilidad de reproducir este campo de sonido en la mayoría de las configuraciones de altavoces (por ejemplo, 5.1.7.1 22.2, etc.). La conversión de la descripción del campo de sonido en señales del altavoz M se puede hacer a través de una matriz de renderización estática con  $(N+1)^2$  entradas y M salidas. En consecuencia, cada configuración de altavoz puede requerir una matriz de renderización dedicada. Pueden existir varios algoritmos para calcular la matriz de renderización para una configuración de altavoces deseada, que puede optimizarse para ciertas medidas objetivas o subjetivas, como los criterios de Gerzon. Para configuraciones de altavoces irregulares, los algoritmos pueden volverse complejos debido a los procedimientos iterativos de optimización numérica, como la optimización convexa. Para calcular una matriz de renderización para diseños de altavoces irregulares sin tiempo de espera, puede ser beneficioso disponer de suficientes recursos de cálculo. Las configuraciones de altavoces irregulares pueden ser comunes en entornos de sala de estar domésticos debido a restricciones arquitectónicas y preferencias estéticas. Por lo tanto, para la mejor reproducción del campo de sonido, se puede preferir una matriz de renderización optimizada para tal escenario, ya que puede permitir la reproducción del campo de sonido con mayor precisión.

**[0041]** Dado que un descodificador de audio en general no requiere muchos recursos computacionales, es posible que el dispositivo no pueda calcular una matriz de renderización irregular en un momento favorable para el consumidor. Diversos aspectos de las técnicas descritas en esta divulgación pueden proporcionar el uso de un enfoque informático basado en la nube de la siguiente manera:

1. El descodificador de audio puede enviar a través de una conexión a Internet las coordenadas del altavoz (y, en algunos casos, también las mediciones de SPL obtenidas con un micrófono de calibración) a un servidor;
2. El servidor basado en la nube puede calcular la matriz de renderización (y posiblemente algunas versiones diferentes, de modo que el cliente pueda luego elegir entre estas versiones diferentes); y
3. El servidor puede entonces enviar la matriz de renderización (o las diferentes versiones) de vuelta al descodificador de audio a través de la conexión a Internet.

**[0042]** Este enfoque puede permitirle al fabricante mantener bajos los costes de fabricación de un descodificador de audio (debido a que puede no ser necesario un procesador potente para calcular estas matrices de renderización irregular), al mismo tiempo que facilita una reproducción de audio más óptima en comparación con matrices de renderización en general diseñadas para configuraciones o geometrías de altavoces regulares. El algoritmo para calcular la matriz de renderización también se puede optimizar después de que se haya enviado un descodificador de audio, lo cual puede reducir los costes de las revisiones de hardware o incluso recuperarlos. Las técnicas también pueden, en algunos casos, recopilar mucha información sobre diferentes configuraciones de altavoces de productos de consumidor que pueden ser beneficiosas para futuros desarrollos de productos.

**[0043]** En algunos casos, el sistema mostrado en la FIG. 3 puede que no señale la información de renderización de audio 2 en el flujo de bits 21 como se describió anteriormente, sino que señale esta información de renderización de audio 2 como metadatos separados del flujo de bits 21. De forma alternativa o junto con lo descrito anteriormente, el sistema mostrado en la FIG. 3 puede señalar una parte de la información de renderización de audio 2 en el flujo de bits 21 como se describió anteriormente y señalar una parte de esta información de renderización de audio 3 como metadatos separados del flujo de bits 21. En algunos ejemplos, el dispositivo de codificación de audio 20 puede emitir estos metadatos, que luego pueden cargarse en un servidor u otro dispositivo. El dispositivo de descodificación de audio 24 puede luego descargar o recuperar de otro modo estos metadatos, que luego se usan para aumentar la información de renderización de audio extraída del flujo de bits 21 por el dispositivo de descodificación de audio 24. El flujo de bits 21 formado de acuerdo con los aspectos de información de renderización de las técnicas se describe a continuación con respecto a los ejemplos de las FIGS. 8A-8D.

**[0044]** La FIG. 3 es un diagrama de bloques que ilustra, con mayor detalle, un ejemplo del dispositivo de codificación de audio 20 mostrado en el ejemplo de la FIG. 2 que puede realizar diversos aspectos de las técnicas descritas en esta divulgación. El dispositivo de codificación de audio 20 incluye una unidad de análisis de contenido 26, una unidad de descomposición basada en vectores 27 y una unidad de descomposición basada en la dirección 28. Aunque se describe brevemente a continuación, en la Publicación de Patente Internacional n.º WO 2014/194099, titulada "INTERPOLATION FOR DECOMPOSED REPRESENTATIONS OF A SOUND FIELD" ["INTERPOLACIÓN PARA REPRESENTACIONES DESCOMPUESTAS DE UN CAMPO DE SONIDO"], presentada el 29 de mayo de 2014, está disponible más información sobre el dispositivo de codificación de audio 20 y los diversos aspectos de compresión o codificación de otro modo de los coeficientes HOA.

**[0045]** La unidad de análisis de contenido 26 representa una unidad configurada para analizar el contenido de los coeficientes HOA 11 para identificar si los coeficientes HOA 11 representan contenido generado a partir de una grabación en directo o un objeto de audio. La unidad de análisis de contenido 26 puede determinar si los coeficientes HOA 11 se generaron a partir de una grabación de un campo sonoro real o de un objeto de audio artificial. En algunos

casos, cuando los coeficientes HOA en tramas 11 se generaron a partir de una grabación, la unidad de análisis de contenido 26 pasa los coeficientes HOA 11 a la unidad de descomposición basada en vectores 27. En algunos casos, cuando los coeficientes HOA por trama 11 se generaron a partir de un objeto de audio sintético, la unidad de análisis de contenido 26 pasa los coeficientes HOA 11 a la unidad de síntesis basada en la dirección 28. La unidad de síntesis basada en la dirección 28 puede representar una unidad configurada para realizar una síntesis basada en la dirección de los coeficientes HOA 11 para generar un flujo de bits basado en la direccionalidad 21.

**[0046]** Como se muestra en el ejemplo de la FIG. 3, la unidad de descomposición basada en vectores 27 puede incluir una unidad de transformada invertible lineal (LIT) 30, una unidad de cálculo de parámetros 32, una unidad de reordenamiento 34, una unidad de selección del primer plano 36, una unidad de compensación de energía 38, una unidad codificadora de audio psicoacústico 40, una unidad de generación de flujo de bits 42, una unidad de análisis del campo sonoro 44, una unidad de reducción de coeficientes 46, una unidad de selección de fondo (BG) 48, una unidad de interpolación espacio-temporal 50 y una unidad de cuantificación 52.

**[0047]** La unidad de transformada invertible lineal (LIT) 30 recibe los coeficientes HOA 11 en forma de canales HOA, siendo cada canal representativo de un bloque o trama de un coeficiente asociado a un orden dado, suborden de las funciones de base esférica (que se puede indicar como  $HOA[k]$ , donde  $k$  puede indicar la trama o bloque actual de muestras). La matriz de coeficientes HOA 11 puede tener las dimensiones  $D: M \times (N+1)^2$ .

**[0048]** La unidad de LIT 30 puede representar una unidad configurada para realizar una forma de análisis denominada descomposición en valores singulares. Aunque se describen con respecto a la SVD, las técnicas descritas en esta divulgación pueden realizarse con respecto a cualquier transformación o descomposición similar que proporcione conjuntos de salida de energía compactada y no correlacionada linealmente. También, la referencia a "conjuntos" en esta divulgación está en general concebida para referirse a conjuntos no nulos, a menos que se indique específicamente lo contrario y no se pretenda referirse a la definición matemática clásica de conjuntos que incluye el denominado "conjunto vacío". Una transformación alternativa puede comprender un análisis de componentes principales, que se denomina a menudo "PCA". Dependiendo del contexto, el PCA puede ser mencionado con varios nombres diferentes, tales como la transformada discreta de Karhunen-Loeve, la transformada de Hotelling, la descomposición ortogonal adecuada (POD) y la descomposición en autovalores (EVD), por nombrar algunos ejemplos. Las propiedades de dichas operaciones que conducen a la meta subyacente de compresión de datos de audio son la "compactación energética" y la "des correlación" de los datos de audio multicanal.

**[0049]** En cualquier caso, suponiendo que la unidad de LIT 30 realiza una descomposición en valores singulares (que, de nuevo, puede denominarse "SVD") con fines de ejemplo, la unidad de LIT 30 puede transformar los coeficientes HOA 11 en dos o más conjuntos de coeficientes HOA transformados. Estos "conjuntos" de coeficientes HOA transformados pueden incluir vectores de coeficientes HOA transformados. En el ejemplo de la FIG. 3, la unidad de LIT 30 puede realizar la SVD con respecto a los coeficientes HOA 11 para generar una matriz llamada  $V$ , una matriz  $S$  y una matriz  $U$ . La SVD, en álgebra lineal, puede representar una factorización de una matriz  $X$  real o compleja de  $y$  por  $z$  (donde  $X$  puede representar datos de audio multicanal, tales como los coeficientes HOA 11) de la siguiente forma:

$$X = USV^*$$

$U$  puede representar una matriz unitaria real o compleja de  $y$  por  $y$ , donde las columnas  $y$  de  $U$  se conocen como los vectores singulares izquierdos de los datos de audio multicanal.  $S$  puede representar una matriz diagonal rectangular de  $y$  por  $z$  con números reales no negativos en la diagonal, donde los valores diagonales de  $S$  se conocen comúnmente como los valores singulares de los datos de audio multicanal.  $V^*$  (que puede indicar una traspuesta conjugada de  $V$ ) puede representar una matriz unitaria real o compleja de  $z$  por  $z$ , donde las columnas  $z$  de  $V^*$  son conocidas como vectores singulares derechos de los datos de audio multicanal.

**[0050]** En algunos ejemplos, la matriz  $V^*$  en la expresión matemática de SVD mencionada anteriormente se indica como la traspuesta conjugada de la matriz  $V$  para reflejar que la SVD se puede aplicar a matrices que comprenden números complejos. Cuando se aplica a matrices que comprenden solo números reales, la conjugada compleja de la matriz  $V$  (o, en otras palabras, la matriz  $V^*$ ) puede considerarse como la traspuesta de la matriz  $V$ . A continuación se supone, con fines de facilitar la ilustración, que los coeficientes HOA 11 comprenden números reales, con el resultado de que la matriz  $V$  se emite mediante la SVD, en lugar de la matriz  $V^*$ . Además, aunque se indica como la matriz  $V$  en esta divulgación, la referencia a la matriz  $V$  debería entenderse como referencia a la traspuesta de la matriz  $V$  cuando corresponda. Aunque se supone que es la matriz  $V$ , las técnicas pueden aplicarse de manera similar a los coeficientes HOA 11 que tienen coeficientes complejos, donde la salida de la SVD es la matriz  $V^*$ . Por consiguiente, las técnicas no deberían estar limitadas a este respecto para proporcionar solamente la aplicación de la SVD para generar una matriz  $V$ , sino que pueden incluir la aplicación de la SVD a coeficientes HOA 11 que tengan componentes complejos para generar una matriz  $V^*$ .

**[0051]** De esta manera, la unidad de LIT 30 puede realizar una SVD con respecto a los coeficientes HOA 11 para emitir vectores  $US[k]$  33 (que pueden representar una versión combinada de los vectores  $S$  y los vectores  $U$ ) que tienen dimensiones  $D$ : Vectores  $M \times (N+1)^2$  y  $V[k]$  35 con dimensiones  $D$ :  $(N+1)^2 \times (N+1)^2$ . Los elementos vectoriales

individuales en la matriz  $US[k]$  también pueden denominarse  $X_{PS}(k)$ , mientras que los vectores individuales de la matriz  $V[k]$  también pueden denominarse  $v(k)$ .

**[0052]** Un análisis de las matrices  $U$ ,  $S$  y  $V$  puede revelar que las matrices llevan o representan características espaciales y temporales del campo sonoro subyacente representado anteriormente por  $X$ . Cada uno de los  $N$  vectores en  $U$  (de muestras de longitud  $M$ ) puede representar señales de audio normalizadas separadas como una función del tiempo (para el período de tiempo representado por  $M$  muestras), que son ortogonales entre sí y que han sido desacopladas de cualquier característica espacial (que también puede denominarse información direccional). Las características espaciales, que representan la forma espacial y la posición ( $r$ ,  $\theta$ ,  $\phi$ ), pueden representarse en cambio por vectores  $r^{\text{ésimos}}$  individuales,  $v^{(i)}(k)$ , en la matriz  $V$  (cada uno de longitud  $(N+1)^2$ ). Los elementos individuales de cada uno de los vectores  $v^{(i)}(k)$  pueden representar un coeficiente HOA que describe la forma (incluida la anchura) y la posición del campo de sonido para un objeto de audio asociado. Tanto los vectores de la matriz  $U$  como los de la matriz  $V$  se normalizan de tal manera que sus energías de raíces de cuadrados medios son iguales a la unidad. La energía de las señales de audio en  $U$  está de este modo representada por los elementos diagonales en  $S$ . Multiplicar  $U$  por  $S$  para formar  $US[k]$  (con los elementos vectoriales individuales  $X_{PS}(k)$ ), representa de este modo la señal de audio con energías. La capacidad de la descomposición SVD para desacoplar las señales de tiempo de audio (en  $U$ ), sus energías (en  $S$ ) y sus características espaciales (en  $V$ ) puede dar soporte a varios aspectos de las técnicas descritas en esta divulgación. Además, el modelo de síntesis de los coeficientes HOA[k] subyacentes,  $X$ , por una multiplicación de vectores de  $US[k]$  y  $V[k]$  da lugar al término "descomposición basada en vectores" que se utiliza en todo este documento.

**[0053]** Aunque se describe como realizada directamente con respecto a los coeficientes HOA 11, la unidad de LIT 30 puede aplicar la transformada invertible lineal a los obtenidos de los coeficientes HOA 11. Por ejemplo, la unidad de LIT 30 puede aplicar una SVD con respecto a una matriz de densidad espectral de potencia obtenida a partir de los coeficientes HOA 11. Mediante la realización de la SVD con respecto a la densidad espectral de potencia (PSD) de los coeficientes HOA, en lugar de los coeficientes en sí, la unidad de LIT 30 puede reducir potencialmente la complejidad de cálculo para realizar la SVD en términos de uno o más ciclos de procesador y espacio de almacenamiento, mientras se consigue la misma eficacia en la codificación de la fuente de audio como si la SVD se aplicara directamente a los coeficientes HOA.

**[0054]** La unidad de cálculo de parámetros 32 representa una unidad configurada para calcular varios parámetros, tales como un parámetro de correlación ( $R$ ), parámetros de propiedades direccionales ( $\theta$ ,  $\phi$ ,  $r$ ) y una propiedad de energía ( $e$ ). Cada uno de los parámetros para la trama actual puede indicarse como  $R[k]$ ,  $\theta[k]$ ,  $\phi[k]$ ,  $r[k]$  y  $e[k]$ . La unidad de cálculo de parámetros 32 puede realizar un análisis de energía y/o correlación (o lo que se denomina correlación cruzada) con respecto a los vectores  $US[k]$  33 para identificar los parámetros. La unidad de cálculo de parámetros 32 también puede determinar los parámetros para la trama anterior, donde los parámetros de tramas anteriores pueden indicarse como  $R[k-1]$ ,  $\theta[k-1]$ ,  $\phi[k-1]$ ,  $r[k-1]$  y  $e[k-1]$ , basándose en la trama anterior del vector  $US[k-1]$  y los vectores  $V[k-1]$ . La unidad de cálculo de parámetros 32 puede emitir los parámetros actuales 37 y los parámetros anteriores 39 para reordenar la unidad 34.

**[0055]** Los parámetros calculados por la unidad de cálculo de parámetros 32 pueden ser utilizados por la unidad de reordenamiento 34 para reordenar los objetos de audio para representar su evaluación natural o su continuidad en el tiempo. La unidad de reordenamiento 34 puede comparar cada uno de los parámetros 37 de los primeros vectores  $US[k]$  33, por orden, con cada uno de los parámetros 39 para los segundos vectores  $US[k-1]$  33. La unidad de reordenamiento 34 puede reordenar (utilizando como ejemplo un algoritmo húngaro) los diversos vectores dentro de la matriz  $US[k]$  33 y la matriz  $V[k]$  35 basándose en los parámetros actuales 37 y los parámetros anteriores 39 para emitir una matriz  $US[k]$  reordenada 33' (que se puede indicar matemáticamente como  $US[k]$ ) y una matriz  $V[k]$  reordenada 35' (que se puede indicar matemáticamente como  $V[k]$ ) a una unidad de selección de sonido de primer plano 36 (o sonido predominante - PS) ("unidad de selección del primer plano 36") y una unidad de compensación de energía 38.

**[0056]** La unidad de análisis del campo sonoro 44 puede representar una unidad configurada para realizar un análisis de campo sonoro con respecto a los coeficientes HOA 11 para alcanzar potencialmente una velocidad de bits deseada 41. La unidad de análisis del campo sonoro 44 puede, basándose en este análisis y/o en una velocidad de bits deseada recibida 41, determinar el número total de instanciaciones del codificador psicoacústico (que puede ser una función del número total de canales ambientales o de fondo ( $BG_{TOT}$ ) y el número de canales de primer plano o, en otras palabras, canales predominantes. El número total de instanciaciones del codificador psicoacústico se puede indicar como  $numHOATransportChannels$ .

**[0057]** La unidad de análisis del campo sonoro 44 también puede determinar, de nuevo para alcanzar potencialmente la velocidad de bits objetivo 41, el número total de canales de primer plano ( $nFG$ ) 45, el orden mínimo del campo sonoro de fondo (o, en otras palabras, ambiente) ( $N_{BG}$  o, de forma alternativa,  $MinAmbHOAorder$ ), el número correspondiente de canales reales representativos del orden mínimo del campo sonoro de fondo ( $nBGa = (MinAmbHOAorder + 1)^2$ ), y los índices ( $i$ ) de los canales adicionales de BG HOA a enviar (que pueden indicarse colectivamente como información de canal de fondo 43 en el ejemplo de la FIG. 3). La información de canal de fondo 42 también puede denominarse información de canal ambiental 43. Cada uno de los canales que queda de

numHOATransportChannels- nBGa puede ser un "canal de fondo/ambiental adicional", un "canal predominante basado en un vector activo", una "señal predominante basada en dirección activa" o estar "completamente inactivo". En un aspecto, los tipos de canales se pueden indicar como un elemento sintáctico ("TipoCanal") con dos bits (por ejemplo, 00: señal basada en dirección; 01: señal predominante basada en vectores; 10: señal ambiental adicional; 11: señal inactiva). El número total de señales de fondo o de ambiente, nBGa, puede estar dado por  $(\text{MinOrdenHoaAmb} + 1)^2$  + el número de veces que el índice 10 (en el ejemplo anterior) aparece como un tipo de canal en el flujo de bits para esa trama.

**[0058]** La unidad de análisis del campo sonoro 44 puede seleccionar el número de canales de fondo (o, en otras palabras, ambientales) y el número de canales de primer plano (o, en otras palabras, predominantes) basándose en la velocidad de bits deseada 41, seleccionando más canales de fondo y/o de primer plano cuando la velocidad de bits deseada 41 es relativamente mayor (por ejemplo, cuando la velocidad de bits deseada 41 es igual o mayor que 512 Kbps). En un aspecto, el numHOATransportChannels se puede fijar en 8 mientras que el MinOrdenHoaAmb puede ser fijado en 1 en la sección de cabecera del flujo de bits. En este escenario, en cada trama se pueden dedicar cuatro canales para representar la parte de fondo o ambiental del campo sonoro, mientras que los otros 4 canales pueden variar, de trama a trama, según el tipo de canal, por ejemplo, usado como un canal de fondo/ambiental adicional o como un canal de primer plano/predominante. Las señales de primer plano/predominantes pueden ser una entre señales basadas en vectores o en la dirección, como se ha descrito anteriormente.

**[0059]** En algunos casos, el número total de señales predominantes basadas en vectores para una trama, puede estar dado por el número de veces que el índice TipoCanal es 01 en el flujo de bits de dicha trama. En el aspecto anterior, para cada canal de fondo/ambiental adicional (por ejemplo, correspondiente a un TipoCanal de 10), información correspondiente de cuál de los posibles coeficientes HOA (más allá de los cuatro primeros) se puede representar en ese canal. La información, para contenido HOA de cuarto orden, puede ser un índice para indicar los coeficientes HOA 5 a 25. Los cuatro primeros coeficientes HOA ambientales 1 a 4 pueden enviarse todo el tiempo cuando minOrdenHOAAmb se fija en 1, por lo tanto, el dispositivo de codificación de audio solo puede necesitar indicar uno de los coeficientes HOA ambientales adicionales que tengan un índice de 5 a 25. Esta información, por tanto, podría ser enviada utilizando un elemento sintáctico de 5 bits (para el contenido de 4.º orden), que puede denominarse "CodedAmbCoeffIdx". En cualquier caso, la unidad de análisis del campo sonoro 44 emite la información de canal de fondo 43 y los coeficientes HOA 11 a la unidad de selección de fondo (BG) 36, la información de canal de fondo 43 a la unidad de reducción de coeficientes 46 y a la unidad de generación del flujo de bits 42, y el nFG 45 a una unidad de selección de primer plano 36.

**[0060]** La unidad de selección de fondo 48 puede representar una unidad configurada para determinar los coeficientes HOA de fondo o ambientales 47 basándose en la información de canal de fondo (por ejemplo, el campo sonoro de fondo (NBG) y el número (nBGa) y los índices (i) de los canales HOA de BG adicionales a enviar). Por ejemplo, cuando N<sub>BG</sub> es igual a uno, la unidad de selección de fondo 48 puede seleccionar los coeficientes HOA 11 para cada muestra de la trama de audio que tenga un orden igual o inferior a uno. La unidad de selección de fondo 48 puede, en este ejemplo, seleccionar entonces los coeficientes HOA 11 que tengan un índice identificado por uno de los índices (i) como coeficientes adicionales HOA de BG, donde el nBGa se proporciona a la unidad de generación del flujo de bits 42 para que se especifique en el flujo de bits 21 para permitir al dispositivo de decodificación de audio, tal como el dispositivo de decodificación de audio 24, que se muestra en el ejemplo de las FIGs. 2 y 4, analizar los coeficientes HOA de fondo 47 del flujo de bits 21. La unidad de selección de fondo 48 puede emitir entonces los coeficientes HOA ambientales 47 a la unidad de compensación de energía 38. Los coeficientes HOA ambientales 47 pueden tener dimensiones D:  $M \times [(N_{BG}+1)^2 + nBGa]$ . Los coeficientes HOA ambientales 47 también pueden denominarse "coeficientes HOA ambientales 47", donde cada uno de los coeficientes HOA ambientales 47 corresponde a un canal HOA ambiental independiente 47, a codificar por la unidad codificadora de audio psicoacústica 40.

**[0061]** La unidad de selección del primer plano 36 puede representar una unidad configurada para seleccionar la matriz US[k] reordenada 33' y la matriz V[k] reordenada 35', que representan componentes de primer plano o distintos, del campo sonoro basándose en nFG 45 (que pueden representar uno o más índices que identifican los vectores de primer plano). La unidad de selección de primer plano 36 puede emitir señales 49 de nFG (que pueden denotarse como  $US[k]_i$  reordenado, ..., nFG 49,  $FG_1, nFG[k] 49$ , o  $X_{PS}^{(1..nFG)}(k) 49$ ) a la unidad codificadora de audio psicoacústica 40, donde las señales 49 de nFG pueden tener dimensiones D:  $M \times nFG$  y cada una representa objetos de mono-audio. La unidad de selección del primer plano 36 también pueden emitir la matriz V[k] reordenada 35' (o  $\sqrt{1..nFG}(k)$  35'), correspondiente a los componentes de primer plano del campo sonoro, a la unidad de interpolación espacio-temporal 50, donde un subconjunto de la matriz V[k] reordenada 35', correspondiente a los componentes de primer plano, se puede indicar como la matriz V[k] de primer plano  $51_k$  (que se puede indicar matemáticamente como  $\bar{V}_1, \dots, nFG[k]$ ) que tiene dimensiones D:  $(N+1)^2 \times nFG$ .

**[0062]** La unidad de compensación de energía 38 puede representar una unidad configurada para realizar compensación de energía con respecto a los coeficientes HOA ambientales 47, para compensar la pérdida de energía debida a la eliminación de varios de los canales HOA por la unidad de selección de fondo 48. La unidad de compensación de energía 38 puede realizar un análisis energético con respecto a una o más entre la matriz

reordenada US[k] 33', la matriz V[k] reordenada 35', las nFG señales 49, los vectores V[k] 51<sub>k</sub> de primer plano y los coeficientes HOA ambientales 47, y luego realizar una compensación de energía basándose en el análisis de energía para generar los coeficientes HOA ambientales compensados en energía 47'. La unidad de compensación de energía 38 puede emitir los coeficientes HOA ambientales compensados en energía 47' a la unidad codificadora de audio psicoacústica 40.

**[0063]** La unidad de interpolación espaciotemporal 50 puede representar una unidad configurada para recibir los vectores V[k] de primer plano 51<sub>k</sub> para la k-ésima trama y los vectores de primer plano V[k-1] 51<sub>k-1</sub> para la trama anterior (de ahí la notación k-1) y realizar una interpolación espaciotemporal para generar los vectores V[k] de primer plano interpolados. La unidad de interpolación espaciotemporal 50 puede recombinar las nFG señales 49 con los vectores V[k] de primer plano 51<sub>k</sub> para recuperar los coeficientes HOA de primer plano reordenados. La unidad de interpolación espaciotemporal 50 puede entonces dividir los coeficientes HOA de primer plano reordenados por los vectores V[k] interpolados para generar nFG señales interpoladas 49'. La unidad de interpolación espaciotemporal 50 también puede emitir los vectores V[k] de primer plano 51<sub>k</sub> que se usaron para generar los vectores V[k] de primer plano interpolados, de modo que un dispositivo de descodificación de audio, tal como el dispositivo de descodificación de audio 24, pueda generar los vectores V[k] de primer plano interpolados y recuperar así los vectores V[k] de primer plano 51<sub>k</sub>. Los vectores V[k] de primer plano 51<sub>k</sub> utilizados para generar los vectores V[k] de primer plano interpolados se indican como los vectores V[k] de primer plano restantes 53. Con el fin de asegurar que los mismos V[k] y V[k-1] se usen en el codificador y el descodificador (para crear los vectores interpolados V[k]) se pueden usar versiones cuantificadas/descuantificadas de los vectores en el codificador y el descodificador. La unidad de interpolación espaciotemporal 50 puede emitir las señales nFG interpoladas 49' a la unidad codificadora de audio psicoacústica 46 y los vectores V[k] de primer plano interpolados 51<sub>k</sub> a la unidad de reducción de coeficientes 46.

**[0064]** La unidad de reducción de coeficientes 46 puede representar una unidad configurada para realizar una reducción de coeficientes con respecto a los restantes vectores V[k] de primer plano 53, basándose en la información de canal de fondo 43 para emitir los vectores de primer plano reducidos V[k] 55 a la unidad de cuantificación 52. Los vectores V[k] de primer plano reducidos 55 pueden tener dimensiones D:  $[(N+1)^2 - (N_{BG}+1)^2 - BG_{TOT}] \times nFG$ . La unidad de reducción de coeficientes 46 puede, a este respecto, representar una unidad configurada para reducir el número de coeficientes de los restantes vectores V[k] de primer plano 53. En otras palabras, la unidad de reducción de coeficientes 46 puede representar una unidad configurada para eliminar los coeficientes en los vectores V[k] de primer plano (que forman los restantes vectores V[k] de primer plano 53) que tienen poca o ninguna información direccional. En algunos ejemplos, los coeficientes de los vectores V[k] distintos o, en otras palabras, de primer plano que corresponden a funciones de base de primer orden y de orden cero (que pueden indicarse como N<sub>BG</sub>) proporcionan poca información direccional y por lo tanto pueden ser retirados de los V vectores de primer plano (mediante un proceso que se puede denominar "reducción de coeficientes"). En este ejemplo, se puede proporcionar mayor flexibilidad para identificar no solo los coeficientes que corresponden a N<sub>BG</sub>, sino para identificar canales HOA adicionales (que pueden indicarse con la variable TotalOfAddAmbHOAChan) a partir del conjunto de  $[(N_{BG}+1)^2+1, (N+1)^2]$ .

**[0065]** La unidad de cuantificación 52 puede representar una unidad configurada para realizar cualquier forma de cuantificación para comprimir los vectores V[k] de primer plano reducidos 55 y generar vectores de primer plano V[k] codificados 57, emitiendo los vectores de primer plano codificados V[k] 57 a la unidad de generación del flujo de bits 42. En funcionamiento, la unidad de cuantificación 52 puede representar una unidad configurada para comprimir un componente espacial del campo sonoro, es decir, uno o más de los vectores V[k] de primer plano reducidos 55 en este ejemplo. La unidad de cuantificación 52 puede realizar cualquiera de los siguientes 12 modos de cuantificación, como se indica mediante un elemento sintáctico de modo de cuantificación denominado "NbitsQ":

Valor de NbitsQ	Tipo de modo de cuantificación
0-3:	Reservado
4:	Cuantificación de vector
5:	Cuantificación escalar sin codificación Huffman
6:	Cuantificación escalar de 6 bits con codificación Huffman
7:	Cuantificación escalar de 7 bits con codificación Huffman
8:	Cuantificación escalar de 8 bits con codificación Huffman
...	...
16:	Cuantificación escalar de 16 bits con codificación Huffman

La unidad de cuantificación 52 también puede realizar versiones predichas de cualquiera de los tipos anteriores de modos de cuantificación, donde se determina una diferencia entre un elemento de (o una ponderación cuando se realiza la cuantificación del vector) del vector V de una trama previa y el elemento (o ponderación cuando se realiza la cuantificación del vector) del vector V de una trama actual. La unidad de cuantificación 52 puede entonces cuantificar



la diferencia entre los elementos o ponderaciones de la trama actual y de la trama anterior en lugar del valor del elemento del vector V de la propia trama actual.

[0066] La unidad de cuantificación 52 puede realizar múltiples formas de cuantificación con respecto a cada uno de los vectores  $V[k]$  de primer plano reducidos 55 para obtener múltiples versiones codificadas de los vectores  $V[k]$  de primer plano reducidos 55. La unidad de cuantificación 52 puede seleccionar una de las versiones codificadas de los vectores  $V[k]$  de primer plano reducidos 55 como el vector  $V[k]$  de primer plano codificado 57. La unidad de cuantificación 52 puede, en otras palabras, seleccionar uno del vector V cuantificado en vector no predicho, el vector V cuantificado en vector predicho, el vector V cuantificado escalar sin codificación Huffman y el vector V cuantificado escalar con codificación Huffman para usar como el vector V cuantificado conmutado de salida basado en cualquier combinación de los criterios analizados en esta divulgación. En algunos ejemplos, la unidad de cuantificación 52 puede seleccionar un modo de cuantificación de un conjunto de modos de cuantificación que incluye un modo de cuantificación vectorial y uno o más modos de cuantificación escalar, y cuantificar un vector V de entrada basándose en (o de acuerdo con) el modo seleccionado. La unidad de cuantificación 52 puede proporcionar entonces uno del vector V cuantificado por vector no predicho (por ejemplo, en términos de valores de ponderación o bits indicativos del mismo), el vector V cuantificado por vector predicho (por ejemplo, en términos de valores de error o bits indicativos del mismo), el vector V cuantificado escalar sin codificación Huffman y el vector V cuantificado escalar con codificación Huffman a la unidad de generación de flujo de bits 52 como los vectores  $V[k]$  de primer plano codificados 57. La unidad de cuantificación 52 también puede proporcionar los elementos sintácticos indicativos del modo de cuantificación (por ejemplo, el elemento sintáctico NbitsQ) y cualquier otro elemento sintáctico utilizado para descuantificar o reconstruir de otro modo el vector V.

[0067] La unidad codificadora de audio psicoacústica 40 incluida dentro del dispositivo de codificación de audio 20 puede representar múltiples instancias de un codificador de audio psicoacústico, cada una de las cuales se utiliza para codificar un objeto de audio diferente o canal HOA de cada uno de los coeficientes HOA ambientales compensados en energía 47' y las nFG señales interpoladas 49' para generar los coeficientes HOA ambientales codificados 59 y las nFG señales codificadas 61. La unidad codificadora de audio psicoacústico 40 puede emitir los coeficientes HOA ambientales codificados 59 y las nFG señales codificadas 61 a la unidad de generación de flujo de bits 42.

[0068] La unidad de generación del flujo de bits 42 incluida dentro del dispositivo de codificación de audio 20 representa una unidad que formatea datos para adaptarse a un formato conocido (que puede referirse a un formato conocido por un dispositivo de decodificación), generando de este modo el flujo de bits basado en vectores 21. El flujo de bits 21 puede, en otras palabras, representar datos de audio codificados, que se han codificado de la manera descrita anteriormente. La unidad de generación de flujo de bits 42 puede representar un multiplexor en algunos ejemplos, que pueden recibir los vectores codificados  $V[k]$  de primer plano 57, los coeficientes HOA codificados 59, las señales nFG codificadas 61 y la información de canal de fondo 43. La unidad de generación del flujo de bits 42 puede generar entonces un flujo de bits 21 basado en los vectores codificados  $V[k]$  de primer plano 57, los coeficientes HOA ambientales codificados 59, las nFG señales codificadas 61 y la información de canal de fondo 43. De esta forma, la unidad de generación de flujo de bits 42 puede especificar de ese modo los vectores 57 en el flujo de bits 21 para obtener el flujo de bits 21. El flujo de bits 21 puede incluir un flujo de bits primario o principal y uno o más flujos de bits de canal lateral.

[0069] Diversos aspectos de las técnicas también pueden permitir que la unidad de generación de flujo de bits 46, como se describió anteriormente, especifique información de renderización de audio 2 en el flujo de bits 21. Mientras que la versión actual del próximo borrador de trabajo de compresión de audio 3D proporciona señalización de matrices de mezcla descendente [downmix] específicas dentro del flujo de bits 21, el borrador de trabajo no proporciona la especificación de renderizadores utilizados en la renderización de coeficientes HOA 11 en el flujo de bits 21. Para el contenido HOA, el equivalente de dicha matriz de mezcla descendente es la matriz de renderización que convierte la representación de HOA en las alimentaciones deseadas del altavoz. Varios aspectos de las técnicas descritas en esta divulgación proponen armonizar adicionalmente los conjuntos de características de contenido de canal y HOA permitiendo que la unidad de generación de flujo de bits 46 señale matrices de renderización de HOA dentro del flujo de bits (como, por ejemplo, información de renderización de audio 2).

[0070] Una solución de señalización a modo de ejemplo basada en el esquema de codificación de matrices de mezcla descendente y optimizada para HOA se presenta a continuación. De forma similar a la transmisión de matrices de mezcla descendente, las matrices de renderización de HOA se pueden señalar dentro de *mpegh3daConfigExtension()*. Las técnicas pueden proporcionar un nuevo tipo de extensión ID\_CONFIG\_EXT\_HOA\_MATRIX como se establece en las siguientes tablas (con letra cursiva y negrita indicando cambios en la tabla existente).

Tabla - Sintaxis de mpeg3daConfigExtension() (Tabla 13 en CD)

Sintaxis	N.º de bits	Mnemónico
<pre> mpeg3daConfigExtension() {     numConfigExtensions = escapedValue(2,4,8) + 1;      for (confExtIdx=0; confExtIdx&lt;numConfigExtensions; confExtIdx++) {         usacConfigExtType[confExtIdx] = escapedValue(4,8,16);         usacConfigExtLength[confExtIdx] = escapedValue(4,8,16);          switch (usacConfigExtType[confExtIdx]) {             case ID_CONFIG_EXT_FILL:                 while (usacConfigExtLength[confExtIdx]--) {                     <b>fill_byte[i];</b> /* should be '10100101' */                 }                 break;             case ID_CONFIG_EXT_DMX_MATRIX:                 DownmixMatrixSet()                 break;             <u>case ID_CONFIG_EXT_HOA_MATRIX:</u>                 <u>HOARenderingMatrixSet()</u>                 <u>break;</u>             case ID_CONFIG_EXT_LOUDNESS_INFO:                 loudnessInfoSet ();                 break;             default:                 while (usacConfigExtLength[confExtIdx]--) {                     <b>tmp;</b>                 }                 break;         }     } } </pre>	8	<b>uimsbf</b>
	8	<b>uimsbf</b>

Tabla - Valor de usacConfigExtType (Tabla 1 en CD)

usacConfigExtType	Valor
ID_CONFIG_EXT_FILL	0
ID_CONFIG_EXT_DMX_MATRIX	1
ID_CONFIG_EXT_LOUDNESS_INFO	2
ID_CONFIG_EXT_HOA_MATRIX	3
/ * reservado para uso ISO * /	4-127
/ * reservado para su uso fuera del alcance ISO * /	128 y superior

**[0071]** El campo de bits *HOARenderingMatrixSet()* puede ser igual en estructura y funcionalidad en comparación con *Downmix-MatrixSet()*. En lugar de *inputCount(audioChannelLayout)*, *HOARenderingMatrixSet()* puede usar el valor "equivalente" *NumOfHoaCoeffs*, calculado en *HOAConfig*. Además, debido a que el orden de los coeficientes HOA puede fijarse dentro del descodificador HOA (ver, por ejemplo, el Anexo G en el CD), el *HOARenderingMatrixSet* no necesita ningún equivalente a *inputConfig(audioChannelLayout)*.

**Tabla 2 - Sintaxis de HOARenderingMatrixSet() (adoptada de la Tabla 15 en CD)**

Sintaxis	N.º de bits	Mnemónico
HOARenderingMatrixSet() { <b>numHOARenderingMatrices</b> for (k=0; k< numHOARenderingMatrices; ++k) { <b>downmixIdx;</b> <b>CICPspeakerLayoutIdx;</b> <b>DmxMatrixLenBits</b> = escapedValue(8,8,12); HOARenderingMatrix (NumOfHoaCoeffs, outputConfig(CICPspeakerLayoutIdx), outputCount(CICPspeakerLayoutIdx) ); } }	<b>5</b>  <b>6</b> <b>6</b> <b>8..28</b> <b>DmxMatrixL</b> <b>enBits</b>	<b>Uimsbf</b>  <b>Uimsbf</b> <b>Uimsbf</b>

**[0072]** Diversos aspectos de las técnicas también pueden permitir que la unidad de generación de flujo de bits 46, al comprimir los datos de audio HOA (por ejemplo, los coeficientes HOA 11 en el ejemplo de la FIG. 4) utilizando un primer esquema de compresión (como el esquema de compresión de descomposición representado por unidad de descomposición basada en vectores 27), especifique el flujo de bits 21 de modo que los bits correspondientes a un segundo esquema de compresión (por ejemplo, el esquema de compresión basado en la direccionalidad o el esquema de compresión basado en la direccionalidad representado por la unidad de descomposición 28 basada en dirección) no estén incluidos en el flujo de bits 21. Por ejemplo, la unidad de generación de flujo de bits 42 puede generar el flujo de bits 21 para no incluir elementos sintácticos o campo HOAPredictionInfo que pueden reservarse para usar para especificar información de predicción entre señales direccionales del esquema de compresión basado en la direccionalidad. Ejemplos del flujo de bits 21 generados de acuerdo con diversos aspectos de las técnicas descritas en esta divulgación se muestran en los ejemplos de las FIGs. 8E y 8F.

**[0073]** En otras palabras, la predicción de señales direccionales puede ser parte de la Síntesis del Sonido Predominante empleada por la unidad de descomposición basada en la dirección 28 y depende de la existencia del TipoCanal 0 (que puede indicar una señal basada en la dirección). Cuando no está presente ninguna señal basada en la dirección dentro de una trama, no se puede realizar ninguna predicción de señales direccionales. Sin embargo, la información de banda lateral asociada *HOAPredictionInfo()* puede, aunque no se utilice, escribirse en cada trama independientemente de la existencia de señales basadas en la dirección. Cuando no existe señal direccional dentro de una trama, las técnicas descritas en esta divulgación pueden permitir que la unidad de generación de flujo de bits 42 reduzca el tamaño de la banda lateral al no señalar HOAPredictionInfo en la banda lateral como se establece en la siguiente Tabla (donde las cursivas con subrayado denotan adiciones):

**Tabla: Sintaxis de HOAFrame**

Sintaxis	N.º de bits	Mnemónico
HOAFrame(usacIndependencyFlag) { NumOfDirSigs = 0; NumOfVecSigs = 0; NumOfContAddHoaChans = 0; if(usacIndependencyFlag){ hoaIndependencyFlag = usacIndependencyFlag; } else{ <b>hoaIndependencyFlag;</b> } for(i=0; i< NumOfAdditionalCoders; ++i){ ChannelSideInfoData(i); <u><i>if (MaxGainCorrAmpExp&gt;0)</i></u> } }	<b>1</b>	<b>bslbf</b>

```

        { HOAGainCorrectionData(i); }
    switch ChannelType[i] {
    case 0:
        DirSigChannelIds[NumOfDirSigs] = i + 1;
        NumOfDirSigs++;
        break;
    case 1:
        VecSigChannelIds[NumOfVecSigs] = i + 1;
        NumOfVecSigs++;
        break;
    case 2:
        if (AmbCoeffTransitionState[i] == 0) {
            ContAddHoaCoeff [NumOfContAddHoaChans] =
                AmbCoeffIdx[i];
            NumOfContAddHoaChans++;
        }
        break;
    }
}
if (MaxGainCorrAmpExp > 0) {
    for ( i= NumOfAdditionalCoders;
        i< NumHOATransportChannels; ++i){
        HOAGainCorrectionData(i);
    }
}
for(i=0; i< NumOfVecSigs; ++i){
    VVectorData ( VecSigChannelIds(i) );
}
if (NumOfDirSigs>0)
    { HOAPredictionInfo( DirSigChannelIds, NumOfDirSigs ); }
byte_alignment();
}

```

**[0074]** A este respecto, las técnicas pueden permitir que un dispositivo, tal como el dispositivo de codificación de audio 20, se configure para, al comprimir datos de audio ambisónicos de orden superior utilizando un primer esquema de compresión, especificar un flujo de bits representativo de una versión comprimida de los datos de audio ambisónicos de orden superior que no incluye los bits correspondientes a un segundo esquema de compresión también utilizado para comprimir los datos de audio ambisónicos de orden superior.

**[0075]** En algunos casos, el primer esquema de compresión comprende un esquema de compresión de descomposición basado en vectores. En estos y otros casos, el esquema de compresión de descomposición basado en vectores comprende un esquema de compresión que implica la aplicación de una descomposición de valores singulares (o equivalentes de los mismos descritos con mayor detalle en esta divulgación) a los datos de audio ambisónicos de orden superior.

**[0076]** En estos y otros casos, el dispositivo de codificación de audio 20 puede configurarse para especificar que el flujo de bits que no incluye los bits corresponde a al menos un elemento sintáctico utilizado para realizar el segundo tipo de esquema de compresión. El segundo esquema de compresión puede, como se indicó anteriormente, comprender un esquema de compresión basado en la direccionalidad.

**[0077]** El dispositivo de codificación de audio 20 también puede configurarse para especificar el flujo de bits 21 de manera que el flujo de bits 21 no incluya los bits correspondientes a un elemento sintáctico HOAPredictionInfo del segundo esquema de compresión.

**[0078]** Cuando el segundo esquema de compresión comprende un esquema de compresión basado en la direccionalidad, el dispositivo de codificación de audio 20 puede configurarse para especificar el flujo de bits 21 de manera que el flujo de bits 21 no incluya los bits correspondientes a un elemento sintáctico HOAPredictionInfo del esquema de compresión basado en la direccionalidad. En otras palabras, el dispositivo de codificación de audio 20 puede configurarse para especificar el flujo de bits 21 de manera que el flujo de bits 21 no incluya los bits correspondientes a al menos un elemento sintáctico utilizado para realizar el segundo tipo de esquemas de compresión, siendo el al menos un elemento sintáctico indicativo de una predicción entre dos o más señales basadas

en la dirección. Nuevamente, una vez más, cuando el segundo esquema de compresión comprende un esquema de compresión basado en la direccionalidad, el dispositivo de codificación de audio 20 puede configurarse para especificar el flujo de bits 21 de manera que el flujo de bits 21 no incluye los bits correspondientes a un elemento sintáctico HOAPredictionInfo del esquema de compresión basado en la direccionalidad, donde el elemento sintáctico HOAPredictionInfo es indicativo de una predicción entre dos o más señales basadas en la dirección.

**[0079]** Diversos aspectos de las técnicas pueden permitir además que la unidad de generación de flujo de bits 46 especifique el flujo de bits 21 en ciertos casos, de manera que el flujo de bits 21 no incluye datos de corrección de ganancia. La unidad de generación de flujo de bits 46 puede, cuando se suprime la corrección de ganancia, especificar el flujo de bits 21 de manera que el flujo de bits 21 no incluya los datos de corrección de ganancia. Se muestran ejemplos del flujo de bits 21 generado de acuerdo con diversos aspectos de las técnicas, como se indicó anteriormente, en los ejemplos de las FIGS. 8E y 8F.

**[0080]** En algunos casos, la corrección de ganancia se aplica cuando se realizan ciertos tipos de codificación psicoacústica dado el rango dinámico relativamente más pequeño de estos ciertos tipos de codificación psicoacústica en comparación con otros tipos de codificación psicoacústica. Por ejemplo, AAC tiene un rango dinámico relativamente más pequeño que la codificación unificada de voz y audio (USAC). Cuando el esquema de compresión (como un esquema de compresión de síntesis basado en vectores o un esquema de compresión basado en la direccionalidad) implica USAC, la unidad de generación de flujo de bits 46 puede indicar en el flujo de bits 21 que la corrección de ganancia ha sido suprimida (por ejemplo, especificando un elemento sintáctico *MaxGainCorrAmbExp* en el *HOAConfig* con un valor de cero en el flujo de bits 21) y luego especificar el flujo de bits 21 para no incluir los datos de corrección de ganancia (en un campo *HOAGainCorrectionData* ()).

**[0081]** En otras palabras, el campo de bits *MaxGainCorrAmbExp* como parte de la *HOAConfig* (véase la Tabla 71 en el CD) puede controlar el grado en que el módulo de control de ganancia automático afecta a las señales del canal de transporte antes de la codificación del núcleo USAC. En algunos casos, este módulo fue desarrollado para RM0 para mejorar el rango dinámico no ideal de la implementación del codificador AAC disponible. Con el cambio del codificador AAC a codificador de núcleo USAC durante la fase de integración, el rango dinámico del codificador central puede mejorar y, por lo tanto, la necesidad de este módulo de control de ganancia puede no ser tan crítica como antes.

**[0082]** En algunos casos, la funcionalidad de control de ganancia se puede suprimir si *MaxGainCorrAmbExp* se establece en 0. En estos casos, la información de banda lateral asociada *HOAGainCorrectionData* ( ) pueden no escribirse en cada trama de HOA según la tabla anterior que ilustra la "Sintaxis de HOAFrame". Para la configuración donde *MaxGainCorrAmbExp* se establece en 0, las técnicas descritas en esta divulgación pueden no indicar el *HOAGainCorrectionData*. Además, en tal escenario, el módulo de control de ganancia inverso puede incluso anularse, reduciendo la complejidad del descodificador en aproximadamente 0,05 MOPS por canal de transporte sin ningún efecto secundario negativo.

**[0083]** A este respecto, las técnicas pueden configurar el dispositivo de codificación de audio 20 para, cuando se suprime la corrección de ganancia durante la compresión de datos de audio ambisónicos de orden superior, especificar el flujo de bits 21 representativo de una versión comprimida de los datos de audio ambisónicos de orden superior de modo que el flujo de bits 21 no incluye datos de corrección de ganancia.

**[0084]** En estos y otros casos, el dispositivo de codificación de audio 20 puede configurarse para comprimir los datos de audio ambisónicos de orden superior de acuerdo con un esquema de compresión de descomposición basado en vectores para generar la versión comprimida de los datos de audio ambisónicos de orden superior. Los ejemplos del esquema de compresión de descomposición pueden implicar la aplicación de una descomposición de valores singulares (o equivalentes de los mismos descritos con mayor detalle anteriormente) a los datos de audio ambisónicos de orden superior para generar la versión comprimida de los datos de audio ambisónicos de orden superior.

**[0085]** En estos y otros casos, el dispositivo de codificación de audio 20 puede configurarse para especificar un elemento sintáctico *MaxGainCorrAmbExp* en el flujo de bits 21 como cero para indicar que se suprime la corrección de ganancia. En algunos casos, el dispositivo de codificación de audio 20 puede configurarse para especificar, cuando se suprime la corrección de ganancia, el flujo de bits 21 de manera que el flujo de bits 21 no incluye un campo de datos *HOAGainCorrection* que almacena los datos de corrección de ganancia. En otras palabras, el dispositivo de codificación de audio 20 puede configurarse para especificar un elemento sintáctico *MaxGainCorrAmbExp* en el flujo de bits 21 como cero para indicar que la corrección de ganancia se suprime y no incluye en el flujo de bits un campo de datos *HOAGainCorrection* que almacena los datos de corrección de ganancia.

**[0086]** En estos y otros casos, el dispositivo de codificación de audio 20 puede configurarse para suprimir la corrección de ganancia cuando la compresión de los datos de audio ambisónicos de orden superior incluye la aplicación de una codificación de audio de voz y voz de audio unificada (USAC) a los datos de audio ambisónicos de orden superior.

**[0087]** Las potenciales optimizaciones anteriores para la señalización de diversa información en el flujo de bits 21 se pueden adaptar o, de otro modo, actualizar de la manera que se describe con mayor detalle a continuación. Las

actualizaciones se pueden aplicar junto con otras actualizaciones que se analizan a continuación o se pueden usar para actualizar solo varios aspectos de las optimizaciones analizadas anteriormente. Como tal, se considera cada posible combinación de actualizaciones a las optimizaciones descritas anteriormente, incluida la aplicación de una actualización única descrita a continuación a las optimizaciones descritas anteriormente o cualquier combinación particular de las actualizaciones descritas a continuación a las optimizaciones descritas anteriormente.

**[0088]** Para especificar una matriz en el flujo de bits, la unidad de generación de flujo de bits 42 especifica un ID\_CONFIG\_EXT\_HOA\_MATRIX en una mpeg3daConfigExtension() del flujo de bits 21, como se muestra a continuación, en negrita y resaltado en la siguiente Tabla. La siguiente tabla es representativa de la sintaxis para especificar la parte mpeg3daConfigExtension() del flujo de bits 21:

**Tabla - Sintaxis de mpeg3daConfigExtension()**

Sintaxis	N.º de bits	Mnemónico
<pre> mpeg3daConfigExtension() {     numConfigExtensions = escapedValue(2,4,8) + 1;      for (confExtIdx=0; confExtIdx&lt;numConfigExtensions; confExtIdx++) {         usacConfigExtType[confExtIdx]          = escapedValue(4,8,16);         usacConfigExtLength[confExtIdx] = escapedValue(4,8,16);          switch (usacConfigExtType[confExtIdx]) {         case ID_CONFIG_EXT_FILL:             while (usacConfigExtLength[confExtIdx]--) {                 <b>fill_byte[i];</b> /* should be '10100101' */             }             break;         case ID_CONFIG_EXT_DOWNMIX:             downmixConfig();             break;         case ID_CONFIG_EXT_LOUDNESS_INFO:             loudnessInfoSet();             break;         case ID_CONFIG_EXT_AUDIOSCENE_INFO:             mae_AudioSceneInfo();             break;         case <b>ID_CONFIG_EXT_HOA_MATRIX :</b>             <b>HoaRenderingMatrixSet();</b>             <b>break;</b>         default:             while (usacConfigExtLength[confExtIdx]--) {                 <b>tmp;</b>             }             break;         }     } } </pre>	<p>8</p> <p>8</p>	<p><b>uimsbf</b></p> <p><b>uimsbf</b></p>

ID\_CONFIG\_EXT\_HOA\_MATRIX en la tabla anterior proporciona un contenedor en el que se especifica la matriz de renderización, con el contenedor denotado como "HoaRenderingMatrixSet()".

**[0089]** El contenido del contenedor HoaRenderingMatrixSet() se define de acuerdo con la sintaxis establecida en la siguiente tabla:

**Tabla - Sintaxis de HoaRenderingMatrixSet()**

Sintaxis	N.º de bits	Mnemónico
<b>HoaRenderingMatrixSet()</b>		
{		
<b>numHoaRenderingMatrices;</b>	<b>5</b>	<b>uimsbf</b>
for (k=0; k< numHoaRenderingMatrices; ++k) {		
<b>HoaRenderingMatrixId;</b>	<b>7</b>	<b>uimsbf</b>
<b>CICPspeakerLayoutIdx;</b>	<b>6</b>	<b>uimsbf</b>
<b>HoaMatrixLenBits = escapedValue(8,8,12);</b>	<b>8..28</b>	
HoaRenderingMatrix(NumOfHoeCoeffs	<b>HoaMatrixLenBits</b>	
outputConfig(CICPspeakerLayoutIdx),		
outputCount(CICPspeakerLayoutIdx) );		
}		
}		

Como se muestra en la Tabla directamente arriba, el HoaRenderingMatrixSet() incluye una cantidad de elementos sintácticos diferentes, incluyendo un numHoaRenderingMatrices, un HoaRendereringMatrixId, un CICPspeakerLayoutIdx, un HoaMatrixLenBits y un HoARenderingMatrix.

[0090] El elemento sintáctico numHoaRenderingMatrices especifica un número de definiciones HoaRendereringMatrixId presentes en el elemento de flujo de bits. El elemento sintáctico HoaRenderingMatrixId representa un campo que define de forma única un Id para una matriz de renderización de HOA predeterminada disponible en el lado del descodificador o una matriz de renderización de HOA transmitida. A este respecto, HoaRenderingMatrixId representa un ejemplo del valor de la señal que incluye dos o más bits que definen un índice que indica que el flujo de bits incluye una matriz utilizada para renderizar coeficientes armónicos esféricos en una pluralidad de alimentaciones de altavoces o el valor de la señal que incluye dos o más bits que definen un índice asociado con una de una pluralidad de matrices usadas para renderizar coeficientes armónicos esféricos en una pluralidad de alimentaciones de altavoces. El elemento sintáctico CICPspeakerLayoutIdx puede representar un valor que describe el diseño de altavoces de salida para la matriz de renderización de HOA dada y puede corresponder a un elemento ChannelConfiguration definido en ISO/IEC 23000 1-8. Los HoaMatrixLenBits (que también se pueden denotar como el elemento sintáctico "HoaRenderingMatrixLenBits") pueden especificar una longitud del siguiente elemento de flujo de bits (por ejemplo, el contenedor HoaRenderingMatrix()) en bits.

[0091] El contenedor HoaRenderingMatrix () incluye un NumOfHoeCoeffs seguido de un contenedor outputConfig() y un contenedor outputCount(). El contenedor outputConfig() incluye vectores de configuración de canal que especifican la información sobre cada altavoz. La unidad de generación de flujo de bits 42 asume que esta información de altavoz se conoce a partir de las configuraciones de canal del diseño de salida. Cada entrada, outputConfig[i], representa una estructura de datos con los siguientes miembros:

AzimuthAngle (que denota el valor absoluto del ángulo de azimut del hablante);

AzimuthDirection (que denota la dirección de azimut usando, como por ejemplo, 0 para la izquierda y 1 para la derecha);

Ángulo de elevación (que denota el valor absoluto de los ángulos de elevación del altavoz);

ElevationDirection (que denota la dirección de elevación utilizando, como por ejemplo, 0 para arriba y 1 para abajo); e isLFE (que indica si el altavoz es un altavoz de efecto de baja frecuencia (LFE)).

La unidad de generación de flujo de bits 42 invoca una función auxiliar, en algunos casos, indicada como "findSymmetricSpeakers", que especifica adicionalmente lo siguiente:

pairType (que almacena un valor de SYMMETRIC (que significa un par simétrico de dos altavoces en algún ejemplo), CENTER o ASYMMETRIC); y

symmetricPair->originalPosition (que denota la posición en la configuración del canal original del segundo altavoz (por ejemplo, derecho) en el grupo, solo para grupos SYMMETRIC).

El contenedor outputCount() especifica una cantidad de altavoces para los cuales se define la matriz de renderización de HOA.

[0092] La unidad de generación de flujo de bits 42 especifica el contenedor HoaRenderingMatrix() de acuerdo con la sintaxis establecida en la siguiente tabla:

Tabla - Sintaxis de HoaRenderingMatrix ()

Sintaxis	N.º de bits	Mnemónico
HoaRenderingMatrix ()		
{		
lfeExist = 0;		
hasLfeRendering = 0;		
for (i=0; i< inputCount; ++i)		
isHoaCoefSparse[i] = 0;		
<b>precisionLevel</b>	<b>1</b>	<b>uimsbf</b>
If ( <b>gainLimitPerHoaOrder</b> ) {	<b>1</b>	<b>uimsbf</b>
for (i = 0; i<(maxHoaOrder+1); ++i ) {		
maxGain[i] = - escapedValue(3, 5, 5);		
minGain[i] = -(escapedValue(4, 5, 6) + 1 - maxGain[i]);		
}		
} else {		
maxGain[0] = - escapedValue (3, 5, 5);		
minGain[0] = -( escapedValue (4, 5, 6) + 1- maxGain[0]);		
for (i = 1; i<(maxHoaOrder+1); ++i ) {		
maxGain[i] = maxGain[0];		
minGain[i] = minGain[0];		
}		
}		
If ( <b>isFullMatrix</b> ) {	<b>1</b>	<b>uimsbf</b>
<b>firstSparseOrder</b>	<b>1</b>	<b>uimsbf</b>
for (i = (firstSparseOrder*firstSparseOrder); i<inputCount;		
++i)		
isHoaCoefSparse[i] = 1;		
}		
for (i=0; i< outputCount; ++i){		
if (outputConfig[i].isLFE)		
lfeExist = 1;		
}		
if (lfeExist)		
<b>hasLfeRendering;</b>	<b>1</b>	<b>uimsbf</b>
numPairs = findSymmetricSpeakers(outputCount, outputConfig,		
hasLfeRendering);		
for (i=0; i<numPairs; ++i) {		
valueSymmetricPairs[i] = 0;		
signSymmetricPairs[i] = 0;		
}		
<b>zerothOrderAlwaysPositive;</b>	<b>1</b>	<b>uimsbf</b>
if ( <b>isAllValueSymmetric</b> ) {	<b>1</b>	<b>uimsbf</b>
for (i=0; i<numPairs; ++i) { valueSymmetricPairs[i] = 1; }		
} else {		
if ( <b>isAnyValueSymmetric</b> ) {	<b>1</b>	<b>uimsbf</b>
for (i=0; i<numPairs; ++i)		
valueSymmetricPairs[i] = <b>boolVal</b> ;	<b>1</b>	<b>uimsbf</b>
If ( <b>isAnySignSymmetric</b> ) {	<b>1</b>	<b>uimsbf</b>
for (i=0; i<numPairs; ++i) {		
if (0==valueSymmetricPairs[i])		
signSymmetricPairs[i] = <b>boolVal</b> ;	<b>1</b>	<b>uimsbf</b>
}		
}		
} else {		
if ( <b>isAllSignSymmetric</b> ) {	<b>1</b>	<b>uimsbf</b>



for (i=0; i<numPairs; ++i)		
signSymmetricPairs[i] = 1;		
} else { /* isAnyValueSymmetric==0 */		
if { <b>isAnySignSymmetric</b> }	<b>1</b>	<b>uimsbf</b>
for (i=0; i<numPairs; ++i)		
signSymmetricPairs[i] = <b>boolVal</b>	<b>1</b>	<b>uimsbf</b>
}		
}		
<b>hasVerticalCoef;</b>	<b>1</b>	<b>uimsbf</b>
DecodeHoaMatrixData()		
}		

Como se muestra en la Tabla directamente arriba, el elemento sintáctico numPairs se establece en el valor de salida al invocar la función auxiliar findSymmetricSpeakers utilizando el outputCount y outputConfig y hasLfeRendering como entradas. Por lo tanto, numPairs indica el número de pares de altavoces simétricos identificados en la configuración del altavoz de salida que se consideran para la codificación de simetría eficiente. El elemento sintáctico de precisionLevel en la tabla anterior puede denotar una precisión utilizada para la cuantificación uniforme de las ganancias de acuerdo con la siguiente tabla:

Tabla - Tamaño de paso de cuantificación uniforme de hoaGain en función del precisionLevel

precisionLevel	tamaño de paso de cuantificación más pequeño [dB]
0	1,0
1	0,5
2	0,25
3	0,125

**[0093]** El elemento sintáctico gainLimitPerHoaOrder que se muestra en la Tabla anterior que establece la sintaxis de HoaRenderingMatrix() puede representar un indicador que indica si maxGain y minGain se especifican individualmente para cada orden o para toda la matriz de renderización de HOA. Los elementos sintácticos maxGain[i] pueden especificar una ganancia real máxima en la matriz para los coeficientes para el orden de HOA i expresada, por ejemplo, en decibelios (dB). Los elementos sintácticos minGain[i] pueden especificar una ganancia real mínima en la matriz para los coeficientes del orden de HOA i expresado, nuevamente por ejemplo, en dB. El elemento sintáctico isFullMatrix representa un indicador que indica si la matriz de renderización de HOA es dispersa o está llena. El elemento sintáctico firstSparseOrder especifica, en el caso de que la matriz de renderización de HOA se haya especificado como dispersa por el elemento sintáctico isFullMatrix, el primer orden de HOA que está dispersamente codificado. El elemento sintáctico isHoaCoefSparse representa un vector de máscara de bits obtenido del elemento sintáctico firstSparseOrder. El elemento sintáctico lfeExists puede representar un indicador indicativo de si existen uno o más LFE en outputConfig. El elemento sintáctico hasLfeRendering indica si la matriz de renderización contiene elementos distintos de cero para uno o más canales LFE. El elemento sintáctico zerothOrderAlwaysPositive puede representar un indicador indicativo de si el orden 0-ésimo de HOA solo tiene valores positivos.

**[0094]** El elemento sintáctico isAllValueSymmetric representa un indicador que indica si todos los pares de altavoces simétricos tienen valores absolutos iguales en la matriz de renderización HOA. El elemento sintáctico isAnyValueSymmetric representa un indicador que indica, cuando es falso, por ejemplo, si algunos de los pares de altavoces simétricos tienen valores absolutos iguales en la matriz de renderización de HOA. El elemento sintáctico valueSymmetricPairs puede representar una máscara de bits de longitud numPairs que indica los pares de altavoces con simetría de valores. El elemento sintáctico isValueSymmetric puede representar una máscara de bits obtenida de la manera que se muestra en la Tabla 3 del elemento sintáctico valueSymmetricPairs. El elemento sintáctico isAllSignSymmetric denota, cuando no hay simetrías de valores en la matriz, si todos los pares de altavoces simétricos tienen al menos simetrías de signos numéricos. El elemento sintáctico isAnySignSymmetric representa un indicador que indica si hay al menos algunos pares de altavoces simétricos con simetrías de signos numéricos. El elemento sintáctico signSymmetricPairs representa una máscara de bits de longitud numPairs que indica los pares de altavoces con simetría de signo. La variable métrica isSignSymmetric representa una máscara de bits obtenida del elemento sintáctico signSymmetricPairs de la manera que se muestra arriba en la Tabla que establece la sintaxis de HoaRenderingMatrix(). El elemento sintáctico hasVerticalCoef puede representar un indicador indicativo de si la matriz es una matriz de renderización HOA horizontal solamente. El elemento sintáctico bootVal puede representar una variable utilizada en el ciclo de descodificación.

**[0095]** En otras palabras, la unidad de generación de flujo de bits 42 puede analizar el renderizador de audio 1 para generar una cualquiera o más de la información de simetría de valor anterior (por ejemplo, cualquier combinación de

uno o más del elemento sintáctico isAllValueSymmetric, el elemento sintáctico isAnyValueSymmetric, el elemento sintáctico valueSymmetricPairs, el elemento sintáctico isValueSymmetric y el elemento sintáctico valueSymmetricPairs) u obtener de otro modo la información de simetría de valor. La unidad de generación de flujo de bits 42 puede especificar la información de renderizador de audio 2 en el flujo de bits 21 de la manera mostrada anteriormente de manera que la información de renderizador de audio 2 incluya la información de simetría de signo de valor.

**[0096]** Además, la unidad de generación de flujo de bits 42 también analiza el renderizador de audio 1 para generar una o más de la información de simetría de signo anterior (por ejemplo, cualquier combinación de uno o más del elemento sintáctico isAllSignSymmetric, el elemento sintáctico isAnySignSymmetric, el elemento sintáctico signSymmetricPairs, el elemento sintáctico isSignSymmetric, y el elemento sintáctico sign-SymmetricPairs) o de otro modo obtener la información de simetría de signo. La unidad de generación de flujo de bits 42 especifica la información de renderizador de audio 2 en el flujo de bits 21 de la manera mostrada anteriormente de modo que la información de renderizador de audio 2 incluya la información de simetría de signo de audio.

**[0097]** Cuando se determina la información de simetría de valor y la información de simetría de signo, la unidad de generación de flujo de bits 42 analiza los diversos valores del renderizador de audio 1, que se puede especificar como una matriz. Una matriz de renderización puede formularse como un pseudoinverso de una matriz R. En otras palabras, para renderizar  $(N+1)^2$  canales de HOA (indicados como Z a continuación) en señales de altavoz L (indicadas por el vector de columna, p, de las señales del altavoz L), se puede dar la siguiente ecuación:

$$Z = R * p.$$

Para llegar a la matriz de renderización que emite las señales del altavoz L, la inversa de la matriz R se multiplica por los canales Z HOA como se muestra en la siguiente ecuación:

$$p = R^{-1} * Z.$$

A menos que el número de canales de altavoz, L, sea el mismo que el número de canales de Z HOA,  $(N+1)^2$ , la matriz R no será cuadrada y no podrá determinarse una inversa perfecta. Como resultado, en lugar de eso, se puede usar el pseudoinverso, que se define de la siguiente manera:

$$\text{pinv}(R) = R^T (R * R^T)^{-1},$$

donde  $R^T$  denota la transposición de la matriz R. Sustituyendo  $R^{-1}$  en la ecuación anterior, la resolución de las señales del altavoz L denotadas por el vector de columna p se puede denotar matemáticamente de la siguiente manera:

$$p = \text{pinv}(R) * Z = R^T (R * R^T)^{-1} * Z.$$

**[0098]** Las entradas de la matriz R son los valores de los armónicos esféricos para las posiciones del altavoz con  $(N+1)^2$  filas para los diferentes armónicos esféricos y L columnas para los altavoces. La unidad de generación de flujo de bits 42 puede determinar pares de altavoces basándose en los valores para los altavoces. Analizando los valores de los armónicos esféricos para las posiciones del altavoz, la unidad de generación de flujo de bits 42 puede determinar basándose en los valores cuales de las posiciones del altavoz son pares (por ejemplo, como pares pueden tener valores similares, casi iguales o iguales, pero con signos opuestos).

**[0099]** Después de identificar los pares, la unidad de generación de flujo de bits 42 puede determinar para cada par, si los pares tienen el mismo valor o casi el mismo valor. Cuando todos los pares tienen el mismo valor, la unidad de generación de flujo de bits 42 puede establecer el elemento sintáctico isAllValueSymmetric en uno. Cuando todos los pares no tienen el mismo valor, la unidad de generación de flujo de bits 42 puede establecer el elemento sintáctico isAllValueSymmetric en cero. Cuando uno o más, pero no todos los pares tienen el mismo valor, la unidad de generación de flujo de bits 42 puede establecer el elemento sintáctico isAnyValueSymmetric en uno. Cuando ninguno de los pares tiene el mismo valor, la unidad de generación de flujo de bits 42 puede establecer el elemento sintáctico isAnyValueSymmetric en cero. Para pares con valores simétricos, la unidad de generación de flujo de bits 42 solo puede especificar un valor en lugar de dos valores separados para el par de altavoces, reduciendo así el número de bits usados para representar la información de renderización de audio 2 (por ejemplo, la matriz en este ejemplo) en el flujo de bits 21.

**[0100]** Cuando no hay simetrías de valores entre los pares, la unidad de generación de flujo de bits 42 también determina para cada par, si los pares de altavoces tienen simetría de signo (lo que significa que un altavoz tiene un valor negativo mientras que el otro altavoz tiene un valor positivo). Cuando todos los pares tienen simetría de signo, la unidad de generación de flujo de bits 42 establece el elemento sintáctico isAllSignSymmetric en uno. Cuando todos los pares no tienen simetría de signo, la unidad de generación de flujo de bits 42 establece el elemento sintáctico isAllSignSymmetric en cero. Cuando uno o más, pero no todos los pares tienen simetría de signo, la unidad de

generación de flujo de bits 42 establece el elemento sintáctico `isAnySignSymmetric` en uno. Cuando ninguno de los pares tiene simetría de signo, la unidad de generación de flujo de bits 42 establece el elemento sintáctico `isAnySignSymmetric` en cero. Para pares con signos simétricos, la unidad de generación de flujo de bits 42 solo especifica uno o ningún signo en lugar de dos signos separados para el par de altavoces, reduciendo así el número de bits utilizados para representar la información de renderización de audio 2 (por ejemplo, la matriz en este ejemplo) en el flujo de bits 21.

**[0101]** La unidad de generación de flujo de bits 42 especifica el contenedor `DecodeHoaMatrixData()` que se muestra en la Tabla estableciendo la sintaxis de `HoaRenderingMatrix()` de acuerdo con la sintaxis que se muestra en la siguiente tabla:

Tabla - Sintaxis de DecodeHoaMatrixData

Sintaxis	N.º de bits	Mnemónico
<pre> DecodeHoaMatrixData () {     j = 0;     for (i=0; i&lt;outputCount; ++i) {         isValueSymmetric[i] = 0;         isSignSymmetric[i] = 0;         if ((outputConfig[i].pairType == SP_PAIR_SYMMETRIC)         &amp;&amp; (outputConfig[i].symmetricPair != NULL)) {             if (0==(outputConfig[i].isLFE &amp;&amp; (0==hasLfeRendering))) {                 isValueSymmetric[i] = valueSymmetricPairs[j];                 isSignSymmetric[i] = signSymmetricPairs[j++];             }         }     }     for (i = 0; i &lt; inputCount; ++i) {         currentHoaOrder = ceil(sqrt(i+1))-1;         for (j = outputCount-1; j &gt;= 0; --j) {             signMatrix[i * outputCount + j] = 1;             hoaMatrix [i * outputCount + j] = 0.0;             if ((vertBitmask[i] &amp;&amp; hasVerticalCoef)    !vertBitmask[i])         {             hasValue = 1;             if (0 == isValueSymmetric[j]) {                 if ((hasLfeRendering &amp;&amp; outputConfig[j].isLFE)    (!outputConfig[j].isLFE)) {                     if (isHoaCoefSparse[i]){                         <b>hasValue</b>                     }                     if (hasValue) {                         hoaMatrix [i * outputCount + j] =  DecodeHoaGainValue(currentHoaOrder);                         if (0==isSignSymmetric[j]) {                             if (hoaMatrix [i * outputCount + j] != 0.0) {                                 if (currentHoaOrder    !zerothOrderAlwaysPositive) {                                     signMatrix[i * outputCount + j] = <b>boolVal</b>*2-1;                                 }                             }                         }                     }                 }             }         }     } } </pre>	<p><b>1</b></p> <p><b>1</b></p>	<p><b>Uimsbf</b></p> <p><b>Uimsbf</b></p>

```

    } else { // isSignSymmetric[i] == 1
        pairIdx = outputConfig[j].symmetricPair-
>originalPosition;
        signMatrix[i * outputCount + j] = symSigns[i] *
signMatrix[i * outputCount + pairIdx];
    }
    }
    }
    else { // isAllValueSymmetric
        pairIdx = outputConfig[j].symmetricPair-
>originalPosition;
        hoaMatrix [i*outputCount+j] =
            hoaMatrix [i*outputCount+pairIdx];
        SignMatrix[i *outputCount+j] = symSigns[i] *

        signMatrix[i*outputCount+pairIdx];
    }
    }
    }
    for (i = 0; i < inputCount; ++i) {
        for (j = 0; j < outputCount; ++j) {
            hoaMatrix[i *outputCount+j] *=
signMatrix[i*outputCount+j];
        }
    }
}

```

**[0102]** El elemento sintáctico hasValue en la Tabla anterior, que establece la sintaxis de DecodeHoaMatrixData, representa un indicador indicativo de si el elemento de matriz está dispersamente codificado. El elemento sintáctico de signMatrix puede representar una matriz con los valores de signo de la matriz de renderización de HOA en, por ejemplo, forma de vector linealizado. El elemento sintáctico hoaMatrix puede representar los valores de la matriz de renderización HOA en, por ejemplo, forma vectorial linealizada. La unidad de generación de flujo de bits 42 puede especificar el contenedor DecodeHoaGainValue() que se muestra en la Tabla que expone la sintaxis de DecodeHoaMatrixData de acuerdo con la sintaxis que se muestra en la siguiente Tabla:

Tabla - Sintaxis de DecodeHoaGainValue

Sintaxis	N.º de bits	Mnemónico
<pre> DecodeHoaGainValue(order) {     nAlphabet = (maxGain[order] - minGain[order]) * 2 ^ precisionLevel + 2;     gainValueIndex = ReadRange(nAlphabet);     gainValue = maxGain[order] - gainValueIndex / 2 ^ precisonLevel;     if (gainValue &lt; minGain) {         gainValue = 0.0;     }else {         gainValue = 10.0 ^ (gainValue / 20.0);     }     return gainValue; } </pre>		

**[0103]** La unidad de generación de flujo de bits 42 puede especificar el contenedor readRange() que se muestra en la Tabla que expone la sintaxis de DecodeHoaGainValue de acuerdo con la sintaxis especificada en la siguiente tabla:

Tabla 7- Sintaxis de ReadRange

Sintaxis	N.º de bits	Mnemónico
<pre> ReadRange(alphabetSize) {   nBits = floor(log2(alphabetSize));   nUnused = 2 ^ (nBits + 1) - alphabetSize;   range;   if (range &gt;= nUnused) {     rangeExtra;     range = range * 2 - nUnused + rangeExtra;   }   return range; } </pre>	<p><b>nBits</b></p> <p><b>1</b></p>	<p><b>uimbsbf</b></p> <p><b>uimbsbf</b></p>

**[0104]** Aunque no se muestra en el ejemplo de la FIG. 3, el dispositivo de codificación de audio 20 también puede incluir una unidad de salida de flujo de bits que conmuta la salida del flujo de bits desde el dispositivo de codificación de audio 20 (por ejemplo, entre el flujo de bits basado en la direccionalidad 21 y el flujo de bits basado en vectores 21), basándose en si una trama actual va a ser codificada usando la síntesis basada en la dirección o la síntesis basada en vectores. La unidad de salida de flujo de bits puede realizar la conmutación basándose en el elemento sintáctico emitido por la unidad de análisis de contenido 26, indicando si se realizó una síntesis basada en la dirección (como resultado de detectar que los coeficientes HOA 11 se generaron a partir de un objeto de audio sintético) o si se realizó una síntesis basada en vectores (como resultado de detectar que se registraron los coeficientes HOA). La unidad de salida del flujo de bits puede especificar la sintaxis de cabecera correcta para indicar la conmutación o codificación actual utilizada para la trama actual, junto con la respectiva de los flujos de bits 21.

**[0105]** Además, como se ha indicado anteriormente, la unidad de análisis de campo de sonido 44 puede identificar  $BG_{TOT}$  coeficientes HOA ambientales 47, que pueden cambiar de trama a trama (aunque a veces  $BG_{TOT}$  puede permanecer constante o igual entre dos o más tramas adyacentes (en el tiempo)). El cambio en  $BG_{TOT}$  puede dar como resultado cambios en los coeficientes expresados en los vectores  $V[k]$  de primer plano reducidos 55. El cambio en  $BG_{TOT}$  puede dar como resultado los coeficientes HOA de fondo (que también pueden denominarse "coeficientes HOA ambientales") que cambian de trama a trama (aunque, de nuevo, a veces  $BG_{TOT}$  puede permanecer constante o igual entre dos o más tramas adyacentes (en el tiempo)). Los cambios a menudo dan como resultado un cambio de energía para los aspectos del campo de sonido representados por la adición o eliminación de los coeficientes HOA ambientales adicionales y la correspondiente eliminación de coeficientes de, o adición de coeficientes a, los vectores  $V[k]$  de primer plano reducidos 55.

**[0106]** Como resultado, la unidad de análisis de campo de sonido 44 puede determinar además cuándo los coeficientes HOA ambientales cambian de trama a trama, y generan un indicador u otro elemento sintáctico indicativo del cambio para el coeficiente HOA ambiental, en términos de ser utilizado para representar los componentes ambientales del campo de sonido (donde el cambio también puede denominarse una "transición" del coeficiente HOA ambiental o una "transición" del coeficiente HOA ambiental). En particular, la unidad de reducción de coeficientes 46 puede generar el indicador (que puede indicarse como un indicador AmbCoeffTransition o un indicador AmbCoeffIdxTransition), proporcionando el indicador a la unidad de generación de flujo de bits 42 para que el indicador pueda incluirse en el flujo de bits 21 (posiblemente como parte de la información de canal lateral).

**[0107]** La unidad de reducción de coeficientes 46 puede, además de especificar el indicador de transición de coeficientes ambientales, también modificar cómo se generan los vectores  $V[k]$  de primer plano reducidos 55. En un ejemplo, al determinar que uno de los coeficientes HOA ambientales está en transición durante la trama actual, la unidad de reducción de coeficientes 46 puede especificar un coeficiente vectorial (que también puede denominarse "elemento vectorial" o "elemento") para cada uno de los vectores  $V$  de los vectores  $V[k]$  de primer plano reducidos 55, que corresponde al coeficiente HOA ambiental en transición. De nuevo, el coeficiente HOA ambiental en transición puede agregarse o eliminarse del número total  $BG_{TOT}$  de coeficientes de fondo. Por lo tanto, el cambio resultante en el número total de coeficientes de fondo afecta si el coeficiente HOA ambiental se incluye o no en el flujo de bits, y si el elemento correspondiente de los vectores  $V$  se incluye para los vectores  $V$  especificados en el flujo de bits en los modos de configuración segundo y tercero, descritos anteriormente. Se proporciona más información sobre cómo la unidad de reducción de coeficientes 46 puede especificar los vectores  $V[k]$  de primer plano reducidos 55 para superar los cambios en la energía en la Solicitud de EE.UU. con n.º de serie 14/594 533, titulada "TRANSITIONING OF AMBIENT HIGHER ORDER AMBISONIC COEFFICIENTS" ["TRANSICIÓN DE COEFICIENTES AMBISÓNICOS DE ORDEN SUPERIOR AMBIENTALES"], presentada 12 de enero de 2015.

**[0108]** La FIG. 4 es un diagrama de bloques que ilustra el dispositivo de decodificación de audio 24 de la FIG. 2 con mayor detalle. Como se muestra en el ejemplo de la FIG. 4, el dispositivo de decodificación de audio 24 puede

incluir una unidad de extracción 72, una unidad de reconstrucción de renderizador 81, una unidad de reconstrucción basada en la direccionalidad 90 y una unidad de reconstrucción basada en vectores 92. Aunque se describe a continuación, en la publicación de solicitud de patente internacional número WO 2014/194099, titulada "INTERPOLATION FOR DECOMPOSED REPRESENTATIONS OF A SOUND FIELD" ["INTERPOLACIÓN PARA RENDERIZACIONES DESCOMPUESTAS DE UN CAMPO DE SONIDO"] presentada el 29 de mayo de 2014, está disponible más información sobre el dispositivo de descodificación de audio 24 y los diversos aspectos de descompresión, o descodificación de otro modo, de los coeficientes HOA.

**[0109]** La unidad de extracción 72 puede representar una unidad configurada para recibir el flujo de bits 21 y extraer información de renderización de audio 2 y las diversas versiones codificadas (por ejemplo, una versión codificada basada en la direccionalidad o una versión codificada basada en vectores) de los coeficientes HOA 11. En otras palabras, las matrices de renderización de ambisónicos de orden superior (HOA) pueden ser transmitidas por el dispositivo de codificación de audio 20 para permitir el control sobre el proceso de renderización de HOA en el sistema de reproducción de audio 16. La transmisión puede ser facilitada por medio de la mpeg3daConfigExtension del tipo ID\_CONFIG\_EXT\_HOA\_MATRIX que se muestra arriba. La mpeg3daConfigExtension puede contener varias matrices de renderización de HOA para diferentes configuraciones de reproducción de altavoces. Cuando se transmiten matrices de renderización de HOA, el dispositivo de codificación de audio 20 señala, para cada señal de matriz de renderización de HOA, el diseño de altavoces objetivo asociado que determina junto con el HoaOrder las dimensiones de la matriz de renderización.

**[0110]** La transmisión de un único HoaRenderingMatrixId permite hacer referencia a una matriz de renderización de HOA predeterminada disponible en el sistema de reproducción de audio 16, o a una matriz de renderización de HOA transmitida desde fuera del flujo de bits de audio 21. En algunos casos, se supone que cada matriz de renderización de HOA se normaliza en N3D y sigue el orden de los coeficientes HOA tal como se define en el flujo de bits 21.

**[0111]** La función findSymmetricSpeakers indica, como se indicó anteriormente, un número y una posición de todos los pares de altavoces dentro de la configuración del altavoz proporcionada que son simétricos con respecto a, por ejemplo, el plano medio de un oyente en el llamado "punto óptimo". Esta función auxiliar se define de la siguiente manera: int findSymmetricSpeakers(int output-Count, SpeakerInformation\* outputConfig, int hasLfeRendering);

La unidad de extracción 72 invoca la función createSymSigns para calcular un vector de valores 1,0 y -1,0 que luego se puede usar para generar los elementos de matriz asociados con altavoces simétricos. Esta función createSymSigns se define de la forma siguiente:

```
void createSymSigns(int* symSigns, int hoaOrder)
{
    int n, m, k = 0;
    for (n = 0; n<=hoaOrder; ++n) {
        for (m = -n; m<=n; ++m)
            symSigns[k++] = ((m>=0)*2)-1;
    }
}
```

**[0112]** La unidad de extracción 72 puede invocar la función create2dBitmask para generar una máscara de bits para identificar los coeficientes HOA que solo se usan en el plano horizontal. La función create2dBitmask puede definirse de la forma siguiente:

```
void create2dBitmask(int* bitmask, int hoaOrder)
{
    int n, m, k = 0;
    bitmask[k++] = 0;
    for (n = 1; n<=hoaOrder; ++n) {
        for (m = -n; m<=n; ++m)
            bitmask[k++] = abs(m)!=n;
    }
}
```

**[0113]** Para descodificar los coeficientes de matriz de renderización de HOA, la unidad de extracción 72 primero extrae el elemento sintáctico HoaRenderingMatrixSet(), que como se indicó anteriormente contiene una o más matrices de renderización de HOA que se aplican para lograr una renderización de HOA a un diseño de altavoces deseado. En algunos casos, un flujo de bits dado puede no contener más de una instancia de HoaRenderingMatrixSet(). El elemento sintáctico HoaRenderingMatrix() contiene la información de matriz de renderización de HOA (que se denota como información de renderizador 2 en el ejemplo de la FIG. 4). La unidad de

extracción 72 lee primero en la información de configuración, que guía el proceso de decodificación. Después, la unidad de extracción 72 lee los elementos de la matriz en consecuencia.

**[0114]** En algunos casos, la unidad de extracción 72, al principio, lee los campos `precisionLevel` y `gainLimitPerOrder`. Cuando se establece el indicador `gainLimitPerOrder`, la unidad de extracción 72 lee y descodifica los campos `maxGain` y `minGain` para cada orden de HOA por separado. Cuando no se establece el indicador `gainLimitPerOrder`, la unidad de extracción 72 lee y descodifica los campos `maxGain` y `minGain` una vez y aplica estos campos a todos los órdenes HOA durante el proceso de decodificación. En algunos casos, el valor de `minGain` debe estar entre 0 dB y -69 dB. En algunos casos, el valor de `maxGain` debe ser entre 1 dB y 111 dB menor que el valor de `minGain`. La FIG. 9 es un diagrama que ilustra un ejemplo de ganancias mínimas y máximas dependientes del orden de HOA dentro de una matriz de renderización de HOA.

**[0115]** La unidad de extracción 72 luego lee el indicador `isFullMatrix`, que puede indicar si una matriz se define como completa o parcialmente dispersa. Cuando la matriz se define como parcialmente dispersa, la unidad de extracción 72 lee el siguiente campo (por ejemplo, el elemento sintáctico `firstSparseOrder`), que especifica el orden de HOA a partir del cual la matriz de renderización de HOA está dispersamente codificada. Las matrices de renderización de HOA a menudo pueden ser densas para orden inferior y volverse dispersas en los órdenes superiores, dependiendo de la configuración de reproducción del altavoz. La FIG. 10 es un diagrama que ilustra una matriz de renderización HOA de sexto orden parcialmente dispersa para 22 altavoces. La dispersión de la matriz mostrada en la FIG. 10 comienza en el coeficiente de HOA 26 (orden de HOA 5).

**[0116]** Dependiendo de si existen uno o más canales de efectos de baja frecuencia (LFE) dentro de la configuración de reproducción del altavoz (indicada por el elemento sintáctico `lfeExists`), la unidad de extracción 72 puede leer el campo `hasLfeRendering`. Cuando `hasLfeRendering` no está configurado, la unidad de extracción 72 está configurada para suponer que los elementos de la matriz relacionados con los canales LFE son ceros digitales. El siguiente campo leído por la unidad de extracción 72 es el indicador `zerothOrderAlwaysPositive`, que indica si los elementos de la matriz asociados con el coeficiente del orden 0 son positivos. En este caso, que `zerothOrderAlwaysPositive` indica que los coeficientes HOA de orden cero son positivos, la unidad de extracción 72 determina que los signos numéricos no están codificados para los coeficientes de matriz de renderización correspondientes a los coeficientes HOA de orden cero.

**[0117]** A continuación, se señalan las propiedades de la matriz de renderización de HOA para los pares de altavoces simétricos con respecto al plano medio. En algunos casos, hay dos propiedades de simetría relacionadas con a) simetría de valor y b) simetría de signo. En el caso de simetría de valor, los elementos de matriz del altavoz izquierdo del par de altavoces simétricos no están codificados, sino que la unidad de extracción 72 obtiene esos elementos de los elementos de matriz decodificados del altavoz derecho empleando la función auxiliar `createSymSigns`, que realiza lo siguiente:

```
pairIdx = outputConfig[j].symmetricPair->originalPosition;
```

```
hoaMatrix[i * outputCount + j] = hoaMatrix[i * outputCount + pairIdx];
```

y

```
signMatrix[i * outputCount + j] = symSigns[i] * signMatrix[i * outputCount + pairIdx].
```

**[0118]** Cuando un par de altavoces no tiene un valor simétrico, entonces los elementos de la matriz pueden ser simétricos con respecto a sus signos numéricos. Cuando un par de altavoces es simétrico, los signos numéricos de los elementos de matriz del altavoz izquierdo del par de altavoces simétricos no están codificados, y la unidad de extracción 72 obtiene estos signos numéricos a partir de los signos numéricos de los elementos de matriz asociados con el altavoz derecho mediante el empleo de la función auxiliar `createSymSigns`, que realiza lo siguiente:

```
pairIdx = outputConfig[j].symmetricPair->originalPosition;
```

```
signMatrix[i * outputCount + j] = symSigns[i] * signMatrix[i * outputCount + pairIdx];
```

**[0119]** La FIG. 11 es un diagrama que ilustra la señalización de las propiedades de simetría. Un par de altavoces no se puede definir como un valor simétrico y un signo simétrico al mismo tiempo. El indicador de decodificación final `hasVerticalCoef` especificado si solo los elementos de la matriz asociados con los coeficientes HOA circulares (es decir, 2D) están codificados. Si `hasVerticalCoef` no está configurado, los elementos de matriz asociados con los coeficientes HOA definidos con la función auxiliar `create2dBitmask` se configuran como cero digital.

**[0120]** Es decir, la unidad de extracción 72 extrae la información de renderización de audio 2 de acuerdo con el proceso expuesto en la FIG. 11. La unidad de extracción 72 puede leer primero el elemento sintáctico



isAllValueSymmetric del flujo de bits 21 (300). Cuando el elemento sintáctico isAllValueSymmetric se establece en uno (o, en otras palabras, un verdadero booleano), la unidad de extracción 72 puede iterar a través del valor del elemento sintáctico numPairs, estableciendo el elemento sintáctico de matriz valueSymmetricPairs en un valor de uno (indicando efectivamente que todos los pares de altavoces tienen un valor simétrico) (302).

**[0121]** Cuando el elemento sintáctico isAllValueSymmetric se pone a cero (o, en otras palabras, un falso booleano), la unidad de extracción 72 luego lee el elemento sintáctico isAnyValueSymmetric (304). Cuando el elemento sintáctico isAnyValueSymmetric se establece en uno (o, en otras palabras, un verdadero booleano), la unidad de extracción 72 itera a través del valor del elemento sintáctico numPairs, estableciendo el elemento sintáctico de matriz valueSymmetricPairs en un bit leído secuencialmente del flujo de bits 21 (306). La unidad de extracción 72 también obtiene el elemento sintáctico isAnySignSymmetric para cualquiera de los pares que tienen un elemento sintáctico valueSymmetricPairs establecido en cero (308). La unidad de extracción 72 a continuación vuelve a iterar a través del número de pares y, cuando el valueSymmetricPairs es igual a cero, establece un bit de SignSymmetricPairs en un valor leído del flujo de bits 21 (310).

**[0122]** Cuando el elemento sintáctico isAnyValueSymmetric se establece en cero (o, en otras palabras, un booleano falso), la unidad de extracción 72 lee el elemento sintáctico isAllSignSymmetric del flujo de bits 21 (312). Cuando el elemento sintáctico isAllSignSymmetric se establece en un valor de uno (o, en otras palabras, un verdadero booleano), la unidad de extracción 72 itera a través del valor del elemento sintáctico numPairs, estableciendo el elemento sintáctico de la matriz signSymmetricPairs en un valor de uno (indicando de manera efectiva que todos los pares de altavoces tienen signo simétrico) (316).

**[0123]** Cuando el elemento sintáctico isAllSignSymmetric se pone a cero (o, en otras palabras, un falso booleano), la unidad de extracción 72 lee el elemento sintáctico isAnySignSymmetric del flujo de bits 21 (316). La unidad de extracción 72 itera a través del valor del elemento sintáctico numPairs, estableciendo el elemento sintáctico de matriz signSymmetricPairs en un bit leído secuencialmente del flujo de bits 21 (318). La unidad de generación de flujo de bits 42 realiza un proceso recíproco al descrito anteriormente con respecto a la unidad de extracción 72 para especificar la información de simetría de valor, la información de simetría de signo o una combinación de la información de simetría de valor y de signo.

**[0124]** La unidad de reconstrucción de renderizador 81 representa una unidad configurada para reconstruir un renderizador basándose en la información de renderización de audio 2. Es decir, usando las propiedades mencionadas anteriormente, la unidad de reconstrucción de renderización 81 lee una serie de valores de ganancia de elemento de matriz. Para leer el valor absoluto de ganancia, la unidad de reconstrucción de renderización 81 invoca la función DecodeGainValue(). La unidad de reconstrucción de renderización 81 invoca la función ReadRange() del índice del alfabeto para descodificar uniformemente los valores de ganancia. Cuando el valor de ganancia descodificado no es un cero digital, la unidad de reconstrucción de renderizador 81 lee además el valor de signo de número (según la Tabla a siguiente). Cuando el elemento de la matriz está asociado con un coeficiente HOA que se señaló como disperso (a través de isHoaCoefSparse), el indicador hasValue precede al gainValueIndex (consulte la Tabla b). Cuando el indicador hasValue es cero, este elemento se establece en cero digital y no se señalan gainValueIndex y el signo.

Tablas a y b - Ejemplos de sintaxis de flujo de bits para descodificar un elemento de matriz

bitfield	gainValueIndex	signo	bitfield	hasValue	gainValueIndex	signo
tamaño	alphabetSize	1	tamaño	1	alphabetSize	1
a)			b)			

**[0125]** Dependiendo de las propiedades de simetría especificadas para los pares de altavoces, la unidad de reconstrucción de renderizadores 81 obtiene los elementos de matriz asociados con el altavoz izquierdo del altavoz derecho. En este caso, la información de renderización de audio 2 en el flujo de bits 21 para descodificar un elemento de matriz para el altavoz izquierdo se reduce o se omite potencialmente por completo en consecuencia.

**[0126]** De esta manera, el dispositivo de descodificación de audio 24 determina la información de simetría para reducir el tamaño de la información de renderización de audio que debe especificarse. En algunos casos, el dispositivo de descodificación de audio 24 determina la información de simetría para reducir el tamaño de la información de renderización de audio que debe especificarse, y obtiene al menos una parte del renderizador de audio basándose en la información de simetría.

**[0127]** En estos y otros casos, el dispositivo de descodificación de audio 24 puede determinar la información de simetría de valor para reducir un tamaño de la información de renderización de audio que debe especificarse. En estos y otros casos, el dispositivo de descodificación de audio 24 puede obtener al menos una parte del renderizador de audio basándose en la información de simetría de valor.

**[0128]** En estos y otros casos, el dispositivo de descodificación de audio 24 determina la información de simetría de la señal para reducir un tamaño de la información de renderización de audio que se va a especificar. En estos y otros casos, el dispositivo de descodificación de audio 24 obtiene al menos una parte del renderizador de audio basándose en la información de simetría de signo.

**[0129]** En estos y otros casos, el dispositivo de descodificación de audio 24 determina la información de dispersión indicativa de una dispersión de una matriz utilizada para renderizar coeficientes armónicos esféricos en una pluralidad de alimentaciones de altavoces.

**[0130]** En estos y otros casos, el dispositivo de descodificación de audio 24 puede determinar un diseño de altavoces para el cual se va a usar una matriz para renderizar coeficientes armónicos esféricos en una pluralidad de alimentaciones de altavoces.

**[0131]** El dispositivo de descodificación de audio 24 puede, a este respecto, a continuación determinar la información de renderización de audio 2 especificada en el flujo de bits. Basándose en el valor de la señal incluido en la información de renderización de audio 2, el sistema de reproducción de audio 16 puede renderizar una pluralidad de alimentaciones de altavoces 25 usando uno de los renderizadores de audio 22. Las alimentaciones de altavoces pueden activar los altavoces 3. Como se indicó anteriormente, el valor de la señal puede incluir en algunos casos una matriz (que se descodifica y proporciona como uno de los renderizadores de audio 22) utilizada para renderizar los coeficientes armónicos esféricos en una pluralidad de alimentaciones de altavoces. En este caso, el sistema de reproducción de audio 16 puede configurar uno de los renderizadores de audio 22 con la matriz, usando este uno de los renderizadores de audio 22 para renderizar las alimentaciones de altavoces 25 basándose en la matriz.

**[0132]** Para extraer y luego descodificar las diversas versiones codificadas de los coeficientes HOA 11 de manera que los coeficientes HOA estén disponibles para ser renderizados usando el renderizador de audio 22 obtenido, la unidad de extracción 72 puede determinar a partir del elemento sintáctico indicado anteriormente si los coeficientes HOA 11 fueron codificados a través de varias versiones basadas en dirección o basadas en vectores. Cuando se realizó una codificación basada en la dirección, la unidad de extracción 72 puede extraer la versión basada en la dirección de los coeficientes HOA 11 y los elementos sintácticos asociados a la versión codificada (que se indica como información basada en dirección 91 en el ejemplo de la FIG. 4), pasando la información basada en dirección 91 a la unidad de reconstrucción basada en dirección 90. La unidad de reconstrucción basada en la dirección 90 puede representar una unidad configurada para reconstruir los coeficientes HOA en forma de coeficientes HOA 11', basándose en la información basada en la dirección 91.

**[0133]** Cuando el elemento sintáctico indica que los coeficientes HOA 11 se codificaron usando una descomposición basada en vectores, la unidad de extracción 72 puede extraer los vectores codificados  $V[k]$  de primer plano 57 (que pueden incluir ponderaciones codificadas 57 y/o índices 63 o vectores  $V$  con cuantificación escalar), los coeficientes HOA ambientales codificados 59 y los correspondientes objetos de audio 61 (que también pueden denominarse las señales codificadas de nFG 61). Cada uno de los objetos de audio 61 corresponden cada uno a uno de los vectores 57. La unidad de extracción 72 puede pasar los vectores codificados  $V[k]$  de primer plano 57 a la unidad de reconstrucción de vector  $V$  74 y los coeficientes HOA ambientales codificados 59, junto con las nFG señales codificadas 61, a la unidad de descodificación psicoacústica 80.

**[0134]** La unidad de reconstrucción de vector  $V$  74 puede representar una unidad configurada para reconstruir los vectores  $V$  a partir de los vectores  $V[k]$  de primer plano codificados. La unidad de reconstrucción de vector  $V$  74 puede funcionar de una manera recíproca a la de la unidad de cuantificación 52.

**[0135]** La unidad de descodificación psicoacústica 80 puede funcionar de una manera recíproca a la unidad de codificador de audio psicoacústica 40 mostrada en el ejemplo de la FIG. 3 para descodificar los coeficientes HOA ambientales codificados 59 y las señales codificadas de nFG 61 y generar de ese modo coeficientes HOA ambientales compensados de energía 47' y las señales nFG interpoladas 49' (que también pueden denominarse objetos de audio nFG interpolados 49'). La unidad de descodificación psicoacústica 80 puede pasar los coeficientes HOA ambientales compensados en energía 47' a la unidad de fundido 770 y las nFG señales 49' a la unidad de formulación de primer plano 78.

**[0136]** La unidad de interpolación espacio-temporal 76 puede funcionar de una manera similar a la descrita anteriormente con respecto a la unidad de interpolación espacio-temporal 50. La unidad de interpolación espacio-temporal 76 puede recibir los vectores  $V[k]$  reducidos de primer plano  $55_k$  y realizar la interpolación espacio-temporal con respecto a los vectores  $V[k]$  de primer plano  $55_k$  y los vectores  $V[k-1]$  reducidos de primer plano  $55_{k-1}$  para generar vectores  $V[k]$  de primer plano interpolados  $55_k$ . La unidad de interpolación espacio-temporal 76 puede remitir los vectores  $V[k]$  de primer plano interpolados  $55_k$  a la unidad de fundido 770.

**[0137]** La unidad de extracción 72 también puede emitir una señal 757 indicativa de cuándo uno de los coeficientes HOA ambientales está en transición hacia la unidad de fundido 770, que puede entonces determinar

cuál de los  $\text{SHC}_{\text{BG}}$  47' (donde los  $\text{SHC}_{\text{BG}}$  47' también se pueden indicar como "canales de HOA ambientales 47'" o "coeficientes HOA ambientales 47'") y los elementos de los vectores  $V[k]$  de primer plano interpolados  $55_k$ " han de realizar un fundido inicial o fundido final. En algunos ejemplos, la unidad de fundido 770 puede funcionar opuesta con respecto a cada uno de los coeficientes HOA ambientales 47' y los elementos de los vectores  $V[k]$  de primer plano interpolados  $55_k$ ". Es decir, la unidad de fundido 770 puede realizar un fundido inicial o un fundido final, o tanto un fundido inicial como un fundido final, con respecto al correspondiente coeficiente HOA ambiental 47', mientras realiza un fundido inicial o un fundido final, o tanto un fundido inicial como un fundido final, con respecto al correspondiente de los elementos de los vectores  $V[k]$  de primer plano interpolados  $55_k$ ". La unidad de fundido 770 puede emitir coeficientes HOA ambientales ajustados 47'" a la unidad de formulación de coeficientes HOA 82 y los vectores  $V[k]$  de primer plano ajustados  $55_k$ " a la unidad de formulación de primer plano 78. A este respecto, la unidad de fundido 770 representa una unidad configurada para realizar una operación de fundido con respecto a diversos aspectos de los coeficientes HOA o sus derivados, por ejemplo, en forma de los coeficientes HOA ambientales 47' y los elementos de los vectores  $V[k]$  interpolados de primer plano  $55_k$ ".

**[0138]** La unidad de formulación del primer plano 78 puede representar una unidad configurada para realizar la multiplicación matricial con respecto a los vectores  $V[k]$  ajustados de primer plano  $55_k$ " y las nFG señales interpoladas 49' para generar los coeficientes HOA de primer plano 65. A este respecto, la unidad de formulación de primer plano 78 puede combinar los objetos de audio 49' (que es otra manera mediante la cual denotar las señales de nFG interpoladas 49') con los vectores  $55_k$ " para reconstruir el primer plano o, en otras palabras, los aspectos predominantes de los coeficientes HOA 11'. La unidad de formulación de primer plano 78 puede realizar una multiplicación matricial de las nFG señales interpoladas 49' por los vectores  $V[k]$  ajustados de primer plano  $55_k$ ".

**[0139]** La unidad de formulación de coeficientes HOA 82 puede representar una unidad configurada para combinar los coeficientes HOA de primer plano 65 a los coeficientes HOA ambientales 47'" para obtener los coeficientes HOA 11'. La notación principal refleja que los coeficientes HOA 11' pueden ser similares pero no iguales a los coeficientes HOA 11. Las diferencias entre los coeficientes HOA 11 y 11' pueden ser el resultado de pérdidas debidas a la transmisión por un medio de transmisión con pérdidas, cuantificación u otras operaciones con pérdidas.

**[0140]** Adicionalmente, la unidad de extracción 72 y el dispositivo de descodificación de audio 24 también pueden estar configurados en general para funcionar de acuerdo con diversos aspectos de las técnicas descritas en esta divulgación para obtener los flujos de bits 21 que están potencialmente optimizados de la forma descrita anteriormente con respecto a no incluir varios elementos sintácticos o campos de datos en ciertas instancias.

**[0141]** En algunos casos, el dispositivo de descodificación de audio 24 puede configurarse para que, al descomprimir datos de audio ambisónicos de orden superior comprimidos usando un primer esquema de compresión, obtener un flujo de bits 21 representativo de una versión comprimida de los datos de audio ambisónicos de orden superior que no incluye bits correspondientes a un segundo esquema de compresión que también se usa para comprimir los datos de audio ambisónicos de orden superior. El primer esquema de compresión puede comprender un esquema de compresión basado en vectores, con el vector resultante definido en el dominio armónico esférico y enviado a través del flujo de bits 21. El esquema de compresión de descomposición basado en vectores puede comprender, en algunos ejemplos, un esquema de compresión que implica la aplicación de una descomposición de valores singulares (o equivalentes de los mismos como se describe con mayor detalle con respecto al ejemplo de la FIG. 3) a los datos de audio ambisónicos de orden superior.

**[0142]** El dispositivo de descodificación de audio 24 puede estar configurado para obtener el flujo de bits 21 que no incluye los bits que corresponden a al menos un elemento sintáctico utilizado para realizar el segundo tipo de esquema de compresión. Como se indicó anteriormente, el segundo esquema de compresión comprende un esquema de compresión basado en la direccionalidad. Más específicamente, el dispositivo de descodificación de audio 24 puede configurarse para obtener el flujo de bits 21 que no incluye los bits correspondientes a los elementos sintácticos HOAPredictionInfo del segundo esquema de compresión. En otras palabras, cuando el segundo esquema de compresión comprende un esquema de compresión basado en la direccionalidad, el dispositivo de descodificación de audio 24 puede configurarse para obtener el flujo de bits 21 que no incluye los bits correspondientes a un elemento sintáctico HOAPredictionInfo del esquema de compresión basado en la direccionalidad. Como se indicó anteriormente, el elemento sintáctico HOAPredictionInfo puede ser indicativo de una predicción entre dos o más señales basadas en la dirección.

**[0143]** En algunos casos, de forma alternativa o conjuntamente con los ejemplos anteriores, el dispositivo de descodificación de audio 24 puede configurarse para, cuando se suprime la corrección de ganancia durante la compresión de datos de audio ambisónicos de orden superior, obtener el flujo de bits 21 representativo de una versión comprimida de los datos de audio ambisónicos de orden superior que no incluyen datos de corrección de ganancia. El dispositivo de descodificación de audio 24 puede, en estos casos, estar configurado para descomprimir los datos de audio ambisónicos de orden superior de acuerdo con un esquema de descompresión de síntesis basado en vectores. La versión comprimida de los datos ambisónicos de orden superior se genera mediante la aplicación de una descomposición de valores singulares (o equivalentes de los mismos descritos con mayor detalle con respecto al ejemplo de la FIG. 3 anterior) a los datos de audio ambisónicos de orden superior. Cuando se aplica SVD o sus equivalentes a los datos de audio HOA, el dispositivo de codificación de audio 20 especifica al menos

uno de los vectores resultantes o bits indicativos del mismo en el flujo de bits 21, donde los vectores describen características espaciales de los correspondientes objetos de audio de primer plano (tales como ancho, ubicación y volumen de los correspondientes objetos de audio de primer plano).

**[0144]** Más específicamente, el dispositivo de descodificación de audio 24 puede configurarse para obtener un elemento sintáctico MaxGainCorrAmbExp del flujo de bits 21 con un valor establecido en cero para indicar que se suprime la corrección de ganancia. Es decir, el dispositivo de descodificación de audio 24 puede configurarse para obtener, cuando se suprime la corrección de ganancia, el flujo de bits de manera que el flujo de bits no incluye un campo de datos HOAGainCorrection que almacena los datos de corrección de ganancia. El flujo de bits 21 puede comprender un elemento sintáctico MaxGainCorrAmbExp que tiene un valor de cero para indicar que la corrección de ganancia está suprimida y no incluye un campo de datos HOAGainCorrection que almacena los datos de corrección de ganancia. La supresión de la corrección de ganancia puede ocurrir cuando la compresión de los datos de audio ambisónicos de orden superior incluye la aplicación de una codificación de voz y voz y audio unificadas (USAC) a los datos de audio ambisónicos de orden superior.

**[0145]** La FIG. 5 es un diagrama de flujo que ilustra el funcionamiento a modo de ejemplo de un dispositivo de codificación de audio, tal como el dispositivo de codificación de audio 20 mostrado en el ejemplo de la FIG. 3, en la realización de diversos aspectos de las técnicas de síntesis basadas en vectores descritas en esta divulgación. Inicialmente, el dispositivo de codificación de audio 20 recibe los coeficientes HOA 11 (106). El dispositivo de codificación de audio 20 puede invocar la unidad de LIT 30, que puede aplicar una LIT con respecto a los coeficientes HOA para emitir los coeficientes HOA transformados (por ejemplo, en el caso de la SVD, los coeficientes HOA transformados pueden comprender los vectores  $US[k]$  33 y los vectores  $V[k]$  35) (107).

**[0146]** El dispositivo de codificación de audio 20 puede invocar, a continuación, la unidad de cálculo de parámetros 32 para realizar el análisis descrito anteriormente con respecto a cualquier combinación de los vectores  $US[k]$  33, los vectores  $US[k-1]$  33, los vectores  $V[k]$  y/o los vectores  $V[k-1]$  35, para identificar varios parámetros de la manera descrita anteriormente. Es decir, la unidad de cálculo de parámetros 32 puede determinar al menos un parámetro basándose en un análisis de los coeficientes HOA transformados 33/35 (108).

**[0147]** El dispositivo de codificación de audio 20 puede entonces invocar la unidad de reordenamiento 34, que puede reordenar los coeficientes HOA transformados (que, de nuevo en el contexto de la SVD, pueden referirse a los vectores  $US[k]$  33 y a los vectores  $V[k]$  35) basándose en el parámetro para generar los coeficientes HOA transformados reordenados 33'/35' (o, en otras palabras, los vectores  $US[k]$  33' y los vectores  $V[k]$  35'), como se ha descrito anteriormente (109). El dispositivo de codificación de audio 20 puede, durante cualquiera de las operaciones anteriores u operaciones subsiguientes, invocar también la unidad de análisis del campo sonoro 44. La unidad de análisis del campo sonoro 44 puede, como se ha descrito anteriormente, realizar un análisis de campo sonoro con respecto a los coeficientes HOA 11 y/o los coeficientes HOA transformados 33/35 para determinar el número total de canales de primer plano ( $n_{FG}$ ) 45, el orden del campo sonoro de fondo ( $N_{BG}$ ) y el número ( $n_{BGa}$ ) y los índices ( $i$ ) de canales adicionales HOA de BG a enviar (que pueden indicarse colectivamente como información de canal de fondo 43 en el ejemplo de la FIG. 3) (109).

**[0148]** El dispositivo de codificación de audio 20 también puede invocar la unidad de selección de fondo 48. La unidad de selección de fondo 48 puede determinar coeficientes HOA de fondo o HOA ambientales 47 basándose en la información de canal de fondo 43 (110). El dispositivo de codificación de audio 20 puede invocar adicionalmente la unidad de selección del primer plano 36, la cual puede seleccionar los vectores  $US[k]$  reordenados 33' y los vectores  $V[k]$  reordenados 35 que representan componentes de primer plano o distintos del campo sonoro basándose en  $n_{FG}$  45 (que puede representar uno o más índices que identifican los vectores de primer plano) (112).

**[0149]** El dispositivo de codificación de audio 20 puede invocar la unidad de compensación de energía 38. La unidad de compensación de energía 38 puede realizar compensación de energía con respecto a los coeficientes HOA ambientales 47 para compensar la pérdida de energía debida a la eliminación de varios de los coeficientes HOA por la unidad de selección de fondo 48 (114) y generar así coeficientes HOA ambientales compensados en energía 47'.

**[0150]** El dispositivo de codificación de audio 20 puede también invocar la unidad de interpolación espaciotemporal 50. La unidad de interpolación espaciotemporal 50 puede realizar interpolación espaciotemporal con respecto a los coeficientes HOA transformados reordenados 33'/35' para obtener las señales de primer plano interpoladas 49' (que también pueden denominarse las "señales  $n_{FG}$  interpoladas 49'") y la información direccional restante de primer plano 53 (que también se puede denominar los "vectores  $V[k]$  53") (116). El dispositivo de codificación de audio 20 puede entonces invocar la unidad de reducción de coeficientes 46. La unidad de reducción de coeficientes 46 puede realizar una reducción de los coeficientes con respecto a los restantes vectores  $V[k]$  de primer plano, basándose en la información de canal de fondo 43 para obtener información direccional reducida de primer plano 55 (que también se puede mencionar como vectores de primer plano reducidos  $V[k]$  55) (118).

**[0151]** El dispositivo de codificación de audio 20 puede entonces invocar la unidad de cuantificación 52 para comprimir, de la manera descrita anteriormente, los vectores  $V[k]$  de primer plano reducidos 55 y generar los

vectores  $V[k]$  de primer plano codificados 57.

[0152] El dispositivo de codificación de audio 20 también puede invocar la unidad codificadora de audio psicoacústico 40. La unidad codificadora de audio psicoacústico 40 puede codificar psicoacústicamente cada vector de los coeficientes HOA ambientales compensados en energía 47' y las nFG señales interpoladas 49' para generar coeficientes HOA ambientales codificados 59 y nFG señales codificadas 61. El dispositivo de codificación de audio puede entonces invocar la unidad de generación de flujo de bits 42. La unidad de generación del flujo de bits 42 puede generar el flujo de bits 21 basándose en la información direccional codificada de primer plano 57, los coeficientes HOA ambientales codificados 59, las nFG señales codificadas 61 y la información de canal de fondo 43.

[0153] La FIG. 6 es un diagrama de flujo que ilustra el funcionamiento a modo de ejemplo de un dispositivo de descodificación de audio, como el dispositivo de descodificación de audio 24 mostrado en la FIG. 4, en la realización de varios aspectos de las técnicas descritas en esta divulgación. Inicialmente, el dispositivo descodificador de audio 24 puede recibir el flujo de bits 21 (130). Al recibir el flujo de bits, el dispositivo de descodificación de audio 24 puede invocar la unidad de extracción 72. Suponiendo, a efectos de análisis, que el flujo de bits 21 indica que se va a realizar una reconstrucción basada en vectores, la unidad de extracción 72 puede analizar el flujo de bits para recuperar la información indicada anteriormente, pasando la información a la unidad de reconstrucción basada en vectores 92.

[0154] En otras palabras, la unidad de extracción 72 puede extraer la información direccional codificada de primer plano 57 (que, de nuevo, también se puede mencionar como los vectores  $V[k]$  de primer plano codificados 57), los coeficientes HOA ambientales codificados 59 y las señales de primer plano codificadas (que también se pueden denominar las nFG señales de primer plano codificadas 59 o los objetos de audio de primer plano codificados 59) del flujo de bits 21, de la manera descrita anteriormente (132).

[0155] El dispositivo de descodificación de audio 24 puede invocar además la unidad de descuantificación 74. La unidad de descuantificación 74 puede descodificar por entropía y descuantificar la información direccional codificada de primer plano 57 para obtener información direccional reducida de primer plano  $55_k$  (136). El dispositivo de descodificación de audio 24 también puede invocar la unidad de descodificación psicoacústica 80. La unidad de codificación de audio psicoacústico 80 puede descodificar los coeficientes HOA ambientales codificados 59 y las señales de primer plano codificadas 61 para obtener coeficientes HOA ambientales compensados en energía 47' y las señales de primer plano interpoladas 49' (138). La unidad de descodificación psicoacústica 80 puede pasar los coeficientes HOA ambientales compensados en energía 47' a la unidad de fundido 770 y las nFG señales 49' a la unidad de formulación de primer plano 78.

[0156] El dispositivo de descodificación de audio 24 puede invocar a continuación la unidad de interpolación espacio-temporal 76. La unidad de interpolación espacio-temporal 76 puede recibir la información direccional de primer plano reordenada  $55_{k'}$  y realizar la interpolación espacio-temporal con respecto a la información direccional reducida de primer plano  $55_{k'}/55_{k-1}$  para generar la información direccional de primer plano interpolada  $55_k$  (140). La unidad de interpolación espacio-temporal 76 puede remitir los vectores  $V[k]$  de primer plano interpolados  $55_k$  a la unidad de fundido 770.

[0157] El dispositivo de descodificación de audio 24 puede invocar la unidad de fundido 770. La unidad de fundido 770 puede recibir u obtener de otro modo elementos sintácticos (por ejemplo, desde la unidad de extracción 72) indicativos de cuándo los coeficientes HOA ambientales compensados en energía 47' están en transición (por ejemplo, el elemento sintáctico AmbCoefTransition). La unidad de fundido 770 puede, basándose en los elementos sintácticos de transición y la información de estado de transición mantenida, incorporar o eliminar los coeficientes HOA ambientales compensados en energía 47', emitiendo coeficientes HOA ambientales ajustados 47" a la unidad de formulación de coeficientes HOA 82. La unidad de fundido 770 puede también, basándose en los elementos sintácticos y la información de estado de transición mantenida, realizar un fundido final o un fundido inicial de los correspondientes uno o más elementos de los vectores de primer plano interpolados  $V[k]$   $55_k$ ", emitiendo los vectores  $V[k]$  ajustados de primer plano  $55_k$ " a la unidad de formulación de primer plano 78 (142).

[0158] El dispositivo de descodificación de audio 24 puede invocar la unidad de formulación del primer plano 78. La unidad de formulación del primer plano 78 puede realizar la multiplicación matricial de las nFG señales 49' por la información direccional de primer plano ajustada  $55_k$ " para obtener los coeficientes HOA de primer plano 65 (144). El dispositivo de descodificación de audio 24 también puede invocar la unidad de formulación de coeficientes HOA 82. La unidad de formulación de coeficientes HOA 82 puede añadir los coeficientes HOA de primer plano 65 a los coeficientes HOA ambientales ajustados 47' para obtener los coeficientes HOA 11' (146).

[0159] La FIG. 7 es un diagrama de flujo que ilustra el funcionamiento de ejemplo de un sistema, tal como el sistema 10 mostrado en el ejemplo de la FIG. 2, en la realización de diversos aspectos de las técnicas descritas en esta divulgación. Como se analizó anteriormente, el dispositivo creador de contenido 12 puede emplear un sistema de edición de audio 18 para crear o editar contenido de audio capturado o generado (que se muestra como los coeficientes HOA 11 en el ejemplo de la FIG. 2). El dispositivo creador de contenido 12 puede renderizar entonces los coeficientes HOA 11 usando el renderizador de audio 1 para las alimentaciones de altavoces multicanal

generadas, como se ha analizado con mayor detalle anteriormente (200). El dispositivo creador de contenido 12 puede entonces reproducir estas alimentaciones de altavoces usando un sistema de reproducción de audio y determinar si se requieren ajustes o ediciones adicionales para capturar, por ejemplo, la intención artística deseada (202). Cuando se desean ajustes adicionales ("SÍ" 202), el dispositivo creador de contenido 12 puede remezclar los coeficientes HOA 11 (204), renderizar los coeficientes HOA 11 (200) y determinar si son necesarios ajustes adicionales (202). Cuando no se desean más ajustes ("NO" 202), el dispositivo de codificación de audio 20 puede codificar el contenido de audio para generar el flujo de bits 21 de la manera descrita anteriormente con respecto al ejemplo de la FIG. 5 (206). El dispositivo de codificación de audio 20 también puede generar y especificar la información de renderización de audio 2 en el flujo de bits 21, como se ha descrito con mayor detalle anteriormente (208).

**[0160]** El dispositivo consumidor de contenido 14 puede entonces obtener la información de renderización de audio 2 del flujo de bits 21 (210). El dispositivo de descodificación 24 puede entonces descodificar el flujo de bits 21 para obtener el contenido de audio (que se muestra como los coeficientes HOA 11' en el ejemplo de la FIG. 2) de la manera descrita anteriormente con respecto al ejemplo de la FIG. 6 (211). El sistema de reproducción de audio 16 puede entonces renderizar los coeficientes HOA 11' basándose en la información de renderización de audio 2 de la manera descrita anteriormente (212) y reproducir el contenido de audio renderizado a través de los altavoces 3 (214).

**[0161]** Las técnicas descritas en esta divulgación pueden por lo tanto permitir, como primer ejemplo, un dispositivo que genera un flujo de bits representativo de contenido de audio multicanal para especificar información de renderización de audio. El dispositivo puede, en este primer ejemplo, incluir medios para especificar información de renderización de audio que incluye un valor de la señal que identifica un renderizador de audio utilizado cuando se genera el contenido de audio multicanal.

**[0162]** El dispositivo del primer ejemplo, en el que el valor de la señal incluye una matriz utilizada para renderizar coeficientes armónicos esféricos en una pluralidad de alimentaciones de altavoces.

**[0163]** En un segundo ejemplo, el dispositivo del primer ejemplo, en el que el valor de la señal incluye dos o más bits que definen un índice que indica que el flujo de bits incluye una matriz utilizada para renderizar coeficientes armónicos esféricos en una pluralidad de alimentaciones de altavoces.

**[0164]** El dispositivo del segundo ejemplo, en el que la información de renderización de audio incluye además dos o más bits que definen un número de filas de la matriz incluidas en el flujo de bits y dos o más bits que definen un número de columnas de la matriz incluidas en el flujo de bits.

**[0165]** El dispositivo del primer ejemplo, en el que el valor de la señal especifica un algoritmo de renderización usado para renderizar objetos de audio en una pluralidad de alimentaciones de altavoces.

**[0166]** El dispositivo del primer ejemplo, en el que el valor de la señal especifica un algoritmo de renderización usado para renderizar coeficientes armónicos esféricos en una pluralidad de alimentaciones de altavoces.

**[0167]** El dispositivo del primer ejemplo, en el que el valor de la señal incluye dos o más bits que definen un índice asociado con una de una pluralidad de matrices usadas para renderizar coeficientes armónicos esféricos en una pluralidad de alimentaciones de altavoces.

**[0168]** El dispositivo del primer ejemplo, en el que el valor de la señal incluye dos o más bits que definen un índice asociado con uno de una pluralidad de algoritmos de renderización utilizados para renderizar objetos de audio en una pluralidad de alimentaciones de altavoces.

**[0169]** El dispositivo del primer ejemplo, en el que el valor de la señal incluye dos o más bits que definen un índice asociado con uno de una pluralidad de algoritmos de renderización usados para renderizar coeficientes armónicos esféricos en una pluralidad de alimentaciones de altavoces.

**[0170]** El dispositivo del primer ejemplo, en el que los medios para especificar la información de renderización de audio comprenden medios para especificar la información de renderización de audio por cada trama de audio en el flujo de bits.

**[0171]** El dispositivo del primer ejemplo, en el que los medios para especificar la información de renderización de audio comprenden medios para especificar la información de renderización de audio una única vez en el flujo de bits.

**[0172]** En un tercer ejemplo, un medio de almacenamiento no transitorio legible por ordenador que tiene almacenadas en el mismo instrucciones que, cuando se ejecutan, hacen que uno o más procesadores especifiquen información de renderización de audio en el flujo de bits, en el que la información de renderización de audio identifica un renderizador de audio utilizado al generar el contenido de audio multicanal.

**[0173]** En un cuarto ejemplo, un dispositivo para renderizar contenido de audio multicanal a partir de un flujo de bits, comprendiendo el dispositivo medios para determinar información de renderización de audio que incluye un valor de la señal que identifica un renderizador de audio utilizado al generar el contenido de audio multicanal y medios para renderizar una pluralidad de alimentaciones de altavoces basándose en la información de renderización de audio especificada en el flujo de bits.

**[0174]** El dispositivo del cuarto ejemplo, en el que el valor de la señal incluye una matriz utilizada para renderizar coeficientes armónicos esféricos en una pluralidad de alimentaciones de altavoces, y en el que los medios para renderizar la pluralidad de alimentaciones de altavoces comprenden medios para renderizar la pluralidad de alimentaciones de altavoces basándose en matriz.

**[0175]** En un quinto ejemplo, el dispositivo del cuarto ejemplo, en el que el valor de la señal incluye dos o más bits que definen un índice que indica que el flujo de bits incluye una matriz utilizada para renderizar coeficientes armónicos esféricos en una pluralidad de alimentaciones de altavoces, en el que el dispositivo comprende además medios para analizar la matriz desde el flujo de bits en respuesta al índice, y en el que los medios para renderizar la pluralidad de alimentaciones de altavoces comprenden medios para renderizar la pluralidad de alimentaciones de altavoces basándose en la matriz analizada.

**[0176]** El dispositivo del quinto ejemplo, en el que el valor de la señal incluye además dos o más bits que definen un número de filas de la matriz incluidas en el flujo de bits y dos o más bits que definen un número de columnas de la matriz incluidas en el flujo de bits, y en el que los medios para analizar la matriz desde el flujo de bits comprenden medios para analizar la matriz desde el flujo de bits en respuesta al índice y basándose en los dos o más bits que definen un número de filas y los dos o más bits que definen el número de columnas.

**[0177]** El dispositivo del cuarto ejemplo, en el que el valor de la señal especifica un algoritmo de renderización usado para renderizar objetos de audio en la pluralidad de alimentaciones de altavoces, y en el que los medios para renderizar la pluralidad de alimentaciones de altavoces comprenden medios para renderizar la pluralidad de alimentaciones de altavoces de los objetos de audio utilizando el algoritmo de renderización especificado.

**[0178]** El dispositivo del cuarto ejemplo, en el que el valor de la señal especifica un algoritmo de renderización usado para renderizar coeficientes armónicos esféricos en la pluralidad de alimentaciones de altavoces, y en el que los medios para renderizar la pluralidad de alimentaciones de altavoces comprenden medios para renderizar la pluralidad de alimentaciones de altavoces desde los coeficientes armónicos esféricos utilizando el algoritmo de renderización especificado.

**[0179]** El dispositivo del cuarto ejemplo, en el que el valor de la señal incluye dos o más bits que definen un índice asociado con una de una pluralidad de matrices usadas para renderizar coeficientes armónicos esféricos en la pluralidad de alimentaciones de altavoces, y en el que los medios para renderizar la pluralidad de las alimentaciones de altavoces comprenden medios para renderizar la pluralidad de alimentaciones de altavoces a partir de los coeficientes armónicos esféricos usando una de la pluralidad de matrices asociadas con el índice.

**[0180]** El dispositivo del cuarto ejemplo, en el que el valor de la señal incluye dos o más bits que definen un índice asociado con uno de una pluralidad de algoritmos de renderización usados para renderizar objetos de audio en la pluralidad de alimentaciones de altavoces, y en el que los medios para renderizar la pluralidad de las alimentaciones de altavoces comprenden medios para renderizar la pluralidad de alimentaciones de altavoces de los objetos de audio usando uno de la pluralidad de algoritmos de renderización asociados con el índice.

**[0181]** El dispositivo del cuarto ejemplo, en el que el valor de la señal incluye dos o más bits que definen un índice asociado con uno de una pluralidad de algoritmos de renderización usados para renderizar coeficientes armónicos esféricos en una pluralidad de alimentaciones de altavoces, y en el que los medios para renderizar la pluralidad de alimentaciones de altavoces comprenden medios para renderizar la pluralidad de alimentaciones de altavoces a partir de los coeficientes armónicos esféricos usando uno de la pluralidad de algoritmos de renderización asociados con el índice.

**[0182]** El dispositivo del cuarto ejemplo, en el que los medios para determinar la información de renderización de audio incluyen medios para determinar la información de renderización de audio por cada trama de audio a partir del flujo de bits.

**[0183]** El dispositivo del cuarto ejemplo, en el que los medios para determinar los medios de información de renderización de audio incluyen la determinación de la información de renderización de audio una sola vez desde el flujo de bits.

**[0184]** En un sexto ejemplo, un medio de almacenamiento no transitorio legible por ordenador que tiene almacenadas en el mismo instrucciones que, cuando se ejecutan, hacen que uno o más procesadores determinen información de renderización de audio que incluye un valor de la señal que identifica un renderizador de audio utilizado al generar contenido del audio multicanal; y renderizar una pluralidad de alimentaciones de altavoces

basándose en la información de renderización de audio especificada en el flujo de bits.

**[0185]** Las Figs. 8A-8D son diagramas que ilustran los flujos de bits 21A-21D formados de acuerdo con las técnicas descritas en esta divulgación. En el ejemplo de la FIG. 8A, el flujo de bits 21A puede representar un ejemplo del flujo de bits 21 mostrado en las FIGs. 2-4 anteriores. El flujo de bits 21A incluye información de renderización de audio 2A que incluye uno o más bits que definen un valor de la señal 554. Este valor de la señal 554 puede representar cualquier combinación de los tipos de información descritos a continuación. El flujo de bits 21A también incluye contenido de audio 558, que puede representar un ejemplo del contenido de audio 7/9.

**[0186]** En el ejemplo de la FIG. 8B, el flujo de bits 21B puede ser similar al flujo de bits 21A donde el valor de la señal 554 de la información de renderización de audio 2B comprende un índice 554A, uno o más bits que definen un tamaño de fila 554B de la matriz señalada, uno o más bits que definen un tamaño de columna 554C de la matriz señalada, y coeficientes de matriz 554D. El índice 554A se puede definir usando de dos a cinco bits, mientras que cada uno de los tamaños de fila 554B y de columna 554C puede definirse usando de dos a dieciséis bits.

**[0187]** La unidad de extracción 72 puede extraer el índice 554A y determinar si el índice indica que la matriz está incluida en el flujo de bits 21B (donde ciertos valores de índice, tales como 0000 o 1111, pueden indicar que la matriz está especificada explícitamente en el flujo de bits 21B). En el ejemplo de la FIG. 8B, el flujo de bits 21B incluye un índice 554A que indica que la matriz está especificada explícitamente en el flujo de bits 21B. Como resultado, la unidad de extracción 72 puede extraer el tamaño de fila 554B y el tamaño de columna 554C. La unidad de extracción 72 puede configurarse para calcular el número de bits a analizar que representan los coeficientes de la matriz como una función del tamaño de fila 554B, el tamaño de columna 554C y un tamaño de bit señalado (no mostrado en la FIG. 8A) o implícito de cada coeficiente de matriz. Usando el número determinado de bits, la unidad de extracción 72 puede extraer los coeficientes de matriz 554D, que el sistema de reproducción de audio 16 puede usar para configurar uno de los renderizadores de audio 22 como se describió anteriormente. Aunque se muestra como señalización de la información de renderización de audio 2B una sola vez en el flujo de bits 21B, la información de renderización de audio 2B puede señalarse múltiples veces en el flujo de bits 21B o al menos parcial o totalmente en un canal fuera de banda separado (como datos opcionales en algunas instancias).

**[0188]** En el ejemplo de la FIG. 8C, el flujo de bits 21C puede representar un ejemplo del flujo de bits 21 mostrado en las FIGs. 2-4 anteriores. El flujo de bits 21C incluye la información de renderización de audio 2C que incluye un valor de la señal 554, que en este ejemplo especifica un índice de algoritmo 554E. El flujo de bits 21C también incluye contenido de audio 558. El índice de algoritmo 554E se puede definir usando de dos a cinco bits, como se indicó anteriormente, donde este índice de algoritmo 554E puede identificar un algoritmo de renderización para ser utilizado cuando se renderiza el contenido de audio 558.

**[0189]** La unidad de extracción 72 puede extraer el índice de algoritmo 554E y determinar si el índice de algoritmo 554E indica que la matriz está incluida en el flujo de bits 21C (donde ciertos valores de índice, como 0000 o 1111, pueden indicar que la matriz está especificada explícitamente en el flujo de bits 21C). En el ejemplo de la FIG. 8C, el flujo de bits 21C incluye el índice de algoritmo 554E que indica que la matriz no está explícitamente especificada en el flujo de bits 21C. Como resultado, la unidad de extracción 72 remite el índice de algoritmo 554E al sistema de reproducción de audio 16, que selecciona el correspondiente (si está disponible) de los algoritmos de renderización (que se designan como renderizadores 22 en el ejemplo de las FIGs. 2-4). Mientras se muestra como información de renderización de audio de señalización 2C una sola vez en el flujo de bits 21C, en el ejemplo de la FIG. 8C, la información de renderización de audio 2C puede señalarse múltiples veces en el flujo de bits 21C o al menos parcial o totalmente en un canal separado fuera de banda (como datos opcionales en algunos casos).

**[0190]** En el ejemplo de la FIG. 8D, el flujo de bits 21D puede representar un ejemplo del flujo de bits 21 mostrado en las FIGs. 2-4 anteriores. El flujo de bits 21D incluye la información de renderización de audio 2D que incluye un valor de la señal 554, que en este ejemplo especifica un índice de matriz 554F. El flujo de bits 21D también incluye contenido de audio 558. El índice de matriz 554F se puede definir utilizando de dos a cinco bits, como se indicó anteriormente, donde este índice de matriz 554F puede identificar un algoritmo de renderización que se utilizará al representar el contenido de audio 558.

**[0191]** La unidad de extracción 72 puede extraer el índice de matriz 554F y determinar si el índice de matriz 554F indica que la matriz está incluida en el flujo de bits 21D (donde ciertos valores de índice, tales como 0000 o 1111, pueden indicar que la matriz está especificada explícitamente en el flujo de bits 21C). En el ejemplo de la FIG. 8D, el flujo de bits 21D incluye el índice de matriz 554F que indica que la matriz no está especificada explícitamente en el flujo de bits 21D. Como resultado, la unidad de extracción 72 remite el índice de matriz 554F al dispositivo de reproducción de audio, que selecciona el correspondiente (si está disponible) de los renderizador 22. Mientras que se muestra como señalización de la información de renderización de audio de 2D una sola vez en el flujo de bits 21D, en el ejemplo de la FIG. 8D, la información de renderización de audio 2D puede señalarse múltiples veces en el flujo de bits 21D o al menos parcial o totalmente en un canal fuera de banda separado (como datos opcionales en algunos casos).

**[0192]** Las FIGs. 8E-8G son diagramas que ilustran partes de la información del flujo de bits o del canal lateral,



que puede especificar los componentes espaciales comprimidos con mayor detalle. La FIG. 8E ilustra un primer ejemplo de una trama 249A' del flujo de bits 21. En el ejemplo de la FIG. 8E, la trama 249A' incluye los campos ChannelSideInfoData (CSID) 154A-154C, los campos HOAGain-CorrectionData (HOAGCD) y los campos VVectorData 156A y 156B. El campo CSID 154A incluye la unitC 267, bb 266 y ba265 junto con el TipoCanal 269, cada uno de los cuales está establecido en los valores correspondientes 01, 1,0 y 01 que se muestran en el ejemplo de la FIG. 8E. El campo CSID 154B incluye la unitC 267, bb 266 y ba265 junto con el TipoCanal 269, cada uno de los cuales está establecido en los valores correspondientes 01, 1,0 y 01 que se muestran en el ejemplo de la FIG. 8E. El campo CSID 154C incluye el campo TipoCanal 269 que tiene un valor de 3. Cada uno de los campos CSID 154A-154C corresponde al canal respectivo de los canales de transporte 1, 2 y 3. En efecto, cada campo CSID 154A-154C indica si la carga útil correspondiente 156A y 156B son señales basadas en la dirección (cuando el TipoCanal correspondiente es igual a cero), señales basadas en vectores (cuando el TipoCanal correspondiente es igual a uno), un coeficiente HOA ambiental adicional (cuando el TipoCanal correspondiente es igual a dos) o vacío (cuando el TipoCanal es igual a tres).

**[0193]** En el ejemplo de la FIG. 8E, la trama 249A incluye dos señales basadas en vectores (dado el TipoCanal 269 igual a 1 en los campos CSID 154A y 154B) y un vacío (dado que el TipoCanal 269 es igual a 3 en el campo CSID 154C). Basándose en una parte de HOAconfig anterior (no mostrada para facilitar la ilustración), el dispositivo de descodificación de audio 24 puede determinar que todos los elementos de vector de 16 V están codificados. Por tanto, los DatosVectorV 156A y 156B incluyen, cada uno, todos los 16 elementos vectoriales, cada uno de ellos uniformemente cuantificado con 8 bits.

**[0194]** Como se muestra adicionalmente en el ejemplo de la FIG. 8E, la trama 249A' no incluye un campo HOAPredictionInfo. El campo HOAPredictionInfo puede representar un campo correspondiente a un segundo esquema de compresión basado en la direccionalidad que puede eliminarse de acuerdo con la técnica descrita en esta divulgación cuando el esquema de compresión basado en vectores se usa para comprimir datos de audio HOA.

**[0195]** La FIG. 8F es un diagrama que ilustra una trama 249A" que es sustancialmente similar a la trama 249A excepto que el HOAGainCorrectionData se ha eliminado de cada canal de transporte almacenado en la trama 249A". El campo HOAGainCorrectionData puede eliminarse de la trama 249A" cuando se suprime la corrección de ganancia de acuerdo con diversos aspectos de las técnicas descritas anteriormente.

**[0196]** La FIG. 8G es un diagrama que ilustra una trama 249A'" que puede ser similar a la trama 249A" excepto que el campo HOAPredictionInfo se elimina. La trama 249A'" representa un ejemplo en el que ambos aspectos de las técnicas se pueden aplicar conjuntamente para eliminar varios campos que pueden no ser necesarios en ciertas circunstancias.

**[0197]** Las técnicas anteriores pueden realizarse con respecto a cualquier cantidad de contextos y ecosistemas de audio diferentes. A continuación se describen un número de contextos a modo de ejemplo, aunque las técnicas deberían limitarse a los contextos a modo de ejemplo. Un ecosistema a modo de ejemplo de audio puede incluir contenido de audio, estudios de cine, estudios de música, estudios de audio de juegos, contenido de audio basado en canales, motores de codificación, derivaciones de audio de juegos, motores de codificación/renderización y sistemas de suministro.

**[0198]** Los estudios de cine, los estudios de música y los estudios de audio de juegos pueden recibir contenido de audio. En algunos ejemplos, el contenido de audio puede representar el resultado de una adquisición. Los estudios de cine pueden emitir contenido de audio basado en canales (por ejemplo, en 2.0, 5.1 y 7.1), tal como utilizando una estación de trabajo de audio digital (DAW). Los estudios de música pueden emitir contenido de audio basado en canales (por ejemplo, en 2.0 y 5.1), tal como utilizando una DAW. En cualquier caso, los motores de codificación pueden recibir y codificar el contenido de audio basado en canales basándose en uno o más códecs (por ejemplo, AAC, AC3, Dolby True HD, Dolby Digital Plus y DTS Master Audio) para emitir por los sistemas de suministro. Los estudios de audio de juegos pueden emitir uno o más fragmentos de audio de juego, por ejemplo, mediante una DAW. Los motores de codificación/reproducción de audios de juego pueden codificar y/o reproducir las derivaciones de audio como contenido de audio basado en canales para su emisión por los sistemas de suministro. Otro ejemplo de contexto en el que se pueden realizar las técnicas comprende un ecosistema de audio que puede incluir grabación difundida de objetos de audio, sistemas de audio profesionales, captura de consumidores en el dispositivo, formato de audio HOA, renderización en el dispositivo, audio del consumidor, televisión y accesorios, y sistemas de audio para el automóvil.

**[0199]** Los objetos de audio de grabación difundida, los sistemas de audio profesionales 320 y la captura de consumidores en el dispositivo pueden codificar todos sus resultados utilizando el formato de audio HOA. De esta manera, el contenido de audio puede codificarse utilizando el formato de audio HOA en una representación única que puede reproducirse utilizando la reproducción en dispositivo, el audio de consumidor, la televisión y accesorios y los sistemas de audio para automóviles. En otras palabras, la representación única del contenido de audio se puede reproducir en un sistema genérico de reproducción de audio (es decir, en lugar de requerir una configuración particular tal como 5.1, 7.1, etc.), tal como un sistema de reproducción de audio 16.

**[0200]** Otros ejemplos de contexto en los que se pueden realizar las técnicas incluyen un ecosistema de audio que puede incluir elementos de adquisición y elementos de reproducción. Los elementos de adquisición pueden incluir dispositivos de adquisición alámbricos y/o inalámbricos (por ejemplo, micrófonos Eigen), captura de sonido envolvente en dispositivo y dispositivos móviles (por ejemplo, teléfonos inteligentes y tablets). En algunos ejemplos, los dispositivos de adquisición alámbricos y/o inalámbricos pueden estar acoplados al dispositivo móvil mediante uno o más canales de comunicación alámbricos y/o inalámbricos.

**[0201]** De acuerdo con una o más técnicas de esta divulgación, el dispositivo móvil puede usarse para adquirir un campo sonoro. Por ejemplo, el dispositivo móvil puede adquirir un campo sonoro mediante dispositivos de adquisición alámbricos y/o inalámbricos y/o la captura de sonido envolvente en el dispositivo (por ejemplo, una pluralidad de micrófonos integrados en el dispositivo móvil). El dispositivo móvil puede entonces codificar el campo de sonido adquirido en los coeficientes HOA para la reproducción por uno o más de los elementos de reproducción. Por ejemplo, un usuario del dispositivo móvil puede grabar (adquirir un campo sonoro de) un suceso en vivo (por ejemplo, una reunión, una conferencia, una obra de teatro, un concierto, etc.) y codificar la grabación en coeficientes HOA.

**[0202]** El dispositivo móvil también puede utilizar uno o más de los elementos de reproducción para reproducir el campo sonoro codificado por HOA. Por ejemplo, el dispositivo móvil puede descodificar el campo sonoro codificado por HOA y emitir una señal a uno o más de los elementos de reproducción, que hace que los uno o más de los elementos de reproducción recreen el campo sonoro. Como ejemplo, el dispositivo móvil puede utilizar los canales de comunicación inalámbricos y/o alámbricos para emitir la señal a uno o más altavoces (por ejemplo, matrices de altavoces, barras de sonido, etc.). Como otro ejemplo, el dispositivo móvil puede utilizar soluciones de acoplamiento para emitir la señal a una o más estaciones de acoplamiento y/o uno o más altavoces acoplados (por ejemplo, sistemas de sonido en coches inteligentes y/u hogares). Como otro ejemplo, el dispositivo móvil puede utilizar la reproducción de auriculares para emitir la señal a un conjunto de auriculares, por ejemplo, para crear un sonido binaural realista.

**[0203]** En algunos ejemplos, un dispositivo móvil particular puede adquirir un campo sonoro tridimensional y reproducir el mismo campo sonoro tridimensional en un momento posterior. En algunos ejemplos, el dispositivo móvil puede adquirir un campo sonoro tridimensional, codificar el campo sonoro tridimensional en HOA y transmitir el campo sonoro tridimensional codificado a otros uno o más dispositivos (por ejemplo, otros dispositivos móviles y/u otros dispositivos no móviles) para su reproducción.

**[0204]** Otro contexto más en el que las técnicas se pueden realizar incluye un ecosistema de audio que puede incluir contenido de audio, estudios de juegos, contenido de audio codificado, motores de reproducción y sistemas de suministro. En algunos ejemplos, los estudios de juego pueden incluir una o más DAW que pueden prestar soporte a la edición de señales HOA. Por ejemplo, las una o más DAW pueden incluir acoples HOA y/o herramientas que pueden ser configuradas para funcionar con (por ejemplo, trabajar con) uno o más sistemas de audio de juegos. En algunos ejemplos, los estudios de juego pueden producir nuevos formatos de derivación que den soporte a HOA. En cualquier caso, los estudios de juego pueden emitir contenido de audio codificado a los motores de reproducción que pueden representar un campo sonoro para su reproducción mediante los sistemas de suministro.

**[0205]** Las técnicas también se pueden realizar con respecto a dispositivos a modo de ejemplo de adquisición de audio. Por ejemplo, las técnicas se pueden realizar con respecto a un micrófono Eigen que puede incluir una pluralidad de micrófonos que están configurados colectivamente para grabar un campo de sonido tridimensional. En algunos ejemplos, la pluralidad de micrófonos del micrófono Eigen puede estar situada en la superficie de una bola esencialmente esférica con un radio de aproximadamente 4 cm. En algunos ejemplos, el dispositivo de codificación de audio 20 puede estar integrado en el micrófono Eigen a fin de emitir un flujo de bits 21 directamente desde el micrófono.

**[0206]** Otro contexto a modo de ejemplo de adquisición de audio puede incluir un camión de producción que puede configurarse para recibir una señal desde uno o más micrófonos, tal como uno o más micrófonos Eigen. El camión de producción también puede incluir un codificador de audio, tal como el codificador de audio 20 de la FIG. 3.

**[0207]** El dispositivo móvil puede incluir también, en algunos casos, una pluralidad de micrófonos que están configurados colectivamente para grabar un campo sonoro tridimensional. En otras palabras, la pluralidad de micrófonos puede tener diversidad X, Y, Z. En algunos ejemplos, el dispositivo móvil puede incluir un micrófono que puede ser girado para proporcionar diversidad X, Y, Z con respecto a otros uno o más micrófonos del dispositivo móvil. El dispositivo móvil también puede incluir un codificador de audio, tal como el codificador de audio 20 de la FIG. 3.

**[0208]** Un dispositivo de captura de vídeo robustecido puede configurarse adicionalmente para grabar un campo sonoro tridimensional. En algunos ejemplos, el dispositivo de captura de vídeo robustecido puede estar unido a un casco de un usuario que participa en una actividad. Por ejemplo, el dispositivo de captura de vídeo resistente puede estar unido a un casco de un usuario que hace rafting en aguas bravas. De esta manera, el dispositivo de captura de

vídeo robustecido puede capturar un campo sonoro tridimensional que representa la acción alrededor del usuario (por ejemplo, agua que se estrella detrás del usuario, otro navegante que habla delante del usuario, etc.).

**[0209]** Las técnicas también se pueden realizar con respecto a un dispositivo móvil mejorado accesorio, que se puede configurar para grabar un campo de sonido tridimensional. En algunos ejemplos, el dispositivo móvil puede ser similar a los dispositivos móviles expuestos anteriormente, con la adición de uno o más accesorios. Por ejemplo, un micrófono Eigen se puede unir al dispositivo móvil mencionado anteriormente para formar un dispositivo móvil mejorado accesorio. De esta manera, el dispositivo móvil mejorado accesorio puede capturar una versión del campo sonoro tridimensional de calidad superior que utilizando únicamente componentes de captura de sonido integrados en el dispositivo móvil mejorado accesorio.

**[0210]** Los dispositivos a modo de ejemplo de reproducción de audio que pueden llevar a cabo diversos aspectos de las técnicas descritas en esta divulgación se exponen adicionalmente a continuación. De acuerdo con una o más técnicas de esta divulgación, los altavoces y/o las barras de sonido pueden disponerse en cualquier configuración arbitraria mientras se sigue reproduciendo un campo sonoro tridimensional. Además, en algunos ejemplos, los dispositivos de reproducción de auriculares pueden estar acoplados a un descodificador 24 mediante una conexión por cable o bien inalámbrica. De acuerdo con una o más técnicas de esta divulgación, se puede utilizar una única representación genérica de un campo sonoro para renderizar el campo sonoro sobre cualquier combinación de altavoces, barras de sonido y dispositivos de reproducción de auriculares.

**[0211]** Un cierto número de diferentes entornos de reproducción de audio de ejemplo también pueden ser adecuados para realizar diversos aspectos de las técnicas descritas en esta divulgación. Por ejemplo, un entorno de reproducción de altavoces 5.1, un entorno de reproducción de altavoces 2.0 (por ejemplo, estéreo), un entorno de reproducción de altavoces 9.1 con altavoces frontales de altura completa, un entorno de reproducción de altavoces 22.2, un entorno de reproducción de altavoces 16.0, un entorno de reproducción de altavoces automotrices y un dispositivo móvil con entorno de reproducción por audífono pueden ser entornos adecuados para realizar diversos aspectos de las técnicas descritas en esta divulgación.

**[0212]** De acuerdo con una o más técnicas de esta divulgación, se puede utilizar una única representación genérica de un campo sonoro para representar el campo sonoro en cualquiera de los entornos de reproducción precedentes. Adicionalmente, las técnicas de esta divulgación permiten a un reproductor reproducir un campo sonoro desde una representación genérica para la reproducción en entornos de reproducción distintos al descrito anteriormente. Por ejemplo, si las consideraciones de diseño prohíben la colocación correcta de los altavoces de acuerdo con un entorno de reproducción de altavoces 7.1 (por ejemplo, si no es posible colocar un altavoz envolvente derecho), las técnicas de esta divulgación permiten que un reproductor compense con los otros 6 altavoces de modo que la reproducción se pueda lograr en un entorno de reproducción de altavoces 6.1.

**[0213]** Además, un usuario puede observar un juego deportivo mientras utiliza los auriculares. De acuerdo con una o más técnicas de esta divulgación, el campo sonoro tridimensional del juego deportivo puede ser adquirido (por ejemplo, uno o más micrófonos Eigen pueden ser colocados en y/o alrededor del estadio de béisbol), los coeficientes HOA correspondientes al campo sonoro tridimensional pueden obtenerse y transmitirse a un descodificador, el descodificador puede reconstruir el campo sonoro tridimensional basándose en los coeficientes HOA y emitir el campo sonoro tridimensional reconstruido a un reproductor, el reproductor puede obtener una indicación en cuanto al tipo de entorno de reproducción (por ejemplo, los auriculares) y reproducir el campo sonoro tridimensional reconstruido en señales que hacen que los auriculares emitan una representación del campo sonoro tridimensional del juego deportivo.

**[0214]** En cada una de los diversos casos descritos anteriormente, debería entenderse que el dispositivo de codificación de audio 20 puede realizar un procedimiento o comprender de otro modo medios para realizar cada etapa del procedimiento, para llevar a cabo el cual el dispositivo de codificación de audio 20 está configurado en algunos casos; estos medios pueden comprender uno o más procesadores. En algunos casos, los uno o más procesadores pueden representar un procesador de propósito especial configurado por medio de instrucciones almacenadas en un medio de almacenamiento no transitorio legible por ordenador. En otras palabras, varios aspectos de las técnicas en cada uno de los conjuntos de ejemplos de codificación pueden proveer un medio de almacenamiento no transitorio legible por ordenador que tiene almacenadas en el mismo instrucciones que, cuando se ejecutan, hacen que los uno o más procesadores realicen el procedimiento a llevar a cabo para el que el dispositivo de codificación de audio 20 ha sido configurado.

**[0215]** En uno o más ejemplos, las funciones descritas pueden implementarse en hardware, software, firmware o cualquier combinación de estos. Si se implementan en software, las funciones, como una o más instrucciones o código, pueden almacenarse en, y transmitirse por, un medio legible por ordenador, y ejecutarse mediante una unidad de procesamiento basada en hardware. Los medios legibles por ordenador pueden incluir medios de almacenamiento legibles por ordenador, que correspondan a un medio tangible tal como medios de almacenamiento de datos. Los medios de almacenamiento de datos pueden ser medios disponibles cualesquiera a los que se puede acceder desde uno o más ordenadores o uno o más procesadores para recuperar instrucciones, código y/o estructuras de datos para la implementación de las técnicas descritas en esta divulgación. Un producto de programa

informático puede incluir un medio legible por ordenador.

**[0216]** Análogamente, en cada uno de los diversos casos descritos anteriormente, debería entenderse que el dispositivo de decodificación de audio 24 puede realizar un procedimiento o comprender de otro modo medios para realizar cada etapa del procedimiento a realizar para el cual está configurado el dispositivo de decodificación de audio 24. En algunos casos, estos medios pueden comprender uno o más procesadores. En algunos casos, los uno o más procesadores pueden representar un procesador de propósito especial configurado por medio de instrucciones almacenadas en un medio de almacenamiento no transitorio legible por ordenador. En otras palabras, varios aspectos de las técnicas en cada uno de los conjuntos de ejemplos de codificación pueden proporcionar un medio de almacenamiento no transitorio legible por ordenador que tiene almacenadas en él instrucciones que, cuando se ejecutan, hacen que uno o más procesadores realicen el procedimiento a realizar para el cual el dispositivo de decodificación de audio 24 ha sido configurado.

**[0217]** A modo de ejemplo, y no de manera limitativa, tales medios de almacenamiento legibles por ordenador pueden comprender RAM, ROM, EEPROM, CD-ROM u otro almacenamiento de disco óptico, almacenamiento de disco magnético u otros dispositivos de almacenamiento magnético, memoria flash o cualquier otro medio que pueda usarse para almacenar código de programa deseado en forma de instrucciones o estructuras de datos y al que pueda accederse mediante un ordenador. Sin embargo, debería entenderse que los medios de almacenamiento legibles por ordenador y los medios de almacenamiento de datos no incluyen conexiones, ondas portadoras, señales u otros medios transitorios, sino que, en cambio, se orientan a medios de almacenamiento tangibles no transitorios. El término disco, tal como se utiliza en el presente documento, incluye un disco compacto (CD), un disco láser, un disco óptico, un disco versátil digital (DVD), un disco flexible y un disco Blu-ray, donde algunos discos habitualmente emiten datos magnéticamente, mientras que otros discos emiten datos ópticamente con láseres. Las combinaciones de los anteriores también deben incluirse dentro del alcance de los medios legibles por ordenador.

**[0218]** Las instrucciones pueden ser ejecutadas por uno o más procesadores, tales como uno o más procesadores de señales digitales (DSP), microprocesadores de propósito general, circuitos integrados específicos de la aplicación (ASIC), matriz de puertas programables in situ (FPGA) u otros circuitos lógicos integrados o discretos equivalentes. Por consiguiente, el término «procesador», como se utiliza en el presente documento, puede referirse a cualquiera de las estructuras anteriores o a cualquier otra estructura adecuada para la implementación de las técnicas descritas en el presente documento. Además, en algunos aspectos, la funcionalidad descrita en el presente documento puede proporcionarse dentro de módulos de hardware y/o software dedicados configurados para la codificación y la decodificación, o incorporarse en un códec combinado. Asimismo, las técnicas podrían implementarse por completo en uno o más circuitos o elementos lógicos.

**[0219]** Las técnicas de la presente divulgación se pueden implementar en una amplia variedad de dispositivos o aparatos, incluidos un teléfono inalámbrico, un circuito integrado (IC) o un conjunto de IC (por ejemplo, un conjunto de chips). Diversos componentes, módulos o unidades se describen en esta divulgación para enfatizar aspectos funcionales de dispositivos configurados para realizar las técnicas divulgadas, pero no requieren necesariamente su realización mediante diferentes unidades de hardware. En cambio, como se ha descrito anteriormente, diversas unidades pueden combinarse en una unidad de hardware de códec o proporcionarse por medio de un grupo de unidades de hardware interoperativas, que incluyen uno o más procesadores como los descritos anteriormente, conjuntamente con software y/o firmware adecuados.

**[0220]** Se han descrito diversos aspectos de las técnicas. Estos y otros aspectos están dentro del alcance de las reivindicaciones siguientes.

# REIVINDICACIONES

1. Un dispositivo configurado para renderizar coeficientes ambisónicos de orden superior, comprendiendo el dispositivo:  
5 uno o más procesadores configurados para:  
  
obtener, a partir de un flujo de bits que incluye una versión codificada de los coeficientes ambisónicos de orden superior, información de dispersión que indica una dispersión de una matriz utilizada para renderizar los coeficientes ambisónicos de orden superior para generar una pluralidad de alimentaciones de altavoces;  
10 obtener, del flujo de bits, información de simetría de signo indicativa de la simetría de signo de la matriz;  
15 obtener, del flujo de bits, un número reducido de bits usados para representar la matriz; y  
  
basándose en la información de dispersión, la información de simetría de signo y el número reducido de bits, reconstruir la matriz; y  
20 una memoria acoplada a los uno o más procesadores, y configurada para almacenar la información de dispersión.
2. El dispositivo de la reivindicación 1, en el que uno o más procesadores están configurados además para determinar un diseño de altavoz para el cual la matriz se va a usar para renderizar la pluralidad de alimentaciones de altavoces a partir de los coeficientes ambisónicos de orden superior.  
25
3. El dispositivo de la reivindicación 1, que comprende además un altavoz configurado para reproducir un campo de sonido representado por los coeficientes ambisónicos de orden superior basándose en la pluralidad de alimentaciones de altavoces.  
30
4. El dispositivo según la reivindicación 1, en el que los uno o más procesadores están configurados además para obtener, a partir del flujo de bits, información de renderización de audio indicativa de un valor de la señal que identifica un renderizador de audio utilizado al generar la pluralidad de alimentaciones de altavoces y renderizar la pluralidad de alimentaciones de altavoces basado en la información de renderización de audio.  
35
5. El dispositivo de la reivindicación 4,  
en el que el valor de la señal incluye un índice asociado con la matriz utilizada para renderizar los coeficientes ambisónicos de orden superior a la pluralidad de alimentaciones de altavoces, y  
en el que los uno o más procesadores están configurados para renderizar la pluralidad de alimentaciones de altavoces basándose en la matriz asociada con el índice incluido en el valor de la señal.  
40
6. Un procedimiento para renderizar coeficientes ambisónicos de orden superior, el procedimiento que comprende:  
45 obtener, a partir de un flujo de bits que incluye una versión codificada de los coeficientes ambisónicos de orden superior, información de dispersión que indica una dispersión de una matriz utilizada para renderizar los coeficientes ambisónicos de orden superior para generar una pluralidad de alimentaciones de altavoces;  
50 obtener, a partir del flujo de bits, información de simetría de signo indicativa de simetría de signo de la matriz;  
  
obtener, a partir del flujo de bits, un número reducido de bits usados para representar la matriz; y  
55 basándose en la información de dispersión, la información de simetría de signo y el número reducido de bits, reconstruir la matriz.
7. El procedimiento según la reivindicación 6, que comprende además determinar un diseño de altavoz para el cual la matriz se va a usar para renderizar los datos de audio multicanal a partir de los coeficientes ambisónicos de orden superior.  
60
8. El procedimiento según la reivindicación 6, que comprende además reproducir un campo de sonido representado por los coeficientes ambisónicos de orden superior basándose en la pluralidad de alimentaciones de altavoces.  
65
9. El procedimiento según la reivindicación 6, que comprende además:

obtener, a partir del flujo de bits, información de renderización de audio indicativa de un valor de la señal que identifica un renderizador de audio utilizado al generar la pluralidad de alimentaciones de altavoces; y

5           renderizar la pluralidad de alimentaciones de altavoces basándose en la información de renderización de audio.

10.       El procedimiento según la reivindicación 9,  
10       en el que el valor de la señal incluye un índice asociado con la matriz utilizada para renderizar los coeficientes ambisónicos de orden superior para generar la pluralidad de alimentaciones de altavoces, y  
10       en el que el procedimiento comprende además renderizar la pluralidad de alimentaciones de altavoces basándose en la matriz asociada con el índice incluido en el valor de la señal.

11.       Un dispositivo configurado para producir un flujo de bits, el dispositivo que comprende:  
15           una memoria configurada para almacenar una matriz utilizada para renderizar un coeficiente ambisónico de orden superior para generar una pluralidad de alimentaciones de altavoces; y  
20           uno o más procesadores acoplados a la memoria, y configurados para:  
20           obtener información de simetría de signo indicativa de simetría de signo de la matriz;  
20           obtener información de dispersión que indica una dispersión de la matriz; y  
25           basándose en la información de simetría de signo y la información de dispersión, determine un número reducido de bits utilizados para representar la matriz  
25           generar el flujo de bits para incluir una versión codificada de los coeficientes ambisónicos de orden superior, la información de simetría de signo, la información de dispersión y el número reducido de  
30           bits.

12.       El dispositivo de la reivindicación 11, en el que los uno o más procesadores están configurados además para determinar un diseño de altavoz para el cual la matriz se va a usar para renderizar la pluralidad de alimentaciones de altavoces a partir de los coeficientes ambisónicos de orden superior.

13.       El dispositivo de la reivindicación 11, que comprende además un micrófono configurado para capturar un campo de sonido representado por los coeficientes ambisónicos de orden superior.

14.       Un procedimiento de producción de un flujo de bits, el procedimiento que comprende:  
40           obtener información de dispersión indicativa de una dispersión de una matriz utilizada para generar coeficientes ambisónicos de orden superior para generar una pluralidad de alimentaciones de altavoces;  
40           obtener información de simetría de signo que indica una simetría de signo de la matriz; y  
45           basándose en la información de simetría de signo y la información de dispersión, determinar un número reducido de bits usados para representar la matriz;  
45           generar el flujo de bits para incluir una versión codificada de los coeficientes ambisónicos de orden superior, la información de simetría de signo, la información de dispersión y el número reducido de bits.

15.       El procedimiento según la reivindicación 14, que comprende además determinar un diseño de altavoz para el cual la matriz se va a usar para renderizar los datos de audio multicanal a partir de los coeficientes ambisónicos de orden superior.

55

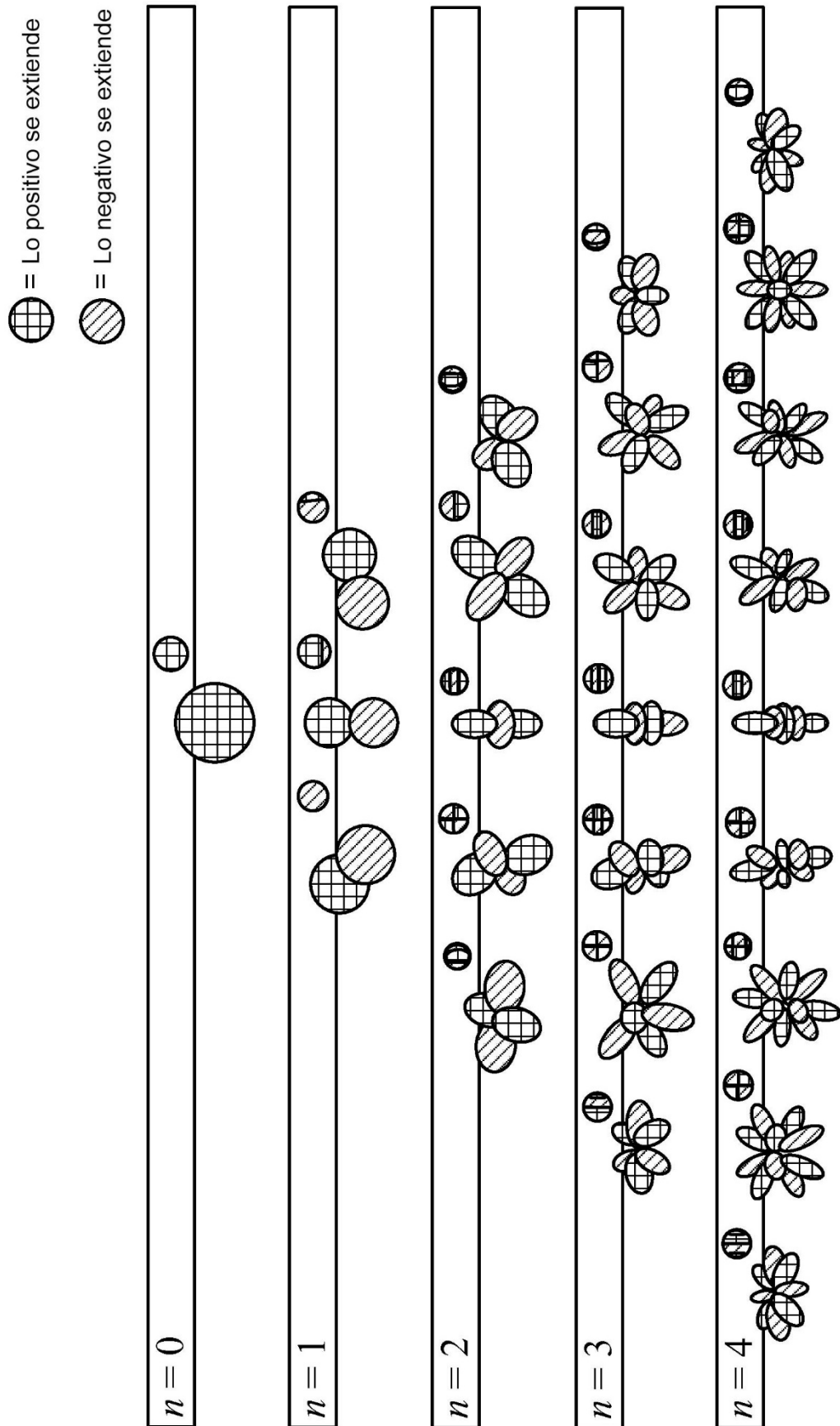


FIG. 1

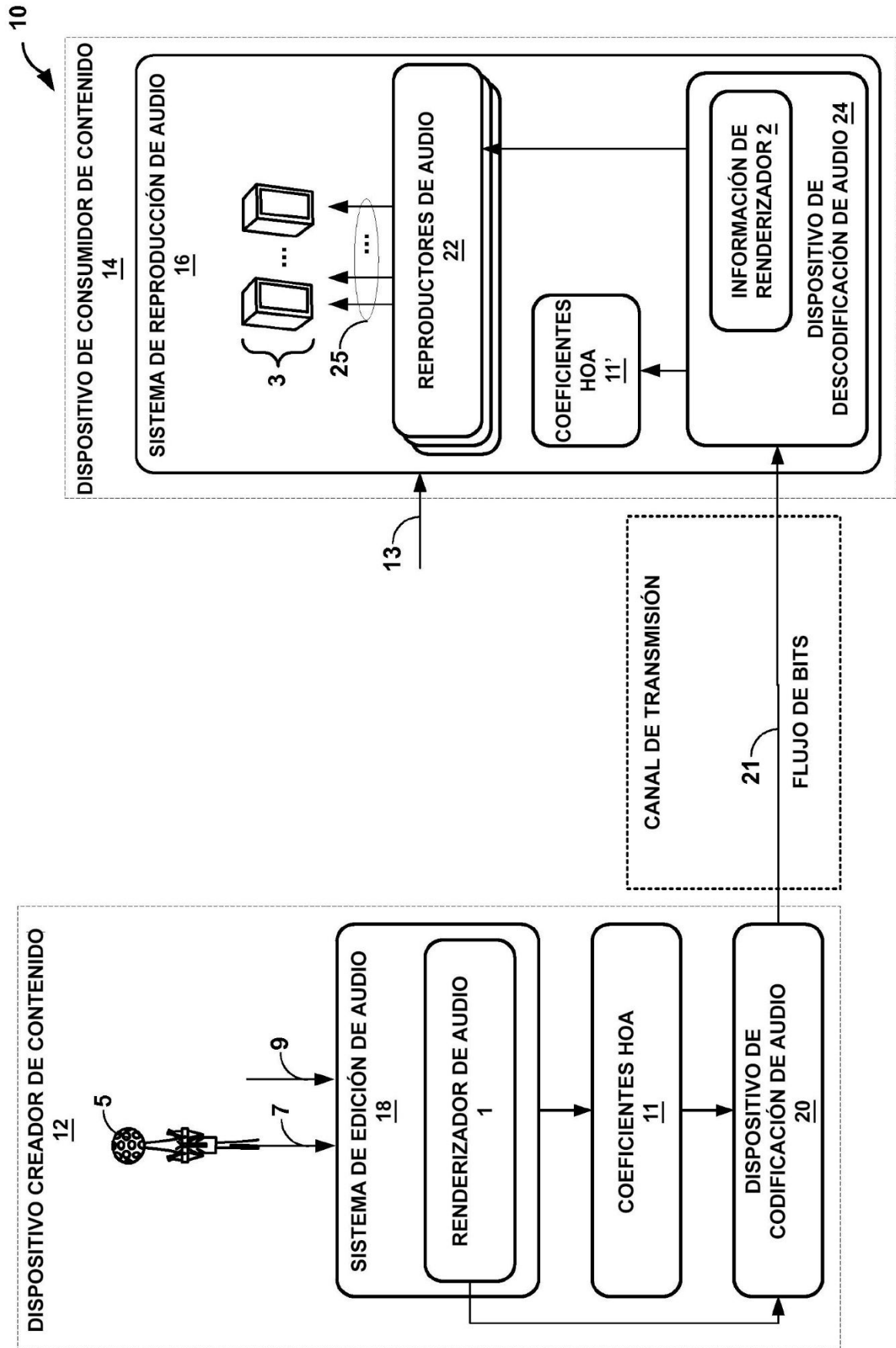
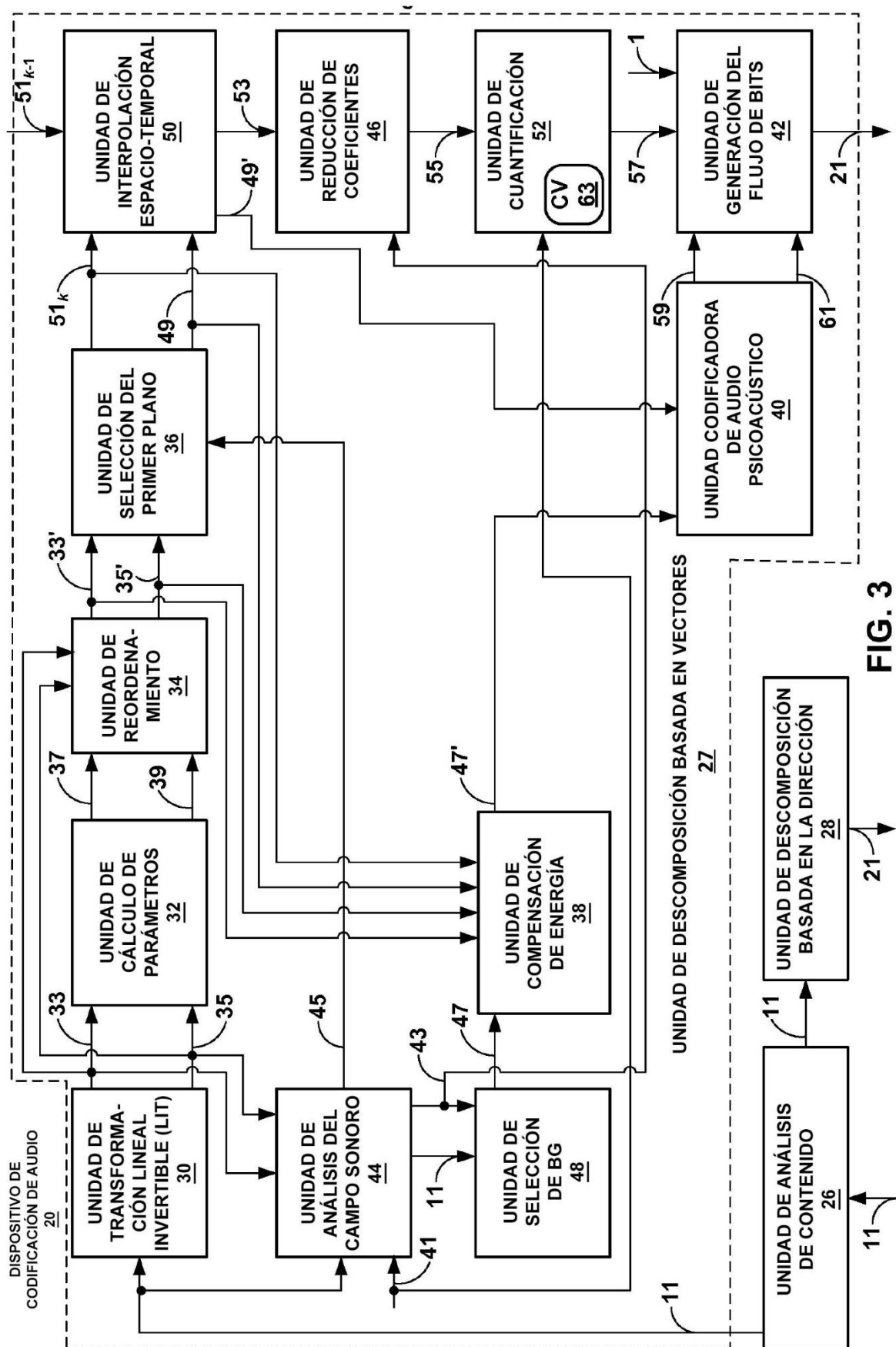


FIG. 2





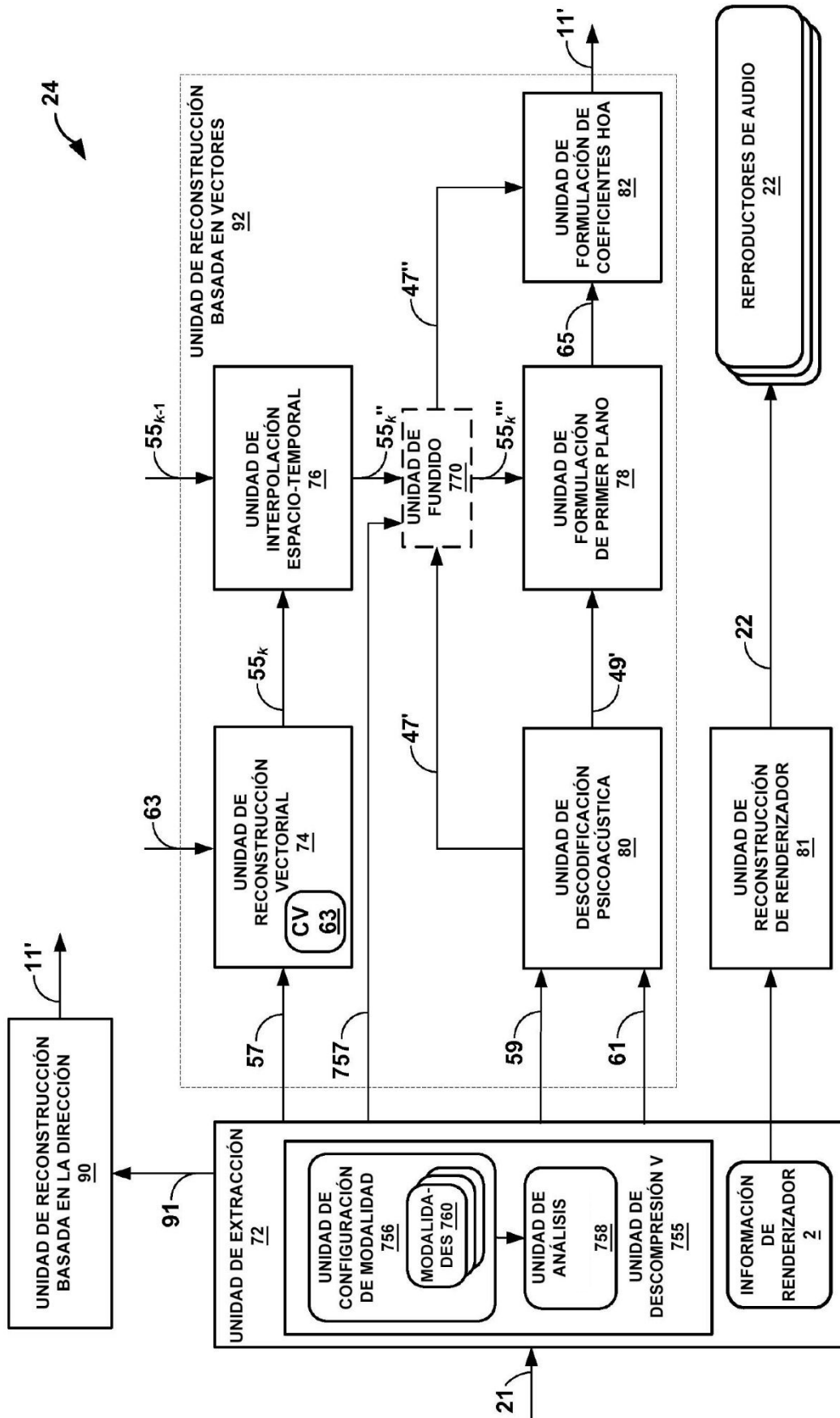


FIG. 4

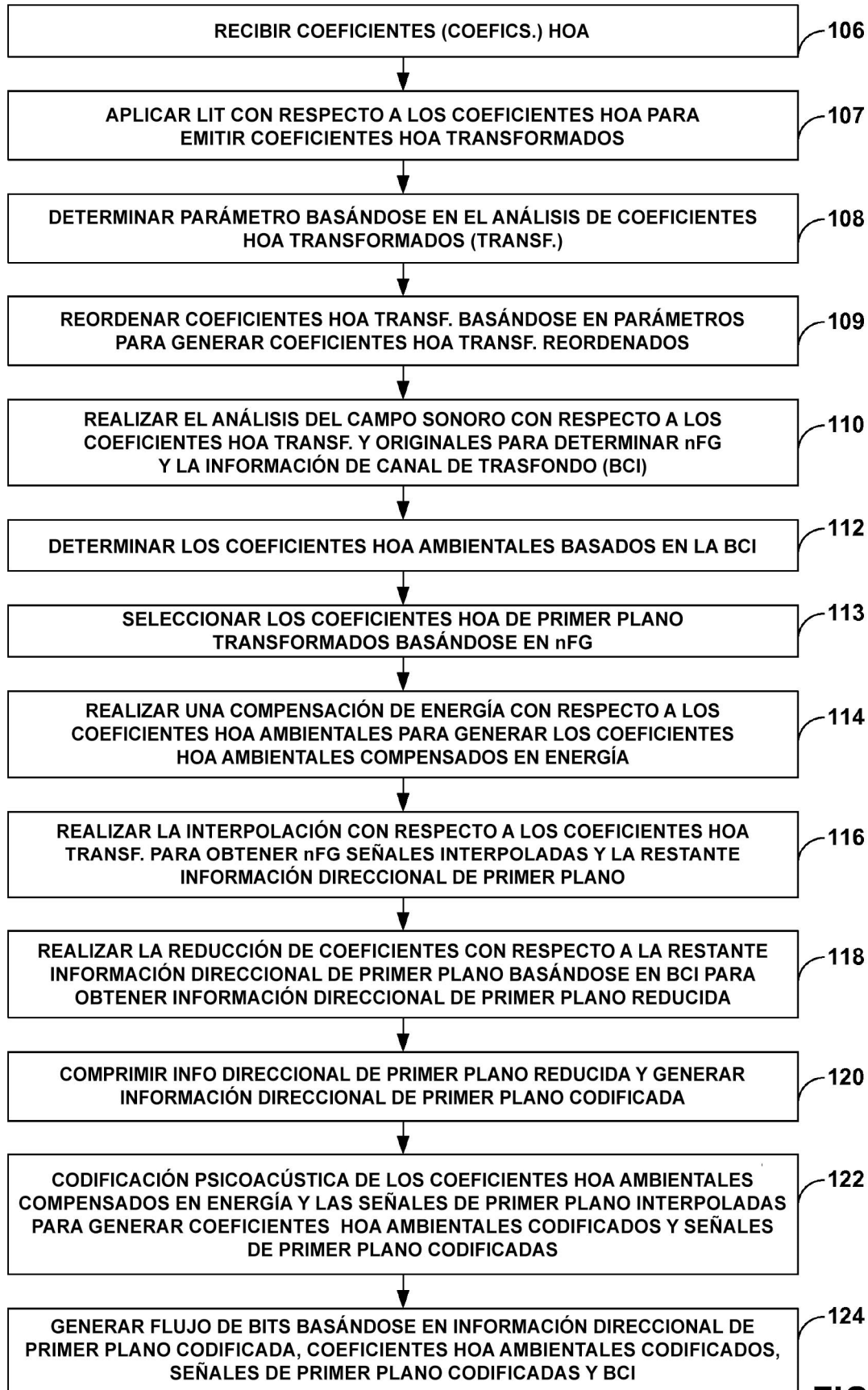


FIG. 5

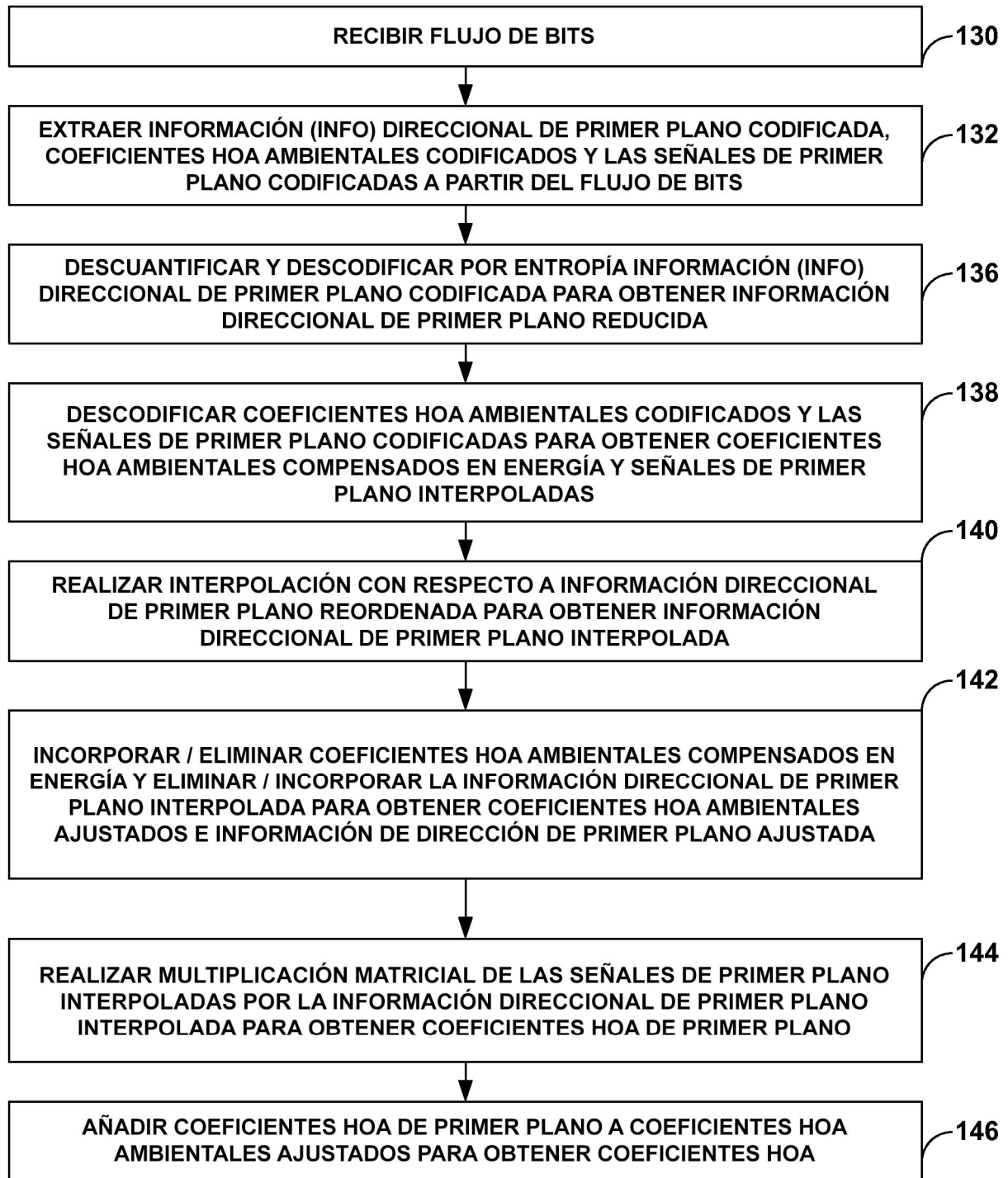


FIG. 6

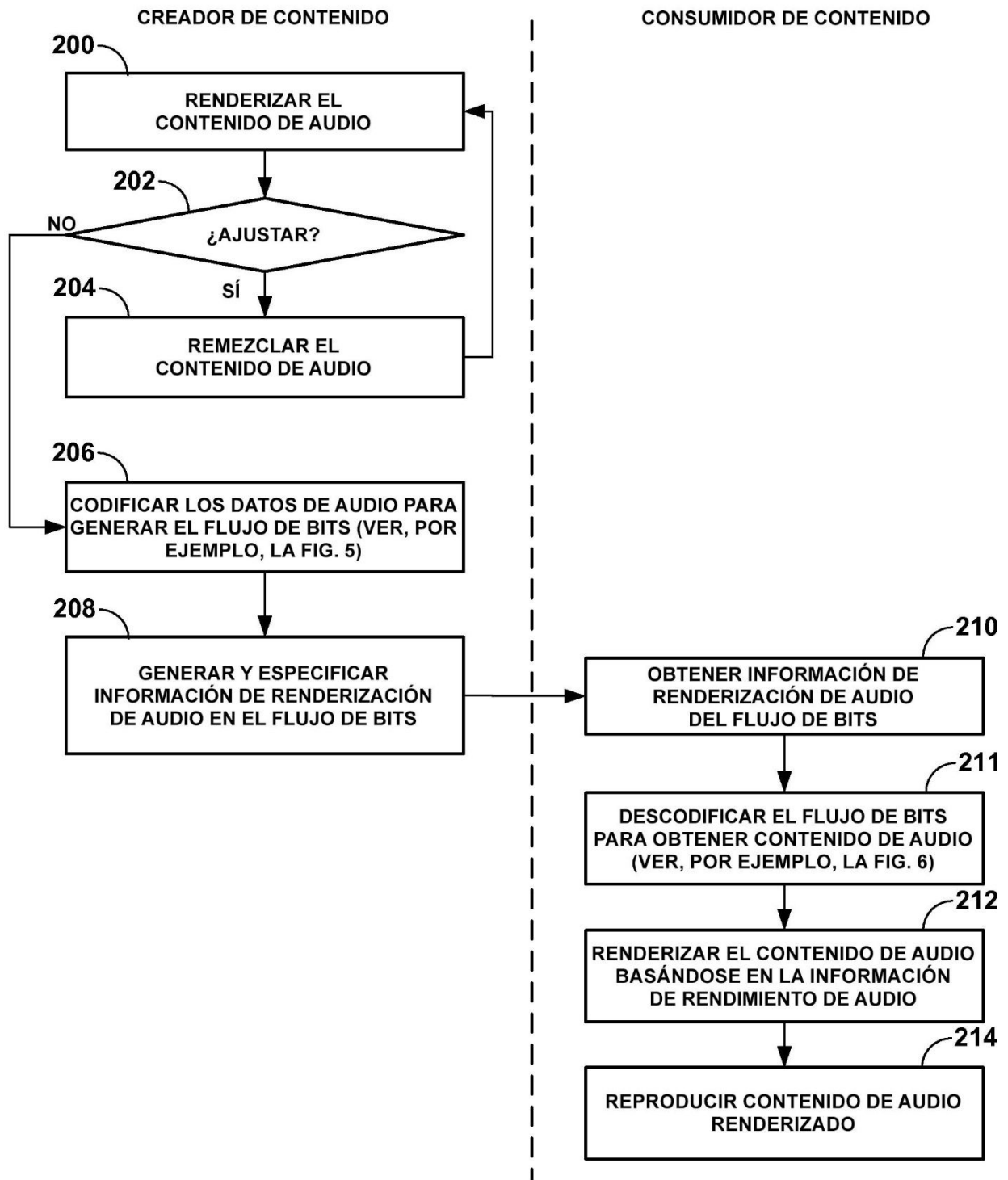


FIG. 7

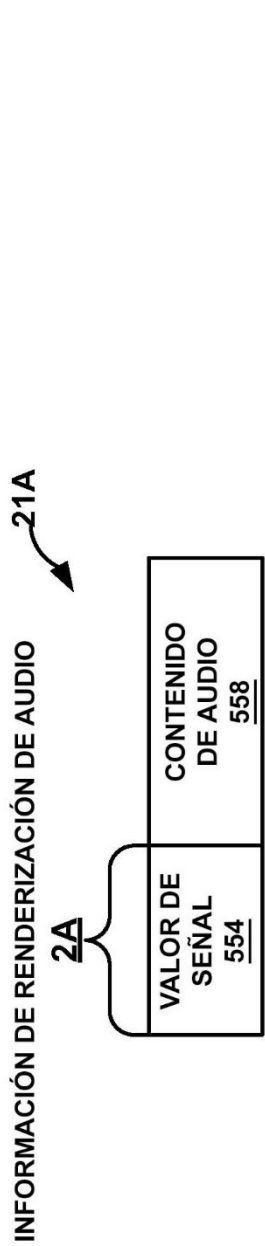


FIG. 8A

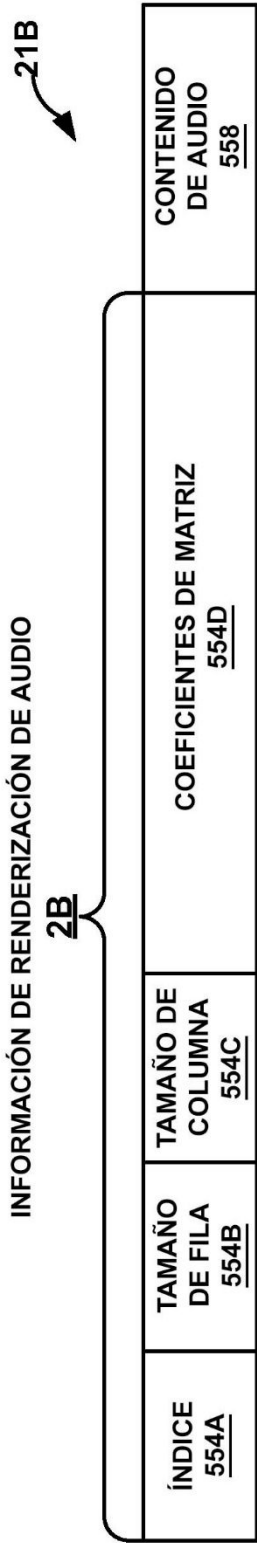


FIG. 8B



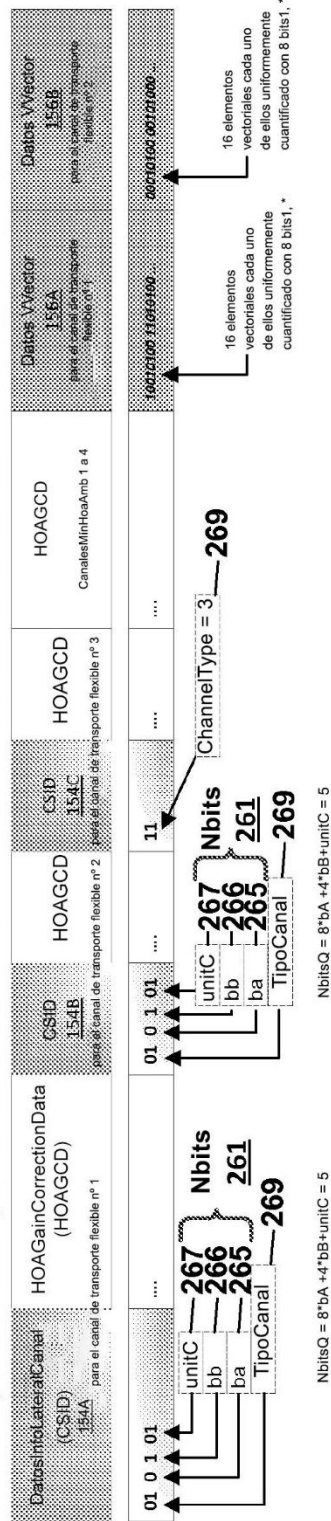
FIG. 8C

FIG. 8D

249A'

### Trama k

En este ejemplo, los canales de transporte flexibles contienen dos canales de primer plano basados en vectores en la trama k y tres vectores de primer plano basados en vectores en la segunda trama k+1



249A''

FIG. 8E

### Trama k

En este ejemplo, los canales de transporte flexibles contienen dos canales de primer plano basados en vectores en la trama k y tres vectores de primer plano basados en vectores en la segunda trama k+1

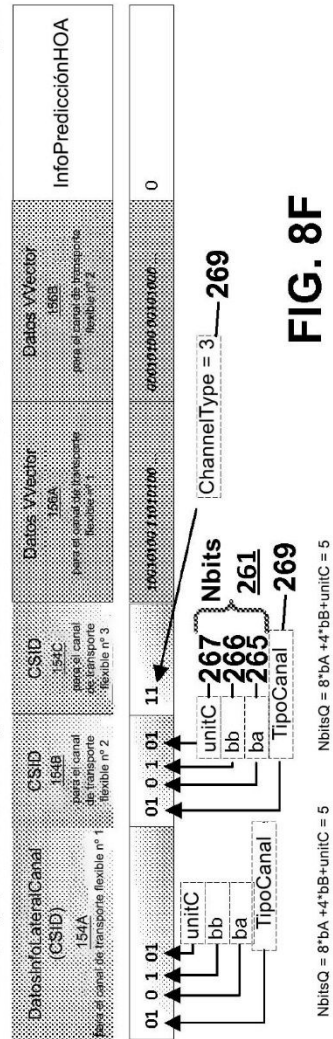


FIG. 8F

249A''

# Trama k

En este ejemplo, los canales de transporte flexibles contienen dos canales de primer plano basados en vectores en la trama k y tres vectores de primer plano basados en vectores en la segunda trama k+1

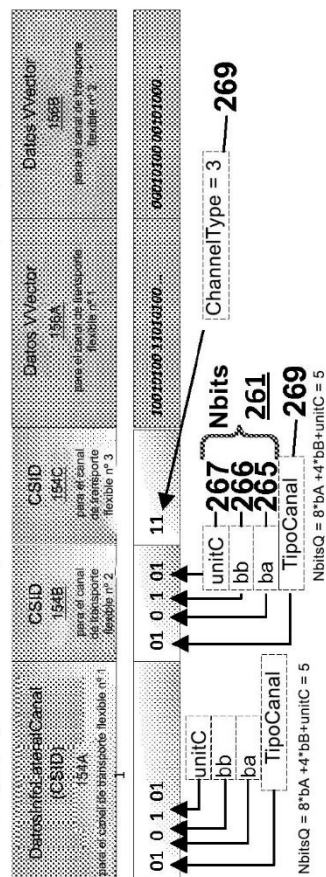
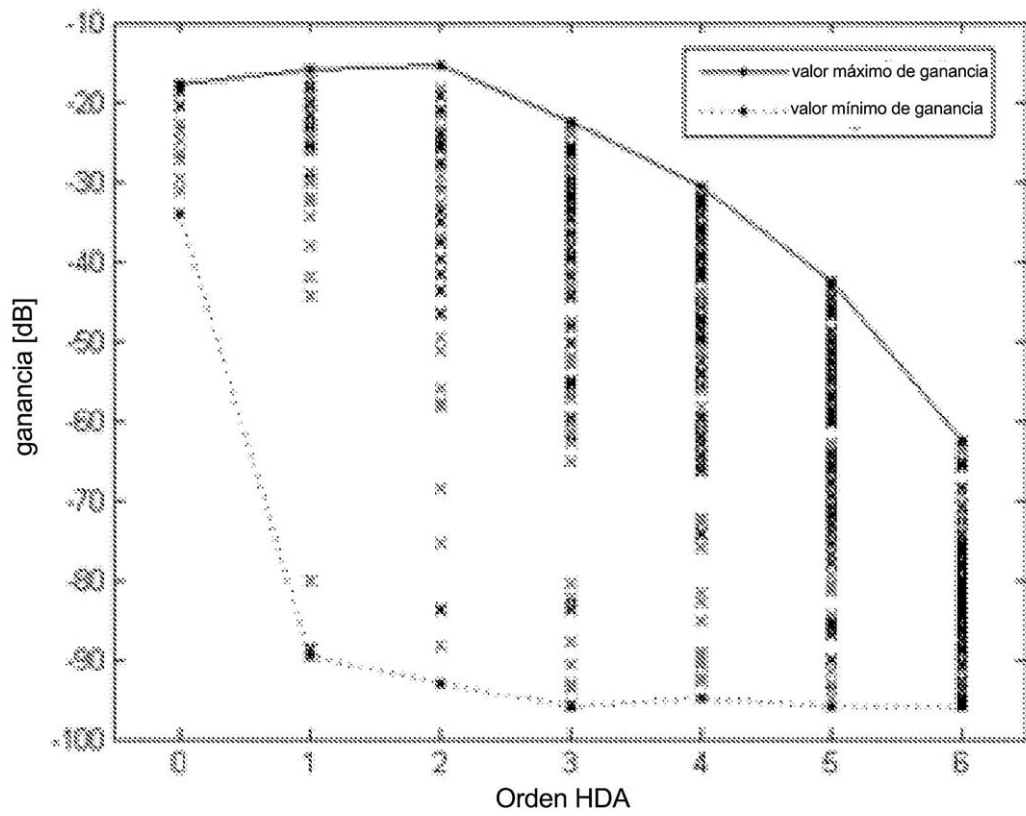
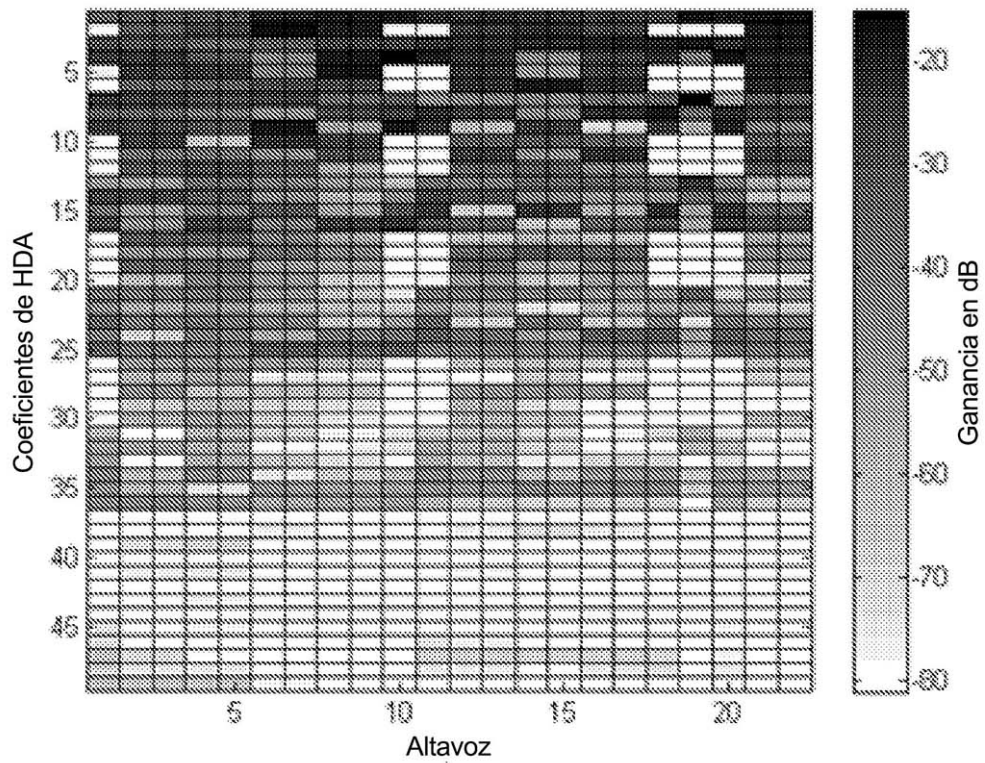


FIG. 8G





**FIG. 9**



**FIG. 10**

