

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 734 512**

51 Int. Cl.:

H04S 3/00 (2006.01)
H04L 29/06 (2006.01)
H04H 20/89 (2008.01)
G10L 19/18 (2013.01)
G10L 19/16 (2013.01)
G10L 19/008 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **18.06.2013** **PCT/US2013/046369**

87 Fecha y número de publicación internacional: **23.01.2014** **WO14014600**

96 Fecha de presentación y número de la solicitud europea: **18.06.2013** **E 13731630 (3)**

97 Fecha y número de publicación de la concesión europea: **10.04.2019** **EP 2873252**

54 Título: **Sistemas, procedimientos, aparatos y medios legibles por ordenador para la codificación de audio compatible con versiones anteriores**

30 Prioridad:

15.07.2012 US 201261671789 P
15.03.2013 US 201313844447

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
10.12.2019

73 Titular/es:

QUALCOMM INCORPORATED (100.0%)
5775 Morehouse Drive
San Diego, CA 92121-1714, US

72 Inventor/es:

SEN, DIPANJAN y
XIANG, PEI

74 Agente/Representante:

FORTEA LAGUNA, Juan José

ES 2 734 512 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

Sistemas, procedimientos, aparatos y medios legibles por ordenador para la codificación de audio compatible con versiones anteriores

ANTECEDENTES

Campo

[0001] Esta divulgación se refiere a la codificación espacial de vídeo.

Antecedentes

[0002] La evolución del sonido envolvente ha dejado disponibles muchos formatos de salida para el entretenimiento de hoy en día. La gama de formatos de sonido envolvente en el mercado incluye el popular formato del sistema de cine en casa 5.1, que ha sido el más exitoso en términos de hacer incursiones en salas de estar más allá del estéreo. Este formato incluye los siguientes seis canales: frontal izquierdo (L), frontal derecho (R), central o frontal central (C), posterior izquierda o envolvente izquierdo (Ls), posterior derecha o envolvente derecho (Rs) y efectos de baja frecuencia (LFE). Otros ejemplos de formatos de sonido envolvente incluyen el creciente formato 7.1 y el formato futurista 22.2 desarrollado por NHK (Nippon Hoso Kyokai o Corporación Difusora de Japón) para su uso, por ejemplo, con la norma de Televisión de Definición Ultra Alta. Puede ser deseable que un formato de sonido envolvente codifique audio en dos dimensiones y / o en tres dimensiones.

[0003] El documento US2004 / 247134 describe un sistema y un procedimiento para la reproducción de sonido envolvente compatible 2D / 3D (esfera completa con altura). El documento EP2469741 describe un procedimiento y un aparato para codificar y decodificar tramas sucesivas de una representación ambisónica de un campo sonoro bidimensional o tridimensional. El documento US2006 / 045275 describe un procedimiento para procesar datos de audio y un dispositivo de adquisición de sonido que implementan este procedimiento. El sonido envolvente tridimensional también se expone en el artículo "Transforming Ambiophonic and Ambisonic 3D Surround Sound to & from ITU 5.1/6.1 [Transformación del sonido envolvente ambiofónico y ambisónico tridimensional hasta y desde la UIT 5.1 / 6.1]", publicado por AES, 60 East 42nd Street, Room 2520, NY.

[0004] El documento EP 2 450 880 A1 describe estructuras de datos para los datos de audio de Ambisónicos de Orden Superior, generados por un codificador de audio y utilizados por un decodificador.

SUMARIO

[0005] Un procedimiento para procesar una pluralidad de coeficientes de función de base que describe un campo de sonido durante un intervalo de tiempo comprende la secuencia de etapas según lo definido en la reivindicación 1. Un procedimiento para obtener una pluralidad de coeficientes de función de base que describe un campo de sonido durante un intervalo de tiempo comprende la secuencia de etapas según lo definido en la reivindicación 4. Además, un medio de almacenamiento no transitorio legible por ordenador que tiene instrucciones almacenadas en el mismo se reivindica según lo definido en la reivindicación 19.

[0006] Un aparato para procesar una pluralidad de coeficientes de función de base que describe un campo de sonido durante un intervalo de tiempo comprende el conjunto de características según lo definido en la reivindicación 7.

[0007] Un aparato para obtener una pluralidad de coeficientes de función de base que describe un campo de sonido durante un intervalo de tiempo comprende el conjunto de características según lo definido en la reivindicación 11.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

[0008]

La figura 1A muestra un ejemplo de codificación basada en escenas sin tener en cuenta la compatibilidad con versiones anteriores.

La figura 1B ilustra una estructura general para la estandarización utilizando un códec del MPEG.

La figura 1C muestra un ejemplo compatible con versiones anteriores para mono / estéreo.

La figura 2A muestra una realización de un proceso de codificación con un diseño compatible con versiones anteriores.

La figura 2B muestra un ejemplo de un proceso de decodificación en un decodificador convencional que no puede decodificar datos basados en escenas.

La figura 3A muestra una realización de un proceso de decodificación con un nuevo dispositivo que puede gestionar datos basados en la escena.

5 La figura 3B muestra un diagrama de flujo de un procedimiento de procesamiento de señales de audio M100, de acuerdo a un modo de realización.

La figura 3C muestra un diagrama de bloques de un aparato MF100 para el procesamiento de señales de audio, según un modo de realización.

10 La figura 3D muestra un diagrama de bloques de un aparato A100 para el procesamiento de señales de audio, según otro modo de realización.

15 La figura 4 muestra ejemplos de gráficos de malla superficial de las magnitudes de funciones de base armónicas esféricas de orden 0 y 1.

La figura 5 muestra ejemplos de gráficos de malla superficial de las magnitudes de funciones de base armónicas esféricas de orden 2.

20 La figura 6 muestra un ejemplo compatible con las normas de los ángulos de altavoces para un sistema 5.1.

La figura 7 muestra un ejemplo de una disposición uniformemente espaciada de ángulos de altavoz para un sistema de destino 5.1.

25 La figura 8 muestra un ejemplo compatible con las normas de los ángulos de los altavoces para un sistema 7.1.

La figura 9 muestra un ejemplo de una disposición uniformemente espaciada de ángulos de altavoz para un sistema de destino 7.1.

30 La figura 10A muestra un diagrama de flujo de un procedimiento de procesamiento de señales de audio M200, de acuerdo a un modo de realización.

La figura 10B muestra un diagrama de flujo de una implementación M210 del procedimiento M200.

35 La figura 11A muestra un diagrama de bloques de un aparato MF200 para el procesamiento de señales de audio, según un modo de realización.

La figura 11B muestra un diagrama de bloques de una implementación MF210 del aparato MF100.

40 La figura 11C muestra un diagrama de bloques de un aparato A200 para el procesamiento de señales de audio, de acuerdo con otro modo de realización.

La figura 11D muestra un diagrama de bloques de una implementación A210 del aparato A100.

45 La figura 11E muestra un diagrama de bloques de un aparato P100 para el procesamiento de señales de audio, de acuerdo con otro modo de realización.

DESCRIPCIÓN DETALLADA

50 **[0009]** A menos que esté limitado expresamente por su contexto, el término "señal" se usa en el presente documento para indicar cualquiera de sus significados comunes, incluyendo un estado de una posición de memoria (o conjunto de posiciones de memoria), según se expresa en un cable, bus u otro medio de transmisión. A menos que esté limitado expresamente por su contexto, el término "generar" se usa en el presente documento para indicar cualquiera de sus significados comunes, tales como calcular o producir de otro modo. A menos que esté limitado expresamente por su contexto, el término "calcular" se usa en el presente documento para indicar cualquiera de sus significados comunes, tales como calcular, evaluar, estimar y/o seleccionar a partir de una pluralidad de valores. A menos que esté limitado expresamente por su contexto, el término "obtener" se usa para indicar cualquiera de sus significados comunes, tales como calcular, obtener, recibir (por ejemplo, desde un dispositivo externo) y/o recuperar (por ejemplo, desde una formación de elementos de almacenamiento). A menos que esté limitado expresamente por su contexto, el término "seleccionar" se usa para indicar cualquiera de sus significados comunes, tales como identificar, indicar, aplicar y/o usar al menos uno, y menos que todos, de un conjunto de dos o más. Cuando se usa el término "que comprende" en la presente descripción y en las reivindicaciones, no excluye otros elementos u operaciones. El término "basado en" (como en "A está basado en B") se usa para indicar cualquiera de sus significados comunes, incluidos los casos (i) "obtenido de" (por ejemplo, "B es un precursor de A"), (ii) "basado al menos en" (por ejemplo, "A está basado al menos en B") y, si corresponde en el contexto particular, (iii) "igual a" (por ejemplo, "A es igual a B"). De manera similar, el término "en respuesta a" se usa para indicar cualquiera de sus significados comunes, incluyendo "en respuesta a al

menos".

[0010] Las referencias a la "ubicación" de un micrófono de un dispositivo de detección de audio de micrófonos múltiples indican la ubicación del centro de una cara del micrófono acústicamente sensible, a menos que el contexto indique otra cosa. El término "canal" se usa a veces para indicar un trayecto de señal y otras veces para indicar una señal transportada por un trayecto de ese tipo, de acuerdo al contexto particular. A menos que se indique otra cosa, el término "serie" se usa para indicar una secuencia de dos o más elementos. El término "logaritmo" se usa para indicar el logaritmo de base diez, aunque las extensiones de dicha operación a otras bases estén dentro del alcance de la presente divulgación. El término "componente de frecuencia" se usa para indicar uno entre un conjunto de frecuencias o bandas de frecuencia de una señal, tal como una muestra de una representación de dominio de frecuencia de la señal (por ejemplo, producida por una transformación rápida de Fourier) o una subbanda de la señal (por ejemplo, una escala de Bark o subbanda con escala Mel).

[0011] A menos que se indique otra cosa, cualquier divulgación de una operación de un aparato que tiene una característica particular también está expresamente concebida para divulgar un procedimiento que tenga una característica análoga (y viceversa), y cualquier divulgación de una operación de un aparato de acuerdo a una configuración particular también está concebida expresamente para divulgar un procedimiento de acuerdo a una configuración análoga (y viceversa). El término "configuración" puede usarse en referencia a un procedimiento, aparato y/o sistema según lo indicado por su contexto particular. Los términos "procedimiento", "proceso" y "técnica" se usan de forma genérica e intercambiable a menos que el contexto particular indique otra cosa. Los términos "aparato" y "dispositivo" también se usan de forma genérica e intercambiable a menos que el contexto particular indique otra cosa. Los términos "elemento" y "módulo" se usan habitualmente para indicar una parte de una configuración mayor. A menos que esté limitado expresamente por su contexto, el término "sistema" se usa en el presente documento para indicar cualquiera de sus significados comunes, incluyendo "un grupo de elementos que interactúan para servir a un propósito común".

[0012] A menos que sea introducido inicialmente por un artículo definido, un término ordinal (por ejemplo, "primero", "segundo", "tercero", etc.) usado para modificar un elemento de reivindicación no indica por sí mismo ninguna prioridad u orden del elemento de reivindicación con respecto a otro, sino más bien distingue simplemente el elemento de reivindicación de otro elemento de reivindicación que tenga un mismo nombre (excepto para el uso del término ordinal). A menos que esté expresamente limitado por su contexto, cada uno de los términos "pluralidad" y "conjunto" se usa en este documento para indicar una cantidad entera que es mayor que uno.

[0013] El estado actual de la técnica en audio de consumo es la codificación espacial que utiliza un sonido envolvente basado en canales, que está concebido para ser reproducido a través de altavoces en posiciones previamente especificadas. Otro enfoque de la codificación de audio espacial es el audio basado en objetos, que implica datos discretos de modulación de código de pulsos (PCM) para objetos de audio individuales con metadatos asociados que contienen coordenadas de ubicación de los objetos en el espacio (entre otra información).

[0014] Un enfoque adicional a la codificación de audio espacial (por ejemplo, a la codificación de sonido envolvente) es el audio basado en escenas, que implica representar el campo de sonido utilizando los coeficientes de las funciones de bases armónicas esféricas. Tales coeficientes también se denominan "coeficientes armónicos esféricos" o SHC. La figura 1A representa un proceso de codificación y decodificación sencillo con un enfoque basado en escenas. En este ejemplo, el codificador SE10 basado en escenas produce una descripción de los SHC que se transmite (y / o se almacena) y se decodifica en el decodificador SD10 basado en escenas para recibir los SHC para su representación (por ejemplo, mediante el representador SR10 de SH). Dicha codificación puede incluir una o más técnicas de codificación con pérdidas o sin pérdidas para la compresión del ancho de banda, tales como la cuantización (por ejemplo, en uno o más índices de libro de códigos), la codificación de corrección de errores, la codificación de redundancia, etc. Adicionalmente o alternativamente, dicha codificación puede incluir la codificación de canales de audio (por ejemplo, salidas de micrófono) en un formato Ambisónico, tal como el formato B, el formato G o Ambisónico de orden superior (HOA). En general, el codificador SE10 puede codificar los SHC utilizando técnicas que aprovechan las redundancias entre los coeficientes y / o las irrelevancias (para la codificación con o sin pérdidas).

[0015] Puede ser deseable proporcionar una codificación de información de audio espacial en un flujo de bits estandarizado y una decodificación posterior que sea adaptable y agnóstica a la geometría del altavoz y las condiciones acústicas en la ubicación del representador. Un enfoque de ese tipo puede brindar el objetivo de una experiencia de audición uniforme, independientemente de la configuración particular que se utilice finalmente para la reproducción. La figura 1B ilustra una estructura general para dicha estandarización, utilizando un códec del MPEG. En este ejemplo, las fuentes de audio de entrada al codificador MP10 pueden incluir uno o más de los siguientes, por ejemplo: fuentes basadas en canales (por ejemplo, 1.0 (monofónico), 2.0 (estereofónico), 5.1, 7.1, 11.1, 22.2), fuentes basadas en objetos y fuentes basadas en escenas (por ejemplo, armónicos esféricos de orden superior, Ambisónicos). De manera similar, la salida de audio producida por el decodificador (y el representador) MP20 puede incluir uno o más de los siguientes, por ejemplo: alimentaciones para formaciones de altavoces monofónicos, estereofónicos, 5.1, 7.1 y / o 22.2; alimentaciones para formaciones de altavoces distribuidos irregularmente; alimentaciones para auriculares; audio interactivo.

[0016] También puede ser deseable seguir una filosofía de "crear una vez, usar muchas" en la que el material de audio es creado una vez (por ejemplo, por un creador de contenido) y codificado en formatos que posteriormente pueden decodificarse y representarse para diferentes salidas y configuraciones de altavoces. A un creador de contenido, tal como un estudio de Hollywood, por ejemplo, por lo general le gustaría producir la banda sonora para una película una vez y no gastar el esfuerzo para volver a mezclarla para cada posible configuración de altavoces.

[0017] La entrada al codificador MP10 del MPEG en el ejemplo estandarizado de la figura 1B es, optativamente, uno entre un formato basado en canales, un formato basado en objetos y un formato basado en escenas. Hay una multitud de ventajas de usar un formato basado en escenas. Sin embargo, una posible desventaja de usar un formato de este tipo es la falta de compatibilidad con versiones anteriores de los sistemas existentes de audio de consumo. La falta de compatibilidad con versiones anteriores al usar los SHC se debe al hecho de que los SHC no son datos de PCM. Los receptores existentes que carecen del nuevo decodificador basado en escenas (también llamados sistemas "heredados") no pueden decodificar tales señales y no podrían reproducir el programa. Esta divulgación describe procedimientos, sistemas y aparatos que pueden usarse para abordar esta falta de compatibilidad con versiones anteriores al usar conjuntos jerárquicos de coeficientes (por ejemplo, los SHC o coeficientes de otros conjuntos de funciones de base) para representar un campo sonoro.

[0018] El enfoque descrito en este documento proporciona una solución a una posible desventaja en el uso de la representación de campos sonoros basada en los SHC. Sin esta solución, la representación basada en los SHC será difícil de implementar, debido a la desventaja significativa impuesta por no poder tener funcionalidad en los millones de sistemas existentes de reproducción heredados.

[0019] La compatibilidad con versiones anteriores era una preocupación incluso cuando se introdujo por primera vez el formato estereofónico, ya que era necesario que los sistemas heredados de reproducción monofónica conservaran la compatibilidad. En este caso, la compatibilidad con versiones anteriores de mono-estéreo se mantuvo utilizando la matrización. El formato estéreo 'M-medio' y 'S-Lateral' es capaz de mantener la compatibilidad con sistemas con capacidad mono usando solo el canal M. Como se muestra en la figura 1C, los sistemas con capacidad estéreo realizan una simple operación matricial de tamaño 2×2 para decodificar los canales 'L-izquierdo' y 'R-derecho'.

[0020] La señal M-S se puede calcular a partir de la señal L-R utilizando la inversa de la matriz anterior (que resulta ser idéntica). De esta manera, el reproductor monofónico heredado conserva la funcionalidad, mientras que el reproductor estereofónico puede decodificar los canales Izquierdo y Derecho con precisión. De manera similar, se puede agregar un tercer canal que conserva la compatibilidad con versiones anteriores, preservando la funcionalidad del reproductor monofónico y del reproductor estereofónico y agregando la funcionalidad de un reproductor de tres canales.

[0021] Las extensiones de un enfoque tradicional de audio matricial basado en canales 5.1 para incluir canales adicionales pueden evitar el problema de la compatibilidad con versiones anteriores incluyendo las muestras 5.1 como un subconjunto del formato de canal extendido. Por ejemplo, la mayoría de los sistemas existentes de audio de consumo aceptan entrada de canales 5.1. En un flujo de bits extendido compatible con versiones anteriores, las muestras 5.1 se encuentran en una ubicación reconocida por los sistemas heredados, y los canales adicionales se ubican, por ejemplo, en una parte extendida de un paquete de tramas que contiene todas las muestras de canal. Otro ejemplo de tal enfoque es un flujo de bits Dolby Digital Plus (Laboratorios Dolby, Inc., San Francisco, CA), que incluye un paquete central con una mezcla 5.1 y un paquete de extensión con cuatro canales envolventes de una mezcla 7.1. Alternativamente, un decodificador puede configurarse para determinar los datos de canal 5.1 mediante la realización de una operación de matrización en el mayor número de canales.

[0022] Un enfoque propuesto para abordar el problema de la compatibilidad con versiones anteriores en un formato basado en objetos es enviar una señal de canal 5.1 mezclada a la baja junto con los objetos. En un escenario de ese tipo, los sistemas heredados basados en canales reproducirían el audio basado en canales 5.1 mezclado a la baja, mientras que los representadores más avanzados usarían una combinación de audio 5.1 y los objetos de audio individuales, o solo los objetos individuales, para representar el campo sonoro. Desafortunadamente, un enfoque de ese tipo para la compatibilidad con versiones anteriores puede utilizar el ancho de banda de manera ineficaz al enviar la misma información de audio en dos formatos diferentes.

[0023] Esta divulgación incluye descripciones de sistemas, procedimientos y aparatos en los que dos partes de un conjunto jerárquico de elementos (por ejemplo, un conjunto de SHC) que describe un campo sonoro se procesan de manera diferente. En este enfoque, un subconjunto de los elementos se convierte (por ejemplo, utilizando una matriz de transformación como se detalla a continuación) para obtener un conjunto de señales de audio de múltiples canales. Un enfoque de ese tipo se puede utilizar para convertir un subconjunto del conjunto jerárquico (un "conjunto básico") en múltiples canales de audio que sean compatibles con un formato tradicional de audio de múltiples canales. Un enfoque de ese tipo también puede configurarse para obtener cualquier número deseado de canales para mantener la compatibilidad con versiones anteriores. Ejemplos de aspectos de un sistema compatible con versiones anteriores de ese tipo se resumen en los diagramas de sistema en las figuras 2A, 2B y 3A, con explicaciones sobre ambas estructuras de codificador y decodificador.

[0024] La figura 2A muestra un ejemplo general de un enfoque de ese tipo de codificación de acuerdo a una

realización. Un uso particular es convertir un conjunto básico de un conjunto completo de los SHC (que de otro modo se podría usar si la compatibilidad con versiones anteriores no fuera un problema) en múltiples canales que representan un formato tradicional de audio de múltiples canales (por ejemplo, según lo producido por la matriz de transformación TM10 y el recodificador / transcodificador TC10). El resto del conjunto de los SHC (por ejemplo, los coeficientes que no formaban parte del subconjunto) se considera como un conjunto extendido que no está convertido y que puede codificarse por separado (por una implementación SE20 del codificador SE10 basado en escenas, por ejemplo, para la compresión de ancho de banda) para su transmisión por un canal de transmisión (por ejemplo, un canal cableado y / o inalámbrico), y / o para su almacenamiento, junto con las señales codificadas de audio de múltiples canales, compatibles con versiones anteriores (por ejemplo, en un paquete o flujo de bits, según lo descrito en este documento y lo producido en este ejemplo por el multiplexor de flujo de bits MX10). Por ejemplo, estos bits codificados se pueden empaquetar en una parte extendida de un flujo de bits o una parte extendida de un paquete para la trama (por ejemplo, una parte definida por el usuario).

[0025] En el extremo receptor, el flujo de bits se demultiplexa (por ejemplo, mediante el demultiplexor DX10), y los sistemas heredados que solo prestan soporte a decodificadores convencionales (por ejemplo, el decodificador heredado LD10) usarían solo el contenido de audio de múltiples canales compatible con versiones anteriores, conservando así la funcionalidad, e ignorarían las partes extendidas del paquete de tramas o del flujo de bits, como se muestra en la figura 2B.

[0026] Un receptor que presta soporte a la decodificación basada en escenas realiza un proceso de decodificación de acuerdo al ejemplo general que se muestra en la figura 3A de acuerdo a otra realización, lo que es recíproco a un enfoque de codificación como se muestra en la figura 2A. Si los canales se han recodificado o transcodificado, se puede realizar una etapa intermedia de decodificación. Por ejemplo, un trans-decodificador TC20 se puede utilizar para convertir el flujo de bits compatible con versiones anteriores en señales de audio de múltiples canales. Posteriormente, se utiliza una matriz de transformación inversa IM10 para convertir las señales de audio de múltiples canales en el subconjunto original del conjunto jerárquico (por ejemplo, un conjunto básico de los SHC). El resto del conjunto jerárquico (por ejemplo, un conjunto extendido de SHC) se recupera mediante una implementación SD20 del decodificador SD10 basado en escenas (por ejemplo, a partir de una parte extendida del paquete o flujo de bits). De esta manera, el representador SH SR10 puede recuperar y procesar el conjunto jerárquico completo (por ejemplo, un conjunto completo de los SHC) para permitir que tengan lugar varios tipos de representación de campos de sonido.

[0027] Usar un conjunto de los SHC para representar un campo de sonido es un ejemplo particular de un enfoque general de uso de un conjunto jerárquico de elementos para representar un campo de sonido. El conjunto jerárquico de elementos, tal como un conjunto de los SHC, es un conjunto 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 del campo de sonido en el espacio se vuelve más detallada.

[0028] Los SHC de origen (por ejemplo, como se muestra en la Figura 3A) pueden ser señales de origen mezcladas por ingenieros de mezcla en un estudio de grabación capacitado para basarse en escenas. Los SHC de origen también puede generarse a partir de señales capturadas por una formación de micrófonos o de una grabación de una presentación sonora por una formación envolvente de altavoces. También se contempla la conversión de un flujo de PCM y de información de ubicación asociada (por ejemplo, un objeto de audio) en un conjunto de origen de los SHC.

[0029] Un campo de sonido se puede representar en términos de SHC usando una expresión tal como la siguiente:

$$p_i(t, r_r, \theta_r, \varphi_r) = \sum_{\omega=0}^{\infty} \left[4\pi \sum_{n=0}^{\infty} j_n(kr_r) \sum_{m=-n}^n A_n^m(k) Y_n^m(\theta_r, \varphi_r) \right] e^{j\omega t}, \quad (1)$$

Esta expresión muestra que la presión p_i en cualquier punto $\{r_r, \theta_r, \phi_r\}$ del campo sonoro puede representarse únicamente por los SHC $A_n^m(k)$. Aquí, el número de onda $k = \frac{\omega}{c}$, c es la velocidad del sonido (~ 343 m / s), $\{r_r, \theta_r, \phi_r\}$ es un punto de referencia (o punto de observación), $j_n(\cdot)$ es la función de Bessel esférica de orden n , y $Y_n^m(\theta_r, \varphi_r)$ son las funciones de base armónicas esféricas de orden n y suborden m (algunas descripciones de las SHC etiquetan n como grado (es decir, del correspondiente polinomio de Legendre) y m como orden). Se puede reconocer que el término entre corchetes es una representación en el dominio de la frecuencia de la señal (es decir, $S(\omega, r_r, \theta_r, \phi_r)$) que puede ser aproximada por varias transformaciones de tiempo-frecuencia, tales como la transformación discreta de Fourier (DFT), la transformación de coseno discreta (DCT) o una transformación de ondículas.

[0030] La figura 4 muestra ejemplos de gráficos de malla superficial de las magnitudes de funciones de base armónicas esféricas de orden 0 y 1. La magnitud de la función Y_0^0 es esférica y omnidireccional. La función Y_1^{-1} tiene lóbulos esféricos positivos y negativos que se extienden, respectivamente, en las direcciones +y y -y. La función Y_1^0

tiene lóbulos esféricos positivos y negativos que se extienden, respectivamente, en las direcciones +z y -z. La función Y_1^1 tiene lóbulos esféricos positivos y negativos que se extienden, respectivamente, en las direcciones +x y -x.

[0031] La figura 5 muestra ejemplos de gráficos de malla superficial de las magnitudes de funciones de base armónicas esféricas de orden 2. Las funciones Y_2^{-2} y Y_2^2 tienen lóbulos que se extienden en el plano x-y. La función Y_2^{-1} tiene lóbulos que se extienden en el plano y-z y la función Y_2^1 tiene lóbulos que se extienden en el plano x-z. La función Y_2^0 tiene lóbulos positivos que se extienden en las direcciones +z y -z y un lóbulo negativo toroidal que se extiende en el plano x-y.

[0032] El número total de los SHC en el conjunto puede depender de varios factores. Para audio basado en escenas, por ejemplo, el número total de los SHC puede estar limitado por el número de transductores de micrófono en la formación de grabación. Para el audio basado en canales y objetos, el ancho de banda disponible puede determinar el número total de los SHC. En un ejemplo, se utiliza una representación de cuarto orden que implica 25 coeficientes (es decir, $0 \leq n \leq 4$, $-n \leq m \leq +n$) para cada frecuencia. Otros ejemplos de conjuntos jerárquicos que pueden utilizarse con el enfoque descrito en este documento incluyen conjuntos de coeficientes de transformación de ondículas y otros conjuntos de coeficientes de funciones de base de múltiple resolución.

[0033] Los SHC $A_n^m(k)$ se pueden obtener de señales que se adquieren físicamente (por ejemplo, se graban) utilizando cualquiera de varias configuraciones de formaciones de micrófonos, tal como una formación de micrófonos tetraédrica o esférica. La entrada de esta forma representa la entrada de audio basada en escenas a un codificador propuesto. En un ejemplo no limitativo, se supone que las entradas al codificador de SHC son los diferentes canales de salida de una formación de micrófonos, tal como un Eigenmike^R (mh acoustics LLC, San Francisco, CA). Un ejemplo de una formación Eigenmike^R es la formación em32, que incluye 32 micrófonos dispuestos en la superficie de una esfera de 8,4 centímetros de diámetro, de manera que cada una de las señales de salida $p_i(t)$, $i = 1$ a 32, es la presión registrada en la muestra cronológica t por el micrófono i. Alternativamente, los SHC $A_n^m(k)$ se pueden derivar de las descripciones basadas en canales o basadas en objetos del campo de sonido. Los coeficientes $A_n^m(k)$ para el campo sonoro correspondiente a un objeto de audio individual pueden expresarse como

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

donde i es $\sqrt{-1}$, $h_n^{(2)}(\cdot)$ es la función de Hankel esférica (de la segunda especie) de orden n, $\{r_s, \theta_s, \phi_s\}$ es la ubicación del objeto, y $g(\omega)$ es la energía de origen como función de la frecuencia.

[0034] Conocer la fuente de energía $g(\omega)$ como función de la frecuencia nos permite convertir cada objeto de PCM y su ubicación en los SHC $A_n^m(k)$. Esta energía de origen se puede obtener, por ejemplo, utilizando técnicas de análisis de tiempo-frecuencia, tal como, por ejemplo, realizando una transformada rápida de Fourier (por ejemplo, una FFT de 256, 512 o 1.024 puntos) en el flujo de PCM. Además, se puede demostrar (ya 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, estos 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 en una representación del campo sonoro global en las proximidades del punto de observación $\{r_r, \theta_r, \phi_r\}$.

[0035] Un experto en la materia reconocerá que pueden usarse representaciones de coeficientes A_n^m (o, equivalentemente, de los correspondientes coeficientes de tiempo-dominio a_n^m) distintas a la representación mostrada en la expresión (2), tales como representaciones que no incluyen el componente radial. Un experto en la técnica reconocerá que se conocen varias definiciones ligeramente diferentes de funciones de base armónicas esféricas (por ejemplo, reales, complejas, normalizadas (por ejemplo, N3D), semi-normalizadas (por ejemplo, SN3D), de Furse-Malham (FuMa o FMH), etc.) y, en consecuencia, la expresión (1) (es decir, la descomposición armónica esférica de un campo de sonido) y la expresión (2) (es decir, la descomposición armónica esférica de un campo de sonido producido por un origen puntual) pueden aparecer en la bibliografía en forma ligeramente diferente. La presente descripción no se limita a ninguna forma particular de las funciones de base armónicas esféricas y, de hecho, es generalmente aplicable asimismo a otros conjuntos jerárquicos de elementos.

[0036] La figura 3B muestra un diagrama de flujo de un procedimiento M100 para procesar una pluralidad de coeficientes de función de base que describe un campo de sonido durante un intervalo de tiempo de acuerdo a una realización. El procedimiento M100 incluye las tareas T100 y T200. La tarea T100 realiza una transformación reversible en un primer grupo (un "conjunto básico") de la pluralidad de coeficientes de función de base para producir una

pluralidad de señales de canal, en donde cada una entre la pluralidad de señales de canal está asociada a una correspondiente región diferente del espacio. En función de la pluralidad de señales de canal, la tarea T200 produce una estructura de datos que incluye (A) una representación de un segundo grupo (un "conjunto extendido") de la pluralidad de coeficientes de función de base, en donde el segundo grupo es diferente al primer grupo y (B) una representación de la pluralidad de señales de canal que es distinta a dicha representación del segundo grupo.

[0037] El diseño de una implementación del procedimiento M100 puede incluir la selección del subconjunto del conjunto jerárquico original (es decir, el conjunto básico) que se ha de convertir en audio de múltiples canales (por ejemplo, en un formato convencional). Puede esperarse que tal enfoque se implemente para mantener la compatibilidad con al menos la capacidad tradicional de 5.1 envolvente / cine en casa. Para el formato 5.1, los canales de audio de múltiples canales son Frontal Izquierdo (L), Centro (C), Frontal Derecho (R), Envolvente Izquierdo (Ls), Envolvente Derecho (Rs) y Efectos de Baja Frecuencia (LFE). El formato 7.1 agrega los canales Trasero Izquierdo (Lb) y Trasero Derecho (Rb). En las exposiciones a continuación, el formato 5.1 se usará como un típico formato de audio de destino de múltiples canales, y se elaborará enfoque ejemplar. Se contempla y se describe por la presente que la misma metodología se puede generalizar a otros formatos de audio de múltiples canales (por ejemplo, 7.1, 22.2, etc.).

[0038] Son posibles varios enfoques para seleccionar el conjunto básico. Dado que cinco señales (correspondientes a audio de banda completa desde ubicaciones específicas) están disponibles en el formato 5.1 (más la señal LFE, que no tiene ninguna ubicación estandarizada y se puede determinar mediante el filtrado de paso bajo de los cinco canales), un enfoque es utilizar cinco de los SHC para convertir al formato 5.1. Además, puede ser deseable seleccionar el conjunto básico de acuerdo a las orientaciones espaciales de las correspondientes funciones de base. Debido a que el formato 5.1 solo es capaz de representación bidimensional, por ejemplo, puede ser deseable que el conjunto básico incluya solo los SHC que contengan alguna información horizontal. En otras palabras, puede ser deseable seleccionar los SHC que correspondan a las funciones básicas cuya energía se concentra a lo largo de al menos una dirección dentro del plano que incluye las ubicaciones de los altavoces (por ejemplo, el plano x-y). Como se ve en la figura 4, el coeficiente $A_1^0(k)$ lleva muy poca información sobre la directividad horizontal, por ejemplo, y por lo tanto puede ser excluido de este subconjunto. Lo mismo es cierto para la parte real o imaginaria de $A_2^1(k)$ (por ejemplo, véanse las funciones básicas mostradas en la figura 5 como Y_2^1 y Y_2^{-1}). Los coeficientes particulares seleccionados para el conjunto básico pueden variar en función, por ejemplo, de la definición de las funciones básicas armónicas esféricas elegidas en la implementación (hay varias definiciones en la bibliografía: reales, imaginarios, complejos o combinaciones).

[0039] De tal manera, cinco coeficientes $A_n^m(k)$ pueden ser seleccionados para su conversión. Como el coeficiente $A_0^0(k)$ lleva la información omnidireccional, puede ser deseable utilizar siempre este coeficiente. De manera similar, puede ser deseable incluir la parte real de $A_1^1(k)$ y la parte imaginaria de $A_1^{-1}(k)$, ya que llevan información significativa de directividad horizontal. Para los dos últimos coeficientes, los posibles candidatos incluyen la parte real e imaginaria de $A_2^2(k)$. Numerosas otras combinaciones también son posibles. Por ejemplo, el conjunto básico puede seleccionarse para incluir solo los tres coeficientes $A_0^0(k)$, la parte real de $A_1^1(k)$ y la parte imaginaria de $A_1^{-1}(k)$. También es posible que el conjunto básico incluya más o menos coeficientes, según el número de canales deseados y también la restricción de invertibilidad en la matriz de transformación, como se expone a continuación.

[0040] Los detalles de un diseño para sistemas compatibles con 5.1, como se describe en este documento, pueden ajustarse de acuerdo a los diferentes formatos de destino. Como ejemplo, para habilitar la compatibilidad con sistemas 7.1, se agregan dos canales adicionales de contenido de audio al requisito compatible, y se pueden agregar dos SHC más al conjunto básico, de modo que la matriz de transformación resultante sea cuadrada y, por lo tanto, pueda ser invertible. Debido a que la mayoría de las disposiciones de altavoces para sistemas 7.1 (por ejemplo, Dolby TrueHD) aún se encuentran en un plano horizontal, la selección de SHC aún puede excluir aquellos con información de altura. De esta manera, la representación de la señal del plano horizontal se beneficiará de los canales de altavoces añadidos en el sistema de representación. En un sistema que incluye altavoces con diversidad de altura (por ejemplo, sistemas 9.1, 11.1 y 22.2), puede ser deseable incluir los SHC con información de altura en el conjunto básico.

[0041] Si bien hemos generalizado a canales múltiples, el énfasis principal en el mercado actual es para los canales 5.1, ya que es el "denominador menos común" para garantizar la funcionalidad de los sistemas heredados de audio de consumo, tales como los decodificadores. Para un número menor de canales, como estéreo y mono, las soluciones 5.1 existentes en muchas técnicas anteriores deberían ser suficientes para cubrir la mezcla a la baja para mantener la información del contenido. Estos casos (es decir, 1.0 y 2.0) se consideran triviales y no se exponen más en esta divulgación.

[0042] Otro problema que puede surgir es cuánto error se produce en la conversión hacia adelante y hacia atrás a partir del conjunto básico (por ejemplo, de los SHC) al audio de múltiples canales y de vuelta al conjunto básico. La tarea T100 realiza una transformación reversible en el conjunto básico de los coeficientes de la función básica para

producir una pluralidad de señales de canal, cada una asociada a una correspondiente región diferente del espacio (por ejemplo, una correspondiente ubicación diferente de altavoz). Puede ser deseable implementar la tarea T100 para aplicar una matriz invertible para convertir el conjunto básico de los SHC (por ejemplo, los cinco coeficientes seleccionados como se ha descrito anteriormente) en las cinco señales de audio de banda completa en el formato 5.1. El deseo de la invertibilidad es permitir la conversión de las cinco señales de audio de banda completa en el conjunto básico de los SHC con poca o ninguna pérdida de resolución.

[0043] Un posible procedimiento para determinar esta matriz de transformación (por ejemplo, la matriz TM10 como se muestra en la Figura 2A) es una operación conocida como "coincidencia de modalidad". Aquí, las fuentes de alimentación de los altavoces se calculan suponiendo que cada altavoz produce una onda esférica. En tal escenario, la presión (como función de la frecuencia) en una determinada posición r, θ, ϕ , debido al ℓ -ésimo altavoz, viene dada por

$$P_{\ell}(\omega, r, \theta, \phi) = g_{\ell}(\omega) \sum_{n=0}^{\infty} j_n(kr) \sum_{m=-n}^n (-4\pi i k) h_n^{(2)}(kr_{\ell}) Y_n^{m*}(\theta_{\ell}, \phi_{\ell}) Y_n^m(\theta, \phi), \quad (3)$$

donde $\{r_{\ell}, \theta_{\ell}, \phi_{\ell}\}$ representa la posición del ℓ -ésimo altavoz y $g_{\ell}(\omega)$ es la alimentación de altavoz del ℓ -ésimo altavoz (en el dominio de la frecuencia). La presión total P_t debida a los cinco altavoces viene así dada por

$$P_t(\omega, r, \theta, \phi) = \sum_{\ell=1}^5 g_{\ell}(\omega) \sum_{n=0}^{\infty} j_n(kr) \sum_{m=-n}^n (-4\pi i k) h_n^{(2)}(kr_{\ell}) Y_n^{m*}(\theta_{\ell}, \phi_{\ell}) Y_n^m(\theta, \phi). \quad (4)$$

[0044] También sabemos que la presión total en términos de los cinco SHC viene dada por la ecuación

$$P_t(\omega, r, \theta, \phi) = 4\pi \sum_{n=0}^{\infty} j_n(kr) \sum_{m=-n}^n A_n^m(k) Y_n^m(\theta, \phi). \quad (5)$$

[0045] La igualación de las dos ecuaciones anteriores nos permite usar una matriz de transformación para expresar las fuentes de los altavoces en términos de los SHC, de la siguiente manera:

$$\begin{bmatrix} A_0^0(\omega) \\ A_1^1(\omega) \\ A_1^{-1}(\omega) \\ A_2^2(\omega) \\ A_2^{-2}(\omega) \end{bmatrix} = -ik \begin{bmatrix} h_0^{(2)}(kr_1) Y_0^{0*}(\theta_1, \phi_1) & h_0^{(2)}(kr_2) Y_0^{0*}(\theta_2, \phi_2) & . & . & . \\ h_1^{(2)}(kr_1) Y_1^{1*}(\theta_1, \phi_1) & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \end{bmatrix} \begin{bmatrix} g_1(\omega) \\ g_2(\omega) \\ g_3(\omega) \\ g_4(\omega) \\ g_5(\omega) \end{bmatrix}. \quad (6)$$

[0046] Esta expresión muestra que existe una relación directa entre las cinco fuentes de alimentación de los altavoces y los SHC elegidos. La matriz de transformación puede variar según, por ejemplo, qué coeficientes se usaron en el subconjunto (por ejemplo, el conjunto básico) y qué definición de funciones de base armónicas esféricas se usa. De manera similar, se puede construir una matriz de transformación para convertir, de un conjunto básico seleccionado a un formato de canal diferente (por ejemplo, 7.1, 22.2). Si bien la matriz de transformación anterior se obtuvo de un criterio de "coincidencia de modalidad", las matrices de transformación alternativas también se pueden obtener de otros criterios, como la coincidencia de presiones, la coincidencia de energías, etc.

[0047] Si bien una matriz de transformación, tal como en la expresión anterior, permite una conversión de las fuentes de altavoz a los SHC, también nos gustaría que la matriz fuera invertible, de manera que, comenzando con los SHC, podamos averiguar las cinco fuentes de canales y luego, en el decodificador, optativamente, podamos volver a convertir a los SHC (cuando están presentes representadores avanzados (es decir, no heredados)). Puede ser deseable, por ejemplo, que la matriz de transformación produzca salidas enteras en respuesta a entradas enteras. Se pueden aprovechar varias formas de manipular el entorno anterior para garantizar la invertibilidad de la matriz. Estas incluyen, pero no se limitan a, técnicas matemáticas, tales como las técnicas de regularización (por ejemplo, la regularización dependiente de la frecuencia) y otras diversas técnicas de manipulación matricial que a menudo funcionan para garantizar rango completo y valores propios bien definidos.

[0048] Las técnicas para asegurar la invertibilidad también incluyen, pero no se limitan a, la variación de las posiciones de uno o más de los altavoces. Tales técnicas pueden incluir, por ejemplo, ajustar las posiciones de uno o más de los cinco altavoces de un sistema 5.1 de modo que aún cumplan una tolerancia angular especificada (por ejemplo, según lo especificado por la recomendación UIT-R BS.775-1, Unión Internacional de Telecomunicaciones,

Ginebra, CH). La figura 6 muestra un ejemplo compatible de ángulos de altavoces para un sistema 5.1 que se pueden usar. Los espaciados regulares de los transductores, tales como los que cumplen el diseño en T, suelen comportarse bien.

[0049] Alternativamente, la tarea T100 puede implementarse para aplicar una matriz de transformación que convierte el conjunto básico en un conjunto de señales de canal que corresponden a ángulos de altavoz que están espaciados uniformemente en el plano. La invertibilidad está habitualmente asegurada para una matriz de transformación de este tipo. La figura 7 muestra un ejemplo de una disposición de este tipo de ángulos de altavoz para un sistema 5.1 de destino. Aunque una disposición así espaciada uniformemente puede no ser exactamente compatible con la norma 5.1, todavía proporciona un conjunto adecuado de alimentaciones de altavoz compatibles con versiones anteriores y también proporciona un procedimiento fiable para obtener una matriz de transformación invertible para recuperar el conjunto básico de coeficientes. Para implementaciones de siete canales de la tarea T100, las figuras 8 y 9 muestran, respectivamente, un ejemplo similar de un conjunto de ángulos de altavoz compatibles con la norma para un sistema 7.1 y un conjunto correspondiente de ángulos de altavoz que están espaciados uniformemente en el plano.

[0050] Para un formato de canal de destino que incluya al menos un canal de efectos de baja frecuencia (LFE), tal como 5.1, 7.1 y 22.2, puede ser deseable implementar el procedimiento M100 para producir el conjunto de señales de canal para incluir asimismo un canal de ese tipo. En un ejemplo, el procedimiento M100 incluye una tarea que produce el canal de LFE mediante el filtrado de paso bajo de las señales de canal producidas por la tarea T100 y la suma de las señales de paso bajo resultantes. Para un caso en el que se necesitan múltiples canales de LFE (por ejemplo, canales de LFE izquierdo y derecho para un formato 22.2), una tarea de ese tipo puede implementarse para producir estos canales sumando las señales de paso bajo que sean espacialmente más cercanas a cada ubicación de destino. Aunque un canal de LFE producido de esta manera es redundante, aún proporciona una señal adecuada para el canal de LFE heredado, y tal enfoque también evita cualquier cambio en la matriz de transformación.

[0051] Puede ser deseable probar (por ejemplo, psico-acústicamente) las señales de canal representadas por la matriz de transformación para garantizar que, después de toda la manipulación, la matriz modificada realmente produzca alimentaciones de altavoces correctas y / o aceptables. Mientras se mantenga la invertibilidad, el problema inverso de garantizar la correcta decodificación de los SHC no suele ser un problema. Habitualmente, es suficiente que se pueda obtener una matriz que permita la transformación entre el conjunto básico (por ejemplo, un subconjunto de los SHC) y el audio tradicional de múltiples canales y también que, después de la manipulación (que no reduce la fidelidad del audio de múltiples canales), también se pueda formular una matriz ligeramente modificada que también sea invertible.

[0052] En función de la pluralidad de señales de canal, la tarea T200 produce una estructura de datos que incluye (A) una representación de un segundo grupo de la pluralidad de coeficientes de funciones de base (por ejemplo, un "conjunto extendido"), en donde el segundo grupo es diferente al primer grupo, y (B) una representación de la pluralidad de señales de canal que es distinta a dicha representación del segundo grupo. La estructura de datos puede ser una estructura en el tiempo (por ejemplo, que incluya una o más flujos de bits en vivo o flujos de paquetes) y / o una estructura en el espacio (por ejemplo, que incluya uno o más flujos de bits almacenados en un medio de almacenamiento, o uno o más paquetes almacenados en un almacén temporal).

[0053] Como se muestra en la expresión (6), la tarea T100 puede implementarse para aplicar la matriz de transformación al conjunto básico en cada frecuencia para producir las señales de canal en un dominio de frecuencia. El procedimiento M100 puede implementarse para incluir una tarea que realice una transformación de frecuencia inversa (por ejemplo, una FFT inversa) en cada una de estas señales de canal en el dominio de la frecuencia, para producir una correspondiente señal de canal en el dominio del tiempo (por ejemplo, un flujo de PCM lineal).

[0054] Las señales de canal pueden transmitirse como flujos de PCM lineales mediante una interfaz de HDMI (Interfaz de Multimedia de Alta Definición, HDMI Licensing, LLC, Sunnyvale, CA). En otro ejemplo, las señales de canal pueden almacenarse como flujos de PCM lineales en un disco óptico, tales como un CD, DVD, DVD-Audio o disco Blu-Ray. Un disco Blu-Ray (por ejemplo, un medio de almacenamiento óptico de datos compatible con la Definición de Aplicación de Disco Blu-Ray BD-J, marzo de 2005, Asociación del Disco Blu-Ray, www.punto-blu-raydisc-punto-com) puede incluir un fichero 'zzzzz.m2ts' que contiene un flujo de transporte de MPEG-2, donde 'zzzzz' es un número de cinco dígitos que asocia el fichero de flujo de AV con un archivo de información de recortes. El fichero de flujo 'zzzzz.m2ts' puede incluir múltiples flujos elementales de audio. La tarea T200 puede implementarse para producir un archivo de flujo de este tipo que incluya versiones en el dominio del tiempo de las señales de canal producidas por la tarea T100 como flujos de LPCM.

[0055] Para reducir el uso de ancho de banda y / o recursos de almacenamiento, puede ser deseable implementar la tarea T200 para comprimir los flujos de canal de LPCM. Para garantizar la capacidad de recuperación del conjunto básico de los SHC, puede ser deseable realizar dicha compresión utilizando un esquema de compresión sin pérdidas. En un ejemplo, la tarea T200 se implementa para codificar los flujos de PCM utilizando el Empaquetamiento sin Pérdidas Meridian (MLP) para producir un flujo de bits que sea compatible con el DVD-Audio. En otro ejemplo, la tarea T200 se implementa para codificar los flujos de PCM utilizando la extensión sin pérdidas SLS (ajustable a escala

hasta sin pérdidas) de MPEG-4 al códec principal AAC. En un ejemplo adicional, la tarea T200 se implementa para producir un fichero de flujo (por ejemplo, un fichero m2ts compatible con Blu-Ray, como se ha descrito anteriormente) que incluye flujos de audio elementales producidos por la codificación sin pérdidas de los flujos de PCM utilizando Dolby TrueHD, que codifica audio 7.1 utilizando una versión mejorada de MLP y / o DTS-HD Master Audio (DTS, Inc., Calabasas, CA), que también codifica audio 7.1 con una opción sin pérdidas.

[0056] La tarea T200 puede implementarse de otro modo para codificar las señales de canal en señales codificadas compatibles con versiones anteriores que describen las señales de canal (por ejemplo, como se ilustra en la Figura 2A mediante la operación del transcodificador TC10). Dicha codificación puede incluir realizar un esquema de compresión con pérdidas en las señales del canal. Los ejemplos de códecs compatibles con versiones anteriores que se pueden usar en tales implementaciones de la tarea T200 incluyen el AC3 (por ejemplo, según lo descrito en la Norma ATSC: Compresión de audio digital, doc. A/52:2012, 23 de marzo de 2012, Comité de Sistemas Avanzados de Televisión, Washington, DC; también llamado ATSC A/52 o Dolby Digital, que utiliza compresión MDCT con pérdidas), Dolby TrueHD (que incluye opciones de compresión con pérdidas), Master Audio DTS-HD (que también incluye opciones de compresión con pérdidas) y Envolvente MPEG (MPS, ISO/IEC 14496-3, también denominado codificación avanzada de audio de alta eficacia o HeAAC, en el que los seis canales de una señal de formato 5.1 se mezclan a la baja en un flujo de PCM mono o estéreo, con la información lateral correspondiente, tal como la diferencia de nivel interaural, la diferencia de tiempo interaural y la coherencia entre canales, que permite la síntesis del resto de los canales en el representador). Estos códecs habitualmente aceptan señales de canal de dominio de tiempo (por ejemplo, un conjunto de flujos de PCM lineales) como entrada. Dicha transcodificación permite que las señales del canal conserven la compatibilidad con versiones anteriores con los decodificadores AC3 que se encuentran en muchos dispositivos de consumo y decodificadores. Por ejemplo, los canales codificados se pueden empaquetar en una parte correspondiente de un paquete que cumple con un correspondiente formato deseado basado en canales.

[0057] En tales casos, el procedimiento M100 puede implementarse para codificar el resto del conjunto jerárquico (por ejemplo, los SHC que no formaban parte del subconjunto) por separado para la transmisión (y / o almacenamiento) en una o más partes extendidas del flujo de bits transcodificado (por ejemplo, una porción "datosaux" de un paquete AC3). Dicha operación se ilustra en la figura 2A, por ejemplo, mediante la operación del multiplexador de flujo de bits MX10. También es posible que tal implementación del procedimiento M100 incluya dos o más operaciones de transcodificación diferentes, cada una de las cuales codifique la señal multicanal en un formato respectivo diferente (por ejemplo, una transcodificación AC3 y una transcodificación Dolby TrueHD), para producir dos flujos de bits diferentes compatibles con versiones anteriores para su transmisión y / o almacenamiento.

[0058] Como se ha indicado anteriormente, un flujo de bits Dolby Digital Plus incluye un paquete central (A) con una mezcla 5.1 y un paquete de extensión (B) con cuatro canales envolventes de una mezcla 7.1. Una implementación del procedimiento M100 para la compatibilidad con un códec de ese tipo puede incluir una implementación de la tarea T100 configurada para transformar un conjunto básico de siete SHC (para cada frecuencia) en siete canales para su codificación en dichos paquetes A y B. En tal caso, la tarea T200 puede implementarse para incluir los SHC octavo y superiores en el conjunto extendido o, alternativamente, para incluir los SHC sexto y superiores en el conjunto extendido, donde el conjunto extendido puede codificarse en uno o más paquetes de extensión adicionales.

[0059] Puede ser deseable implementar la tarea T200 para producir paquetes que correspondan a intervalos de tiempo de diferentes longitudes. Por ejemplo, se puede usar un intervalo más corto para codificar sucesos transitorios en la entrada de audio, mientras que se puede usar un intervalo más largo para codificar señales estacionarias. En un ejemplo de ese tipo, la tarea T200 se implementa para codificar un intervalo de tiempo constante en los canales de PCM, para la compatibilidad con versiones anteriores, pero para codificar diferentes longitudes de intervalo en el conjunto extendido. En otro ejemplo de ese tipo, la tarea T200 se implementa para codificar intervalos de diferentes longitudes en los flujos de PCM, junto con una indicación de la duración del intervalo de tiempo correspondiente. Para un caso en el que se utiliza la codificación AC-3, por ejemplo, la tarea T200 puede implementarse para conmutar entre las dos opciones de longitud de bloque de 256 y 512 muestras.

[0060] Como se ilustra en la figura 2A, el conjunto extendido de coeficientes puede codificarse antes de asociarse (por ejemplo, multiplexarse) con las señales de canal codificadas para su transmisión y / o almacenamiento. Dicha codificación puede incluir compresión de ancho de banda. Por ejemplo, el conjunto extendido puede codificarse (por ejemplo, mediante el codificador SE20) aplicando una o más técnicas de codificación con pérdidas o sin pérdidas, tales como la cuantización (por ejemplo, en uno o más índices de libro de códigos), la codificación de corrección de errores, la codificación de redundancia, etc., y / o la paquetización. Adicional o alternativamente, dicha codificación puede incluir codificación en un formato Ambisónico, tal como el formato B, el formato G o Ambisónicos de orden superior (HOA). En un ejemplo, el conjunto extendido de coeficientes se codifica en formato B de HOA y luego las señales en formato B se codifican utilizando la Codificación de audio avanzada (AAC; por ejemplo, como se define en ISO/IEC 14496-3:2009, "Tecnología de la información -- Codificación de objetos audiovisuales. Parte 3: Audio, "Organización Internacional de Estandarización, Ginebra, CH). Las descripciones de otros procedimientos para codificar conjuntos de SHC que pueden realizarse en el conjunto extendido (por ejemplo, mediante el codificador SE20) se pueden encontrar, por ejemplo, en la Publicación Estadounidense Solicitudes de Patente N° 2012/0155653 A1 (Jax et al.) y 2012/0314878 A1 (Daniel et al.). El conjunto extendido de coeficientes puede codificarse, por ejemplo, como diferencias entre coeficientes de diferentes órdenes y / o diferencias entre coeficientes

del mismo orden en diferentes momentos.

[0061] Puede ser deseable implementar el codificador de MPEG MP10, como se muestra en la figura 1B, para realizar una implementación del procedimiento M100 según lo descrito en este documento (por ejemplo, para producir un flujo de bits para su transmisión, difusión, multidifusión, almacenamiento y / o masterización de medios (por ejemplo, masterización de CD, DVD y / o disco Blu-Ray™)).

[0062] La obtención del conjunto completo de coeficientes de función de base a partir de la estructura de datos compatible con versiones anteriores se puede realizar de manera análoga, eliminando cualquier capa de codificación y / o transformación de dominio intervinientes (por ejemplo, mediante el demultiplexador DX10, el trans-decodificador TC20 y el decodificador basado en escenas SD20) para obtener las señales de canal y luego aplicar una inversa de la transformación reversible descrita anteriormente (por ejemplo, la matriz IM10 según se muestra en la Figura 3A) a las señales de canal. La figura 10A muestra un diagrama de flujo de un procedimiento M200, de acuerdo a una configuración general, de obtención de una pluralidad de coeficientes de función de base que describen un campo de sonido durante un intervalo de tiempo, de acuerdo a una realización adicional. Como se ha descrito anteriormente, cada uno entre la pluralidad de coeficientes de función de base corresponde a una función única entre un conjunto de funciones de base ortogonales (por ejemplo, a una función única entre un conjunto de funciones de base armónicas esféricas).

[0063] El procedimiento M200 incluye las tareas T300 y T400. A partir de una estructura de datos, la tarea T300 obtiene (A) una representación de un segundo grupo (el "conjunto extendido") de la pluralidad de coeficientes de función de base y (B) una representación de una pluralidad de señales de canal que es distinta a dicha representación del segundo grupo. Cada una entre un subconjunto de la pluralidad de señales de canal se asocia a una correspondiente región diferente del espacio (por ejemplo, las direcciones coplanares de un formato 5.1 o 7.1), mientras que la pluralidad de señales de canal también puede incluir una o más señales no direccionales (u omnidireccionales), como los canales de LFE. Puede ser deseable implementar la tarea T300 para realizar una FFT en un conjunto de flujos de LPCM para obtener el subconjunto de la pluralidad de señales de canal en el dominio de la frecuencia. La tarea T400 realiza una transformación en el subconjunto de la pluralidad de señales de canal para producir un primer grupo (el "conjunto básico") de la pluralidad de coeficientes de función base, en donde el primer grupo es diferente al segundo grupo. Puede ser deseable ejecutar el procedimiento M200 solo al detectar que una representación de un conjunto extendido de coeficientes está presente en la estructura de datos y, de lo contrario, representar las señales de canal de forma compatible con versiones anteriores (por ejemplo, como alimentaciones de altavoces) sin transformarlas en un conjunto básico de coeficientes.

[0064] El procedimiento M200 puede implementarse para representar el conjunto completo de coeficientes de función base en una geometría de altavoz que es diferente a la asociada a la pluralidad de señales de canal de las cuales se recuperó el conjunto básico. La figura 10B muestra un diagrama de flujo de una implementación M210 del procedimiento M200 que incluye una tarea T500. Basándose en la pluralidad de coeficientes de función base, la tarea T500 produce una segunda pluralidad de señales de canal. En este caso, cada señal del subconjunto de la pluralidad de señales de canal está asociada a una correspondiente dirección diferente entre un conjunto de direcciones coplanares, mientras que cada una entre la segunda pluralidad de señales de canal está asociada a una correspondiente dirección diferente entre una serie de direcciones que abarcan un espacio tridimensional. Por ejemplo, la segunda pluralidad de señales de canal puede incluir uno o más canales que corresponden a ubicaciones de altavoces que están por encima del plano de referencia o que de otro modo proporcionan información de altura. En un ejemplo, la tarea T500 se implementa para aplicar una matriz de representación para una geometría de altavoz particular que se obtiene de manera análoga a la obtención de la matriz de transformación para convertir el conjunto básico de coeficientes en las señales de canal compatibles con versiones anteriores, como se ha descrito anteriormente (por ejemplo, con referencia a las expresiones (3) a (6) pero, habitualmente, sin ninguna restricción de invertibilidad).

[0065] Por lo tanto, lo anterior representa un mecanismo sin pérdidas para convertir entre un conjunto jerárquico de elementos (por ejemplo, un conjunto de SHC) y múltiples canales de audio. No se incurre en errores siempre que las señales de audio de múltiples canales no estén sujetas a ruido adicional de codificación. En caso de que estén sujetas a ruido de codificación (por ejemplo, mediante una operación de compresión con pérdidas), la conversión a los SHC puede incurrir en errores. Sin embargo, es posible tener en cuenta estos errores monitorizando los valores de los coeficientes y emprendiendo la acción adecuada para reducir su efecto. Estos procedimientos pueden tener en cuenta características de los SHC, incluida la redundancia inherente en la representación de los SHC.

[0066] La figura 3C muestra un diagrama de bloques de un aparato MF100 para procesar una pluralidad de coeficientes de función base que describen un campo de sonido durante un intervalo de tiempo, de acuerdo a una realización adicional. El aparato MF100 incluye medios F100 para realizar una transformación reversible en un primer grupo (un "conjunto básico") de la pluralidad de coeficientes de función base, para producir una pluralidad de señales de canal, en donde cada una entre la pluralidad de señales de canal está asociada a una correspondiente región diferente del espacio (por ejemplo, como se describe en este documento con referencia a implementaciones de la tarea T100 y la matriz de transformación TM10). El aparato MF100 también incluye medios F200 para producir una estructura de datos, basándose en la pluralidad de señales de canal, que incluye (A) una representación de un

segundo grupo (un "conjunto extendido") de la pluralidad de coeficientes de función base, en donde el segundo grupo es diferente al primer grupo, y (B) una representación de la pluralidad de señales de canal que es distinta a dicha representación del segundo grupo (por ejemplo, como se describe en este documento con referencia a las implementaciones de la tarea T200, el trans-codificador TC10, el codificador SE20 y el multiplexador MX10).

[0067] La figura 3D muestra un diagrama de bloques de un aparato A100 para procesar una pluralidad de coeficientes de función base que describen un campo de sonido durante un intervalo de tiempo, de acuerdo a una realización adicional. El aparato A100 incluye una calculadora 100 configurada para realizar una transformación reversible en un primer grupo (un "conjunto básico") de la pluralidad de coeficientes de función base, para producir una pluralidad de señales de canal, en donde cada una entre la pluralidad de señales de canal está asociada a una correspondiente región diferente del espacio (por ejemplo, como se describe en este documento con referencia a las implementaciones de la tarea T100 y la matriz de transformación TM10). El aparato A100 también incluye un formateador de datos 200 configurado para producir una estructura de datos, basándose en la pluralidad de señales de canal, que incluye (A) una representación de un segundo grupo (un "conjunto extendido") de la pluralidad de coeficientes de función base, en donde el segundo grupo es diferente al primer grupo, y (B) una representación de la pluralidad de señales de canal que es distinta a dicha representación del segundo grupo (por ejemplo, como se describe en este documento con referencia a implementaciones de la tarea T200, el trans-codificador TC10, el codificador SE20 y el multiplexador MX10). El formateador 200 puede incluir un empaquetador.

[0068] La figura 11A muestra un diagrama de bloques de un aparato MF200, de acuerdo a una configuración general, para obtener una pluralidad de coeficientes de función base que describen un campo de sonido durante un intervalo de tiempo, de acuerdo a una realización adicional. El aparato MF200 incluye medios F300 para obtener, a partir de una estructura de datos, (A) una representación de un segundo grupo (el "conjunto extendido") de la pluralidad de coeficientes de función base y (B) una representación de una pluralidad de señales de canal que es distinta a dicha representación del segundo grupo (por ejemplo, como se describe en este documento con referencia a las implementaciones de la tarea T300, el demultiplexador DX10, el trans-codificador TC20 y el decodificador SD20). El aparato MF200 también incluye medios F400 para realizar una transformación en el subconjunto de la pluralidad de señales de canal, para producir un primer grupo (el "conjunto básico") de la pluralidad de coeficientes de función base, en donde el primer grupo es diferente al segundo grupo (por ejemplo, como se describe en este documento con referencia a las implementaciones de la tarea T400 y la matriz de transformación inversa IM10). La figura 11B muestra un diagrama de flujo de una implementación MF210 del aparato MF200 que incluye medios F500 para producir una segunda pluralidad de señales de canal basándose en la pluralidad de coeficientes de función base (por ejemplo, como se describe en este documento con referencia a las implementaciones de la tarea T500 y el representador SR10).

[0069] La figura 11C muestra un diagrama de bloques de un aparato A200, de acuerdo a una configuración general de otra realización, para obtener una pluralidad de coeficientes de función base que describen un campo de sonido durante un intervalo de tiempo. El aparato A200 incluye un extractor de datos 300 configurado para obtener, a partir de una estructura de datos, (A) una representación de un segundo grupo (el "conjunto extendido") entre la pluralidad de coeficientes de función base y (B) una representación de una pluralidad de señales de canal que es distinta a dicha representación del segundo grupo (por ejemplo, como se describe en este documento con referencia a las implementaciones de la tarea T300, el demultiplexador DX10, el trans-codificador TC20 y el decodificador SD20). El extractor 300 puede incluir un desempaquetador. El aparato A200 también incluye una calculadora 400 configurada para realizar una transformación en el subconjunto de la pluralidad de señales de canal, para producir un primer grupo (el "conjunto básico") de la pluralidad de coeficientes de función base, en donde el primer grupo es diferente al segundo grupo (por ejemplo, como se describe en el presente documento con referencia a las implementaciones de la tarea T400 y la matriz de transformación inversa IM10). La figura 11D muestra un diagrama de flujo de una implementación A210 del aparato A200 que incluye un representador 500 configurado para producir una segunda pluralidad de señales de canal basándose en la pluralidad de coeficientes de función base (por ejemplo, como se describe en este documento con referencia a las implementaciones de la tarea T500 y el representador SR10).

[0070] En un ejemplo adicional, el procedimiento M100 se implementa para incluir una tarea que combina dos o más conjuntos de coeficientes de función base, donde cada conjunto describe un correspondiente campo de sonido durante un correspondiente intervalo de tiempo, y proporciona el conjunto combinado como entrada para la tarea T100. Por ejemplo, una tarea de este tipo puede implementarse para agregar la pluralidad de conjuntos de coeficientes (por ejemplo, para realizar la suma vectorial de coeficientes de los vectores de SHC), para producir una descripción de un campo de sonido combinado. Cada conjunto de coeficientes de función base puede representar, por ejemplo, un objeto de audio. Un conjunto de dichos coeficientes de función base (por ejemplo, el vector de los SHC para un objeto) puede tener un orden superior (por ejemplo, una longitud más larga) que otro conjunto de coeficientes de función base (por ejemplo, para otro de los objetos). Por ejemplo, un objeto en primer plano (por ejemplo, la voz de un actor principal) puede representarse con un conjunto de orden superior al de un objeto en segundo plano (por ejemplo, un efecto de sonido).

[0071] La figura 11E muestra un diagrama de bloques para un aparato P100 de acuerdo a una configuración general de una realización. El aparato P100 incluye un procesador P10 configurado para realizar una implementación del procedimiento M100, como se describe en este documento, en una pluralidad de coeficientes de función base que

describen un campo de sonido durante un intervalo de tiempo, para producir una correspondiente estructura de datos. El aparato P100 también incluye una memoria ME10 configurada para almacenar la estructura de datos producida por el procesador P10.

- 5 **[0072]** Puede ser deseable realizar un procedimiento como se describe en este documento (por ejemplo, el procedimiento M100 o M200) dentro de un dispositivo portátil de detección de audio que tiene una formación de dos o más micrófonos configurados para recibir señales acústicas. Los ejemplos de un dispositivo portátil de detección de audio que puede implementarse para incluir una formación de este tipo y pueden usarse para grabaciones de audio y / o aplicaciones de comunicaciones de voz incluyen un equipo telefónico de mano (por ejemplo, un teléfono celular);
- 10 un auricular con cable o inalámbrico (por ejemplo, un auricular Bluetooth); una grabadora de audio y / o vídeo de mano (por ejemplo, una videocámara); un reproductor de medios personal, configurado para grabar contenido de audio y / o vídeo; un asistente digital personal (PDA) u otro dispositivo informático de mano; y un ordenador portátil, un ordenador plegable, un ordenador portátil en red, un ordenador de tableta u otro dispositivo informático portátil.
- 15 **[0073]** Los procedimientos y los aparatos divulgados en el presente documento pueden aplicarse en general en cualquier aplicación transceptora y/o detectora de audio, incluyendo casos portátiles de otro modo de dichas aplicaciones y / o detección de componentes de señal desde orígenes de campos lejanos. Por ejemplo, el rango de configuraciones divulgadas en el presente documento incluye dispositivos de comunicaciones que residen en un sistema de comunicación de telefonía inalámbrica, configurado para emplear una interfaz aérea de acceso múltiple por división de código (CDMA). Sin embargo, los expertos en la técnica entenderían que un procedimiento y un aparato que tengan características según lo descrito en el presente documento pueden residir en cualquiera de los diversos sistemas de comunicación que empleen una amplia gama de tecnologías conocidas por los expertos en la técnica, tales como sistemas que emplean Voz sobre IP (VoIP) por canales de transmisión cableados y/o inalámbricos (por ejemplo, de CDMA, TDMA, FDMA y/o TD-SCDMA).
- 20 **[0074]** Se contempla expresamente y se divulga por la presente que los dispositivos de comunicaciones divulgados en el presente documento (por ejemplo, teléfonos inteligentes, ordenadores de tableta) pueden adaptarse para su uso en redes que estén conmutadas por paquetes (por ejemplo, redes cableadas y/o inalámbricas dispuestas para transportar transmisiones de audio de acuerdo a protocolos tales como VoIP) y/o conmutadas por circuitos. También se contempla expresamente y se divulga por la presente que los dispositivos de comunicaciones divulgados en el presente documento pueden adaptarse para su uso en sistemas de codificación de banda estrecha (por ejemplo, sistemas que codifiquen un rango de frecuencia de audio de aproximadamente cuatro o cinco kilohercios) y/o para su uso en sistemas de codificación de banda ancha (por ejemplo, sistemas que codifiquen frecuencias de audio superiores a cinco kilohercios), incluyendo los sistemas de codificación de banda ancha de banda completa y los
- 25 sistemas de codificación de banda ancha de banda dividida.
- [0075]** Se proporciona la presentación anterior de las configuraciones descritas con el fin de permitir que cualquier experto en la técnica realice o use los procedimientos y otras estructuras divulgados en el presente documento. Los diagramas de flujo, los diagramas de bloques y otras estructuras mostradas y descritas en el presente documento son solamente ejemplos, y otras variantes de estas estructuras también están dentro del alcance de la divulgación. Son posibles diversas modificaciones de estas configuraciones, y los principios genéricos presentados en el presente documento pueden aplicarse también a otras configuraciones.
- 30 **[0076]** Los expertos en la técnica entenderán que la información y las señales pueden representarse usando cualquiera entre varias tecnologías y técnicas diferentes. Por ejemplo, los datos, las instrucciones, los comandos, la información, las señales, los bits y los símbolos que puedan haberse mencionado a lo largo de la descripción anterior pueden representarse mediante tensiones, corrientes, ondas electromagnéticas, campos o partículas magnéticos, campos o partículas ópticos, o cualquier combinación de los mismos.
- 35 **[0077]** Los requisitos importantes de diseño para la implementación de una configuración según lo divulgado en el presente documento pueden incluir minimizar el retardo del procesamiento y/o la complejidad de cálculo (medidos habitualmente en millones de instrucciones por segundo o MIPS), especialmente para aplicaciones intensivos en cálculos, tales como la reproducción de audio comprimido o de información audiovisual (por ejemplo, un fichero o flujo codificada de acuerdo a un formato de compresión, tal como uno de los ejemplos identificados en el presente documento) o aplicaciones para comunicaciones de banda ancha (por ejemplo, comunicaciones de voz a frecuencias de muestreo superiores a ocho kilohercios, tales como 12, 16, 32, 44,1, 48 o 192 kHz).
- 40 **[0078]** Las metas de un sistema de procesamiento de múltiples micrófonos pueden incluir lograr de diez a doce dB en la reducción general de ruido, preservar el nivel de voz y el color durante el movimiento de un altavoz deseado, obtener una percepción de que el ruido se ha trasladado al fondo en lugar de una eliminación agresiva del ruido, eliminación de la reverberación del habla y / o habilitación de la opción de pos-procesamiento para una reducción de ruido más agresiva.
- 45 **[0079]** Un aparato como el divulgado en el presente documento (por ejemplo, cualquiera entre los aparatos A100, A200, A210, MF100, MF200, MF210 y P100) se puede implementar en cualquier combinación de hardware con software, y/o con firmware, que se considere adecuada para la aplicación deseada. Por ejemplo, los elementos de un
- 50
- 55
- 60
- 65

aparato de ese tipo pueden fabricarse como dispositivos electrónicos y/u ópticos que residan, por ejemplo, en el mismo chip o entre dos o más chips en un conjunto de chips. Un ejemplo de un dispositivo de ese tipo es una formación fija o programable de elementos lógicos, tales como transistores o compuertas lógicas, y cualquiera de estos elementos puede implementarse como una o más de dichas formaciones. Dos o más cualesquiera de, o incluso todos, estos elementos pueden implementarse dentro de las mismas una o más formaciones. Tales una o más formaciones pueden implementarse dentro de uno o más chips (por ejemplo, dentro de un conjunto de chips que incluya dos o más chips).

[0080] Uno o más elementos de las diversas implementaciones de los aparatos divulgados en el presente documento (por ejemplo, cualquiera de los aparatos A100, A200, A210, MF100, MF200, MF210 y P100) puede implementarse total o parcialmente como uno o más conjuntos de instrucciones dispuestas para ejecutarse en una o más formaciones fijas o programables de elementos lógicos, tales como microprocesadores, procesadores incorporados, núcleos del IP, procesadores de señales digitales, FPGA (formaciones de compuertas programables en el terreno), ASSP (productos estándar específicos de la aplicación) y ASIC (circuitos integrados específicos de la aplicación). Cualquiera de los diversos elementos de una implementación de un aparato según lo divulgado en el presente documento también puede realizarse como uno o más ordenadores (por ejemplo, máquinas que incluyan una o más formaciones programadas para ejecutar uno o más conjuntos o secuencias de instrucciones, también denominadas "procesadores"), y dos o más cualesquiera de, o incluso todos, estos elementos pueden implementarse dentro de dichos mismos uno o más ordenadores.

[0081] Un procesador u otros medios para procesar, según lo divulgado en el presente documento (por ejemplo, el procesador P10), pueden fabricarse como uno o más dispositivos electrónicos y/u ópticos que residan, por ejemplo, en el mismo chip o entre dos o más chips en un conjunto de chips. Un ejemplo de un dispositivo de ese tipo es una formación fija o programable de elementos lógicos, tales como transistores o compuertas lógicas, y cualquiera de estos elementos puede implementarse como una o más de dichas formaciones. Tales una o más formaciones pueden implementarse dentro de uno o más chips (por ejemplo, dentro de un conjunto de chips que incluya dos o más chips). Los ejemplos de dichas formaciones incluyen formaciones fijas o programables de elementos lógicos, tales como microprocesadores, procesadores integrados, núcleos del IP, DSP, FPGA, ASSP y ASIC. Un procesador u otros medios para el procesamiento, como se divulga en el presente documento, también pueden realizarse como uno o más ordenadores (por ejemplo, máquinas que incluyan una o más formaciones programadas para ejecutar uno o más conjuntos o secuencias de instrucciones) u otros procesadores. Es posible que un procesador como se describe en el presente documento sea usado para realizar tareas o ejecutar otros conjuntos de instrucciones que no estén directamente relacionadas con un procedimiento de codificación de audio según lo descrito en este documento, tal como una tarea relacionada con otra operación de un dispositivo o sistema en el que el aparato esté incorporado (por ejemplo, un dispositivo de detección de audio). También es posible que parte de un procedimiento como se divulga en el presente documento sea realizada por un procesador del dispositivo de detección de audio y que otra parte del procedimiento sea realizado bajo el control de otros uno o más procesadores.

[0082] Los expertos en la técnica apreciarán que los diversos módulos, bloques lógicos, circuitos y pruebas ilustrativos y otras operaciones descritas en relación con las configuraciones divulgadas en el presente documento puedan implementarse como hardware electrónico, software informático o combinaciones de ambos. Tales módulos, bloques lógicos, circuitos y operaciones pueden implementarse o realizarse con un procesador de propósito general, un procesador de señales digitales (DSP), un ASIC o un ASSP, una FPGA u otro dispositivo lógico programable, compuerta discreta o lógica de transistor, componentes de hardware discretos o cualquier combinación de los mismos diseñada para producir la configuración según lo divulgado en el presente documento. Por ejemplo, una configuración de ese tipo puede implementarse, al menos parcialmente, como un circuito cableado, como una configuración de circuito fabricada en un circuito integrado específico de la aplicación o como un programa de firmware cargado en un almacenamiento no volátil o un programa de software cargado desde, o en, un medio de almacenamiento de datos como código legible por máquina, siendo dicho código instrucciones ejecutables por una formación de elementos lógicos tales como un procesador de propósito general u otra unidad de procesamiento de señales digitales. Un procesador de propósito general puede ser un microprocesador pero, como alternativa, el procesador puede ser cualquier procesador, controlador, microcontrolador o máquina de estados convencional. Un procesador también puede implementarse como una combinación de dispositivos informáticos, por ejemplo, una combinación de un DSP y un microprocesador, una pluralidad de microprocesadores, uno o más microprocesadores junto con un núcleo de DSP o cualquier otra configuración de este tipo. Un módulo de software puede residir en un medio de almacenamiento no transitorio tal como RAM (memoria de acceso aleatorio), ROM (memoria de sólo lectura), RAM no volátil (NVRAM) tal como memoria RAM flash, ROM programable borrable (EPROM), ROM programable eléctricamente borrable (EEPROM), registros, disco duro, un disco extraíble o un CD-ROM; o en cualquier otra forma de medio de almacenamiento conocido en la técnica. Un medio de almacenamiento ilustrativo está acoplado al procesador de tal manera que el procesador pueda leer información de, y escribir información en, el medio de almacenamiento. De forma alternativa, el medio de almacenamiento puede estar integrado en el procesador. El procesador y el medio de almacenamiento pueden residir en un ASIC. El ASIC puede residir en un terminal de usuario. De forma alternativa, el procesador y el medio de almacenamiento pueden residir como componentes discretos en un terminal de usuario.

[0083] Se observa que los diversos procedimientos divulgados en el presente documento (por ejemplo, cualquiera entre los procedimientos M100, M200 y M210) pueden ser realizados por una formación de elementos lógicos tales como un procesador, y que los diversos elementos de un aparato, según lo descrito en el presente documento, pueden

implementarse como módulos diseñados para ejecutarse en dicha formación. Como se usa en el presente documento, el término "módulo" o "submódulo" puede referirse a cualquier procedimiento, aparato, dispositivo, unidad o medio de almacenamiento de datos legible por ordenador que incluya instrucciones de ordenador (por ejemplo, expresiones lógicas) en forma de software, hardware o firmware. Ha de entenderse que múltiples módulos o sistemas pueden combinarse en un módulo o sistema y que un módulo o sistema puede separarse en múltiples módulos o sistemas para realizar las mismas funciones. Cuando se implementan en software o en otras instrucciones ejecutables por ordenador, los elementos de un proceso son esencialmente los segmentos de código para realizar las tareas relacionadas, tales como rutinas, programas, objetos, componentes, estructuras de datos y similares. El término "software" debería entenderse como incluyente de código fuente, código de lenguaje ensamblador, código de máquina, código binario, firmware, macrocódigo, microcódigo, uno o más conjuntos o secuencias cualesquiera de instrucciones ejecutables mediante una formación de elementos lógicos, y cualquier combinación de dichos ejemplos. El programa o los segmentos de código pueden ser almacenados en un medio de almacenamiento legible por procesador o ser transmitidos por una señal de datos de ordenador, realizada en una onda portadora por un medio de transmisión o enlace de comunicación.

[0084] Las implementaciones de procedimientos, sistemas y técnicas divulgadas en el presente documento también pueden realizarse de forma tangible (por ejemplo, en uno o más medios legibles por ordenador, según lo enumerado en el presente documento) como uno o más conjuntos de instrucciones legibles y/o ejecutables por una máquina que incluya una formación de elementos lógicos (por ejemplo, un procesador, un microprocesador, un microcontrolador u otra máquina de estados finitos). El término "medio legible por ordenador" puede incluir cualquier medio que pueda almacenar o transferir información, incluyendo medios volátiles, no volátiles, extraíbles y no extraíbles. Los ejemplos de un medio legible por ordenador incluyen un circuito electrónico, un dispositivo de memoria semiconductora, una ROM, una memoria flash, una ROM borrable (EROM), un disquete u otro almacenamiento magnético, un CD-ROM/DVD u otro almacenamiento óptico, un disco duro, un medio de fibra óptica, un enlace de radiofrecuencia (RF) o cualquier otro medio que pueda usarse para almacenar la información deseada y a la que se pueda acceder. La señal de datos de ordenador puede incluir cualquier señal que pueda propagarse por un medio de transmisión tal como canales de red electrónica, fibras ópticas, aire, medios electromagnéticos, enlaces de RF, etc. Los segmentos de código pueden descargarse mediante redes informáticas tales como Internet o una intranet.

[0085] Cada una de las tareas de los procedimientos descritos en el presente documento puede incorporarse directamente en hardware, en un módulo de software ejecutado por un procesador o en una combinación de los dos. En una aplicación típica de una implementación de un procedimiento según lo divulgado en el presente documento, una formación de elementos lógicos (por ejemplo, compuertas lógicas) está configurada para realizar una de, más de una de, o incluso todas, las diversas tareas del procedimiento. Una o más de (posiblemente todas) las tareas también pueden implementarse como código (por ejemplo, uno o más conjuntos de instrucciones), incorporarse en un producto de programa informático (por ejemplo, uno o más medios de almacenamiento de datos, tales como discos, memoria flash u otras tarjetas de memoria no volátiles, chips de memoria semiconductores, etc.), que sea legible y/o ejecutable por una máquina (por ejemplo, un ordenador) que incluya una formación de elementos lógicos (por ejemplo, un procesador, un microprocesador, un micro-controlador u otra máquina de estados finitos). Las tareas de una implementación de un procedimiento según lo divulgado en el presente documento también pueden ser realizadas por más de una de formación o máquina de ese tipo. En estas u otras implementaciones, las tareas pueden ser realizadas dentro de un dispositivo para comunicaciones inalámbricas, tales como un teléfono celular u otro dispositivo que tenga dicha capacidad de comunicaciones. Un dispositivo de ese tipo puede configurarse para comunicarse con redes conmutadas por circuitos y/o conmutadas por paquetes (por ejemplo, usando uno o más protocolos tales como VoIP). Por ejemplo, un dispositivo de ese tipo puede incluir circuitos de RF configurados para recibir y/o transmitir tramas codificadas.

[0086] Se divulga expresamente que los diversos procedimientos divulgados en el presente documento pueden ser realizados por un dispositivo de comunicaciones portátil, tal como un equipo de mano, un auricular o un asistente digital portátil (PDA), y que los diversos aparatos descritos en este documento pueden incluirse dentro de un dispositivo de ese tipo. Una aplicación típica en tiempo real (por ejemplo, en línea) es una conversación telefónica realizada usando un dispositivo móvil de ese tipo.

[0087] En uno o más modos de realización ejemplares, las operaciones descritas en el presente documento pueden implementarse en hardware, software, firmware o cualquier combinación de los mismos. Si se implementan en software, dichas operaciones, como una o más instrucciones o código, pueden almacenarse en, o transmitirse por, un medio legible por ordenador. El término "medios legibles por ordenador" incluye tanto medios de almacenamiento legibles por ordenador como medios de comunicación (por ejemplo, de transmisión). A modo de ejemplo, y no de limitación, los medios de almacenamiento legibles por ordenador pueden comprender una formación de elementos de almacenamiento, tales como memoria semiconductora (que puede incluir, sin limitación, RAM dinámica o estática, ROM, EEPROM y/o RAM flash) o memoria ferroeléctrica, magnetorresistiva, ovónica, polimérica o de cambio de fase; CD-ROM u otro almacenamiento en disco óptico; y/o dispositivos de almacenamiento en disco magnético u otros dispositivos de almacenamiento magnético. Dichos medios de almacenamiento pueden almacenar información en forma de instrucciones o estructuras de datos a las que pueda accederse mediante un ordenador. Los medios de comunicación pueden comprender cualquier medio que pueda usarse para transportar el código de programa deseado en forma de instrucciones o de estructuras de datos y al que pueda accederse mediante un ordenador, incluyendo

cualquier medio que facilite la transferencia de un programa informático desde un lugar a otro. Además, cualquier conexión recibe adecuadamente la denominación de medio legible por ordenador. Por ejemplo, si el software se transmite desde una sede de la Red, un servidor u otro origen remoto, usando un cable coaxial, un cable de fibra óptica, un par trenzado, una línea de abonado digital (DSL) o tecnología inalámbrica tal como infrarrojos, radio y/o microondas, entonces el cable coaxial, el cable de fibra óptica, el par trenzado, la DSL o la tecnología inalámbrica tal como infrarrojos, radio y/o microondas, se incluyen en la definición de medio. Los discos, como se usan en el presente documento, incluyen el disco compacto (CD), el disco láser, el disco óptico, el disco versátil digital (DVD), el disco flexible y el disco Blu-ray™ (Blu-Ray Disc Association, Universal City, CA), donde algunos discos reproducen usualmente datos de forma magnética, mientras que otros reproducen los datos de forma óptica con láser. Las combinaciones de lo anterior también se deberían incluir dentro del alcance de los medios legibles por ordenador.

[0088] Un aparato de procesamiento de señales acústicas según lo descrito en el presente documento (por ejemplo, el aparato A100 o MF100) puede incorporarse en un dispositivo electrónico que acepte una entrada de voz con el fin de controlar ciertas operaciones, o puede beneficiarse de otro modo de la separación entre ruidos deseados y ruidos de fondo, tales como dispositivos de comunicaciones. Muchas aplicaciones pueden beneficiarse de la mejora o de la separación del sonido claro deseado con respecto a los sonidos de fondo procedentes de múltiples direcciones. Dichas aplicaciones pueden incluir interfaces hombre-máquina en dispositivos electrónicos o informáticos que incorporen capacidades tales como el reconocimiento y la detección de voces, la mejora y la separación del habla, el control activado por voz y similares. Puede ser deseable implementar dicho aparato de procesamiento de señales acústicas para que sea adecuado en dispositivos que solo proporcionen capacidades de procesamiento limitadas.

[0089] Los elementos de las diversas implementaciones de los módulos, elementos y dispositivos descritos en el presente documento pueden fabricarse como dispositivos electrónicos y/u ópticos que residen, por ejemplo, en el mismo chip o entre dos o más chips en un conjunto de chips. Un ejemplo de dicho dispositivo es una formación fija o programable de elementos lógicos, tales como transistores o compuertas. Uno o más elementos de las diversas implementaciones del aparato descrito en el presente documento pueden implementarse también, total o parcialmente, como uno o más conjuntos de instrucciones dispuestas para ejecutarse en una o más formaciones fijas o programables de elementos lógicos tales como microprocesadores, procesadores incorporados, núcleos del IP, procesadores de señales digitales, FPGA, ASSP y ASIC.

[0090] Es posible que uno o más elementos de una implementación de un aparato, según lo descrito en el presente documento, se usen para realizar tareas o ejecutar otros conjuntos de instrucciones que no estén directamente relacionados con una operación del aparato, tal como una tarea relacionada con otra operación de un dispositivo o sistema en el que el aparato esté incorporado. También es posible que uno o más elementos de una implementación de un aparato de ese tipo tengan una estructura en común (por ejemplo, un procesador usado para ejecutar partes de código correspondientes a diferentes elementos en momentos diferentes, un conjunto de instrucciones ejecutadas para realizar las tareas correspondientes a diferentes elementos en momentos diferentes o una disposición de dispositivos electrónicos y/u ópticos que realicen operaciones para diferentes elementos en momentos diferentes).

REIVINDICACIONES

1. Un procedimiento (M100) de procesamiento de una pluralidad de coeficientes de función base, que describe un campo sonoro durante un intervalo de tiempo, comprendiendo dicho procedimiento:
5 realizar (T100) una transformación reversible en un primer grupo de la pluralidad de coeficientes de función base, para producir una pluralidad de señales de canal, en donde cada una entre la pluralidad de señales de canal está asociada a una correspondiente región diferente del espacio; **caracterizado por**
10 producir (T200) datos configurados para almacenamiento y / o transmisión, comprendiendo los datos una estructura de datos que tiene una primera parte que codifica un segundo grupo de la pluralidad de coeficientes de función base, en donde el segundo grupo es diferente al primer grupo, y una segunda parte, diferente a la primera parte, que codifica la pluralidad de señales de canal.
2. El procedimiento según la reivindicación 1, en el que dicha pluralidad de coeficientes de función base es una pluralidad de coeficientes de funciones base armónicas esféricas.
3. El procedimiento según la reivindicación 1, en el que dicha estructura de datos incluye un primer flujo que incluye dicha representación del segundo grupo y un segundo flujo que incluye la representación de la pluralidad de señales de canal.
20
4. Un procedimiento (M200) de obtención de una pluralidad de coeficientes de función base que describe un campo de sonido durante un intervalo de tiempo a partir de datos que se han almacenado y / o transmitido en una estructura de datos que comprende una primera parte que codifica un segundo grupo de la pluralidad de coeficientes de función base y una segunda parte, distinta a la primera parte, que codifica una pluralidad de señales de canal, en donde un subconjunto de la pluralidad de señales de canal se basa en un primer grupo de la pluralidad de coeficientes de función base, estando cada señal del subconjunto de la pluralidad de señales de canal asociada a una correspondiente región diferente del espacio, comprendiendo el procedimiento:
25 realizar (T400) una transformación sobre datos de la segunda parte de la estructura de datos que comprende el subconjunto de la pluralidad de señales de canal, para producir el primer grupo entre la pluralidad de coeficientes de función base, en donde el primer grupo es diferente al segundo grupo; y obtener la pluralidad de coeficientes de función base que describen el campo de sonido durante el intervalo de tiempo, desde el primer grupo y el segundo grupo.
30
5. El procedimiento de acuerdo a la reivindicación 4, en el que cada uno entre dicha pluralidad de coeficientes de función base corresponde a una única función entre un conjunto de funciones base armónicas esféricas.
35
6. El procedimiento de acuerdo a la reivindicación 4, en el que dicho procedimiento comprende, basándose en dicha pluralidad de coeficientes de función base, producir una segunda pluralidad de señales de canal, en donde cada señal del subconjunto entre la pluralidad de señales de canal está asociado a una correspondiente dirección diferente entre un conjunto de direcciones coplanares, y en donde cada una entre la segunda pluralidad de señales de canal está asociada a una correspondiente dirección diferente entre un conjunto de direcciones que abarcan un espacio tridimensional.
40
7. Un aparato (MF100) para procesar una pluralidad de coeficientes de función base que describen un campo de sonido durante un intervalo de tiempo, comprendiendo dicho aparato:
45 medios (F100) para realizar una transformación reversible en un primer grupo entre la pluralidad de coeficientes de función base, para producir una pluralidad de señales de canal, en donde cada una entre la pluralidad de señales de canal está asociada a una correspondiente región diferente del espacio;
50 **caracterizado por** medios (F200) para producir datos configurados para el almacenamiento y / o la transmisión, comprendiendo los datos una estructura de datos que tiene una primera parte que codifica un segundo grupo entre la pluralidad de coeficientes de función base, en donde el segundo grupo es diferente al primer grupo, y una segunda parte, distinta a la primera parte, que codifica la pluralidad de señales de canal que es distinta a dicha representación del segundo grupo.
55
8. El aparato de la reivindicación 7, que comprende además un transcodificador (TC10), acoplado a la salida de los medios (F100) para realizar una transformación reversible en un primer grupo entre la pluralidad de coeficientes de función base, estando el transcodificador configurado para producir, a partir de la pluralidad de señales de canal, un flujo de bits compatible con la norma ATSC A/52:2012, o la codificación de audio avanzada de alta eficacia ISO/IEC 14496-3.
60
9. El aparato de la reivindicación 7, en el que el segundo grupo entre la pluralidad de funciones básicas son coeficientes armónicos esféricos obtenibles de señales que son señales de audio adquiridas físicamente,
65

capturadas por una formación de micrófonos.

- 5 **10.** El aparato de la reivindicación 7, que comprende además un codificador (SE20) para codificar el segundo grupo de coeficientes armónicos esféricos, para producir la primera parte de la estructura de datos.
- 10 **11.** Un aparato (MF200) para obtener una pluralidad de coeficientes de función base que describe un campo sonoro durante un intervalo de tiempo que comprende:

 medios para recibir una estructura de datos, comprendiendo la estructura de datos una primera parte que codifica un segundo grupo entre la pluralidad de coeficientes de función base, y una segunda parte, distinta a la primera parte, que codifica una pluralidad de señales de canal, en donde un subconjunto entre la pluralidad de las señales de canal se basa en un primer grupo entre la pluralidad de coeficientes de función base, estando cada uno de los subconjuntos entre la pluralidad de señales de canal asociado a una correspondiente región diferente del espacio;

15 medios (F300) para obtener, a partir de la estructura de datos, el segundo grupo entre la pluralidad de coeficientes de función base y la pluralidad de señales de canal;

20 medios (F400) para realizar una transformación en la pluralidad de señales de canal para producir el primer grupo entre la pluralidad de coeficientes de función base, en donde el primer grupo es diferente al segundo grupo; y

25 medios para obtener la pluralidad de coeficientes de función base que describen el campo de sonido durante el intervalo de tiempo desde el primer grupo y el segundo grupo entre la pluralidad de coeficientes de función base.
- 30 **12.** El aparato de la reivindicación 11, que comprende además un decodificador basado en escenas (SD20), configurado para decodificar la primera parte de la estructura de datos.
- 35 **13.** El aparato de la reivindicación 12, que comprende además un representador (500; SR10) para producir una segunda pluralidad de señales de canal basadas en la pluralidad de coeficientes de función base.
- 40 **14.** El aparato de la reivindicación 13, en el que los medios para realizar la transformación en el subconjunto de la pluralidad de señales de canal están configurados además para:

 determinar si la primera parte está presente en la estructura de datos; y

 realizar la transformación en el subconjunto de la pluralidad de señales de canal, solo cuando la primera parte está presente en la estructura de datos; y

 en el que el aparato comprende además medios para representar la pluralidad de señales de canal como alimentaciones de altavoz cuando la primera parte no está presente en la estructura de datos.
- 45 **15.** El aparato de la reivindicación 11, en el que la pluralidad de señales de canal son compatibles con un formato de audio de múltiples canales según la norma ATSC A/52:2012 o la codificación de audio avanzada de alta eficacia ISO/IEC 14496-3.
- 50 **16.** El aparato de la reivindicación 11, en el que cada uno de los subconjuntos de la pluralidad de señales de canal está asociado a una correspondiente región diferente del espacio, asociada a diferentes ubicaciones de altavoces.
- 55 **17.** El aparato de la reivindicación 16, en el que las diferentes ubicaciones de los altavoces son parte de un formato 9.1, 11.1 o 22.2.
- 60 **18.** El aparato de la reivindicación 11, en el que cada uno de los subconjuntos entre la pluralidad de señales de canal está asociado a una correspondiente región diferente del espacio, asociada a direcciones coplanares de un formato 5.1 o 7.1.
- 19.** Un medio de almacenamiento de datos no transitorio legible por ordenador que tiene instrucciones almacenadas en el mismo que, al ejecutarse, hacen que una máquina realice un procedimiento de acuerdo a cualquiera de las reivindicaciones 1 a 6.

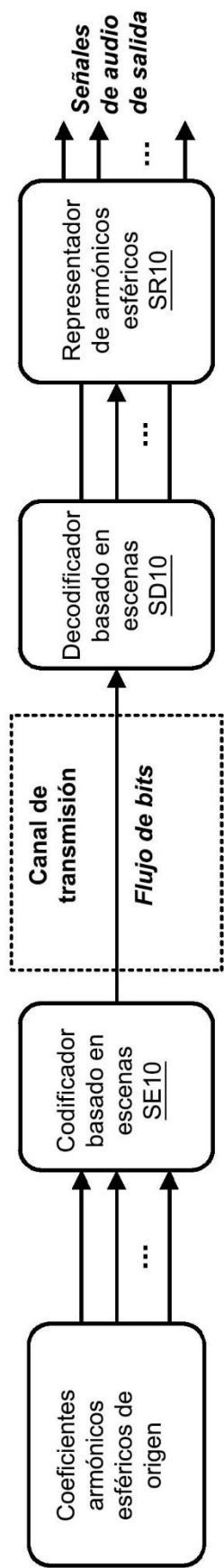


FIG. 1A

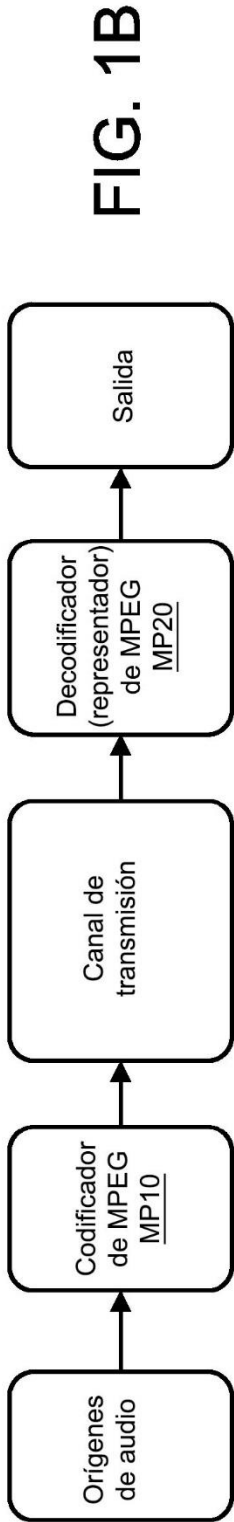


FIG. 1B

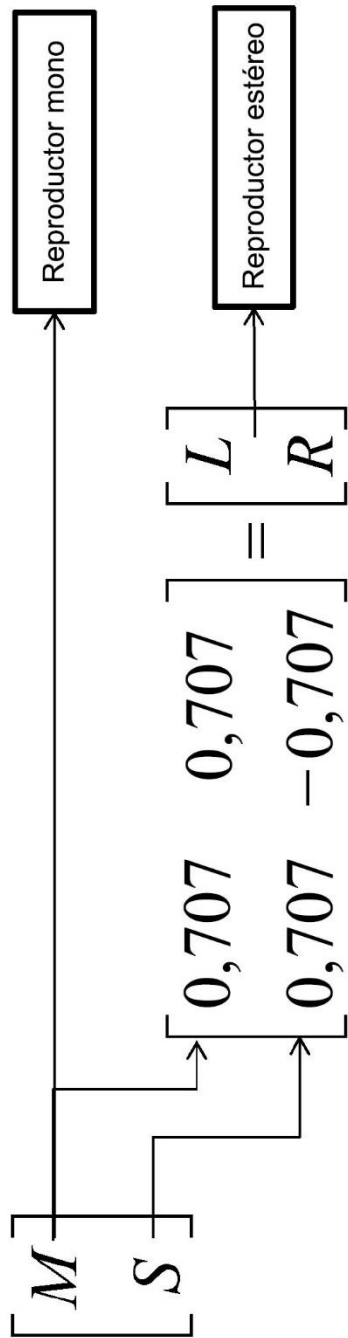


FIG. 1C

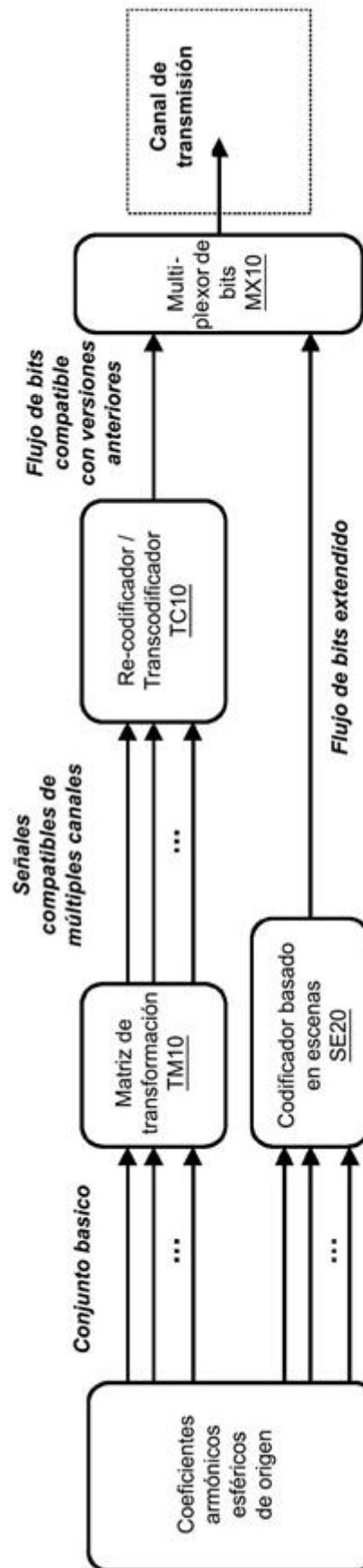


FIG. 2A

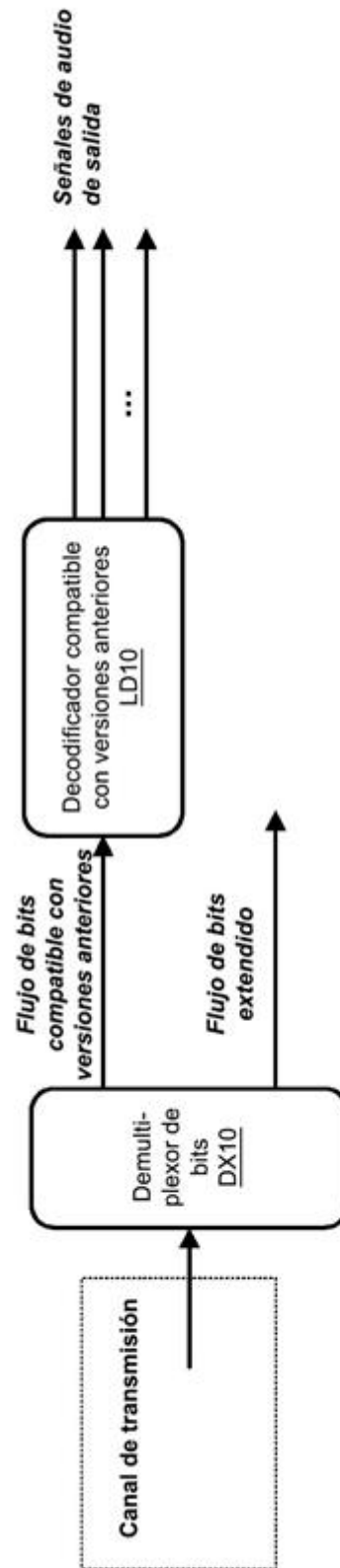
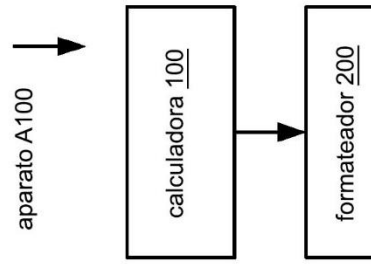
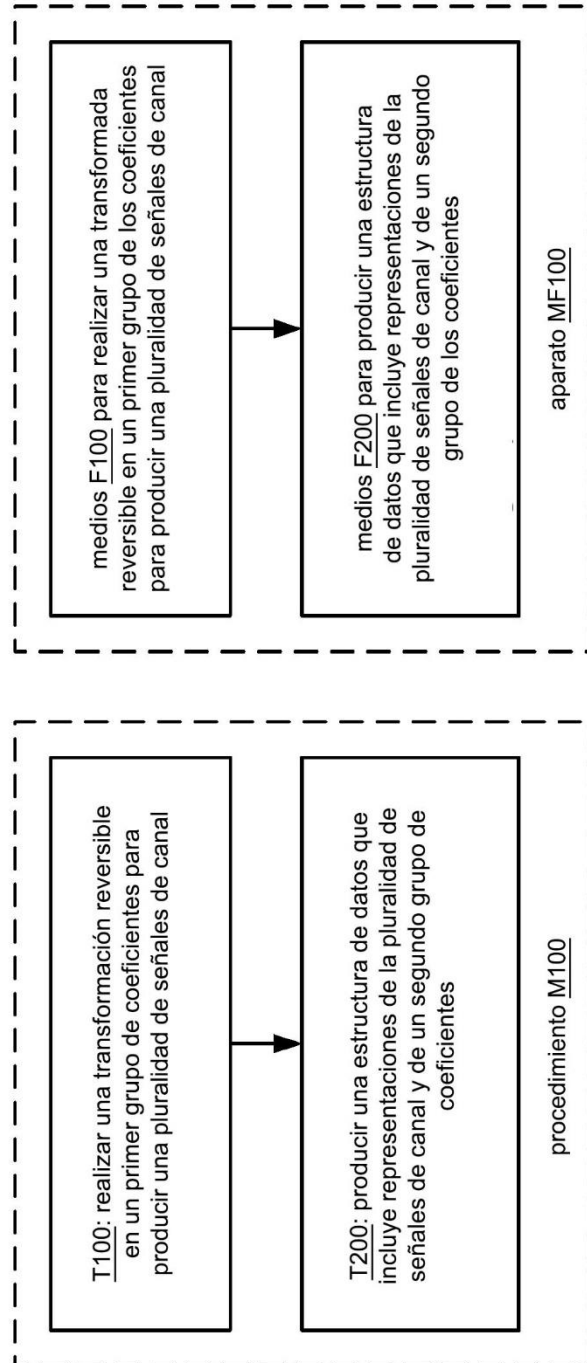
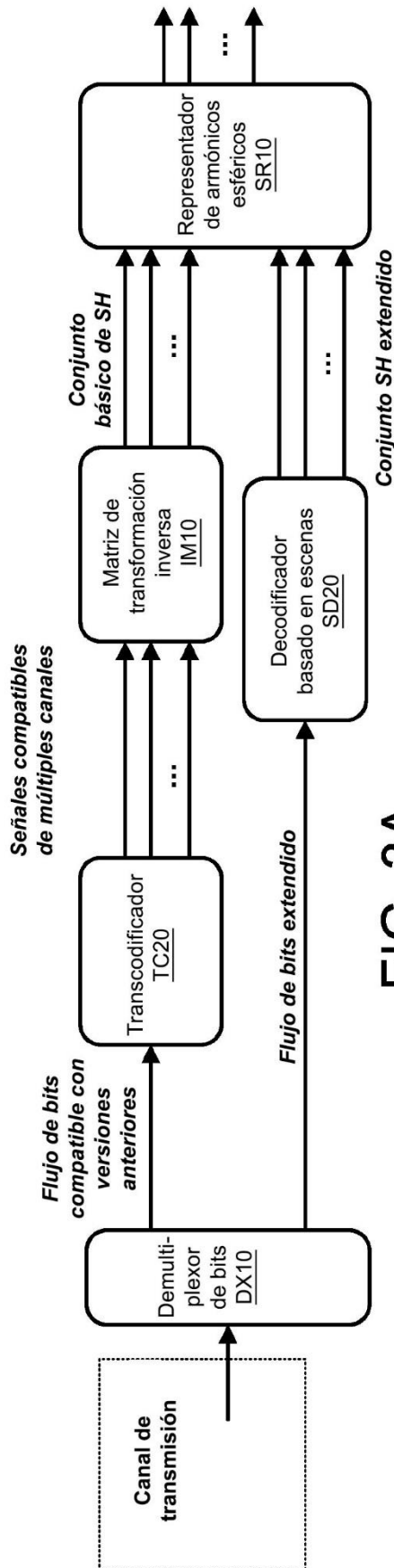


FIG. 2B



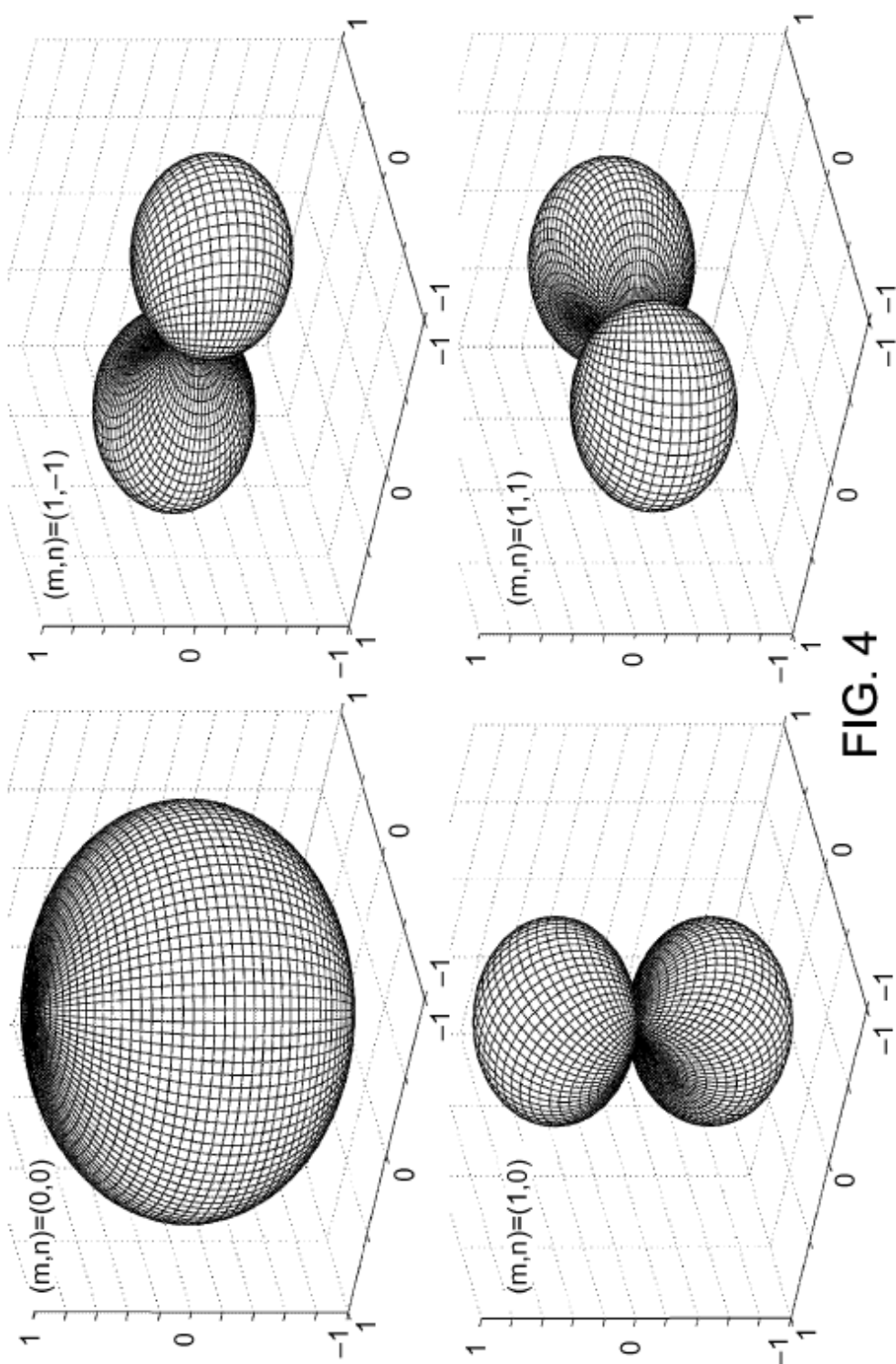


FIG. 4

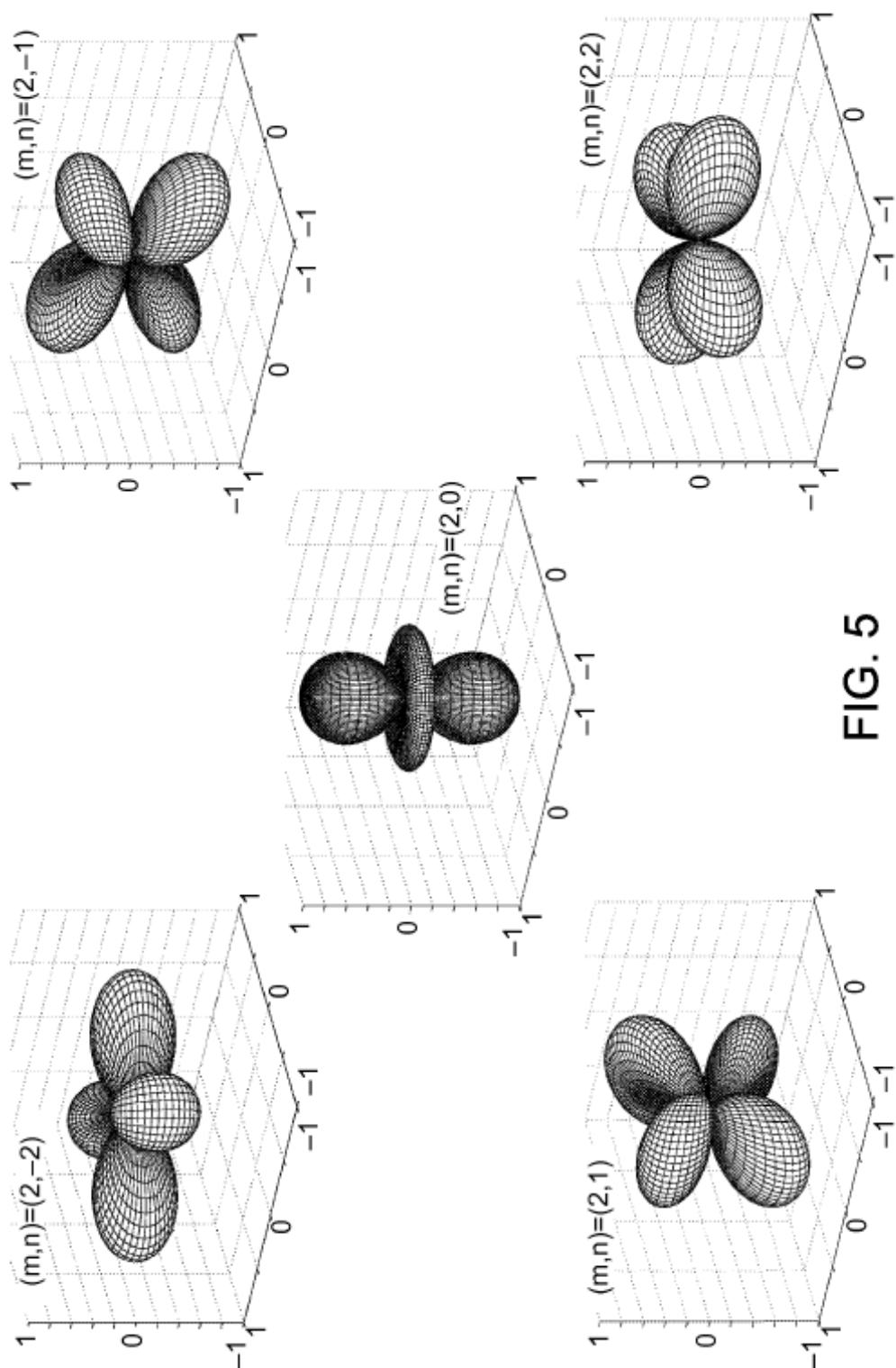


FIG. 5

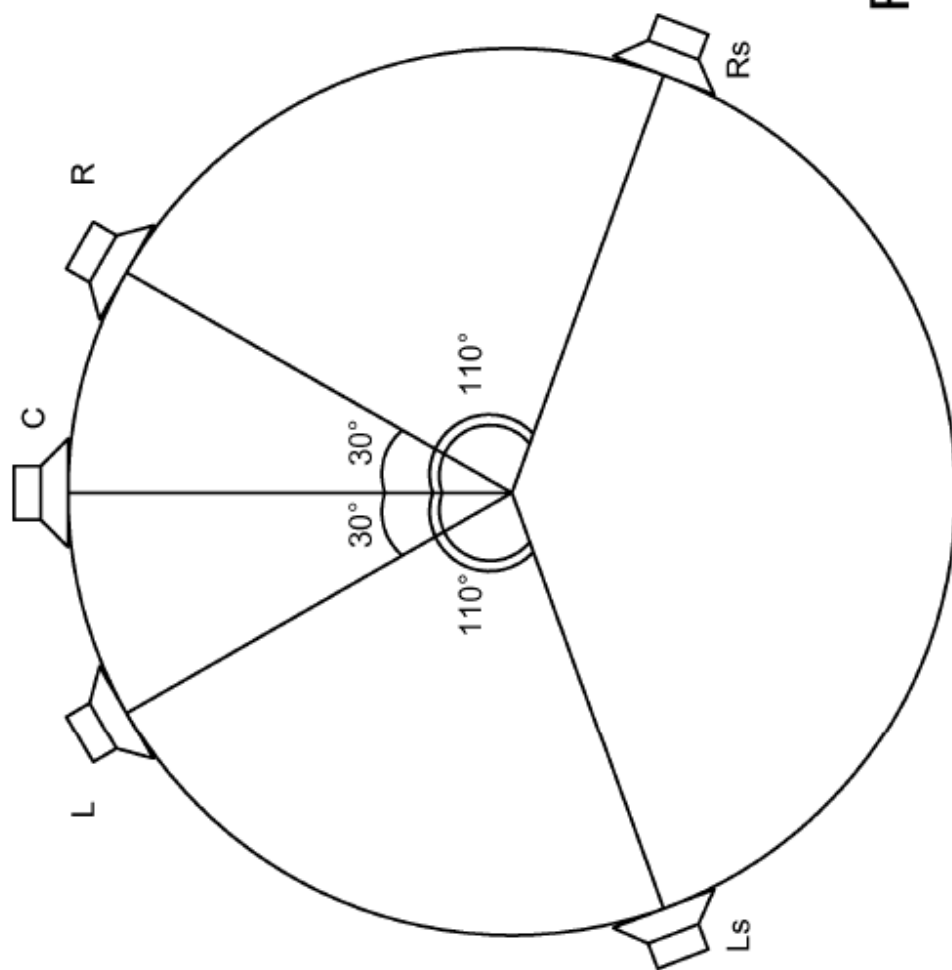


FIG. 6

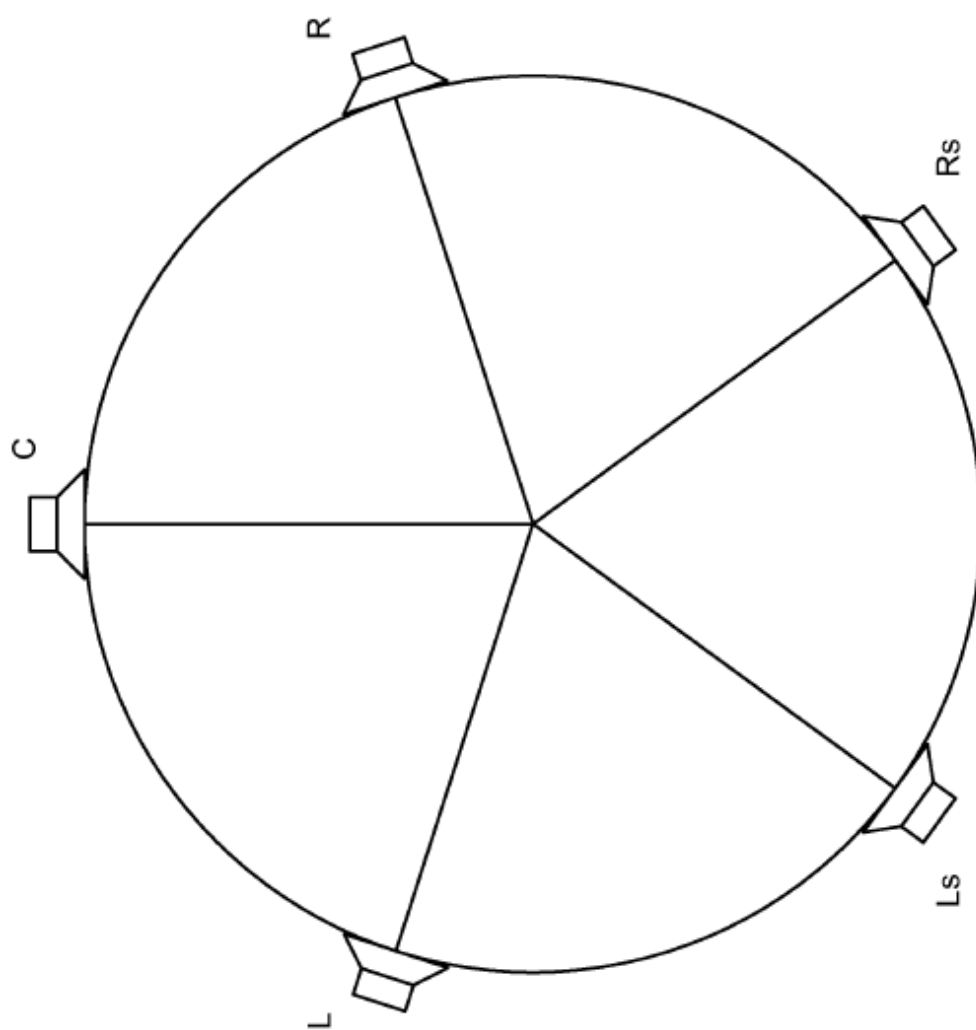


FIG. 7

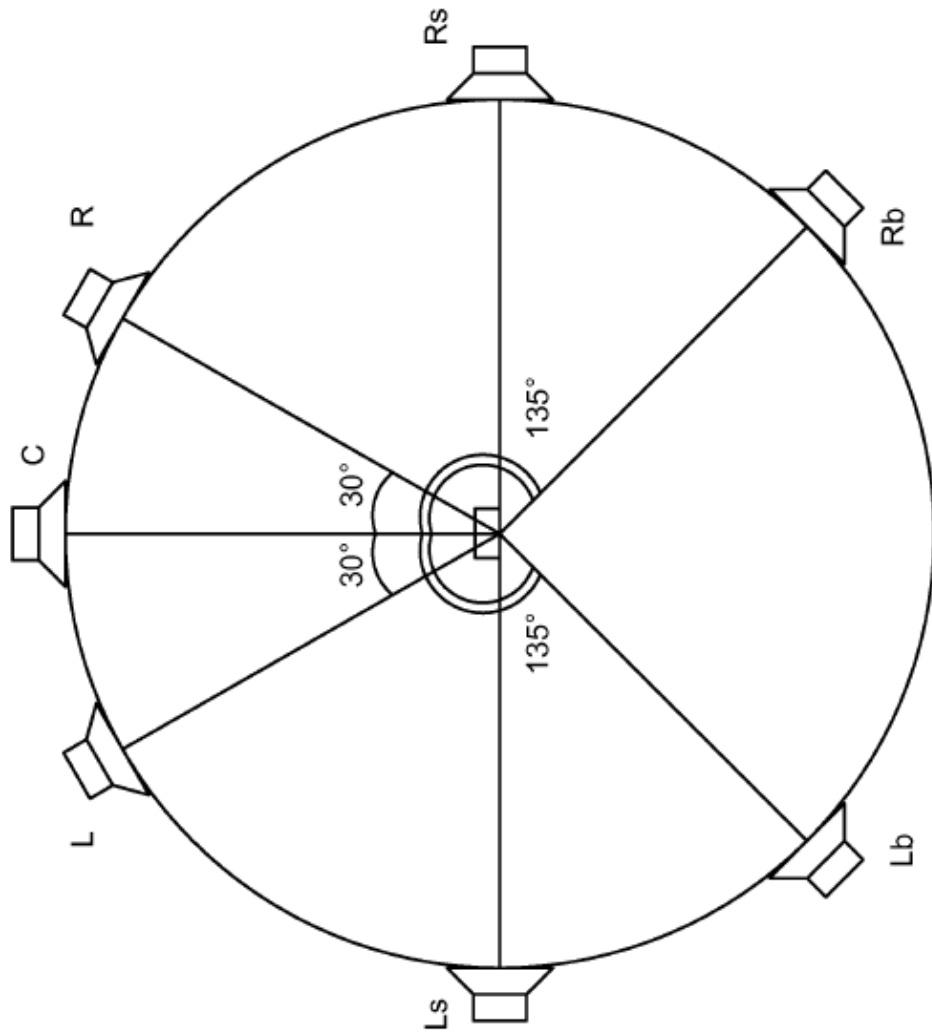


FIG. 8

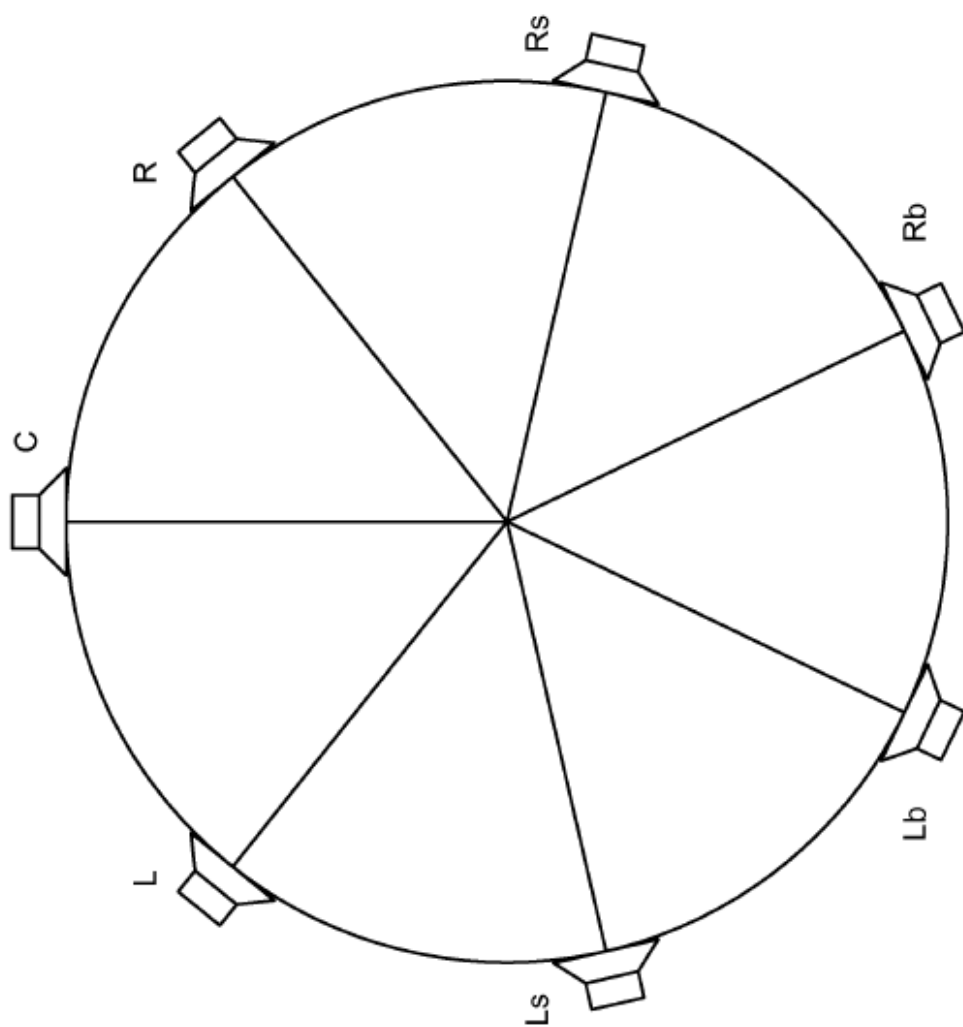


FIG. 9

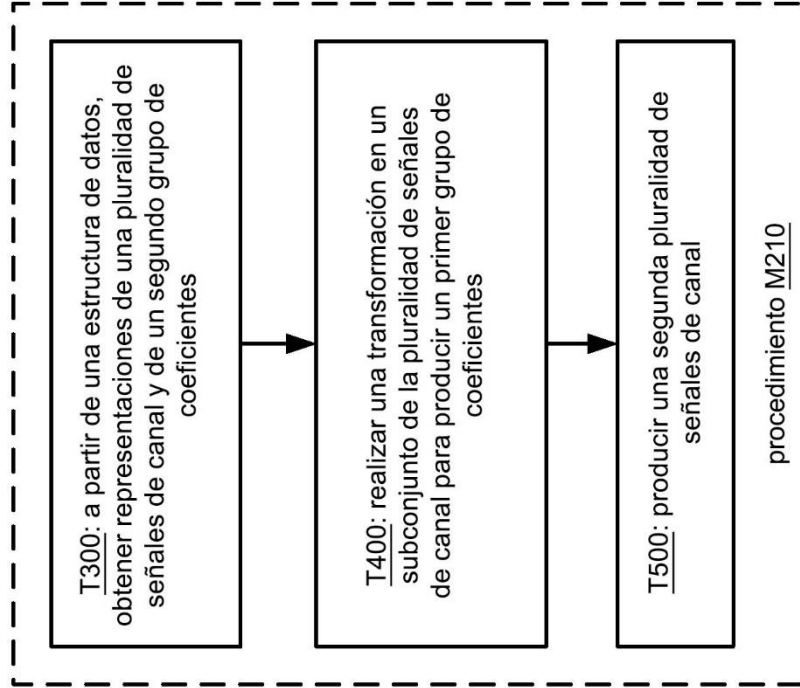


FIG. 10B

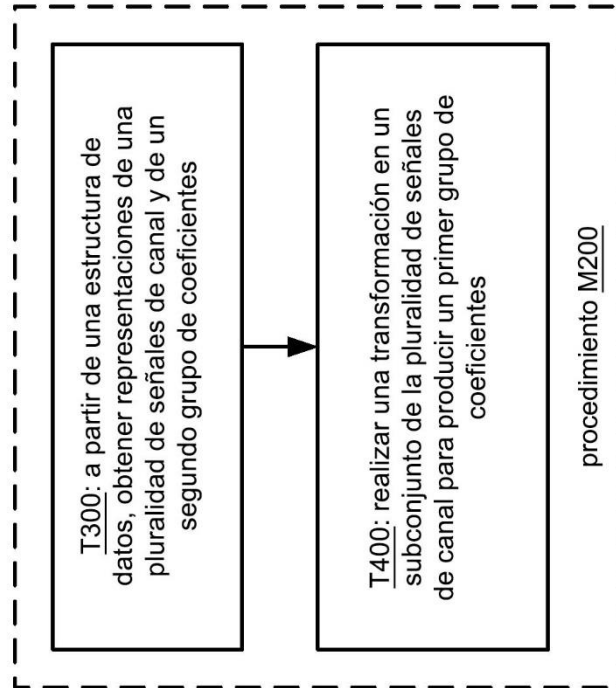


FIG. 10A

