

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 396 072**

51 Int. Cl.:

G10L 19/00 (2006.01)

H04N 7/15 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **24.04.2007 E 09010510 (7)**

97 Fecha y número de publicación de la concesión europea: **07.11.2012 EP 2112652**

54 Título: **Aparato para combinar múltiples fuentes de audio paramétricamente codificadas**

30 Prioridad:

07.07.2006 US 819419 P
24.04.2007 US 739544

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
19.02.2013

73 Titular/es:

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

72 Inventor/es:

HILPERT, JOHANNES;
HERRE, JUERGEN;
LINZMEIER, KARSTEN;
HELLMUTH, OLIVER y
KASTNER, THORSTEN

74 Agente/Representante:

PONTI SALES, Adelaida

ES 2 396 072 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

DESCRIPCIÓN

Aparato para combinar múltiples fuentes de audio paramétricamente codificadas

5 **Campo de la Invención**

[0001] La presente invención se refiere a codificación de audio multicanales y, en particular, a un concepto para combinar corrientes de audio paramétricamente codificadas de una manera flexible y eficiente.

10 **ANTECEDENTES DE LA INVENCION Y TÉCNICA ANTERIOR**

15 [0002] El reciente desarrollo en el área de la codificación de audio ha dado lugar a varias técnicas de codificación paramétrica de audio para codificar conjuntamente una señal de audio multicanales (e.g., 5.1 canales) en uno (o más) canales de sub-mezcla más una corriente de información lateral. En general, la corriente de información lateral tiene parámetros relacionados con las propiedades de los canales originales de la señal multicanal, ya sea con respecto a otros canales originales de la señal multicanal, o con respecto al canal de sub-mezcla. La definición particular de los parámetros del canal de referencia a los cuales se refieren estos parámetros, depende de la implementación específica. Algunas de las técnicas conocidas en la técnica son "codificación de indicación binaural", "codificación de audio espacial" y "estéreo paramétrico".

20 [0003] Para detalles de estas implementaciones particulares, se hace referencia en la presente a las publicaciones relacionadas. Por ejemplo, la codificación de indicación binaural se detalla en:

25 C. Faller y F. Baumgarte, "Efficient representation of spatial audio using perceptual parametrization" (Representación eficiente del audio espacial utilizando parametrización perceptual), IRRR WASPAA, Mohonk, NY, Octubre 2001; F. Baumgarte y C Faller, "Estimation of auditory spatial cues for binaural cue coding" (Estimación de las señales de entrada espaciales auditivas para la codificación de indicación binaural), ICASSP, Orlando, FL, Mayo 2002; C. Faller y F. Baumgarte, "Binaural cue coding: a novel and efficient representation of spatial audio" (Codificación de indicación binaural: una nueva y eficiente representación del audio espacial), ICASSP, Orlando, FL, Mayo 2002; C. Faller y F. Baumgarte, "Binaural cue coding applied to audio compression with flexible rendering" (Codificación de indicación binaural aplicada a la compresión de audio con representación flexible), AES 113th Convention, Los Angeles, Preimpresión 5686, Octubre 2002; C. Faller y F. Baumgarte, "Binaural cue coding – Part II: Schemes and applications", (Codificación de indicación binaural – Parte II: Esquemas y aplicaciones) IEEE Trans. On Speech and Audio Proc., vol., 11, no. 6, Noviembre 2003, y J. Herre, C. Faller et al., "Spatial Audio Coding: Next-generation efficient and compatible coding of multi-channel audio" (Codificación de audio espacial: codificación de siguiente generación, eficiente y compatible de audio multicanales), Audio Engineering Society Convention Paper, Octubre 28, 2004, San Francisco, CA. EUA.

40 [0004] Aunque la codificación de indicación binaural utiliza múltiples canales originales, el estéreo paramétrico es una técnica relacionada para la codificación paramétrica de una señal de estéreo de dos canales dando como resultado una mono señal transmitida e información lateral de parámetro, por ejemplo, como se revisa en las siguientes publicaciones: J. Breebaart, S. van de Par, A. Kohlrausch, E. Schuijers, "High-Quality Parametric Spatial Audio Coding at Low Bit rates" (Codificación de audio espacial paramétrica de alta calidad a bajas velocidades de transmisión), AES 116th Convention, Berlín, Preimpresión 6072, Mayo 2004; E. Schuijers, J. Breebaart, H. Purnhagen, J. Engdegard, "Low Complexity Parametric Stereo Coding" (Codificación de estéreo paramétrica de baja complejidad), AES 116th Convention, Berlín, Preimpresión 6073, Mayo 2004.

50 [0005] Otras tecnologías se basan en la multiplexión de números arbitrarios de fuentes u objetos de audio en un solo canal de transmisión de audio. Los esquemas a base de multiplexión, por ejemplo, se introducen como "representación flexible" en publicaciones relacionadas con BCC (codificación de indicación binaural), o, más recientemente, mediante un esquema llamado "codificación de fuente de unión" (JSC). Las publicaciones relacionadas son, por ejemplo: C. Faller, "Parametric Joint Coding of Audio Sources" (Codificación de unión paramétrica de fuentes de audio), Convention Paper 6752, 120th AES Convention, París, Mayo 2006. Similar a los esquemas de codificación de indicación paramétrica de estéreo y binaural, estas técnicas se destinan a codificar múltiples objetos de audio originales (canales) para la transmisión mediante menos canales de sub-mezcla. Mediante la derivación adicional de los parámetros en base a objetos para cada canal de entrada, que pueden encontrarse codificados a una muy baja proporción de datos y que también se transmiten a un receptor, estos objetos pueden separarse en el lado receptor y representarse (mezclarse) para cierto número de dispositivos de salida, como por ejemplo, audífonos, altavoces estéreo de dos canales o instalaciones de altavoz multicanales. Este procedimiento permite el ajuste y la redistribución (panorámica) del nivel de los diferentes objetos de audio hacia diferentes ubicaciones en la instalación de reproducción, i.e., en el lado receptor.

60 [0006] Básicamente, tales técnicas operan como transmisor M-k-N, siendo M el número de objetos de audio en la entrada, siendo k el número de canales de sub-mezcla transmitidos, típicamente $k \leq 2$. N es el número de canales de audio en la salida de representación, i.e., por ejemplo, el número de altavoces. Es decir, $N = 2$ para una

representación estéreo o $N = 6$ para una instalación multicanales 5.1. En términos de eficiencia de compresión, los valores típicos son, e.g., 64 kbps o menos para un canal de sub-mezcla perceptualmente codificado (que consiste de canales de audio k) y de aproximadamente 3 kbps para los parámetros de objeto por objeto de audio transmitido.

5 **[0007]** Los escenarios de aplicación para las técnicas anteriores son, por ejemplo, codificar las escenas de audio espacial relacionadas con cine-película-producciones para permitir la reproducción espacial del sonido en un sistema de teatro casero. Los ejemplos comunes son las pistas de sonido envolvente 5.1 y 7.1 ampliamente conocidas en un medio de películas tal como DVD y lo similar. Las producciones de películas se han vuelto más y más complejas con respecto a las escenas de audio, destinadas a proporcionar una experiencia de percepción auditiva espacial y por tanto tienen que mezclarse con gran cuidado. Diferentes ingenieros de sonido pueden comisionarse con la mezcla de las diferentes fuentes envolventes o efectos de sonido y, en consecuencia, es deseable la transmisión de escenarios multicanales paramétricamente codificados entre los ingenieros de sonido individuales para transportar las corrientes de audio de los ingenieros de sonido individuales de manera eficiente.

10 **[0008]** Otro escenario de aplicación para tal tecnología es la tele-conferencia con múltiples hablantes en cualquier extremo de una conexión de punto-a-punto tal como se describe en US 2005/ 0062843. Para ahorrar el ancho de banda, la mayoría de las instalaciones de tele-conferencia operan con transmisión monofónica. Por ejemplo, puede lograrse utilizando codificación de fuente unida o una de las otras técnicas de codificación multicanales para la transmisión, redistribución y alineación de nivel de los diferentes hablantes en el extremo receptor (cada extremo) y por tanto, la inteligibilidad y el balance de los hablantes se mejora gastando una velocidad de transmisión marginalmente incrementada en comparación con un sistema monofónico. La ventaja de la inteligibilidad incrementada se hace particularmente evidente en el caso especial de la asignación a cada individuo participante de la conferencia a un canal único (y, por tanto al orador) de una instalación de servicio multicanales en un extremo receptor. Sin embargo, este es un caso especial. En general, el número de participantes no igualará el número de oradores en el extremo receptor. Sin embargo, utilizando la instalación de servicio existente es posible representar la señal asociada con cada participante de manera que parezca originarse desde cualquier posición deseada. Es decir, el participante individual no solo se reconoce por su voz diferente, sino también por la ubicación de la fuente de audio relacionada con el participante hablante.

15 **[0009]** Aunque el estado de las técnicas de la técnica implementa conceptos acerca de cómo codificar eficientemente canales u objetos de audio múltiples, todas las técnicas actualmente conocidas carecen de la posibilidad de combinar dos o más de estas corrientes de audio transmitidas de manera eficiente para derivar una corriente de salida (señal de salida), que es una representación de todas las corrientes de audio de entrada (señales de audio de entrada).

20 **[0010]** El problema surge, por ejemplo, cuando se considera un escenario de tele-conferencia con más de dos ubicaciones, teniendo cada ubicación uno o más oradores. Entonces, se requiere una instancia intermedia para recibir las señales de entrada de audio de las fuentes individuales y para generar una señal de salida de audio para cada ubicación de tele-conferencia que tenga solamente la información de las ubicaciones de tele-conferencia restantes. Es decir, la instancia intermedia tiene que generar una señal de salida, que se deriva de una combinación de dos o más señales de entrada de audio y que permite una reproducción de los canales de audio u objetos de audio individuales de las dos o más señales de entrada.

25 **[0011]** Puede presentarse un escenario similar cuando dos ingenieros de audio en una producción de cine-película desean combinar sus señales de audio espacial para verificar la impresión de percepción auditiva generada por ambas señales. Entonces, puede ser deseable combinar directamente dos señales multicanales codificadas para verificar la impresión de percepción auditiva combinada. Es decir, una señal combinada necesita ser tal que se asemeje a todos los objetos (fuentes) de audio de los dos ingenieros de audio.

30 **[0012]** Sin embargo, de acuerdo con las técnicas de la técnica anterior, tal combinación es solo posible decodificando las señales de audio (corrientes). Entonces, las señales de audio decodificadas pueden re-codificarse de nuevo mediante los codificadores multicanales de la técnica anterior para generar una señal combinada en la cual todos los canales de audio u objetos de audio originales se representan apropiadamente.

35 **[0013]** Esto tiene la desventaja de una alta complejidad computacional, desperdiciando así mucha energía y algunas veces haciendo imposible aplicar el concepto, especialmente en escenarios en tiempo real. Además, una combinación mediante subsecuente decodificación y re-codificación del audio puede ocasionar un retraso considerable debido a las dos etapas de procesamiento, lo cual es inaceptable para ciertas aplicaciones, tales como tele-conferencia/telecomunicación.

Resumen de la Invención

40 **[0014]** Un objetivo de la presente invención es proporcionar un concepto para combinar eficientemente múltiples señales de audio paramétricamente codificadas.

45 **[0015]** De acuerdo con un primer aspecto de la presente invención, este objetivo se logra mediante un generador de señal de audio de acuerdo con la reivindicación 1 para generar una señal de salida de audio, comprendiendo el generador de señal de audio: un receptor de señal de audio para recibir una primera señal de audio que comprende un primer canal de sub-mezcla que tiene información sobre dos o más primeros canales originales y que comprende un

parámetro original asociado con uno de los primeros canales originales que describe una propiedad de uno de los primeros canales originales con respecto a un canal de referencia; y una segunda señal de audio que comprende un segundo canal de sub-mezcla que tiene información sobre al menos un segundo canal original; un combinador de canales para derivar un canal de sub-mezcla combinado al combinar el primer canal de sub-mezcla y el segundo canal de sub-mezcla; un calculador de parámetro para derivar un primer parámetro combinado que describe la propiedad de uno de los primeros canales originales con respecto a un canal de referencia común, y un segundo parámetro combinado que describe la propiedad de otro de los primeros canales originales o del al menos un segundo canal original con respecto al canal de referencia común; y una interfaz de salida para producir la señal de salida de audio que comprende el canal de sub-mezcla combinado y los parámetros combinados primero y segundo.

[0016] De acuerdo con un segundo aspecto de la presente invención, este objetivo se logra mediante un procedimiento para generar una señal de salida de audio de acuerdo con la reivindicación 10, comprendiendo el procedimiento: recibir una primera señal de audio que comprende un primer canal de sub-mezcla que tiene información sobre dos o más primeros canales originales y que comprende un parámetro original asociado con uno de los primeros canales originales que describe una propiedad de uno de los primeros canales originales con respecto a un canal de referencia; y una segunda señal de audio que comprende un segundo canal de sub-mezcla que tiene información sobre al menos un segundo canal original; derivar un canal de sub-mezcla combinado al combinar el primer canal de sub-mezcla y el segundo canal de sub-mezcla; derivar un primer parámetro combinado que describe la propiedad de uno de los primeros canales originales con respecto a un canal de referencia común, y un segundo parámetro combinado que describe la propiedad de otro de los primeros canales originales o del al menos un segundo canal original con respecto al canal de referencia común; y producir la señal de salida de audio que comprende el canal de sub-mezcla combinado y los parámetros combinados primero y segundo.

[0017] De acuerdo con un tercer aspecto de la presente invención, este objetivo se logra mediante una corriente de datos de una representación de tres o más canales de audio tal como se define en la reivindicación 11, que comprende: un canal de sub-mezcla combinado que es una combinación de un primer canal de sub-mezcla que tiene información sobre al menos dos primeros canales originales y un segundo canal de sub-mezcla que tiene información sobre al menos un segundo canal original; un primer parámetro que describe una propiedad de uno de los al menos dos primeros canales originales con respecto a un canal de referencia; y un segundo parámetro que describe la propiedad de otro canal de los primeros canales originales o la propiedad del al menos un segundo canal original con respecto al canal de referencia.

[0018] De acuerdo con un cuarto aspecto de la presente invención, este objetivo se logra mediante un programa de computadora que implementa un procedimiento para generar una señal de salida de audio tal como se define en la reivindicación 13, comprendiendo el procedimiento: recibir una primera señal de audio que comprende un primer canal de sub-mezcla que tiene información sobre dos o más primeros canales originales y que comprende un parámetro original asociado con uno de los primeros canales originales que describe una propiedad de uno de los primeros canales originales con respecto a un canal de referencia; y una segunda señal de audio que comprende un segundo canal de sub-mezcla que tiene información sobre al menos un segundo canal original; derivar un canal de sub-mezcla combinado al combinar el primer canal de sub-mezcla y el segundo canal de sub-mezcla; derivar un primer parámetro combinado que describe la propiedad de uno de los primeros canales originales con respecto a un canal de referencia común, y un segundo parámetro combinado que describe la propiedad de otro de los primeros canales originales o del al menos un segundo canal original con respecto al canal de referencia común; y producir la señal de salida de audio que comprende el canal de sub-mezcla combinado y los parámetros combinados primero y segundo.

[0019] De acuerdo con un quinto aspecto de la presente invención, este objetivo se logra mediante un sistema de conferencia que tiene un generador de señal de audio para generar una señal de salida de audio, que comprende: un receptor de señal de audio recibir una primera señal de audio que comprende un primer canal de sub-mezcla que tiene información sobre dos o más primeros canales originales y que comprende un parámetro original asociado con uno de los primeros canales originales que describe una propiedad de uno de los primeros canales originales con respecto a un canal de referencia; y una segunda señal de audio que comprende un segundo canal de sub-mezcla que tiene información sobre al menos un segundo canal original; un combinador de canal para derivar un canal de sub-mezcla combinado al combinar el primer canal de sub-mezcla y el segundo canal de sub-mezcla; un calculador de parámetro para derivar un primer parámetro combinado que describe la propiedad de uno de los primeros canales originales con respecto a un canal de referencia común, y un segundo parámetro combinado que describe la propiedad de otro de los primeros canales originales o del al menos un segundo canal original con respecto al canal de referencia común; y una interfaz de salida para producir la señal de salida de audio que comprende el canal de sub-mezcla combinado y los parámetros combinados primero y segundo.

[0020] Según algunas realizaciones de la presente invención, el calculador de parámetros es operativo para usar un canal predeterminado de los primeros canales originales o el al menos un segundo canal original como canal de referencia común.

[0021] Según algunas realizaciones de la presente invención, el canal de referencia común se escoge para ser el canal de referencia de la primera señal de audio.

[0022] Según realizaciones adicionales de la presente invención, el canal de referencia común se escoge para ser el canal combinado de sub-mezcla.

[0023] Según realizaciones adicionales de la presente invención, el canal de referencia común se escoge para ser el canal original que tiene la mayor energía.

[0024] Según realizaciones adicionales de la presente invención, el receptor de señal de audio es operativo para recibir señales de audio que comprenden canales de sub-mezcla representados por parámetros de muestreo muestreados con una frecuencia de muestreo predeterminada.

[0025] La presente invención se basa en el descubrimiento de que múltiples señales de audio paramétricamente codificadas pueden combinarse eficientemente utilizando un generador de señal de audio o un combinador de señal de audio, que genera una señal de salida de audio al combinar los canales de sub-mezcla y los parámetros asociados de las señales de entrada de audio directamente dentro del dominio de parámetro, i.e., sin reconstruir o decodificar las señales de entrada de audio individuales antes de la generación de la señal de salida de audio. Para ser más específicos, esto se logra mezclando directamente los canales de sub-mezcla asociados de las señales de entrada individuales, por ejemplo mediante la suma o formación de una combinación lineal de los mismos. Una característica clave de la presente invención es que la combinación de los canales de sub-mezcla se logra mediante simples operaciones aritméticas computacionalmente no costosas, tales como sumas.

[0026] Lo mismo se aplica para la combinación de los parámetros que se asocian a los canales de sub-mezcla. Dado que al menos un sub-conjunto de los parámetros asociados tendrá que alterarse durante la combinación de las señales de audio de entrada, lo más importante es que los cálculos llevados a cabo para alterar los parámetros sean simples y, por tanto, no necesiten energía computacional significativa ni incurrir en retrasos adicionales, e.g., utilizando bancos de filtro u otras operaciones que impliquen memoria.

[0027] De acuerdo con una realización de la presente invención, se implementa un generador de señal de audio para generar una señal de salida de audio para combinar una primera y una segunda señal de audio, encontrándose ambas paramétricamente codificadas. Para generar la señal de salida de audio, el generador de señal de audio de la invención extrae los canales de sub-mezcla de las señales de audio de entrada y genera un canal de sub-mezcla combinado formando una combinación lineal de los dos canales de sub-mezcla. Es decir, los canales individuales se agregan aplicando pesos adicionales.

[0028] En una realización preferida de la presente invención, los pesos aplicados se derivan mediante operaciones aritméticas extremadamente simples, por ejemplo, utilizando el número de canales representados por la primera señal de audio y la segunda señal de audio como una base para el cálculo.

[0029] En una realización adicional preferida, el cálculo de peso se lleva a cabo asumiendo que cada canal de audio original de las señales de entrada contribuye a la energía de señal total con la misma cantidad. Es decir, los pesos aplicados son proporciones simples de los números de canal de las señales de entrada y del número total de canales.

[0030] En una realización adicional preferida de la presente invención, los pesos de los canales de sub-mezcla individuales se calculan en base a la energía contenida dentro de los canales de sub-mezcla a fin de permitir una reproducción más auténtica del canal de sub-mezcla combinado incluido en la señal de audio de salida generada.

[0031] En una realización adicional preferida de la presente invención, el esfuerzo computacional disminuye además en que solamente los parámetros asociados con una de las dos señales de audio se alteran. Es decir, los parámetros de la otra señal de audio se transmiten inalterados, en consecuencia sin ocasionar ninguna computación y por tanto, minimizando la carga en el generador de señal de audio de la invención.

[0032] En los siguientes párrafos, el concepto de la invención se detallará principalmente para un esquema de codificación utilizando codificación de fuente de unión (JSC). En ese sentido, la presente invención extiende esta tecnología para conectar múltiples transceptores monofónicos o habilitados por JSC a estaciones remotas, mezclando las señales de sub-mezcla de JSC y la información del objeto dentro del dominio de parámetro. Como lo han mostrado las consideraciones anteriores, el concepto de la invención no se restringe en modo alguno al uso de codificación JSC, sino que también podría implementarse con codificación BCC u otros esquemas de codificación multicanales, tales como codificación MPEG de audio espacial (MPEG envolvente) y lo similar.

[0033] Dado que el concepto de la invención se detallará principalmente utilizando codificación JSC, la codificación JSC se revisará brevemente dentro de los siguientes párrafos a fin de señalar más claramente la flexibilidad del concepto de la invención y las mejoras que pueden lograrse sobre la técnica anterior al aplicar el concepto de la invención a esquemas existentes de codificación de audio multicanales.

Breve Descripción de los Dibujos

[0034]

La Figura 1 muestra un ejemplo de un esquema de codificación JSC;

La Figura 2 muestra un ejemplo de un representador de JSC;

La Figura 3 muestra un escenario de tele-conferencia con dos ubicaciones;

La Figura 4 muestra un escenario de tele-conferencia con tres ubicaciones;

La Figura 5 muestra un ejemplo de tele-conferencia utilizando un generador de señal de audio de la invención;

5 La Figura 6 muestra un ejemplo adicional de tele-conferencia utilizando un generador de señal de audio de la invención;

La Figura 6b muestra la compatibilidad inversa del concepto de la invención; y

La Figura 7 muestra un ejemplo para un generador de señal de audio de la invención.

10 **[0035]** Para la explicación de la codificación JSC, se hará referencia en lo siguiente a las Figuras 1 y 2. Dentro de las siguientes figuras, los componentes funcionalmente idénticos comparten las mismas marcas de referencia, indicando que los componentes individuales que proporcionan la misma funcionalidad pueden intercambiarse entre las realizaciones individuales de la presente invención sin perder o restringir la funcionalidad y sin limitar el alcance de la presente invención.

15 **[0036]** La Figura 1 muestra un diagrama en bloque del esquema de codificación de la fuente de unión, un codificador 2 correspondiente y un decodificador 4 correspondiente.

[0037] El codificador 2 recibe entradas de audio discretas $S_i(N)$, 6a, 6b, y 6c, y crea una señal de sub-mezcla $s(n)$ 8, por ejemplo, mediante la suma de las formas de onda.

20 **[0038]** Adicionalmente, un extractor de parámetro 10 dentro del codificador 2 extrae la información lateral paramétrica para cada objeto individual (señal 6a, 6b y 6c). Aunque no se muestra en la Figura 1, la señal de sub-mezcla 8 puede comprimirse adicionalmente mediante un codificador de habla o audio y se transmite con la información lateral paramétrica al decodificador JSC 4. Un módulo de síntesis 12, dentro del decodificador 4 regenera los estimados 14a, 14b y 14c ($s_i(n)$) de los objetos de entrada (canales 6a, 6b y 6c).

25 **[0039]** A fin de reconstruir los estimados 14a, 14b y 14c, siendo perceptualmente similares a los objetos de entrada discretos (canales de entrada) 6a, 6b y 6c, tiene que extraerse la información lateral paramétrica apropiada para cada canal. Dado que se suman los canales individuales para la generación de la señal de sub-mezcla 8, las proporciones de energía entre los canales son tales cantidades adecuadas. En consecuencia, la información paramétrica para los diferentes objetos o canales consiste de proporciones de energía delta p de cada objeto para el primer objeto (objeto de referencia).

30 **[0040]** Esta información se deriva en el dominio de frecuencia en bandas de frecuencia no uniformemente separadas (sub-bandas) que corresponden a la resolución de banda crítica de la percepción auditiva humana. Este es un concepto descrito en mayor detalle, por ejemplo, en: J. Blauert, "Spatial Hearing: The Psychophysics of Human Sound Localization" (Audición espacial: la psicofísica de la localización sonora humana), The MIT Press, Cambridge, MA, edición revisada 1997.

35 **[0041]** Es decir, los canales de audio de entrada de banda ancha se filtran hacia varias bandas de frecuencia de ancho de banda finito y para cada una de las bandas de frecuencia individuales se llevan a cabo los siguientes cálculos. Como ya se mencionó, la energía de banda del primer objeto (objeto de referencia o canal de referencia) actúa como un valor de referencia.

$$\Delta p_i(n) = 10 \log_{10} \frac{E\{\hat{s}_i^2(n)\}}{E\{s^2(n)\}}, \quad i = 2 \dots M \quad \text{Ecuación 1}$$

40 **[0042]** Para evitar la introducción adicional de artefactos, por ejemplo, introducidos por medio de una división entre cero, estas proporciones de energía (en la representación logarítmica) pueden limitarse además a un máximo de, por ejemplo, 24 dB en cada sub-banda. La proporción de energía puede además cuantificarse antes de someterse para ahorrar adicionalmente el ancho de banda de transmisión.

45 **[0043]** No es necesario transmitir explícitamente la energía del primer objeto. Por el contrario, este valor puede derivarse de la presunción de que, para objetos estadísticamente independientes, la suma de las energías de las señales sintetizadas $s_i(n)$ es igual a la energía de la señal de sub-mezcla $s(n)$. En términos de una expresión matemática, esto significa:

$$E\{s^2(n)\} = \sum_{i=1}^M E\{\hat{s}_i^2(n)\} \quad \text{Ecuación 2}$$

[0044] En base a esta presunción y ecuación, las energías de sub-banda para el primer objeto (objeto de referencia o

canal de referencia) pueden reconstruirse, como se describirá adicionalmente a continuación al detallar el concepto de la invención.

[0045] Para resumir, una señal de audio o corriente de audio, de acuerdo con JSC, comprende un canal de sub-mezcla y parámetros asociados, describiendo los parámetros proporciones de energía de los canales originales con respecto a un canal de referencia original. Puede notarse que este escenario puede alterarse fácilmente en que otros canales se seleccionan para ser el canal de referencia. Por ejemplo, el canal de referencia por sí mismo puede ser el canal de referencia, requiriendo la transmisión de un parámetro adicional, relacionando la energía del primer canal de referencia anterior con la energía del canal de sub-mezcla. También, el canal de referencia puede seleccionarse para ser variable, en que el único canal que tiene la mayor energía se selecciona para ser el canal de referencia. De aquí que, dado que la energía dentro de los canales individuales puede cambiar con el tiempo, el canal de referencia también puede variar con el tiempo. También, debido al hecho de que todo el procesamiento se lleva a cabo típicamente de una manera de frecuencia selectiva, el canal de referencia puede ser diferente para diferentes bandas de frecuencia.

[0046] La Figura 2 muestra un esquema adicional mejorado de la codificación JSC, en base al esquema de la Figura 1. Las características detalladas con respecto a la Figura 1, se encuentran incluidas con la caja de almacenamiento o transmisión 20, que recibe los canales de entrada 6 que van a codificarse y los estimados 14 que salen de los canales de entrada 6. El esquema de la Figura 2 es mejorado en que comprende además un mezclador 22 que recibe los estimados. Es decir, los objetos sintetizados 14 no salen directamente como señales de audio únicas, sino representados a los canales de salida N en el módulo de mezclador. Tal mezclados puede implementarse de diferentes maneras, por ejemplo, recibiendo los parámetros de mezclado 24 adicionales como entrada, para dirigir el mezclado de los objetos sintetizados 14. Solo como ejemplo, puede considerarse un escenario de tele-conferencia, en el cual cada uno de los canales de salida 26 se atribuye a un participante de la conferencia. En consecuencia, un participante en el extremo receptor tiene la posibilidad de separar virtualmente a los otros participantes asignando sus voces a posiciones individuales. Por tanto, no solamente la voz puede servir como criterio para distinguir entre los diferentes participantes de una conferencia telefónica, sino también la dirección desde la cual el que escucha recibe la voz de un participante. Además el que escucha puede disponer el canal de salida de tal manera que todos los participantes desde la misma ubicación de tele-conferencia se agrupan en la misma dirección, mejorando aún más la experiencia perceptual.

[0047] Como se muestra en la Figura 2, $s_1(n) \dots s_M(n)$ denota los objetos de audio discretos en la entrada del codificador JSC. En la salida del decodificador JSC $s_1(n) \dots s_M(n)$ representa los objetos de audio virtualmente separados que se alimentan en el mezclador. Los parámetros de mezclado 24 pueden modificarse interactivamente en el lado del receptor para colocar los diferentes objetos en una etapa de sonido que se reproduce mediante los canales de salida $s_1(n) \dots s_N(n)$.

[0048] La Figura 3 muestra la aplicación de esquemas de codificación de audio multicanales a un escenario básico de tele-conferencia, teniendo lugar entre dos ubicaciones. Aquí, una primera ubicación 40 se comunica con una segunda ubicación 42. La primera ubicación puede tener A participantes, i.e., A objetos de audio, la segunda ubicación tiene B participantes u objetos de audio. Para tele-conferencia de punto-a-punto, la tecnología descrita de codificación JSC puede aplicarse directamente para transmitir las señales de audio de objetos múltiples en cada ubicación a la estación remota correspondiente. Es decir, los parámetros (A-1) a_i y una sub-mezcla asociada, se transfieren a la ubicación 42. En la dirección opuesta, los parámetros (B-1) b_i se transmiten conjuntamente con una sub-mezcla asociada a la ubicación 40.

[0049] Para tele-conferencia con más de dos puntos de extremo, la situación es completamente diferente como se ilustra en la Figura 4.

[0050] La Figura 4 muestra, además de las ubicaciones 40 y 42, una tercera ubicación 44. Como puede observarse en la Figura 4, tal escenario necesita un distribuidor central para las señales de audio asociadas, generalmente llamado unidad de control de puntos múltiples, MCU. Cada una de las ubicaciones (sitios) 40, 42 y 44 se encuentra conectado a la MCU 46. Para cada sitio 40, 42 y 44, existe una corriente ascendente única para la MCU que contiene la señal del sitio. Dado que cada sitio individual necesita recibir las señales de los sitios restantes, la corriente descendente para cada sitio 40, 42 y 44 es una mezcla de las señales de los otros sitios, excluyendo la señal propia del sitio, que también se refiere como la señal (N-1). Generalmente, para cumplir con los requerimientos de la instalación y para mantener el ancho de banda de transmisión razonablemente bajo, no es posible transmitir las corrientes codificadas JSC N-1 desde la MCU a cada sitio. Esto, por supuesto, sería la opción directa.

[0051] El estado de la técnica que apunta a derivar las corrientes descendentes individuales es para sintetizar todas las corrientes entrantes (objetos) dentro de la MCU 46 utilizando un decodificador JSC. Entonces, los objetos de audio re-sintetizados podrían reagruparse y re-codificarse a fin de proporcionar cada sitio con corrientes de audio que comprenden los objetos de audio o los canales de audio deseados. Incluso dentro de este simple escenario, esto significaría tres tareas de decodificación y tres de codificación, las cuales deben llevarse a cabo simultáneamente dentro de la MCU 46. A pesar de las demandas computacionales significativas, pueden esperarse adicionalmente artefactos audibles mediante este proceso paramétrico de "codificación en serie" (codificación/decodificación repetida). El incremento del número de sitios incrementaría adicionalmente el número de corrientes y, por tanto, el número de procesos de codificación o decodificación requerido, haciendo imposibles los procedimientos directos para escenarios en tiempo real.

[0052] De acuerdo con la presente invención, en consecuencia, se desarrolla un esquema para mezclar diferentes

corrientes paramétricamente codificadas (corrientes JSC en este ejemplo particular) directamente dentro del dominio de sub-mezcla y del objeto para tal escenario tipo MCU, creando las señales de salida deseadas (corrientes de audio de salida) con un mínimo de esfuerzo computacional y de pérdida de calidad.

[0053] Dentro de los siguientes párrafos, el concepto de la invención de mezclar directamente las corrientes de audio multicanales paramétricamente codificadas dentro del dominio de parámetro, se detalla para corrientes de audio codificadas por JSC.

[0054] El concepto de la invención se explica con la combinación de dos señales de audio originales (corrientes) en una señal de salida. Al unir tres o más corrientes juntas pueden derivarse fácilmente en el caso de combinar dos corrientes. Las siguientes consideraciones matemáticas se ilustran por la Figura 5, que muestra un caso en donde tienen que combinarse tres canales de audio del sitio A con cuatro canales de audio del sitio B. Este es, por supuesto, solamente un ejemplo para visualizar el concepto de la invención. Al utilizar la codificación JSC, el sitio 50 (A) que tiene tres participantes de conferencia (hablantes) 52a a 52c, que generan señales s_{Ax} , transmite una corriente de audio o señal de audio 54. La señal de audio 54 tiene un canal de sub-mezcla s_A y parámetros a_2 y a_3 , que relacionan la energía de los canales 52b y 52c con la energía del canal 52a. De manera equivalente, el sitio 56 (B) transmite una señal de audio 58 que tiene un canal de sub-mezcla s_B y tres parámetros b_2 , b_3 y b_4 , que son la representación codificada por JSC de cuatro hablantes 60a a 60d. La MCU 46 combina las señales de audio 54 y 58 para derivar una señal de salida 62 que tiene un canal de sub-mezcla combinado s_Y y seis parámetros y_2, \dots, y_7 .

[0055] En el lado receptor, el receptor 64 decodifica la señal de salida 62 para derivar representaciones de los 7 objetos de audio o canales de audio de los sitios 50 y 56.

[0056] En términos generales, el objetivo es formar una sola representación combinada 62 de dos corrientes de JSC 54 y 58, representando cada una un número de objetos mediante una señal de sub-mezcla común s_Y y un conjunto de parámetros de objeto que caracterizan los objetos. Idealmente, la representación de JSC combinada debe ser idéntica a la que se obtendría codificando el conjunto total de señales fuente originales que unen ambas corrientes JSC en una sola corriente JSC en una etapa.

[0057] Para mantener simples las siguientes ecuaciones, asumimos que las proporciones de energía relativas de la Ecuación 1 no se encuentran disponibles en el dominio logarítmico, sino solo como proporciones de energía. Cada parámetro de objeto $r_i(n)$ de un cierto objeto i puede derivarse como

$$r_i(n) = \frac{E\{s_{r_i}^2(n)\}}{E\{s_{r_i}^2(n)\}} \quad \text{Ecuación 3}$$

[0058] La transposición en el dominio logarítmico puede aplicarse posteriormente a cada parámetro a fin de permitir la cuantificación utilizando una escala logarítmica de energía.

[0059] Todas las señales siguientes se asumen descompuestas en una representación de sub-banda, por tanto, cada uno de los cálculos se aplica para cada sub-banda por separado.

[0060] Tenemos la corriente A, con su señal de sub-mezcla s_A y los parámetros (proporciones de energía relativas) para los objetos U $a_2 \dots a_U$. La corriente B consiste de la señal de sub-mezcla s_B y los parámetros para los objetos V $b_2 \dots b_V$.

[0061] La señal de sub-mezcla combinada s_Y puede formarse como una combinación lineal de ambas señales de sub-mezcla s_A y s_B . Para asegurar el nivelado de volumen correcto de las diferentes contribuciones de objeto, pueden aplicarse los factores de ganancia g_A y g_B .

$$s_Y = g_A \cdot s_A + g_B \cdot s_B$$

con

$$g_A = \frac{U}{(U + V)}, g_B = \frac{V}{(U + V)}$$

[0062] Este tipo de escala puede ser significativa si se han sumado fuentes de sonido individuales de igual energía promedio y si se han normalizado a la escala total de la trayectoria de sub-mezcla.

[0063] Alternativamente, podría utilizarse un procedimiento de conservación de energía para los factores de ganancia con

$$g_A = \sqrt{\frac{U}{(U + V)}}, g_B = \sqrt{\frac{V}{(U + V)}}$$

[0064] Otra posibilidad es seleccionar el factor de ganancia de tal manera que ambas señales de sub-mezcla contribuyan a la misma energía promedio para la sub-mezcla combinada, i.e., seleccionando

$$\frac{g_B}{g_A} = \sqrt{\frac{E\{s_A^2(n)\}}{E\{s_B^2(n)\}}}.$$

[0065] Los parámetros de objeto y_i para la corriente combinada s_y deben representar todos los objetos $U + V$.

[0066] Dado que los parámetros asociados con los canales de sub-mezcla son proporciones de energía relativa, los parámetros a_2, \dots, a_U pueden utilizarse como se encuentran (inalterados) y los parámetros para los objetos de B pueden concatenarse a los parámetros a_2, \dots, a_U . Una vez seleccionado el primer objeto de la señal A para ser el objeto de referencia o canal de referencia, los parámetros originales b_i tienen que transformarse para relacionarse con ese canal de referencia. Puede notarse que solamente los parámetros de una corriente tienen que recalcularse, disminuyendo adicionalmente la carga computacional dentro de una MCU 46.

[0067] Puede notarse además que de ningún modo es necesario utilizar el canal de referencia de una de las corrientes de audio originales como nuevo canal de referencia. El concepto de la invención de combinar corrientes de audio paramétricamente codificadas dentro del dominio de parámetro, puede implementarse también muy bien con otros canales de referencia, seleccionados del número de canales originales de los sitios A o B. Una posibilidad adicional sería utilizar el canal de sub-mezcla combinado como un nuevo canal de referencia.

[0068] Siguiendo este procedimiento de utilizar el canal de referencia original del sitio A como nuevo canal de referencia (canal de referencia combinado), la energía poder) del primer objeto (canal) de cada señal A y B tiene que calcularse primero, dado que estos se encuentran solo implícitamente disponibles.

[0069] La representación de energía para la señal de sub-mezcla A, asumiendo las fuentes estadísticamente independientes, proporciona:

$$E\{S_A^2(n)\} = \sum_{i=1}^U E\{\hat{S}_i^2(n)\}.$$

[0070] Las energías de señal $E\{S_{A_2}^2(n)\} \dots E\{S_{A_U}^2(n)\}$, se definen con sus energías relativas $\alpha_2 \dots \alpha_U$ hasta

$$E\{S_{A_2}^2(n)\}:$$

[0071]

$$E\{S_{A_2}^2(n)\} = a_2 \cdot E\{S_{A_1}^2(n)\}$$

$$E\{S_{A_3}^2(n)\} = a_3 \cdot E\{S_{A_1}^2(n)\}$$

...

$$E\{S_{A_U}^2(n)\} = a_U \cdot E\{S_{A_1}^2(n)\}$$

[0072] Esto conduce a la energía de s_{A1} como:

5

$$E\left\{S_{A_1}^2(n)\right\} = \frac{E\left\{S_A^2(n)\right\}}{(1 + a_2 + a_3 + \dots + a_U)}$$

[0073] Al aplicar lo mismo para la señal de sub-mezcla s_B , se puede calcular la energía del objeto s_{B1} como:

10

$$E\left\{S_{B_1}^2(n)\right\} = \frac{E\left\{S_B^2(n)\right\}}{(1 + b_2 + b_3 + \dots + b_V)}$$

15

[0074] Ahora se puede construir el nuevo conjunto de parámetros para todos los objetos de la señal s_y :

y_1 : (no transmitido, objeto de referencia, implícitamente disponible)

$y_2 = a_2$

$y_3 = a_3$

...

$y_U = a_U$

20

25

$$y_{U+1} = \frac{g_B^2}{g_A^2} \cdot \frac{E\left\{S_{B_1}^2(n)\right\}}{E\left\{S_{A_1}^2(n)\right\}},$$

30

(proporción de energía del primer objeto de la señal B con respecto al objeto de referencia A1)

35

$$y_{U+2} = b_2 \cdot \frac{g_B^2}{g_A^2} \cdot \frac{E\left\{S_{B_1}^2(n)\right\}}{E\left\{S_{A_1}^2(n)\right\}},$$

40

(proporción de energía del segundo objeto de la señal B renormalizada a la energía del objeto de referencia A1)

45

50

$$y_{U+3} = b_3 \cdot \frac{g_B^2}{g_A^2} \cdot \frac{E\left\{S_{B_1}^2(n)\right\}}{E\left\{S_{A_1}^2(n)\right\}}$$

$$y_{U+V} = b_V \cdot \frac{g_B^2}{g_A^2} \cdot \frac{E\left\{S_{B_1}^2(n)\right\}}{E\left\{S_{A_1}^2(n)\right\}}$$

65

[0075] Como lo han mostrado los párrafos anteriores, el concepto de la invención permite la generación de una corriente de audio combinada utilizando solamente operaciones aritméticas simples, siendo, por tanto computacionalmente

extremadamente eficiente. Por tanto, la combinación de múltiples corrientes de audio paramétricamente codificadas puede llevarse a cabo en tiempo real.

[0076] Para enfatizar adicionalmente la gran flexibilidad del concepto de la invención, la Figura 6 muestra cómo una señal monofónica 70, ocasionada por un solo hablante en el sitio 56, puede combinarse de acuerdo con la invención, con dos o más señales codificadas por JSC de los hablantes en el sitio 50. Es decir, debido a la flexibilidad del concepto de la invención, las señales monofónicas de los sistemas de tele-conferencia arbitrarios pueden combinarse de acuerdo con la invención, con fuentes multicanales paramétricamente codificadas (multi-objetos) para generar una señal de audio codificada por JSC que representa todos los canales de audio originales (objetos).

[0077] Al extender la compatibilidad también con estaciones remotas que no son capaces de transmitir objetos JSC, sino señales tradicionales monofónicas, esta técnica también es aplicable para insertar un objeto monofónico, e.g., desde un dispositivo de conferencia legado hacia la corriente basada en el objeto.

[0078] El ejemplo anterior con la corriente JSC A (sub-mezcla s_A , parámetros $a_2...a_U$) y un objeto monofónico C (sub-mezcla s_C) conduce a una señal combinada Z con la señal de sub-mezcla

$$S_Z = g_A \cdot S_A + g_C \cdot S_C$$

con factores de ganancia como se trató previamente y sus parámetros de objeto:

$$\begin{aligned} y_1: & \text{no transmitido (canal de referencia, implícitamente disponible)} \\ y_2 &= a_2 \\ y_3 &= a_3 \\ &\dots \\ y_U &= a_U \end{aligned}$$

$$Y_{U+1} = \frac{g_C^2 E\{S_C^2(n)\}}{g_A^2 E\{S_{A_1}^2(n)\}},$$

(proporción de energía de la señal C con respecto al objeto de referencia A1)

[0079] El ejemplo antes mencionado para transcodificar/fusionar dos corrientes JSC, depende de la representación de la energía de los objetos como se proporciona en la Ecuación 1. No obstante, el mismo esquema de la invención puede aplicarse también a otras formas de representar esta información.

[0080] La Figura 6b enfatiza de nuevo la gran flexibilidad del concepto de la invención incorporando una fuente de audio monofónica. La Figura 6b se basa en el escenario multicanales de la Figura 4 y además muestra cuán fácilmente puede integrarse un codificador de audio monofónico de la técnica anterior en una fuente de audio c(44), en una conferencia de audio multicanales utilizando la MCU 46 de la invención.

[0081] Como se mencionó previamente, el concepto de la invención no se restringe a la codificación JSC teniendo un canal de referencia fijo predeterminado. En consecuencia, en un ejemplo alternativo, la proporción de energía puede computarse con respecto a un canal de referencia, que es variable con el tiempo, siendo el canal de referencia uno que tiene la mayor energía dentro de un intervalo de tiempo predeterminado.

[0082] En lugar de normalizar los valores de energía de la señal en forma de banda con la energía de la banda correspondiente de un canal (objeto) de referencia fijo y de transponer el resultado al dominio logarítmico (dB) como se señala en la Ecuación 1, la normalización puede tener lugar en relación a la máxima energía sobre todos los objetos en una cierta banda de frecuencia:

$$pnorm_i(n) = \frac{E\{S_i^2(n)\}}{\max_i \left(E\{S_i^2(n)\} \right)}, i = 1..M \quad \text{Ecuación 4}$$

[0083] Estos valores de energía normalizados (que se proporcionan en una representación lineal) no necesitan ninguna limitación adicional para un cierto límite superior, dado que de manera innata pueden tomar solo los valores entre 0 y 1.

Esta ventaja supera la desventaja de tener que transmitir un parámetro adicional para el canal de referencia no conocido ya a priori.

[0084] El proceso de mezclado para este escenario incluiría las siguientes etapas (que de nuevo tienen que llevarse a cabo para cada sub-banda por separado):

5 Tenemos una corriente A con su señal de sub-mezcla s_A y los parámetros (valores de energía normalizados, Ecuación 3, Ecuación 1) para los objetos U $a_1 \dots a_U$.

La corriente B consiste de la señal de sub-mezcla s_B y los parámetros para los objetos V $b_1 \dots b_V$.

[0085] Una señal de sub-mezcla combinada puede formarse de acuerdo con una de las opciones ya mostradas:

10

$$s_Y = g_A \cdot s_A + g_B \cdot s_B$$

[0086] Todos los valores de energía normalizados para la representación combinada y_i tienen que ajustarse en relación con el objeto con la energía más alta de todos los objetos de la señal Y. No existen dos candidatos para ser este "objeto máximo" de Y, tanto el objeto máximo de A como el objeto máximo de B, pueden identificarse teniendo una proporción de energía normalizada de "1".

15

[0087] Esta decisión puede efectuarse comparando la energía absoluta de ambos candidatos. De nuevo puede utilizarse la relación para la energía de las señales de sub-mezcla (Ecuación 2) para obtener:

20

25

$$E\left\{S_{A_{max}}^2(n)\right\} = \frac{E\left\{S_A^2(n)\right\}}{\sum_{i=1}^U a_i} \quad y \quad E\left\{S_{B_{max}}^2(n)\right\} = \frac{E\left\{S_B^2(n)\right\}}{\sum_{i=1}^V b_i}$$

30

35

[0088] Ahora, podemos comparar las energías máximas de objeto ponderadas con los factores de ganancia del proceso de sub-mezcla:

40

$$g_A^2 \cdot E\left\{S_{A_{max}}^2(n)\right\} > g_B^2 \cdot E\left\{S_{B_{max}}^2(n)\right\}?$$

[0089] Cualquiera que sea la energía más alta del objeto, este objeto servirá como "objeto máximo" para los parámetros combinados y_i .

45

[0090] Como un ejemplo, sea a_2 el objeto de energía máxima total a_{max} de ambas señales A y B, entonces todos los otros parámetros pueden combinarse como:

50

$$\begin{aligned} Y_1 &= a_1 \\ Y_2 &= a_2 \\ &\dots \\ y_U &= a_U \end{aligned}$$

55

$$y_{U+1} = \frac{g_B^2 \cdot E\left\{S_{B_{max}}^2(n)\right\}}{g_A^2 \cdot E\left\{S_{A_{max}}^2(n)\right\}},$$

60

(proporción de energía del primer objeto de la señal B con respecto al "objeto máximo", aquí a_2)

$$Y_{U+2} = b_2 \cdot \frac{g_B^2}{\gamma} \cdot \frac{E \left\{ S_{B_{max}}^2(n) \right\}}{\gamma},$$

$$Y_{U+V} = b_V \cdot \frac{g_B^2}{g_A^2} \cdot \frac{E \left\{ S_{B_{max}}^2(n) \right\}}{E \left\{ S_{A_{max}}^2(n) \right\}},$$

[0091] Para este ejemplo, todos los parámetros para los objetos A pueden permanecer sin cambio, dado que la señal A contuvo el objeto máximo total.

[0092] También en esta representación, la inserción de un objeto monofónico puede efectuarse por consiguiente, e.g., asumiendo que $V = 1$.

[0093] Generalmente, el proceso de transcodificación se lleva a cabo de tal manera que su resultado logra el resultado que se habría obtenido si todos los objetos originales para ambas corrientes se hubieran codificado en una sola corriente JSC en primer lugar.

[0094] La Figura 7 muestra un ejemplo para un generador de señal de audio de la invención para generar una señal de salida de audio, como puede utilizarse dentro de la MCU 46 para implementar el concepto de la invención.

[0095] El generador de señal de audio 100 comprende un receptor de señal de audio 102, un combinador de canales 104, un calculador de parámetro 106 y una interfaz de salida 108.

[0096] El receptor de señal de audio 103 recibe una primera señal de audio 110 que comprende un primer canal de sub-mezcla 110a que tiene información sobre dos o más de los primeros canales originales y que comprende un parámetro original 110b asociado con uno de los primeros canales originales que describe una propiedad de uno de los primeros canales originales con respecto a un canal de referencia. El receptor de señal de audio 102 recibe además una segunda señal de audio 112 que comprende un segundo canal de sub-mezcla 112a que tiene información sobre al menos un segundo canal original.

[0097] El receptor de señal de audio da salida al primer canal de sub-mezcla 110a y al segundo canal de sub-mezcla 112a hacia una entrada del combinador de canales 104 y al primer canal de sub-mezcla 110a, el segundo canal de sub-mezcla 112a y el parámetro original 110b al calculador de parámetro 106.

[0098] El combinador de canales 104 deriva un canal de sub-mezcla combinado 114 al combinar el primer canal de sub-mezcla 110a y un segundo canal de sub-mezcla 112b, i.e., al combinar los canales de sub-mezcla directamente sin reconstruir los canales de audio originales subyacentes.

[0099] El calculador de parámetro 106 deriva un primer parámetro combinado 116a que describe la propiedad de uno de los primeros canales originales con respecto a un canal de referencia común y un segundo parámetro combinado 116b que describe la propiedad de otro de los primeros canales originales o del al menos un segundo canal original con respecto al mismo canal de referencia común. Los parámetros combinados primero y segundo se introducen en la interfaz de salida 108, que además recibe el canal de sub-mezcla combinado 114 desde el combinador de canales 104. Finalmente, la interfaz de salida da salida a una señal de salida 120 que comprende el canal de sub-mezcla combinado 114 y los parámetros combinados primero y segundo 116a y 116b.

[0100] La señal de salida de audio se ha derivado, por tanto, sin la reconstrucción total de las señales de audio de entrada y por tanto, sin operaciones computacionalmente costosas.

[0101] Dentro de los párrafos anteriores, se ha mostrado el concepto general de mezclar dos o más señales, cada una basada en un procedimiento paramétrico de JSC. Particularmente, las ecuaciones anteriores muestran cómo aplicar estas técnicas para un caso, en donde la información paramétrica consiste de proporciones de energía relativas. No obstante, esta técnica no se restringe a una representación específica de los parámetros de objeto. En consecuencia, pueden utilizarse también los parámetros que describen mediciones de amplitud u otras propiedades de los canales de audio individuales, tales como las correlaciones. Las proporciones de energía también pueden computarse con respecto al canal de sub-mezcla combinado, al costo de transmitir un parámetro adicional. Por otra parte, uno se beneficia de este escenario alternativo por la reducida complejidad computacional durante el mezclado de las corrientes de audio, dado que la reconstrucción de la energía del canal de referencia, que no se transmite explícitamente en JSC "genérico", es obsoleta.

[0102] Además, la invención no se limita a un escenario de tele-conferencia, sino que puede aplicarse cuando se desea

la multiplexión de objetos paramétricos en una sola corriente. Este, por ejemplo, puede ser el caso en esquemas de codificación BCC, envolventes espaciales MPEG y otros.

[0103] Como se ha mostrado, el concepto de la invención permite incluso de manera inconsútil incluir estaciones remotas legadas que proporcionan una sola señal monofónica en el escenario en base al objeto. Además de la combinación de diferentes corrientes de objetos, el concepto de la invención muestra también cuantas formas diferentes de representar los datos paramétricos pueden generarse, de manera que sean adecuadas para permitir procesos de combinación computacionalmente eficientes. Como tal, es una característica ventajosa de una sintaxis de corriente de bits paramétrico de la invención expresar las propiedades del objeto de tal manera que dos corrientes puedan combinarse llevando a cabo meramente operaciones simples.

[0104] En consecuencia, el concepto de la invención muestra también cómo crear corrientes de bits apropiadas o formatos de corriente de bits para codificar paraméricamente múltiples canales de audio (objetos de audio), adhiriéndose a los siguientes criterios:

- La señal de sub-mezcla combinada se forma simplemente a partir de las señales de sub-mezcla parcial.
- La información lateral paramétrica combinada se forma de la combinación de la información lateral paramétrica individual y alguna simple para computar las características de las señales de sub-mezcla (e.g., energía).
- En ningún caso tiene que efectuarse una operación compleja tal como una etapa de decodificación/codificación para los objetos de audio.

[0105] En consecuencia, la representación paramétrica que describe los objetos debe seleccionarse de tal manera que una combinación (“adición”) de dos o más corrientes de objeto sea posible utilizando solamente campos de corriente de bits que se encuentren disponibles como parte de la información lateral paramétrica, y posiblemente simples para computar las métricas de las señales de sub-mezcla (e.g., energía, valor pico).

[0106] Un ejemplo para tal representación podría ser utilizar valores de energía normalizados (Ecuación 4) para cada objeto. Éstos podrían transformarse en una representación logarítmica (dB) y después cuantificarse a cierto número de etapas de cuantificador o a sus índices de cuantificador representativos. La sintaxis de corriente de bits debe permitir el incremento fácil (o disminución) del número de parámetros de objeto en una corriente, e.g., simplemente concatenando, insertando o retirando los parámetros.

[0107] Resumiendo, el concepto de la invención permite una combinación más flexible y computacionalmente eficiente de las corrientes de audio paraméricamente codificadas. Debido a la alta eficiencia computacional, el concepto de la invención no se restringe a un número máximo de canales a combinar. Principalmente, los canales, que pueden combinarse en tiempo real, pueden proporcionarse a un generador de señal de audio de la invención en números arbitrarios. También, la representación paramétrica precisa (JSC) utilizada para ilustrar el concepto de la invención no es obligatoria. Además, como ya se mencionó, otros esquemas de codificación paramétrica, tales como los esquemas envolventes comúnmente conocidos, pueden ser la base para la aplicación del concepto de la invención.

[0108] Además, las computaciones necesarias no necesariamente tienen que aplicarse en software. Las implementaciones de hardware utilizando, por ejemplo, DSPs, ASICs y otros circuitos integrados, también pueden utilizarse para llevar a cabo los cálculos, que incrementarán aún más la velocidad del concepto de la invención, permitiendo la aplicación del concepto de la invención en escenarios en tiempo real.

[0109] Debido a la flexibilidad del concepto de la invención, las corrientes de audio de la invención pueden basarse en diferentes representaciones paramétricas. Los parámetros que van a transmitirse podrían, por ejemplo, ser también mediciones de amplitud, diferencias de tiempo entre los canales de audio originales, mediciones de coherencia y otros.

[0110] Por tanto, se ha mostrado el concepto general de mezclar dos o más señales que se basan cada una en un procedimiento paramétrico tipo JSC.

[0111] Las ecuaciones anteriores muestran cómo aplicar esta técnica para un caso, en donde la información paramétrica consiste de proporciones de energía relativas. No obstante, esta técnica no se restringe a una representación específica de los parámetros de objeto.

[0112] Además, la invención no se limita a un escenario de tele-conferencia, sino que puede aplicarse en cualquier caso en donde sea ventajosa la multiplexión de objetos paramétricos en una sola corriente JSC.

[0113] Además, esta técnica permite de manera inconsútil incluir estaciones remotas legadas que proporcionan una sola señal monofónica en el escenario en base al objeto.

[0114] Además del proceso real para combinar diferentes corrientes de objeto, la invención muestra también cuántas formas diferentes de representar los datos paramétricos son adecuadas para permitir este proceso de combinación. Dado que no todas las posibles representaciones paramétricas permiten tal proceso de combinación descrito sin la decodificación/recodificación total de los objetos, es una característica ventajosa de la sintaxis de corriente de bits

paramétrica expresar las propiedades del objeto de tal manera que puedan combinarse dos corrientes llevando a cabo meramente operaciones simples.

5 [0115] Dependiendo de ciertos requerimientos de implementación de los procedimientos de la invención, los procedimientos de la invención pueden implementarse en hardware o software. La implementación puede llevarse a cabo utilizando un medio de almacenamiento digital, en particular, un disco, DVD o un CD que tiene señales de control electrónicamente legibles almacenadas en el mismo, que cooperan con un sistema computarizado programable de tal manera que se llevan a cabo los procedimientos de la invención. Generalmente, la presente invención es, en consecuencia, un producto de programa de computadora con un código de programa almacenado en un vehículo legible en máquina, siendo el código del programa operativo para llevar a cabo los procedimientos de la invención, cuando el producto de programa de computadora corre en una computadora. En otras palabras, los procedimientos de la invención, en consecuencia, son un programa de computadora que tiene un código de programa para llevar a cabo al menos uno de los procedimientos de la invención cuando el programa de computadora corre en una computadora.

10
15 [0116] Aunque lo anterior se ha mostrado particularmente y se ha descrito con referencia a realizaciones particulares del mismo, se entenderá por los expertos en la técnica, que otros varios cambios en forma y detalles, pueden efectuarse sin apartarse del alcance de las mismas. Se entenderá que pueden efectuarse varios cambios para adaptarse a diferentes realizaciones sin apartarse de los más amplios conceptos descritos en la presente y comprendidos por las reivindicaciones siguientes.

REIVINDICACIONES

1. Un generador de señal de audio (100) para generar una señal de salida de audio, que comprende:

5 un receptor (102) de señal de audio para recibir:

una primera señal de audio (100) que comprende un primer canal de sub-mezcla (110a) que tiene información sobre dos o más primeros canales originales y que comprende al menos un parámetro original (110b) asociado con uno de los primeros canales originales que describe una relación de potencia de uno de los primeros canales originales con respecto a un canal de referencia; y

una segunda señal de audio (112) que comprende un segundo canal de sub-mezcla (112a) que tiene información sobre al menos un segundo canal original;

15 un combinador de canales (104) para derivar un canal de sub-mezcla combinado (114), al combinar el primer canal de sub-mezcla (110a) y el segundo canal de sub-mezcla (112a);

un calculador de parámetro (106) para derivar, empleando la energía $E\{S_A^2(n)\}$ primer canal de sub-mezcla (110a) y la energía $E\{S_B^2(n)\}$ del al menos un parámetro original (110b), o como alternativa una relación de potencia con respecto a un canal de referencia común, siendo el canal de referencia común el canal original entre los primeros canales originales y teniendo el al menos un segundo canal original la máxima energía dentro de un determinado intervalo de tiempo predeterminado, un primer parámetro combinado (116a) que describe la relación de potencia de uno de los primeros canales originales con respecto a un canal de referencia común, y un segundo parámetro combinado (116b) que describe la relación de potencia de otro de los primeros canales originales o del al menos un segundo canal original con respecto al canal de referencia común; y

una interfaz de salida para producir la señal de salida de audio (120) que comprende el canal de sub-mezcla combinado (114), y los parámetros combinados (116a, 116b) primero y segundo.

30 2. El generador de señal de audio (100) de acuerdo con la reivindicación 1, en el cual el combinador de canales (104) es operativo para derivar el canal de sub-mezcla combinado (114) utilizando una combinación lineal del primero (110a) y el segundo canal de sub-mezcla (110b).

35 3. El generador de señal de audio (100) de acuerdo con la reivindicación 2, en el cual el combinador de canales (104) es operativo para utilizar una combinación lineal que tiene coeficientes que dependen del número U de los primeros canales originales y del número V de los segundos canales originales.

40 4. El generador de señal de audio (100) de acuerdo con la reivindicación 3, en el cual el combinador de canales (104) es operativo para utilizar una combinación lineal que tiene un coeficiente g_A para el primer canal de sub-mezcla (110a), y un coeficiente g_B para el segundo canal de sub-mezcla (112a) derivado utilizando una de las siguientes ecuaciones:

45
$$g_A = \frac{U}{(U + V)}, g_B = \frac{V}{(U + V)}$$

50
$$g_A = \sqrt{\frac{U}{(U + V)}}, g_B = \sqrt{\frac{V}{(U + V)}}$$

55 5. El generador de señal de audio de acuerdo con cualquiera de las reivindicaciones previas, en el cual el calculador de parámetro (106) es operativo para calcular la energía $E\{S_{Aref}\}$ del canal de referencia al derivar la energía $E\{S_A^2\}$ del primer canal de sub-mezcla (110a) y los parámetros $a_i\{i = 1, \dots, n\}$ asociados con canales diferentes al canal de referencia de acuerdo con la ecuación:

5

$$E\left\{S_{A_{ref}}^2\right\} = \frac{E\left\{S_A^2(n)\right\}}{1 + \sum_i a_i} \bullet$$

10 6. El generador de señal de audio (100) de acuerdo con cualquiera de las reivindicaciones previas, en el cual el calculador de parámetro (106) es operativo para utilizar el canal de referencia como el canal de referencia común y el parámetro original a_2 como el primer parámetro combinado y_u y para derivar el segundo parámetro combinado y_{u+1} para el al menos un segundo canal original con respecto al canal de referencia.

15 7. El generador de señal de audio (100) de acuerdo con cualquiera de las reivindicaciones anteriores, en el cual el calculador de parámetro (106) es operativo para utilizar además los coeficientes g_A asociados con el primer canal de sub-mezcla (110a) y g_B asociados con el segundo canal de sub-mezcla (112a), utilizándose los coeficientes para la combinación lineal de la primera y segunda sub-mezcla utilizada por el combinador de canales (104).

20 8. El generador de señal de audio (100) según cualquiera de las reivindicaciones anteriores, en el cual el calculador de parámetro (106) es operativo para calcular el segundo parámetro combinado y_{u+1} para el al menos un segundo canal original de acuerdo con la siguiente ecuación:

25

$$y_{u+1} = \frac{g_B^2 E\left\{S_{B_1}^2(n)\right\}}{g_A^2 E\left\{S_{A_1}^2(n)\right\}},$$

30

en donde $E\left\{S_{A_1}^2(n)\right\}$ es la energía del canal de referencia derivada utilizando la energía del primer canal de sub-mezcla $E\left\{S_A^2(n)\right\}$ de acuerdo con la siguiente fórmula:

35

$$E\left\{S_{A_1}^2(n)\right\} = \frac{E\left\{S_A^2(n)\right\}}{(1 + a_2)},$$

40

en donde a_2 es el parámetro original que relaciona un primer canal original con el canal de referencia.

45 9. El generador de señal de audio (100) de acuerdo con cualquiera de las reivindicaciones previas, en el cual el calculador de parámetro (106) es operativo para procesar las porciones de frecuencia del primero y el segundo canales de sub-mezcla asociados con intervalos de frecuencia separados de tal manera que los parámetros combinados se derivan para cada intervalo de frecuencia separado.

50 10. Un procedimiento para generar una señal de salida de audio, comprendiendo el procedimiento:

recibir una primera señal de audio (110) que comprende un primer canal de sub-mezcla (110a) que tiene información sobre dos o más primeros canales originales y que comprende al menos un parámetro original (110b) asociado con uno de los primeros canales originales que describe una relación de potencia de uno de los primeros canales originales con respecto a un canal de referencia; y una segunda señal de audio (112) que comprende un segundo canal de sub-mezcla (112a) que tiene información sobre al menos un segundo canal original;

55 derivar un canal de sub-mezcla combinado (114) al combinar el primer canal de sub-mezcla (110) y el segundo canal de sub-mezcla (112);

60 derivar, empleando la energía $E\{S_A^2(n)\}$ del primer canal de sub-mezcla (110a) y la energía $E\{S_B^2(n)\}$ del segundo canal de sub-mezcla (112b) y el al menos un parámetro original (110b), o como alternativa una relación de potencia con respecto a un canal de referencia común, siendo el canal de referencia común el canal original entre los primeros canales originales y teniendo el al menos un segundo canal original la máxima energía dentro de un determinado

intervalo de tiempo predeterminado, un primer parámetro combinado (116a) que describe la relación de potencia de uno de los primeros canales originales con respecto a un canal de referencia común, y un segundo parámetro combinado (116b) que describe la relación de potencia de otro de los primeros canales originales o del al menos un segundo canal original con respecto al canal de referencia común; y

producir la señal de salida de audio (120) que comprende el canal de sub-mezcla combinado (114), y los parámetros combinados (116a, 116b) primero y segundo.

11. Una corriente de datos de una representación de tres o más canales de audio (120), que comprende:

un canal de sub-mezcla combinado (114) que es una combinación de un primer canal de sub-mezcla que tiene información sobre al menos dos primeros canales originales y un segundo canal de sub-mezcla que tiene información sobre al menos un segundo canal original;

un primer parámetro (116a) que depende de la energía $E\{S_A^2(n)\}$ del primer canal de sub-mezcla (110a) y la energía $E\{S_B^2(n)\}$ del segundo canal de sub-mezcla (112b) que describe una relación de potencia de uno de los al menos dos primeros canales originales con respecto a un canal de referencia y que depende de al menos un parámetro original (110b) asociado con uno de los primeros canales que describen una relación de potencia de uno de los primeros canales originales con respecto a un canal de referencia; y un segundo parámetro (116b) que depende de la energía $E\{S_A^2(n)\}$ del primer canal de sub-mezcla (110a) y la energía $E\{S_B^2(n)\}$ del segundo canal de sub-mezcla (112b) que describe la relación de potencia de otro canal de los primeros canales originales o la relación de potencia del al menos un segundo canal original con respecto al canal de referencia.

12. Sistema de conferencia que comprende un generador de señal de audio (100) para generar una señal de salida de audio de acuerdo con la reivindicación 1.

13. Un programa de computadora para implementar, al ejecutarse en una computadora, un procedimiento para generar una señal de salida de audio, comprendiendo el procedimiento:

recibir una primera señal de audio que comprende un primer canal de sub-mezcla que tiene información sobre dos o más primeros canales originales y que comprende al menos un parámetro original asociado con uno de los primeros canales originales que describe una relación de potencia de uno de los primeros canales originales con respecto a un canal de referencia y una segunda señal de audio que comprende un segundo canal de sub-mezcla que tiene información sobre al menos un segundo canal original;

derivar un canal de sub-mezcla combinado al combinar el primer canal de sub-mezcla y el segundo canal de sub-mezcla;

derivar, empleando la energía $E\{S_A^2(n)\}$ del primer canal de sub-mezcla (110a) y la energía $E\{S_B^2(n)\}$ del segundo canal de sub-mezcla (112b) y el al menos un parámetro original (110b), o como alternativa una relación de potencia con respecto a un canal de referencia común, siendo el canal de referencia común el canal original entre los primeros canales originales y teniendo el al menos un segundo canal original la máxima energía dentro de un determinado intervalo de tiempo predeterminado, un primer parámetro combinado que describe la relación de potencia de uno de los primeros canales originales con respecto a un canal de referencia común, y un segundo parámetro combinado que describe la relación de potencia de otro de los primeros canales originales o del al menos un segundo canal original con respecto al canal de referencia común; y

producir la señal de salida de audio que comprende el canal de sub-mezcla combinado, y los parámetros combinados primero y segundo.

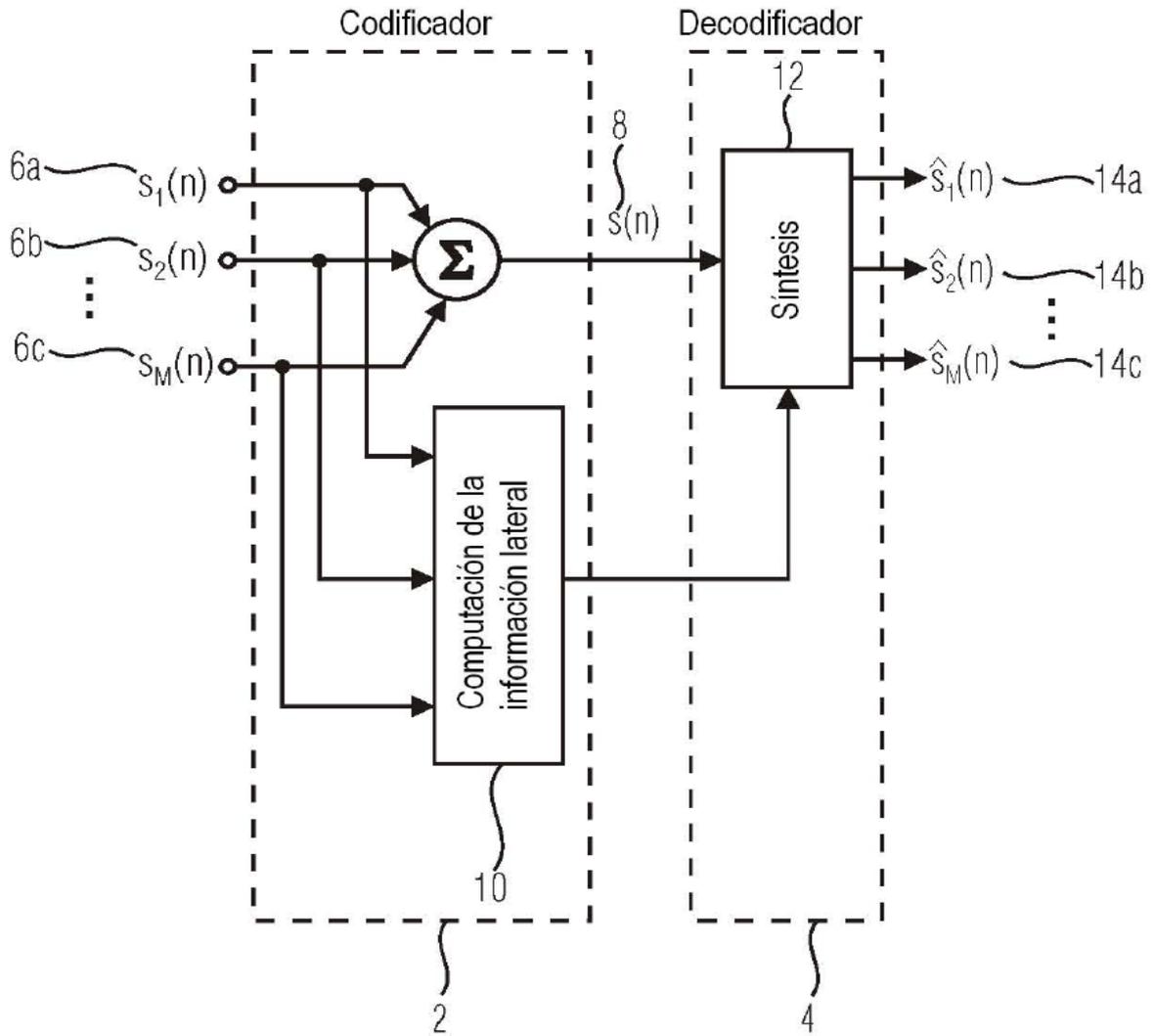


FIG 1

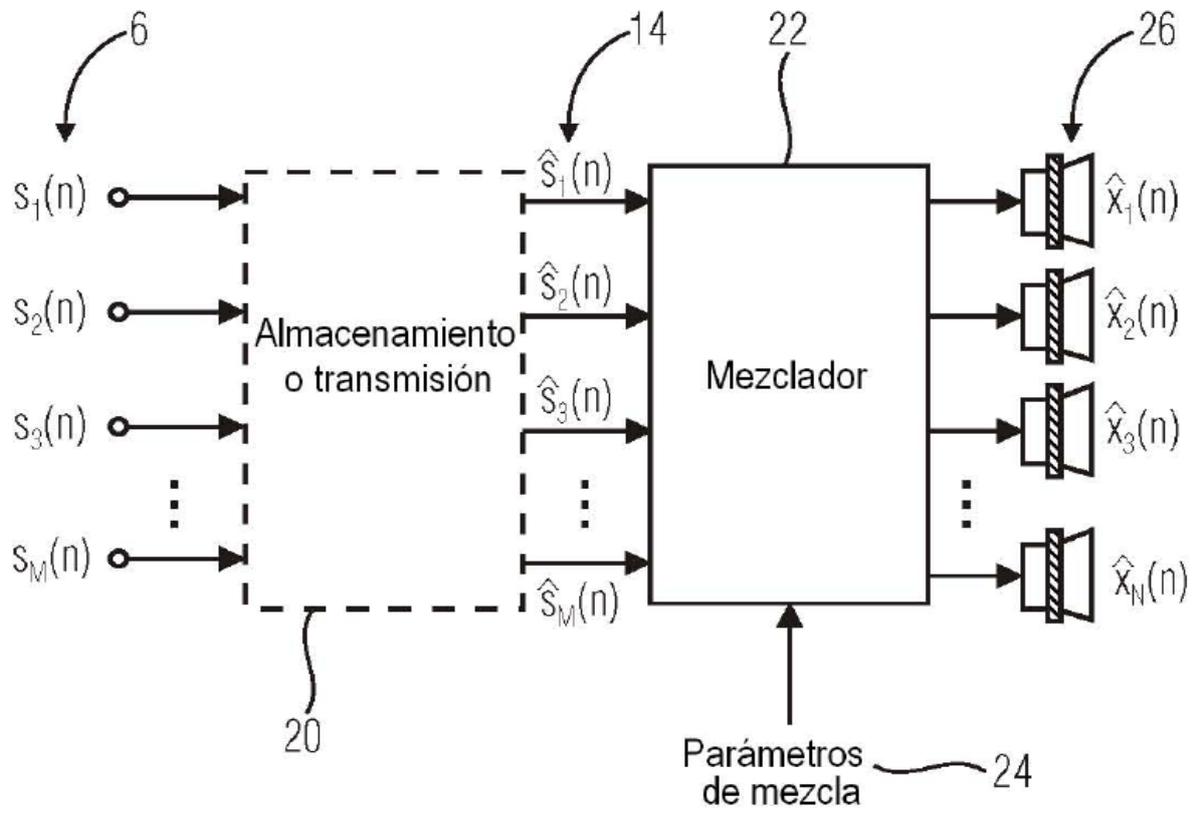


FIG 2

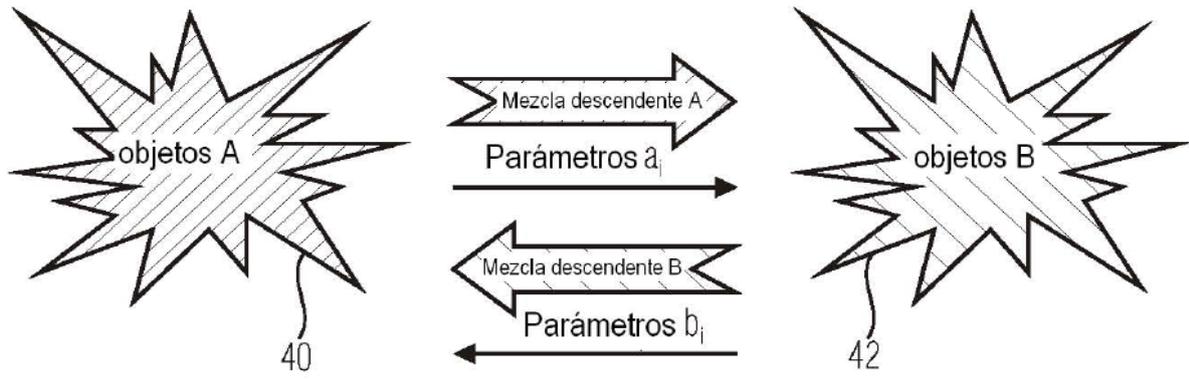


FIG 3

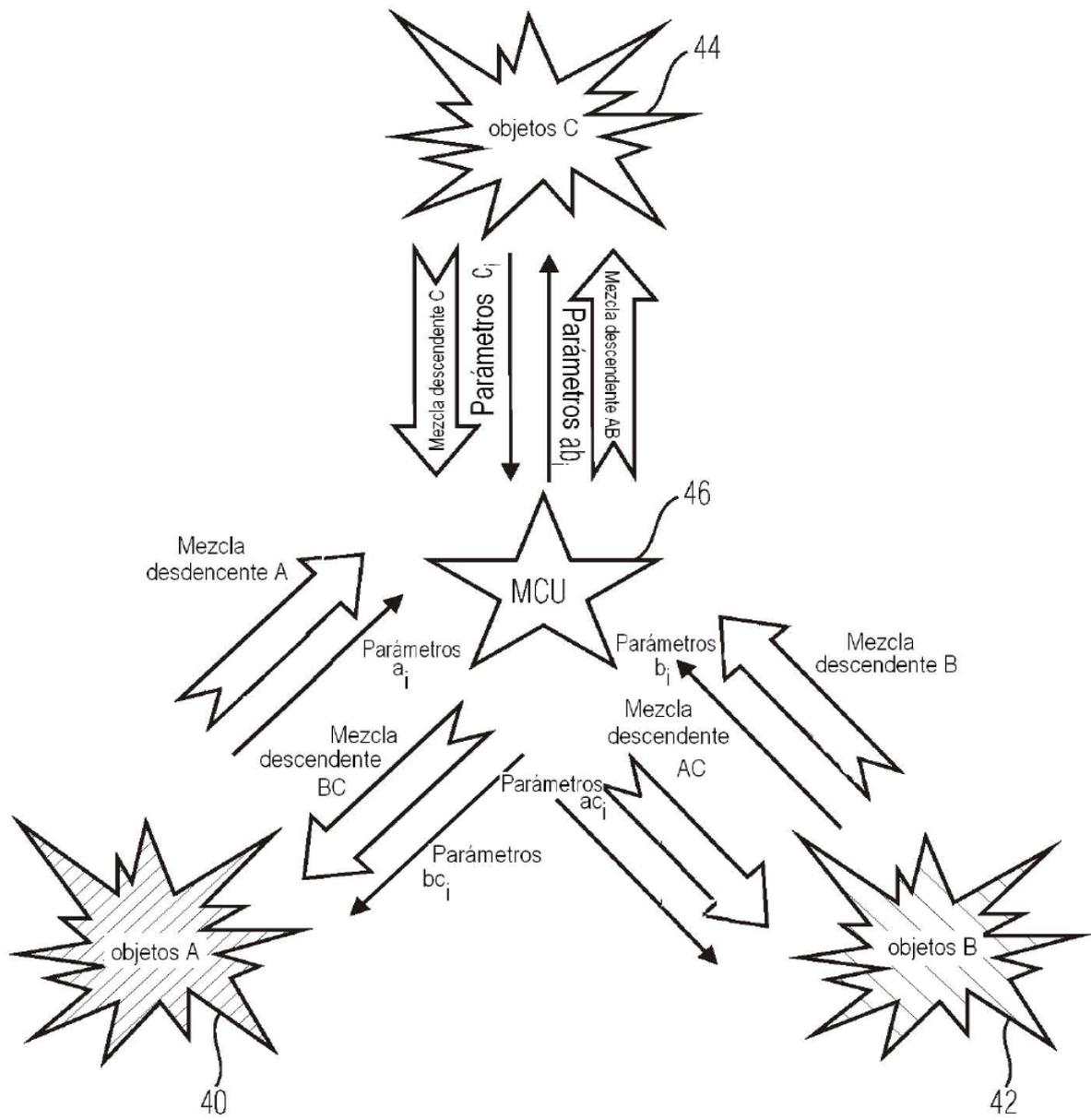


FIG. 4

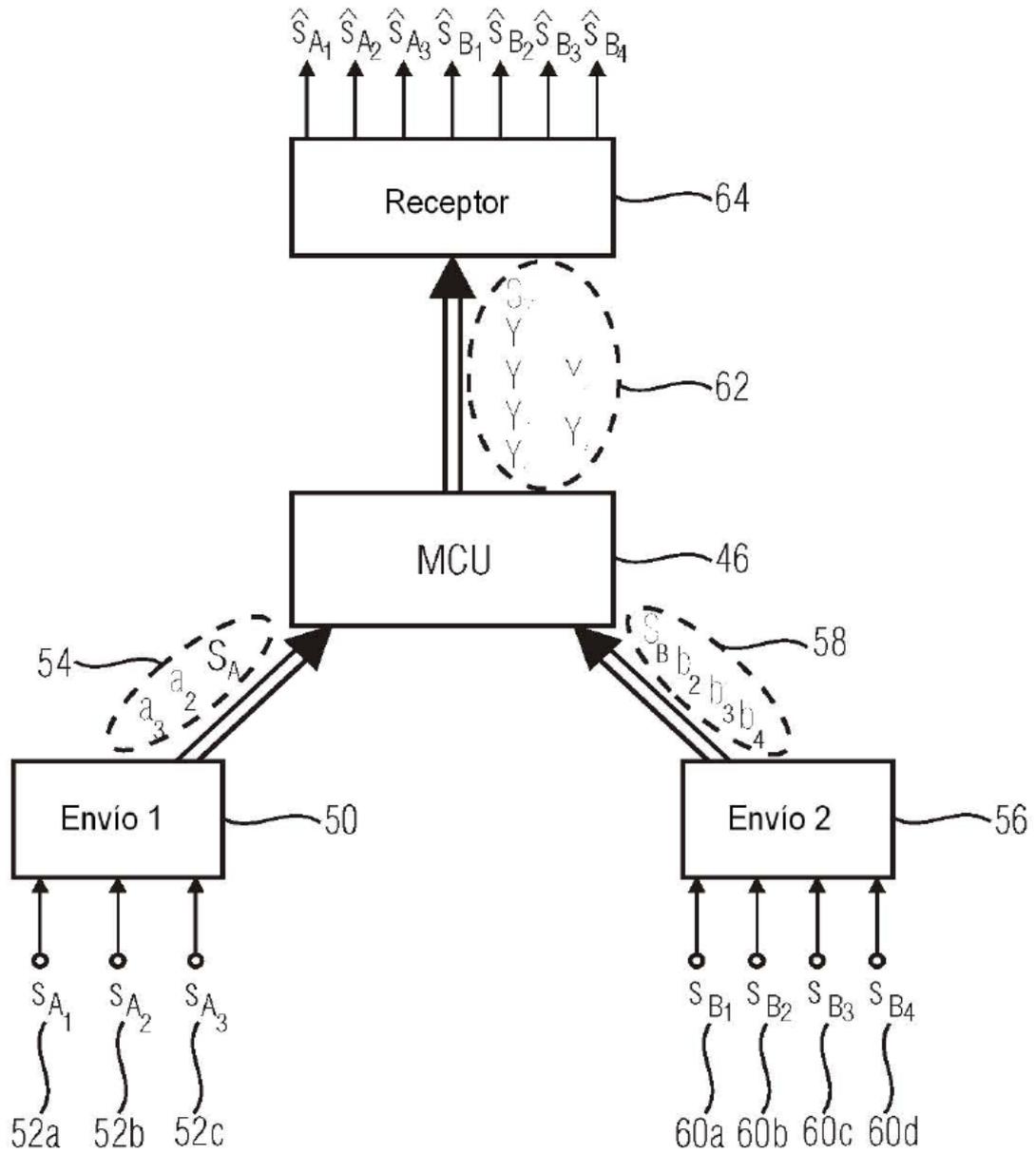


FIG 5

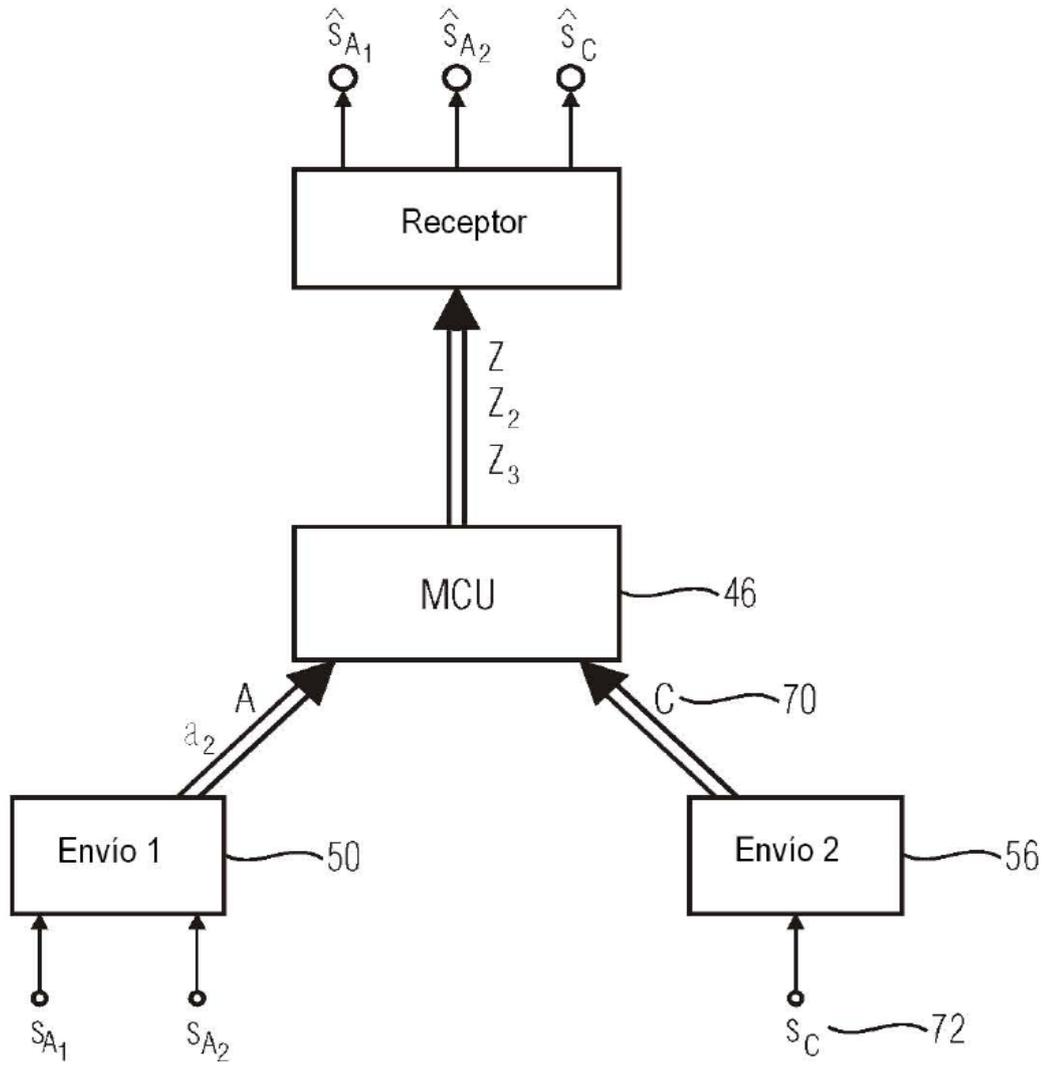


FIG 6

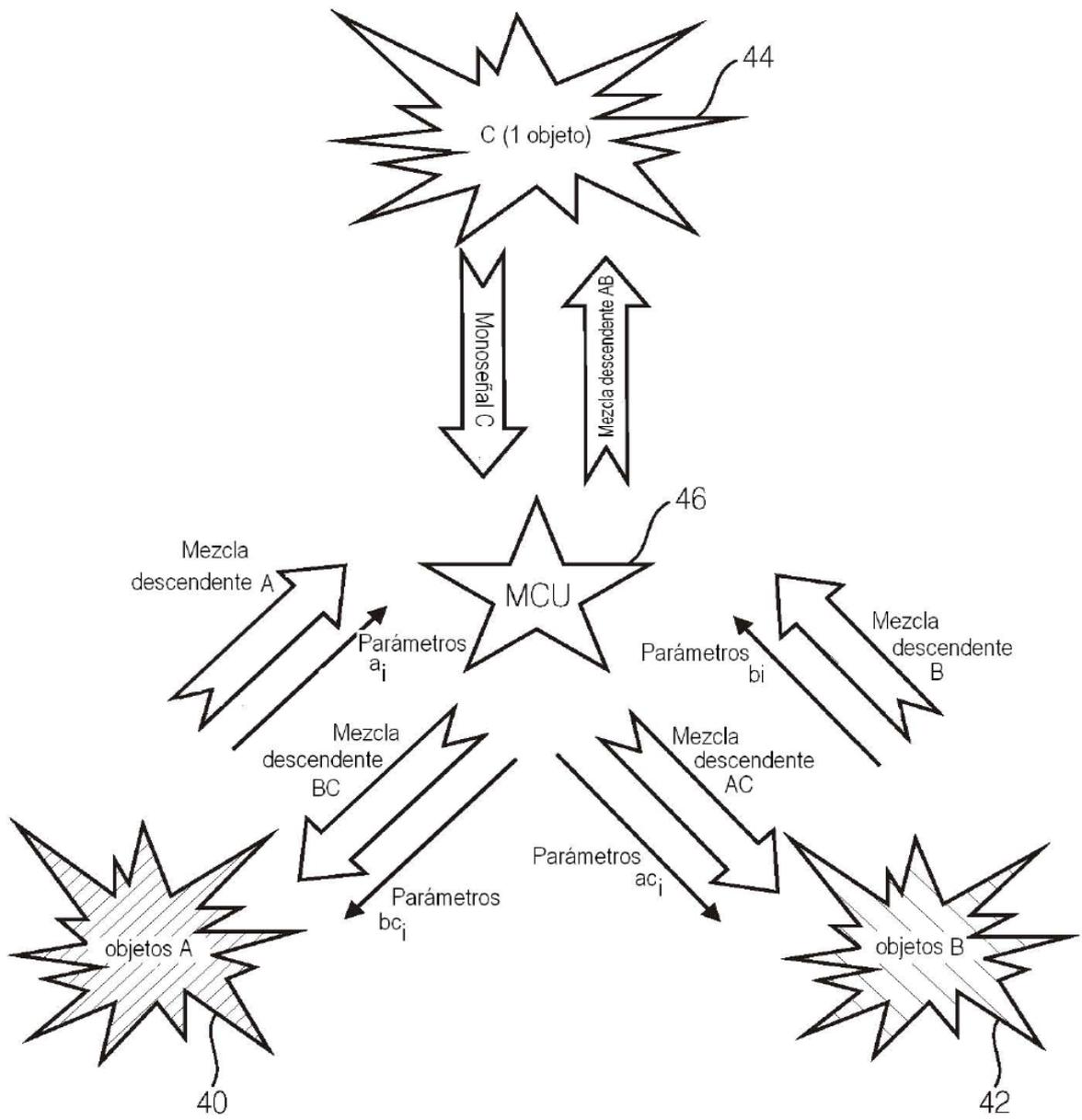


FIG 6B

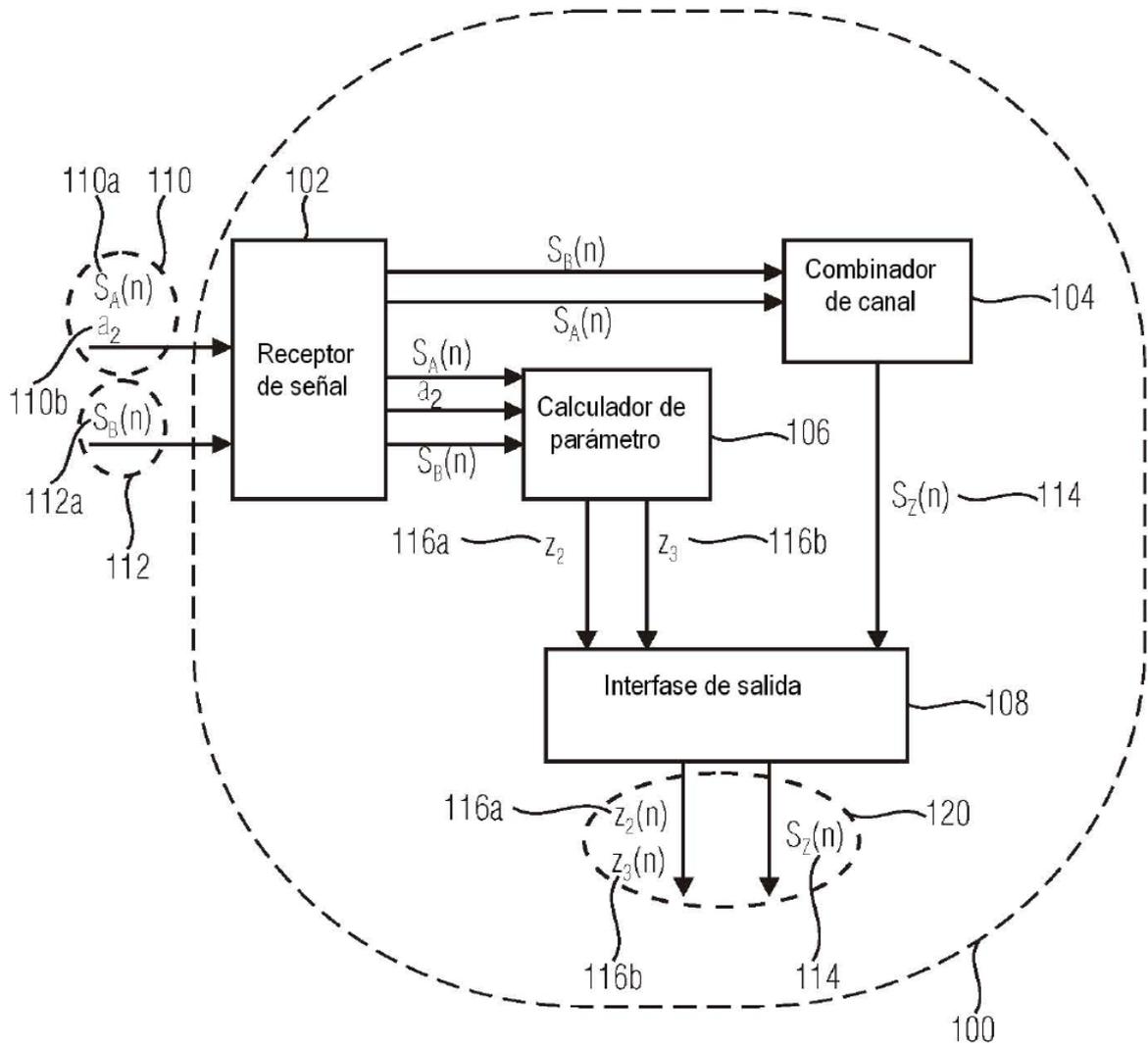


FIG 7