

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 380 059**

51 Int. Cl.:
G10L 19/00 (2006.01)
H04N 7/15 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 96 Número de solicitud europea: **07724528 .0**
96 Fecha de presentación: **24.04.2007**
97 Número de publicación de la solicitud: **2038878**
97 Fecha de publicación de la solicitud: **25.03.2009**

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

30 Prioridad:
07.07.2006 US 819419 P
24.04.2007 US 739544

45 Fecha de publicación de la mención BOPI:
08.05.2012

45 Fecha de la publicación del folleto de la patente:
08.05.2012

73 Titular/es:
**Fraunhofer-Gesellschaft zur Förderung der
angewandten Forschung e.V.**
Hansastraße 27c
80686 München, DE

72 Inventor/es:
HILPERT, Johannes;
HERRE, Juergen;
LINZMEIER, Karsten;
HELLMUTH, Oliver y
KASTNER, Thorsten

74 Agente/Representante:
Arizti Acha, Monica

ES 2 380 059 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 y método para combinar múltiples fuentes de audio codificadas paramétricamente.

Campo de la invención

5 La presente invención se refiere a codificación de audio multicanal y, en particular, a un concepto para combinar flujos de audio codificados paramétricamente de una manera flexible y eficaz.

Antecedentes de la invención y técnica anterior

10 El reciente desarrollo en el área de la codificación de audio ha dado lugar a varias técnicas de codificación de audio paramétrica para codificar conjuntamente una señal de audio multicanal (por ejemplo, 5.1 canales) en uno (o más) canales de mezcla descendente más un flujo de información lateral. En general, el flujo 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 mezcla descendente. La definición particular de los parámetros del canal de referencia al que 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”.

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

20 C. Faller y F. Baumgarte, “Efficient representation of spatial audio using perceptual parametrization”, IEEE WASPAA, Mohonk, NY, octubre de 2001; F. Baumgarte y C. Faller, “Estimation of auditory spatial cues for binaural cue coding”, ICASSP, Orlando, FL, mayo de 2002; C. Faller y F. Baumgarte, “Binaural cue coding: a novel and efficient representation of spatial audio”, ICASSP, Orlando, FL, mayo de 2002; C. Faller y F. Baumgarte, “Binaural cue coding applied to audio compression with flexible rendering”, AES 113th Convention, Los Ángeles, preimpresión 5686, octubre de 2002; C. Faller y F. Baumgarte, “Binaural cue coding – Part II: Schemes and applications”, IEEE Trans. On Speech and Audio Proc., vol., 11, n.º 6, noviembre de 2003, y J. Herre, C. Faller *et al.*, “Spatial Audio Coding: Next-generation efficient and compatible coding of multi-channel audio”, Audio Engineering Society Convention Paper, octubre de 28, 2004, San Francisco, CA, EE.UU.

25 Aunque la codificación de indicación binaural utiliza múltiples canales originales, el estéreo paramétrico es una técnica relacionada para la codificación paramétrica de una señal estéreo de dos canales que da como resultado una señal mono transmitida e información lateral de parámetro, como se considera, por ejemplo, en las siguientes publicaciones: J. Breebaart, S. van de Par, A. Kohlrausch, E. Schuijers, “High-Quality Parametric Spatial Audio Coding at Low Bitrates”, AES 116th Convention, Berlín, preimpresión 6072, mayo de 2004; E. Schuijers, J. Breebaart, H. Purnhagen, J. Engdegard, “Low Complexity Parametric Stereo Coding”, AES 116th Convention, Berlín, preimpresión 6073, mayo de 2004.

30 Otras tecnologías se basan en la multiplexación de números arbitrarios de fuentes u objetos de audio en un solo canal de audio de transmisión. Los esquemas basados en multiplexación se introducen, por ejemplo, como “presentación flexible” en publicaciones relacionadas con BCC (codificación de indicación binaural) o, más recientemente, mediante un esquema llamado “codificación de fuente conjunta” (JSC). Las publicaciones relacionadas son, por ejemplo: C. Faller, “Parametric Joint Coding of Audio Sources”, Convention Paper 6752, 120th AES Convention, París, mayo de 2006. De manera similar a los esquemas de codificación de indicación binaural y estéreo paramétrica, estas técnicas están previstas para codificar múltiples objetos (canales) de audio originales para la transmisión por menos canales de mezcla descendente. Mediante la derivación adicional de los parámetros basados en objetos para cada canal de entrada, que pueden codificarse a una tasa de transmisión de datos muy baja y que también se transmiten a un receptor, estos objetos pueden separarse en el lado del receptor y presentarse (mezclarse) a un cierto número de dispositivos de salida, como por ejemplo, auriculares, altavoces estéreo de dos canales o instalaciones de altavoz multicanal. Este procedimiento permite un ajuste de nivel y redistribución (panorámica) de los diferentes objetos de audio hacia diferentes ubicaciones en la instalación de reproducción, es decir, en el lado del receptor.

35 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 mezcla descendente transmitidos, normalmente k es ≤ 2 . N es el número de canales de audio en la salida de presentación, es decir, por ejemplo, el número de altavoces. Es decir, N = 2 para una presentación estéreo o N = 6 para una instalación de altavoz multicanal 5.1. En términos de eficiencia de compresión, los valores típicos son, por ejemplo, 64 kbps o menos para un canal de mezcla descendente codificado de manera perceptiva (que consiste en k canales de audio) y aproximadamente 3 kbps para parámetros de objeto por objeto de audio transmitido.

40 Los escenarios de aplicación para las técnicas anteriores son, por ejemplo, la codificación de escenas de audio espacial relacionadas con producciones de cine-películas para permitir una reproducción espacial del sonido en un sistema de cine en casa. Ejemplos comunes son las pistas de sonido envolvente 5.1 y 7.1 ampliamente conocidas en un medio de película tal como DVD y similar. Las producciones de películas se están haciendo cada vez más complejas con respecto

a las escenas de audio, previstas para proporcionar una experiencia de escucha espacial y que por tanto tienen que mezclarse con mucho cuidado. A diferentes ingenieros de sonido se les puede encargar la mezcla de diferentes fuentes envolventes o efectos de sonido y, en consecuencia, es deseable la transmisión de escenarios multicanal codificados paramétricamente entre los ingenieros de sonido individuales para transportar los flujos de audio de los ingenieros de sonido individuales de manera eficaz.

Otro escenario de aplicación para tal tecnología es la teleconferencia con múltiples hablantes en cada extremo de una conexión punto a punto tal como se describe en el documento US2005/0062843. Para ahorrar ancho de banda, la mayoría de las instalaciones de teleconferencia funcionan con transmisión monofónica. Utilizando, por ejemplo, codificación de fuente conjunta o una de las otras técnicas de codificación multicanal para la transmisión, puede lograrse una redistribución y alineación de nivel de los diferentes hablantes en el extremo receptor (cada extremo) y por tanto se mejora la inteligibilidad y el equilibrio de los hablantes gastando una tasa de transmisión de bits ligeramente aumentada en comparación con un sistema monofónico. La ventaja de la inteligibilidad aumentada se hace particularmente evidente en el caso especial de asignar a cada individuo participante de la conferencia un solo canal (y por tanto hablante) de una instalación de hablante multicanal en un extremo receptor. Sin embargo, éste es un caso especial. En general, el número de participantes no igualará al número de hablantes en el extremo receptor. Sin embargo, utilizando la instalación de hablante existente es posible presentar la señal asociada con cada participante de manera que parezca que se origina desde cualquier posición deseada. Es decir, el participante individual no sólo se reconoce por su voz diferente, sino también por la ubicación de la fuente de audio relacionada con el participante hablante.

Aunque las técnicas del estado de la técnica implementan conceptos acerca de cómo codificar de manera eficaz múltiples canales u objetos de audio, todas las técnicas actualmente conocidas carecen de la posibilidad de combinar dos o más de estos flujos de audio transmitidos de manera eficaz para derivar un flujo de salida (señal de salida), que es una representación de todos los flujos de audio de entrada (señales de audio de entrada).

El problema surge, por ejemplo, cuando se considera un escenario de teleconferencia con más de dos ubicaciones, teniendo cada ubicación uno o más hablantes. 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 teleconferencia que tenga solamente la información de las ubicaciones de teleconferencia 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.

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 escucha generada por ambas señales. Entonces, puede ser deseable combinar directamente dos señales multicanal codificadas para verificar la impresión de escucha combinada. Es decir, es necesario que una señal combinada sea tal que se asemeje a todos los objetos (fuentes) de audio de los dos ingenieros de audio.

Sin embargo, según las técnicas de la técnica anterior, tal combinación es sólo posible decodificando las señales (flujos) de audio. Entonces, las señales de audio decodificadas pueden volver a codificarse de nuevo mediante codificadores multicanal de la técnica anterior para generar una señal combinada en la que todos los canales de audio u objetos de audio originales se representan apropiadamente.

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 decodificación y nueva codificación de audio subsiguiente puede ocasionar un retraso considerable debido a las dos etapas de procesamiento, lo cual es inaceptable para determinadas aplicaciones, tales como teleconferencia/telecomunicación.

Sumario de la invención

El objetivo de la presente invención es proporcionar un concepto para combinar de manera eficaz múltiples señales de audio codificadas paramétricamente.

Según un primer aspecto de la presente invención, este objetivo se logra mediante un generador de señal de audio según 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 mezcla descendente 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 mezcla descendente que tiene información sobre al menos un segundo canal original; un combinador de canales para derivar un canal de mezcla descendente combinado combinando el primer canal de mezcla descendente y el segundo canal de mezcla descendente; 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 mezcla descendente combinado, los parámetros combinados primero y segundo.

5 Según un segundo aspecto de la presente invención, este objetivo se logra mediante un método para generar una señal de salida de audio según la reivindicación 15, comprendiendo el método: recibir, según la reivindicación 15, una primera
 10 señal de audio que comprende un primer canal de mezcla descendente 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
 15 señal de audio que comprende un segundo canal de mezcla descendente que tiene información sobre al menos un segundo canal original; derivar un canal de mezcla descendente combinado combinando el primer canal de mezcla descendente y el segundo canal de mezcla descendente; derivar un primer parámetro combinado que describe la
 20 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 a un canal de referencia común; y producir la señal de salida de audio que comprende el canal de mezcla descendente combinado y los parámetros combinados primero y segundo.

15 Según un tercer aspecto de la presente invención, este objetivo se logra mediante un programa informático que implementa un método para generar una señal de salida de audio según la reivindicación 17, comprendiendo el método: recibir una primera señal de audio que comprende un primer canal de mezcla descendente que tiene información sobre
 20 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 mezcla descendente que tiene información sobre al menos un segundo canal original; derivar un canal de mezcla descendente combinado combinando
 25 el primer canal de mezcla descendente y el segundo canal de mezcla descendente; 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 a un canal de referencia común; y producir la señal de salida de audio que comprende el canal de mezcla descendente combinado y los parámetros combinados primero y segundo.

Según 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
 30 audio para recibir una primera señal de audio que comprende un primer canal de mezcla descendente 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 mezcla descendente que tiene información sobre al menos un segundo canal original; un combinador de canal para derivar un canal de mezcla descendente combinado combinando el primer canal de mezcla descendente y el segundo canal de mezcla
 35 descendente; 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 mezcla descendente combinado, los parámetros combinados primero y segundo.

40 La presente invención se basa en el descubrimiento de que múltiples señales de audio codificadas paramétricamente pueden combinarse de manera eficaz utilizando un generador de señal de audio o un combinador de señal de audio, que genera una señal de salida de audio combinando los canales de mezcla descendente y los parámetros asociados de las señales de entrada de audio directamente dentro del dominio de parámetro, es decir, 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
 45 específicos, esto se logra mezclando directamente los canales de mezcla descendente 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 mezcla descendente se logra mediante simples operaciones aritméticas económicas desde el punto de vista computacional, tales como sumas.

50 Lo mismo se aplica para la combinación de los parámetros que asocian los canales de mezcla descendente. Dado que en general al menos un subconjunto 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 que, por tanto, no necesiten energía computacional significativa ni incurrir en retrasos adicionales, por ejemplo, utilizando bancos de filtros u otras operaciones que impliquen memoria.

55 Según 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, codificándose ambas paramétricamente. Para generar la señal de salida de audio, el generador de señal de audio de la invención extrae los canales de mezcla descendente de las señales de audio de entrada y genera un canal de mezcla descendente combinado formando una combinación lineal de los dos canales de mezcla descendente. Es decir, los canales individuales se agregan aplicando pesos adicionales.

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 base para el cálculo.

5 En una realización preferida adicional, el cálculo del peso se lleva a cabo suponiendo 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.

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

En una realización preferida adicional de la presente invención, el esfuerzo computacional se disminuye adicionalmente alterando solamente los parámetros asociados a una de las dos señales de audio. 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.

15 En los siguientes párrafos, se detallará el concepto de la invención principalmente para un esquema de codificación que utiliza codificación de fuente conjunta (JSC). En este sentido, la presente invención extiende esta tecnología para conectar múltiples transceptores monofónicos o habilitados para JSC a estaciones remotas mezclando señales de mezcla descendente de JSC e información de objeto dentro del dominio de parámetro. Como han mostrado las consideraciones anteriores, el concepto de la invención no se limita en modo alguno al uso de codificación JSC, sino que también podría implementarse con codificación BCC u otros esquemas de codificación multicanal, tales como codificación MPEG de audio espacial (MPEG envolvente) y similar.

20 Dado que el concepto de la invención se detallará principalmente utilizando codificación JSC, la codificación JSC se comentará brevemente dentro de los siguientes párrafos a fin de destacar más claramente la flexibilidad del concepto de la invención y las mejoras que pueden lograrse frente a la técnica anterior al aplicar el concepto de la invención a esquemas existentes de codificación de audio multicanal.

25 Breve descripción de los dibujos

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

la figura 2 muestra un ejemplo de un presentador de JSC;

la figura 3 muestra un escenario de teleconferencia con dos ubicaciones;

30 la figura 4 muestra un escenario de teleconferencia con tres ubicaciones;

la figura 5 muestra un ejemplo de teleconferencia utilizando un generador de señal de audio de la invención;

la figura 6 muestra un ejemplo adicional de teleconferencia utilizando un generador de señal de audio de la invención;

la figura 6b muestra la retrocompatibilidad del concepto de la invención; y

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

35 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 los mismos números 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 limitar la funcionalidad y sin limitar el alcance de la presente invención.

40 La figura 1 muestra un diagrama de bloques del esquema de codificación de fuente conjunta, un codificador 2 correspondiente y un decodificador 4 correspondiente.

El codificador 2 recibe entradas 6a, 6b, y 6c de audio discretas $s_i(n)$, y crea una señal 8 de mezcla descendente $s(n)$, por ejemplo, mediante la suma de las formas de onda.

45 Adicionalmente, un extractor 10 de parámetro dentro del codificador 2 extrae información lateral paramétrica para cada objeto individual (señal 6a, 6b y 6c). Aunque no se muestra en la figura 1, la señal 8 de mezcla descendente puede comprimirse adicionalmente mediante un codificador de voz o audio y se transmite con la información lateral paramétrica adyacente al decodificador 4 JSC. Un módulo 12 de síntesis dentro del decodificador 4 regenera los estimados 14a, 14b y 14c ($\hat{S}_i(n)$) de los objetos de entrada (canales 6a, 6b y 6c).

A fin de reconstruir los estimados 14a, 14b y 14c, que desde el punto de vista perceptivo son similares a los objetos 6a, 6b y 6c de entrada discretos (canales de entrada), tiene que extraerse información lateral paramétrica apropiada para

cada canal. Dado que se suman los canales individuales para la generación de la señal 8 de mezcla descendente, las proporciones de energía entre los canales son tales cantidades adecuadas. En consecuencia, la información paramétrica para los diferentes objetos o canales consiste en proporciones de energía Δp de cada objeto en relación con el primer objeto (objeto de referencia).

5 Esta información se deriva en el dominio de frecuencia en bandas (subbandas) de frecuencia no separadas igualmente 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", The MIT Press, Cambridge, MA, edición revisada 1997.

10 Es decir, los canales de audio de entrada de banda ancha se filtran para dar 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 anteriormente, la energía de banda del primer objeto (objeto de referencia o canal de referencia) actúa como valor de referencia.

$$\Delta p_i(n) = 10 \log_{10} \frac{E\{s_i^2(n)\}}{E\{s_1^2(n)\}}, \quad i = 2 \dots M \quad \text{Ecuación 1}$$

15 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 subbanda. La proporción de energía puede cuantificarse adicionalmente antes de someterse para ahorrar adicionalmente ancho de banda de transmisión.

20 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 ($\hat{s}_i(n)$) es igual a la energía de la señal de mezcla descendente $s(n)$. En términos de una expresión matemática, esto significa:

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

25 Basándose en esta presunción y ecuación, las energías de subbanda para el primer objeto (el objeto de referencia o canal de referencia) pueden reconstruirse, como se describirá adicionalmente a continuación al detallar el concepto de la invención.

30 Para resumir, una señal de audio o flujo de audio según JSC comprende un canal de mezcla descendente y parámetros asociados, describiendo los parámetros proporciones de energía de canales originales con respecto a un canal de referencia original. Puede observarse que este escenario puede alterarse fácilmente seleccionando otros canales para ser el canal de referencia. Por ejemplo, el propio canal de mezcla descendente 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 mezcla descendente. También puede seleccionarse el canal de referencia para ser variable, seleccionando el canal que tiene la mayor energía para ser el canal de referencia. Por tanto, 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 normalmente de manera selectiva en frecuencia, el canal de referencia puede ser diferente para diferentes bandas de frecuencia.

40 La figura 2 muestra un esquema adicional mejorado de la codificación JSC, basado en el esquema de la figura 1. Las características detalladas con respecto a la figura 1 se incluyen en el cuadro 20 de almacenamiento o transmisión, que recibe los canales 6 de entrada que van a codificarse y que produce los estimados 14 de los canales 6 de entrada. El esquema de la figura 2 se mejora porque comprende además un mezclador 22 que recibe los estimados. Es decir, los objetos 14 sintetizados no se producen como señales de audio únicas directamente, sino que se presentan a N canales de salida en el módulo de mezclador. Tal mezclador puede implementarse de diferentes maneras, por ejemplo, recibiendo los parámetros 24 de mezclado adicionales como entrada, para dirigir el mezclado de los objetos 14 sintetizados. Sólo como ejemplo, puede considerarse un escenario de teleconferencia, en el que a cada uno de los canales 26 de salida se le atribuye 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 sólo la voz puede servir como criterio para distinguir entre 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 teleconferencia se agrupen en la misma dirección, mejorando aún más la experiencia perceptiva.

Como se muestra en la figura 2, $s_1(n) \dots s_M(n)$ denotan los objetos de audio discretos en la entrada del codificador JSC.

En la salida del decodificador JSC $\hat{s}_1(n) \dots \hat{s}_M(n)$ representan los objetos de audio separados virtualmente que se alimentan al mezclador. Los parámetros 24 de mezclado 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 $\hat{x}_1(n) \dots \hat{x}_N(n)$.

La figura 3 muestra la aplicación de esquemas de codificación de audio multicanal a un escenario básico de teleconferencia, que tiene lugar entre dos ubicaciones. En este caso, una primera ubicación 40 se comunica con una segunda ubicación 42. La primera ubicación puede tener A participantes, es decir, A objetos de audio, la segunda ubicación tiene B participantes u objetos de audio. Para una teleconferencia punto a punto, la tecnología descrita de codificación JSC puede aplicarse directamente para transmitir señales de audio de múltiples objetos en cada ubicación a la estación remota correspondiente. Es decir, se transfieren (A-1) parámetros a_i y una mezcla descendente asociada a la ubicación 42. En la dirección opuesta se transmiten (B-1) parámetros b_i conjuntamente con una mezcla descendente asociada a la ubicación 40.

Para una teleconferencia con más de dos puntos de extremo, la situación es completamente diferente, tal como se ilustra en la figura 4.

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 denominado unidad de control de puntos múltiples, MCU. Cada una de las ubicaciones (sitios) 40, 42 y 44 se conecta a la MCU 46. Para cada sitio 40, 42 y 44, existe un único flujo ascendente a la MCU que contiene la señal desde el sitio. Dado que cada sitio individual necesita recibir las señales desde los sitios restantes, el flujo descendente para cada sitio 40, 42 y 44 es una mezcla de las señales de los otros sitios, excluyendo la señal del propio sitio, a la que también se hace referencia como 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 N-1 flujos con codificación JSC desde la MCU a cada sitio. Esto, por supuesto, sería la opción directa.

El enfoque del estado de la técnica para derivar los flujos descendentes individuales es resintetizar todos los flujos entrantes (objetos) dentro de la MCU 46 utilizando un decodificador JSC. Entonces, los objetos de audio resintetizados podrían reagruparse y recodificarse a fin de proporcionar a cada sitio flujos de audio que comprendan los objetos de audio o los canales de audio deseados. Incluso dentro de este escenario simple, esto significaría tres tareas de decodificación y tres de codificación, que deben llevarse a cabo simultáneamente dentro de la MCU 46. A pesar de las demandas computacionales significativas, pueden esperarse adicionalmente artefactos audibles por este proceso paramétrico de "codificación en tándem" (codificación/decodificación repetida). El incremento del número de sitios incrementaría adicionalmente el número de flujos y, por tanto, el número de procesos de codificación o decodificación requeridos, haciendo que ninguno de los enfoques directos sea posible para los escenarios en tiempo real.

Según la presente invención, en consecuencia, se desarrolla un esquema para mezclar diferentes flujos codificados paramétricamente (flujos JSC en este ejemplo particular) directamente dentro del dominio de parámetro de objeto y mezcla descendente para tal escenario tipo MCU, creando las señales de salida deseadas (flujos de audio de salida) con un esfuerzo computacional y pérdida de calidad mínimos.

En los siguientes párrafos, se detalla el concepto de la invención de mezclar directamente flujos de audio multicanal codificados paramétricamente dentro del dominio de parámetro para flujos de audio codificados por JSC.

El concepto de la invención se explica con la combinación de dos señales (flujos) de audio originales para dar una señal de salida. La unión de tres o más flujos entre sí puede derivarse fácilmente del caso de combinar dos flujos. Las siguientes consideraciones matemáticas se ilustran por la figura 5, que muestra un caso en el que tienen que combinarse tres canales de audio del sitio A con cuatro canales de audio del sitio B. Esto es, por supuesto, solamente un ejemplo para visualizar el concepto de la invención. Cuando se usa codificación JSC, el sitio 50 (A) que tiene tres participantes 52a a 52c (hablantes) de conferencia, que generan señales s_{Ax} , transmite un flujo de audio o señal 54 de audio. La señal 54 de audio tiene un canal de mezcla descendente 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 58 de audio que tiene un canal de mezcla descendente 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 54 y 58 de audio para derivar una señal 62 de salida que tiene un canal de mezcla descendente combinado s_Y y seis parámetros y_2, \dots, y_7 .

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

En términos generales, el objetivo es formar una sola representación 62 combinada de dos flujos 54 y 58 JSC, representando cada uno un número de objetos mediante una señal de mezcla descendente 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 subyacentes a ambos flujos JSC para dar un solo flujo JSC en una etapa.

5 Para que las siguientes ecuaciones sigan siendo simples, se supone 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 determinado objeto i puede derivarse como

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

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

10 Se supone que todas las señales siguientes se descomponen en una representación de subbanda, por tanto, cada uno de los cálculos se aplica para cada subbanda por separado.

Se dispone del flujo A con su señal de mezcla descendente s_A y los parámetros (proporciones de energía relativas) para U objetos $a_2... a_U$. El flujo B consiste en la señal de mezcla descendente s_B y los parámetros para V objetos $b_2...b_V$.

15 La señal de mezcla descendente combinada s_Y puede formarse como una combinación lineal de ambas señales de mezcla descendente s_A y s_B . Para asegurar el correcto nivelado del volumen 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

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

20 Este tipo de ajuste a escala puede ser significativo 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 mezcla descendente.

Alternativamente, podría utilizarse un enfoque de conservación de energía para los factores de ganancia con

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

Otra posibilidad es seleccionar el factor de ganancia de tal manera que ambas señales de mezcla descendente contribuyan con la misma energía promedio para la mezcla descendente combinada, es decir, seleccionando

$$25 \quad \frac{g_B}{g_A} = \sqrt{\frac{E\{s_A^2(n)\}}{E\{s_B^2(n)\}}}$$

Los parámetros de objeto y_i para el flujo combinado s_Y deben representar todos $U + V$ objetos.

30 Dado que los parámetros asociados a los canales de mezcla descendente son proporciones de energía relativas, 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 observarse que solamente tienen que recalcularse los parámetros de un flujo, disminuyendo adicionalmente la carga computacional dentro de una MCU 46.

35 Puede observarse además que en ningún caso es necesario utilizar el canal de referencia de uno de los flujos de audio originales como nuevo canal de referencia. El concepto de la invención de combinar flujos de audio codificados paramétricamente 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 mezcla descendente combinado como nuevo canal de referencia.

Siguiendo este enfoque de utilizar el canal de referencia original del sitio A como nuevo canal de referencia (canal de referencia combinado), la energía (potencia) del primer objeto (canal) de cada señal A y B tiene que calcularse primero, dado que se están disponibles sólo de manera implícita.

5 La conservación de energía para la señal de mezcla descendente A, suponiendo fuentes estadísticamente independientes, proporciona:

$$E\{s_A^2(n)\} = \sum_{i=1}^U E\{s_i^2(n)\}.$$

Las energías de señal $E\{s_{A_2}^2(n)\} \dots E\{s_{A_U}^2(n)\}$ se definen con sus energías relativas $a_2 \dots a_U$ para $E\{s_{A_1}^2(n)\}$:

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

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

...

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

Esto lleva a la energía de s_{A1} como:

$$E\{s_{A_1}^2(n)\} = \frac{E\{s_A^2(n)\}}{(1 + a_2 + a_3 + \dots + a_U)}$$

10

Aplicando lo mismo para la señal de mezcla descendente s_B , puede calcularse la energía del objeto s_{B1} como:

$$E\{s_{B_1}^2(n)\} = \frac{E\{s_B^2(n)\}}{(1 + b_2 + b_3 + \dots + b_U)}$$

Ahora puede construirse el nuevo conjunto de parámetros para todos los objetos de la señal s_y :

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

15

$$y_2 = a_2$$

$$y_3 = a_3$$

...

$$y_U = a_U$$

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

20

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

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

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

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

...

$$y_{U+v} = b_v \cdot \frac{g_B^2}{g_A^2} \cdot \frac{E\{s_{B_1}^2(n)\}}{E\{s_{A_1}^2(n)\}}$$

5

Como han mostrado los párrafos anteriores, el concepto de la invención permite la generación de un flujo de audio combinado utilizando solamente operaciones aritméticas simples, siendo, por tanto extremadamente eficaz desde el punto de vista computacional. Por tanto, la combinación de múltiples flujos de audio codificados paramétricamente puede llevarse a cabo en tiempo real.

10

Para enfatizar adicionalmente la gran flexibilidad del concepto de la invención, la figura 6 muestra cómo una señal 70 monofónica, provocada por un solo hablante en el sitio 56 puede combinarse según 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 sistemas de teleconferencia arbitrarios pueden combinarse según la invención con fuentes multicanal (multiobjeto) codificadas paramétricamente para generar una señal de audio codificada por JSC que representa todos los canales (objetos) de audio originales.

15

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, por ejemplo, desde un dispositivo de conferencia legado en el flujo basado en objeto.

El ejemplo anterior con el flujo A JSC (mezcla descendente s_A , parámetros $a_2...a_U$) y un objeto monofónico C (mezcla descendente s_C) conduce a una señal combinada Z con la señal de mezcla descendente

$$s_Z = g_A \cdot s_A + g_C \cdot s_C$$

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

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

20

$$y_2 = a_2$$

$$y_3 = a_3$$

...

$$y_U = a_U$$

$$y_{U+1} = \frac{g_C^2}{g_A^2} \cdot \frac{E\{s_C^2(n)\}}{E\{s_{A_1}^2(n)\}},$$

25

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

El ejemplo mencionado anteriormente para transcodificar/fusionar dos flujos 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.

30

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 multicanal de la figura 4 y además muestra la facilidad con la que un codificador de audio monofónico de la técnica anterior presente en la fuente de audio C(44) puede integrarse en una conferencia de audio multicanal utilizando la MCU 46 de la invención.

35

Como se mencionó previamente, el concepto de la invención no se restringe a la codificación JSC que tiene 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 el canal que tiene la mayor parte de energía dentro de un intervalo de tiempo predeterminado dado.

En lugar de normalizar los valores de energía de señal de banda para dar 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 con la máxima energía sobre todos los objetos en una cierta banda de frecuencia:

$$pnorm_i(n) = \frac{E\{s_i^2(n)\}}{\max_i(E\{s_i^2(n)\})}, \quad i = 1 \dots M \quad \text{Ecuación 4}$$

5 Estos valores de energía normalizados (que se proporcionan en una representación lineal) no necesitan ninguna limitación adicional para un determinado 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 que ya no se conoce a priori.

El proceso de mezclado para este escenario incluiría las siguientes etapas (que de nuevo tienen que llevarse a cabo para cada subbanda por separado):

10 se dispone de un flujo A con su señal de mezcla descendente s_A y los parámetros (valores de energía normalizados, ecuación 3, ecuación 1) para U objetos $a_1 \dots a_U$.

El flujo B consiste en la señal de mezcla descendente s_B y los parámetros para V objetos $b_1 \dots b_V$.

Una señal de mezcla descendente combinada puede formarse según una de las opciones ya mostradas:

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

15 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. Hay dos candidatos para ser este "objeto máximo" de Y, bien el objeto máximo de A o bien el objeto máximo de B, ambos pueden identificarse teniendo una proporción de energía normalizada de "1".

20 Esta decisión puede realizarse comparando la energía absoluta de ambos candidatos. De nuevo puede utilizarse la relación con respecto a la energía de las señales de mezcla descendente (ecuación 2) para obtener:

$$E\{s_{A_{\max}}^2(n)\} = \frac{E\{s_A^2(n)\}}{\sum_{i=1}^U a_i} \quad \text{y} \quad E\{s_{B_{\max}}^2(n)\} = \frac{E\{s_B^2(n)\}}{\sum_{i=1}^V b_i}$$

Ahora, se puede comparar las energías máximas de objeto ponderadas con los factores de ganancia del proceso de mezcla descendente:

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

25 Independientemente del objeto cuya energía sea más alta, este objeto servirá como "objeto máximo" para los parámetros combinados y_i .

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

$$Y_1 = a_1$$

30 $Y_2 = a_2$

$$y_U = a_U$$

$$y_{U+1} = b_1 \frac{g_B^2}{g_A^2} \cdot \frac{E\{s_{B_{\max}}^2(n)\}}{E\{s_{A_{\max}}^2(n)\}},$$

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

$$y_{U+2} = b_2 \cdot \frac{g_B^2}{g_A^2} \cdot \frac{E\{s_{B_{\max}}^2(n)\}}{E\{s_{A_{\max}}^2(n)\}}$$

...

$$y_{U+V} = b_V \cdot \frac{g_B^2}{g_A^2} \cdot \frac{E\{s_{B_{\max}}^2(n)\}}{E\{s_{A_{\max}}^2(n)\}}$$

Para este ejemplo, todos los parámetros para los objetos de A pueden permanecer sin cambio, dado que la señal A llevó el objeto máximo total.

5 También en esta representación, la inserción de un objeto monofónico puede efectuarse de manera correspondiente, por ejemplo, suponiendo que $V = 1$.

Generalmente, el proceso de transcodificación se lleva a cabo de tal manera que su resultado se aproxima al resultado que se habría obtenido si todos los objetos originales para ambos flujos se hubieran codificado para dar un solo flujo JSC en primer lugar.

10 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.

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

15 El receptor 102 de señal de audio recibe una primera señal 110 de audio que comprende un primer canal 110a de mezcla descendente que tiene información sobre dos o más primeros canales originales y que comprende un parámetro 110b original asociado a 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 102 de señal de audio recibe además una segunda señal 112 de audio que comprende un segundo canal 112a de mezcla descendente que tiene información sobre al menos un segundo canal original.

20 El receptor de señal de audio emite el primer canal 110a de mezcla descendente y el segundo canal 112a de mezcla descendente a una entrada del combinador 104 de canales y el primer canal 110a de mezcla descendente, el segundo canal 112a de mezcla descendente y el parámetro 110b original al calculador 106 de parámetro.

El combinador 104 de canales deriva un canal 114 de mezcla descendente combinado combinando el primer canal 110a de mezcla descendente y un segundo canal 112b de mezcla descendente, es decir, combinando los canales de mezcla descendente directamente sin reconstruir los canales de audio originales subyacentes.

25 El calculador 106 de parámetro deriva un primer parámetro 116a combinado que describe la propiedad de uno de los primeros canales originales con respecto a un canal de referencia común y un segundo parámetro 116b combinado 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 108 de salida, que además recibe el canal 114 de mezcla descendente combinado desde el combinador 104 de canales. Finalmente, la interfaz de salida emite una señal 120 de salida que comprende el canal 114 de mezcla descendente combinado y los parámetros 116a y 116b combinados primero y segundo.

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 costosas desde el punto de vista computacional.

35 En los párrafos anteriores, se ha mostrado el concepto general de mezclar dos o más señales, basándose cada una en un enfoque paramétrico de JSC. Particularmente, las ecuaciones anteriores muestran cómo aplicar esta técnica para un caso, en el que la información paramétrica consiste en 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 parámetros que describan medidas de amplitud u otras propiedades de canales de audio individuales, tales como correlaciones. Las proporciones de energía también pueden computarse con respecto al canal de mezcla descendente combinado, a expensas de transmitir un parámetro adicional. Por otra parte, es ventajoso en este escenario alternativo la reducida complejidad computacional durante el mezclado de los flujos de audio, dado que la reconstrucción de la energía del canal de referencia, que no se transmite explícitamente en JSC "genérico", está obsoleta.

45 Además, la invención no se limita a un escenario de teleconferencia, sino que puede aplicarse cuando se desee la multiplexación de objetos paramétricos para dar un solo flujo. Este, por ejemplo, puede ser el caso en esquemas de codificación BCC, envolvente espacial MPEG y otros.

5 Como se ha mostrado, el concepto de la invención permite incluso incluir sin interrupción estaciones remotas legadas que proporcionen una sola señal monofónica en el escenario basándose en objeto. Además de la combinación de diferentes flujos de objetos, el concepto de la invención muestra también cómo pueden generarse diferentes formas de representar datos paramétricos, de manera que sean adecuadas para permitir procesos de combinación eficaces desde el punto de vista computacional. Como tal, es una característica ventajosa de una sintaxis de flujo de bits paramétrica de la invención expresar las propiedades del objeto de tal manera que puedan combinarse dos flujos llevando a cabo meramente operaciones simples.

10 En consecuencia, el concepto de la invención también enseña cómo crear flujos de bits apropiados o formatos de flujos de bits para codificar paraméricamente múltiples canales de audio (objetos de audio) originales, adhiriéndose a los siguientes criterios:

- La señal de mezcla descendente combinada se forma simplemente a partir de las señales de mezcla descendente parciales.
- La información lateral paramétrica combinada se forma a partir de la combinación de información lateral paramétrica individual y alguna simple para computar las características de las señales de mezcla descendente (por ejemplo, energía).
- En ningún caso tiene que llevarse a cabo una operación compleja tal como una etapa de decodificación/recodificación para los objetos de audio.

20 En consecuencia, la representación paramétrica que describe los objetos debe seleccionarse de tal manera que sea posible una combinación ("adición") de dos o más flujos de objeto utilizando solamente campos de flujo de bits que se encuentren disponibles como parte de la información lateral paramétrica, y posiblemente simples para computar las métricas de las señales de mezcla descendente (por ejemplo, energía, valor pico).

25 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 para dar un determinado número de etapas de cuantificador o sus índices de cuantificador representativos. La sintaxis de flujo de bits debe permitir el aumento (o disminución) sencillo del número de parámetros de objeto en un flujo, por ejemplo, simplemente concatenando, insertando o eliminando parámetros.

30 Resumiendo, el concepto de la invención permite la combinación más flexible y eficaz desde el punto de vista computacional de los flujos de audio codificados paraméricamente. Debido a la alta eficacia computacional, el concepto de la invención no se restringe a un número máximo de canales para su combinación. Principalmente, los canales, que pueden combinarse en tiempo real, pueden proporcionarse a un generador de señal de audio de la invención en números arbitrarios. La representación paramétrica precisa (JSC) utilizada para ilustrar el concepto de la invención tampoco 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 y el concepto de la invención.

35 Además, las computaciones necesarias no tienen que aplicarse necesariamente en software. También pueden utilizarse implementaciones de hardware utilizando, por ejemplo, DSP, ASIC y otros circuitos integrados para llevar a cabo los cálculos, que aumentarán aún más la velocidad del concepto de la invención, permitiendo la aplicación del concepto de la invención en escenarios en tiempo real.

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

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

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

Además, la invención no se limita a un escenario de teleconferencia, sino que puede aplicarse en cualquier caso en el que sea ventajosa la multiplexación de objetos paramétricos para dar un solo flujo JSC.

Además, esta técnica permite incluir sin interrupción estaciones remotas legadas que proporcionen una sola señal monofónica en el escenario basado en objeto.

50 Además del proceso real de combinar diferentes flujos de objeto, la invención muestra también cómo diferentes formas de representar 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 una sintaxis de flujo de bits

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

5
10
Dependiendo de determinados requerimientos de implementación de los métodos de la invención, los métodos 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 tenga señales de control electrónicamente legibles almacenadas en el mismo, que actúen conjuntamente con un sistema informático programable de tal manera que se lleven a cabo los métodos de la invención. Generalmente, la presente invención es, en consecuencia, un producto de programa informático con un código de programa almacenado en un soporte legible por máquina, siendo el código de programa operativo para llevar a cabo los métodos de la invención cuando el producto de programa informático se ejecuta en un ordenador. En otras palabras, los métodos de la invención, en consecuencia, son un programa informático que tiene un código de programa para llevar a cabo al menos uno de los métodos de la invención cuando el programa informático se ejecuta en un ordenador.

15
Aunque lo anterior se ha mostrado y descrito particularmente con referencia a realizaciones particulares del mismo, los expertos en la técnica entenderán pueden realizarse diversos otros cambios en la forma y los detalles sin apartarse del alcance de las mismas. Se entenderá que pueden realizarse diversos cambios en la adaptación a diferentes realizaciones sin apartarse de los conceptos más amplios dados a conocer en el presente documento y abarcados por las siguientes reivindicaciones.

REIVINDICACIONES

1. Generador (100) de señal de audio para generar una señal de salida de audio, que comprende:
un receptor (102) de señal de audio para recibir:

5 una primera señal (110) de audio que comprende un primer canal (110a) de mezcla descendente que tiene información sobre dos o más primeros canales originales, y que comprende al menos un parámetro (110b) 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 (112) de audio que comprende un segundo canal (112a) de mezcla descendente que tiene información sobre al menos un segundo canal original;

10 un combinador (104) de canales para derivar un canal (114) de mezcla descendente combinado combinando el primer canal (110a) de mezcla descendente y el segundo canal (112a) de mezcla descendente utilizando una combinación lineal del primer canal (110a) de mezcla descendente y el segundo (112b), en el que los coeficientes de la combinación lineal dependen de la energía $E\{s_A^2(n)\}$ dentro del primer canal (110a) de mezcla descendente y de la energía $E\{s_B^2(n)\}$ dentro del segundo canal (112a) de mezcla descendente; y

15 un calculador (106) de parámetro para derivar, utilizando la energía $E\{s_A^2(n)\}$ del primer canal (110a) de mezcla descendente, la energía $E\{s_B^2(n)\}$ del segundo canal (112b) de mezcla descendente y el al menos un parámetro (110b) original,

un primer parámetro (116a) combinado que describe la propiedad de uno de los primeros canales originales con respecto a un canal de referencia común, y

20 un segundo parámetro (116b) combinado que describe o bien la propiedad de otro de los primeros canales originales con respecto al canal de referencia común o bien la propiedad del al menos un segundo canal original con respecto al canal de referencia común; y

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

25 2. Generador (100) de señal de audio según la reivindicación 1, en el que el combinador (104) de canales es operativo para utilizar una combinación lineal que tiene un coeficiente g_A para el primer canal de mezcla descendente, y un coeficiente g_B para el segundo canal de mezcla descendente derivado utilizando la siguiente ecuación:

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

30 3. Generador (100) de señal de audio según las reivindicaciones 1 ó 2, en el que el combinador (104) de canales 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.

35 4. Generador (100) de señal de audio según la reivindicación 3, en el que el combinador (104) de canales es operativo para utilizar una combinación lineal que tiene un coeficiente g_A del primer canal (110a) de mezcla descendente y un coeficiente g_B del segundo canal (112a) de mezcla descendente derivado según una de las siguientes ecuaciones:

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

o

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

40 5. Generador (100) de señal de audio según las reivindicaciones 1 a 4, en el que el calculador (106) de parámetro es operativo para utilizar un canal predeterminado de los primeros canales originales o del al menos un segundo canal original como canal de referencia común.

6. Generador (100) de señal de audio según las reivindicaciones 1 a 4, en el que el calculador (106) de parámetro es operativo para utilizar el canal de referencia de la primera señal (110) de audio como canal de referencia común.

5 7. Generador (100) de señal de audio según las reivindicaciones 1 a 4, en el que el calculador (106) de parámetro es operativo para utilizar el canal (114) de mezcla descendente combinado como canal de referencia común.

8. Generador (100) de señal de audio según las reivindicaciones 1 a 4, en el que el calculador (106) de parámetro es operativo para utilizar el canal original como canal de referencia común que tiene la energía más alta.

10 9. Generador (100) de señal de audio según cualquiera de las reivindicaciones anteriores, en el que el calculador (106) de parámetro es operativo para calcular la energía $E\{s_{Aref}\}$ del canal de referencia derivando la energía $E\{s_A^2\}$ del primer canal (110a) de mezcla descendente y parámetros a_i ($i = 1, \dots, n$) asociados a canales diferentes del canal de referencia según la ecuación:

$$E\left\{s_{Aref}\right\} = \frac{E\left\{s_A^2(n)\right\}}{1 + \sum_i a_i}.$$

15 10. Generador (100) de señal de audio según cualquiera de las reivindicaciones anteriores, en el que el calculador (106) de parámetro es operativo para utilizar el canal de referencia como canal de referencia común y el al menos un parámetro original a_2 como 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.

20 11. Generador (100) de señal de audio según la reivindicación 1, en el que el calculador (106) de parámetro es operativo para utilizar además los coeficientes g_A asociados al primer canal (110a) de mezcla descendente y g_B asociados al segundo canal (112a) de mezcla descendente, los coeficientes utilizados para la combinación lineal de la primera y segunda mezcla descendente utilizada por el combinador (104) de canales.

25 12. Generador (100) de señal de audio según la reivindicación 13, en el que el calculador (106) de parámetro es operativo para calcular el segundo parámetro combinado y_{u+1} para el al menos un segundo canal original según la siguiente ecuación:

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

en la que $E\left\{s_{A_1}^2(n)\right\}$ es la energía del canal de referencia derivada utilizando la energía del primer canal de mezcla descendente $E\left\{s_A^2(n)\right\}$ según la siguiente fórmula:

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

30 en la que a_2 es el al menos un parámetro original que relaciona un primer canal original con el canal de referencia.

35 13. Generador (100) de señal de audio según cualquiera de las reivindicaciones anteriores, en el que el calculador (106) de parámetro es operativo para procesar las partes de frecuencia del primer y el segundo canal de mezcla descendente asociados con intervalos de frecuencia discretos de tal manera que los parámetros combinados se derivan para cada intervalo de frecuencia discreto.

14. Generador (100) de señal de audio según cualquiera de las reivindicaciones anteriores, en el que el receptor de señal de audio es operativo para recibir señales (110, 112) de audio que comprenden canales (110a, 112a) de mezcla descendente representados por parámetros de muestreo muestreados con una frecuencia de muestra predeterminada.

40 15. Método para generar una señal de salida de audio, comprendiendo el método:

- 5 recibir una primera señal (110) de audio que comprende un primer canal (110a) de mezcla descendente que tiene información sobre dos o más primeros canales originales, y que comprende al menos un parámetro (110b) 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 (112) de audio que comprende un segundo canal (112a) de mezcla descendente que tiene información sobre al menos un segundo canal original;
- 10 derivar un canal (114) de mezcla descendente combinado combinando el primer canal (110) de mezcla descendente y el segundo canal (112) de mezcla descendente utilizando una combinación lineal del primer canal (110a) de mezcla descendente y el segundo (110b), en el que los coeficientes de la combinación lineal dependen de la energía $E\{s_A^2(n)\}$ dentro del primer canal (110a) de mezcla descendente y de la energía $E\{s_B^2(n)\}$ dentro del segundo canal (112a) de mezcla descendente; y
- 15 derivar, utilizando la energía $E\{s_A^2(n)\}$ del primer canal (110a) de mezcla descendente, la energía $E\{s_B^2(n)\}$ del segundo canal (112b) de mezcla descendente y el al menos un parámetro (110b) original,
- un primer parámetro (116a) combinado que describe la propiedad de uno de los primeros canales originales con respecto a un canal de referencia común, y
- un segundo parámetro (116b) combinado que describe o bien la propiedad de otro de los primeros canales originales con respecto al canal de referencia común o bien la propiedad del al menos un segundo canal original con respecto a un canal de referencia común; y
- 20 producir la señal (120) de salida de audio que comprende el canal (114) de mezcla descendente combinado y los parámetros combinados primero (116a) y segundo (116b).
16. Sistema de conferencia que comprende un generador (100) de señal de audio para generar una señal de salida de audio según la reivindicación 1.
17. Programa informático para implementar, cuando se ejecute en un ordenador, un método según la reivindicación 15.

25

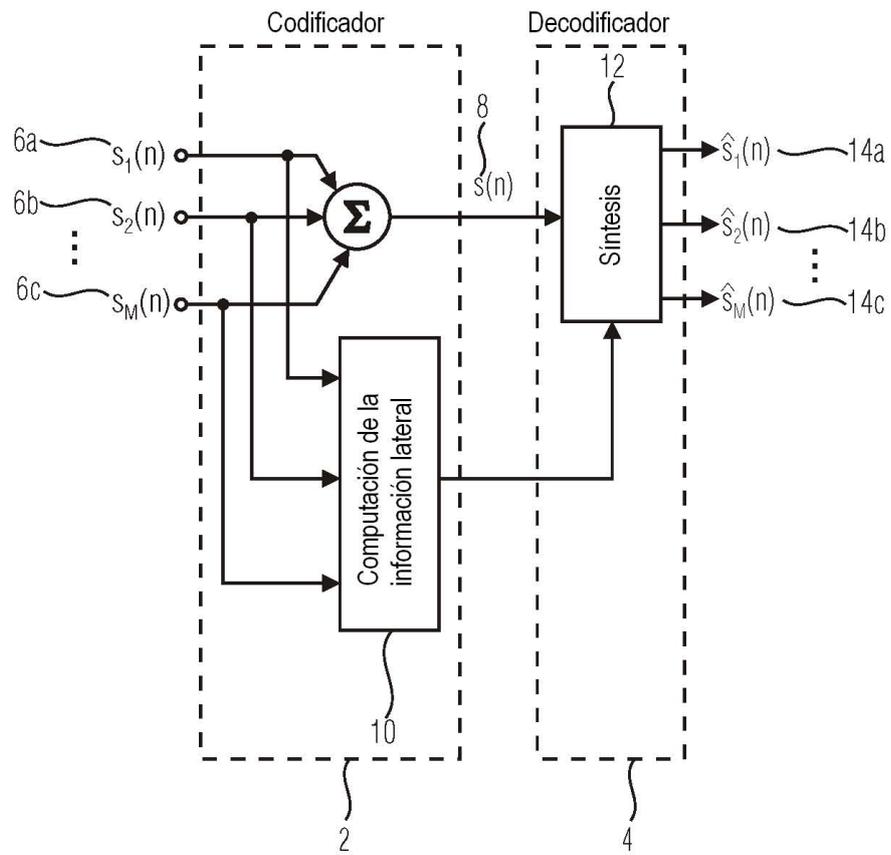


FIG 1

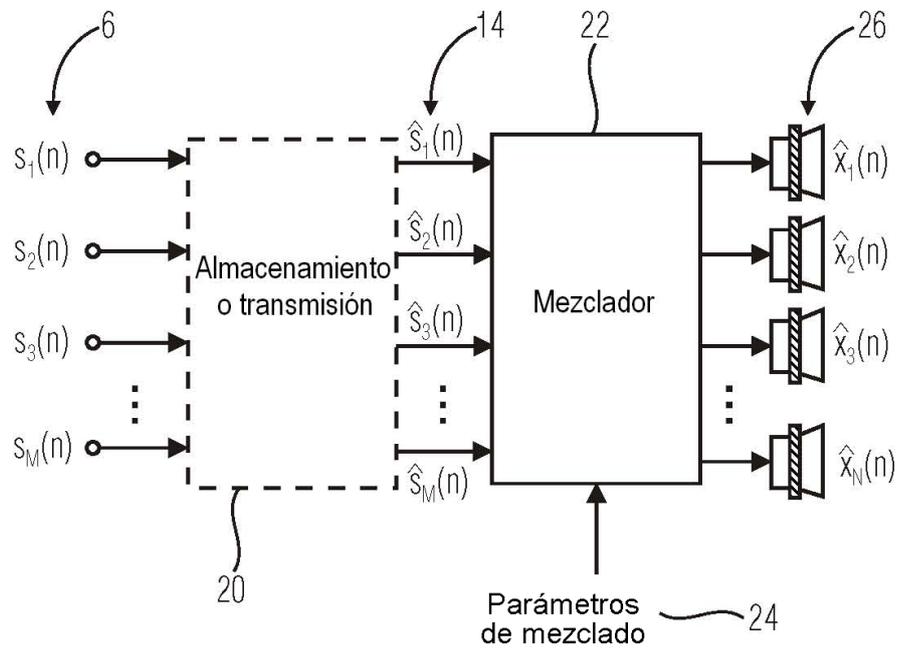


FIG 2

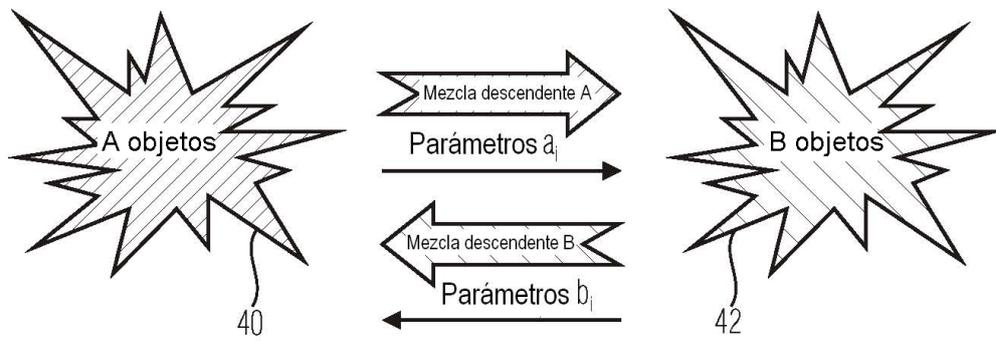


FIG 3

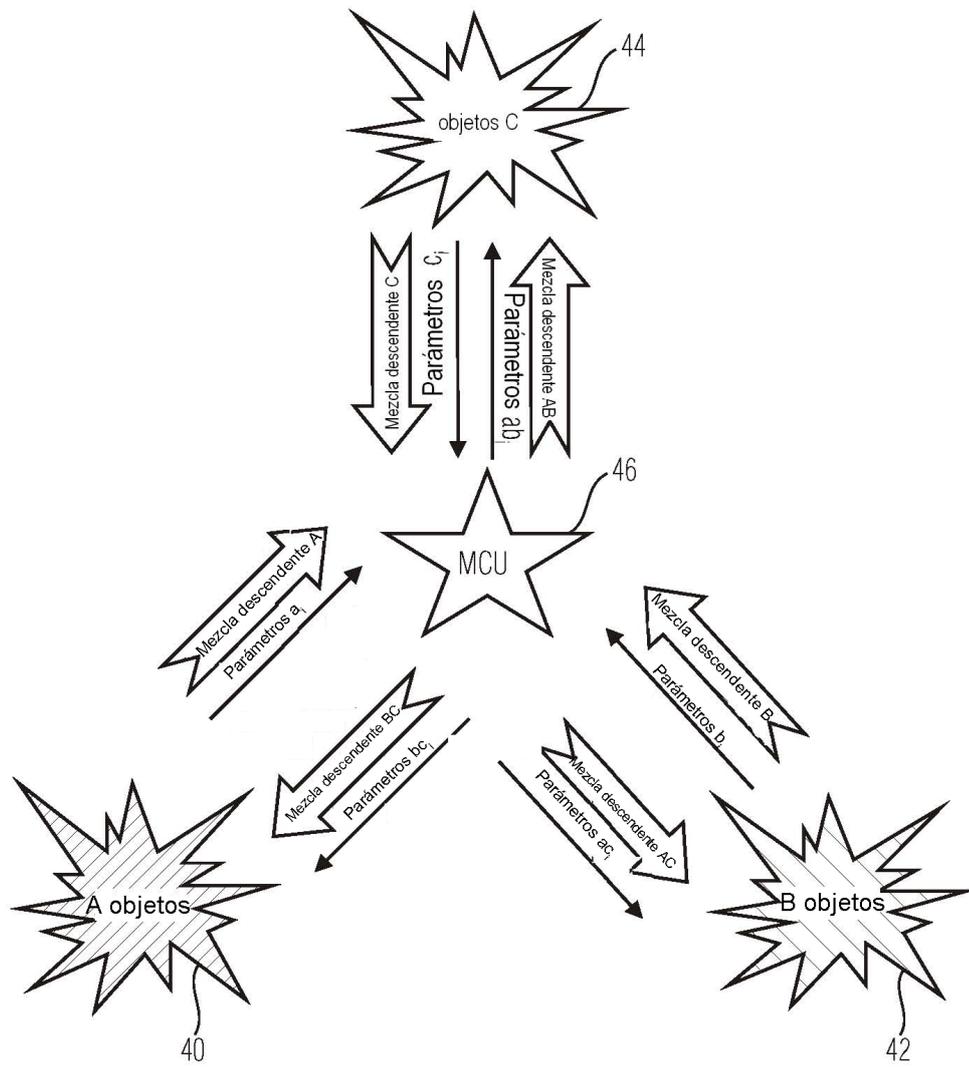


FIG 4

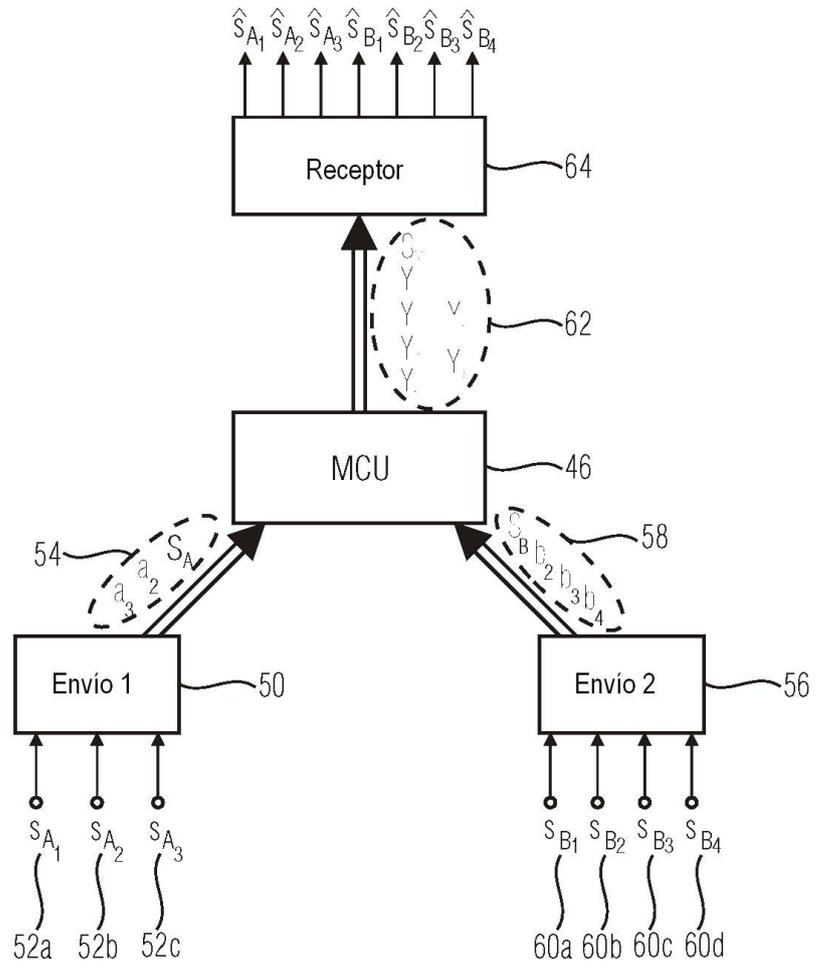


FIG 5

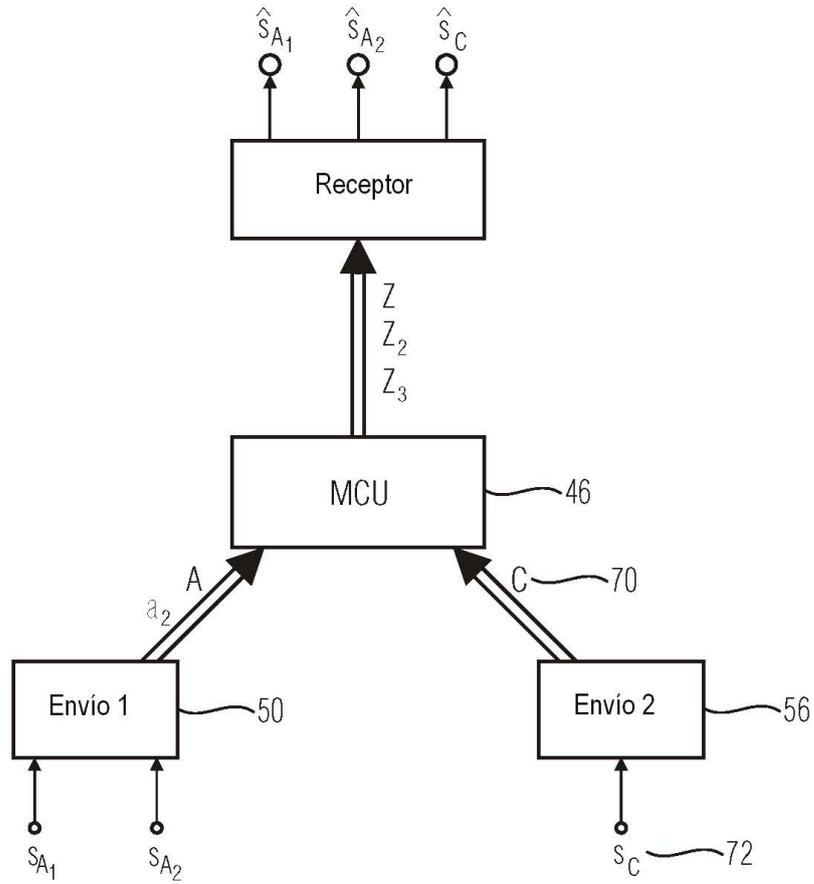


FIG 6

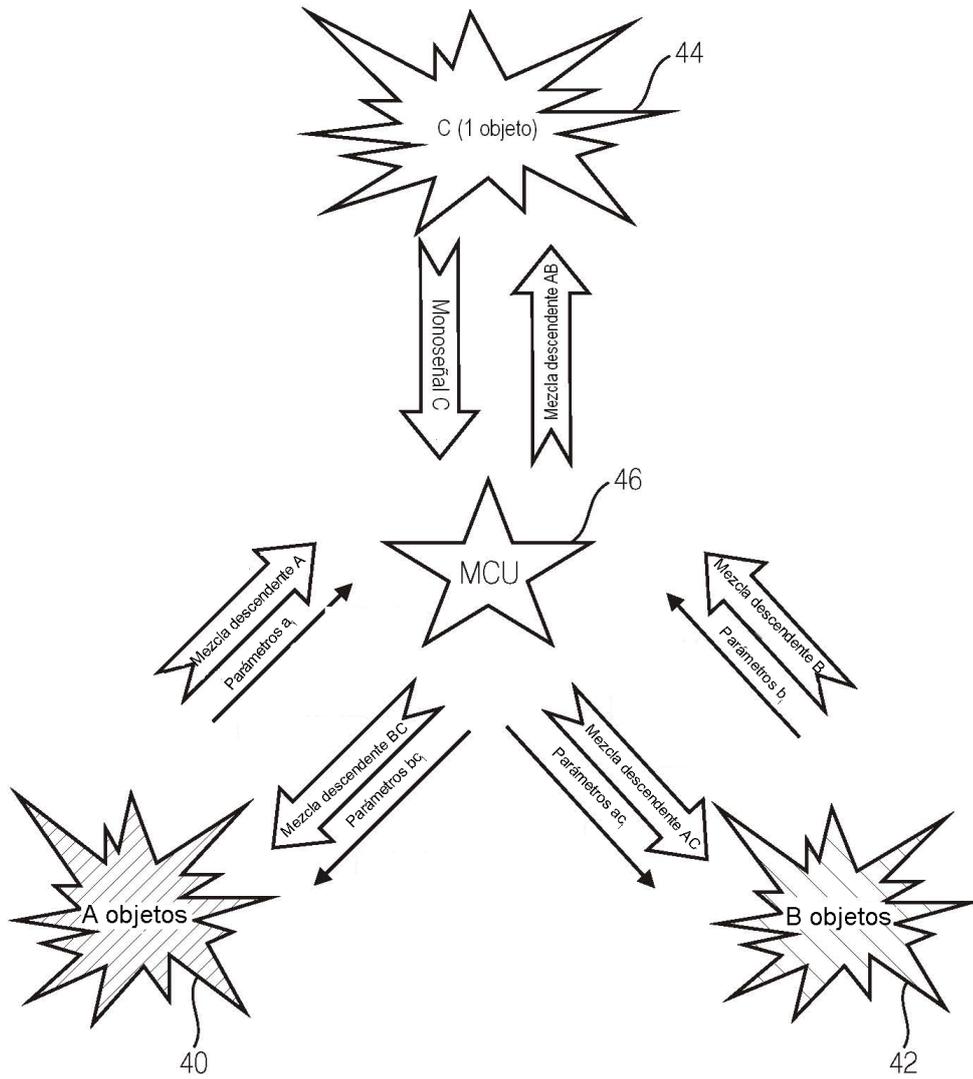


FIG 6B

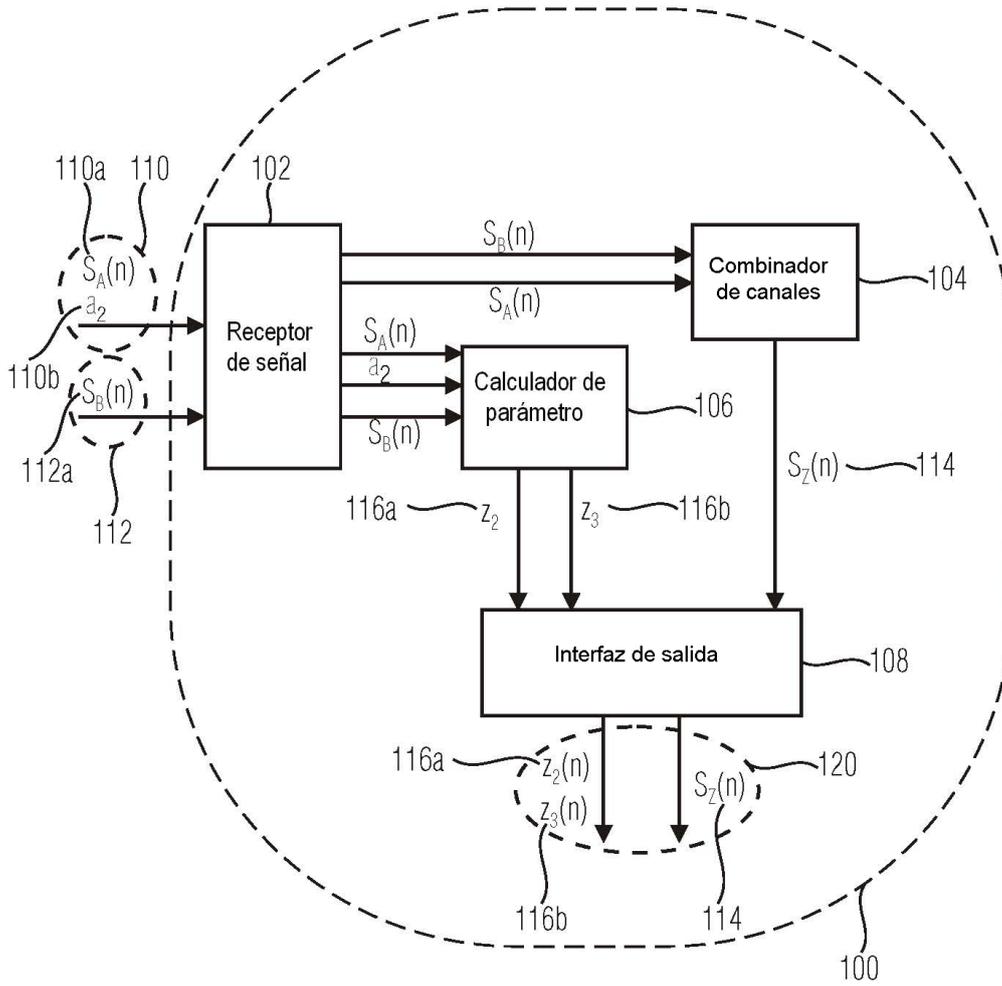


FIG 7