

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 709 117**

51 Int. Cl.:

G10L 21/0364 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **01.10.2015 PCT/EP2015/072666**

87 Fecha y número de publicación internacional: **07.04.2016 WO16050899**

96 Fecha de presentación y número de la solicitud europea: **01.10.2015 E 15771962 (6)**

97 Fecha y número de publicación de la concesión europea: **05.12.2018 EP 3201916**

54 Título: **Codificador y decodificador de audio**

30 Prioridad:

01.10.2014 US 201462058157 P

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

15.04.2019

73 Titular/es:

**DOLBY INTERNATIONAL AB (100.0%)
Apollo Building, 3E, Herikerbergweg 1-35
1101 CN Amsterdam Zuidoost, NL**

72 Inventor/es:

**KOPPENS, JEROEN;
VILLEMOES, LARS;
HIRVONEN, TONI y
KJOERLING, KRISTOFER**

74 Agente/Representante:

ELZABURU, S.L.P

ES 2 709 117 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

Codificador y decodificador de audio

Campo técnico

5 La descripción en este documento se refiere, en general, a la codificación de audio. En particular, se refiere a un método y a un aparato para mejorar el diálogo en un decodificador en un sistema de audio. La descripción se refiere, además, a un método y a un aparato para codificar una pluralidad de objetos de audio, los cuales incluyen al menos un objeto que representa un diálogo.

Técnica anterior

10 En los sistemas de audio convencionales, se emplea un enfoque basado en canales. Por ejemplo, cada canal puede representar el contenido de un altavoz o de un conjunto de altavoces. Los posibles esquemas de codificación para tales sistemas incluyen la codificación multicanal discreta o la codificación paramétrica, como MPEG Surround.

15 Más recientemente, se ha desarrollado un nuevo enfoque. Este enfoque está basado en objetos, lo que puede ser ventajoso al codificar escenas de audio complejas, por ejemplo, en aplicaciones de cine. En los sistemas que emplean el enfoque basado en objetos, una escena de audio tridimensional se representa mediante objetos de audio, con sus metadatos asociados (por ejemplo, metadatos de posición). Estos objetos de audio se mueven en la escena de audio tridimensional durante la reproducción de la señal de audio. El sistema puede incluir, además, los llamados canales de lecho, que pueden describirse como señales que se asignan directamente a ciertos canales de salida, por ejemplo, de un sistema de audio convencional como se describió anteriormente.

20 La mejora del diálogo es una técnica para mejorar o aumentar el nivel de diálogo en relación con otros componentes, tales como música, sonidos de fondo y efectos de sonido. El contenido de audio basado en objetos puede ser adecuado para la mejora del diálogo, ya que el diálogo puede representarse mediante objetos separados. Por ejemplo, en su *“Proposal for extension of SAOC technology for Advanced Clean Audio functionality”* [Propuesta para la extensión de la tecnología SAOC para funcionalidad avanzada de audio limpio], presentada como contribución m29208 en la reunión de MPEG2013 en Incheon, Hellmuth y colaboradores proponen extender el estándar de codificación de objetos de audio espacial (SAOC) para permitir la modificación de la ganancia relativa entre los objetos de primer plano, tales como el diálogo, y los objetos de fondo. Otro ejemplo se provee en Engdegård y colaboradores, que en *“Spatial Audio Object (SAOC) - The Upcoming MPEG Standard on Parametric Object Based Audio Coding”* [“Objeto de audio espacial (SAOC) - El estándar MPEG que se viene sobre codificación de audio basada en objetos paramétricos”, presentado como el documento 7377 en la 124.ª convención de la Audio Engineering Society, donde se sugiere que mediante el uso de las técnicas de SAOC, el usuario puede remezclar interactivamente los diferentes sonidos en una mezcla, como por ejemplo, cambiar el nivel de diálogo con respecto a la música de fondo. Sin embargo, en algunas situaciones, la escena de audio puede comprender una gran cantidad de objetos. Para reducir la complejidad y la cantidad de datos requeridos para representar la escena de audio, la escena de audio puede simplificarse reduciendo el número de objetos de audio, es decir, por agrupación de objetos.

35 Este enfoque puede introducir la mezcla entre el diálogo y otros objetos en algunos de los grupos de objetos.

Al incluir las posibilidades de mejora del diálogo para dichos grupos de audio en un decodificador en un sistema de audio, la complejidad computacional del decodificador puede incrementarse.

Breve descripción de los dibujos

A continuación se describirán realizaciones ejemplares con referencia a los dibujos adjuntos, en los cuales:

40 La figura 1 muestra un diagrama de bloques generalizado de un decodificador de alta calidad, para mejorar el diálogo en un sistema de audio, de acuerdo con las realizaciones ejemplares.

La figura 2 muestra un primer diagrama de bloques generalizado de un decodificador de baja complejidad, para mejorar el diálogo en un sistema de audio, de acuerdo con las realizaciones ejemplares.

45 La figura 3 muestra un segundo diagrama de bloques generalizado de un decodificador de baja complejidad, para mejorar el diálogo en un sistema de audio de acuerdo con las realizaciones ejemplares.

La figura 4 describe un método para codificar una pluralidad de objetos de audio, que incluye al menos un objeto que representa un diálogo, de acuerdo con las realizaciones ejemplares.

La figura 5 muestra un diagrama de bloques generalizado de un codificador para codificar una pluralidad de objetos de audio, que incluye al menos un objeto que representa un diálogo de acuerdo con las realizaciones ejemplares.

50 Todas las figuras son esquemáticas y, en general, solo muestran las partes que son necesarias para dilucidar la descripción, mientras que otras partes pueden omitirse o simplemente sugerirse. A menos que se indique lo contrario, los números de referencia similares se refieren a partes similares en las diferentes figuras.

Descripción detallada

En vista de lo anterior, el objetivo reside en proporcionar codificadores y decodificadores y métodos asociados, con el objeto de reducir la complejidad de la mejora del diálogo en el decodificador.

I. Generalidades: decodificador

5 De acuerdo con un primer aspecto, las realizaciones ejemplares proponen métodos de decodificación, decodificadores y productos de programas informáticos para la decodificación. Los métodos, decodificadores y productos de programas informáticos propuestos generalmente pueden tener las mismas características y ventajas.

10 Según las realizaciones ejemplares, se proporciona un método para mejorar el diálogo en un decodificador en un sistema de audio, que comprende las siguientes etapas: recibir una pluralidad de señales de mezcla descendente, en donde las señales de mezcla descendente son una mezcla descendente de una pluralidad de objetos de audio, que incluyen al menos un objeto que representa un cuadro de diálogo; recibir información complementaria, indicativa de los coeficientes que permiten la reconstrucción de la pluralidad de objetos de audio a partir de la pluralidad de señales de mezcla descendente; recibir datos que identifican cuál de la pluralidad de objetos de audio representa un cuadro de diálogo; modificar los coeficientes utilizando un parámetro mejorador y los datos que identifican cuál de la pluralidad de objetos de audio representa un diálogo, y reconstruir el único objeto como mínimo que representa un diálogo, utilizando los coeficientes modificados.

20 El parámetro mejorador suele ser una configuración del usuario, disponible en el decodificador. Por ejemplo, un usuario puede usar un control remoto para subir el volumen del diálogo. En consecuencia, el parámetro mejorador no suele proporcionarse al decodificador mediante un codificador en el sistema de audio. En muchos casos, el parámetro mejorador se traduce en una ganancia del diálogo, pero también puede traducirse en una atenuación del diálogo. Además, el parámetro mejorador puede relacionarse con ciertas frecuencias del diálogo, por ejemplo, una ganancia dependiente de la frecuencia o atenuación del diálogo.

25 Por el término “diálogo”, en el contexto de la presente memoria descriptiva, debe entenderse que en algunas realizaciones, solo se mejora el diálogo relevante —y no, por ejemplo, la charla de fondo ni cualquier versión reverberante del diálogo. Un diálogo puede comprender una conversación entre personas, pero también un monólogo, una narración u otro discurso.

30 Como se usa aquí, el “objeto de audio” se refiere a un elemento de una escena de audio. Un objeto de audio típicamente comprende una señal de audio e información adicional, como la posición del objeto en un espacio tridimensional. La información adicional se suele utilizar para renderizar de forma óptima el objeto de audio en un sistema de reproducción determinado. La frase “objeto de audio” también abarca un grupo de objetos de audio, es decir, un “grupo de objetos”. Un “grupo de objetos” representa una mezcla de al menos dos objetos de audio y generalmente comprende la mezcla de los objetos de audio como una señal de audio e información adicional, tal como la posición del grupo de objetos en un espacio tridimensional. Los únicos dos objetos como mínimo en un grupo de objetos pueden mezclarse en función de que sus posiciones espaciales individuales estén cerca y de que la posición espacial del grupo de objetos se elija como un promedio de las posiciones individuales de los objetos.

35 Como se usa en este documento, una “señal de mezcla descendente” se refiere a una señal que es una combinación del único objeto como mínimo de audio de la pluralidad de objetos de audio. Otras señales de la escena de audio, como los canales de lecho, también pueden combinarse en la señal de mezcla descendente. El número de señales de mezcla descendente por lo general (aunque no necesariamente) es menor que la suma del número de objetos de audio y canales de lecho, lo que explica por qué las señales de mezcla descendente se denominan mezcla descendente. Una señal de mezcla descendente también puede denominarse un “grupo de mezcla descendente”.

Como se usa en este documento, la “información complementaria” también se puede denominar “metadatos”.

45 En el contexto de la presente memoria descriptiva, por la expresión “información complementaria indicativa de coeficientes” debe entenderse que los coeficientes están directamente presentes en la información complementaria enviada, por ejemplo, en un flujo de bits desde el codificador, o que se calculan a partir de los datos presentes en la información complementaria.

50 De acuerdo con el presente método, los coeficientes que permiten la reconstrucción de la pluralidad de objetos de audio se modifican para mejorar el único objeto de audio como mínimo reconstruido en último término que representa un diálogo. En comparación con el método convencional de realizar la mejora del único objeto de audio como mínimo reconstruido y que representa un diálogo después de que ha sido reconstruido —es decir, sin modificar los coeficientes que permiten la reconstrucción— el presente método ofrece una menor complejidad matemática y, por lo tanto, una complejidad computacional del decodificador que implementa el presente método.

55 De acuerdo con las realizaciones ejemplares, la etapa de modificar los coeficientes utilizando el parámetro mejorador comprende multiplicar los coeficientes que permiten la reconstrucción del único objeto como mínimo que

representa un diálogo con el parámetro mejorador. Esta es una operación de baja complejidad desde el punto de vista computacional, para modificar los coeficientes que aún mantiene la relación mutua entre los coeficientes.

5 De acuerdo con las realizaciones ejemplares, el método comprende además: calcular los coeficientes que permiten la reconstrucción de la pluralidad de objetos de audio a partir de la pluralidad de señales de mezcla descendente provenientes de la información complementaria.

De acuerdo con las realizaciones ejemplares, la etapa de reconstruir el único objeto como mínimo que representa un diálogo comprende reconstruir solo el único objeto como mínimo que representa un diálogo.

10 En muchos casos, las señales de mezcla descendente pueden corresponder a una renderización o salida de la escena de audio a una configuración de altavoz dada, por ejemplo, una configuración 5.1 estándar. En tales casos, la decodificación de baja complejidad se puede lograr solo reconstruyendo los objetos de audio que representan el diálogo a mejorar, es decir, no realizar una reconstrucción completa de todos los objetos de audio.

15 De acuerdo con las realizaciones ejemplares, la reconstrucción de solo el único objeto como mínimo que representa un diálogo no implica la descorrelación de las señales de mezcla descendente. Esto reduce la complejidad de la etapa de reconstrucción. Además, dado que no todos los objetos de audio se reconstruyen, es decir, la calidad del contenido de audio que se va a re-renderizar puede reducirse para esos objetos de audio, el uso de la descorrelación al reconstruir el único objeto como mínimo que representa el diálogo no mejoraría la calidad de audio percibida del contenido de audio renderizado que se ha mejorado. En consecuencia, la decorrelación puede omitirse.

20 De acuerdo con las realizaciones ejemplares, el método comprende, además, la etapa de: fusionar el único objeto como mínimo que representa el diálogo reconstruido con las señales de mezcla descendente como una única señal separada como mínimo. En consecuencia, el único objeto como mínimo reconstruido no necesita mezclarse o combinarse otra vez con las señales de mezcla descendente. Por consiguiente, de acuerdo con esta realización, no se necesita información que describa cómo se mezcló el único objeto como mínimo que representa un diálogo con la pluralidad de señales de mezcla descendente mediante un codificador en el sistema de audio.

25 De acuerdo con las realizaciones ejemplares, el método comprende, además, recibir datos con información espacial correspondiente a las posiciones espaciales para la pluralidad de señales de mezcla descendente y para el único objeto como mínimo que representa un diálogo, y renderizar la pluralidad de señales de mezcla descendente y el único objeto como mínimo que representa un diálogo reconstruido en función de los datos con información espacial.

30 De acuerdo con las realizaciones ejemplares, el método comprende, además, combinar las señales de mezcla descendente y el único objeto como mínimo que representa un diálogo reconstruido utilizando información que describe cómo se mezcló el único objeto como mínimo que representa un diálogo en la pluralidad de señales de mezcla descendente mediante un codificador en el sistema de audio. Las señales de mezcla descendente se pueden mezclar en forma descendente como soporte de salida de audio siempre (AAO, *always audio out*) para una determinada configuración de altavoz (por ejemplo, una configuración 5.1 o una configuración 7.1), es decir, las señales de mezcla descendente se pueden utilizar directamente para la reproducción en una configuración de este tipo de altavoz. Al combinar las señales de mezcla descendente y el único objeto como mínimo que representa un cuadro de diálogo reconstruido, la mejora del diálogo se logra al mismo tiempo que todavía se admite AAO. En otras palabras, de acuerdo con algunas realizaciones, el único objeto como mínimo que representa un diálogo reconstruido y con una mejora en el diálogo, se mezcla nuevamente en las señales de mezcla descendente para seguir siendo compatible con AAO.

40 De acuerdo con las realizaciones ejemplares, el método comprende, además, renderizar la combinación de las señales de mezcla descendente y el único objeto como mínimo que representa un diálogo reconstruido.

45 De acuerdo con las realizaciones ejemplares, el método comprende, además, recibir información que describe cómo se mezcló el único objeto como mínimo que representa un diálogo en la pluralidad de señales de mezcla descendente mediante un codificador en el sistema de audio. El codificador en el sistema de audio ya puede tener este tipo de información al mezclar en forma descendente la pluralidad de objetos de audio, que incluye al único objeto como mínimo que representa un diálogo; alternativamente, la información puede calcularse fácilmente mediante el codificador.

50 De acuerdo con las realizaciones ejemplares, la información recibida que describe cómo se mezcló el único objeto como mínimo que representa un diálogo en la pluralidad de señales de mezcla descendente se codifica por codificación de entropía. Esto puede reducir la tasa de bits requerida para transmitir la información.

55 De acuerdo con las realizaciones ejemplares, el método comprende, además, las etapas de: recibir datos con información espacial correspondiente a las posiciones espaciales para la pluralidad de señales de mezcla descendente y para el único objeto como mínimo que representa un diálogo, y calcular la información que describe cómo se mezcló el único objeto como mínimo que representa un diálogo en la pluralidad de señales de mezcla descendente mediante un codificador en el sistema de audio, basándose en los datos con información espacial. Una ventaja de esta realización puede ser que se reduce la tasa de bits requerida para transmitir el flujo de bits que incluye las señales de mezcla descendente y la información complementaria al codificador, ya que la información

espacial correspondiente a las posiciones espaciales para la pluralidad de señales de mezcla descendente y para el único objeto como mínimo que representa un diálogo puede recibirse de todas maneras y no es necesario que el decodificador reciba más información o datos.

5 De acuerdo con las realizaciones ejemplares, la etapa de calcular la información que describe cómo se mezcló el único objeto como mínimo que representa un diálogo en la pluralidad de señales de mezcla descendente comprende la aplicación de una función que mapea la posición espacial para el único objeto como mínimo que representa un diálogo en las posiciones espaciales para la pluralidad de señales de mezcla descendente. La función puede ser, por ejemplo, un algoritmo de panoramización 3D tal como un algoritmo de panoramización de amplitud base vectorial (VBAP, *Vector Base Amplitude Panning*). Se puede utilizar cualquier otra función adecuada.

10 De acuerdo con las realizaciones ejemplares, la etapa de reconstruir el único objeto como mínimo que representa un diálogo comprende reconstruir la pluralidad de objetos de audio. En ese caso, el método puede comprender recibir datos con información espacial correspondiente a las posiciones espaciales para la pluralidad de objetos de audio, y representar la pluralidad reconstruida de objetos de audio basándose en los datos con información espacial. Dado que la mejora del diálogo se realiza en los coeficientes que permiten la reconstrucción de la pluralidad de objetos de audio, como se describió anteriormente, la reconstrucción de la pluralidad de objetos de audio y la renderización objeto de audio reconstruido, que son operaciones de matriz, pueden combinarse en una sola operación que reduce la complejidad de las dos operaciones.

De acuerdo con las realizaciones ejemplares, se proporciona un medio legible por computadora que comprende instrucciones de códigos informáticos adaptadas para llevar a cabo cualquier método del primer aspecto cuando se ejecuta en un dispositivo que tiene capacidad de procesamiento.

20 Según realizaciones ejemplares, se proporciona un decodificador para mejorar el diálogo en un sistema de audio. El decodificador comprende una etapa de recepción configurada para: recibir una pluralidad de señales de mezcla descendente, en donde las señales de mezcla descendente son una mezcla descendente de una pluralidad de objetos de audio, que incluyen al único objeto como mínimo que representa un diálogo; recibir información complementaria, indicativa de los coeficientes que permiten la reconstrucción de la pluralidad de objetos de audio a partir de la pluralidad de señales de mezcla descendente y recibir datos que identifican cuál de la pluralidad de objetos de audio representa un diálogo. El decodificador comprende, además, una etapa de modificación configurada para modificar los coeficientes utilizando un parámetro mejorador y los datos que identifican cuál de la pluralidad de objetos de audio representa un diálogo. El decodificador comprende, además, una etapa de reconstrucción, configurada para reconstruir el único objeto como mínimo que representa un diálogo utilizando los coeficientes modificados.

II. Generalidades: codificador

Según un segundo aspecto, las realizaciones ejemplares proponen métodos de codificación, codificadores y productos de programas informáticos para la codificación. Los métodos, codificadores y productos de programas de computadora propuestos generalmente pueden tener las mismas características y ventajas. En general, las características del segundo aspecto pueden tener las mismas ventajas que las características correspondientes del primer aspecto.

De acuerdo con las realizaciones ejemplares, se proporciona un método para codificar una pluralidad de objetos de audio, que incluye al único objeto como mínimo que representa un diálogo, que comprende las siguientes etapas: determinar una pluralidad de señales de mezcla descendente, que es una mezcla descendente de la pluralidad de objetos de audio que incluyen al menos un objeto representando un diálogo; determinar la información complementaria indicativa de los coeficientes que permiten la reconstrucción de la pluralidad de objetos de audio a partir de la pluralidad de señales de mezcla descendente; determinar los datos que identifican cuál de la pluralidad de objetos de audio representa un diálogo y formar un flujo de bits que comprende la pluralidad de señales de mezcla descendente, la información complementaria y los datos que identifican cuál de la pluralidad de objetos de audio representa un diálogo.

De acuerdo con las realizaciones ejemplares, el método comprende, además, las etapas para determinar la información espacial correspondiente a las posiciones espaciales para la pluralidad de señales de mezcla descendente y para el único objeto como mínimo que representa un diálogo, e incluir dicha información espacial en el flujo de bits.

Según realizaciones ejemplares, la etapa de determinar una pluralidad de señales de mezcla descendente comprende, además, determinar información que describe cómo se mezcla el único objeto como mínimo que representa un diálogo en la pluralidad de señales de mezcla descendente. Esta información que describe cómo se mezcla el único objeto como mínimo que representa un diálogo con la pluralidad de señales de mezcla descendente está de acuerdo con esta realización incluida en el flujo de bits.

De acuerdo con las realizaciones ejemplares, la información determinada que describe cómo se mezcla el único objeto como mínimo que representa un diálogo en la pluralidad de señales de mezcla descendente se codifica utilizando la codificación de entropía.

5 De acuerdo con las realizaciones ejemplares, el método comprende, además, las etapas de determinar la información espacial correspondiente a las posiciones espaciales para la pluralidad de objetos de audio, e incluir la información espacial correspondiente a las posiciones espaciales para la pluralidad de objetos de audio en el flujo de bits.

10 De acuerdo con las realizaciones ejemplares, se proporciona un medio legible por computadora que comprende instrucciones de códigos informáticos adaptadas para llevar a cabo cualquier método del segundo aspecto, cuando se ejecuta en un dispositivo que tiene capacidad de procesamiento.

15 De acuerdo con las realizaciones ejemplares, se proporciona un codificador para codificar una pluralidad de objetos de audio que incluyen al único objeto como mínimo que representa un diálogo. El codificador comprende una etapa de mezcla descendente, configurada para: determinar una pluralidad de señales de mezcla descendente que es una mezcla descendente de la pluralidad de objetos de audio que incluye al único objeto como mínimo que representa un diálogo; determinar información complementaria que comprende indicativos de coeficientes que permiten la reconstrucción de la pluralidad de objetos de audio a partir de la pluralidad de señales de mezcla descendente y una etapa de codificación configurada para: formar un flujo de bits que comprende la pluralidad de señales de mezcla descendente y la información complementaria, en donde el flujo de bits comprende, además, datos que identifican cuál de la pluralidad de objetos de audio representa un diálogo.

20 III. Realizaciones ejemplares

Como se describió anteriormente, la mejora del diálogo se trata de aumentar el nivel de diálogo en relación con los otros componentes de audio. Cuando se organiza correctamente desde la creación de contenido, el contenido del objeto es adecuado para la mejora del diálogo, ya que el diálogo puede representarse por objetos separados. La codificación paramétrica de los objetos (es decir, grupos de objetos o señales de mezcla descendente) puede introducir la mezcla entre el diálogo y otros objetos.

25 A continuación se describirá un decodificador para mejorar el diálogo mezclado en tales grupos de objetos, junto con las figuras 1-3. La figura 1, muestra un diagrama de bloques generalizado de un decodificador 100 de alta calidad, para mejorar el diálogo en un sistema de audio de acuerdo con las realizaciones ejemplares. El decodificador 100 recibe un flujo de bits 102 en una etapa de recepción 104. La etapa de recepción 104 también puede verse como un decodificador núcleo, que decodifica al flujo de bits 102 y genera el contenido decodificado del flujo de bits 102. El flujo de bits 102 puede comprender, por ejemplo, una pluralidad de señales de mezcla descendente 110 o grupos de mezcla descendente, que son una mezcla descendente de una pluralidad de objetos de audio que incluyen al único objeto como mínimo que representa un diálogo. Por lo general, la etapa de recepción comprende un componente de decodificador de mezcla descendente que puede adaptarse para decodificar partes del flujo de bits 30 102 para formar las señales de mezcla descendente 110, de manera que sean compatibles con el sistema de decodificación de sonido del decodificador, como Dolby Digital Plus o los estándares MPEG, tales como AAC, USAC o MP3. El flujo de bits 102 puede comprender, además, información complementaria 108 indicativa de coeficientes que permiten la reconstrucción de la pluralidad de objetos de audio a partir de la pluralidad de señales de mezcla descendente. Para una mejora eficiente del diálogo, el flujo de bits 102 puede comprender, además, los datos 108 35 que identifican cuál de la pluralidad de objetos de audio representa un diálogo. Estos datos 108 pueden incorporarse en la información complementaria 108, o pueden estar separados de la información complementaria 108. Como se explica en detalle a continuación, la información complementaria 108 típicamente comprende mezcla seca ascendente de los coeficientes que pueden traducirse en una matriz de mezcla ascendente seca **C** y coeficientes de mezcla ascendente húmeda que se pueden traducir en una matriz de mezcla ascendente húmeda **P**.

40 El decodificador 100 comprende, además, una etapa de modificación 112, que está configurada para modificar los coeficientes indicados en la información complementaria 108, mediante el uso de un parámetro mejorador 140 y los datos 108 que identifican cuál de la pluralidad de objetos de audio representa un diálogo. El parámetro mejorador 140 se puede ser recibido en la etapa de modificación 112, de cualquier manera adecuada. De acuerdo con las realizaciones, la etapa de modificación 112 modifica tanto la matriz de mezcla ascendente seca **C** como la matriz de 45 mezcla ascendente húmeda **P**, en donde al menos los coeficientes correspondientes al diálogo.

La etapa de modificación 112 está aplicando así la mejora de diálogo deseada a los coeficientes correspondientes al o a los objetos de diálogo. Según una realización, la etapa de modificar los coeficientes utilizando el parámetro mejorador 140 comprende multiplicar los coeficientes que permiten la reconstrucción del único objeto como mínimo que representa un diálogo con el parámetro mejorador 140. En otras palabras, la modificación comprende una 50 amplificación fija de los coeficientes correspondientes a los objetos de diálogo.

En algunas formas de realización, el decodificador 100 comprende, además, una etapa de descorrelación previa 114 y una etapa de decorrelación 116. Estas dos etapas 114,116 forman conjuntamente versiones descorrelacionadas de las combinaciones de las señales de mezcla descendente 110, que se utilizarán más adelante para la

reconstrucción (por ejemplo, en la mezcla ascendente) de la pluralidad de objetos de audio provenientes de la pluralidad de señales de mezcla descendente 110. Como se puede ver en la figura 1, la información complementaria 108 se puede enviar a la etapa de descorrelación previa 114 antes de la modificación de los coeficientes en la etapa de modificación 112. De acuerdo con las realizaciones, los coeficientes indicados en la información complementaria 108 se traducen en una matriz modificada de mezcla ascendente seca 120, una matriz modificada de mezcla ascendente húmeda 142 y una matriz de descorrelación previa **Q**, indicada como referencia 144 en la figura 1. La matriz modificada de mezcla ascendente húmeda se usa para mezclar de manera ascendente las señales de descorrelación 122 en una etapa de reconstrucción 124, como se describe a continuación.

La matriz de descorrelación previa **Q** se usa en la etapa de descorrelación previa 114 y, de acuerdo con las realizaciones, puede calcularse de la siguiente manera:

$$\mathbf{Q} = (\mathbf{abs P})^T \mathbf{C}$$

donde **abs P** denota la matriz obtenida tomando valores absolutos de los elementos de la matriz de mezcla húmeda ascendente **P** no modificada, y **C** denota la matriz de mezcla ascendente seca no modificada.

Se contemplan formas alternativas de calcular los coeficientes de descorrelación previa **Q**, basados en la matriz de mezcla ascendente seca **C** y la matriz de mezcla ascendente húmeda **P**. Por ejemplo, puede calcularse como $\mathbf{Q} = (\mathbf{abs P}_0)^T \mathbf{C}$, donde la matriz **P**₀ se obtiene al normalizar cada columna de **P**.

El cálculo de la matriz de descorrelación previa **Q** solo implica cálculos con una complejidad relativamente baja y, por lo tanto, puede emplearse de un modo conveniente del lado del decodificador. Sin embargo, según algunas realizaciones, la matriz de descorrelación previa **Q** se incluye en la información complementaria 108.

En otras palabras, el decodificador puede configurarse para calcular los coeficientes que permiten la reconstrucción de la pluralidad de objetos de audio 126 a partir de la pluralidad de señales de mezcla descendente provenientes de la información complementaria. De esta manera, la matriz de descorrelación previa no se ve influenciada por ninguna modificación hecha a los coeficientes en la etapa de modificación, lo que puede ser ventajoso ya que, si se modifica la matriz de descorrelación previa, el proceso de descorrelación en la etapa de descorrelación previa 114 y en la etapa de decorrelación 116 puede introducir una mejora adicional del diálogo que podría no desearse. De acuerdo con otras realizaciones, la información complementaria se introduce en la etapa de descorrelación previa 114 después de la modificación de los coeficientes en la etapa 112 de modificación. Dado que el decodificador 100 es un decodificador de alta calidad, puede configurarse para reconstruir la pluralidad completa de objetos de audio. Esto se realiza en la etapa de reconstrucción 124. La etapa de reconstrucción 124 del decodificador 100 recibe así las señales de mezcla descendente 110, las señales descorrelacionadas 122 y los coeficientes modificados 120, 142, que permiten la reconstrucción de la pluralidad de objetos de audio a partir de la pluralidad de señales de mezcla descendente 110. Así, la etapa de reconstrucción puede reconstruir paramétricamente los objetos de audio 126 antes de renderizar los objetos de audio en la configuración de salida del sistema de audio, por ejemplo una salida de canal 7.1.4. Sin embargo, normalmente esto no ocurrirá en muchos casos, ya que la reconstrucción de los objetos de audio en la etapa de reconstrucción 124 y la renderización en la etapa de renderización 128 son operaciones matriciales que pueden combinarse (denotadas por la línea discontinua 134) para una implementación computacionalmente eficiente. Para renderizar los objetos de audio en una posición correcta en un espacio tridimensional, el flujo de bits 102 comprende, además, los datos 106 con información espacial correspondiente a las posiciones espaciales para la pluralidad de objetos de audio.

Puede observarse que, según algunas realizaciones, el decodificador 100 se configurará para proporcionar los objetos reconstruidos como una salida, de manera que puedan procesarse y renderizarse fuera del decodificador. De acuerdo con esta realización, el decodificador 100 genera en consecuencia los objetos de audio reconstruidos 126 y no comprende la etapa de renderización 128.

La reconstrucción de los objetos de audio se realiza típicamente en un dominio de frecuencia, por ejemplo, un dominio de filtros espejo en cuadratura (QMF, *Quadrature Mirror Filters*). Sin embargo, es posible que el audio deba emitirse en un dominio de tiempo. Por esta razón, el decodificador comprende, además, una etapa de transformación 132, en la que las señales renderizadas 130 se transforman en el dominio del tiempo, por ejemplo, aplicando un banco de filtros espejo de cuadratura inversa (IQMF, *Inverse Quadrature Mirror Filters*). De acuerdo con algunas realizaciones, la transformación en la etapa de transformación 132 al dominio del tiempo puede realizarse antes de renderizar las señales en la etapa de renderización 128.

En resumen, la implementación del decodificador descrita en conjunto con la figura 1 implementa de manera eficiente la mejora del diálogo, al modificar los coeficientes que permiten la reconstrucción de la pluralidad de objetos de audio, a partir de la pluralidad de señales de mezcla descendente, antes de la reconstrucción de los objetos de audio. Realizar la mejora de los coeficientes cuesta unas pocas multiplicaciones por cuadro, una para cada coeficiente relacionado con el diálogo multiplicado por el número de bandas de frecuencia. Lo más probable es que, en casos típicos, el número de multiplicaciones sea igual al número de canales de mezcla descendente (por ejemplo, 5-7) multiplicado por el número de bandas de parámetros (por ejemplo, 20-40), pero podría ser mayor si el cuadro de diálogo también obtiene una contribución de descorrelación. En comparación, la solución de la técnica

anterior para realizar una mejora del diálogo en los objetos reconstruidos resulta en una multiplicación por cada muestra multiplicada por el número de bandas de frecuencia, dos para una señal compleja. Por lo general, esto llevará a una multiplicación de $16 * 64 * 2 = 2048$ por cuadro, a menudo más.

5 Los sistemas de codificación/decodificación de audio generalmente dividen el espacio tiempo-frecuencia en mosaicos de tiempo/frecuencia, por ejemplo, aplicando bancos de filtros adecuados a las señales de audio de entrada. Por mosaico de tiempo/frecuencia se entiende una porción del espacio tiempo-frecuencia correspondiente a un intervalo de tiempo y a una banda de frecuencia. El intervalo de tiempo suele corresponder a la duración de un cuadro de tiempo utilizado en el sistema de codificación/decodificación de audio. La banda de frecuencia es parte de todo el rango de frecuencia de la señal de audio/del objeto que se está codificando o decodificando. La banda de frecuencia puede corresponder típicamente a una o a varias bandas de frecuencia vecinas, definidas por un banco de filtros utilizado en el sistema de codificación/decodificación. En el caso de que la banda de frecuencia corresponda a varias bandas de frecuencia vecinas definidas por el banco de filtros, esto permite tener bandas de frecuencia no uniformes en el proceso de decodificación de la señal de audio, por ejemplo, bandas de frecuencia más anchas para frecuencias más altas de la señal de audio.

15 En un modo de salida alternativo, para guardar la complejidad del decodificador, los objetos de mezcla descendente no se reconstruyen. Las señales de mezcla descendente se consideran en esta realización como señales que se van a renderizar directamente en la configuración de salida, por ejemplo, una configuración de salida 5.1. Esto también se conoce como un modo de operación de salida de audio siempre (AAO). Las figuras 2 y 3 describen los decodificadores 200, 300 que permiten mejorar el diálogo incluso para esta realización de baja complejidad.

20 La Figura 2 describe un decodificador 200 de baja complejidad para mejorar el diálogo en un sistema de audio de acuerdo con las primeras realizaciones ejemplares. El decodificador 100 recibe el flujo de bits 102 en la etapa de recepción 104 o el decodificador núcleo. La etapa de recepción 104 se puede configurar como se describe en conjunto con la figura 1. En consecuencia, la etapa de recepción emite información complementaria 108 y señales de mezcla descendente 110. Los coeficientes indicados por la información complementaria 108 son modificados por el parámetro mejorador 140, como se describió antes, mediante la etapa de modificación 112, con la diferencia de que debe tenerse en cuenta que el diálogo ya está presente en la señal de mezcla descendente 110 y, en consecuencia, el parámetro mejorador puede tener que reducirse antes de ser utilizado para modificar la información complementaria 108, como se describe a continuación. Una diferencia adicional puede residir en que, dado que no se emplea la descorrelación en el decodificador 200 de baja complejidad (como se describe a continuación), la etapa de modificación 112 solo modifica los coeficientes de mezcla ascendente seca en la información complementaria 108 y, por lo tanto, ignora cualquier coeficiente de mezcla ascendente húmeda presente en la información complementaria 108. En algunas formas de realización, la corrección puede tener en cuenta una pérdida de energía en la predicción del objeto de diálogo, causada por la omisión de la contribución de descorrelación. La modificación por la etapa de modificación 112 garantiza que los objetos de diálogo se reconstruyan como señales de mejora que, cuando se combinan con las señales de mezcla descendente, dan como resultado un diálogo mejorado. Los coeficientes modificados 218 y las señales de mezcla descendente se ingresan en una etapa de reconstrucción 204. En la etapa de reconstrucción, solo el único objeto como mínimo que representa un diálogo puede reconstruirse usando los coeficientes modificados 218. Para reducir aún más la complejidad de decodificación del decodificador 200, la reconstrucción del único objeto como mínimo que representa un diálogo en la etapa de reconstrucción 204 no implica una descorrelación de las señales de mezcla descendente 110. La etapa de reconstrucción 204 genera así una o más señales de mejora de diálogo 206. En muchas realizaciones, la etapa de reconstrucción 204 es una parte de la etapa de reconstrucción 124, estando dicha parte relacionada con la reconstrucción del único objeto como mínimo que representa un diálogo.

45 Para seguir emitiendo señales de acuerdo con la configuración de salida admitida, es decir, la configuración de salida con la que se mezclaron en forma descendente las señales de mezcla descendente 110 para admitir (por ejemplo, señales envolventes 5.1 o 7.1), las señales 206 mejoradas en el diálogo deben mezclarse en forma descendente o combinarse con las señales de mezcla descendente 110 otra vez. Por esta razón, el decodificador comprende una etapa de mezcla adaptativa 208, que usa la información 202 que describe cómo se mezcló el único objeto como mínimo que representa un diálogo en la pluralidad de señales de mezcla descendente, mediante un codificador en el sistema de audio para mezclar los objetos de mejora de diálogo nuevamente en una la representación 210, que corresponde a cómo se representan los objetos de diálogo en las señales de mezcla descendente 110. Esta representación se combina 212 luego, con la señal de mezcla descendente 110, de manera que las señales combinadas resultantes 214 comprendan un diálogo mejorado.

55 Las etapas conceptuales descritas anteriormente para mejorar el diálogo en una pluralidad de señales de mezcla descendente pueden implementarse mediante una única operación matricial en la matriz D , que representa un mosaico de tiempo-frecuencia de la pluralidad de señales de mezcla descendente 110:

$$D_b = D + MD \quad \text{ecuación 1}$$

donde D_b es una mezcla descendente 214 modificada, que incluye las partes de diálogo potenciadas. La matriz modificadora M se obtiene mediante:

$$M = GC \quad \text{ecuación 2}$$

5 donde **G** es una matriz de [número de canales de mezcla descendente, número de objetos de diálogo] de ganancias de mezcla descendente, es decir, la información 202 que describe cómo se mezcló el único objeto como mínimo que representa un diálogo en el mosaico de tiempo-frecuencia **D** decodificado actualmente de la pluralidad de señales de mezcla descendente 110. **C** es una matriz [número de objetos de diálogo, número de canales de mezcla descendente] de los coeficientes modificados 218.

10 Una implementación alternativa para mejorar el diálogo en una pluralidad de señales de mezcla descendente puede implementarse mediante una operación matricial en el vector de columna **X** [número de canales de mezcla descendente], en donde cada elemento representa una única muestra de tiempo-frecuencia de la pluralidad de señales de mezcla descendente 110:

$$X_b = EX \quad \text{ecuación 3}$$

donde **X_b** es una mezcla descendente modificada 214, que incluye las partes de diálogo mejoradas. La matriz modificadora **E** se obtiene mediante:

$$E = I + GC \quad \text{ecuación 4}$$

15 donde **I** es la matriz de identidad [número de los canales de mezcla descendente, número de los canales de mezcla descendente], **G** es una matriz [número de los canales de mezcla descendente, número de objetos de diálogo] de ganancia de mezcla descendente, es decir, la información 202 que describe cómo se mezcló el único objeto como mínimo que representa una diálogo con la pluralidad actualmente decodificada de señales de mezcla descendente 110, y **C** es una matriz [número de objetos de diálogo, número de canales de mezcla descendente] de los coeficientes modificados 218.

20 La matriz **E** se calcula para cada banda de frecuencia y muestra de tiempo en el cuadro. Normalmente, los datos para la matriz **E** se transmiten una vez por cuadro y la matriz se calcula para cada muestra de tiempo en el mosaico de tiempo-frecuencia mediante la interpolación con la matriz correspondiente en el cuadro anterior.

25 Según algunas realizaciones, la información 202 es parte del flujo de bits 102 y comprende los coeficientes de mezcla descendente que fueron utilizados por el codificador en el sistema de audio para mezclar los objetos de diálogo en las señales de mezcla descendente.

30 En algunas realizaciones, las señales de mezcla descendente no corresponden a los canales de una configuración de altavoz. En tales realizaciones, es beneficioso renderizar las señales de mezcla descendente en ubicaciones correspondientes a los altavoces de la configuración utilizada para la reproducción. Para estas realizaciones, el flujo de bits 102 puede transportar datos de posición para la pluralidad de señales de mezcla descendente 110.

Ahora se describirá una sintaxis ejemplar del flujo de bits correspondiente a dicha información 202 recibida. Los objetos de diálogo pueden mezclarse con más de una señal de mezcla descendente. Los coeficientes de mezcla descendente para cada canal de mezcla descendente pueden codificarse así en el flujo de bits de acuerdo con la siguiente tabla:

Sintaxis del flujo de bits	Coeeficiente de mezcla descendente	Sintaxis del flujo de bits	Coeeficiente de mezcla descendente	Sintaxis del flujo de bits	Coeeficiente de mezcla descendente
0	0	10101	6/15	11011	12/15
10000	1/15	10110	7/15	11100	13/15
10001	2/15	10111	8/15	11101	14/15
10010	3/15	11000	9/15	1111	1
10011	4/15	11001	10/15		
10100	5/15	11010	11/15		

Tabla 1, sintaxis de coeficientes de mezcla descendente

Un flujo de bits que representa los coeficientes de mezcla descendente para un objeto de audio que se mezcla en forma descendente de manera tal que la 5.^a de 7 señales de mezcla descendente comprenda solo el objeto de diálogo, se ve entonces así: 0000111100. Correspondientemente, un flujo de bits que representa los coeficientes de

mezcla descendente para un objeto de audio que mezcla en forma descendente a 1/15 en la quinta señal de mezcla descendente y 14/15 en la séptima señal de mezcla descendente, se ve entonces así: 000010000011101.

5 Con esta sintaxis, lo más frecuente es que se transmita el valor 0, ya que los objetos de diálogo por lo general no se encuentran en todas las señales de mezcla descendente y, muy probablemente, están en una sola señal de mezcla descendente. Por lo tanto, los coeficientes de mezcla descendente se pueden codificar de manera ventajosa mediante la codificación de entropía definida en la tabla anterior. Gastar un bit más en coeficientes distintos de cero y solo 1 para el valor 0 lleva a la longitud de palabra promedio por debajo de los 5 bits para la mayoría de los casos. Por ejemplo, $1/7 * (1 \text{ [bit]} * 6 \text{ [coeficientes]} + 5 \text{ [bit]} * 1 \text{ [coeficiente]}) = 1,57 \text{ bit por coeficiente}$ en promedio, cuando un objeto de diálogo está presente en una de las 7 señales de mezcla descendente. Si se codificaran todos los coeficientes de manera directa con 4 bits, el costo sería $1/7 * (4 \text{ [bits]} * 7 \text{ [coeficientes]}) = 4 \text{ bits por coeficiente}$. Solo si los objetos de diálogo están en 6 o 7 señales de mezcla descendente (de 7 señales de mezcla descendente) es más costoso que una codificación directa. El uso de la codificación de entropía como se describió anteriormente reduce la tasa de bits requerida para transmitir los coeficientes de mezcla descendente.

15 De un modo alternativo, es posible usar la codificación de Huffman para transmitir los coeficientes de mezcla descendente.

De acuerdo con otras realizaciones, la información 202 que describe cómo se mezcló el único objeto como mínimo que representa un diálogo en la pluralidad de señales de mezcla descendente mediante un codificador en el sistema de audio no es recibida por el decodificador, sino que se calcula en la etapa de recepción 104, o en otra etapa apropiada del decodificador 200. Esto reduce la tasa de bits requerida para transmitir el flujo de bits 102 recibido por el decodificador 200. Este cálculo puede basarse en datos con información espacial correspondiente a posiciones espaciales para la pluralidad de señales de mezcla descendente 110 y para el único objeto como mínimo que representa un diálogo. Dichos datos ya suelen ser conocidos por el decodificador 200, ya que normalmente están incluidos en el flujo de bits 102 mediante un codificador en el sistema de audio. El cálculo puede comprender aplicar una función que mapea la posición espacial para el único objeto como mínimo que representa un diálogo en las posiciones espaciales para la pluralidad de señales de mezcla descendente 110. El algoritmo puede ser un algoritmo de panoramización 3D, por ejemplo, un algoritmo de panoramización de amplitud basada en vectores (VBAP). El VBAP es un método para posicionar fuentes de sonido virtuales, por ejemplo, objetos de diálogo, en direcciones arbitrarias, utilizando una configuración de múltiples fuentes de sonido físicas, por ejemplo, altavoces, es decir, la configuración de salida del altavoz. Por lo tanto, tales algoritmos pueden reutilizarse para calcular los coeficientes de mezcla descendente utilizando las posiciones de las señales de mezcla descendente como posiciones de los altavoces.

Utilizando la fórmula de las ecuaciones 1 y 2 anteriores, \mathbf{G} se calcula permitiendo que $\mathbf{rendCoef} = \mathbf{R}(\mathbf{spkPos}, \mathbf{sourcePos})$ —en donde \mathbf{R} un algoritmo de panoramización 3D (por ejemplo, VBAP)— proporcione el vector de coeficiente de renderización $\mathbf{rendCoef} = [\mathbf{nbrSpeakers} \times \mathbf{1}]$ para un objeto de diálogo ubicado en $\mathbf{sourcePos}$ (por ejemplo, coordenadas cartesianas) renderizado en los canales de mezcla descendente $\mathbf{nbrSpeakers}$ ubicados en \mathbf{spkPos} (matriz en la que cada fila corresponde a las coordenadas de una señal de mezcla descendente). Entonces se obtiene \mathbf{G} por:

$$\mathbf{G} = [\mathbf{rendCoef}_1, \mathbf{rendCoef}_2, \dots, \mathbf{rendCoef}_n] \text{ ecuación 5}$$

en la que $\mathbf{rendCoef}_i$ son los coeficientes de renderización para el objeto de diálogo i , de n objetos de diálogo.

40 Dado que la reconstrucción de los objetos de audio se realiza típicamente en un dominio QMF, como se describió anteriormente en conjunto con la figura 1, y que el sonido puede necesitar ser emitido en un dominio de tiempo, el decodificador 200 comprende, además, una etapa de transformación 132, en la que las señales combinadas 214 se transforman en señales 216 en el dominio del tiempo, por ejemplo, aplicando un QMF inverso.

45 De acuerdo con las realizaciones, el decodificador 200 puede comprender, además, una etapa de renderización (no mostrada) arriba de la etapa de transformación 132 o abajo de la etapa de transformación 132. Como se discutió anteriormente, las señales de mezcla descendente, en algunos casos, no corresponden a los canales de la configuración de un altavoz. En tales realizaciones, es beneficioso renderizar las señales de mezcla descendente en ubicaciones correspondientes a los altavoces de la configuración utilizada para la reproducción. Para estas realizaciones, el flujo de bits 102 puede transportar datos de posición para la pluralidad de señales de mezcla descendente 110.

55 En la figura 3, se muestra una realización alternativa de un decodificador de baja complejidad para mejorar el diálogo en un sistema de audio. La principal diferencia entre el decodificador 300 que se muestra en la figura 3 y el decodificador 200 descrito anteriormente es que los objetos 206 con mejora del diálogo reconstruidos no se combinan con las señales de mezcla descendente 110 una vez más después de la etapa de reconstrucción 204. En cambio, el único objeto con mejora de diálogo reconstruido como mínimo 206 se fusiona con las señales de mezcla descendente 110, como una única señal separada como mínimo. La información espacial para el único objeto de diálogo como mínimo, que normalmente ya es conocida por el decodificador 300, como se ha descrito anteriormente, se utiliza para renderizar la señal adicional 206 junto con la renderización de las señales de mezcla

descendente según la información de posición espacial 304 para la pluralidad de señales de mezcla descendente, después o antes de que la señal adicional 206 se haya transformado en el dominio de tiempo, mediante la etapa de transformación 132, como se ha descrito con anterioridad.

5 Para las dos realizaciones del decodificador 200, 300 descritas en conjunto con las figuras 2-3, se debe tener en cuenta que el diálogo ya está presente en la señal de mezcla descendente 110, y que los objetos de diálogo reconstruidos 206 mejorados se agregan a esto, sin importar si se combinan con las señales de mezcla descendente 110, como se describe en conjunto con la figura 2, o si se fusionan con las señales de mezcla descendente 110, como se describe en conjunto con la figura 3. En consecuencia, el parámetro mejorador g_{DE} debe restarse, por ejemplo, en 1 si la magnitud del parámetro mejorador se calcula basándose en que el diálogo existente en las
10 señales de mezcla descendente tiene la magnitud 1.

La figura 4 describe un método 400 para codificar una pluralidad de objetos de audio, que incluye al único objeto como mínimo que representa un diálogo de acuerdo con las realizaciones ejemplares. Debe observarse que el orden de las etapas del método 400 que se muestra en la figura 4 se ilustra sólo como ejemplo.

15 Una primera etapa del método 400 es una etapa opcional para determinar S401 la información espacial correspondiente a las posiciones espaciales para la pluralidad de objetos de audio. Normalmente, el audio del objeto va acompañado de una descripción de dónde se debe renderizar cada objeto. Esto suele hacerse en términos de coordenadas (por ejemplo, cartesianas, polares, etc.).

20 Una segunda etapa del método es la etapa de determinar S402 una pluralidad de señales de mezcla descendente, que es una mezcla descendente de la pluralidad de objetos de audio que incluye al único objeto como mínimo que representa un diálogo. Esto también puede denominarse etapa de mezcla descendente.

Por ejemplo, cada una de las señales de mezcla descendente puede ser una combinación lineal de la pluralidad de objetos de audio. En otras realizaciones, cada banda de frecuencia en una señal de mezcla descendente puede comprender diferentes combinaciones de la pluralidad de objetos de audio. Un sistema de codificación de audio que implementa este método comprende así un componente de mezcla descendente, que determina y codifica las
25 señales de mezcla descendente de los objetos de audio. Las señales de mezcla descendente codificadas pueden ser, por ejemplo, señales envolventes 5.1 o 7.1, que sean retrocompatibles con los sistemas de decodificación de sonido establecidos, como Dolby Digital Plus o los estándares MPEG, como AAC, USAC o MP3, de modo que se logre la AAO.

30 La etapa de determinar S402 una pluralidad de señales de mezcla descendente puede comprender opcionalmente determinar S404 la información que describe cómo se mezcla el único objeto como mínimo que representa un diálogo con la pluralidad de señales de mezcla descendente. En muchas realizaciones, los coeficientes de mezcla descendente siguen el procesamiento en la operación de mezcla descendente. En algunas realizaciones, esto se puede hacer comparando el o los objetos de diálogo con las señales de mezcla descendente utilizando un algoritmo de error cuadrático medio mínimo (MMSE, *minimum mean square error*).

35 Hay muchas maneras de mezclar en forma descendente objetos de audio; por ejemplo, se puede usar un algoritmo que mezcle en forma descendente los objetos que están muy juntos espacialmente. De acuerdo con este algoritmo, se determina en qué posiciones en el espacio hay concentraciones de objetos. Estos se utilizan luego como centroides para las posiciones de señal de mezcla descendente. Esto es solo un ejemplo. Otros ejemplos incluyen mantener los objetos de diálogo separados de los otros objetos de audio, si es posible, al mezclar en forma
40 descendente, para mejorar la separación del diálogo y simplificar aún más la mejora del diálogo del lado del decodificador.

45 El cuarto paso del método 400 es el paso opcional de determinar S406 la información espacial correspondiente a las posiciones espaciales para la pluralidad de señales de mezcla descendente. En el caso de que se haya omitido la etapa opcional de determinar S401 la información espacial correspondiente a las posiciones espaciales para la pluralidad de objetos de audio, la etapa S406 comprende, además, determinar la información espacial correspondiente a las posiciones espaciales para el único objeto como mínimo que representa un diálogo.

La información espacial se conoce típicamente cuando se determina S402 la pluralidad de señales de mezcla descendente como se describió anteriormente.

50 La siguiente etapa en el método es el paso de determinar S408 la información complementaria indicativa de los coeficientes que permiten la reconstrucción de la pluralidad de objetos de audio, a partir de la pluralidad de señales de mezcla descendente. Estos coeficientes también pueden denominarse parámetros de mezcla ascendente. Los parámetros de mezcla ascendente se pueden determinar, por ejemplo, a partir de las señales de mezcla descendente y los objetos de audio, por ejemplo, por optimización MMSE. Los parámetros de mezcla ascendente típicamente comprenden coeficientes de mezcla ascendente seca y coeficientes de mezcla ascendente húmeda. Los
55 coeficientes de mezcla ascendente seca definen un mapeo lineal de la señal de mezcla descendente que se aproxima a las señales de audio a codificar. Los coeficientes de mezcla ascendente seca son, por lo tanto, coeficientes que definen las propiedades cuantitativas de una transformación lineal, que toma las señales de mezcla

descendente como entrada y genera un conjunto de señales de audio que se aproximan a las señales de audio que deben codificarse. El conjunto determinado de coeficientes de mezcla ascendente seca, por ejemplo, puede definir un mapeo lineal de la señal de mezcla descendente correspondiente a una aproximación de error cuadrático medio mínimo de la señal de audio, es decir, entre el conjunto de mapeos lineales de la señal de mezcla descendente, el conjunto determinado de coeficientes de mezcla ascendente seca puede definir el mapeo lineal que mejor se aproxima a la señal de audio en un sentido de mínimo cuadrático medio.

Los coeficientes de mezcla ascendente húmeda se pueden determinar, por ejemplo, en función de una diferencia entre, o comparando, una covarianza de las señales de audio tal como se reciben y una covarianza de las señales de audio según se aproxima mediante el mapeo lineal de la señal de mezcla descendente.

En otras palabras, los parámetros de mezcla ascendente pueden corresponder a elementos de una matriz de mezcla ascendente que permite la reconstrucción de los objetos de audio a partir de las señales de mezcla descendente. Los parámetros de mezcla ascendente se calculan normalmente en función de la señal de mezcla descendente y los objetos de audio, con respecto a los mosaicos de tiempo/frecuencia individuales. Por lo tanto, los parámetros de mezcla ascendente se determinan para cada mosaico de tiempo/frecuencia. Por ejemplo, se puede determinar una matriz de mezcla ascendente (incluidos los coeficientes de mezcla ascendente seca y los coeficientes de mezcla ascendente húmeda) para cada mosaico de tiempo/frecuencia.

La sexta etapa del método para codificar una pluralidad de objetos de audio, que incluye al único objeto como mínimo que representa un diálogo mostrado en la figura 4, es la etapa de determinar S410 los datos que identifican cuál de la pluralidad de objetos de audio representa un diálogo. Normalmente, la pluralidad de objetos de audio puede ir acompañada de metadatos que indican qué objetos contienen diálogo. Alternativamente, se puede usar un detector de voz como se conoce en la técnica.

La etapa final del método descrito es la etapa S412 de formar un flujo de bits que comprende al menos la pluralidad de señales de mezcla descendente, según lo determinado por la etapa de mezcla descendente S402, la información complementaria determinada por la etapa S408, donde se determinan los coeficientes para la reconstrucción y los datos que identifican cuál de la pluralidad de objetos de audio representa un diálogo como se describió anteriormente, en conjunto con la etapa S410. El flujo de bits también puede comprender los datos emitidos o determinados por las etapas opcionales S401, S404, S406, S408 anteriores.

En la figura 5, se muestra un diagrama de bloques de un codificador 500 a modo de ejemplo. El codificador está configurado para codificar una pluralidad de objetos de audio que incluyen al único objeto como mínimo que representa un diálogo, y finalmente para transmitir un flujo de bits 520 que puede ser recibido por cualquiera de los decodificadores 100, 200, 300, como se describe en conjunto con las figuras 1-3 antes explicadas.

El decodificador comprende una etapa de mezcla descendente 503, que comprende un componente de mezcla descendente 504 y un componente de cálculo de parámetros de reconstrucción 506. El componente de mezcla descendente recibe una pluralidad de objetos de audio 502, que incluyen al menos un objeto que representa un diálogo y determina una pluralidad de señales de mezcla descendente 507, que es una mezcla descendente de la pluralidad de objetos de audio 502. Las señales de mezcla descendente pueden ser, por ejemplo, señales envolventes 5.1 o 7.1. Como se describió antes, la pluralidad de objetos de audio 502 puede ser, en realidad, una pluralidad de grupos de objetos 502. Esto significa que, arriba del componente de mezcla descendente 504, puede existir un componente de agrupamiento (no mostrado) que determina una pluralidad de grupos de objetos de una pluralidad mayor de objetos de audio.

El componente de mezcla descendente 504 puede determinar, además, la información 505 que describe cómo se mezcla el único objeto como mínimo que representa un diálogo en la pluralidad de señales de mezcla descendente.

La pluralidad de señales de mezcla descendente 507 y la pluralidad de objetos de audio (o grupos de objetos) son recibidas por el componente que calcula los parámetros de reconstrucción 506, que determina, por ejemplo utilizando una optimización de error cuadrático medio mínimo (MMSE), la información complementaria 509 indicativa de los coeficientes que permiten la reconstrucción de la pluralidad de objetos de audio, a partir de la pluralidad de señales de mezcla descendente. Como se describió anteriormente, la información complementaria 509 suele comprender coeficientes de mezcla ascendente seca y coeficientes de mezcla ascendente húmeda.

El codificador ejemplar 500 puede comprender, además, un componente de codificador de mezcla descendente 508, que puede adaptarse para codificar las señales de mezcla descendente 507, de modo que sean retrocompatibles con los sistemas de decodificación de sonido establecidos, tales como Dolby Digital Plus o los estándares MPEG, tales como AAC, USAC o MP3.

El codificador 500 comprende, además, un multiplexor 518, que combina al menos las señales 510 de mezcla descendente codificadas, la información complementaria 509 y los datos 516 que identifican cuál de la pluralidad de objetos de audio representa un diálogo en un flujo de bits 520. El flujo de bits 520 también puede comprender la información 505 que describe cómo se mezcla el único objeto como mínimo que representa un diálogo en la pluralidad de señales de mezcla descendente que pueden codificarse por codificación de entropía. Además, el flujo

de bits 520 puede comprender información espacial 514 correspondiente a posiciones espaciales para la pluralidad de señales de mezcla descendente y para el único objeto como mínimo que representa un diálogo. Además, el flujo de bits 520 puede comprender información espacial 512 correspondiente a posiciones espaciales para la pluralidad de objetos de audio en el flujo de bits.

- 5 En resumen, esta descripción se inscribe en el campo de la codificación de audio, en particular, se relaciona con el campo de la codificación de audio espacial, donde la información de audio está representada por múltiples objetos de audio que incluyen al menos un objeto de diálogo. En particular, la descripción proporciona un método y un aparato para mejorar el diálogo en un decodificador en un sistema de audio. Además, esta descripción proporciona un método y un aparato para codificar tales objetos de audio, a fin de permitir que el decodificador del sistema de audio mejore el diálogo.

Equivalentes, extensiones, alternativas y varios

- 15 Otras realizaciones de la presente invención resultarán evidentes para una persona experta en la técnica después de estudiar la descripción anterior. Aunque la presente descripción y los dibujos revelan realizaciones y ejemplos, la invención no está restringida a estos ejemplos específicos. Se pueden realizar numerosas modificaciones y variaciones sin apartarse del alcance de la presente invención, que se define en las reivindicaciones adjuntas. Cualquier signo de referencia que aparezca en las reivindicaciones no debe entenderse como una limitación de su alcance.

- 20 Además, el experto en la materia puede comprender y realizar variaciones en las realizaciones descritas al llevar la invención a la práctica, a partir de un estudio de los dibujos, la invención y las reivindicaciones adjuntas. En las reivindicaciones, la frase “que comprende/n” no excluye otros elementos o etapas, y el artículo indefinido “un” o “una” no excluye una pluralidad. El mero hecho de que ciertas medidas se enumeren en reivindicaciones dependientes mutuamente diferentes no indica que no se pueda usar una combinación de estas medidas de un modo ventajoso.

- 25 Los sistemas y métodos descritos aquí con anterioridad pueden implementarse como *software*, *firmware*, *hardware* o una combinación de los mismos. En una implementación de *hardware*, la división de tareas entre las unidades funcionales mencionadas en la descripción anterior no necesariamente corresponde a la división en unidades físicas; por el contrario, un componente físico puede tener múltiples funcionalidades, y una tarea puede ser realizada por varios componentes físicos en cooperación. Ciertos componentes o todos los componentes pueden implementarse como *software* ejecutado por un procesador de señal digital o microprocesador, o implementarse como *hardware* o como un circuito integrado específico de la aplicación. Dicho *software* puede distribuirse en medios legibles por computadora, que pueden comprender medios de almacenamiento informáticos (o medios no transitorios) y medios de comunicación (o medios transitorios). Como es bien conocido por los expertos en la técnica, la frase medios de almacenamiento informático incluye tanto medios volátiles como no volátiles, extraíbles y no extraíbles, implementados en cualquier método o tecnología para el almacenamiento de información, como instrucciones legibles por computadora, estructuras de datos, módulos de programas u otros datos. Los medios de almacenamiento informáticos incluyen, aunque no taxativamente, RAM, ROM, EEPROM, memoria *flash* u otra tecnología de memoria, CD-ROM, discos versátiles digitales (DVD) u otro almacenamiento en disco óptico, casetes magnéticos, cinta magnética, almacenamiento en disco magnético u otros dispositivos de almacenamiento magnético, o cualquier otro medio que pueda usarse para almacenar la información deseada y al que se pueda acceder desde una computadora. Además, los expertos saben muy bien que los medios de comunicación típicamente incorporan instrucciones legibles por computadora, estructuras de datos, módulos de programa u otros datos en una señal de datos modulada, como una onda portadora u otro mecanismo de transporte, e incluye cualquier medio de suministro de información.

REIVINDICACIONES

1. Un método para mejorar el diálogo en un decodificador (100, 200, 300) en un sistema de audio, que comprende las siguientes etapas:
- 5 recibir una pluralidad de señales de mezcla descendente (110), donde las señales de mezcla descendente (110) son una mezcla descendente de una pluralidad de objetos de audio (126), que incluyen al único objeto como mínimo que representa un diálogo,
- recibir información complementaria (108) indicativa de los coeficientes que permiten la reconstrucción de la pluralidad de objetos de audio (126), a partir de la pluralidad de señales de mezcla descendente (110),
- recibir datos (108) que identifican cuál de la pluralidad de objetos de audio representa un diálogo,
- 10 caracterizado por
- modificar los coeficientes utilizando un parámetro mejorador (140) y los datos (108), identificando cuál de la pluralidad de objetos de audio representa un diálogo y
- reconstruir al único objeto como mínimo que representa un diálogo (126, 206) usando los coeficientes modificados (120, 142, 218).
- 15 2. El método según la reivindicación 1, en el que la etapa de modificar los coeficientes usando el parámetro mejorador (140) comprende multiplicar los coeficientes que permiten la reconstrucción del único objeto como mínimo que representa un diálogo con el parámetro mejorador (140).
3. El método según una cualquiera de las reivindicaciones 1-2, que comprende, además, la siguiente etapa:
- 20 calcular los coeficientes que permiten la reconstrucción de la pluralidad de objetos de audio (126) a partir de la pluralidad de señales de mezcla descendente (110) de la información complementaria (108).
4. El método según una cualquiera de las reivindicaciones 1-3, en el que la etapa de reconstruir el único objeto como mínimo que representa un diálogo comprende reconstruir solo el único objeto como mínimo que representa un diálogo.
- 25 5. El método según la reivindicación 4, en el que los coeficientes que permiten la reconstrucción de la pluralidad de objetos de audio comprenden coeficientes de mezcla ascendente seca y coeficientes de mezcla ascendente húmeda, siendo los coeficientes de mezcla ascendente húmeda para mezclar de manera ascendente versiones descorrelacionadas (122) de combinaciones de la pluralidad de señales de mezcla descendente (110), en el que, en la etapa de modificar los coeficientes, solo se modifican los coeficientes de mezcla ascendente seca, y en la que, en la etapa de reconstruir solo el único objeto como mínimo que representa un diálogo, el único objeto como mínimo de
- 30 audio que representa un diálogo se reconstruye a partir de coeficientes de mezcla ascendente seca modificados (218) y la pluralidad de señales de mezcla descendente (110).
6. El método según la reivindicación 4 o 5, que comprende, además, la siguiente etapa:
- combinar las señales de mezcla descendente (110) y el único objeto como mínimo que representa un diálogo (206) reconstruido usando información (202) que describe cómo se mezcló el único objeto como mínimo que representa un diálogo en la pluralidad de señales de mezcla descendente mediante un codificador en el sistema de audio.
- 35 7. El método según la reivindicación 6, que comprende, además, las siguientes etapas:
- renderizar la combinación (214) de las señales de mezcla descendente (110) y el único objeto como mínimo que representa un diálogo (206) reconstruido.
8. El método según la reivindicación 6 o 7, que comprende, además, la siguiente etapa:
- 40 recibir información que describe cómo se mezcló el único objeto como mínimo que representa un diálogo en la pluralidad de señales de mezcla descendente, mediante un codificador en el sistema de audio.
9. El método según la reivindicación 8, en el que la información recibida que describe cómo se mezcló el único objeto como mínimo que representa un diálogo en la pluralidad de señales de mezcla descendente se codifica mediante codificación de entropía.
- 45 10. El método según la reivindicación 6 o 7, que comprende, además, las siguientes etapas:
- recibir datos con información espacial correspondiente a posiciones espaciales para la pluralidad de señales de mezcla descendente (110) y para el único objeto como mínimo que representa un diálogo (206) y

calcular la información que describe cómo se mezcló el único objeto como mínimo que representa un diálogo (206) con la pluralidad de señales de mezcla descendente (110) mediante un codificador en el sistema de audio, en función de los datos con información espacial.

5 11. El método según la reivindicación 10, en el que la etapa de calcular comprende aplicar una función, preferiblemente un algoritmo de panoramización 3D, que mapea la posición espacial para el único objeto como mínimo que representa un diálogo (206), sobre las posiciones espaciales para la pluralidad de señales de mezcla descendente (110).

12. El método según la reivindicación 1, en el que la etapa de reconstruir el único objeto como mínimo que representa un diálogo comprende reconstruir la pluralidad de objetos de audio.

10 13. El método según la reivindicación 12, que comprende, además, las siguientes etapas:

recibir datos (106) con información espacial correspondiente a posiciones espaciales para la pluralidad de objetos de audio (126) y

renderizar la pluralidad reconstruida de objetos de audio (126) sobre la base de los datos (106) con información espacial.

15 14. Un producto de un programa informático que comprende un medio legible por computadora con instrucciones para llevar a cabo el método según una cualquiera de las reivindicaciones 1-13 cuando dicho producto de programa se ejecuta en un ordenador.

15. Un decodificador (100, 200, 300) para mejorar el diálogo en un sistema de audio, decodificador que comprende lo siguiente:

20 una etapa de recepción (104) configurada para lo siguiente:

recibir una pluralidad de señales de mezcla descendente (110), en donde las señales de mezcla descendente son una mezcla descendente de una pluralidad de objetos de audio (126), que incluyen al único objeto como mínimo que representa un diálogo,

25 recibir información complementaria (108) indicativa de los coeficientes que permiten la reconstrucción de la pluralidad de objetos de audio (126), a partir de la pluralidad de señales de mezcla descendente (110) y

recibir datos (108) que identifican cuál de la pluralidad de objetos de audio representa un diálogo,

caracterizado por

una etapa de modificación (112) configurada para lo siguiente:

30 modificar los coeficientes usando un parámetro mejorador (140) y los datos (108) que identifican cuál de la pluralidad de objetos de audio representa un diálogo,

una etapa de reconstrucción (124, 204) configurada para lo siguiente:

reconstruir al menos al único objeto como mínimo que representa un diálogo (126, 206) utilizando los coeficientes modificados (120, 142, 218).

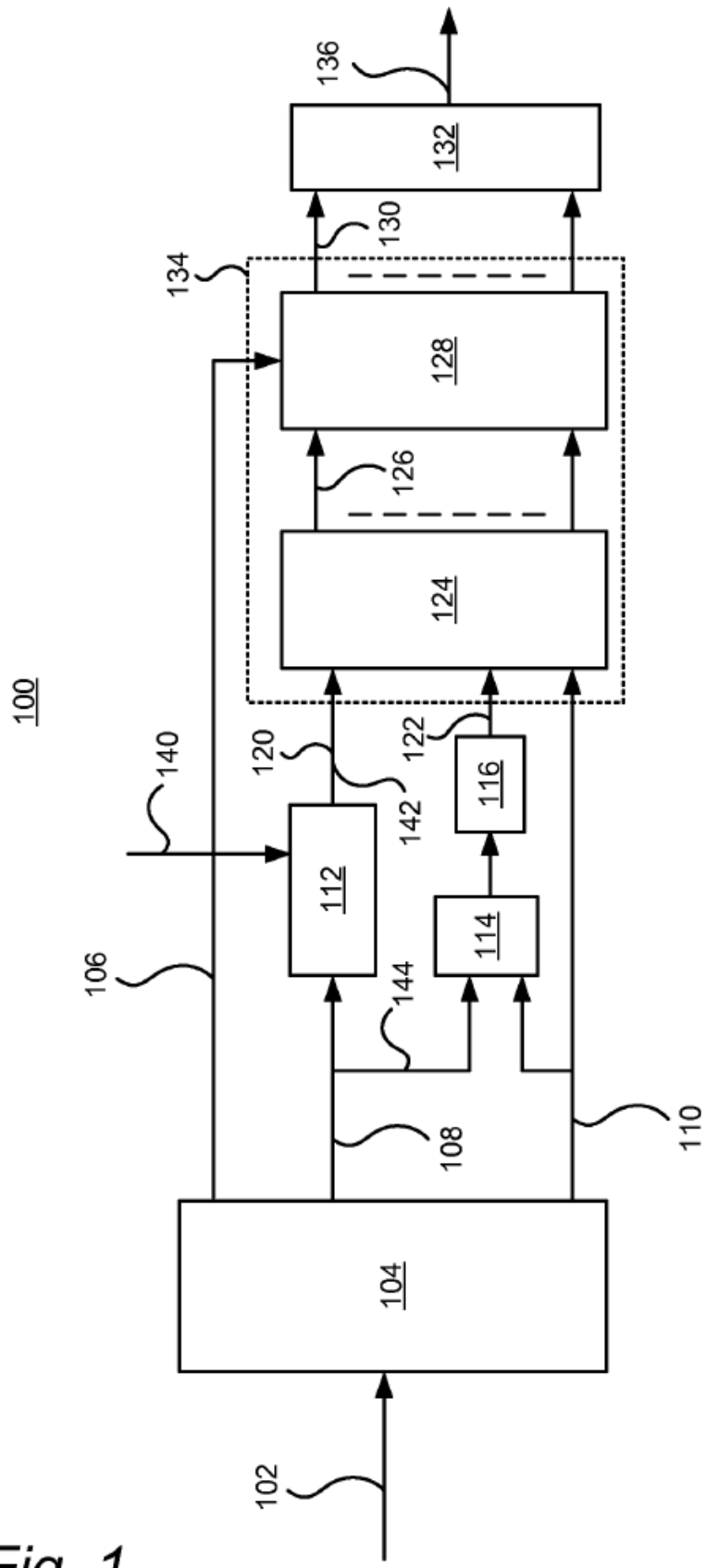


Fig. 1

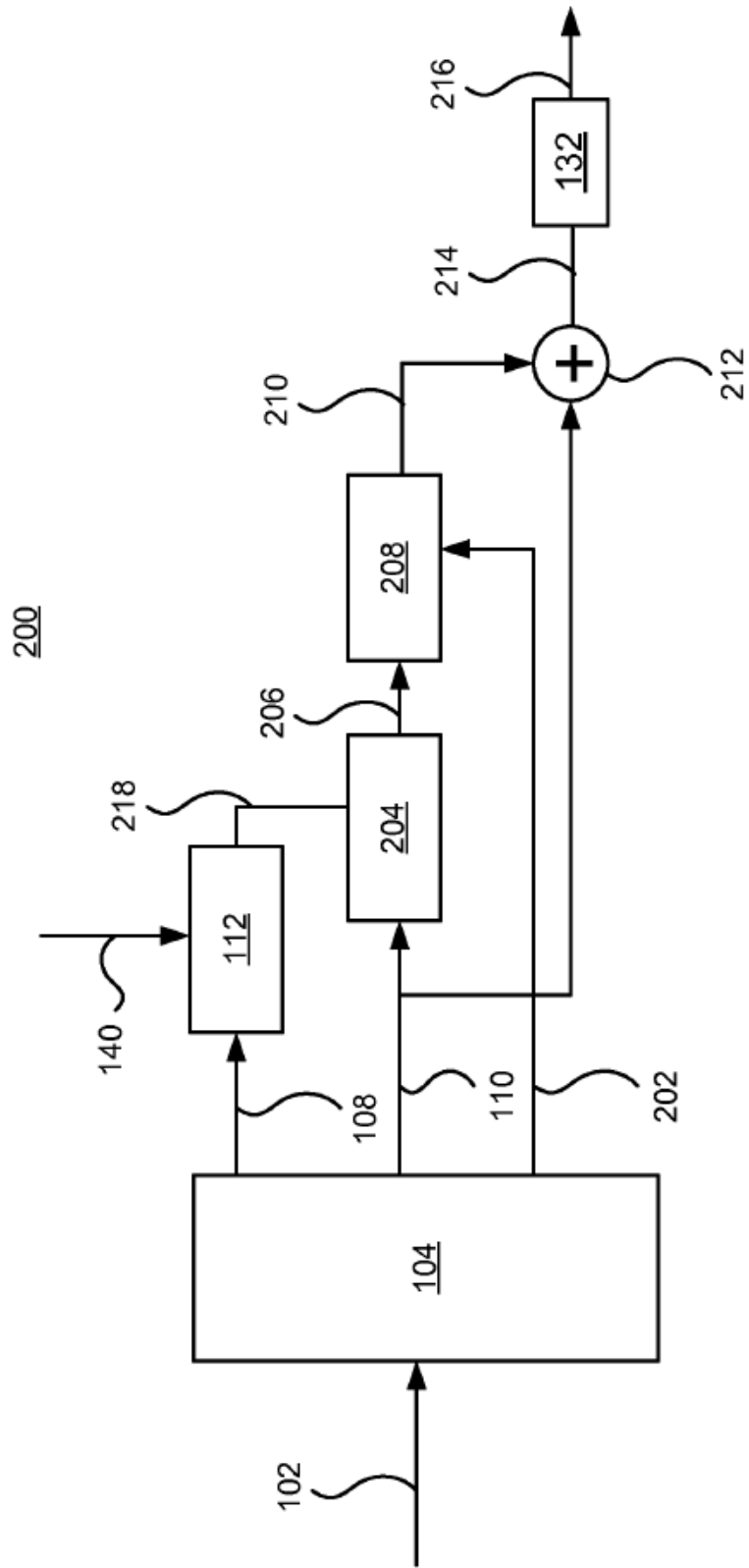


Fig. 2

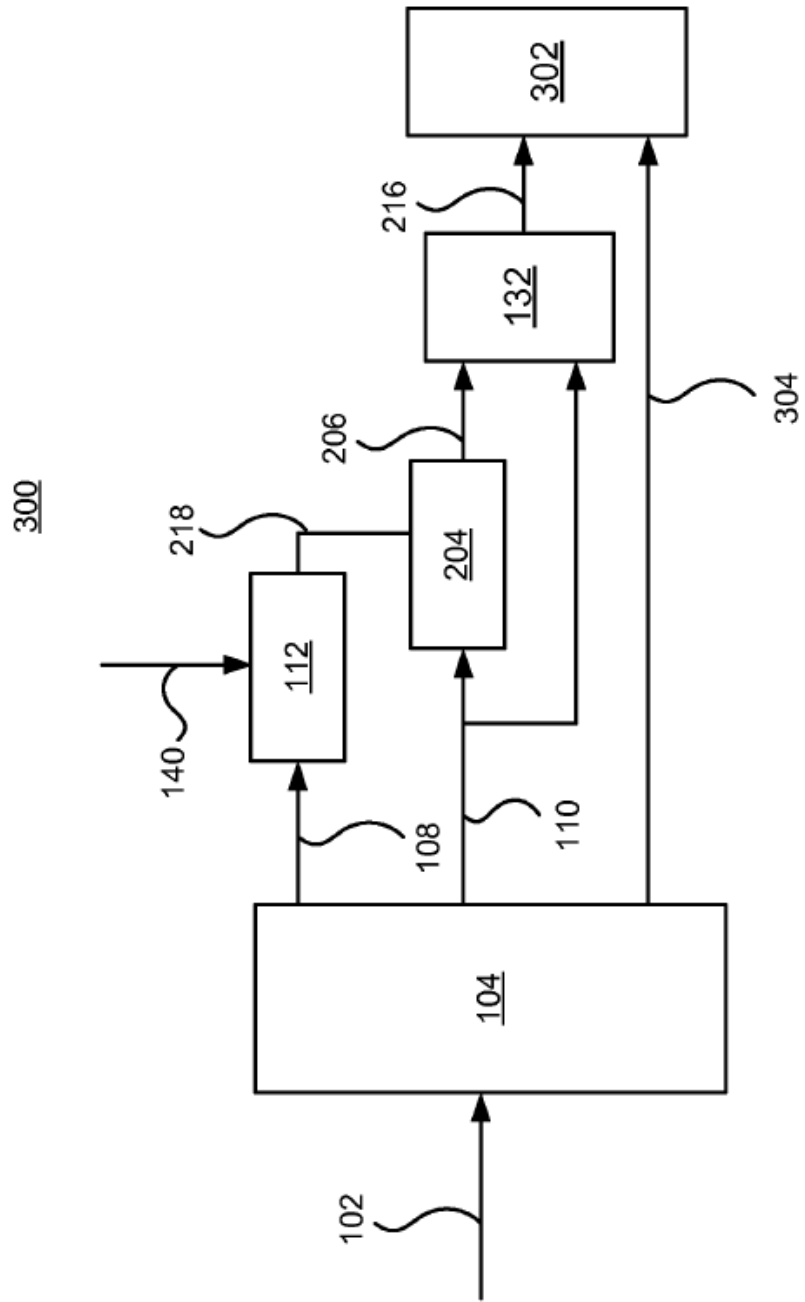


Fig. 3

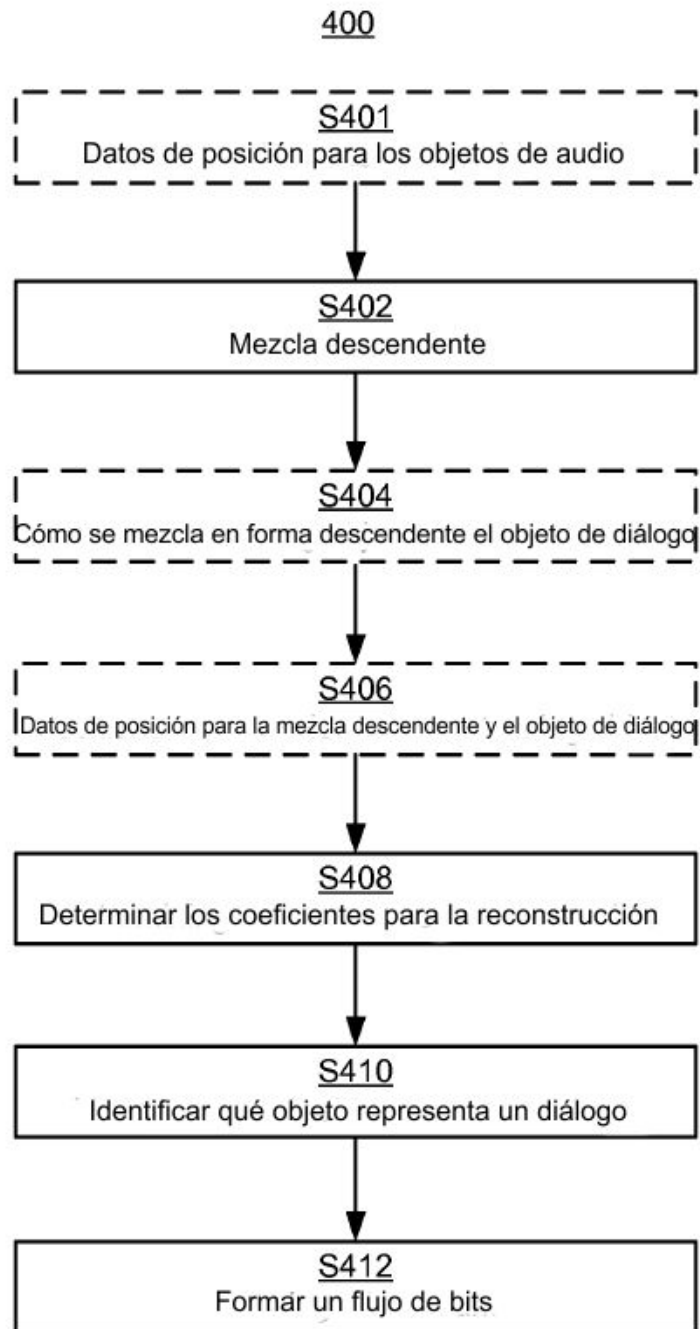


Fig. 4

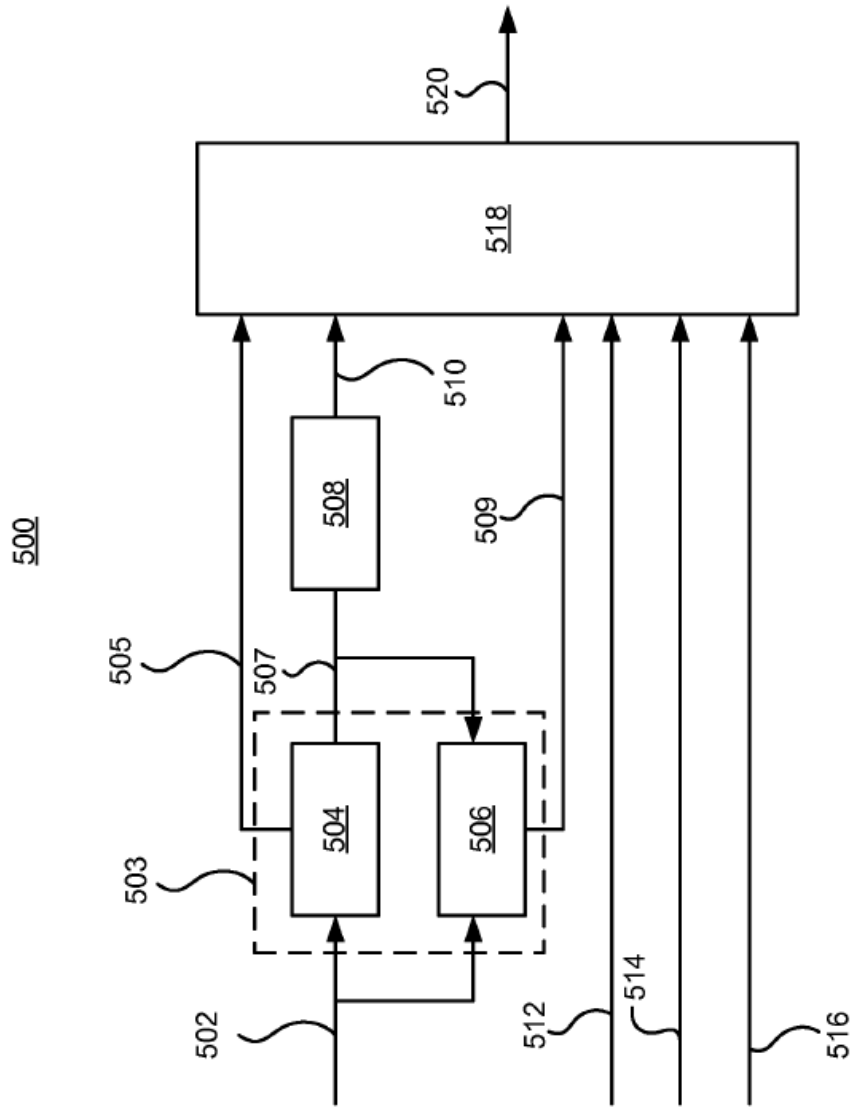


Fig. 5