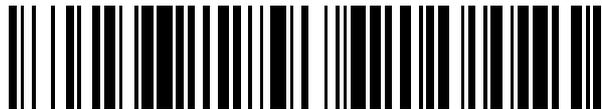


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 784 752**

51 Int. Cl.:

G10L 19/008 (2013.01)

G10L 19/24 (2013.01)

G10L 19/16 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 86 Fecha de presentación y número de la solicitud internacional: **07.10.2016 PCT/EP2016/073970**
- 87 Fecha y número de publicación internacional: **13.04.2017 WO17060411**
- 96 Fecha de presentación y número de la solicitud europea: **07.10.2016 E 16787751 (3)**
- 97 Fecha y número de publicación de la concesión europea: **11.03.2020 EP 3360135**

54 Título: **Codificación en capas para representaciones de sonido o de campo sonido comprimidas**

30 Prioridad:

08.10.2015 EP 15306590
13.07.2016 US 201662361809 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
30.09.2020

73 Titular/es:

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

72 Inventor/es:

KORDON, SVEN y
KRUEGER, ALEXANDER

74 Agente/Representante:

ELZABURU, S.L.P

ES 2 784 752 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

Codificación en capas para representaciones de sonido o de campo sonido comprimidas

Referencia cruzada a aplicaciones relacionadas

5 La presente solicitud reivindica prioridad a la solicitud de patente europea N° 15306590.9 presentada el 8 de Octubre de 2015 y la solicitud de patente de los Estados Unidos N° 62/361.809.

Campo técnico

10 El presente documento se refiere a procedimientos y a aparatos para la codificación de audio en capas. En particular, el presente documento se refiere a procedimientos y a aparatos para la codificación de audio en capas de representaciones de sonido (o campo de sonido) comprimidas, por ejemplo, representaciones de sonido (o campo de sonido) Ambisonics de orden superior (Higher-Order Ambisonics, HOA).

Antecedentes

15 Para la transmisión continua de una representación de sonido (o campo de sonido) a través de un canal de transmisión con condiciones variables en el tiempo, la codificación en capas es un medio para adaptar la calidad de la representación de sonido recibida a las condiciones de transmisión y, en particular, para evitar interrupciones de señal no deseadas.

20 Para la codificación en capas, la representación de sonido (o campo de sonido) normalmente se subdivide en una capa base de alta prioridad de un tamaño relativamente pequeño y capas de mejora adicionales con prioridades decrecientes y tamaños arbitrarios. Típicamente, se supone que cada capa de mejora contiene información incremental para complementar la de todas las capas inferiores con el fin de mejorar la calidad de la representación del sonido (o campo de sonido). La cantidad de protección contra errores para la transmisión de las capas individuales se controla según su prioridad. En particular, la capa base está provista de una alta protección contra errores, lo cual es razonable y asequible debido a su pequeño tamaño.

Sin embargo, existe una necesidad de esquemas de codificación en capas para (versiones extendidas de) tipos especiales de representaciones comprimidas de sonido o campos de sonido, tales como, por ejemplo, sonido HOA comprimido o representaciones de campo de sonido.

25 El presente documento aborda los problemas anteriores. En particular, se describen procedimientos y codificadores/decodificadores para codificación en capas de representaciones de sonido o de campo de sonido comprimidas.

30 El documento EP 2 922 057 A1 describe un procedimiento para comprimir una señal HOA que es una representación HOA de entrada con tramas temporales de entrada $C(k)$ de secuencias de coeficientes HOA que comprende una codificación HOA espacial de las tramas de tiempo de entrada y una codificación perceptual posterior y una codificación fuente.

35 El documento US 2015/248889 A1 describe un formato de codificación de audio en capas con una capa monofónica y al menos una capa de campo de sonido. Se descomponen múltiples señales de audio, según los parámetros de descomposición que controlan las propiedades cuantitativas de una transformación de compactación de energía ortogonal, en señales de audio rotadas. Además, se deriva un perfil de ganancia variable en el tiempo que especifica de manera constructiva cómo pueden procesarse las señales de audio rotadas para atenuar el contenido de audio no deseado. La capa monofónica puede comprender una de las señales rotadas y el perfil de ganancia. La capa de campo de sonido puede comprender las señales rotadas y los parámetros de descomposición. En un ejemplo, el perfil de ganancia comprende un perfil de ganancia de limpieza con el objetivo principal de eliminar componentes que no son de voz y/o el ruido. El perfil de ganancia puede comprender también ganancias de banda ancha mutuamente independientes.

40 Se hace referencia también a Deep Sen et al., "Thoughts on layered/scalable coding for HOA", 110th MPEG meeting, 20-24 de Octubre de 2014, Estrasburgo, ISO/IEC JTC1/SC29/WG11, N° m35160, 15 de Octubre de 2014, y a Erik Hellerud et al., "Spatial redundancy in Higher Order Ambisonics and its use for low delay compression", International Conference on Acoustics, Speech and Signal Processing, 2009, IEEE, 19 de Abril de 2009, págs. 269-272.

Sumario

50 En vista de la necesidad anterior, la invención proporciona un procedimiento para decodificar una representación HOA comprimida de un campo de sonido, un aparato para decodificar una representación HOA comprimida de un campo de sonido y un medio legible por ordenador no transitorio correspondiente, que tienen las características de las reivindicaciones independientes respectivas. Las realizaciones preferidas se describen en las reivindicaciones

dependientes.

Según un ejemplo que es útil para comprender la invención, se describe un procedimiento de codificación en capas de una representación de sonido comprimida de un sonido o un campo de sonido. La representación de sonido comprimida puede incluir una representación de sonido comprimida básica que incluye múltiples componentes. Los múltiples componentes pueden ser componentes complementarios. La representación de sonido comprimida puede incluir además información secundaria básica para decodificar la representación de sonido comprimida básica a una representación de sonido reconstruida básica del sonido o del campo de sonido. La representación de sonido comprimida puede incluir además información secundaria de mejora que incluye parámetros para mejorar (por ejemplo, mejorar) la representación de sonido reconstruida básica. El procedimiento puede incluir subdividir (por ejemplo, agrupar) los múltiples componentes en múltiples grupos de componentes. El procedimiento puede incluir además asignar (por ejemplo, añadir) cada uno de los múltiples grupos a una capa respectiva de entre múltiples capas jerárquicas. La asignación puede indicar una correspondencia entre los grupos y capas respectivos. Puede decirse que los componentes asignados a una capa respectiva están incluidos en esa capa. El número de grupos puede corresponder (por ejemplo, puede ser igual) al número de capas. Las múltiples capas pueden incluir una capa base y una o más capas de mejora jerárquicas. Las múltiples capas jerárquicas pueden estar ordenadas, desde la capa base, pasando por la primera capa de mejora, la segunda capa de mejora, y así sucesivamente, hasta una capa de mejora más alta global (capa más alta global). El procedimiento puede incluir además añadir la información secundaria básica a la capa base (por ejemplo, incluir la información secundaria básica en la capa base, o asignar la información secundaria básica a la capa base, por ejemplo, para los propósitos de transmisión o de almacenamiento). El procedimiento puede incluir además determinar múltiples partes de información secundaria de mejora a partir de la información secundaria de mejora. El procedimiento puede incluir además asignar (por ejemplo, añadir) cada una de las múltiples partes de información secundaria de mejora a una capa respectiva de entre las múltiples capas. Cada parte de información secundaria de mejora puede incluir parámetros para mejorar una representación de sonido reconstruida (por ejemplo, descomprimida) obtenible de los datos incluidos en (por ejemplo, asignados o añadidos a) la capa respectiva y cualquier capa más baja que la capa respectiva. La codificación en capas puede realizarse para propósitos de transmisión a través de un canal de transmisión o para propósitos de almacenamiento en un medio de almacenamiento adecuado, tal como un CD, DVD o Blu-ray Disc™, por ejemplo.

Configurado como se ha indicado anteriormente, el procedimiento propuesto permite aplicar eficientemente la codificación en capas a las representaciones de sonido comprimidas que comprenden múltiples componentes, así como una primera información y una información secundaria de mejora (por ejemplo, información secundaria básica independiente e información secundaria de mejora) que tienen las propiedades establecidas anteriormente. En particular, el procedimiento propuesto garantiza que cada capa incluya información secundaria adecuada para reconstruir una representación de sonido reconstruida a partir de los componentes incluidos en cualquier capa hasta la capa en cuestión. Aquí, se entiende que las capas hasta la capa en cuestión incluyen, por ejemplo, la capa base, la primera capa de mejora, la segunda capa de mejora, y así sucesivamente, hasta la capa en cuestión. De esta manera, independientemente de una capa utilizable más alta real (por ejemplo, la capa debajo de la capa más baja que no se ha recibido de manera válida, de manera que todas las capas debajo de la capa utilizable más alta y la propia capa utilizable más alta hayan sido recibidas de manera válida), se habilitaría un decodificador para mejorar una representación de sonido reconstruida, aunque la representación de sonido reconstruida pueda ser diferente de la representación de sonido completa (por ejemplo, completa). En particular, independientemente de la capa utilizable más alta real, es suficiente que el decodificador decodifique una carga útil de información secundaria de mejora para sólo una única capa (es decir, para la capa utilizable más alta) para mejorar la representación de sonido reconstruida que puede obtenerse en base a todos los componentes incluidos en las capas hasta la capa utilizable más alta real. Es decir, para cada intervalo de tiempo (por ejemplo, trama) solo debe decodificarse una única carga útil de información secundaria de mejora. Por otra parte, el procedimiento propuesto permite aprovechar al máximo la reducción del ancho de banda requerido que puede conseguirse con la aplicación de la codificación en capas.

En algunas implementaciones de este ejemplo, los componentes de la representación de sonido comprimida básica pueden corresponder a señales monoaurales (por ejemplo, señales de transporte o señales de transporte monoaurales). Las señales monoaurales pueden representar señales de sonido predominantes o secuencias de coeficientes de una representación HOA. Las señales monoaurales pueden cuantificarse.

En algunas implementaciones de este ejemplo, la información secundaria básica puede incluir información que especifica la decodificación (por ejemplo, descompresión) de una o más de los múltiples componentes de manera individual, independientemente de otros componentes. Por ejemplo, la información secundaria básica puede representar información secundaria relacionada con las señales monoaurales individuales, independientemente de otras señales monoaurales. De esta manera, la información secundaria básica puede denominarse información secundaria básica independiente.

- 5 En algunas implementaciones de este ejemplo, la información secundaria de mejora puede representar información secundaria de mejora. La información secundaria de mejora puede incluir parámetros de predicción para la representación de sonido comprimida básica para mejorar (por ejemplo, mejorar) la representación de sonido reconstruida básica que puede obtenerse a partir de la representación de sonido comprimida básica y la información secundaria básica.
- 10 En algunas implementaciones de este ejemplo, el procedimiento puede incluir además generar un flujo de transporte para la transmisión de los datos de las múltiples capas (por ejemplo, datos asignados o añadidos a las capas respectivas, o sino incluidos en las capas respectivas). La capa base puede tener la más alta prioridad de transmisión y las capas de mejora jerárquicas pueden tener prioridades de transmisión decrecientes. Es decir, la prioridad de transmisión puede disminuir desde la capa base a la primera capa de mejora, desde la primera capa de mejora a la segunda capa de mejora, y así sucesivamente. Puede controlarse una cantidad de protección contra errores para la transmisión de los datos de las múltiples capas según las prioridades de transmisión respectivas. De esta manera, puede garantizarse que al menos una serie de capas inferiores se transmiten de manera fiable, mientras que, por otra parte, se reduce el ancho de banda global requerido al no aplicar una protección contra errores excesiva a las capas superiores.
- 15 En algunas implementaciones de este ejemplo, el procedimiento puede incluir, además, para cada una de las múltiples capas, generar un paquete de capa de transporte que incluye los datos de la capa respectiva. Por ejemplo, para cada intervalo de tiempo (por ejemplo, trama), puede generarse un paquete de capa de transporte respectivo para cada una de las múltiples capas.
- 20 En algunas implementaciones de este ejemplo, la representación de sonido comprimida puede incluir además información secundaria básica adicional para decodificar la representación de sonido comprimida básica a la representación de sonido reconstruida básica. La información secundaria básica adicional puede incluir información que especifica la decodificación de una o más de los múltiples componentes dependiendo de otros componentes respectivos. El procedimiento puede incluir además descomponer la información secundaria básica adicional en múltiples partes de información secundaria básica adicional. El procedimiento puede incluir además añadir las partes de información secundaria básica adicional a la capa base (por ejemplo, puede incluir las partes de información secundaria básica adicional en la capa base, o asignar las partes de información secundaria básica adicional a la capa base, por ejemplo, para propósitos de transmisión o de almacenamiento). Cada parte de información secundaria básica adicional puede corresponder a una capa respectiva y puede incluir información que especifica la decodificación de uno o más componentes asignados a la capa respectiva que dependen (solo) de otros componentes respectivos asignados a la capa respectiva y cualquier capa más baja que la capa respectiva. Es decir, cada parte de información secundaria básica adicional especifica componentes en la capa respectiva a la que corresponde esa parte de información secundaria básica adicional sin referencia a ningún otro componente asignado a capas superiores a la capa respectiva.
- 25 Configurado de esta manera, el procedimiento propuesto evita la fragmentación de la información secundaria básica adicional añadiendo todas las partes a la capa base. En otras palabras, todas las partes de información secundaria básica adicional se incluyen en la capa base. La descomposición de la información secundaria básica adicional garantiza que para cada capa haya disponible una parte de información secundaria básica adicional que no requiere conocimiento de los componentes en las capas superiores. De esta manera, independientemente de una capa utilizable más alta real, es suficiente que el decodificador decodifique la información secundaria básica adicional incluida en las capas hasta la capa utilizable más alta.
- 30 En algunas implementaciones de este ejemplo, la información secundaria básica adicional puede incluir información que especifica la decodificación (por ejemplo, descompresión) de una o más de los múltiples componentes que dependen de otros componentes. Por ejemplo, la información secundaria básica adicional puede representar información secundaria relacionada con señales monoaurales individuales que dependen de otras señales monoaurales. De esta manera, la información secundaria básica adicional puede denominarse información secundaria básica dependiente.
- 35 En algunas implementaciones de este ejemplo, la representación de sonido comprimida puede procesarse para intervalos de tiempo sucesivos, por ejemplo, intervalos de tiempo de igual tamaño. Los intervalos de tiempo sucesivos pueden ser tramas. De esta manera, el procedimiento puede funcionar sobre tramas, es decir, la representación de sonido comprimida puede codificarse en base a trama. La representación de sonido comprimida puede estar disponible para cada intervalo de tiempo sucesivo (por ejemplo, para cada trama). Es decir, la operación de compresión mediante la cual se ha obtenido la representación de sonido comprimido puede operar en base a trama.
- 40 En algunas implementaciones de este ejemplo, el procedimiento puede incluir además generar información de configuración que indica, para cada capa, los componentes de la representación de sonido comprimida básica asignados a esa capa. De esta manera, el decodificador puede acceder fácilmente a la información necesaria para la decodificación sin realizar un análisis sintáctico innecesario de las cargas útiles de datos recibidas.
- 45
- 50
- 55

Según otro ejemplo que es útil para comprender la invención, se describe un procedimiento de codificación en capas de una representación de sonido comprimida de un sonido o un campo de sonido. La representación de sonido comprimida puede incluir una representación de sonido comprimida básica que incluye múltiples componentes. Los múltiples componentes pueden ser componentes complementarios. La representación de sonido comprimida puede incluir además información secundaria básica (por ejemplo, información secundaria básica independiente) y una tercera información (por ejemplo, información secundaria básica dependiente) para decodificar la representación de sonido comprimida básica a una representación de sonido reconstruida básica del sonido o del campo de sonido. La información secundaria básica puede incluir información que especifica la decodificación de uno o más de los múltiples componentes individualmente, independientemente de otros componentes. La información secundaria básica adicional puede incluir información que especifica la decodificación de uno o más de los múltiples componentes que dependen de otros componentes respectivos. El procedimiento puede incluir subdividir (por ejemplo, agrupar) los múltiples componentes en múltiples grupos de componentes. El procedimiento puede incluir además asignar (por ejemplo, añadir) cada uno de los múltiples grupos a una capa respectiva de las múltiples capas jerárquicas. La asignación puede indicar una correspondencia entre los grupos y las capas respectivos. Puede decirse que los componentes asignados a una capa respectiva están incluidos en esa capa. El número de grupos puede corresponder (por ejemplo, puede ser igual) al número de capas. Las múltiples capas pueden incluir una capa base y una o más capas de mejora jerárquicas. El procedimiento puede incluir además añadir la información secundaria básica a la capa base (por ejemplo, incluir la información secundaria básica en la capa base, o asignar la información secundaria básica a la capa base, por ejemplo, para propósitos de transmisión o de almacenamiento). El procedimiento puede incluir además descomponer la información secundaria básica adicional en múltiples partes de información secundaria básica adicional y añadir las partes de información secundaria básica adicional a la capa base (por ejemplo, incluir las partes de información secundaria básica adicional en la capa base, o asignar las partes de información secundaria básica adicional a la capa base, por ejemplo, para propósitos de transmisión o de almacenamiento). Cada parte de información secundaria básica adicional puede corresponder a una capa respectiva e y puede incluir información que especifica la decodificación de uno o más componentes asignados a la capa respectiva que dependen de otros componentes respectivos asignados a la capa respectiva y cualquier capa más baja que la capa respectiva.

Configurado de esta manera, el procedimiento propuesto asegura que, para cada capa, la información secundaria básica apropiada esté disponible para decodificar los componentes incluidos en cualquier capa hasta la capa respectiva, sin requerir una recepción o decodificación válida (o en general, conocimiento) de cualquier capa superior. En el caso de una representación HOA comprimida, el procedimiento propuesto asegura que, en el modo de codificación vectorial, hay disponible un vector V adecuado para todos los componentes que pertenecen a capas hasta la capa utilizable más alta. En particular, el procedimiento propuesto excluye el caso en el que los elementos de un vector V correspondientes a componentes en capas superiores no estén explícitamente señalizados. Por consiguiente, la información incluida en las capas hasta la capa utilizable más alta es suficiente para decodificar (por ejemplo, descomprimir) cualquier componente que pertenezca a las capas hasta la capa utilizable más alta. De esta manera, se garantiza una descompresión apropiada de las representaciones HOA reconstruidas respectivas para las capas inferiores, incluso si las capas superiores pueden no haber sido recibidas de manera válida por el decodificador. Por otra parte, el procedimiento propuesto permite aprovechar al máximo la reducción del ancho de banda requerido que puede conseguirse cuando se aplica la codificación en capas.

Las implementaciones de este ejemplo pueden estar relacionadas con las implementaciones del ejemplo anterior.

Según otro ejemplo que es útil para comprender la invención, se describe un procedimiento para decodificar una representación de sonido comprimida de un sonido o un campo de sonido. La representación de sonido comprimida puede haber sido codificada en múltiples capas jerárquicas. Las múltiples capas jerárquicas pueden incluir una capa base y una o más capas de mejora jerárquicas. Las múltiples capas pueden tener asignados a las mismas componentes de una representación de sonido comprimida básica de un sonido o un campo de sonido. En otras palabras, las múltiples capas pueden incluir los componentes de la información secundaria comprimida básica. Los componentes pueden asignarse a capas respectivas en grupos de componentes respectivos. Los múltiples componentes pueden ser componentes complementarios. La capa base puede incluir información secundaria básica para decodificar la representación de sonido comprimida básica. Cada capa puede incluir una parte de información secundaria de mejora que incluye parámetros para mejorar una representación de sonido reconstruida básica obtenible a partir de los datos incluidos en la capa respectiva y cualquier capa más baja que la capa respectiva. El procedimiento puede incluir recibir cargas útiles de datos que corresponden respectivamente a las múltiples capas jerárquicas. El procedimiento puede incluir además determinar un primer índice de capa que indica una capa utilizable más alta de entre las múltiples capas a ser usadas para decodificar la representación de sonido comprimida básica a la representación de sonido reconstruida básica del sonido o del campo de sonido. El procedimiento puede incluir además obtener la representación de sonido reconstruida básica a partir de los componentes asignados a la capa utilizable más alta y cualquier capa más baja que la capa utilizable más alta, usando la información secundaria básica. El procedimiento puede incluir además determinar un segundo índice de capa que sea indicativo de qué parte de la información secundaria de mejora debería usarse para mejorar (por ejemplo, mejorar) la representación de sonido reconstruida básica. El procedimiento puede incluir además

obtener una representación de sonido reconstruida del sonido o del campo de sonido a partir de la representación de sonido reconstruida básica, haciendo referencia al segundo índice de capa.

5 Configurado de esta manera, el procedimiento propuesto garantiza que la representación de sonido reconstruida tenga una calidad óptima, usando la información disponible (por ejemplo, recibida de manera válida) en la mejor medida posible.

En las implementaciones de este ejemplo, los componentes de la representación de sonido comprimida básica pueden corresponder a señales monoaurales (por ejemplo, señales de transporte monoaurales). Las señales monoaurales pueden representar señales de sonido predominantes o secuencias de coeficientes de una representación HOA. Las señales monoaurales pueden ser cuantificadas.

10 En las implementaciones de este ejemplo, la información secundaria básica puede incluir información que especifica la decodificación (por ejemplo, descompresión) de uno o más de los múltiples componentes individualmente, independientemente de los otros componentes. Por ejemplo, la información secundaria básica puede representar información secundaria relacionada con señales monoaurales individuales, independientemente de las otras señales monoaurales. De esta manera, la información secundaria básica puede denominarse información secundaria básica independiente.

15 En las implementaciones de este ejemplo, la información secundaria de mejora puede representar información secundaria de mejora. La información secundaria de mejora puede incluir parámetros de predicción para la representación de sonido comprimida básica para mejorar (por ejemplo, mejorar) la representación de sonido reconstruida básica que puede obtenerse a partir de la representación de sonido comprimida básica y la información secundaria básica.

20 En las implementaciones de este ejemplo, el procedimiento puede incluir además determinar, para cada capa, si la capa respectiva se ha recibido de manera válida o no. El procedimiento puede incluir además determinar el primer índice de capa como el índice de capa de una capa inmediatamente debajo de la capa más baja que no ha sido recibida de manera válida.

25 En las implementaciones de este ejemplo, la determinación del segundo índice de capa puede implicar determinar si el segundo índice de capa es igual al primer índice de capa, o determinar un valor de índice como el segundo índice de capa que indica que no se debe usar ninguna información secundaria de mejora cuando se obtiene la representación de sonido reconstruida. En el último caso, la representación de sonido reconstruida puede ser igual a la representación de sonido reconstruida básica.

30 En las implementaciones de este ejemplo, las cargas útiles de datos pueden ser recibidas y procesadas durante intervalos de tiempo sucesivos, por ejemplo, intervalos de tiempo de igual tamaño. Los intervalos de tiempo sucesivos pueden ser tramas. De esta manera, el procedimiento puede funcionar en base a trama. El procedimiento puede incluir, además, si las representaciones de sonido comprimidas para los intervalos de tiempo sucesivos pueden decodificarse independientemente unas de otras, determinar que el segundo índice de capa es igual al primer índice de capa.

35 En las implementaciones de este ejemplo, las cargas útiles de datos pueden ser recibidas y procesadas durante intervalos de tiempo sucesivos, por ejemplo, intervalos de tiempo de igual tamaño. Los intervalos de tiempo sucesivos pueden ser tramas. De esta manera, el procedimiento puede funcionar en base a trama. El procedimiento puede incluir, además, para un intervalo de tiempo determinado entre los intervalos de tiempo sucesivos, si las representaciones de sonido comprimidas para los intervalos de tiempo sucesivos no pueden decodificarse independientemente unas de otras, determinar, para cada capa, si la capa respectiva ha sido recibida de manera válida. El procedimiento puede incluir además determinar el primer índice de capa para el intervalo de tiempo determinado como el más pequeño de entre el primer índice de capa del intervalo de tiempo que precede al intervalo de tiempo determinado y el índice de capa de una capa inmediatamente debajo de la capa más baja que no ha sido recibida de manera válida.

40 En las implementaciones de este ejemplo, el procedimiento puede incluir, además, para el intervalo de tiempo determinado, si las representaciones de sonido comprimidas para los intervalos de tiempo sucesivos no pueden decodificarse independientemente unas de otras, determinar si el primer índice de capa para el intervalo de tiempo determinado es igual al primer índice de capa para el intervalo de tiempo anterior. El procedimiento puede incluir, además, si el primer índice de capa para el intervalo de tiempo determinado es igual al primer índice de capa para el intervalo de tiempo precedente, determinar que el segundo índice de capa para el intervalo de tiempo determinado sea igual al primer índice de capa para el intervalo de tiempo determinado. El procedimiento puede incluir, además, si el primer índice de capa para el intervalo de tiempo determinado no es igual al primer índice de capa para el intervalo de tiempo precedente, determinar un valor de índice como el segundo índice de capa que indica que no se debe usar ninguna información secundaria de mejora cuando se obtiene la representación de sonido reconstruida.

En las implementaciones de este ejemplo, la capa base puede incluir al menos una parte de información secundaria básica adicional correspondiente a una capa respectiva y que incluye información que especifica la decodificación de uno o más componentes de entre los componentes asignados a la capa respectiva que dependen de otros componentes asignados a la capa respectiva y cualquier capa más baja que la capa respectiva. El procedimiento puede incluir, además, para cada parte de información secundaria básica adicional, decodificar la parte de información secundaria básica adicional haciendo referencia a los componentes asignados a su capa respectiva y a cualquier capa más baja que la capa respectiva. El procedimiento puede incluir además corregir la parte de información secundaria básica adicional haciendo referencia a los componentes asignados a la capa utilizable más alta y cualquier capa entre la capa utilizable más alta y la capa respectiva. La representación de sonido reconstruida básica puede obtenerse a partir de los componentes asignados a la capa utilizable más alta y cualquier capa más baja que la capa utilizable más alta, usando la información secundaria básica y partes corregidas de la información secundaria básica adicional obtenida a partir de partes de información secundaria básica adicional correspondientes a las capas hasta la capa utilizable más alta.

En las implementaciones de este ejemplo, la información secundaria básica adicional puede incluir información que especifica la decodificación (por ejemplo, descompresión) de uno o más de los múltiples componentes que dependen de otros componentes. Por ejemplo, la información secundaria básica adicional puede representar información secundaria relacionada con señales monoaurales individuales que dependen de otras señales monoaurales. De esta manera, la información secundaria básica adicional puede denominarse información secundaria básica dependiente.

Según otro ejemplo que es útil para comprender la invención, se describe un procedimiento para decodificar una representación de sonido comprimida de un sonido o un campo de sonido. La representación de sonido comprimida puede haber sido codificada en múltiples capas jerárquicas. Las múltiples capas jerárquicas pueden incluir una capa base y una o más capas de mejora jerárquicas. Las múltiples capas pueden tener asignados a las mismas componentes de una representación de sonido comprimida básica de un sonido o un campo de sonido. En otras palabras, las múltiples capas pueden incluir los componentes de la información secundaria comprimida básica. Los componentes pueden asignarse a capas respectivas en grupos de componentes respectivos. Los múltiples componentes pueden ser componentes complementarios. La capa base puede incluir información secundaria básica para decodificar la representación de sonido comprimida básica. La capa base puede incluir además al menos una parte de información secundaria básica adicional que corresponde a una capa respectiva e incluye información que especifica la decodificación de uno o más componentes de entre los componentes asignados a la capa respectiva que dependen de otros componentes asignados a la capa respectiva y cualquier capa más baja que la capa respectiva. El procedimiento puede incluir recibir cargas útiles de datos que corresponden respectivamente a las múltiples capas jerárquicas. El procedimiento puede incluir además determinar un índice de primera capa que indica una capa utilizable más alta de entre las múltiples capas a ser usadas para decodificar la representación de sonido comprimida básica a la representación de sonido reconstruida básica del sonido o del campo de sonido. El procedimiento puede incluir, además, para cada parte de información secundaria básica adicional, decodificar la parte de información secundaria básica adicional haciendo referencia a los componentes asignados a su capa respectiva y cualquier capa más baja que la capa respectiva. El procedimiento puede incluir, además, para cada parte de información secundaria básica adicional, corregir la parte de información secundaria básica adicional haciendo referencia a los componentes asignados a la capa utilizable más alta y cualquier capa entre la capa utilizable más alta y la capa respectiva. La representación de sonido reconstruida básica puede obtenerse a partir de los componentes asignados a la capa utilizable más alta y cualquier capa más baja que la capa utilizable más alta, usando la información secundaria básica y partes corregidas de la información secundaria básica adicional obtenida a partir de partes de información secundaria básica adicional correspondiente a las capas hasta la capa utilizable más alta. El procedimiento puede comprender además determinar un segundo índice de capa que es igual al primer índice de capa o que indica la omisión de la información secundaria de mejora durante la decodificación.

Configurado de esta manera, el procedimiento propuesto garantiza que la información secundaria básica adicional que se usa eventualmente para decodificar la representación de sonido comprimida básica no incluye elementos redundantes, lo que hace que la decodificación real de la representación de sonido comprimida básica sea más eficiente.

Las implementaciones de este ejemplo pueden estar relacionadas con las implementaciones del ejemplo anterior.

Según otro ejemplo que es útil para comprender la invención, se describe un codificador para la codificación en capas de una representación de sonido comprimida de un sonido o un campo de sonido. La representación de sonido comprimida puede incluir una representación de sonido comprimida básica que incluye múltiples componentes. Los múltiples componentes pueden ser componentes complementarios. La representación de sonido comprimida puede incluir además información secundaria básica para decodificar la representación de sonido comprimida básica a una representación de sonido reconstruida básica del sonido o del campo de sonido. La representación de sonido comprimida puede incluir además información secundaria de mejora que incluye parámetros para mejorar (por ejemplo, mejorar) la representación de sonido reconstruida básica. El codificador puede incluir un procesador configurado para realizar algunas o todas las etapas de procedimiento de los procedimientos según el primer ejemplo indicado anteriormente y el segundo ejemplo indicado anteriormente.

Según otro ejemplo que es útil para comprender la invención, se describe un decodificador para decodificar una representación de sonido comprimida de un sonido o un campo de sonido. La representación de sonido comprimida puede haber sido codificada en múltiples capas jerárquicas. Las múltiples capas jerárquicas pueden incluir una capa base y una o más capas de mejora jerárquicas. Las múltiples capas pueden tener asignados a las mismas componentes de una representación de sonido comprimida básica de un sonido o un campo de sonido. En otras palabras, las múltiples capas pueden incluir los componentes de la información secundaria comprimida básica. Los componentes pueden asignarse a capas respectivas en grupos de componentes respectivos. Los múltiples componentes pueden ser componentes complementarios. La capa base puede incluir información secundaria básica para decodificar la representación de sonido comprimida básica. Cada capa puede incluir una parte de información secundaria de mejora que incluye parámetros para mejorar (por ejemplo, mejorar) una representación de sonido reconstruida básica obtenible a partir de los datos incluidos en la capa respectiva y cualquier capa más baja que la capa respectiva. El decodificador puede incluir un procesador configurado para realizar algunas o todas las etapas de procedimiento de los procedimientos según el tercer ejemplo indicado anteriormente y el cuarto ejemplo indicado anteriormente.

Según otros ejemplos, los procedimientos, aparatos y sistemas se refieren a la decodificación de una representación sonora comprimida Ambisonics de orden superior (HOA) de un sonido o un campo de sonido. El aparato puede tener un receptor configurado para, o el procedimiento puede, recibir una secuencia de bits que contiene la representación HOA comprimida correspondiente a múltiples capas jerárquicas que incluyen una capa base y una o más capas de mejora jerárquicas. Las múltiples capas tienen asignados a las mismas componentes de una representación de sonido comprimida básica del sonido o del campo de sonido, estando asignados los componentes a las capas respectivas en los grupos de componentes respectivos. El aparato puede tener un decodificador configurado para, o el procedimiento puede, decodificar la representación HOA comprimida en base a la información secundaria básica asociada con la capa base y en base a la información secundaria de mejora que está asociada con una o más capas de mejora jerárquicas. La información secundaria básica puede incluir información secundaria independiente básica relacionada con las primeras señales monoaurales individuales que se decodificarán independientemente de otras señales monoaurales. Cada una de entre las una o más capas de mejora jerárquicas puede incluir una parte de la información secundaria de mejora que incluye parámetros para mejorar una representación de sonido reconstruida básica obtenible a partir de los datos incluidos en las capas respectivas y cualquier capa más baja que la capa respectiva.

La información secundaria independiente básica puede indicar que las primeras señales monoaurales individuales representan una señal direccional con una dirección de incidencia. La información secundaria básica puede incluir además información secundaria dependiente básica relacionada con segundas señales monoaurales individuales que serán decodificadas en función de otras señales monoaurales. La información secundaria dependiente básica puede incluir señales basadas en vectores que están distribuidas de manera direccional en el interior del campo de sonido, donde la distribución direccional se especifica por medio de un vector. Los componentes del vector se establecen a cero y no forman parte de la representación vectorial comprimida.

Los componentes de la representación de sonido comprimida básica pueden corresponder a señales monoaurales que representan señales de sonido predominantes o secuencias de coeficientes de una representación HOA. La secuencia de bits incluye cargas útiles de datos que corresponden respectivamente a las múltiples capas jerárquicas. La información secundaria de mejora puede incluir parámetros relacionados con al menos uno de entre: predicción espacial, síntesis de señales direccionales sub-banda y replicación de ambiente paramétrica. La información secundaria de mejora puede incluir información que permite la predicción de partes faltantes del sonido o del campo de sonido a partir de señales direccionales. Además, puede determinarse para cada capa, si la capa respectiva se ha recibido de manera válida y un índice de capa de una capa inmediatamente debajo de una capa más baja que no se ha recibido de manera válida.

Según otro ejemplo, se describe un programa de software. El programa de software puede adaptarse para su ejecución en un procesador y para realizar algunas o todas las etapas de procedimiento descritas en la presente memoria cuando son llevadas a cabo en un dispositivo informático.

Según todavía otro ejemplo, se describe un medio de almacenamiento. El medio de almacenamiento puede comprender un programa de software adaptado para su ejecución en un procesador y para realizar algunas o todas las etapas de procedimiento descritas en la presente memoria cuando se llevan a cabo en un dispositivo informático.

Las afirmaciones realizadas con relación a cualquiera de los aspectos anteriores o sus realizaciones se aplican también a otros aspectos o sus realizaciones respectivas, tal como apreciará la persona experta. La repetición de estas afirmaciones para cada aspecto o realización se ha omitido en aras de la brevedad.

Los procedimientos y los aparatos que incluyen sus realizaciones preferidas tal como se describen en la presente memoria pueden usarse de manera independiente o en combinación con los otros procedimientos y sistemas descritos en la presente memoria. Además, todos los aspectos de los procedimientos y los aparatos descritos en la presente memoria pueden combinarse de manera arbitraria. En particular, las características de las reivindicaciones pueden

combinarse entre sí de manera arbitraria.

Las etapas del procedimiento y las características del aparato pueden intercambiarse de muchas maneras. En particular, los detalles del procedimiento descrito pueden implementarse como un aparato adaptado para ejecutar algunas o todas las etapas del procedimiento, y viceversa, tal como apreciará la persona experta.

5 **Descripción de los dibujos**

La invención se explica a continuación de manera ejemplar con referencia a los dibujos adjuntos, en los que:

La Fig. 1 es un diagrama de flujo que ilustra un ejemplo de un procedimiento de codificación en capas según las realizaciones de la descripción;

10 La Fig. 2 es un diagrama de bloques que ilustra esquemáticamente un ejemplo de una etapa de codificador según las realizaciones de la descripción;

La Fig. 3 es un diagrama de flujo que ilustra un ejemplo de un procedimiento para decodificar una representación de sonido comprimida de un sonido o un campo de sonido que ha sido codificada en múltiples capas jerárquicas, según las realizaciones de la descripción;

15 La Fig. 4A y la Fig. 4B son diagramas de bloques que ilustran esquemáticamente ejemplos de una etapa de decodificador según las realizaciones de la descripción;

La Fig. 5 es un diagrama de bloques que ilustra esquemáticamente un ejemplo de una implementación de hardware de un codificador según las realizaciones de la descripción; y

La Fig. 6 es un diagrama de bloques que ilustra esquemáticamente un ejemplo de una implementación de hardware de un decodificador según las realizaciones de la descripción.

20 **Descripción detallada**

En primer lugar, se describirá una representación de sonido (o campo de sonido) comprimida (en adelante denominada representación de sonido comprimida por brevedad) a la que pueden aplicarse los procedimientos y los codificadores/decodificadores según la presente descripción. En general, la representación de sonido (o campo de sonido) comprimida completa (en adelante denominada representación de sonido comprimida completa por brevedad) puede comprender (por ejemplo, consiste en) los tres componentes siguientes: una representación de sonido (o campo de sonido) comprimida básica (en adelante denominada representación de sonido comprimida básica por brevedad), información secundaria básica e información secundaria de mejora.

30 La propia representación de sonido comprimida básica comprende (por ejemplo, consiste en) una serie de componentes (por ejemplo, componentes complementarios). La representación de sonido comprimida básica puede representar el porcentaje distintivamente mayor de la representación de sonido comprimida completa. La representación de sonido comprimida básica puede consistir en señales de transporte monoaurales que representan señales de sonido predominantes o secuencias de coeficientes de la representación HOA original.

35 La información secundaria básica es necesaria para decodificar la representación de sonido comprimida básica y puede suponerse que tiene un tamaño mucho menor en comparación con la representación de sonido comprimida básica. En su mayor puede comprender parte partes disjuntas, cada una de las cuales especifica la descompresión de solo un componente particular de la representación de sonido comprimida básica. La información secundaria básica puede comprender una primera parte que puede denominarse información secundaria básica independiente y una segunda parte que puede denominarse información secundaria básica adicional.

40 Ambas partes primera y segunda, la información secundaria básica independiente y la información secundaria básica adicional, pueden especificar la descompresión de componentes particulares de la representación de sonido comprimida básica. La segunda parte es opcional y puede omitirse. En este caso, puede decirse que la representación de sonido comprimida comprende la primera parte (por ejemplo, información secundaria básica).

45 La primera parte (por ejemplo, Información secundaria básica) puede contener información secundaria que describe componentes individuales (complementarios) de la representación de sonido comprimida básica independientemente de otros componentes (complementarios). En particular, la primera parte (por ejemplo, información secundaria básica) puede especificar la decodificación de uno o más de los múltiples componentes de manera individual, independientemente de otros componentes. De esta manera, la primera parte puede denominarse información secundaria básica independiente.

La segunda parte (opcional) puede contener información secundaria, denominada también información secundaria

básica adicional, puede describir componentes individuales (complementarios) de la representación de sonido comprimida básica que dependen de otros componentes (complementarios). Esta segunda parte puede denominarse también información secundaria básica dependiente. En particular, la dependencia puede tener las siguientes propiedades:

- 5 – La información secundaria básica dependiente para cada componente individual (complementario) de la representación de sonido comprimida básica puede alcanzar su mayor extensión cuando no hay otros componentes determinados (complementarios) contenidos en la representación de sonido comprimida básica.
- En el caso en el que se añaden determinados componentes (complementarios) adicionales a la representación de sonido comprimida básica, la información secundaria básica dependiente para el componente individual (complementario) considerado puede convertirse en un subconjunto de la información secundaria básica dependiente original, reduciendo de esta manera su tamaño.

La información secundaria de mejora es también opcional. Puede usarse para mejorar (por ejemplo, mejorar de manera paramétrica) la representación de sonido comprimida básica. Puede suponerse también que su tamaño es mucho más pequeño que el de la representación de sonido comprimida básica.

15 De esta manera, en las realizaciones, la representación de sonido comprimida puede comprender una representación de sonido comprimida básica que comprende múltiples componentes, información secundaria básica para decodificar (por ejemplo, descomprimir) la representación de sonido comprimida básica en una representación de sonido reconstruida básica del sonido o del campo de sonido, e información secundaria de mejora que incluye parámetros para mejorar (por ejemplo, mejorar de manera paramétrica) la representación de sonido básica reconstruida. La representación de sonido comprimida puede comprender además información secundaria básica adicional para decodificar (por ejemplo, descomprimir) la representación de sonido comprimida básica a la representación de sonido reconstruida básica, que puede incluir información que especifica la decodificación de uno o más de los múltiples componentes en función de los otros componentes respectivos.

20 Un ejemplo de dicho tipo de representación de sonido comprimida completa lo proporciona la representación de campo de sonido comprimida Ambisonics de orden superior (HOA) tal como se especifica en la versión preliminar del estándar de audio MPEG-H 3D (Referencia 1), Capítulo 12 y Anexo C. 5. Es decir, la representación de sonido comprimida puede corresponder a una representación de sonido (o campo de sonido) HOA comprimido de un sonido o un campo de sonido.

25 Para este ejemplo, la representación de campo de sonido comprimida básica (representación de sonido comprimida básica) puede comprender (por ejemplo, puede identificarse con) una serie de componentes. Los componentes pueden ser (por ejemplo, pueden corresponder a) señales monoaurales. Las señales monoaurales pueden ser señales monoaurales cuantizadas. Las señales monoaurales pueden representar señales de sonido predominantes o secuencias de coeficientes de un componente de campo de sonido ambiente HOA.

30 La información secundaria básica puede describir, entre otras cosas, para cada una de estas señales monoaurales, cómo contribuye espacialmente al campo de sonido. Por ejemplo, la información secundaria básica puede especificar una señal de sonido predominante como una señal puramente direccional, lo que significa una onda plana general con una determinada dirección de incidencia. De manera alternativa, la información secundaria básica puede especificar una señal monoaural como una secuencia de coeficientes de la representación HOA original que tiene un cierto índice. La información secundaria básica puede separarse adicionalmente en una primera parte y una segunda parte, tal como se ha indicado anteriormente.

35 La primera parte es información secundaria (por ejemplo, información secundaria básica independiente) relacionada con señales monoaurales individuales específicas. Esta información secundaria básica independiente es independiente de la existencia de otras señales monoaurales. Dicha información secundaria puede especificar, por ejemplo, una señal monoaural para representar una señal direccional (por ejemplo, que significa una onda plana general) con una determinada dirección de incidencia. De manera alternativa, una señal monoaural puede especificarse como una secuencia de coeficientes de la representación HOA original que tiene un determinado índice. La primera parte puede denominarse información secundaria básica independiente. En general, la primera parte (por ejemplo, información secundaria básica) puede especificar la decodificación de una o más de las múltiples señales monoaurales de manera individual, independientemente de otras señales monoaurales.

40 La segunda parte es información secundaria (por ejemplo, información secundaria básica adicional) relacionada con señales monoaurales individuales específicas. Esta información secundaria depende de la existencia de otras señales monoaurales. Dicha información secundaria puede utilizarse, por ejemplo, si se especifica que las señales monoaurales son señales basadas en vectores (véase, por ejemplo, la Referencia 1, Sección 12.4.2.4.4). Estas señales se distribuyen direccionalmente en el interior del campo de sonido, donde la distribución direccional puede especificarse mediante un

- vector. En cierto modo (ver, por ejemplo, $\text{CodedVVecLength} = 1$), los componentes particulares de este vector se establecen implícitamente a cero y no son parte de la representación vectorial comprimida. Estos componentes son aquellos con índices iguales a los de las secuencias de coeficientes de la representación HOA original y parte de la representación de sonido comprimida básica. Eso significa que, si los componentes individuales del vector están codificados, su número total puede depender de la representación de sonido comprimida básica. En particular, el número total puede depender de qué secuencias de coeficientes contiene la representación HOA original.
- Si ninguna secuencia de coeficientes de la representación HOA original está contenida en la representación de sonido comprimida básica, la información secundaria básica dependiente para cada señal basada en vectores consiste en todos los componentes de vector y tiene su mayor tamaño. En el caso en el que secuencias de coeficientes de la representación HOA original con ciertos índices se añaden a la representación de sonido comprimida básica, los componentes de vector con esos índices se eliminan de la información secundaria para cada señal basada en vectores, reduciendo de esta manera el tamaño de la información secundaria básica dependiente para las señales basadas en vectores.
- La información secundaria de mejora (por ejemplo, información secundaria de mejora) puede comprender parámetros relacionados con la predicción espacial (banda ancha) (véase la Referencia 1, Sección 12.4.2.4.3) y/o los parámetros relacionados con la síntesis de señales direccionales de sub-banda y la replicación de ambiente paramétrica.
- Los parámetros relacionados con la predicción espacial (banda ancha) pueden usarse para predecir (linealmente) las partes faltantes del campo de sonido a partir de las señales direccionales.
- La síntesis de señales direccionales de sub-banda y la replicación de ambiente paramétrica son herramientas de compresión que se han introducido recientemente en el estándar de audio 3D MPEG-H con la modificación [véase Referencia 2, Sección 1]. Estas dos herramientas permiten que una predicción paramétrica dependiente de la frecuencia de señales monoaurales adicionales esté distribuida espacialmente con el fin de complementar una representación HOA comprimida espacialmente incompleta o deficiente. La predicción puede basarse en secuencias de coeficientes de la representación de sonido comprimida básica.
- Es importante señalar que la contribución complementaria indicada anteriormente al campo de sonido está representada en la representación HOA comprimida no mediante señales cuantificadas adicionales, sino más bien mediante información secundaria adicional de un tamaño comparativamente mucho más pequeño. Por lo tanto, las dos herramientas de codificación indicadas son especialmente adecuadas para la compresión de representaciones HOA a bajas velocidades de datos.
- Un segundo ejemplo de una representación comprimida de una o más señales monoaurales con la estructura indicada anteriormente puede comprender información espectral codificada para bandas de frecuencia disjuntas hasta una determinada frecuencia superior, que puede considerarse como una representación comprimida básica; información secundaria básica que especifica la información espectral codificada (por ejemplo, por el número y la anchura de las bandas de frecuencia codificadas); e información secundaria de mejora que comprende (por ejemplo, que consiste en) parámetros de una replicación de banda espectral (SBR), que describe cómo reconstruir paraméricamente a partir de la representación comprimida básica la información espectral para las bandas de frecuencia más altas que no se consideran en la representación comprimida básica.
- La presente descripción propone un procedimiento para la codificación en capas de una representación de sonido (o campo de sonido) comprimida completa que tiene la estructura indicada anteriormente.
- La compresión puede basarse en tramas en el sentido de que proporciona representaciones comprimidas (en la forma de paquetes de datos o cargas de tramas equivalentes) para intervalos de tiempo sucesivos. Los intervalos de tiempo pueden tener tamaños iguales o diferentes. Puede suponerse que estos paquetes de datos contienen un indicador de validez, un valor que indica su tamaño, así como los datos de representación comprimidos reales. A continuación, sin propósitos limitativos, se supondrá que la compresión está basada en tramas. Además, a menos que se indique lo contrario y sin propósitos limitativos, se centrará la atención en el tratamiento de una única trama y, por lo tanto, se omitirá el índice de trama.
- Se supone que cada carga útil de fotogramas de la representación de sonido (o campo de sonido) comprimida completa bajo consideración contiene J paquetes de datos (o cargas útiles de trama), cada uno para un componente de una representación de sonido comprimida básica, que se indican mediante BSRC_j , $j = 1, \dots, J$. Además, se supone que contiene un paquete con información secundaria básica independiente (información secundaria básica) indicada mediante BSI_i que especifica componentes BSRC_j particulares de la representación de sonido comprimida básica independientemente de otros componentes. Opcionalmente, puede suponerse también que contiene un paquete con información complementaria básica dependiente (información secundaria básica adicional) indicada mediante BSI_o que especifica componentes BSRC_j particulares de la representación de sonido comprimida básica que dependen de otros

componentes.

La información contenida dentro de los dos paquetes BSI_I y BSI_D de datos puede agruparse opcionalmente en un único paquete BSI de datos de información secundaria básica. Podría decirse que el único paquete BSI de datos contiene, entre otras cosas, J partes, cada una de las cuales especifica un componente BSRC, particular de la representación de sonido comprimida básica. A su vez, puede decirse que cada una de estas partes contiene una parte de información secundaria independiente y, opcionalmente, una parte de información secundaria dependiente.

Eventualmente, puede incluir una carga útil de información secundaria de mejora (información secundaria de mejora) indicada mediante ESI con una descripción de cómo mejorar el sonido (o campo de sonido) reconstruido a partir de la representación de sonido comprimida básica completa.

La solución propuesta para la codificación en capas aborda las etapas requeridas para habilitar tanto la parte de compresión que incluye el empaquetado de los paquetes de datos para la transmisión como la parte de receptor y de descompresión. Cada parte se describirá en detalle a continuación.

En primer lugar, se describirá la compresión y el empaquetado (por ejemplo, para la transmisión). En particular, se describirán los componentes y los elementos de la representación de sonido (o campo de sonido) comprimida completa en el caso de codificación en capas.

La Fig. 1 ilustra esquemáticamente un diagrama de flujo de un ejemplo de un procedimiento para la compresión y el empaquetado (por ejemplo, un procedimiento de codificación, o un procedimiento de codificación en capas de una representación de sonido comprimida de un sonido o un campo de sonido). La asignación (por ejemplo, asignación) de las cargas útiles individuales a la capa base y (M-1) capas de mejora puede conseguirse mediante un empaquetador de capas de transporte. La Fig. 2 ilustra esquemáticamente un diagrama de bloques de un ejemplo de la asignación de las cargas individuales.

Tal como se ha indicado anteriormente, la representación 2100 de sonido comprimida completa puede estar relacionada, por ejemplo, con una representación HOA comprimida que comprende una representación de sonido comprimida básica. La representación 2100 de sonido comprimida completa puede comprender múltiples componentes (por ejemplo, señales monoaurales) 2110-1,... 2110-J, información 2120 secundaria básica independiente (información secundaria básica), información secundaria 2140 de mejora opcional (información secundaria de mejora) e información 2130 secundaria básica dependiente opcional (información secundaria básica adicional). La información 2120 secundaria básica puede ser información para decodificar la representación sonora comprimida básica a una representación de sonido reconstruida básica del sonido o del campo de sonido. La información 2120 secundaria básica puede incluir información que especifica la decodificación de uno o más componentes (por ejemplo, señales monoaurales) individualmente, independientemente de otros componentes. La información 2140 secundaria de mejora puede incluir parámetros para mejorar (por ejemplo, mejorar) la representación de sonido reconstruida básica. La información 2130 secundaria básica adicional puede ser información (adicional) para decodificar la representación de sonido comprimida básica a la representación de sonido reconstruida básica, y puede incluir información que especifica la decodificación de uno o más de los múltiples componentes en función de otros componentes respectivos.

La Fig. 2 ilustra una suposición subyacente en la que hay múltiples capas jerárquicas, que incluyen una capa base (capa básica) y una o más capas de mejora (jerárquicas). Por ejemplo, puede haber M capas en total, es decir, una capa base y M-1 capas de mejora. Las múltiples capas jerárquicas tienen un índice de capa que se incrementa de manera sucesiva. El valor más bajo del índice de capa (por ejemplo, índice de capa 1) corresponde a la capa base. Se entiende además que las capas están ordenadas, desde la capa base, a través de las capas de mejora, hasta la capa de mejora más alta total (es decir, la capa más alta total).

El procedimiento propuesto puede realizarse en base a trama (es decir, en base a trama). En particular, la representación 2100 de sonido comprimida puede comprimirse durante intervalos de tiempo sucesivos, por ejemplo, intervalos de tiempo de igual tamaño. Cada intervalo de tiempo puede corresponder con una trama. Las etapas que se describen a continuación pueden realizarse para cada intervalo de tiempo sucesivo (por ejemplo, trama).

En S1010 en la Fig. 1, los múltiples componentes 2110 se subdividen en múltiples grupos de componentes. Cada uno de entre los múltiples grupos se asigna a continuación (por ejemplo, se añade o se asigna) a una capa respectiva de entre múltiples capas jerárquicas. Aquí, el número de grupos corresponde al número de capas. Por ejemplo, el número de grupos puede ser igual al número de capas, de manera que haya un grupo de componentes para cada capa. Tal como se ha indicado anteriormente, las múltiples capas pueden incluir una capa base y una o más (por ejemplo, M-1) capas de mejora jerárquicas.

En otras palabras, la representación de sonido comprimida básica se subdivide en partes a ser asignadas a las capas individuales. Sin pérdida de generalidad, la agrupación puede describirse mediante M+1 números J_m , $m = 0, \dots, M$ con $J_0 =$

1 y $J_M = J+1$ de manera que los componentes $BSRC_j$ se asignen a la m -ésima capa para $J_{m-1} \leq j < J_m$.

En S1020, los grupos de componentes se asignan a sus capas respectivas. En S1030, la información 2120 secundaria básica se añade (por ejemplo, se asigna) a la capa base (es decir, la más baja de las múltiples capas jerárquicas).

5 Es decir, debido a su pequeño tamaño, se propone incluir la información secundaria básica completa (información secundaria básica e información secundaria básica adicional opcional) a la capa base para evitar su fragmentación innecesaria.

10 Si la representación de sonido comprimida bajo consideración comprende información secundaria básica dependiente (información secundaria básica adicional), el procedimiento puede comprender además (no mostrado en la Fig. 1) descomponer la información secundaria básica adicional en múltiples partes 2130-1,..., 2130-M de información secundaria básica adicional. Las partes de información secundaria básica adicional pueden añadirse a continuación (por ejemplo, pueden asignarse) a la capa base. En otras palabras, las partes de información secundaria básica adicional pueden incluirse en la capa base. Cada parte de información secundaria básica adicional puede corresponder a una capa respectiva y puede incluir información que especifique la decodificación de uno o más componentes asignados a la capa respectiva en función de otros componentes asignados a la capa respectiva y cualquier capa más baja que la capa respectiva.

15 De esta manera, aunque la información 2120 BSI_i secundaria básica independiente (información secundaria básica) no se modifica para la asignación, la información secundaria básica dependiente debe manipularse especialmente para la codificación en capas, con el fin de permitir una decodificación correcta en el lado del receptor, por una parte, y reducir el tamaño de la información secundaria básica dependiente a ser transmitida por otra parte. Se propone descomponer la información secundaria básica dependiente en M partes (partes) indicadas mediante $BSI_{D,m}$, $m = 1, \dots, M$, donde la m -ésima parte contiene información secundaria básica dependiente para cada uno de los componentes $BSRC_j$, $J_{m-1} \leq j < J_m$, de la representación de sonido comprimida básica asignada a la m -ésima capa, suponiendo que exista la información secundaria básica dependiente opcional para la representación de sonido comprimida bajo consideración. En el caso en el que no exista la información secundaria dependiente respectiva, puede suponerse que la representación de sonido comprimida de las partes $BSI_{D,m}$ está vacía. Cada parte de información $BSI_{D,m}$ secundaria básica dependiente puede depender de todos los componentes $BSRC_j$, $1 \leq j < J_m$, contenidos en todas las capas hasta la m -ésima, (es decir, contenidos en todas las capas $j = 1, \dots, m$).

20 Si el paquete BSI_i de información secundaria básica independiente es de un tamaño negligible, es razonable mantenerlo como un todo y añadirlo (asignarlo) a la capa base. Opcionalmente, una descomposición similar a la usada para la información secundaria básica dependiente puede realizarse también para la información secundaria básica independiente, proporcionando los paquetes $BSI_{I,m} = 1, \dots, M$. Esto es útil para reducir el tamaño de la capa base mediante la adición (asignación) de partes de la información secundaria básica independiente a las capas con los componentes correspondientes de la representación de sonido comprimida básica.

25 En S1040, pueden determinarse múltiples partes 2140-1,..., 2140-M de información secundaria de mejora. Cada parte de información secundaria de mejora puede incluir parámetros para mejorar (por ejemplo, mejorar) una representación de sonido reconstruida obtenible a partir de los datos incluidos en la capa respectiva y cualquier capa más baja que la capa respectiva.

30 La razón para realizar esta etapa es que, en el caso de la codificación en capas, es importante darse cuenta de que la información secundaria de mejora debe calcularse para cada capa adicional, ya que está destinada a mejorar el sonido (o campo de sonido) descomprimido de manera preliminar que, sin embargo, depende de las capas disponibles para la descompresión. En particular, el sonido (o campo de sonido) descomprimido de manera preliminar para una capa decodificable más alta determinada (capa utilizable más alta) depende de los componentes incluidos en la capa decodificable más alta y cualquier capa debajo de la capa decodificable más alta. Por lo tanto, la compresión debe proporcionar M paquetes de datos de información secundaria de mejora individuales (partes de información secundaria de mejora), indicadas mediante ESI_m , $m = 1, \dots, M$, donde la información secundaria de mejora en el m -ésimo paquete ESI_m de datos se calcula para mejorar la representación de sonido (o campo de sonido) obtenida a partir de todos los datos contenidos en la capa base y las capas de mejora con índices más bajos que m (por ejemplo, todos los datos contenidos en la m -ésima capa y cualquier capa debajo de la m -ésima capa).

35 En S1050, las múltiples partes 2140-1,..., 2140-M de información secundaria de mejora se asignan (por ejemplo, se añaden o se asignan) a las múltiples capas. Cada una de las múltiples partes de información secundaria de mejora se asigna a una capa respectiva de las múltiples capas. Por ejemplo, cada una de las múltiples capas incluye una parte de información secundaria de mejora respectiva.

La asignación de información secundaria básica y/o de mejora a las capas respectivas puede indicarse en la información de configuración generada por el procedimiento de codificación. En otras palabras, la correspondencia entre la

información secundaria básica y/o de mejora y las capas respectivas puede indicarse en la información de configuración. Además, la información de configuración puede indicar, para cada capa, los componentes de la representación de sonido comprimida básica que están asignados (por ejemplo, incluidos en) a esa capa. Las partes de información secundaria básica adicional se incluyen en la capa base, pero pueden corresponder a capas distintas de la capa base.

5 En resumen, en la etapa de compresión se proporciona un paquete de datos de trama, indicado mediante FRAME, que tiene la siguiente composición:

$$\text{FRAME} = [\text{BSRC}_1 \dots \text{BSRC}_J \text{ BSI}_1 \text{ BSI}_{D,1} \dots \text{BSI}_{D,M} \text{ ESI}_1 \dots \text{ESI}_M] \quad (1)$$

Además, los paquetes BSI₁ y BSI_{D,m} para m = 1,..., M podrían combinarse en un único paquete BSI en cuyo caso el paquete de datos de trama, indicado mediante FRAME, tendría la siguiente composición:

10
$$\text{FRAME} = [\text{BSRC}_1 \text{ BSRC}_2 \dots \text{BSRC}_J \text{ BSI} \text{ ESI}_1 \text{ ESI}_2 \dots \text{ESI}_M] \quad (2)$$

La ordenación de las cargas útiles individuales con el paquete de datos de trama puede ser generalmente arbitraria.

15 Los paquetes de datos individuales pueden agruparse a continuación en el interior de las cargas útiles, que se definen como paquetes de datos especiales que contienen un indicador de validez, un valor que indica su tamaño, así como los datos de representación comprimidos reales. El uso de cargas útiles permite una demultiplexación simple en el lado del receptor, lo que ofrece la ventaja de poder descartar cargas obsoletas, sin la necesidad de analizar las mismas. Una posible agrupación es proporcionada

- asignando (por ejemplo, asignando) cada paquete BSRC_j, j = 1,...,J, a una carga útil individual indicada mediante \overline{BP}_j .
- asignando (por ejemplo, asignando) el m-ésimo paquete de datos de información secundaria de mejora ESI_m y el m-ésimo paquete de datos de información secundaria dependiente BSI_{D,m} a una carga útil de mejora indicada mediante \overline{EP}_m , m = 1,..., M.
- asignando el paquete de información BSI₁ secundaria básica independiente a una carga útil de información secundaria separada indicada mediante \overline{BSIP} .

25 Opcionalmente, si el tamaño de la información secundaria básica independiente es grande, cada m-ésimo de sus componentes, BSI_{1,m}, m = 1,...,M, puede asignarse (por ejemplo, asignarse) a la carga útil de mejora \overline{EP}_m . En este caso, la carga útil de información secundaria \overline{BSIP} está vacía y puede ser ignorada.

Otra opción es asignar todos los paquetes de datos de información secundaria básicos dependientes BSI_{D,m} a la carga útil de información secundaria \overline{BSIP} , lo cual es razonable si el tamaño de la información secundaria básica dependiente es pequeño.

30 Eventualmente, puede proporcionarse un paquete de datos de trama, indicado mediante FRAME, que tiene la siguiente composición

$$\text{FRAME} = [\overline{BP}_1 \dots \overline{BP}_J \overline{BSIP} \overline{EP}_1 \dots \overline{EP}_M] \quad (3)$$

La ordenación de las cargas individuales con el paquete de datos de trama puede ser generalmente arbitraria.

35 El procedimiento puede comprender además (no mostrado en la Fig. 1) generar, para cada una de las múltiples capas, un paquete de capa de transporte (por ejemplo, un paquete 2200 de capa base y M-1 paquetes 2300-1,..., 2300-(M-1) de capa de mejora) que incluye los datos de la capa respectiva (por ejemplo, componentes, información secundaria básica e información secundaria de mejora para la capa base, o componentes e información secundaria de mejora para una o más capas de mejora).

40 Los paquetes de capa de transporte para diferentes capas pueden tener prioridades de transmisión diferentes. De esta manera, el procedimiento puede comprender además (no mostrado en la Fig. 1), generar una secuencia de transporte para la transmisión de los datos de las múltiples capas, en la que la capa base tiene la prioridad de transmisión más alta y las capas de mejora jerárquicas tienen prioridades de transmisión decrecientes. De esta manera, una mayor prioridad de transmisión puede corresponder a un mayor grado de protección contra errores, y viceversa.

45 A menos que las etapas requieran otras etapas determinadas como requisitos previos, las etapas indicadas anteriormente pueden realizarse en cualquier orden y se entiende que el orden ejemplar ilustrado en la Fig. 1 no es limitativo.

La Fig. 3 ilustra un procedimiento de decodificación de una representación de sonido comprimida de un sonido o un campo de sonido) para decodificar o descomprimir (desempaquetar). Los ejemplos del receptor y de etapa de descompresión correspondientes se ilustran esquemáticamente en los diagramas de bloques de Fig. 4A y la Fig. 4B.

5 Tal como se deduce de lo indicado anteriormente, la representación de sonido comprimida puede codificarse en las múltiples capas jerárquicas. Las múltiples capas pueden tener asignados a las mismas (por ejemplo, pueden incluir) los componentes de la representación de sonido comprimida básica, asignándose los componentes a las capas respectivas en los grupos de componentes respectivos. La capa base puede incluir la información secundaria básica para decodificar la representación de sonido comprimida básica. Cada capa puede incluir una de las partes de información secundaria de mejora indicadas anteriormente que incluyen parámetros para mejorar una representación de sonido reconstruida básica
10 obtenible a partir de los datos incluidos en la capa respectiva y cualquier capa más baja que la capa respectiva.

El procedimiento propuesto puede realizarse en base a trama (es decir, de trama en trama). En particular, puede generarse una representación restaurada del sonido o del campo de sonido para intervalos de tiempo sucesivos, por ejemplo, intervalos de tiempo de igual tamaño. Los intervalos de tiempo pueden ser tramas, por ejemplo. Las etapas que se describen a continuación pueden realizarse para cada intervalo de tiempo sucesivo (por ejemplo, tramas).

15 En S3010, se reciben las cargas útiles de datos (por ejemplo, paquetes de capa de transporte) correspondientes a las múltiples capas. Las cargas útiles de datos pueden recibirse como parte de una secuencia de bits que contiene la representación HOA comprimida de un sonido o un campo de sonido, correspondiendo la representación a las múltiples capas jerárquicas. Las capas jerárquicas incluyen una capa base y una o más capas de mejora jerárquicas. Las múltiples capas tienen asignados a las mismas componentes de una representación de sonido comprimida básica del sonido o del campo de sonido. Los componentes se asignan a capas respectivas en los grupos de componentes respectivos.
20

Los paquetes de capa individuales pueden multiplexarse para proporcionar el paquete de trama recibido de la representación de sonido comprimida completa. El paquete de trama recibido puede indicarse mediante

$$(4) \quad [BSI_1 \quad BSI_{D,1} \quad \dots \quad BSI_{D,M} \quad ESI_1 \quad BSRC_1 \quad \dots \quad BSRC_{(J_1)-1} \quad \dots \quad ESI_M \quad BSRC_{J(M-1)} \quad \dots \quad BSRC_J]$$

25 En el caso alternativo de los paquetes BSI_i y BSI_{D,m} para m = 1,...,M combinados en un único paquete BSI, los paquetes de capa individuales pueden multiplexarse para proporcionar el paquete de trama recibido de la representación de sonido comprimida completa indicada mediante

$$[BSI \quad ESI_1 \quad BSRC_1 \quad \dots \quad BSRC_{(J_1)-1} \quad \dots \quad ESI_M \quad BSRC_{J(M-1)} \quad \dots \quad BSRC_J] \quad (5)$$

En términos de cargas útiles, el paquete de trama recibido puede venir determinado por

$$30 \quad \text{FRAME} = [\overline{BP}_1 \quad \dots \quad \overline{BP}_J \quad \overline{BSIP} \quad \overline{EP}_1 \quad \dots \quad \overline{EP}_M] \quad (6)$$

El paquete de trama recibido puede pasarse a continuación a un descompresor o decodificador 4100. Si la transmisión de una capa individual no ha tenido errores, el indicador de validez de al menos la parte de carga \overline{EP}_m útil de información secundaria de mejora contenida (por ejemplo, correspondiente a una parte de la información secundaria de mejora) se establece a "true/verdadero". En caso de un error debido a la transmisión de una capa individual, el indicador de validez dentro de al menos la carga útil de información secundaria de mejora en esta capa se establece a "false/falso". Por lo tanto, la validez de un paquete de capa puede determinarse a partir de la validez de la carga útil de información secundaria de mejora contenida (por ejemplo, a partir de su indicador de validez).
35

En el descompresor 4100, el paquete de trama recibido puede ser demultiplexado. Para este propósito, la información acerca del tamaño de cada carga útil puede aprovecharse para evitar un análisis sintáctico innecesario a través de los datos de las cargas individuales.
40

En S3020, se determina un primer índice de capa que indica una capa más alta (por ejemplo, capa utilizable más alta o capa decodificable más alta) de entre las múltiples capas a ser usadas para decodificar la representación de sonido comprimida básica a la representación de sonido reconstruida básica del sonido o del campo de sonido.

Además, en S3020, puede seleccionarse el valor (por ejemplo, índice de capa) N_B de la capa más alta (capa utilizable más alta) que se usará para la descompresión de la representación de sonido básica. La capa de mejora más alta a ser usada realmente para la descompresión de la representación de sonido básica viene dada por N_B - 1. Debido a que cada capa contiene exactamente una carga útil de información secundaria de mejora (parte de la información secundaria de mejora), en base a la carga útil de información secundaria de mejora puede determinarse si la capa contenedora es válida o no (por ejemplo, se ha recibido de manera válida). Por lo tanto, la selección puede conseguirse usando todas las
45 cargas ESI_m útiles de información secundaria de mejora, m = 1,...,M (o de manera correspondiente, \overline{EP}_m m = 1,...,M).
50

En S3030, se obtiene una representación de sonido reconstruida básica. La representación de sonido básica reconstruida puede obtenerse a partir de componentes asignados a la capa utilizable más alta indicada por el primer índice de capa y cualquier capa más baja que esta capa utilizable más alta, usando la información secundaria básica (o en general, usando la información secundaria básica).

5 Las cargas útiles de los componentes $BSRC_1, \dots, BSRC_J$ de representación de sonido comprimida básica pueden proporcionarse, junto con (todas) las cargas útiles de información secundaria básica (por ejemplo, BSI o BSI_l y $BSI_{D,m}$, $m = 1, \dots, M$) y el valor N_B , a una unidad 4200 de procesamiento de descompresión de representación básica. La unidad 4200 de procesamiento de descompresión de representación básica (ilustrada en las Figs. 4A y 4B), reconstruye la representación de sonido (o campo de sonido) básica usando solo aquellos componentes de la representación de sonido comprimida básica contenidos dentro de las N_B capas más bajas, es decir, la capa base y $N_B - 1$ capas de mejora (es decir, las capas hasta la capa indicada por el primer índice de capa). De manera alternativa, solo las cargas útiles de los componentes de la representación de sonido comprimida básica contenidos en las N_B capas más bajas junto con las cargas útiles de información secundaria básica respectivas pueden proporcionarse a la unidad 4200 de procesamiento de descompresión de representación básica.

10 La información requerida acerca de qué componentes de la representación de sonido (o campo de sonido) comprimida básica están contenidos en las capas individuales se supone que es conocida por el descompresor 4100 a partir de un paquete de datos con información de configuración, que se supone que se envía y se recibe antes que los paquetes de datos de trama.

15 Con el fin de proporcionar los paquetes $BSI_{D,m}$ de datos de información secundaria dependiente, $m = 1, \dots, N_B$ y el paquete ESI_{NE} de datos de información secundaria de mejora, todas las cargas útiles de mejora pueden introducirse a un analizador 4400 sintáctico parcial (véase la Fig. 4B) del descompresor 4100 junto con el valor N_E y el valor N_B . El analizador sintáctico puede descartar todas las cargas útiles y todos los paquetes de datos que no se usarán para la descompresión real. Si el valor de N_E es igual a cero, puede suponerse que todos los paquetes de datos de información secundaria de mejora están vacíos.

20 Si la capa base incluye al menos una carga útil de información secundaria básica dependiente (parte de información secundaria básica adicional) correspondiente a una capa respectiva, la decodificación de cada carga útil de información secundaria básica dependiente individual (por ejemplo, $BSI_{D,m}$, $m = 1, \dots, N_B$ (parte de información secundaria básica adicional)) puede incluir (i) decodificar la parte de información secundaria básica adicional haciendo referencia a los componentes asignados a su capa respectiva y a cualquier capa más baja que la capa respectiva (decodificación preliminar), y (ii) corregir la parte de información secundaria básica adicional haciendo referencia a los componentes asignados a la capa utilizable más alta y cualquier capa entre la capa utilizable más alta y la capa respectiva (corrección). En este documento, la información secundaria básica adicional correspondiente a una capa respectiva incluye información que especifica la decodificación de uno o más componentes de entre los componentes asignados a la capa respectiva en función de otros componentes asignados a la capa respectiva y cualquier capa más baja que la capa respectiva.

25 A continuación, la representación de sonido básica reconstruida puede obtenerse (por ejemplo, generarse) a partir de los componentes asignados a la capa utilizable más alta y cualquier capa más baja que la capa utilizable más alta, usando la información secundaria básica y partes corregidas de la información secundaria básica adicional obtenida a partir de partes de información secundaria básica adicional correspondiente a las capas hasta la capa utilizable más alta.

30 En particular, la decodificación preliminar de cada carga $BSI_{D,m}$ útil, $m = 1, \dots, N_B$, puede implicar aprovechar su dependencia de los primeros $J_m - 1$ componentes $BSRC_1, \dots, BSRC_{(J_m)-1}$ de la representación de sonido comprimida básica contenidos en las primeras m capas, que se supusieron en la etapa de codificación.

35 La corrección sucesiva de cada carga $BSI_{D,m}$ útil, $m = 1, \dots, N_B$, puede implicar considerar que el componente de sonido básico se reconstruye finalmente a partir de los primeros $J_{N_B} - 1$ componentes $BSRC_1, \dots, BSRC_{(J_{N_B})-1}$ de la representación de sonido comprimida básica contenidos en las primeras $N_B > m$ capas, que son más componentes de los supuestos para la decodificación preliminar. Por lo tanto, la corrección puede conseguirse descartando información obsoleta, lo cual es posible debido a la propiedad supuesta inicialmente de la información secundaria básica dependiente de que, si ciertos componentes complementarios se añaden a la representación de sonido comprimida básica, la información secundaria básica dependiente para cada componente individual (complementario) se convierte en un subconjunto de la original.

40 En S3040, puede determinarse un segundo índice de capa. El segundo índice de capa puede indicar la parte o las partes de información secundaria de mejora que deberían usarse para mejorar (por ejemplo, mejorar) la representación de sonido reconstruida básica.

Además del primer índice de capa, puede determinarse un índice (segundo índice de capa) N_E de la carga útil de información secundaria de mejora (parte de la segunda información de mejora) a ser usado para la descompresión. El segundo índice N_E de capa puede ser siempre igual al primer índice N_B de capa o igual a cero. La mejora puede conseguirse siempre según la representación de sonido básica obtenida a partir de la capa utilizable más alta o nunca.

5 En S3050, se obtiene (por ejemplo, se genera) una representación de sonido reconstruida del sonido o del campo de sonido a partir de la representación de sonido reconstruida básica, haciendo referencia al segundo índice de capa.

Es decir, la representación de sonido reconstruida se obtiene mejorando (paramétricamente) la representación de sonido reconstruida básica, tal como mediante el uso de la información secundaria de mejora (parte de la información secundaria de mejora) indicada por el segundo índice de capa. Tal como se indica más adelante, el segundo índice de capa puede indicar que no debe usarse ninguna información secundaria de mejora en absoluto en esta etapa. Entonces, la representación de sonido reconstruida correspondería a la representación de sonido reconstruida básica.

10 Para este propósito, la representación de sonido básica reconstruida junto con todas las cargas ES_{I_1, \dots, I_M} útiles de información secundaria de mejora, las cargas útiles de información secundaria básica (por ejemplo, BSI o BSI_l y $BSI_{D,m}$, $m = 1, \dots, M$) y el valor N_E se proporciona a una unidad 4300 de procesamiento de descompresión de representación mejorada (ilustrada en las Figs. 4A y 4B), que calcula la representación 2100' de sonido (o campo de sonido) mejorada final usando solo la carga $ES_{I_{N_E}}$ útil de información secundaria de mejora y descarta todas las demás cargas útiles de información secundaria de mejora. De manera alternativa, puede proporcionarse solo la carga $ES_{I_{N_E}}$ útil de información secundaria de mejora en lugar de todas las cargas útiles de información secundaria de mejora a la unidad 4300 de procesamiento de descompresión de representación mejorada. Si el valor de N_E es igual a cero, todas las cargas útiles de información secundaria de mejora se descartan (o, de manera alternativa, no se proporciona carga útil de información secundaria de mejora) y la representación 2100' de sonido mejorada reconstruida final es igual a la representación de sonido básica reconstruida. La carga $ES_{I_{N_E}}$ útil de información secundaria de mejora puede haber sido obtenida por el analizador 4400 sintáctico parcial.

15 La Fig. 3 ilustra también en general la decodificación de la representación HOA comprimida en base a la información secundaria básica asociada con la capa base y en base a la información secundaria de mejora asociada con las una o más capas de mejora jerárquicas.

A menos que las etapas requieran otras etapas determinadas como requisitos previos, las etapas indicadas anteriormente pueden realizarse en cualquier orden y se entiende que el orden ejemplar ilustrado en la Fig. 3 no es limitativo.

20 A continuación, se describirán los detalles de la selección de capa para la descompresión (selección de los índices de capa primero y segundo) en las etapas S3020 y S3040.

La determinación del primer índice de capa puede implicar determinar, para cada capa, si la capa respectiva ha sido recibida de manera válida. La determinación del primer índice de capa puede implicar además determinar el primer índice de capa como el índice de capa de una capa inmediatamente debajo de la capa más baja que no ha sido recibida de manera válida. Puede determinarse si una capa se ha recibido o no de manera válida evaluando si la carga útil de información secundaria de mejora de esa capa se ha recibido de manera válida. A su vez, esto puede realizarse evaluando los indicadores de validez dentro de las cargas útiles de información secundaria de mejora.

25 La determinación del segundo índice de capa puede implicar generalmente determinar si el segundo índice de capa es igual al primer índice de capa o determinar un valor de índice como el segundo índice de capa (por ejemplo, el valor de índice 0) que indica que no debe usarse ninguna información secundaria de mejora cuando al obtener la representación de sonido reconstruida.

En el caso en el que todos los paquetes de datos de trama pueden descomprimirse independientemente unos de los otros, tanto el número N_B de la capa más alta (capa utilizable más alta) a ser usada realmente para la descompresión de la representación de sonido básica y el índice N_E de la carga útil de información secundaria de mejora a ser usada para la descompresión pueden establecerse al número L más alto de una carga útil de información secundaria de mejora válida, el cual puede determinarse evaluando los indicadores de validez dentro de las cargas útiles de información secundaria de mejora. Aprovechando el conocimiento del tamaño de cada carga útil de información secundaria de mejora, puede evitarse un complicado análisis sintáctico de los datos reales de las cargas útiles para determinar su validez.

30 Es decir, puede determinarse que el segundo índice de capa es igual al primer índice de capa si las representaciones de sonido comprimidas para los intervalos de tiempo sucesivos pueden decodificarse de manera independiente. En este caso, la representación de sonido básica reconstruida puede mejorarse en base a la carga útil de información secundaria de mejora de la capa utilizable más alta.

En el caso en el que se emplea una descompresión diferencial con dependencias entre tramas, debe considerarse además la decisión de la trama anterior. Cabe señalar que, con la descompresión diferencial, normalmente, los paquetes de datos de trama independientes se transmiten a intervalos de tiempo regulares con el fin de permitir iniciar la descompresión desde estos instantes de tiempo, donde la determinación de los valores N_B y N_E se vuelve independiente de la trama y se realiza tal como se ha descrito anteriormente.

Para explicar en detalle la decisión dependiente de trama propuesta, el número más alto (por ejemplo, índice de capa) de una carga útil de información secundaria de mejora válida para una k -ésima trama se indica mediante $L(k)$, el número de capa más alto (por ejemplo, índice de capa) a ser seleccionado y usado para la descompresión de la representación de sonido básica mediante $N_B(k)$ y el número (por ejemplo, índice de capa) de la carga útil de información secundaria de mejora a ser usada para la descompresión mediante $N_E(k)$.

Usando esta notación, el número de capa más alto a ser usado para la descompresión de la representación de sonido básica mediante $N_B(k)$ puede calcularse según

$$N_B(k) = \min(N_B(k - 1), L(k)). \quad (7)$$

Eligiendo $N_B(k)$ de manera que no sea mayor que $N_B(k - 1)$ y $L(k)$, se garantiza que toda la información requerida para la descompresión diferencial de la representación de sonido básica esté disponible.

Es decir, si las representaciones de sonido comprimidas para los intervalos de tiempo sucesivos (por ejemplo, tramas) no pueden decodificarse de manera independiente unos de otros, la determinación del primer índice de capa puede comprender determinar, para cada capa, si la capa respectiva se ha recibido de manera válida, y determinar el primer índice de capa para el intervalo de tiempo determinado como el más pequeño de entre el primer índice de capa del intervalo de tiempo que precede al intervalo de tiempo determinado y el índice de capa de una capa inmediatamente debajo de la capa más baja que no se ha recibido de manera válida.

El número $N_E(k)$ de la carga útil de información secundaria de mejora a ser usado para la descompresión puede determinarse según

$$N_E(k) = \begin{cases} N_B(k) & \text{si } N_B(k) = N_B(k - 1) \\ 0 & \text{si no} \end{cases} \quad (8)$$

Aquí, la elección de 0 para $N_E(k)$ indica que la representación de sonido básica reconstruida no debe mejorarse usando información secundaria de mejora.

Esto significa en particular que, mientras el número $N_B(k)$ de capa más alto a ser usado para la descompresión de la representación de sonido básica no cambie, se selecciona el mismo número de capa de mejora correspondiente. Sin embargo, en el caso de un cambio de $N_B(k)$, la mejora se deshabilita estableciendo $N_E(k)$ a cero. Debido a la descompresión diferencial supuesta de la información secundaria de mejora, su cambio según $N_B(k)$ no es posible ya que requeriría la descompresión de la capa de información secundaria de mejora correspondiente en la trama anterior que se supone que no se ha realizado.

Es decir, si las representaciones de sonido comprimidas para los intervalos de tiempo sucesivos (por ejemplo, tramas) no pueden decodificarse de manera independiente unas de otras, la determinación del segundo índice de capa puede comprender determinar si el primer índice de capa para el intervalo de tiempo determinado es igual al primer índice de capa para el intervalo de tiempo precedente. Si el primer índice de capa para el intervalo de tiempo determinado es igual al primer índice de capa para el intervalo de tiempo precedente, el segundo índice de capa para el intervalo de tiempo determinado puede determinarse (por ejemplo, seleccionarse) de manera que sea igual al primer índice de capa para el intervalo de tiempo determinado. Por otra parte, si el primer índice de capa para el intervalo de tiempo determinado no es igual al primer índice de capa para el intervalo de tiempo precedente, puede determinarse (por ejemplo, seleccionarse) un valor de índice como el segundo índice de capa que indica de no debe usarse ninguna información secundaria de mejora al obtener la representación de sonido reconstruida.

De manera alternativa, si en la descompresión, todas las cargas útiles de información secundaria de mejora con números hasta $N_E(k)$ se descomprimen en paralelo, la regla de selección en la ecuación (4) puede reemplazarse por

$$N_E(k) = N_B(k). \quad (9)$$

Finalmente, cabe señalar que, para la descompresión diferencial, el número de la capa N_B utilizada más alta solo puede aumentar en los paquetes de datos de trama independientes, mientras que es posible una disminución en cada trama.

Se entiende que el procedimiento propuesto de codificación en capas de una representación de sonido comprimida puede implementarse mediante un codificador para una codificación en capas de una representación de sonido

comprimida. Dicho codificador puede comprender unidades respectivas adaptadas para realizar las etapas respectivas descritas anteriormente. Un ejemplo de dicho codificador 5000 se ilustra esquemáticamente en la Fig. 5. Por ejemplo, dicho codificador 5000 puede comprender una unidad 5010 de subdivisión de componentes adaptada para realizar la etapa S1010 indicada anteriormente, una unidad 5020 de asignación de componentes adaptada para realizar la etapa S1020 indicado anteriormente, una unidad 5030 de asignación de información secundaria básica adaptada para realizar la etapa S1030 indicada anteriormente, una unidad 5040 de división de información secundaria de mejora adaptada para realizar la etapa S1040 indicada anteriormente, y una unidad 5050 de asignación de información secundaria de mejora adaptada para realizar la etapa S1050 indicada anteriormente. Se entiende además que las unidades respectivas de dicho codificador pueden materializarse en un procesador 5100 de un dispositivo informático que está adaptado para realizar el procesamiento llevado a cabo por cada una de dichas unidades respectivas, es decir, que está adaptado para llevar a cabo algunas o todas las etapas indicadas anteriormente, así como cualquier etapa adicional del procedimiento de codificación propuesto. El codificador o dispositivo informático puede comprender además una memoria 5200 que es accesible por el procesador 5100.

Se entiende además que el procedimiento propuesto de decodificación de una representación de sonido comprimida que está codificada en múltiples capas jerárquicas puede implementarse mediante un decodificador para decodificar una representación de sonido comprimida que está codificada en múltiples capas jerárquicas. Dicho decodificador puede comprender unidades respectivas adaptadas para llevar a cabo las etapas respectivas descritas anteriormente. Un ejemplo de dicho decodificador 6000 se ilustra esquemáticamente en la Fig. 6. Por ejemplo, dicho decodificador 6000 puede comprender una unidad 6010 de recepción adaptada para realizar la etapa S3010 indicada anteriormente, una unidad 6020 de determinación del primer índice de capa adaptada para realizar la etapa S3020 indicada anteriormente, una unidad 6030 de reconstrucción básica adaptada para realizar la etapa S3030 indicada anteriormente, una segunda unidad 6040 de determinación de índice de capa adaptada para realizar la etapa S3040 indicada anteriormente, y una unidad 6050 de reconstrucción mejorada adaptada para realizar la etapa S3050 indicada anteriormente. Se entiende además que las unidades respectivas de dicho decodificador pueden materializarse en un procesador 6100 de un dispositivo informático que está adaptado para realizar el procesamiento llevado a cabo por cada una de dichas unidades respectivas, es decir, que está adaptado para llevar a cabo algunas o todas las etapas indicadas anteriormente, así como cualquier etapa adicional del procedimiento de decodificación propuesto. El decodificador o dispositivo informático puede comprender además una memoria 6200 que es accesible por el procesador 6100.

Cabe señalar que la descripción y los dibujos simplemente ilustran los principios de los procedimientos y aparatos propuestos, mientras que el alcance de la invención está definido únicamente por las reivindicaciones adjuntas. De esta manera, se apreciará que las personas expertas en la materia serán capaces de idear diversas disposiciones que, aunque no se describen o muestran explícitamente en la presente memoria, incorporan los principios de la invención y están incluidas dentro de su alcance. Además, todos los ejemplos mencionados en la presente memoria están destinados principal y expresamente a propósitos pedagógicos para ayudar al lector a comprender los principios de los procedimientos y aparatos propuestos y los conceptos a los que han contribuido los inventores para hacer avanzar la técnica, y deben interpretarse como no limitados a dichos ejemplos y condiciones recitados específicamente. Además, todas las afirmaciones en la presente memoria que mencionan principios, aspectos y realizaciones de la invención, así como ejemplos específicos de la misma, pretenden abarcar equivalentes de la misma.

Los procedimientos y aparatos descritos en el presente documento pueden implementarse como software, firmware y/o hardware. Ciertos componentes pueden implementarse, por ejemplo, como software que se ejecuta en un procesador de señal digital o un microprocesador. Otros componentes pueden implementarse, por ejemplo, como hardware y/o como circuitos integrados específicos de la aplicación. Las señales encontradas en los procedimientos y los aparatos descritos pueden almacenarse en medios tales como memoria de acceso aleatorio o medios de almacenamiento óptico. Pueden transferirse a través de redes, tales como redes de radio, redes satelitales, redes inalámbricas o redes cableadas, por ejemplo, Internet.

Referencia 1: ISO/IEC JTC1/SC29/WG11 23008-3:2015 (E). Information technology - High efficiency coding and media delivery in heterogeneous environments - Part 3: 3D audio, Febrero de 2015.

Referencia 2: ISO/IEC JTC1/SC29/WG11 23008-3: 2015/PDAM3. Information technology - High efficiency coding and media delivery in heterogeneous environments - Part 3: 3D audio, AMENDMENT 3: MPEG-H 3D Audio Phase 2, Julio de 2015.

REIVINDICACIONES

1. Procedimiento de decodificación de una representación (2100) Ambisonics de orden superior, HOA, comprimida de un sonido o un campo de sonido, comprendiendo el procedimiento:

5 recibir una secuencia de bits que contiene la representación (2100) HOA comprimida correspondiente a múltiples capas jerárquicas que incluyen una capa base y dos o más capas de mejora jerárquicas, y que contiene información (2120) secundaria básica que está asociada con la capa base e información (2140) secundaria de mejora que está asociada con las dos o más capas de mejora jerárquicas,

10 en el que las múltiples capas tienen asignados a las mismas componentes de una representación de sonido comprimida básica del sonido o del campo de sonido, asignándose los componentes a las capas respectivas en grupos de componentes respectivos, y

en el que las dos o más capas de mejora jerárquicas comprenden una capa de mejora jerárquica utilizable más alta,

caracterizado por que

15 cada una de las dos o más capas de mejora jerárquicas incluye una parte de la información (2140) secundaria de mejora que incluye parámetros para mejorar una representación de sonido reconstruida básica obtenible a partir de los datos incluidos en la capa respectiva y cualquier capa más baja que la capa respectiva; y

20 el procedimiento comprende además decodificar la representación (2100) HOA comprimida en base a la información (2120) secundaria básica que está asociada con la capa base, en base a la parte de la información (2140) secundaria de mejora que está asociada con la capa de mejora jerárquica utilizable más alta, y no en base a la parte de la información (2140) secundaria de mejora que está asociada con cualquier otra capa de las dos o más capas de mejora jerárquicas.

2. Procedimiento según la reivindicación 1, en el que la información (2140) secundaria de mejora incluye parámetros relacionados con al menos una de entre: predicción espacial, síntesis de señales direccionales sub-banda y replicación de ambiente paramétrica; y/o

25 en el que la información (2140) secundaria de mejora incluye información que permite la predicción de partes faltantes del sonido o del campo de sonido a partir de señales direccionales.

3. Procedimiento según cualquiera de las reivindicaciones 1-2, que comprende, además:

determinar, para cada capa, si la capa respectiva se ha recibido o no de manera válida; y

30 determinar un índice de capa de una capa inmediatamente debajo de una capa más baja que no se ha recibido de manera válida.

4. Aparato (6000) de decodificación de una representación Ambisonics de orden superior, HOA, comprimida, de un sonido o campo de sonido, comprendiendo el aparato (6000):

35 un receptor (6010) para recibir una secuencia de bits que contiene la representación (2100) HOA comprimida correspondiente a múltiples capas jerárquicas que incluyen una capa base y dos o más capas de mejora jerárquicas, y que contiene información (2120) secundaria básica que está asociada con la capa base y la información (2140) secundaria de mejora que está asociada con las dos o más capas de mejora jerárquicas,

40 en el que las múltiples capas tienen asignados a las mismas componentes de una representación de sonido comprimida básica del sonido o del campo de sonido, asignándose los componentes a las capas respectivas en grupos de componentes respectivos, y

en el que las dos o más capas de mejora jerárquicas comprenden una capa de mejora jerárquica utilizable más alta,

caracterizado por que

45 cada una de las dos o más capas de mejora jerárquicas incluye una parte de la información (2140) secundaria de mejora que incluye parámetros para mejorar una representación de sonido reconstruida básica obtenible a partir de los datos incluidos en las capas respectivas y cualquier capa más baja que la

capa respectiva; y

5 el aparato (600) comprende además un decodificador (6020, 6030, 6040, 6050) para decodificar la representación (2100) HOA comprimida en base a la información (2120) secundaria básica que está asociada con la capa base, en base a la parte de la información (2140) secundaria de mejora que está asociada con la capa de mejora jerárquica utilizable más alta, y no en base a la parte de la información (2140) secundaria de mejora que está asociada con cualquier otra capa de las dos o más capas de mejora jerárquicas.

10 5. Aparato (6000) según la reivindicación 4, en el que la información (2140) secundaria de mejora incluye parámetros relacionados con al menos uno de entre: predicción espacial, síntesis de señales direccionales sub-banda y replicación de ambiente paramétrica; y/o

en el que la información (2140) secundaria de mejora incluye información que permite la predicción de partes faltantes del sonido o del campo de sonido a partir de señales direccionales.

6. Aparato (6000) según cualquiera de las reivindicaciones 4-5, configurado para:

determinar, para cada capa, si la capa respectiva se ha recibido de manera válida o no; y

15 determinar un índice de capa de una capa inmediatamente debajo de una capa más baja que no se ha recibido de manera válida.

7. Medio no transitorio legible por ordenador que comprende instrucciones interpretables por ordenador que, cuando son ejecutadas por uno o más procesadores de un dispositivo informático, causan que el dispositivo informático realice el procedimiento según una cualquiera de las reivindicaciones 1 a 3.

20

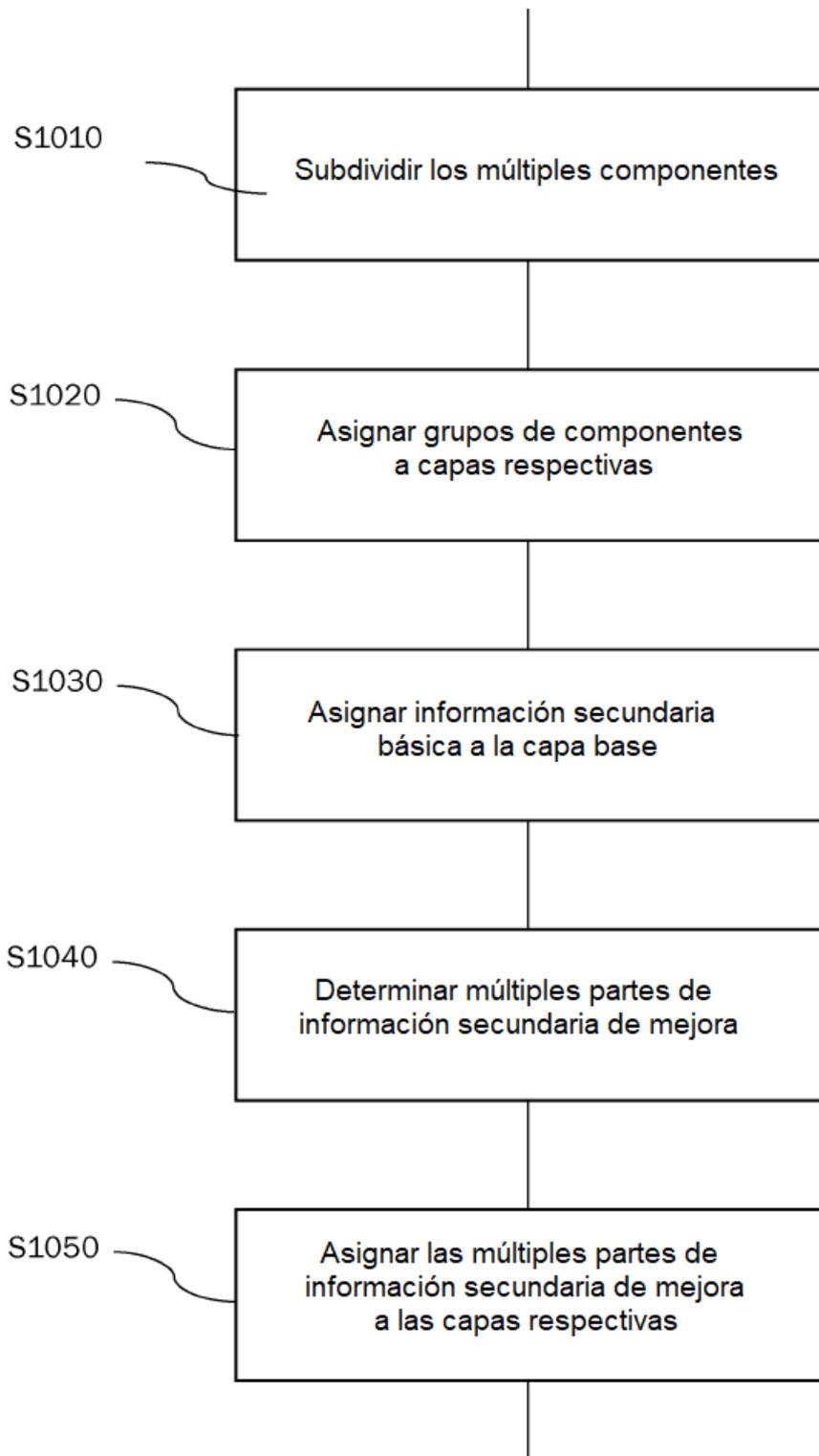


Fig. 1

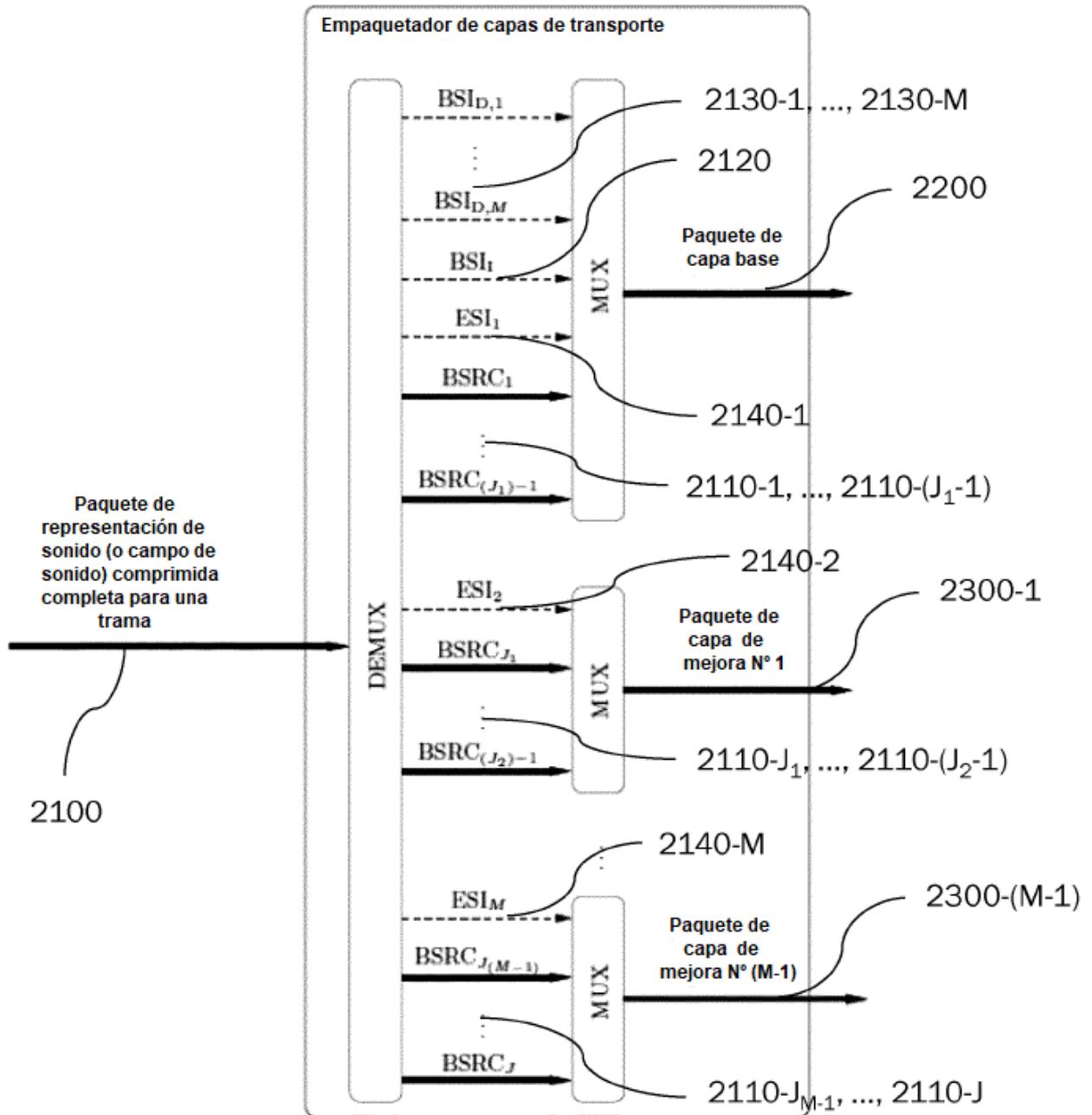


Fig. 2

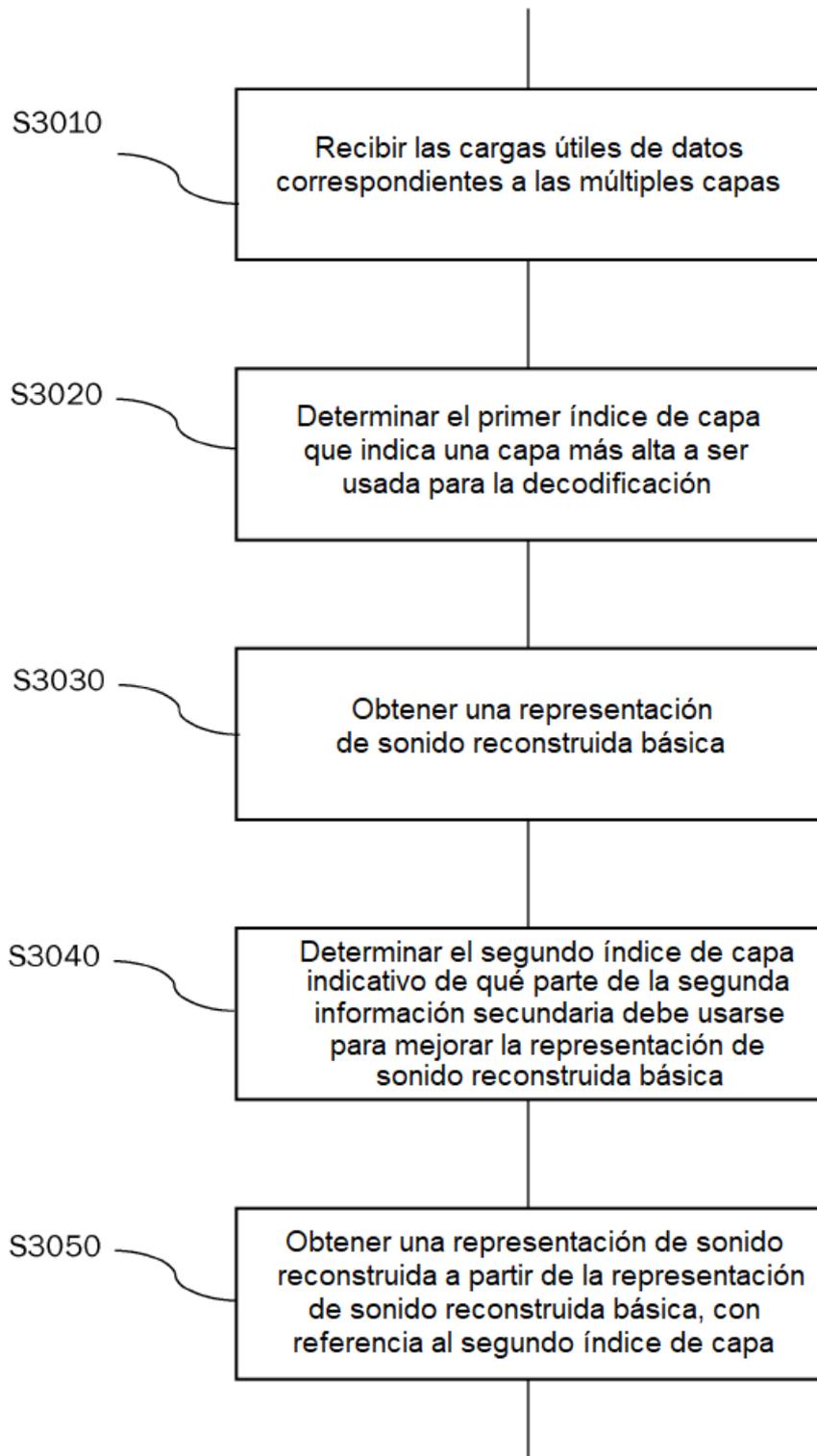


Fig. 3

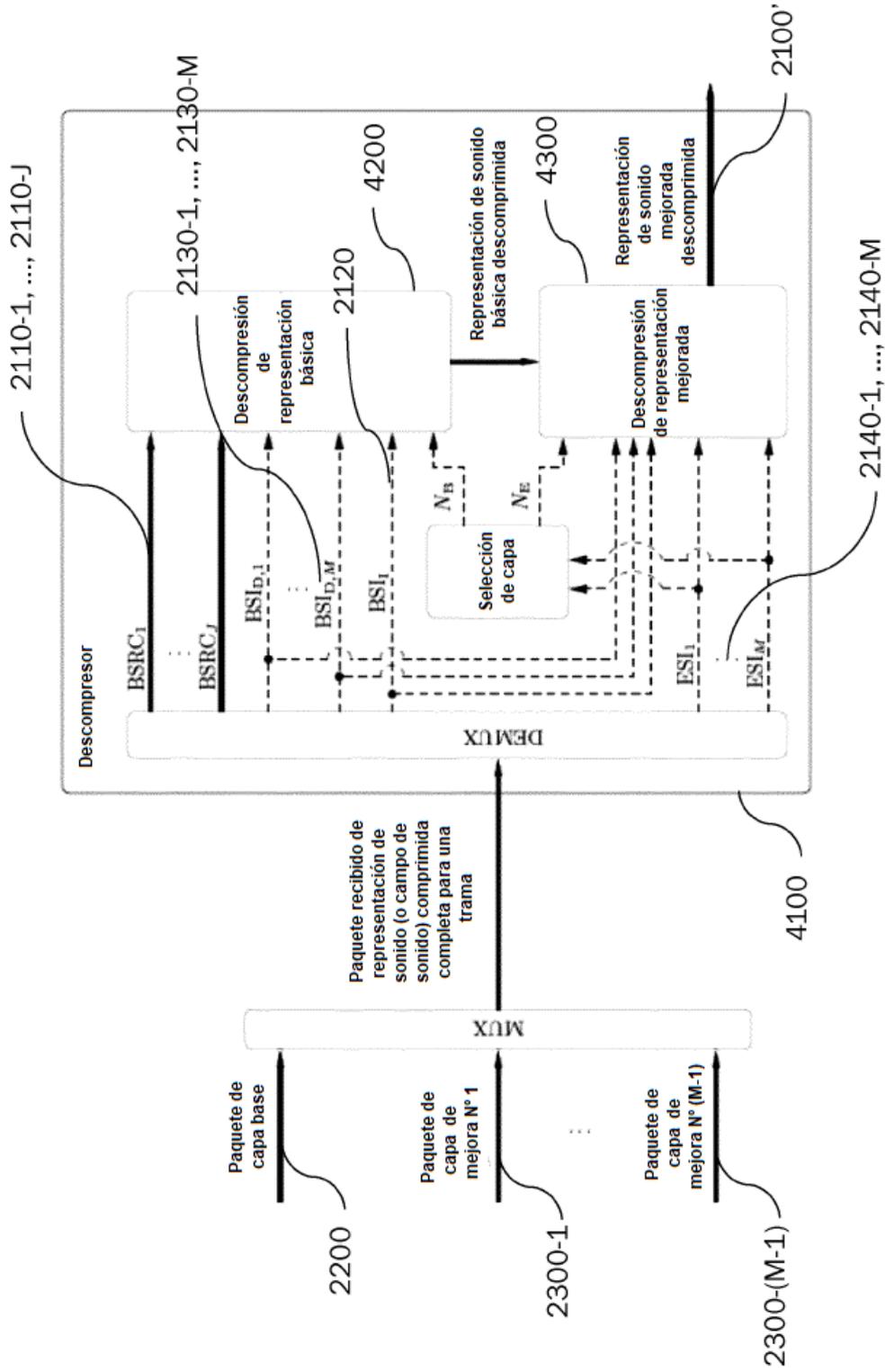


Fig. 4A

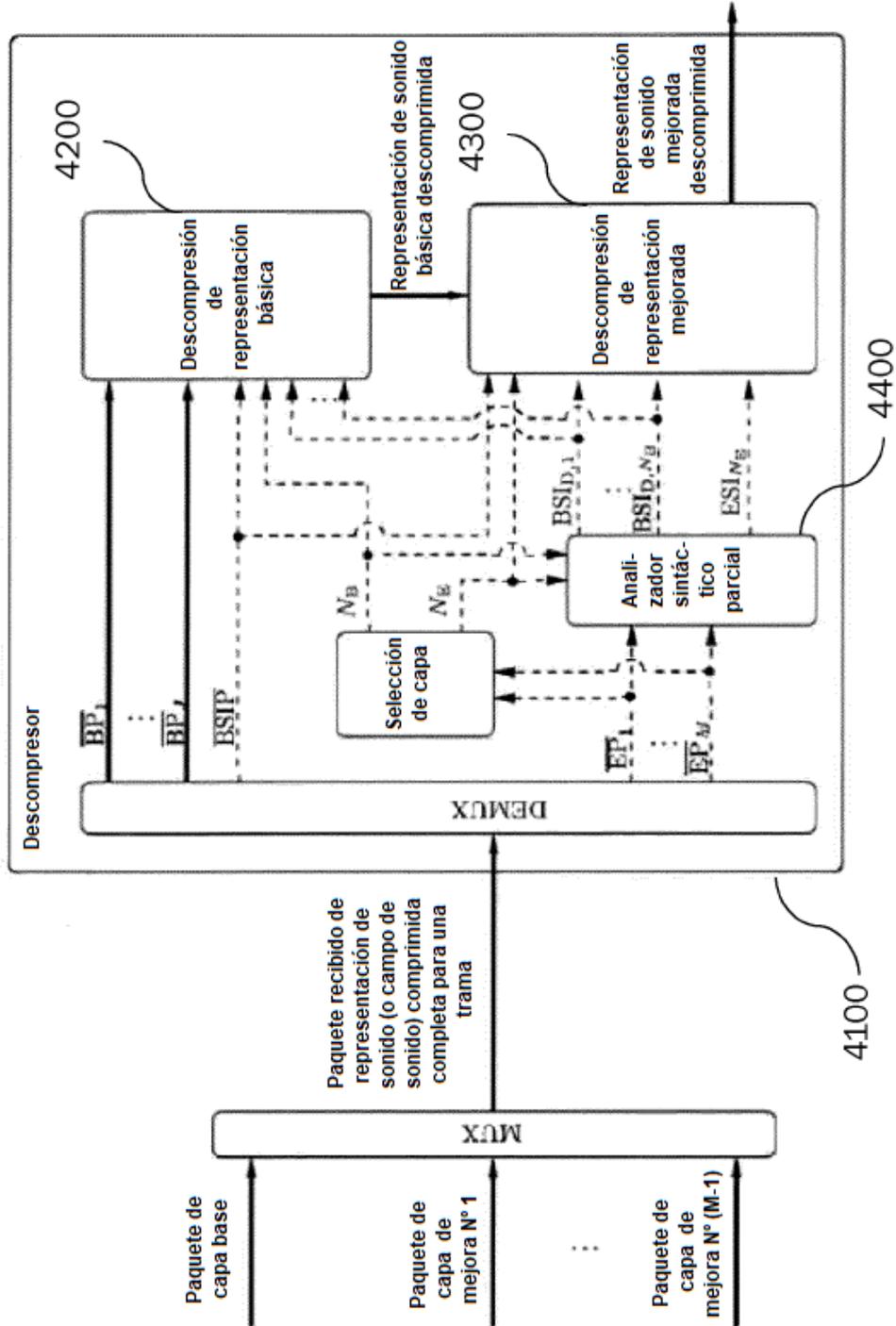


Fig. 4B

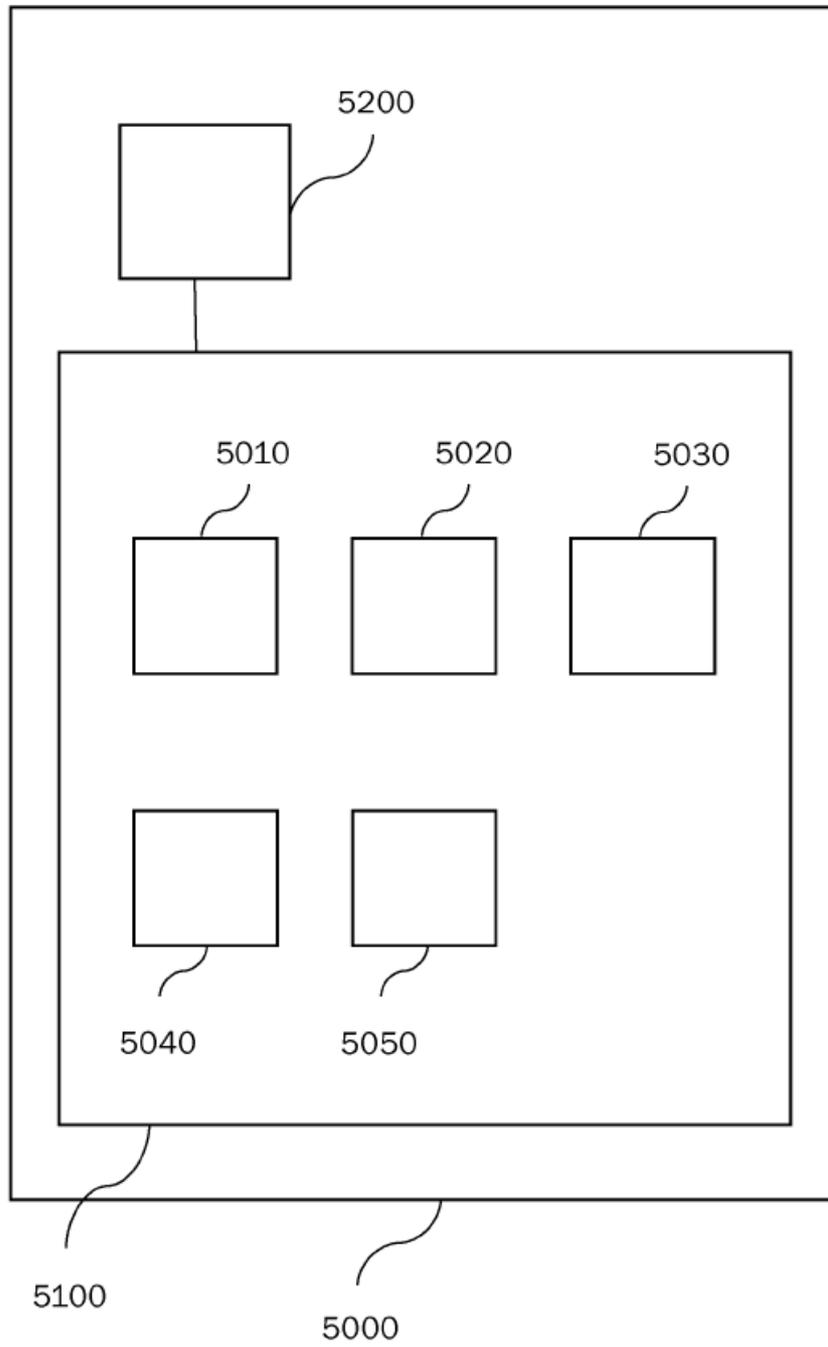


Fig. 5

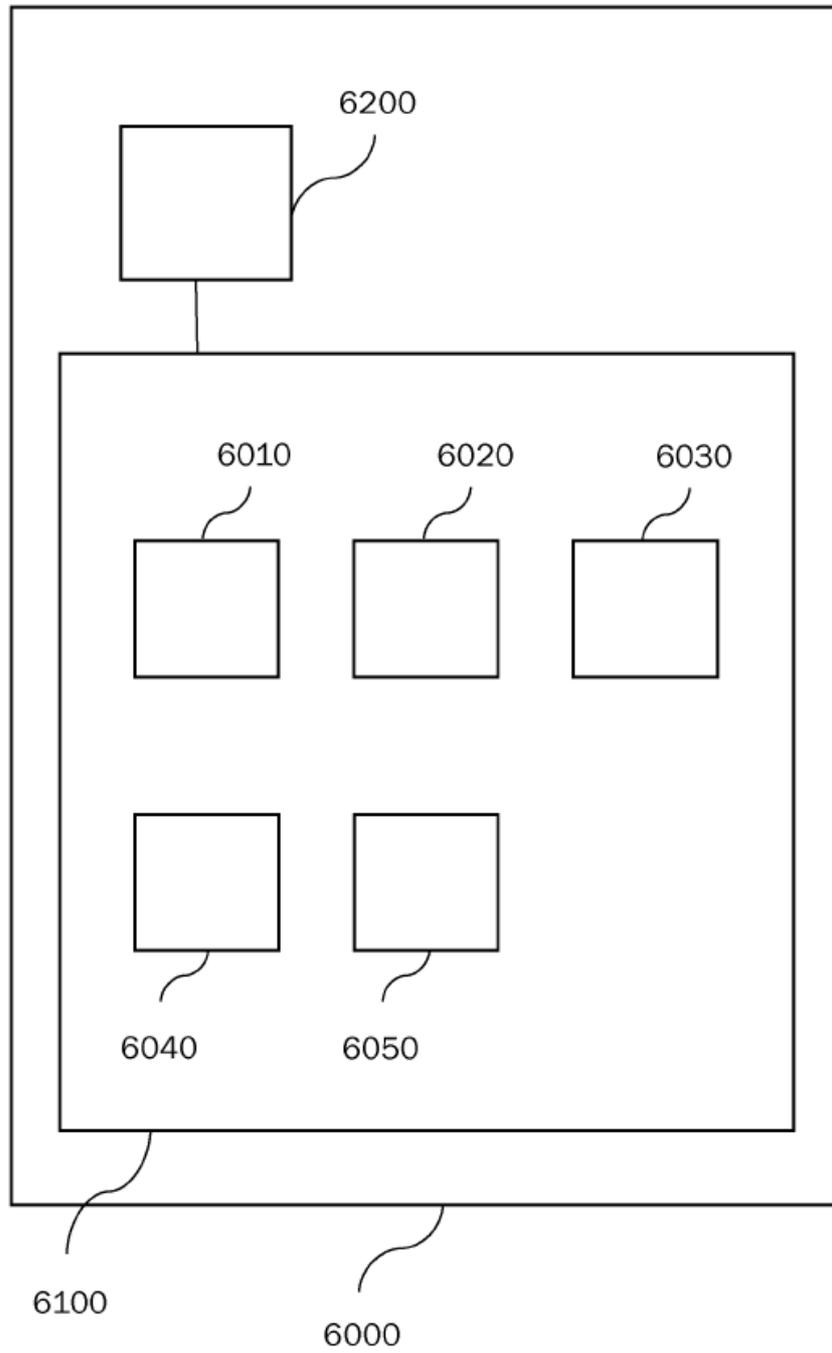


Fig. 6