

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 640 815**

51 Int. Cl.:

G10L 19/008 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **23.05.2014 PCT/EP2014/060733**

87 Fecha y número de publicación internacional: **27.11.2014 WO14187990**

96 Fecha de presentación y número de la solicitud europea: **23.05.2014 E 14730451 (3)**

97 Fecha y número de publicación de la concesión europea: **09.08.2017 EP 3005356**

54 Título: **Codificación eficiente de escenas de audio que comprenden objetos de audio**

30 Prioridad:

24.05.2013 US 201361827246 P
21.10.2013 US 201361893770 P
01.04.2014 US 201461973623 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
06.11.2017

73 Titular/es:

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

72 Inventor/es:

PURNHAGEN, HEIKO;
KJOERLING, KRISTOFER;
HIRVONEN, TONI;
VILLEMOES, LARS;
BREEBAART, DIRK JEROEN y
SAMUELSSON, LEIF JONAS

74 Agente/Representante:

LEHMANN NOVO, María Isabel

ES 2 640 815 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 eficiente de escenas de audio que comprenden objetos de audio

5 REFERENCIA CRUZADA A SOLICITUDES DE PATENTE RELACIONADAS

La presente solicitud reivindica el beneficio de la fecha de presentación de la solicitud de patente provisional de los Estados Unidos nº 61/827,246 presentada con fecha 24 de mayo de 2013, la solicitud de patente provisional de los Estados Unidos nº 61/893,770, presentada con fecha 21 de octubre de 2013 y la solicitud de patente provisional de los Estados Unidos nº 61/973,623 presentada con fecha 1 de abril de 2014.

CAMPO DE LA INVENCION

La idea inventiva se refiere, en general, a la codificación de una escena de audio que comprende objetos de audio. En particular, se refiere a un codificador, un decodificador y métodos asociados para la codificación y decodificación de objetos de audio.

ANTECEDENTES DE LA INVENCION

Una escena de audio puede comprender, en general, objetos de audio y canales de audio. Un objeto de audio es una señal de audio que tiene una posición espacial asociada que puede variar en el tiempo. Un canal de audio es una señal de audio que corresponde directamente a un canal de una configuración de altavoces multicanales, tal como una así denominada configuración de altavoces 5.1 con tres altavoces frontales, dos altavoces circundantes y un altavoz de efectos de baja frecuencia.

Puesto que el número de objetos de audio normalmente puede ser muy grande, a modo de ejemplo, en el orden de magnitud de centenares de objetos de audio, existe una necesidad para métodos de codificación que permiten que los objetos de audio sean reconstruidos eficientemente en el lado del decodificador. Han existido sugerencias para combinar los objetos de audio en una mezcla reducida multicanal (esto es, en una pluralidad de canales de audio que corresponden a los canales de una determinada configuración de altavoces multicanal tal como una configuración 5.1) en un lado del decodificador, y para reconstruir los objetos de audio de forma paramétrica a partir de la mezcla reducida multicanal en un lado de decodificador.

Un inconveniente de dicho método es que un decodificador de legado que no soporta la reconstrucción de objetos de audio puede utilizar la mezcla reducida multicanal directamente para la reproducción en una configuración de altavoces multicanal. A modo de ejemplo, una mezcla reducida 5.1 puede reproducirse directamente en los altavoces de una configuración 5.1.

Un inconveniente con este método es, sin embargo, que la mezcla reducida multicanal no puede proporcionar una reconstrucción suficientemente adecuada de los objetos de audio en el lado del decodificador. A modo de ejemplo, considérese dos objetos de audio que tienen la misma posición horizontal que el altavoz frontal izquierdo de una configuración 5.1, pero una posición vertical diferente. Estos objetos de audio normalmente se combinarían en el mismo canal de una mezcla reducida 5.1. Lo que antecede constituiría una situación difícil de superar la reconstrucción de objetos de audio en el lado del decodificador, que tendría que reconstruir aproximaciones de los dos objetos de audio a partir del mismo canal de mezcla reducida, lo que es un proceso que no puede garantizar una reconstrucción perfecta y que, a veces, incluso daría lugar a los así denominados artefactos audibles.

Existe, por lo tanto, una necesidad de métodos de codificación/decodificación que proporcionen una reconstrucción eficiente y mejorada de los objetos de audio.

La información lateral o metadatos se suele utilizar durante la reconstrucción de objetos de audio a partir de, p.ej., una mezcla reducida. La forma y contenido de dicha información lateral puede, a modo de ejemplo, afectar a la fidelidad de los objetos de audio reconstruidos y/o la complejidad de cálculo para realizar la reconstrucción. Por lo tanto, sería deseable proporcionar métodos de codificación/decodificación con una información lateral nueva y alternativa que permita aumentar la fidelidad de los objetos de audio reconstruidos y/o que permita reducir la complejidad de cálculo de la reconstrucción.

I. Citas de informes de búsqueda internacionales

El Informe de Búsqueda Internacional que corresponde al presente documento cita, *inter alia* los dos documentos siguientes:

Publicación de solicitud de patente internacional (PCT) nº WO 2010/125104 A1, que da a conocer un aparato para proporcionar uno o más parámetros ajustados para una provisión de una representación de señal de mezcla ascendente sobre la base de una representación de señal de mezcla reducida y una información paramétrica relacionada con el objeto comprende un ajustador de parámetros. El ajustador de parámetros está configurado para

5 recibir uno o más parámetros de entrada y para proporcionar, sobre su base, uno o más parámetros ajustados. El ajustador de parámetros está configurado para proporcionar los uno o más parámetros ajustados en función de los uno o más parámetros de entrada y la información paramétrica relacionada con el objeto, de modo que una distorsión de la representación de señal de mezcla ascendente causada por el uso de parámetros no óptimos se reduzca al menos para los parámetros de entrada que se desvían de los parámetros óptimos en más de una desviación predeterminada.

10 La publicación de solicitud de patente internacional (PCT) nº WO 2014/015299 A1, que da a conocer técnicas para agrupar objetos de audio en agrupamientos. En algunos ejemplos, un dispositivo para procesamiento de una señales de audio comprende un módulo de análisis de agrupamientos configurado para agrupar, sobre la base de la información espacial para cada uno de los N objetos de audio, una pluralidad de objetos de audio que incluye los N objetos de audio en L agrupamientos, en donde L es menor que N, en donde el módulo de análisis de agrupamientos está configurado para recibir información procedente de al menos uno de entre un canal de transmisión, un decodificador, y un dispositivo de presentación, y en donde un valor máximo para L está basado en la información recibida. El dispositivo comprende, además, un módulo de mezcla reducida configurado para mezclar la pluralidad de objetos de audio en los L flujos de audio y un módulo de mezcla reducida de metadatos configurado para producir, sobre la base de la información espacial y del agrupamiento, metadatos que indican la información espacial para cada uno de los L flujos de audio.

20 BREVE DESCRIPCIÓN DE LOS DIBUJOS

Formas de realización, a modo de ejemplo, se describirán a continuación haciendo referencia a los dibujos adjuntos, en los que:

25 La Figura 1 es una ilustración esquemática de un codificador en conformidad con formas de realización a modo de ejemplo;

La Figura 2 es una ilustración esquemática de un decodificador que soporta la reconstrucción de objetos de audio en conformidad con formas de realización a modo de ejemplo;

30 La Figura 3 es una ilustración esquemática de un decodificador de baja complejidad que no soporta la reconstrucción de objetos de audio en conformidad con formas de realización a modo de ejemplo;

35 La Figura 4 es una ilustración esquemática de un codificador que comprende un componente de agrupamiento secuencialmente dispuesto para simplificación de una escena de audio en conformidad con formas de realización a modo de ejemplo;

40 La Figura 5 es una ilustración esquemática de un codificador que comprende un componente de agrupamiento dispuesto en paralelo para simplificación de una escena de audio en conformidad con formas de realización a modo de ejemplo;

La Figura 6 ilustra un proceso conocido típico para calcular una matriz de presentación para un conjunto de instancias de metadatos;

45 La Figura 7 ilustra la derivación de una curva de coeficientes utilizada en la presentación de las señales de audio;

La Figura 8 ilustra un método de interpolación de instancias de metadatos, en conformidad con una forma de realización a modo de ejemplo;

50 Las Figuras 9 y 10 ilustran ejemplos de introducción de instancias de metadatos adicionales, en conformidad con formas de realización a modo de ejemplo; y

La Figura 11 ilustra un método de interpolación que utiliza un circuito de tipo *muestreo y retención* (muestreo y retención) con un filtro de paso bajo, en conformidad con una forma de realización a modo de ejemplo.

55 Todas las figuras son esquemáticas y en general, solamente ilustran partes que son necesarias con el fin elucidar la idea inventiva, mientras que otras partes pueden omitirse o meramente sugerirse. A no ser que se indique de otro modo, las referencias numéricas similares se refieren a partes similares en diferentes Figuras.

60 DESCRIPCIÓN DETALLADA DE LA INVENCION

65 Considerando lo que antecede, es un objetivo dar a conocer métodos de codificación y decodificación, un producto de programa informático y un decodificador, que permiten una reconstrucción eficiente y mejorada de objetos de audio y/o que permiten aumentar la fidelidad de los objetos de audio reconstruidos y/o que permite reducir la complejidad de cálculo de la reconstrucción.

Dicho objetivo se resuelve por las reivindicaciones independientes. Formas de realización específicas se definen en las reivindicaciones subordinadas.

I. Descripción general – Codificador

5 Según un primer aspecto de la idea inventiva, se da a conocer un método de codificación, un codificador y un producto de programa informático para codificar objetos de audio, según se estipula en las reivindicaciones.

10 En conformidad con las formas de realización a modo de ejemplo, se da a conocer un método para codificar objetos de audio en un flujo de datos, que comprende; *inter alia*:

la recepción de N objetos de audio, en donde $N > 1$;

15 calcular M señales de mezcla reducida, en donde $M \leq N$, mediante la formación de combinaciones de los N objetos de audio en conformidad con un criterio que es independiente de cualquier configuración de altavoces;

calcular la información lateral que incluye parámetros que permiten la reconstrucción de un conjunto de objetos de audio formado sobre la base de los N objetos de audio a partir de las M señales de mezcla reducida; y

20 incluir las M señales de mezcla reducida y la información lateral en el flujo de datos para su transmisión a un decodificador.

25 Con la disposición antes citada, las M señales de mezcla reducida son formadas, de este modo, a partir de los N objetos de audio con independencia de cualquier configuración de altavoces. Esto implica que las M señales de mezcla reducida no están restringidas a señales de audio que son adecuadas para la reproducción en los canales de una configuración de altavoces con M canales. En lugar de ello, las M señales de mezcla reducida pueden seleccionarse, con mayor libertad, de conformidad con un criterio de modo que se adapten, a modo de ejemplo, a la dinámica de los N objetos de audio y mejorar así la reconstrucción de los objetos de audio en el lado del decodificador.

30 Volviendo al ejemplo con dos objetos de audio que tienen la misma posición horizontal como el altavoz frontal izquierdo de una configuración 5.1, pero una posición de verificación diferente, el método propuesto permite poner el primer objeto de audio en una primera señal de mezcla reducida, y el segundo objeto de audio en la segunda señal de mezcla reducida. Lo que antecede permite una reconstrucción perfecta de los objetos de audio en el decodificador. En general, dicha reconstrucción perfecta es posible en tanto que el número de objetos de audio activos no supere el número de señales de mezcla reducida. Si el número de objetos de audio activos es más alto, en tal caso, el método propuesto permite la selección de los objetos de audio que han de mezclarse en la misma señal de mezcla reducida, de modo que los posibles errores de aproximación que se produzcan en el objeto de audio reconstruido en el decodificador no tengan ninguno o el más pequeño impacto perceptual posible sobre la escena de audio reconstruida.

35 Una segunda ventaja de las M señales de mezcla reducida que son adaptativas, es la capacidad para mantener algunos objetos de audio estrictamente separados de otros objetos de audio. A modo de ejemplo, puede ser ventajoso mantener cualquier objeto de diálogo separado de objetos de fondo, para garantizar que se presente el diálogo con exactitud en términos de atributos espaciales, y permite el procesamiento de objetos en el decodificador, tal como la mejora del diálogo o el aumento de la intensidad sonora del diálogo para mejora de la inteligibilidad. Para otras aplicaciones (p.ej., *karaoke*), puede ser ventajoso permitir el silenciamiento completo de uno o más objetos, lo que requiere también que dichos objetos no estén mezclados con otros objetos. Los métodos convencionales que utilizan una mezcla reducida multicanal correspondiente a una configuración de altavoces específica no permiten un silenciamiento completo de objetos de audio presentes en una mezcla de otros objetos de audio.

40 El término de señal de mezcla reducida refleja que una señal de mezcla reducida es una mezcla, esto es, una combinación, de otras señales. El término "reducida" indica que el número M de señales de mezcla reducida suele ser inferior al número N de objetos de audio.

55 En conformidad con formas de realización a modo de ejemplo, el método puede comprender, además, asociar cada señal de mezcla reducida con una posición espacial e incluye las posiciones espaciales de las señales de mezcla reducida en el flujo de datos como metadatos para las señales de mezcla reducida. Esto es conveniente por cuanto que permite la utilización de una decodificación de baja complejidad en caso de un sistema de reproducción de legado. Más concretamente, los metadatos asociados con las señales de mezcla reducida pueden utilizarse en un lado del decodificador para la presentación de las señales de mezcla reducida hacia los canales de un sistema de reproducción de legado.

60 De conformidad con formas de realización a modo de ejemplo, los N objetos de audio están asociados con metadatos que incluyen posiciones espaciales de los N objetos de audio, y las posiciones espaciales asociadas con las señales de mezcla reducida se calculan sobre la base de las posiciones espaciales de los N objetos de audio. De

este modo, las señales de mezcla reducida pueden interpretarse como objetos de audio que tienen una posición espacial que depende de las posiciones espaciales de los N objetos de audio.

Además, las posiciones espaciales de los N objetos de audio y las posiciones espaciales asociadas con las M señales de mezcla reducida pueden variar en el tiempo, es decir, pueden variar entre tramas temporales de datos de audio. Dicho de otro modo, las señales de mezcla reducida pueden interpretarse como objetos de audio dinámicos que tienen una posición asociada que varía entre tramas temporales. Lo que antecede contrasta con los sistemas de la técnica anterior en donde las señales de mezcla reducida corresponden a posiciones de altavoces espaciales fijas.

En condiciones normales, la información lateral es también variable en el tiempo lo que permite que varíen temporalmente los parámetros que rigen la reconstrucción de los objetos de audio.

El codificador puede aplicar diferentes criterios para el cálculo de las señales de mezcla reducida. De conformidad con formas de realización a modo de ejemplo en donde los N objetos de audio están asociados con metadatos que incluyen posiciones espaciales de los N objetos de audio, el criterio para calcular las M señales de mezcla reducida puede basarse en la proximidad espacial de los N objetos de audio. A modo de ejemplo, los objetos de audio que están próximos entre sí pueden combinarse en la misma señal de mezcla reducida.

En conformidad con formas de realización a modo de ejemplo en donde los metadatos asociados con los N objetos de audio comprende, además, valores de importancia que indican la importancia de los N objetos de audio en relación entre sí, el criterio para calcular las M señales de mezcla reducida puede estar basado, además, en los valores de importancia de los N objetos de audio. A modo de ejemplo, los más importantes de los N objetos de audio pueden ser objeto de mapeado de correspondencia directamente para una señal de mezcla reducida, mientras que los objetos de audio restantes se combinan para formar las señales de mezcla reducida restantes.

En particular, en conformidad con formas de realización a modo de ejemplo, la etapa de calcular las M señales de mezcla reducida comprende un primer procedimiento de agrupamiento que incluye la asociación de los N objetos de audio con M agrupamientos basados en la proximidad espacial y en los valores de importancia, si fuera aplicable, de los N objetos de audio, y calcular una señal de mezcla reducida para cada agrupamiento mediante la formación de una combinación de objetos de audio asociados con el agrupamiento. En algunos casos, un objeto de audio puede formar parte de, como máximo, un agrupamiento. En otros casos, un objeto de audio puede formar parte de varios agrupamientos. De este modo, grupos diferentes, esto es, agrupamientos, se forman a partir de los objetos de audio. Cada agrupamiento puede, a su vez, representarse por una señal de mezcla reducida que puede considerarse como un objeto de audio. El método de agrupamiento permite la asociación de cada señal de mezcla reducida con una posición espacial que se calcula sobre la base de las posiciones espaciales de los objetos de audio asociados con el agrupamiento correspondiente a la señal de mezcla reducida. Con esta interpretación, el primer procedimiento de agrupamiento reduce así la dimensionalidad de los N objetos de audio a M objetos de audio de una manera flexible.

La posición espacial asociada con cada señal de mezcla reducida puede calcularse, a modo de ejemplo, como un centroide o un centroide ponderado de las posiciones espaciales de los objetos de audio asociados con el agrupamiento correspondiente a la señal de mezcla reducida. Los valores de ponderación pueden basarse, a modo de ejemplo, en los valores de importancia de los objetos de audio.

En conformidad con formas de realización a modo de ejemplo, los N objetos de audio están asociados con los M agrupamientos aplicando un denominado algoritmo *K-means* que tiene las posiciones espaciales de los N objetos de audio como entrada.

Puesto que una escena de audio puede comprender un gran número de objetos de audio, el método puede tomar medidas adicionales para reducir la dimensionalidad de la escena de audio, con lo que se reduce la complejidad de cálculo en el lado del decodificador cuando se reconstruyen los objetos de audio. En particular, el método puede comprender, además, un segundo procedimiento de agrupamiento para reducir una primera pluralidad de objetos de audio a una segunda pluralidad objetos de audio.

En conformidad con una forma de realización, el segundo procedimiento de agrupamiento se realiza antes del cálculo de las M señales de mezcla reducida. En esa forma de realización, la primera pluralidad de objetos de audio corresponden, por lo tanto, a los objetos de audio originales de la escena de audio, y la segunda, reducida, pluralidad de objetos corresponde a los N objetos de audio sobre la base sobre la que se calculan las M señales de mezcla reducida. Además, en dicha forma de realización, el conjunto de objetos de audio (a reconstruirse en el decodificador) formados sobre la base de los N objetos de audio corresponden, esto es, es igual a, los N objetos de audio.

En conformidad con otra forma de realización, el segundo procedimiento de agrupamiento se realiza en paralelo con el cálculo de las M señales de mezcla reducida. En dicha forma de realización, los N objetos de audio sobre la base con la que se calculan las M señales de mezcla reducida, así como la primera pluralidad de objetos de audio que son entrada para el segundo procedimiento de agrupamiento corresponden a los objetos de audio originales de la

escena de audio. Además, en dicha forma de realización, el conjunto de objetos de audio (a reconstruirse en el decodificador) formados sobre la base de los N objetos de audio corresponden a la segunda pluralidad de objetos de audio. Con este método, las M señales de mezcla reducida se calculan, por lo tanto, sobre la base de los objetos de audio originales de la escena de audio y no sobre la base de un número reducido de objetos de audio.

5 En conformidad con forma de realización a modo de ejemplo, el segundo procedimiento de agrupamiento comprende:

la recepción de la primera pluralidad de objetos de audio y sus posiciones espaciales asociadas;

10 la asociación de la primera pluralidad de objetos de audio con al menos un agrupamiento sobre la base de una proximidad espacial de la primera pluralidad de objetos de audio,

15 la generación de la segunda pluralidad de objetos de audio representando cada uno de los al menos un agrupamiento por un objeto de audio que es una combinación de los objetos de audio asociados con el agrupamiento,

20 el cálculo de metadatos que incluyen las posiciones espaciales para la segunda pluralidad de objetos de audio, en donde la posición espacial de cada objeto de audio de la segunda pluralidad de objetos de audio se calcula sobre la base de las posiciones espaciales de los objetos de audio asociados con el agrupamiento correspondiente; y

la inclusión de los metadatos para la segunda pluralidad de objetos de audio en el flujo de datos.

25 Dicho de otro modo, el segundo procedimiento de agrupamiento utiliza la redundancia espacial presente en la escena de audio, tal como objetos que tienen localizaciones iguales o muy similares. Además, los valores de importancia de los objetos de audio pueden tenerse en cuenta cuando se genera la segunda pluralidad de objetos de audio.

30 Según se mencionó con anterioridad, la escena de audio puede comprender también canales de audio. Dichos canales de audio pueden considerarse como un objeto de audio que está asociado con una posición estática, p.ej., la posición del altavoz correspondiente al canal de audio. En más detalle, el segundo procedimiento de agrupamiento puede comprender, además:

35 la recepción de al menos un canal de audio;

la conversión de cada uno de los al menos un canal de audio en un objeto de audio que tiene una posición espacial estática correspondiente a una posición de altavoz de ese canal de audio; y

40 la inclusión del al menos un canal de audio convertido en la primera pluralidad de objetos de audio.

De este modo, el método permite la codificación de una escena de audio que comprende canales de audio, así como objetos de audio.

45 En conformidad con formas de realización a modo de ejemplo, se da a conocer un producto de programa informático que comprende un soporte legible por ordenador con instrucciones para realizar el método de decodificación en conformidad con las formas de realización a modo de ejemplo.

50 En conformidad con formas de realización ejemplo, se da a conocer un codificador para codificar objetos de audio en un flujo de datos, que comprende:

un componente de recepción configurado para recibir N objetos de audio, en donde $N > 1$,

55 un componente de mezcla reducida configurado para calcular las M señales de mezcla reducida, en donde $M \leq N$, mediante la formación de combinaciones de los N objetos de audio en conformidad con un criterio que es independiente de cualquier configuración de altavoces;

60 un componente de análisis configurado para calcular información lateral que incluye parámetros que permiten una reconstrucción del conjunto de objetos de audio formados sobre la base de los N objetos de audio a partir de las M señales de mezcla reducida; y

un componente de multiplexación configurado para incluir las M señales de mezcla reducida y la información lateral en el flujo de datos para su transmisión a un decodificador.

65 II. Descripción general – Decodificador

En conformidad con un segundo aspecto de la idea inventiva, se da a conocer un método de decodificación, un

decodificador y un producto de programa informático para decodificar un contenido de audio multicanal, según se indica en las reivindicaciones.

5 El segundo aspecto de la idea inventiva puede tener, en general, las mismas características y ventajas que el primer aspecto.

En conformidad con formas de realización a modo de ejemplo, se da a conocer un método en un decodificador para decodificar un flujo de datos que incluye objetos de audio codificados, que comprende, *inter alia*:

10 la recepción de un flujo de datos que comprende M señales de mezcla reducida que son combinaciones de N objetos de audio calculados en conformidad con un criterio que es independiente de cualquier configuración de altavoces, en donde $M \leq N$ y la información lateral incluye parámetros que permiten la reconstrucción de un conjunto de objetos de audio formados sobre la base de los N objetos de audio a partir de las M señales de mezcla reducida; y

15 la reconstrucción del conjunto de objetos de audio formados sobre la base de los N objetos de audio a partir de las M señales de mezcla reducida y de la información lateral.

20 En conformidad con formas de realización a modo de ejemplo, el flujo de datos comprende, además, metadatos para las M señales de mezcla reducida que incluyen posiciones espaciales asociadas con las M señales de mezcla reducida, comprendiendo el método, además:

25 en una condición en donde el decodificador está configurado para soportar una reconstrucción de objeto de audio, realizar la etapa de reconstruir el conjunto de objetos de audio formado sobre la base de los N objetos de audio a partir de las M señales de mezcla reducida y de la información lateral; y

30 en una condición en donde el decodificador no está configurado para soportar una reconstrucción de objeto de audio, utilizar los metadatos para las M señales de mezcla reducida para la presentación de las M señales de mezcla reducida hacia los canales de salida de un sistema de reproducción.

De conformidad con las formas de realización a modo de ejemplo, las posiciones espaciales asociadas con las M señales de mezcla reducida varían en el tiempo.

35 En conformidad con formas de realización a modo de ejemplo, la información lateral varía en el tiempo.

En conformidad con formas de realización a modo de ejemplo, el flujo de datos comprende, además, metadatos para el conjunto de objetos de audio formados sobre la base de los N objetos de audio que incluyen las posiciones espaciales del conjunto de objetos de audio formados sobre la base de los N objetos de audio, comprendiendo el método, además:

40 la utilización de los metadatos para el conjunto de objetos de audio formado sobre la base de los N objetos de audio para la presentación del conjunto reconstruido de objetos de audio formado sobre la base de los N objetos de audio hacia los canales de salida de un sistema de reproducción.

45 De conformidad con formas de realización a modo de ejemplo, el conjunto de objetos de audio formado sobre la base de los N objetos de audio es igual a los N objetos de audio.

50 En conformidad con formas de realización a modo de ejemplo, el conjunto de objetos de audio formado sobre la base de los N objeto de audio comprende una pluralidad de objetos de audio que son combinaciones de los N objetos de audio y cuyo número es inferior a N.

55 En conformidad con formas de realización a modo de ejemplo, se da a conocer un producto de programa informático que comprende un soporte con instrucciones para realizar el método de decodificación en conformidad con formas de realización a modo de ejemplo.

En conformidad con formas de realización a modo de ejemplo, se da a conocer un decodificador para decodificar un flujo de datos que incluyen objetos de audio decodificados, que comprende:

60 un componente de recepción configurado para recibir un flujo de datos que comprende M señales de mezcla reducida que son combinaciones de N objetos de audio calculados en conformidad con un criterio que es independiente de cualquier configuración de altavoces, en donde $M \leq N$, y la información lateral que incluye parámetros que permiten la reconstrucción de un conjunto de objetos de audio formado sobre la base de los N objetos de audio a partir de las M señales de mezcla reducida; y

65 un componente de reconstrucción configurado para la reconstrucción del conjunto de objetos de audio formado sobre la base de los N objetos de audio a partir de las M señales de mezcla reducida y de la información lateral.

III. Descripción general – Formato para la información lateral y los metadatos

5 De conformidad con un tercer aspecto de la idea inventiva, se da a conocer un método de codificación, un codificador y un producto de programa informático para codificar objetos de audio.

10 Los métodos, codificadores y producto de programa informático según el tercer aspecto de la idea inventiva pueden tener, en general, características y ventajas en común con los métodos, codificadores y producto de programa informático de conformidad con el primer aspecto de la idea inventiva.

15 De conformidad con las formas de realización a modo de ejemplo, se da a conocer un método para codificar objetos de audio como un flujo de datos. El método comprende *inter alia*:

la recepción de N objetos de audio, en donde $N > 1$;

el cálculo de M señales de mezcla reducida, en donde $M \leq N$, mediante la formación de combinaciones de los N objetos de audio;

20 el cálculo de información lateral variable en el tiempo que incluye parámetros que permiten la reconstrucción de un conjunto de objetos de audio formados sobre la base de los N objetos de audio a partir de las M señales de mezcla reducida; y

la inclusión de las M señales de mezcla reducida y de la información lateral en un flujo de datos para su transmisión a un decodificador.

25 En las presentes formas de realización a modo de ejemplo, el método comprende, además, incluir, en el flujo de datos:

30 una pluralidad de instancias de información lateral que especifican respectivos establecimientos de reconstrucción deseados para reconstruir el conjunto de objetos de audio formado sobre la base de los N objetos de audio; y

35 para cada instancia de información lateral, datos de transición que incluyen dos partes independientemente asignables que, en combinación, definen un punto en el tiempo para iniciar una transición desde un establecimiento de reconstrucción actual al establecimiento de reconstrucción deseado que se especifica por la instancia de información lateral, y un punto en el tiempo para completar la transición.

40 En la presente forma de realización a modo de ejemplo, la información lateral es variable en el tiempo, p.ej., que varía en el tiempo, permitiendo a los parámetros regir la reconstrucción de los objetos de audio para variar con respecto al tiempo, lo que se refleja por la presencia de las instancias de información lateral. Utilizando un formato de información lateral que incluye datos de transición que definen puntos en el tiempo para iniciar y punto en el tiempo para completar las transiciones desde los establecimientos de reconstrucción actuales a los establecimientos de reconstrucción deseados respectivos, las instancias de información lateral se realizan más independientes entre sí en el sentido de que puede realizarse interpolación sobre la base de un establecimiento de reconstrucción actual y un establecimiento de reconstrucción deseado especificado por una instancia de información lateral única, esto es,

45 sin conocimiento de cualesquiera otras instancias de información lateral. El formato de información lateral proporcionada, por lo tanto, facilita el cálculo/introducción de instancias de información lateral adicional entre las instancias de información lateral existentes. En particular, el formato de información lateral proporcionada permite el cálculo/introducción de instancias de información lateral adicionales sin afectar a la calidad de la reproducción. En esta invención, el proceso de cálculo/introducción de nuevas instancias de información lateral entre instancias de información lateral existentes se refiere como "remuestreo" de la información lateral. El remuestreo de información lateral suele requerirse durante algunas tareas de procesamiento de audio. A modo de ejemplo, cuando se edita un contenido de audio, mediante, por ejemplo, corte/fusión/mezcla, dichas ediciones pueden ocurrir entre instancias de información lateral. En este caso, el remuestreo de la información lateral puede ser requerido. Otro de dichos casos es cuando las señales de audio y la información lateral asociada se codifican con un *códec* de audio basado en

50 tramas. En este caso, es deseable tener al menos una instancia de información lateral para cada trama de *códec* de audio, preferentemente, con una marca temporal en el inicio de esa trama de *códec*, para mejorar la resistencia a las pérdidas de tramas durante la transmisión. A modo de ejemplo, las señales/objetos de audio pueden ser parte de una señal audiovisual o una señal multimedia que incluye un contenido de vídeo. En dichas aplicaciones, puede ser deseable modificar la tasa de tramas del contenido de audio para la adaptación de una tasa de tramas del contenido de audio, en cuyo caso puede ser deseable un remuestreo correspondiente de la información lateral.

El flujo de datos en el que se incluyen la señal de mezcla reducida y la información lateral puede ser, a modo de ejemplo, un flujo de bits, en particular, un flujo de bits memorizado o transmitido.

65 Ha de entenderse que el cálculo de las M señales de mezcla reducida mediante la formación de combinaciones de los N objetos de audio significa que cada una de las M señales de mezcla reducida se obtiene formando una

combinación, esto es, una combinación lineal, del contenido de audio de los uno o más de los N objetos de audio. Dicho de otro, cada uno de los N objetos de audio no necesita necesariamente contribuir a cada una de las M señales de mezcla reducida.

5 El término de señal de mezcla reducida refleja que una señal de mezcla reducida es una mezcla, esto es, una combinación, de otras señales. La señal de mezcla reducida puede, a modo de ejemplo, ser una mezcla aditiva de otras señales. El término "reducida" o descendente, indica que el número de M de señales de mezcla reducida suele ser inferior al número N de objetos de audio.

10 Las señales de mezcla reducida pueden, a modo de ejemplo, calcularse formando combinaciones de las N señales de audio según un criterio que es independiente de cualquier configuración de altavoces, de conformidad con cualquiera de las formas de realización ejemplo dentro del primer aspecto de la idea inventiva. Como alternativa, las señales de mezcla reducida pueden, a modo de ejemplo, calcularse mediante la formación de combinación de las N
15 señales de audio, de modo que las señales de mezcla reducida sean adecuadas para la reproducción en los canales de una configuración de altavoces con M canales, referida en esta descripción como una mezcla reducida compatible con versiones anteriores.

Mediante los datos de transición que incluyen dos partes independientemente asignables se indica que las dos partes son mutuamente asignables de forma independiente, esto es, se pueden asignar con independencia entre sí.
20 Sin embargo, ha de entenderse que las partes de los datos de transición pueden, a modo de ejemplo, coincidir con partes de datos de transición para otros tipos de información lateral de metadatos.

En la presente forma de realización a modo de ejemplo, las dos partes independientemente asignables de los datos de transición, en combinación, definen el punto en el tiempo para iniciar la transición y el punto en el tiempo para
25 completar la transición, esto es, estos dos puntos en el tiempo son derivables de las dos partes independientemente asignables de los datos de transición.

En conformidad con una forma de realización a modo de ejemplo, el método puede incluir, además, un procedimiento de agrupamiento para reducir una primera pluralidad de objetos de audio a una segunda pluralidad de
30 objetos de audio, en donde los N objetos de audio constituyen la primera pluralidad de objetos de audio o la segunda pluralidad de objetos de audio, y en donde el conjunto de objetos de audio formado sobre la base de los N objetos de audio coincide con la segunda pluralidad de objetos de audio. En la presente forma de realización a modo de ejemplo, el procedimiento de agrupamiento puede comprender:

35 el cálculo de los metadatos del agrupamiento variable en el tiempo que incluye las posiciones espaciales para la segunda pluralidad de objetos de audio; y

la inclusión adicional, en el flujo de datos, para su transmisión al decodificador:

40 de una pluralidad de instancias de metadatos de agrupamiento que especifican los respectivos establecimientos de presentación deseado para presentar el segundo conjunto de objetos de audio; y

para cada instancia de metadatos de agrupamiento, los datos de transición que incluyen dos partes independientemente asignables que, en combinación, definen un punto en el tiempo para iniciar una transición
45 desde un establecimiento de presentación actual al establecimiento de presentación deseado que se especifica por la instancia de metadatos del agrupamiento, y un punto en el tiempo para completar la transición al establecimiento de presentación deseado que se especifica por la instancia de metadatos del agrupamiento.

Puesto que una escena de audio puede comprender un amplio número de objetos de audio, el método en conformidad con la presente forma de realización a modo de ejemplo, adopta medios adicionales para reducir la dimensionalidad de la escena de audio reduciendo la primera pluralidad de objetos de audio a una segunda pluralidad de objetos de audio. En la presente forma de realización a modo de ejemplo, el conjunto de objetos de audio, que se forma sobre la base de los N objetos de audio y que ha de reconstruirse en un lado del decodificador sobre la base de las señales de mezcla reducida y de la información lateral, coincide con la segunda pluralidad de
50 objetos de audio, que corresponde a una simplificación y/o representación dimensional más baja de la escena de audio representada por la primera pluralidad de señales de audio, y se reduce la complejidad de cálculo para la reconstrucción en un lado del decodificador.

La inclusión de los metadatos de agrupamiento en el flujo de datos permite la presentación del segundo conjunto de señales de audio en un lado del decodificador, p.ej., después de que el segundo conjunto de señales de audio se haya reconstruido sobre la base de las señales de mezcla reducida y la información lateral.

De forma similar a la información lateral, los metadatos del agrupamiento en la presente forma de realización a modo de ejemplo, es variable en el tiempo, p.ej., varía en el transcurso del tiempo, lo que permite que los parámetros que rigen la presentación de la segunda pluralidad de objetos de audio varíen con respecto al tiempo. El formato para los metadatos de mezcla reducida pueden ser análogos al formato de la información lateral y puede tener las mismas o
65

correspondientes ventajas. En particular, la forma de los metadatos del agrupamiento que se proporciona en la forma de realización a modo de ejemplo presente, facilita el remuestreo de los metadatos del agrupamiento. El remuestreo de las metodologías del agrupamiento puede, a modo de ejemplo, utilizarse para proporcionar puntos comunes en el tiempo para iniciar y completar las transiciones respectivas asociadas con los metadatos del agrupamiento y la información lateral, y/o para ajustar los metadatos del agrupamiento para una tasa de tramas de las señales de audio asociadas.

En conformidad con una forma de realización a modo de ejemplo, el procedimiento de agrupamiento puede comprender, además:

la recepción de la primera pluralidad de objetos de audio y sus posiciones espaciales;

la asociación de la primera pluralidad de objetos de audio con al menos un agrupamiento sobre la base de la proximidad espacial de la primera pluralidad de objetos de audio;

la generación de la segunda pluralidad de objetos de audio representando cada uno de los al menos un agrupamiento mediante un objeto de audio que es una combinación de los objetos de audio asociados con el agrupamiento; y

el cálculo de la posición espacial de cada objeto de audio de la segunda pluralidad de objetos de audio sobre la base de las posiciones espaciales de los objetos de audio asociados con el agrupamiento respectivo, esto es, con el agrupamiento que representa el objeto de audio.

Dicho de otro modo, el procedimiento de agrupamiento utiliza la redundancia espacial presente en la escena de audio, tal como objetos de audio que tienen localizaciones iguales o muy similares. Además, los valores de importancia de los objetos de audio pueden tenerse en cuenta cuando se genera la segunda pluralidad de objetos de audio, según se describe con respecto a las formas de realización a modo de ejemplo dentro del primer aspecto de la idea inventiva.

La asociación de la primera pluralidad de objetos de audio con al menos un agrupamiento incluye la asociación de cada uno de la primera pluralidad de objetos de audio con uno o más del al menos un agrupamiento. En algunos casos, un objeto de audio puede formar parte de, como máximo, un agrupamiento, mientras que, en otros casos, un objeto de audio puede formar parte de varios agrupamientos. Dicho de otro modo, en algunos casos, un objeto de audio puede dividirse entre varios agrupamientos como parte del procedimiento de agrupamiento.

La proximidad espacial de la primera pluralidad de objetos de audio puede relacionarse con distancias entre, y/o posiciones relativas de, los respectivos objetos de audio en la primera pluralidad de objetos de audio. A modo de ejemplo, los objetos de audio que estén próximos entre sí pueden asociarse con el mismo agrupamiento.

Mediante un objeto de audio que es una combinación de los objetos de audio asociados con el agrupamiento se indica que el contenido de audio/señal asociado con el objeto de audio puede formarse como una combinación de los contenidos de audio/señales asociadas con los respectivos objetos de audio que se asocian con el agrupamiento.

De conformidad con una forma de realización a modo de ejemplo, los respectivos puntos en el tiempo definidos por los datos de transición para las respectivas instancias de metadatos de agrupamiento pueden coincidir con los respectivos puntos en el tiempo definidos por los datos de transición para las instancias de información lateral correspondientes.

Utilizando los mismos puntos en el tiempo para iniciar y para completar las transiciones asociadas con la información lateral y los metadatos de agrupamiento, se facilita el procesamiento conjunto de la información lateral y de los metadatos de agrupamiento, tal como un remuestreo conjunto.

Además, el uso de puntos comunes en el tiempo para iniciar y para completar las transiciones asociadas con la información lateral y los metadatos del agrupamiento facilita una reconstrucción conjunta y la presentación en un lado del decodificador. Si, a modo de ejemplo, una reconstrucción y presentación se realizan como una operación conjunta en un lado del decodificador, los establecimientos conjuntos para la reconstrucción y presentación pueden determinarse para instancia de información lateral e instancia de metadatos y/o interpolación entre establecimientos conjuntos para la reconstrucción y presentación, pueden utilizarse en lugar de realizar una interpolación por separado para los respectivos establecimientos. Dicha interpolación conjunta puede reducir la complejidad de cálculo en el lado del decodificador puesto que menos coeficientes/parámetros necesitan interpolarse.

En conformidad con una forma de realización a modo de ejemplo, el procedimiento de agrupamiento puede realizarse antes del cálculo de las M señales de mezcla reducida. En la presente forma de realización a modo de ejemplo, la primera pluralidad de objetos de audio corresponde a objetos de audio originales de la escena de audio, y los N objetos de audio sobre cuya base se calculan las M señales de mezcla reducida constituyen la segunda,

reducida, pluralidad de objetos de audio. Por consiguiente, en la presente forma de realización a modo de ejemplo, el conjunto de objetos de audio (a reconstruirse en el lado del decodificador) formado sobre la base de los N objetos de audio coincide con los N objetos de audio.

5 Como alternativa, el procedimiento de agrupamiento puede realizarse en paralelo con el cálculo de las M señales de mezcla reducida. En conformidad con la presente alternativa, los N objetos de audio sobre cuya base se calculan las M señales de mezcla reducida constituyen la primera pluralidad de objetos de audio que corresponde con los objetos de audio originales de la escena de audio. Con este método, las M señales de mezcla reducida se calculan, por lo tanto, sobre la base de los objetos de audio originales de la escena de audio y no sobre la base de un número
10 reducido de objetos de audio.

En conformidad con una forma de realización a modo de ejemplo, el método puede comprender, además:

15 la asociación de cada señal de mezcla reducida con una posición espacial variable en el tiempo para la presentación de las señales de mezcla reducida, y

la inclusión, además, en el flujo de datos, de metadatos de mezcla reducida que incluyen las posiciones espaciales de las señales de mezcla reducida,

20 en donde el método comprende, además, la inclusión, en el flujo de datos de:

una pluralidad de instancias de metadatos de mezcla reducida que especifican respectivos establecimientos de presentación de mezcla reducida deseados para la presentación de las señales de mezcla reducida; y

25 para cada instancia de metadatos de mezcla reducida, datos de transición que incluyen dos partes independientemente asignables que, en combinación, definen un punto en el tiempo para iniciar una transición desde un establecimiento de presentación de mezcla reducida actual al establecimiento de presentación de mezcla reducida deseado que se especifica por la instancia de metadatos de mezcla reducida, y un punto en el tiempo para completar la transición al establecimiento de presentación de mezcla reducida deseado que se especifica por la
30 instancia de metadatos de mezcla reducida.

La inclusión de metadatos de mezcla reducida en el flujo de datos es ventajosa por cuanto que permite utilizarse una decodificación de baja complejidad en caso de un equipo de reproducción de legado. Más concretamente, los metadatos de mezcla reducida pueden utilizarse en un lado del decodificador para presentar las señales de mezcla reducida a los canales de un sistema de reproducción de legado, esto es, sin reconstruir la pluralidad de objetos de audio formados sobre la base de los N objetos, que suele ser una operación de cálculo más complejo.
35

En conformidad con la presente forma de realización a modo de ejemplo, las posiciones espaciales asociadas con las M señales de mezcla reducida pueden ser variables en el tiempo, esto es, variar en el transcurso del tiempo, y las señales de mezcla reducida pueden interpretarse como objetos de audio dinámicos que tienen una posición asociada que puede cambiar entre las tramas temporales o las instancias de metadatos de mezcla reducida. Lo que antecede contrasta con los sistemas de la técnica anterior, en donde las señales de mezcla reducida corresponden a posiciones de altavoces espaciales fijas. Se recuerda que el mismo flujo de datos puede reproducirse en una forma orientada al objeto de audio en un sistema de decodificación con capacidades más evolucionadas.
40
45

En algunas formas de realización a modo de ejemplo, los N objetos de audio pueden asociarse con metadatos que incluyen posiciones espaciales de los N objetos de audio, y las posiciones espaciales asociadas con las señales de mezcla reducida pueden, a modo de ejemplo, calcularse sobre la base de las posiciones espaciales de los N objetos de audio. De este modo, las señales de mezcla reducida pueden interpretarse como objetos de audio que tienen posiciones espaciales que dependen de las posiciones espaciales de los N objetos de audio.
50

En conformidad con una forma de realización a modo de ejemplo, los respectivos puntos en el tiempo definidos por los datos de transición para las respectivas instancias de metadatos de mezcla reducida pueden coincidir con los respectivos puntos en el tiempo definidos por los datos de transición para las instancias de información lateral correspondientes. La utilización de los mismos puntos en el tiempo para iniciar y para completar las transiciones asociadas con la información lateral y los metadatos de mezcla reducida facilita un procesamiento conjunto, p.ej., un remuestreo, de la información lateral y de los metadatos de mezcla reducida.
55

De conformidad con una forma de realización a modo de ejemplo, los respectivos puntos en el tiempo definidos por los datos de transición para las respectivas instancias de metadatos de mezcla reducida pueden coincidir con los respectivos puntos en el tiempo definidos por los datos de transición para las instancias de metadatos de agrupamiento correspondientes. La utilización de los mismos puntos en el tiempo para iniciar y para finalizar las transiciones asociadas con los metadatos de agrupamiento y los metadatos de mezcla reducida facilita un procesamiento conjunto, a modo de ejemplo, un remuestreo, de los metadatos del agrupamiento y de los metadatos de la mezcla reducida.
60
65

ES 2 640 815 T3

En conformidad con las formas de realización a modo de ejemplo, se da a conocer un codificador para codificar N objetos de audio como un flujo de datos, en donde $N > 1$. El codificador comprende:

5 un componente de mezcla reducida configurado para calcular las M señales de mezcla reducida, en donde $M \leq N$, mediante la formación de combinaciones de los N objetos de audio;

10 un componente de análisis configurado para calcular información lateral variable en el tiempo, que incluye parámetros que permiten la reconstrucción de un conjunto de objetos de audio formados sobre la base de los N objetos de audio a partir de las M señales de mezcla reducida; y

10 un componente de multiplexación configurado para incluir las M señales de mezcla reducida y la información lateral en un flujo de datos para su envío a un decodificador,

15 en donde el componente de multiplexación está configurado, además, para incluir, en el flujo de datos, para su transmisión al decodificador:

una pluralidad de instancias de información lateral que especifican respectivos establecimientos de reconstrucción deseados para reconstruir el conjunto de objetos de audio formados sobre la base de los N objetos de audio; y

20 para cada instancia de información lateral, datos de transición que incluyen dos partes independientemente asignables que, en combinación, definen un punto en el tiempo para iniciar una transición desde un establecimiento de reconstrucción actual al establecimiento de reconstrucción deseado que se especifica por la instancia de información lateral, y un punto en el tiempo para completar la transición.

25 En conformidad con un cuarto aspecto de la idea inventiva, se da a conocer un método de decodificación, un decodificador y un producto de programa informático para decodificar un contenido de audio multicanal.

30 Los métodos, decodificadores y los productos de programa informático en conformidad con el cuarto aspecto de la idea inventiva están previstos para la cooperación con los métodos, codificadores y productos de programa informático de conformidad con el tercer aspecto, y pueden tener características y ventajas en correspondencia.

35 Los métodos, decodificadores y productos de programa informático de conformidad con el cuarto aspecto, pueden tener, en general, características y ventajas en común con los métodos, decodificadores y productos de programa informático de conformidad con el segundo aspecto.

En conformidad con formas de realización a modo de ejemplo, se da a conocer un método para reconstruir objetos de audio sobre la base de un flujo de datos. El método comprende:

40 la recepción de un flujo de datos que comprende M señales de mezcla reducida que son combinaciones de N objetos de audio, en donde $N > 1$ y $M \leq N$, y la información lateral variable en el tiempo incluye parámetros que permiten la reconstrucción de un conjunto de objetos de audio formados sobre la base de los N objetos de audio a partir de las M señales de mezcla reducida; y

45 la reconstrucción, sobre la base de las M señales de mezcla reducida y de la información lateral, del conjunto de objetos de audio formado sobre la base de los N objetos de audio,

50 en donde el flujo de datos comprende una pluralidad de instancias de información lateral, en donde el flujo de datos comprende, además, para cada instancia de información lateral, datos de transición que incluyen dos partes independientemente asignables que, en combinación, definen un punto en el tiempo para iniciar una transición desde un establecimiento de reconstrucción actual a un establecimiento de reconstrucción deseado especificado por la instancia de información lateral, y un punto en el tiempo para completar la transición, y en donde la reconstrucción del conjunto de objetos de audio formado sobre la base de los N objetos de audio comprende:

55 la realización de la reconstrucción en conformidad con un establecimiento de reconstrucción actual;

el inicio, en un punto en el tiempo definido por los datos de transición para una instancia de información lateral, una transición desde el establecimiento de reconstrucción actual a un establecimiento de reconstrucción deseado, que se especifica por la instancia de información lateral; y

60 completar la transición en el punto en el tiempo definido por los datos de transición para la instancia de información lateral.

65 Según se describió con anterioridad, la utilización de un formato de información lateral que incluye datos de transición que definen puntos en el tiempo para iniciar y puntos en el tiempo para completar las transiciones de los establecimientos de reconstrucción actuales a los establecimientos de reconstrucción deseados respectivos, p.ej., facilita el remuestreo de la información lateral.

El flujo de datos puede, a modo de ejemplo, recibirse en la forma de un flujo de bits, p.ej., generado en un lado del codificador.

5 La reconstrucción, sobre la base de las M señales de mezcla reducida y de la información lateral, del conjunto de objetos de audio formado sobre la base de los N objetos de audio, puede, a modo de ejemplo, incluir la formación de al menos una combinación lineal de las señales de mezcla reducida que emplean coeficientes determinados sobre la base de la información lateral. La reconstrucción, sobre la base de las M señales de mezcla reducida y la información lateral, del conjunto de objetos de audio formados sobre la base de los N objetos de audio, puede, a modo de ejemplo, incluir la formación de combinaciones lineales de las señales de mezcla reducida y, de modo opcional, una o más señales adicionales (p.ej., descorrelacionadas) derivadas de las señales de mezcla reducida, que utilizan coeficientes determinados sobre la base de la información lateral.

15 En conformidad con una forma de realización a modo de ejemplo, el flujo de datos puede comprender, además, metadatos de agrupamiento variables en el tiempo para el conjunto de objetos de audio formado sobre la base de los N objetos de audio, los metadatos de agrupamiento que incluyen las posiciones espaciales para el conjunto de objetos de audio formado sobre la base de los N objetos de audio. El flujo de datos puede comprender una pluralidad de instancias de metadatos de agrupamiento, y el flujo de datos puede comprender, además, para cada instancia de metadatos de agrupamiento, datos de transición que incluyen dos partes independientemente asignables que, en combinación, definen un punto en el tiempo para iniciar una transición desde un establecimiento de presentación actual a un establecimiento de presentación deseado especificado por la instancia de metadatos de agrupamiento, y un punto en el tiempo para completar la transición al establecimiento de presentación deseado especificado por la existencia de metadatos de agrupamiento. El método puede incluir, además:

25 la utilización de los metadatos de agrupamiento para la presentación del conjunto reconstruido de objetos de audio formado sobre la base de los N objetos de audio hacia los canales de salida de una configuración de canal predefinida, cuya presentación comprende:

30 realizar una presentación de conformidad con un establecimiento de presentación actual;
 iniciar, en un punto en el tiempo definido por los datos de transición para una instancia de metadatos de agrupamiento, una transición desde el establecimiento de presentación actual a un establecimiento de presentación deseado que se especifica por la instancia de metadatos de agrupamiento; y

35 completar la transición al establecimiento de presentación deseado en un punto en el tiempo definido por los datos de transición para la instancia de metadatos de agrupamiento.

40 La configuración de canales predefinida puede, a modo de ejemplo, corresponder a una configuración de los canales de salida compatible con un sistema de reproducción particular, esto es, adecuado para la reproducción en un sistema de reproducción particular.

45 La presentación del conjunto reconstruido de objetos de audio formado sobre la base de los N objetos de audio hacia los canales de salida de una configuración de canal predefinida puede, a modo de ejemplo, incluir el mapeado de correspondencia, en un dispositivo de presentación, del conjunto reconstruido de señales de audio formadas sobre la base de los N objetos de audio hacia (una configuración predefinida de) canales de salida del dispositivo de presentación bajo el control de los metadatos del agrupamiento.

50 La presentación del conjunto reconstruido de objetos de audio formados sobre la base de los N objetos de audio hacia los canales de salida de una configuración de canal predefinida puede, a modo de ejemplo, incluir la formación de combinaciones lineales del conjunto reconstruido de objetos de audio formado sobre la base de los N objetos de audio, utilizando coeficientes determinados sobre la base de los metadatos de agrupamiento.

55 En conformidad con una forma de realización a modo de ejemplo, los puntos respectivos en el tiempo definidos por los datos de transición para las respectivas instancias de metadatos de agrupamiento pueden coincidir con los respectivos puntos en el tiempo definidos por los datos de transición para las correspondientes instancias de información lateral.

De conformidad con una forma de realización, el método puede comprender, además:

60 realizar al menos parte de la reconstrucción y al menos parte de la presentación como una operación combinada correspondiente a una primera matriz formada como un producto matricial de una matriz de reconstrucción y una matriz de presentación asociada con un establecimiento de reconstrucción actual y un establecimiento de presentación actual, respectivamente;

65 iniciar, en un punto en el tiempo definido por los datos de transición para una instancia de información lateral y una instancia de metadatos del agrupamiento, una transición combinada desde los estabas de reconstrucción y

presentación actuales a los establecimientos de reconstrucción y presentación deseados que se especifican por la instancia de información lateral y la instancia de metadatos del agrupamiento, respectivamente; y

5 realizar la transición combinada en un punto en el tiempo definido por los datos de transición para la instancia de información lateral y la instancia de metadatos de agrupamiento, en donde la transición combinada incluye la interpolación entre elementos matriciales de la primera matriz y elementos matriciales de una segunda matriz formada como un producto matricial de una matriz de reconstrucción y de una matriz de presentación asociadas con el establecimiento de reconstrucción deseado y el establecimiento de presentación deseado, respectivamente.

10 Realizando una transición combinada en el sentido anterior, en lugar de transiciones separadas de establecimientos de reconstrucción y de establecimientos de presente, menos parámetros/coeficientes necesitan interpolarse, lo que permite una reducción de la complejidad del cálculo.

15 Ha de entenderse que una matriz, tal como una matriz de reconstrucción o una matriz de presentación, según son objeto de referencia en la presente forma de realización a modo de ejemplo, pueden, por ejemplo, estar constituidas por una fila única o una columna única y, por lo tanto, pueden corresponder a un vector.

20 La reconstrucción de objetos de audio a partir de las señales de mezcla reducida se suele realizar utilizando diferentes matrices de reconstrucción en diferentes bandas de frecuencia, mientras que la presentación se suele realizar utilizando la misma matriz de presentación para todas las frecuencias. En tales casos, una matriz correspondiente a una operación combinada de reconstrucción y de presentación, a modo de ejemplo, las primera y segunda matrices objeto de referencia en la presente forma de realización a modo de ejemplo, pueden normalmente ser dependientes de la frecuencia, esto es, diferentes valores para los elementos matriciales se pueden emplear normalmente para diferentes bandas de frecuencias.

25 De conformidad con una forma de realización a modo de ejemplo, el conjunto de objetos de audio formado sobre la base de los N objetos de audio puede coincidir con los N objetos de audio, esto es, el método puede comprender la reconstrucción de los N objetos de audio sobre la base de las M señales de mezcla reducida y de la información lateral.

30 Como alternativa, el conjunto de objetos de audio formado sobre la base de los N objetos de audio puede comprender una pluralidad de objetos de audio que son combinaciones de los N objetos de audio, y cuyo número es menor que N, esto es, el método puede comprender la reconstrucción de estas combinaciones de los N objetos de audio sobre la base de las M señales de mezcla reducida y de la información lateral.

35 En conformidad con una forma de realización a modo de ejemplo, el flujo de datos puede comprender, además, metadatos de mezcla reducida para las M señales de mezcla reducida que incluyen posiciones espaciales variables en el tiempo asociadas con las M señales de mezcla reducida. El flujo de datos puede comprender una pluralidad de instancias de metadatos de mezcla reducida, y el flujo de datos puede comprender, además, para cada instancia de metadatos de mezcla reducida, datos de transición incluyendo dos partes independientemente asignables que, en combinación, definen un punto en el tiempo para iniciar una transición desde un establecimiento de presentación de mezcla reducida actual a un establecimiento de presentación de mezcla reducida deseado especificado por la instancia de metadatos de mezcla reducida, y un punto en el tiempo para completar la transición al establecimiento de presentación de mezcla reducida deseado que se especifica por la instancia de metadatos de mezcla reducida. El método puede comprender, además:

50 en una condición en la que el decodificador es utilizable (o está configurado) para soportar una reconstrucción de objetos de audio, realizar la etapa de reconstrucción, sobre la base de las M señales de mezcla reducida y de la información lateral, del conjunto de objetos de audio formado sobre la base de los N objetos de audio; y

en una condición en la que el decodificador no es utilizable (o no está configurado) para soportar una reconstrucción de objetos de audio, proporcionar las metodologías de mezcla reducida y las M señales de mezcla reducida para la presentación de las M señales de mezcla reducida.

55 En caso de que el decodificador sea utilizable para soportar una reconstrucción de objetos de audio y el flujo de datos comprende, además, metadatos del agrupamiento asociados con el conjunto de objetos de audio formado sobre la base de los N objetos de audio, el decodificador puede, p.ej., proporcionar el conjunto reconstruido de objetos de audio para los metadatos del agrupamiento para la presentación del conjunto reconstruido de objetos de audio.

60 En caso de que el decodificador no sea utilizable para soportar una reconstrucción de objetos de audio, puede descartarse, a modo de ejemplo, la información lateral y, si fuera aplicable, los metadatos del agrupamiento, y proporcionar los metadatos de mezcla reducida y las M señales de mezcla reducida como salida. En tal caso, la salida puede utilizarse por un dispositivo de presentación para presentar las M señales de mezcla reducida hacia los canales de salida del dispositivo de presentación.

65

De modo opcional, el método puede comprender, además, la presentación de las M señales de mezcla reducida hacia canales de salida de una configuración de salida predefinida, p.ej., hacia los canales de salida de un dispositivo de presentación, o hacia los canales de salida del decodificador (en caso de que el decodificador tenga capacidades de presentación), sobre la base de los metadatos de mezcla reducida.

5 En conformidad con forma de realización a modo de ejemplo, se da a conocer un decodificador para la reconstrucción de objetos de audio sobre la base de un flujo de datos. El decodificador comprende:

10 un componente de recepción configurado para recibir un flujo de datos que comprende M señales de mezcla reducida que son combinaciones de N objetos de audio, en donde $N > 1$ y $M \leq N$, e información lateral variable en el tiempo que incluye parámetros que permiten la reconstrucción de un conjunto de objetos de audio formado sobre la base de los N objetos de audio a partir de las M señales de mezcla reducida; y

15 un componente de reconstrucción configurado para reconstruir, sobre la base de las M señales de mezcla reducida y de la información lateral, el conjunto de objetos de audio formado sobre la base de los N objetos de audio,

20 en donde el flujo de datos comprende una pluralidad de instancias de información lateral asociadas, y en donde el flujo de datos comprende, además, para cada instancia de información lateral, datos de transición que incluyen dos partes independientemente asignables que, en combinación, definen un punto en el tiempo para iniciar una transición desde un establecimiento de reconstrucción actual a un establecimiento de reconstrucción deseado, que se especifica por la instancia de información lateral, y un punto en el tiempo para completar la transición. El componente de reconstrucción está configurado para reconstruir el conjunto de objetos de audio formado sobre la base de los N objetos de audio por al menos:

25 la realización de una reconstrucción en conformidad con un establecimiento de reconstrucción actual;

la iniciación, en un punto en el tiempo definido por los datos de transición para una instancia de información lateral, de una transición desde el establecimiento de reconstrucción actual a un establecimiento de reconstrucción deseado especificado por la instancia de información lateral; y

30 completar la transición en un punto en el tiempo definido por los datos de transición para la instancia de información lateral.

35 De conformidad con una forma de realización a modo de ejemplo, el método dentro del tercer o de cuarto aspecto de la idea inventiva puede comprender, además, la generación de una o más instancias de información lateral adicionales que especifican prácticamente el mismo establecimiento de reconstrucción como una instancia de información lateral que precede directamente o sucede directamente a las una o más instancias de información lateral adicionales. Formas de realización a modo de ejemplo son también consideradas con instancias de metadatos agrupamiento adicional y/o instancias de metadatos de mezcla reducida que se generan en una forma análoga.

40 Según se describió con anterioridad, el remuestreo de la información lateral generando más instancias de información lateral puede ser ventajoso en varias situaciones, tales como cuando las señales de audio/objetos de audio y la información lateral asociada se codifican utilizando un dispositivo *códec* de audio basado en tramas; desde entonces, es deseable tener al menos una instancia de información lateral para cada trama de *códec* de audio. En un lado del codificador, las instancias de información lateral proporcionadas por un componente de análisis pueden, a modo de ejemplo, distribuirse en el tiempo de tal manera que no coincidan con una tasa de tramas de las señales de mezcla reducida proporcionadas por un componente de mezcla reducida, y la información lateral puede, por lo tanto, ventajosamente ser objeto de remuestreo introduciendo nuevas instancias de información lateral de modo que exista al menos una instancia de información lateral para cada trama de las señales de mezcla reducida. De modo similar, en un lado del decodificador, las instancias de información lateral recibidas pueden, a modo de ejemplo, distribuirse en el tiempo de tal manera que no coincidan una tasa de tramas de las señales de mezcla reducida recibidas y la información lateral puede, por lo tanto, ventajosamente ser objeto de remuestreo introduciendo nuevas instancias de información lateral de modo que exista al menos una instancia de información lateral para cada trama de las señales de mezcla reducida.

55 Una instancia de información lateral adicional puede, a modo de ejemplo, generarse para un punto en el tiempo seleccionado: copiando la instancia de información lateral que sucede directamente la instancia de información lateral adicional y determinando los datos de transición para la instancia de información lateral adicional sobre la base del punto en el tiempo seleccionado y de los puntos en el tiempo definidos por los datos de transición para la instancia de información lateral en sucesión.

60 En conformidad con un quinto aspecto de la idea inventiva, se da a conocer un método, un dispositivo y un producto de programa informático para transcodificar información lateral codificada junto con M señales de audio en un flujo de datos.

65

Los métodos, los dispositivos y productos de programa informático en conformidad con el quinto aspecto de la idea inventiva están previstos para la cooperación con los métodos, codificadores, decodificador y productos de programa informático de conformidad con el tercero y cuarto aspectos de la idea inventiva pueden tener características y ventajas correspondientes.

5 En conformidad con formas de realización a modo de ejemplo, se da a conocer un método para la transcodificación de información lateral codificada junto con M señales de audio en un flujo de datos. El método comprende:

la recepción de un flujo de datos;

10 la extracción, desde el flujo de datos, de M señales de audio e información lateral variable en el tiempo asociada que incluye parámetros que permiten la reconstrucción de un conjunto de objetos de audio a partir de las M señales de audio en donde $M \geq 1$ y en donde la información lateral extraída incluye:

15 una pluralidad de instancias de información lateral que especifican respectivos establecimientos de reconstrucción deseados para reconstruir los objetos de audio, y

20 para cada instancia de información lateral, datos de transición que incluyen dos partes independientemente asignables que, en combinación, definen un punto en el tiempo para iniciar una transición desde un establecimiento de reconstrucción actual al establecimiento de reconstrucción deseado que se especifica por la instancia de información lateral, y un punto en el tiempo para completar la transición;

25 la generación de una o más instancias de información lateral adicionales que especifican prácticamente el mismo establecimiento de reconstrucción que una instancia de información lateral que precede o sucede, directamente, a las una o más instancias de información lateral adicionales; y

la inclusión de las M señales de audio y de la información lateral en un flujo de datos.

30 En la presente forma de realización a modo de ejemplo, las una o más instancias de información lateral adicionales pueden generarse después de que la información lateral haya sido extraída desde el flujo de datos recibido, y las una o más instancias de información lateral generadas pueden incluirse entonces en un flujo de datos junto con las M señales de audio y las otras instancias de información lateral.

35 Según se describió con anterioridad en relación con el tercer aspecto de la idea inventiva, el remuestreo de la información lateral generando más instancias de información lateral puede ser ventajoso en varias situaciones, tales como cuando las señales de audio/objetos y la información lateral asociada se codifican utilizando un dispositivo *códec* de audio basado en tramas, y desde entonces, es deseable tener al menos una instancia de información lateral para cada trama de *códec* de audio.

40 Formas de realización son también consideradas en las que el flujo de datos comprende, además, metadatos de agrupamiento y/o metadatos de mezcla reducida, según se describe en relación con el tercer y cuarto aspectos de la idea inventiva, y en donde el método comprende, además, la generación de instancias de metadatos de mezcla reducida adicionales y/o instancias de metadatos de agrupamiento, de forma análoga a cómo se generan las instancias de información lateral adicionales.

45 En conformidad con una forma de realización a modo de ejemplo, las M señales de audio pueden codificarse en el flujo de datos recibidos en conformidad con una primera tasa de tramas y el método puede comprender, además:

50 el procesamiento de las M señales de audio para cambiar la tasa de tramas en conformidad con la cual se codifican las M señales de mezcla reducida para una segunda tasa de tramas diferente de la primera tasa de tramas; y

un remuestreo de la información lateral para coincidir, y/o para ser compatible con, la segunda tasa de trama generando al menos las una o más instancias de información lateral adicional.

55 Según se describió con anterioridad en relación con el tercer aspecto, puede ser ventajoso en varias situaciones procesar las señales de audio con el fin de cambiar la tasa de tramas utilizada para su codificación, a modo de ejemplo, de modo que la tasa de tramas modificada coincida con la tasa de tramas de contenido de vídeo de una señal audiovisual a la que pertenecen las señales de audio. La presencia de datos de transición parará codificada instancia de información lateral facilita el remuestreo de la información lateral, según se describió con anterioridad en relación con el tercer aspecto. La información lateral puede ser objeto de remuestreo para coincidir la nueva tasa de trama p.ej., generando instancias de información lateral adicionales de modo que exista al menos una instancia de información lateral para cada trama de las señales de audio procesadas.

65 En conformidad con formas de realización a modo de ejemplo, se da a conocer un dispositivo para la transcodificación de información lateral codificada junto con M señales de audio en un flujo de datos. El dispositivo comprende:

- 5 un componente de recepción configurado para recibir un flujo de datos y para extraer, a partir del flujo de datos, M señales de audio e información lateral variable en el tiempo que incluyen parámetros que permiten la reconstrucción de un conjunto de objetos de audio a partir de las M señales de audio, en donde $M \geq 1$ y en donde la información lateral extraída incluye:
- una pluralidad de instancias de información lateral que especifican los establecimientos de reconstrucción deseados respectivos para reconstruir los objetos de audio, y
- 10 para cada instancia de información lateral, datos de transición que incluyen dos partes independientemente asignables que, en combinación, definen un punto en el tiempo para iniciar una transición desde un establecimiento de reconstrucción actual al establecimiento de reconstrucción deseado que se especifica por la instancia de información lateral y un punto en el tiempo para completar la transición.
- 15 El dispositivo comprende, además:
- un componente de remuestreo configurado para generar una o más instancias de información lateral adicionales que especifican prácticamente el mismo establecimiento de reconstrucción que una instancia de información lateral que precede, o sucede, directamente, a las una o más instancias de información lateral adicionales; y
- 20 un componente de multiplexación configurado para incluir las M señales de audio y la información lateral en un flujo de datos.
- 25 En conformidad con una forma de realización a modo de ejemplo, el método dentro del tercer, cuarto o quinto aspecto puede incluir, además: el cálculo de una diferencia entre un primer establecimiento de reconstrucción deseado que se especifica por una primera instancia de información lateral y uno o más establecimientos de reconstrucción deseados especificados por una o más instancias de información lateral que suceden directamente a la primera instancia de información lateral; y la eliminación de las una o más instancias de información lateral en respuesta a la diferencia calculada que es inferior a un valor umbral predefinido. Formas de realización a modo de
- 30 ejemplo son también consideradas en las que las instancias de metadatos del agrupamiento y/o las instancias de metadatos de mezcla reducida se eliminan de una forma análoga.
- Eliminando las instancias de información lateral en conformidad con la presente forma de realización a modo de ejemplo, pueden evitarse cálculos innecesarios sobre la base de estas instancias de información lateral, p.ej., durante la reconstrucción en un lado del decodificador. Estableciendo el valor umbral predefinido a un nivel adecuado (p.ej., bastante bajo), las instancias de información lateral pueden eliminarse mientras que la calidad de la reproducción y/o la fidelidad de las señales de audio reconstruidas se mantiene al menos aproximadamente.
- 35 La diferencia entre los establecimientos de reconstrucción deseados puede, a modo de ejemplo, calcularse sobre la base de las diferencias entre los respectivos valores para un conjunto de coeficientes utilizados como parte de la reconstrucción.
- 40 De conformidad con las formas de realización a modo de ejemplo dentro de los tercero, cuarto o quinto aspectos, las dos partes independientemente asignables de los datos de transición para cada instancia de información lateral pueden ser:
- 45 una marca temporal que indica el punto en el tiempo para iniciar la transición al establecimiento de reconstrucción deseado y una marca temporal que indica el punto en el tiempo para completar la transición al establecimiento de reconstrucción deseado;
- 50 una marca temporal que indica el punto en el tiempo para iniciar la transición al establecimiento de reconstrucción deseado y un parámetro de duración de interpolación que indica una duración para alcanzar el establecimiento de reconstrucción deseado desde el punto en el tiempo para iniciar la transición al establecimiento de reconstrucción deseado; o
- 55 una marca temporal que indica el punto en el tiempo para completar la transición al establecimiento de reconstrucción deseado y un parámetro de duración de interpolación que indica una duración para alcanzar el establecimiento de reconstrucción deseado desde el punto en el tiempo para iniciar la transición al establecimiento de reconstrucción deseado.
- 60 Dicho de otro modo, los puntos en el tiempo para iniciar y para finalizar una transición pueden definirse en los datos de transición por dos marcas temporales que indican los puntos en el tiempo respectivos, o una combinación de una de las marcas temporales y un parámetro de duración de interpolación que indica una duración de la transición.
- 65 Las respectivas marcas temporales pueden indicar, a modo de ejemplo, los respectivos puntos en el tiempo haciendo referencia a una base de tiempos utilizada para representar las M señales de mezcla reducida y/o los N

objetos de audio.

En conformidad con forma de realización a modo de ejemplo dentro del tercero, cuarto o quinto aspectos, las dos partes independientemente asignables de los datos de transición para cada instancia de metadatos del agrupamiento pueden ser:

una marca temporal que indica el punto en el tiempo para iniciar la transición al establecimiento de presentación deseado y una marca temporal que indica el punto en el tiempo para completar la transición el establecimiento de presentación deseado;

una marca temporal que indica el punto en el tiempo para iniciar la transición al establecimiento de presentación deseado y un parámetro de duración de interpolación que indica una duración para alcanzar el establecimiento de presentación deseado desde el punto en el tiempo para iniciar la transición al establecimiento de presentación deseado; o

una marca temporal que indica el punto en el tiempo para completar la transición al establecimiento de presentación deseado y un parámetro de duración de interpolación que indica una duración para alcanzar el establecimiento de presentación deseado desde el punto en el tiempo para iniciar la transición al establecimiento de presentación deseado.

En conformidad con formas de realización a modo de ejemplo dentro del tercer, cuarto o quinto aspecto, las dos partes independientemente asignables de los datos de transición para cada instancia de metadatos de mezcla reducida pueden ser:

una marca temporal que indica el punto en el tiempo para iniciar la transición al establecimiento de presentación de mezcla reducida deseado y una marca temporal que indica el punto en el tiempo para completar la transición al establecimiento de presentación de mezcla reducida deseado;

una marca temporal que indica el punto en el tiempo para iniciar la transición al establecimiento de presentación de mezcla reducida deseado y un parámetro de duración de interpolación que indica una duración para alcanzar el establecimiento de presentación de mezcla reducida deseado desde el punto en el tiempo para iniciar la transición al establecimiento de presentación de mezcla reducida deseado; o

una marca temporal que indica el punto en el tiempo para completar la transición al establecimiento de presentación de mezcla reducida deseado y un parámetro de duración de interpolación que indica una duración para alcanzar el establecimiento de presentación de mezcla reducida deseado desde el punto en el tiempo para iniciar la transición al establecimiento de presentación de mezcla reducida deseado.

En conformidad con las formas de realización a modo de ejemplo, se da a conocer un producto de programa informático que comprende un soporte con instrucciones para realizar el método de cualquiera de los métodos dentro del tercero, cuarto o quinto aspectos de la idea inventiva.

IV. Formas de realización ejemplo

La Figura 1 ilustra un codificador 100 para codificar objetos de audio 120 en un flujo de datos 140 en conformidad con una a modo de ejemplo. El codificador 100 comprende un componente de recepción (no ilustrado), un componente de mezcla reducida 102, un componente de decodificador 104, un componente de análisis 106 y un componente de multiplexación 108. El funcionamiento del codificador 100 para codificar una trama temporal de datos de audio se describe a continuación. Sin embargo, ha de entenderse que el método siguiente se repite sobre la base de trama temporal. Lo mismo se aplica también a la descripción de las Figuras 2 a 5.

El componente de recepción recibe una pluralidad de objetos de audio (N objetos de audio) 120 y metadatos 122 asociados con los objetos de audio 120. Un objeto de audio, tal como aquí se utiliza, se refiere a una señal de audio que tiene una posición espacial asociada que suele variar en el tiempo (entre tramas temporales), esto es, la posición espacial es dinámica. Los metadatos 122 asociados con los objetos de audio 120 suelen comprender información que describe cómo los objetos de audio 120 han de presentarse para su reproducción en el lado del decodificador. En particular, los metadatos 122 asociados con los objetos de audio 120 incluyen información sobre la posición espacial de los objetos de audio 120 en el espacio tridimensional de la escena de audio. Las posiciones espaciales pueden representarse en coordenadas Cartesianas o por medio de ángulos de dirección, tales como azimut y elevación, opcionalmente aumentados con la distancia. Los metadatos 122 asociados con los objetos de audio 120 pueden comprender, además, el tamaño del objeto, el volumen del objeto, la importancia del objeto, el tipo de contenido del objeto, instrucciones de presentación específicas tales como aplicación de mejora de diálogo o exclusión de algunos altavoces de la presentación (así denominadas, máscaras de zonas) y/o otras propiedades de los objetos.

Según se describirá haciendo referencia a la Figura 4, los objetos de audio 120 pueden corresponder a una

representación simplificada de una escena de audio.

Los N objetos de audio 120 constituyen una entrada para el componente de mezcla reducida 102. El componente de mezcla reducida 102 calcula un número de M de señales de mezcla reducida 124 mediante la formación de combinaciones, que suelen ser combinaciones lineales, de los N objetos de audio 120. En la mayoría de los casos, el número de señales de mezcla reducida 124 es menor que el número de objetos de audio 120, esto es, $M < N$, de modo que se reduce la cantidad de datos que se incluyen en el flujo de datos 140. Sin embargo, para las aplicaciones en donde la tasa binaria objetivo del flujo de datos 140 es alta, el número de señales de mezcla reducida 124 puede ser igual al número de objetos de audio 120, esto es, $M = N$.

El componente de mezcla reducida 102 puede calcular, además, una o más señales de audio auxiliares 127, aquí etiquetadas por L señales de audio auxiliares 127. La función de las señales de audio auxiliares 127 es mejorar la reconstrucción de los N objetos de audio 120 en el lado del decodificador. Las señales de audio auxiliares 127 pueden corresponder a uno o más de los N objetos de audio 120, bien sea directamente, bien sea como una combinación de ellos. A modo de ejemplo, las señales de audio auxiliares 127 pueden corresponder a objetos particularmente importantes de los N objetos de audio 120, de modo que un objeto de audio 120 corresponda a un diálogo. La importancia puede reflejarse por, o derivarse a partir de los metadatos 122 asociados con los N objetos de audio 120.

Las M señales de mezcla reducida 124 y las L señales auxiliares 127 si están presentes, pueden codificarse posteriormente por el componente de codificador 104, aquí etiquetado como codificador base, para generar las M señales de mezcla reducida codificadas 126 y L señales auxiliares codificadas 129. El componente de codificador 104 puede ser un dispositivo *códec* de audio perceptual según se conoce en esta técnica. Ejemplos de dispositivos *códec* de audio perceptuales incluyen Dolby Digital y MPEG AAC.

En algunas formas de realización, el componente de mezcla reducida 102 puede asociar, además, las M señales de mezcla reducida 124 con metadatos 125. En particular, el componente de mezcla reducida 102 puede asociar cada señal de mezcla reducida 124 con una posición espacial e incluir la posición espacial en los metadatos 125. De modo similar a los metadatos 122 asociados con los objetos de audio 120, los metadatos 125 asociados con las señales de mezcla reducida 124 pueden comprender, también, parámetros relacionados con el tamaño, volumen, importancia y/o otras propiedades.

En particular, las posiciones espaciales asociadas con las señales de mezcla reducida 124 pueden calcularse sobre la base de las posiciones espaciales de los N objetos de audio 120. Puesto que las posiciones espaciales de los N objetos de audio 120 pueden ser dinámicas, esto es, variables en el tiempo, también las posiciones espaciales asociadas con las M señales de mezcla reducida 124 pueden ser dinámicas. Dicho de otro modo, las M señales de mezcla reducida 124 pueden por sí mismas interpretarse como objetos de audio.

El componente de análisis 106 calcula la información lateral 128 que incluye parámetros que permiten la reconstrucción de los N objetos de audio 120 (o una aproximación perceptualmente adecuada de los N objetos de audio 120) desde las M señales de mezcla reducida 124 y las L señales auxiliares 129, si están presentes. También la información lateral 128 puede ser variable en el tiempo. A modo de ejemplo, el componente de análisis 106 puede calcular la información lateral 128 analizando las M señales de mezcla reducida 124, las L señales auxiliares 129, si están presentes, y los N objetos de audio 120 en conformidad con cualquier técnica conocida para la codificación paramétrica. Como alternativa, el componente de análisis 106 puede calcular la información lateral 128 analizando los N objetos de audio y la información sobre cómo las M señales de mezcla reducida se crearon a partir de los N objetos de audio, por ejemplo, proporcionando una matriz de mezcla reducida (variable en el tiempo). En ese caso, las M señales de mezcla reducida 124 no se requieren estrictamente como una entrada para el componente de análisis 106.

Las M señales de mezcla reducida codificadas 126, las L señales auxiliares codificadas 129, la información lateral 128, los metadatos 122 asociados con los N objetos de audio y los metadatos 125 asociados con las señales de mezcla reducida son entonces objeto de entrada al componente de multiplexación 108 que incluye sus datos de entrada en un flujo de datos único 140 utilizando técnicas de multiplexación. El flujo de datos 140 puede, de este modo, incluir cuatro tipos de datos:

- a) M señales de mezcla reducida 126 (y de modo opcional, L señales auxiliares 129)
- b) Metadatos 125 asociados con las M señales de mezcla reducida,
- c) Información lateral 128 para la reconstrucción de los N objetos de audio a partir de las M señales de mezcla reducida, y
- d) Metadatos 122 asociados con los N objetos de audio.

Según se indicó con anterioridad, algunos sistemas de la técnica anterior para la codificación de objetos de audio

requieren que las M señales de mezcla reducida sean seleccionadas de modo que sean adecuadas para la reproducción en los canales de una configuración de altavoces como M canales, referida aquí como una mezcla reducida compatible con versiones anteriores. Dicho requisito de la técnica anterior restringe el cálculo de las señales de mezcla reducida por cuanto que los objetos de audio solamente pueden combinarse en una manera preferida. En consecuencia, en conformidad con la técnica anterior, las señales de mezcla reducida no se seleccionan a partir del punto de vista de optimar la reconstrucción de los objetos de audio en un lado del decodificador.

De forma opuesta a los sistemas de la técnica anterior, el componente de mezcla reducida 102 calcula las M señales de mezcla reducida 124 en una manera adaptativa de señales con respecto a los N objetos de audio. En particular, el componente de mezcla reducida 102 puede, para cada trama temporal, calcular las M señales de mezcla reducida 124 como la combinación de los objetos de audio 120 que optimiza actualmente algún criterio. El criterio se suele definir de modo que sea independiente con respecto a cualquier configuración de altavoces, tal como 5.1 u otra configuración de altavoces. Lo que antecede implica que las M señales de mezcla reducida 124, o al menos una de ellas, no estén restringidas a señales de audio que sean adecuadas para su reproducción en los canales de una configuración de altavoces con M canales. En consecuencia, el componente de mezcla reducida 102 puede adaptar las M señales de mezcla reducida 124 a la variación temporal de los N objetos de audio 120 (incluyendo la variación temporal de los metadatos 122 que incluyen posiciones espaciales de los N objetos de audio), con el fin de, p.ej., mejorar la reconstrucción de los objetos de audio 120 en el lado del decodificador.

El componente de mezcla reducida 102 puede aplicar diferentes criterios con el fin de calcular las M señales de mezcla reducida. En conformidad con un ejemplo, las M señales de mezcla reducida pueden calcularse de modo que la reconstrucción de los N objetos de audio, sobre la base de las M señales de mezcla reducida, sea optimizada. A modo de ejemplo, el componente de mezcla reducida 102 puede minimizar un error de reconstrucción formado a partir de los N objetos de audio 120 y una reconstrucción de los N objetos de audio basada en las M señales de mezcla reducida 124.

En conformidad con otro ejemplo, el criterio está basado en las posiciones espaciales, y en particular, en la proximidad espacial, de los N objetos de audio 120. Según describió con anterioridad, los N objetos de audio 120 tienen metadatos asociados 122 que incluyen las posiciones espaciales de los N objetos de audio 120. Sobre la base de los metadatos 122, se puede derivar la proximidad espacial de los N objetos de audio 120.

En más detalle, el componente de mezcla reducida 102 puede aplicar un primer procedimiento de agrupamiento con el fin de determinar las M señales de mezcla reducida 124. El primer procedimiento de agrupamiento puede comprender la asociación de los N objetos de audio 120 con M agrupamientos sobre la base de la proximidad espacial. Propiedades adicionales de los N objetos de audio 120 según se representa por los metadatos asociados 122, que incluyen el tamaño del objeto, el volumen del objeto, la importancia del objeto, pueden tenerse en cuenta también durante la asociación de los objetos de audio 120 con los M agrupamientos.

En conformidad con un ejemplo, el algoritmo denominado *K-means* bien conocido, con los metadatos 122 (posiciones espaciales) de los N objetos de audio como entrada, puede utilizarse para asociar los N objetos de audio 120 con los M agrupamientos sobre la base de la proximidad espacial. Las propiedades adicionales de los N objetos de audio 120 pueden utilizarse como factores de ponderación en el algoritmo *K-means*.

En conformidad con otra realización a modo de ejemplo, el primer procedimiento de agrupamiento puede basarse en un procedimiento de selección que utiliza la importancia de los objetos de audio, según se proporciona por los metadatos 122, como un criterio de selección. Con más detalle, el componente de mezcla reducida 102 puede pasar a través de los más importantes objetos de audio 120 de modo que una o más de las M señales de mezcla reducida correspondan a uno o más de los N objetos de audio 120. Los restantes, menos importantes, objetos de audio pueden asociarse con agrupamientos basados en la proximidad espacial según se describió con anterioridad.

Ejemplos adicionales de agrupamiento de objetos de audio se proporcionan en la solicitud de patente provisional de los Estados Unidos n° 61/865,072 o solicitudes posteriores que reivindican la prioridad de esa solicitud.

En conformidad con otro ejemplo de realización, el primer procedimiento de agrupamiento puede asociar un objeto de audio 120 con más de uno de los M agrupamientos. A modo de ejemplo, un objeto de audio 120 puede distribuirse sobre los M agrupamientos, en donde la distribución p.ej., depende de la posición espacial del objeto de audio 120 y de modo opcional, también de las propiedades del objeto de audio incluyendo el tamaño del objeto, el volumen del objeto, la importancia del objeto, etc. La distribución puede reflejarse por porcentajes, de modo que un objeto de audio, a modo de ejemplo, sea distribuido en tres agrupamientos en conformidad con los porcentajes del 20 %, 30 %, 50 %.

Una vez que los N objetos de audio 120 han sido asociados con los M agrupamientos, el componente de mezcla reducida 102 calcula una señal de mezcla reducida 124 para cada agrupamiento mediante la formación de una combinación, normalmente una combinación lineal, de los objetos de audio 120 asociados con el agrupamiento. En condiciones normales, el componente de mezcla reducida 102 puede utilizar parámetros incluidos en los metadatos

122 asociados con los objetos de audio 120 como factores de ponderación cuando se forma la combinación. A modo de ejemplo, los objetos de audio 120 que están asociados por un agrupamiento pueden ponderarse en conformidad con un tamaño de objeto de audio, el volumen del objeto, importancia del objeto, posición del objeto, distancia desde un objeto con respecto a una posición espacial asociada con el agrupamiento (véase detalles a continuación), etc.

5 En el caso en donde los objetos de audio 120 son distribuidos entre los M agrupamientos, los porcentajes que reflejan la distribución pueden utilizarse como factores de ponderación cuando se forme la combinación.

El primer procedimiento de agrupamiento es ventajoso por cuanto que permite facilitar la asociación de cada una de las M señales de mezcla reducida 124 con una posición espacial. A modo de ejemplo, el componente de mezcla reducida 120 puede calcular una posición espacial de una señal de mezcla reducida 124 que corresponde a un agrupamiento basado en las posiciones espaciales de los objetos de audio 120 que se asocian con el agrupamiento. El centroide o un centroide ponderado de las posiciones espaciales de los objetos de audio que se asocian con el agrupamiento pueden utilizarse para esta finalidad. En caso de un centroide ponderado, los mismos factores de ponderación pueden utilizarse como cuando se forma la combinación de los objetos de audio 120 asociados con el agrupamiento.

10

15

La Figura 2 ilustra un decodificador 200 que corresponde al codificador 100 de la Figura 1. El decodificador 200 es del tipo que soporta la reconstrucción de objetos de audio. El decodificador 200 comprende un componente de recepción 208, un componente de decodificador 204 y un componente de reconstrucción 206. El decodificador 200 puede comprender, además, un dispositivo de presentación 210. Como alternativa, el decodificador 200 puede estar acoplado a un dispositivo de presentación 210 que forma parte de un sistema de reproducción.

20

El componente de recepción 208 está configurado para recibir un flujo de datos 240 a partir del codificador 100. El componente de recepción 208 comprende un componente de demultiplexación configurado para demultiplexar el flujo de datos recibido 240 en sus componentes, en este caso, M señales de mezcla reducida codificadas 226, opcionalmente L señales auxiliares codificadas 229, información lateral 228 para la reconstrucción de N objetos de audio a partir de las M señales de mezcla reducida y las L señales auxiliares, y metadatos 222 asociados con los N objetos de audio.

25

El componente de decodificador 204 procesa las M señales de mezcla reducida codificadas 226 para generar M señales de mezcla reducida 224 y, opcionalmente, L señales auxiliares 227. Según se describió con anterioridad, las M señales de mezcla reducida 224 se formaron, de forma adaptativa, sobre el lado del codificador a partir de los N objetos de audio, es decir, mediante la formación de combinaciones de los N objetos de audio en conformidad con un criterio que es independiente de cualquier configuración de altavoces.

30

El componente de reconstrucción de objetos 206 reconstruye luego los N objetos de audio 220 (o una aproximación perceptualmente adecuada de estos objetos de audio) sobre la base de las M señales de mezcla reducida 224 y opcionalmente, las L señales auxiliares 227 guiadas por la información lateral 228 derivada en el lado del codificador. El componente de reconstrucción de objeto 206 puede aplicar cualquier técnica conocida para dicha reconstrucción paramétrica de los objetos de audio.

35

40

Los N objetos de audio reconstruidos 220 se procesan luego por el dispositivo de presentación 210 utilizando los metadatos 222 asociados con los objetos de audio 222 y el conocimiento sobre la configuración de canales del sistema de reproducción con el fin de generar una señal de salida multicanal 230 adecuada para la reproducción. Configuraciones de reproducción en altavoces típicas incluyen 22.2 y 11.1. La reproducción en los sistemas de altavoces de barra acústica o auriculares (presentación binaural) es también posible con los dispositivos de presentación dedicados para dichos sistemas de reproducción.

45

La Figura 3 ilustra un decodificador de baja complejidad 300 que corresponde al codificador 100 de la Figura 1. El decodificador 300 no soporta la reconstrucción de objetos de audio. El decodificador 300 comprende un componente de recepción 308 y un componente de decodificación 304. El decodificador 300 puede comprender, además, un dispositivo de presentación 310. Como alternativa, el decodificador está acoplado a un dispositivo de presentación 310 que forma parte de un sistema de reproducción.

50

Según se describió con anterioridad, los sistemas de la técnica anterior que utilizan una mezcla reducida compatible con versiones anteriores (tal como una mezcla reducida 5.1) es decir, una mezcla reducida que comprende M señales de mezcla reducida que son adecuadas para la reproducción directa en un sistema de reproducción con M canales, permiten fácilmente una decodificación de baja complejidad para sistemas de reproducción de legado (que, p.ej., solamente soportan un establecimiento de altavoces multicanal 5.1). Dichos sistemas de la técnica anterior suelen decodificar las señales de mezcla reducida compatibles hacia atrás y desechan las partes adicionales del flujo de datos tales como la información lateral (véase elemento 228 ilustrado en la Figura 2) y los metadatos asociados con los objetos de audio (véase elemento de referencia 222 de la Figura 2). Sin embargo, cuando las señales de mezcla reducida se forman, de forma adaptativa, según se describió con anterioridad, las señales de mezcla reducida no suelen ser adecuadas para la reproducción directa en un sistema de legado.

55

60

65

El decodificador 300 es un ejemplo de un decodificador que permite una decodificación de baja complejidad de M

señales de mezcla reducida que están formadas, de forma adaptativa, para la reproducción en un sistema de reproducción de legado que solamente soporta una configuración de reproducción particular.

5 El componente de recepción 308 recibe un flujo binario 340 procedente de un codificador, tal como el codificador 100 representado en la Figura 1. El componente de recepción 308 demultiplexa el flujo binario 340 en sus componentes. En este caso, el componente de recepción 308 solamente mantendrá las M señales de mezcla reducida codificadas 326 y los metadatos 325 asociados con las M señales de mezcla reducida. Los otros componentes del flujo de datos 340, tal como las L señales auxiliares (véase elemento de referencia 229 en la Figura 2), los metadatos asociados con los N objetos de audio (véase elemento de referencia 222 en la Figura 2) y la información lateral (véase elemento de referencia 228 de la Figura 2) son descartados.

15 El componente de decodificación 304 decodifica las M señales de mezcla reducida codificadas 326 para generar M señales de mezcla reducida 324. Las M señales de mezcla reducida son luego, junto con los metadatos de mezcla reducida, objeto de entrada para el dispositivo de presentación 310 que presenta las M señales de mezcla reducida hacia una salida multicanal 330 correspondiente a un formato de reproducción de legado (que suele tener M canales). Puesto que los metadatos de mezcla reducida 325 comprenden posiciones espaciales de las M señales de mezcla reducida 324, el dispositivo de presentación 310 puede ser normalmente similar al dispositivo de presentación 210 ilustrado en la Figura 2, con la única diferencia de que el dispositivo de presentación 310 toma ahora las M señales de mezcla reducida 324 y los metadatos 325 asociados con las M señales de mezcla reducida 324 como una entrada en lugar de los objetos de audio 220 en sus metadatos asociados 222.

Según se describió con anterioridad en relación con la Figura 1, los N objetos de audio 120 pueden corresponder a una representación simplificada de una escena de audio.

25 En general, una escena de audio puede comprender objetos de audio y canales de audio. Por el término de un canal de audio se considera aquí una señal de audio que corresponde a un canal de una configuración de altavoces multicanal. Ejemplos de dicha configuración de altavoces multicanal incluyen una configuración 22.2, una configuración 11.1, etc. Un canal de audio puede interpretarse como un objeto de audio estático que tiene una posición espacial correspondiente a la posición del altavoz del canal.

30 En algunos casos, el número de objetos de audio y de canales de audio en la escena de audio pueden ser grandes, tal como más de 100 objetos de audio y 1 a 24 canales de audio. Si la totalidad de estos objetos de audio/canales han de reconstruirse en el lado del decodificador, se requiere una alta potencia de cálculo. Además, la tasa de datos resultante asociados con los metadatos de objetos y la información lateral serán generalmente muy altos si se proporcionan numerosos objetos como entrada. Por este motivo, es ventajoso simplificar la escena de audio con el fin de reducir el número de objetos de audio a reconstruirse en el lado del decodificador. Para esta finalidad, el codificador puede comprender un componente de agrupamiento que reduce el número de objetos de audio en la escena de audio sobre la base de un segundo procedimiento de agrupamiento. El segundo procedimiento de agrupamiento tiene como objetivo utilizar la redundancia espacial presente en la escena de audio, tal como los objetos de audio que tienen localizaciones iguales o muy similares. Además, la importancia perceptual de los objetos de audio puede tenerse también en cuenta. Por lo general, dicho componente de agrupamiento puede disponerse en secuencia o en paralelo con el componente de mezcla reducida 102 de la Figura 1. La disposición secuencial se describirá con referencia a la Figura 4 y la disposición en paralelo se describirá haciendo referencia a la Figura 5.

45 La Figura 4 ilustra un codificador 400. Además de los componentes descritos con referencia a la Figura 1, el codificador 400 comprende un componente de agrupamiento 409. El componente de agrupamiento 409 está dispuesto en secuencia con el componente de mezcla reducida 102, lo que significa que la salida del componente de agrupamiento 409 es objeto de entrada para el componente de mezcla reducida 102.

50 El componente de agrupamiento 409 toma objetos de audio 421a y/o los canales de audio 421b como entrada junto con los metadatos asociados 423 incluyendo las posiciones espaciales de los objetos de audio 421a. El componente de agrupamiento 409 convierte los canales de audio 421b en objetos de audio estáticos asociando cada canal de audio 421b con la posición espacial de la posición de altavoz correspondiente al canal de audio 421b. Los objetos de audio 421a y los objetos de audio estáticos formados a partir de los canales de audio 421b pueden considerarse como una primera pluralidad de objetos de audio 421.

El componente de agrupamiento 409 suele reducir la primera pluralidad de objetos de audio 421 a una segunda pluralidad de objetos de audio, que corresponde en este caso a los N objetos de audio 120 de la Figura 1. Para esta finalidad, el componente de agrupamiento 409 puede aplicar un segundo procedimiento de agrupamiento.

60 El segundo procedimiento de agrupamiento suele ser similar al primer procedimiento de agrupamiento anteriormente descrito con respecto al componente de mezcla reducida 102. La descripción del primer procedimiento de agrupamiento se aplica también, por lo tanto, al segundo procedimiento de agrupamiento.

65 En particular, el segundo procedimiento de agrupamiento implica la asociación de la primera pluralidad de objetos de audio 121 con al menos un agrupamiento, en este caso, N agrupamientos, sobre la base de la proximidad espacial

de la primera pluralidad de los objetos de audio 121. Según se describió como anterior, la asociación con los agrupamientos puede basarse también en otras propiedades de los objetos de audio según se representa por los metadatos 423. Cada agrupamiento se representa luego por un objeto que es una combinación (lineal) de los objetos de audio asociados con ese agrupamiento. En el ejemplo ilustrado, existen N agrupamientos y por ello, se generan N objetos de audio 120. El componente de agrupamiento 409 calcula, además, metadatos 122 para los N objetos de audio así generados 120. Los metadatos 122 incluyen posiciones espaciales de los N objetos de audio 120. La posición espacial de cada uno de los N objetos de audio 120 puede calcularse sobre la base de las posiciones espaciales de los objetos de audio asociados con el agrupamiento correspondiente. A modo de ejemplo, la posición espacial puede calcularse como un centroide o un centroide de ponderación de las posiciones espaciales de los objetos de audio asociados con el agrupamiento según se explicó con anterioridad haciendo referencia a la Figura 1.

Los N objetos de audio 120 generados por el componente de agrupamiento 409 son luego objeto de entrada para el componente de mezcla reducida 120 según se describió, además, haciendo referencia a la Figura 1.

La Figura 5 ilustra un codificador 500. Además de los componentes descritos con referencia a la Figura 1, el codificador 500 comprende un componente de agrupamiento 509. El componente de agrupamiento 509 está dispuesto en paralelo con el componente de mezcla reducida 102, lo que significa que el componente de mezcla reducida 102 y el componente de agrupamiento 509 tienen la misma entrada.

La entrada comprende una primera pluralidad de objetos de audio, correspondientes a los N objetos de audio 120 de la Figura 1, junto con los metadatos asociados 122 que incluyen las posiciones espaciales de la primera pluralidad de objetos de audio. La primera pluralidad de objetos de audio 120 puede, de modo similar a la primera pluralidad de objetos de audio 121 de la Figura 4, comprender objetos de audio y canales de audio que se convierten en objetos de audio estáticos. A diferencia de la disposición secuencial de la Figura 4, en donde el componente de mezcla reducida 102 opera sobre un número reducido de objetos de audio correspondientes a una versión simplificada de la escena de audio, el componente de mezcla reducida 102 de la Figura 5 opera sobre el contenido de audio completo de la escena de audio con el fin de generar M señales de mezcla reducida 124.

El componente de agrupamiento 509 es similar, en funcionalidad, al componente de agrupamiento 409 descrito con referencia a la Figura 4. En particular, el componente de agrupamiento 509 reduce la primera pluralidad de objetos de audio 120 a una segunda pluralidad de objetos de audio 521, aquí ilustrados por K objetos de audio en donde normalmente $M < K < N$ (para aplicaciones binarias altas $M \leq K \leq N$), aplicando el segundo procedimiento de agrupamiento descrito con anterioridad. La segunda pluralidad de objetos de audio 521 es, de este modo, un conjunto de objetos de audio formados sobre la base de los N objetos de audio 126. Además, el componente de agrupamiento 509 calcula los metadatos 522 para la segunda pluralidad de objetos de audio 521 (los K objetos de audio) incluyendo las posiciones espaciales de la segunda pluralidad de los objetos de audio 521. Los metadatos 522 se incluyen en el flujo de datos 540 por el componente de demultiplexación 108. El componente de análisis 106 calcula la información lateral 528 que permite la reconstrucción de la segunda pluralidad de objetos de audio 521, esto es, el conjunto de objetos de audio formado sobre la base de los N objetos de audio (en este caso, los K objetos de audio), a partir de las M señales de mezcla reducida 124. La información lateral 528 se incluye en el flujo de datos 540 por el componente de multiplexación 108. Como se examinó con anterioridad, el componente de análisis 106 puede, a modo de ejemplo, derivar la información lateral 528 analizando la segunda pluralidad de objetos de audio 521 y las M señales de mezcla reducida 124.

El flujo de datos 540 generado por el codificador 500 puede decodificarse, por lo general, por el decodificador 200 de la Figura 2 o el decodificador 300 de la Figura 3. Sin embargo, los objetos de audio reconstruidos 220 de la Figura 2 (etiquetados N objetos de audio) corresponden ahora a la segunda pluralidad de objetos de audio 521 (etiquetados K objetos de audio) de la Figura 5 y los metadatos 222 asociados con los objetos de audio (etiquetados como metadatos de N objetos de audio) corresponden ahora a los metadatos 522 de la segunda pluralidad de objetos de audio (etiquetados como metadatos de K objetos de audio) de la Figura 5.

En los sistemas de codificación/decodificación de audio basados en el objeto, la información lateral o los metadatos asociados con los objetos se suelen actualizar con relativamente poca frecuencia (escasamente) en tiempo para limitar la tasa de datos asociada. Los intervalos de actualización típicos para las posiciones de objetos pueden variar entre 10 y 500 milisegundos, dependiendo de la velocidad del objeto, de la exactitud de la posición requerida, del ancho de banda disponible para memorizar o transmittir metadatos, etc. Dichas escasas, o incluso irregulares actualizaciones de metadatos requieren una interpolación de metadatos y/o matrices de presentación (esto es, matrices utilizadas en la presentación) para muestras de audio entre dos instancias de metadatos posteriores. Sin interpolación, los cambios graduales consecuentes en la matriz de presentación pueden causar la presencia de artefactos de conmutación indeseables, sonidos de chasquidos, ruidos de cierres u otros artefactos indeseables como resultado de la denominada "salpicadura espectral" introducida por las actualizaciones matriciales graduales.

La Figura 6 ilustra un proceso conocido típico para calcular las matrices de presentación para presentar señales de audio u objetos de audio, sobre la base de un conjunto de instancias de metadatos. Según se ilustra en la Figura 6, un conjunto de instancias de metadatos (m_1 a m_4) 610 corresponde a un conjunto de puntos en el tiempo (t_1 a t_4)

que se indican por su posición a lo largo del eje de tiempos 620. Posteriormente, cada instancia de metadatos se convierte en una respectiva matriz de presentación (c1 a c4) 630, o establecimiento de presentación, que tiene validez en el mismo punto en el tiempo que la instancia de metadatos. De este modo, según se ilustra, la instancia de metadatos m1 crea una matriz de presentación c1 en el tiempo t1, la instancia de metadatos m2 crea una matriz de presentación c2 en el tiempo t2 y así sucesivamente. Por simplicidad, la Figura 6 ilustra solamente una matriz de presentación para cada instancia de metadatos m1 a m4. En sistemas prácticos, sin embargo, una matriz de presentación c1 puede comprender un conjunto de coeficientes de matrices de presentación o coeficientes de ganancias $c_{1,ij}$ a aplicarse a las respectivas señales de audio $x_i(t)$ para crear señales de salida $y_j(t)$:

$$y_j(t) = \sum_i x_i(t)c_{1,ij}.$$

Las matrices de presentación 630 suelen comprender coeficientes que representan valores de ganancias en diferentes puntos en el tiempo. Las instancias de metadatos se definen en algunos puntos en el tiempo discretos y para las muestras de audio entre los puntos en el tiempo de metadatos, la matriz de presentación es objeto de interpolación, según se indica por la línea de trazos 640 que conecta las matrices de presentación 630. Dicha interpolación puede realizarse de forma lineal, pero también se pueden utilizar otros métodos de interpolación (tal como interpolación de banda limitada, interpolación senoidal/cosenoidal y etc.). El intervalo de tiempo entre las instancias de metadatos (y las matrices de presentación correspondientes) se refiere como una "duración de interpolación" y dichos intervalos pueden ser uniformes o pueden ser diferentes, tal como la más larga duración de interpolación entre los tiempos t3 y t4 en comparación con la duración de interpolación entre los tiempos t2 y t3.

En numerosos casos, el cálculo de los coeficientes de matrices de presentación a partir de las instancias de metadatos es bien definido, pero el proceso inverso de calcular instancias de metadatos dada una matriz de presentación (interpolada) suele ser difícil o incluso imposible. A este respecto, el proceso de generar una matriz de presentación a partir de metadatos puede considerarse, a veces, como una función unidireccional criptográfica. El proceso de calcular nuevas instancias de metadatos entre instancias de metadatos existentes se refiere como un "remuestreo" de los metadatos. El remuestreo de metadatos se suele requerir durante algunas tareas de procesamiento de audio. A modo de ejemplo, cuando se edita un contenido de audio, mediante corte/fusión/mezcla y así sucesivamente, dichas ediciones pueden producirse entre instancias de metadatos. En este caso, se requiere un remuestreo de los metadatos. Otro de dichos casos es cuando las señales de audio y los metadatos asociados se codifican con un dispositivo *códec* de audio basado en la trama. En este caso, es deseable tener al menos una instancia de metadatos para cada trama del *códec* de audio, preferentemente con una marca temporal al inicio de esa trama de *códec*, para mejorar la resistencia a las pérdidas de tramas durante la transmisión. Además, la interpolación de metadatos es también ineficaz para algunos tipos de metadatos, tales como metadatos con valoraciones binarias, en donde las técnicas estándar derivarían el valor incorrecto más o menos cada segunda vez. A modo de ejemplo, si los indicadores binarios tales como máscaras de exclusión de zonas se utilizan para excluir algunos objetos de la presentación en algunos puntos en el tiempo, es prácticamente imposible estimar un conjunto válido de metadatos a partir de los coeficientes de la matriz de presentación o de las instancias próximas de metadatos. Este hecho se ilustra en la Figura 6 como un intento fallido para extrapolar o derivar una instancia de metadatos m3a a partir de los coeficientes de matrices de presentación en la duración de interpolación entre los tiempos t3 y t4. Según se ilustra en la Figura 6, las instancias de metadatos m_x son solamente definidas en algunos puntos discretos en el tiempo t_x , lo que, a su vez, produce el conjunto asociado de coeficientes de matrices c_x . Entre estos tiempos discretos t_x , los conjuntos de coeficientes matriciales deben interpolarse sobre la base de instancias de metadatos pasadas o futuras. Sin embargo, según se describió con anterioridad, los esquemas de interpolación de metadatos actuales sufren de la pérdida de calidad de audio espacial debido a inexactitudes inevitables en los procesos de interpolación de metadatos. Esquemas de interpolación alternativos, de conformidad con formas de realización a modo de ejemplo se describirán a continuación haciendo referencia a las Figura 7 a 11.

En las formas de realización a modo de ejemplo descritas con referencia a las Figuras 1 a 5, los metadatos 122, 222 asociados con los N objetos de audio 120, 220 y los metadatos 522 asociados con los K objetos 522 se originan, al menos en algunas formas de realización a modo de ejemplo, a partir de componentes de agrupamiento 409 y 509 y pueden referirse como metadatos de agrupamiento. Además, los metadatos 125, 325 asociados con la señal de mezcla reducida 124, 324 pueden referirse como metadatos de mezcla reducida.

Según se describe con referencia a las Figuras 1, 4 y 5, el componente de mezcla reducida 102 puede calcular las M señales de mezcla reducida 124 mediante la formación de combinaciones de los N objetos de audio 120 en una manera adaptativa de las señales, es decir, en conformidad con un criterio que es independiente de cualquier configuración de altavoces. Dicha operación del componente de mezcla reducida 102 es característica de formas de realización a modo de ejemplo dentro de un primer aspecto. En conformidad con formas de realización ejemplo dentro de otros aspectos de la idea inventiva, el componente de mezcla reducida 102 puede, a modo de ejemplo, calcular las M señales de mezcla reducida 124 mediante la formación de combinaciones de los N objetos de audio 120 en una manera adaptativa de señales o, como alternativa, tal como las M señales de mezcla reducida son adecuadas para la reproducción en los canales de una configuración de altavoces con M canales, esto es, como una mezcla reducida compatible con versiones anteriores.

En una forma de realización ejemplo, el codificador 400 descrito con referencia a la Figura 4 utiliza un formato de metadatos e información lateral particularmente adecuado para un remuestreo, esto es, para generar instancias de información lateral y metadatos adicionales. En la presente forma de realización ejemplo, el componente de análisis 106 calcula la información lateral 128 en una forma que incluye una pluralidad de instancias de información lateral que especifican establecimientos de reconstrucción deseados respectivos para reconstruir los N objetos de audio 120 y, para cada instancia de información lateral, datos de transición que incluyen dos partes independientemente asignables que, en combinación, definen un punto en el tiempo en para iniciar una transición desde un establecimiento de reconstrucción actual al establecimiento de reconstrucción deseado especificado por la instancia de información lateral, y un punto en el tiempo para completar la transmisión. En la forma de realización ejemplo actual, las dos partes independientemente asignables de los datos de transición para cada instancia de información lateral son: una marca temporal que indica el punto en el tiempo para iniciar la transición al establecimiento de reconstrucción deseado y un parámetro de duración de interpolación que indica una duración para alcanzar el establecimiento de reconstrucción deseado desde el punto en el tiempo para iniciar la transición al establecimiento de reconstrucción deseado. El intervalo durante el cual ha de tener lugar una transición está, en la presente forma de realización ejemplo, definido, de forma única, por el tiempo en el que ha de iniciarse la transición y la duración del intervalo de transición. Esta forma particular de la información lateral 128 se describirá a continuación haciendo referencia a las Figuras 7 a 11. Ha de entenderse que existen varias otras maneras para definir, de forma única, este intervalo de transición. A modo de ejemplo, un punto de referencia en la forma de un punto de inicio, final o intermedio del intervalo, acompañado por la duración del intervalo, puede utilizarse en los datos de transición para definir, de forma única, el intervalo. Como alternativa, los puntos inicial y final del intervalo pueden utilizarse en los datos de transición para definir, de forma única, el intervalo.

En la presente forma de realización a modo de ejemplo, el componente de agrupamiento 409 reduce la primera pluralidad de objetos de audio 421 a una segunda pluralidad de objetos de audio. En este caso, correspondientes a los N objetos de audio 120 de la Figura 1. El componente de agrupamiento 409 calcula los metadatos de agrupamiento 122 para los N objetos de audio generados 120 que permite la presentación de los N objetos de audio 122 en un dispositivo de presentación 210 en un lado del decodificador. El componente de agrupamiento 409 proporciona los metadatos de agrupamiento 122 en una forma que incluye una pluralidad de instancias de metadatos de agrupamiento que especifican los respectivos establecimientos de presentación deseados para presentar los N objetos de audio 120, para cada instancia de metadatos de agrupamiento, datos de transición que incluyen dos partes independientemente asignables que, en combinación, definen un punto en el tiempo para iniciar una transición desde un establecimiento de presentación actual al establecimiento de presentación deseado que se especifica por la instancia de metadatos de agrupamiento, y un punto en el tiempo para completar la transición para el establecimiento de presentación deseado. En la presente forma de realización ejemplo, las dos partes independientemente asignables de los datos de transición para cada instancia de metadatos de agrupamiento son: una marca temporal que indica el punto en el tiempo para iniciar la transición al establecimiento de presentación deseado y un parámetro de duración de interpolación que indica una duración para alcanzar el establecimiento de presentación deseado desde el punto en el tiempo para iniciar la transición al establecimiento de presentación deseado. Esta forma particular de los metadatos de agrupamiento 122 se describirá a continuación haciendo referencia a las Figuras 7 a 11.

En la presente forma de realización ejemplo, el componente de mezcla reducida 102 asocia cada señal de mezcla reducida 124 con una posición espacial e incluye la posición espacial en los metadatos de mezcla reducida 125 que permite la presentación de las M señales de mezcla reducida en un dispositivo de presentación 310 en un lado del decodificador. El componente de mezcla reducida 102 proporciona los metadatos de mezcla reducida 125 en una forma que incluye una pluralidad de instancias de metadatos de mezcla reducida que especifican los respectivos establecimientos de presentación de mezcla reducida deseados para presentar las señales de mezcla reducida y, para cada instancia de metadatos de mezcla reducida, datos de transición que incluyen dos partes independientemente asignables que, en combinación, definen un punto en el tiempo para iniciar una transición desde un establecimiento de presentación de mezcla reducida actual al establecimiento de presentación de mezcla reducida deseado que se especifica por la instancia de metadatos de mezcla reducida y un punto en el tiempo para completar la transición al establecimiento de presentación de mezcla reducida deseado. En la presente forma de realización ejemplo, las dos partes independientemente asignables de los datos de transición para cada instancia de metadatos de mezcla reducida son: una marca temporal que indica el punto en el tiempo para iniciar la transición al establecimiento de presentación de mezcla reducida deseado y un parámetro de duración de interpolación que indica una duración para alcanzar el establecimiento de presentación de mezcla reducida deseado desde el punto en el tiempo para iniciar la transición al establecimiento de presentación de mezcla reducida deseado.

En la presente forma de realización ejemplo, se utiliza el mismo formato para la información lateral 128, los metadatos de agrupamiento 122 y los metadatos de mezcla reducida 125. Este formato se describirá ahora con referencia a las Figuras 7 a 11 en términos de metadatos para presentación de señales de audio. Sin embargo, ha de entenderse que en los siguientes ejemplos descritos con referencia a las Figuras 7 a 11, términos o expresiones similares a “metadatos para presentación de señales de audio” pueden solamente sustituirse por los términos o expresiones correspondientes tales como “información lateral para reconstrucción de objetos de audio”, “metadatos de agrupamiento para presentación de objetos de audio” o “metadatos de mezcla reducida para presentación de señales de mezcla reducida”.

La Figura 7 ilustra la derivación, basada en metadatos, de curvas de coeficientes utilizadas en la presentación de señales de audio, de conformidad con una forma de realización ejemplo. Según se ilustra en la Figura 7, un conjunto de instancias de metadatos m_x generadas en diferentes puntos en el tiempo t_x p.ej., asociadas con marcas temporales únicas, se convierte por un convertidor 710 en conjuntos correspondientes de valores de coeficientes matriciales c_x . Estos conjuntos de coeficientes representan valores de ganancias, también referidos como factores de ganancias a utilizarse para la presentación de las señales de audio para diversos altavoces y controladores en un sistema de reproducción para el que ha de presentarse el contenido de audio. Un interpolador 720 interpola luego los factores de ganancia c_x para producir una curva de coeficientes entre los tiempos discretos t_x . En una forma de realización, las marcas temporales t_x asociadas con cada instancia de metadatos m_x puede corresponder a puntos aleatorios en el tiempo, puntos síncronos en el tiempo generados por un circuito de reloj, eventos temporales relacionados con el contenido de audio, tales como límites de tramas o cualquier evento temporizado adecuado. Conviene señalar que, según se describió con anterioridad, la descripción proporcionada con referencia a la Figura 7 se aplica, de forma análoga, a la información lateral para la reconstrucción de objetos de audio.

La Figura 8 ilustra un formato de metadatos en conformidad con una forma de realización (y según se describió con anterioridad, la siguiente descripción se aplica, de forma análoga, a un formato de información lateral correspondiente), que resuelve al menos algunos de los problemas de interpolación asociados con los métodos actuales, según se describió con anterioridad, definiendo una marca temporal como el tiempo de inicio de una transición o una interpolación, y aumentando cada instancia de metadatos con un parámetro de duración de interpolación que representa la duración de transición o la duración de interpolación (también referido como "tamaño de rampa"). Según se ilustra en la Figura 8, un conjunto de instancias de metadatos m2 a m4 (810) especifica un conjunto de matrices de presentación c2 a c4 (830). Cada instancia de metadatos se genera en un punto en el tiempo particular t_x y cada instancia de metadatos se define con respecto a su marca temporal m2 a t2, m3 a t3, y así sucesivamente. Las matrices de presentación asociadas 830 se generan después de realizar transiciones durante las respectivas duraciones de interpolación d2, d3, d4 (830), a partir de la marca temporal asociada (t1 a t4) de cada instancia de metadatos 810. Un parámetro de duración que indica la duración de interpolación (o tamaño de rampa) se incluye con cada instancia de metadatos, esto es, la instancia de metadatos m2 incluye d2, m3 incluye d3 y así sucesivamente. De forma esquemática, lo que antecede puede representarse como sigue: $m_x = (\text{metadatos } (t_x), d_x) \rightarrow c_x$. De esta manera, los metadatos proporcionan esencialmente una forma esquemática de cómo proceder a partir de un establecimiento de presentación actual (p.ej., la matriz de presentación actual resultante de metadatos anteriores) a un nuevo establecimiento de presentación (p.ej., la nueva matriz de presentación resultante de los metadatos actuales). Cada instancia de metadatos se considera que tiene efecto en un punto en el tiempo especificado en el futuro relativo al menos en que la instancia de metadatos fue recibida y la curva de coeficientes se deriva a partir del estado anterior del coeficiente. De este modo, en la Figura 8, m2 genera c2 después de una duración d2, m3 genera c3 después de una duración d3 y m4 genera c4 después de una duración d4. En este esquema para interpolación, los metadatos anteriores no necesitan ser conocidos, solamente se requiere la matriz de presentación anterior o el estado de presentación. La interpolación utilizada puede ser lineal o no lineal dependiendo de las restricciones y configuraciones del sistema.

El formato de metadatos de la Figura 8 permite un remuestreo sin pérdidas de metadatos, según se ilustra en la Figura 9. La Figura 9 ilustra un primer ejemplo de un procesamiento sin pérdidas de metadatos, de conformidad con una forma de realización a modo de ejemplo (y según se describió con anterioridad, la siguiente descripción se aplica de forma análoga a un formato de información lateral correspondiente). La Figura 9 ilustra instancias de metadatos m2 a m4 que se refieren a las matrices de presentación futuras c2 a c4, respectivamente, incluyendo las duraciones de interpolación d2 a d4. Las marcas temporales de las instancias de metadatos m2 a m4 se proporcionan como t2 a t4. En el ejemplo ilustrado en la Figura 9, una instancia de metadatos m4a, en el tiempo t4a, se añade a este respecto. Dichos metadatos pueden añadirse por varios motivos, tales como para mejorar la resistencia a errores del sistema o para sincronizar instancias de metadatos con el inicio/final de una trama de audio. A modo de ejemplo, el tiempo t4a puede representar el tiempo en que un dispositivo *códec* de audio utilizado para codificar un contenido de audio asociado con los metadatos inicia una nueva trama. Para funcionamiento sin pérdidas, los valores de metadatos de m4a son idénticos a los de m4 (esto es, ambos describen una matriz de presentación objetivo c4), pero el tiempo d4a para alcanzar ese punto ha sido reducido por d4-d4a. Dicho de otro modo, la instancia de metadatos m4a es idéntica a la instancia de metadatos anterior m4 de modo que la curva de interpolación entre c3 y c4 no es cambiada. Sin embargo, la nueva duración de interpolación d4a es más corta que la duración original d4. Esta circunstancia aumenta efectivamente la tasa de datos de las instancias de metadatos, lo que puede ser ventajoso en algunas circunstancias, tales como corrección de errores.

Un segundo ejemplo de interpolación de metadatos sin pérdidas se ilustra en la Figura 10 (y según se describió con anterioridad, la siguiente descripción se aplica, de forma análoga, a un formato de información lateral correspondiente). En este ejemplo, el objetivo es incluir un nuevo conjunto de metadatos m3a entre dos instancias de metadatos m3 y m4. La Figura 10 ilustra un caso en donde la matriz de presentación permanece invariable durante un período de tiempo. Por lo tanto, en esta situación, los valores del nuevo conjunto de metadatos m3a son idénticos a los de los metadatos anteriores m3, excepto para la duración de interpolación d3a. El valor de la duración de interpolación d3a debe establecerse al valor correspondiente de t4-t3a, esto es, a la diferencia entre el tiempo t4 asociado con la siguiente instancia de metadatos m4 y el tiempo t3a asociado con el nuevo conjunto de metadatos

m3a. El caso ilustrado en la Figura 10 puede presentarse, por ejemplo, cuando un objeto de audio es estático y una herramienta de autoría creadora interrumpe el envío de nuevos metadatos para objeto debido a esta naturaleza estática. En tal caso, puede ser deseable insertar nuevas instancias de metadatos m3a p.ej., para sincronizar los metadatos con tramas de *códec*.

5 En los ejemplos ilustrados en las Figuras 8 a 10, la interpolación desde una matriz de presentación actual a una matriz de presentación deseada o un estado de presentación fue realizada mediante interpolación lineal. En otras formas de realización a modo de ejemplo, se pueden utilizar también diferentes sistemas de interpolación. Uno de dichos esquemas de interpolación alternativos utiliza un circuito de muestreo y retención combinado con un filtro de paso bajo posterior. La Figura 11 ilustra un esquema de interpolación utilizando un circuito de muestreo y retención con un filtro de paso bajo, de conformidad con una forma de realización a modo de ejemplo (y según se describió con anterioridad, la siguiente descripción se aplica, de forma análoga, a un formato de información lateral correspondiente). Según se ilustra en la Figura 11, las instancias de metadatos m2 a m4 se convierten a coeficientes de matriz de presentación de muestreo y retención c2 y c3. El proceso de muestreo y retención causa que los estados de los coeficientes pasen inmediatamente al estado deseado, lo que da lugar a una curva gradual 1110, según se ilustra. Esta curva 1110 es luego filtrada con un filtro de paso bajo posteriormente para obtener una curva interpolada suave 1120. Los parámetros del filtro de interpolación (p.ej., frecuencia de corte o constante de tiempo) pueden señalizarse como parte de los metadatos, además de las marcas temporales y los parámetros de duración de interpolación. Ha de entenderse que diferentes parámetros pueden utilizarse dependiendo de los requisitos del sistema y de las características de la señal de audio.

25 En una forma de realización a modo de ejemplo, la duración de interpolación o tamaño de rampa pueden tener cualquier valor práctico, incluyendo un valor de prácticamente próximo a cero. Dicha pequeña duración de interpolación es especialmente de utilidad para casos tales como inicialización con el fin de permitir el establecimiento de la matriz de presentación inmediatamente en la primera muestra de un fichero, o permitir ediciones, empalmes o concatenación de flujos. Con este tipo de ediciones destructivas, que tienen la posibilidad de cambiar instantáneamente la matriz de presentación puede ser beneficioso mantener las propiedades espaciales del contenido después de la edición.

30 En una forma de realización a modo de ejemplo, el esquema de interpolación aquí descrito es compatible con la eliminación de instancias de metadatos (y, de forma análoga, con la eliminación de instancias información lateral, según se describió con anterioridad), tal como en un esquema de declive operativo que reduce las tasas binarias de metadatos. La eliminación de instancias de metadatos permite al sistema un remuestreo a una tasa de tramas que es inferior a una tasa de tramas inicial. En este caso, las instancias de metadatos y sus datos de duración de interpolación asociados que se proporcionan por un codificador pueden eliminarse sobre la base algunas características. A modo de ejemplo, un componente de análisis en un codificador puede analizar la señal de audio para determinar si existe un período de inmovilismo significativo de la señal y en tal caso, eliminar algunas instancias de metadatos ya generadas para reducir los requisitos del ancho de banda para la transmisión de datos a un lado del decodificador. La eliminación de instancias de metadatos puede realizarse de forma alternativa o adicional, en un componente separado del codificador, tal como en un decodificador o en un transcodificador. Un transcodificador puede eliminar instancias de metadatos que hayan sido generadas o añadidas por el codificador, y pueden utilizarse en un convertidor de tasas de datos que remuestrea una señal de audio a partir de una primera tasa a una segunda tasa, en donde la segunda tasa puede ser, o no, un múltiplo entero de la primera tasa. Como alternativa al análisis de la señal de audio con el fin de determinar qué instancias de metadatos eliminar, el codificador, decodificador o transcodificador puede analizar los metadatos. A modo de ejemplo, haciendo referencia a la Figura 10, una diferencia puede calcularse entre un primer establecimiento de reconstrucción deseado c3 (o matriz de reconstrucción), que se especifica por una primera instancia de metadatos m3 y los establecimientos de reconstrucción deseados c3a y c4 (o matrices de reconstrucción) que se especifican por instancias de metadatos m3a y m4 que suceden directamente a la primera instancia de metadatos m3. La diferencia puede calcularse, a modo de ejemplo, utilizando una norma matricial a las respectivas matrices de presentación. Si la diferencia es inferior a un valor umbral predefinido, p.ej., correspondiente a una distorsión tolerada de las señales de audio reconstruidas, las instancias de metadatos m3a y m4 que suceden a la primera instancia de metadatos m2 pueden eliminarse. En el ejemplo ilustrado en la Figura 10, la instancia de metadatos m3a que sucede directamente a la primera instancia de metadatos m3 especifica los mismos establecimientos de presentación c3 = c3a que la primera instancia de metadatos m3 y, por lo tanto, será eliminada, mientras que el siguiente establecimiento de metadatos m4 especifica un establecimiento de presentación diferente c4 y puede, dependiendo del valor umbral utilizado, mantenerse como metadatos.

60 En el decodificador 200 descrito con referencia a la Figura 2, el componente de reconstrucción de objeto 206 puede emplear la interpolación como parte de la reconstrucción de los N objetos de audio 220 sobre la base de las M señales de mezcla reducida 224 y la información lateral 228. En analogía con el esquema de interpolación descrito con referencia a las Figuras 7 a 11, la reconstrucción de los N objetos de audio 220 puede incluir, a modo de ejemplo: realizar una reconstrucción de conformidad con un establecimiento de reconstrucción actual; iniciar, en un punto en el tiempo definido por los datos de transición para una instancia de información lateral, una transición desde el establecimiento de reconstrucción actual a un establecimiento de reconstrucción deseado especificado por la instancia de información lateral; y completar la transición para el establecimiento de reconstrucción deseado en un

punto en el tiempo definido por los datos de transición para la instancia de información lateral.

De modo similar, el dispositivo de presentación 210 puede utilizar la interpolación como parte de la presentación de los N objetos de audio reconstruidos 220 con el fin de generar la señal de salida multicanal 230 adecuada para la reproducción. En analogía con el esquema de interpolación descrito con referencia a las Figuras 7 a 11, la presentación puede incluir: realizar una presentación de configuración con un establecimiento de presentación actual; iniciar, en un punto en el tiempo definido por los datos de transición para una instancia de metadatos de agrupamiento, una transición desde el establecimiento de presentación actual a un establecimiento de presentación deseado especificado por la instancia de metadatos del agrupamiento y completar la transición al establecimiento de presentación deseado en un punto en el tiempo definido por los datos de transición para la instancia de metadatos del agrupamiento.

En algunas formas de realización a modo de ejemplo, la sección de reconstrucción de objetos 206 y el dispositivo de presentación 210 pueden ser unidades separadas y/o pueden corresponder a operaciones realizadas como procesos separados. En otras formas de realización a modo de ejemplo, la sección de reconstrucción de objetos 206 y el dispositivo de presentación 210 pueden materializarse como una unidad o proceso único en el que el restablecimiento y la reconstrucción se realizan como una operación combinada. En dicha forma de realización ejemplo, las matrices empleadas para la reconstrucción y presentación pueden combinarse en una matriz única que puede ser interpolada, en lugar de realizar una interpolación sobre una matriz de presentación y una matriz de reconstrucción, por separado.

En el decodificador de baja complejidad 300, descrito con referencia a la Figura 3, el dispositivo de presentación 310 puede realizar una interpolación como parte de la presentación de las M señales de mezcla reducida 324 a la salida multicanal 330. En analogía con el esquema de interpolación descrito con referencia a las Figuras 7 a 11, la presentación puede incluir: realizar una presentación de conformidad con un establecimiento de presentación de mezcla reducida actual; iniciar, en un punto en el tiempo definido por los datos de transición para una instancia de metadatos de mezcla reducida, una transición desde el establecimiento de presentación de mezcla reducida actual a un establecimiento de presentación de mezcla reducida deseado especificado por la instancia de metadatos de mezcla reducida; y completar la transición para el establecimiento de presentación de mezcla reducida deseado en un punto en el tiempo definido por los datos de transición para la instancia de metadatos de mezcla reducida. Según se describió con anterioridad, el dispositivo de presentación 310 puede estar incluido en el decodificador 300 o puede ser un dispositivo/unidad separada. En formas de realización a modo de ejemplo en donde el dispositivo de presentación 310 está separado del decodificador 300, el decodificador puede proporcionar, a la salida, los metadatos de mezcla reducida 325 y las M señales de mezcla reducida 324 para la presentación de las M señales de mezcla reducida en el dispositivo de presentación 310.

Equivalentes, extensiones, alternativas y disposiciones varias

Formas de realización adicionales de la presente invención se harán evidentes para un experto en esta técnica después de estudiar la descripción anterior. Aun cuando la presente descripción y los dibujos adjuntos dan a conocer formas de realización y ejemplos, la invención no está restringida a estos ejemplos específicos. Numerosas modificaciones y variaciones pueden realizarse sin desviarse por ello del alcance de la presente invención, que se define por las reivindicaciones adjuntas. Cualquier signo de referencia que aparece en las reivindicaciones no ha de entenderse como limitación de su alcance.

Además, las variaciones a las formas de realización dadas a conocer pueden entenderse y efectuarse por el experto en esta técnica en la práctica de la invención, a partir de un estudio de los dibujos, la descripción y las reivindicaciones adjuntas. En las reivindicaciones, la expresión “que comprende” no excluye otros elementos o etapas y el artículo indefinido “un” o “una” no excluye una pluralidad. El mero hecho de que algunas medidas se indiquen en realizaciones dependientes mutuamente distintas no indica que no se pueda utilizar una combinación de estas medidas de forma ventajosa.

Los sistemas y métodos aquí dados a conocer pueden realizarse como software, firmware, hardware o una de sus combinaciones. En una puesta en práctica de hardware, la división de tareas entre unidades funcionales referidas en la anterior descripción no corresponden necesariamente a la división en unidades físicas; al contrario, una componente física puede tener múltiples funcionalidades y una tarea puede realizarse por varios componentes físicos en cooperación. Algunos componentes o todos los componentes pueden ponerse en práctica como software ejecutado por un procesador de señal digital o microprocesador, o ponerse en práctica como hardware o como un circuito integrado específico de la aplicación. Dicho software puede distribuirse en soporte legible por ordenador que puede comprender soportes de memorización legibles por ordenador (o soportes no transitorios) y soportes de comunicaciones (o soportes transitorios). Como es bien conocido por un experto en esta técnica, el término de soporte de memorización legible por ordenador incluye soportes volátiles y no volátiles, extraíbles y no extraíbles, puestos en práctica en cualquier método o tecnología para almacenamiento de información tal como instrucciones legibles por ordenador, estructuras de datos, módulos de programas u otros datos. Los soportes de memorización legibles por ordenador incluyen, sin limitación, a memoria RAM, memoria ROM, memoria EEPROM, memoria instantánea u otra tecnología de memoria, CD-ROM, discos versátiles digitales (DVD) u otro tipo de almacenamiento

5 de disco óptico, casete magnética, cinta magnética, memoria de disco magnético u otros dispositivos de memorización magnética o cualquier otro soporte que pueda utilizarse para memorizar la información deseada y que pueda ser objeto de acceso por un ordenador. Además, es bien conocido por los expertos en esta técnica que los soportes de comunicaciones suelen materializar instrucciones legibles por ordenador, estructuras de datos, módulos de programas u otros datos en una señal de datos modulada tal como una onda portadora u otro mecanismo de transporte e incluyen cualquier soporte de entrega de información.

10 Todas las Figuras son de tipo esquemático y por lo general, solamente muestran partes que son necesarias para poder elucidar la idea inventiva, mientras que otras partes pueden omitirse o simplemente sugerirse. A no ser que se indique de otro modo, las referencias numéricas similares se refieren a partes similares en diferentes figuras.

15

REIVINDICACIONES

1. Un método para codificar objetos de audio en un flujo de datos (140), que comprende:
- 5 la recepción de N objetos de audio (120), en donde $N > 1$;
- el cálculo de M señales de mezcla reducida (124), en donde $M \leq N$, mediante la formación de combinaciones de los N objetos de audio en conformidad con un criterio que es independiente de cualquier configuración de altavoces de M canales para la reproducción de las M señales de mezcla reducida, en donde los N objetos de audio están asociados con metadatos que incluyen posiciones espaciales de los N objetos de audio y valores de importancia que indican la importancia de los N objetos de audio en relación entre sí, en donde el criterio para calcular las M señales de mezcla reducida está basado en la proximidad espacial de los N objetos de audio y en los valores de importancia de los N objetos;
- 10
- 15 calcular información lateral (128) que incluye parámetros que permiten la reconstrucción de un conjunto de objetos de audio formado sobre la base de los N objetos de audio a partir de las M señales de mezcla reducida; y
- incluir las M señales de mezcla reducida y la información lateral en un flujo de datos (140) para su transmisión a un decodificador.
- 20
2. El método según la reivindicación 1, en donde una de las M señales de mezcla reducida corresponde a uno solo de los N objetos de audio, siendo dicho uno solo de los N objetos de audio el objeto de audio de los N objetos de audio que reviste la mayor importancia en relación con los otros entre los N objetos de audio.
- 25
3. El método según cualquiera de las reivindicaciones 1 a 2, que comprende, además, asociar cada señal de mezcla reducida con una posición espacial e incluir las posiciones espaciales de las señales de mezcla reducida en el flujo de datos como metadatos para las señales de mezcla reducida.
- 30
4. El método según la reivindicación 3, en donde los N objetos de audio están asociados con metadatos que incluyen las posiciones espaciales de los N objetos de audio, y las posiciones espaciales asociadas con las señales de mezcla reducida se calculan sobre la base de las posiciones espaciales de los N objetos de audio de audio; y
- de modo opcional, en donde las posiciones espaciales de los N objetos de audio y las posiciones espaciales asociadas con las M señales de mezcla reducida varían en el tiempo.
- 35
5. El método según cualquiera de las reivindicaciones precedentes, en donde la información lateral varía en el tiempo.
- 40
6. El método según cualquiera de las reivindicaciones precedentes, en donde la etapa de cálculo de las M señales de mezcla reducida comprende un primer procedimiento de reagrupamiento que incluye la asociación de los N objetos de audio con M agrupamientos basados en la proximidad espacial y en los valores de importancia de los N objetos de audio, y calcular una señal de mezcla reducida para cada agrupamiento mediante la formación de una combinación de objetos de audio asociados con el agrupamiento.
- 45
7. El método según la reivindicación 6 en donde cada señal de mezcla reducida está asociada con una posición espacial que se calcula sobre la base de las posiciones espaciales de los objetos de audio asociados con el reagrupamiento correspondiente a la señal de mezcla reducida, y
- de modo opcional, en donde la posición espacial asociada con cada señal de mezcla reducida se calcula como un centroide o un centroide ponderado de las posiciones espaciales de los objetos de audio asociados con el reagrupamiento correspondiente a la señal de mezcla reducida.
- 50
8. Un método para decodificar un flujo de datos que incluye objetos de audio codificado, que comprende:
- 55 la recepción de un flujo de datos que comprende M señales de mezcla reducida que son combinaciones de los N objetos de audio calculados en conformidad con un criterio que es independiente de cualquier configuración de altavoces de canal M para la reproducción de las M señales de mezcla reducida, en donde $M \leq N$, estando el criterio para calcular las M señales de mezcla reducida basado en la proximidad espacial de los N objetos de audio y sobre los valores de importancia de los N objetos de audio que indican la importancia de los N objetos de audio en relación entre sí,
- 60
- la recepción de información lateral que incluye parámetros que permiten la reconstrucción de un conjunto de objetos de audio formados sobre la base de los N objetos de audio a partir de las M señales de mezcla reducida; y
- 65 la reconstrucción del conjunto de objetos de audio formado sobre la base de los N objetos de audio a partir de las M señales de mezcla reducida y de la información lateral.

- 5 **9.** El método según la reivindicación 8, en donde una de las M señales de mezcla reducida corresponde a uno solo de los N objetos de audio, en donde dicho uno solo de los N objetos de audio es el objeto de audio de los N objetos de audio que es el más importante en relación con los otros de entre los N objetos de audio.
- 10 **10.** El método según la reivindicación 8 o la reivindicación 9, en donde el flujo de datos comprende, además, metadatos para las M señales de mezcla reducida que incluyen posiciones espaciales asociadas con las M señales de mezcla reducida, comprendiendo, dicho método, además:
- 15 en una condición en la que el decodificador está configurado para soportar la reconstrucción de objetos de audio, realizar la etapa de reconstrucción del conjunto de objetos de audio formados sobre la base de los N objetos de audio a partir de las M señales de mezcla reducida y la información lateral; y
- 20 en una condición en la que el decodificador no está configurado para soportar la reconstrucción de objetos de audio, utilizar los metadatos para las M señales de mezcla reducida para la presentación de las M señales de mezcla reducida hacia los canales de salida de un sistema de reproducción; y
- de modo opcional, en donde las posiciones espaciales asociadas con las M señales de mezcla reducida varían en el tiempo.
- 25 **11.** El método según cualquiera de las reivindicaciones 8 a 10, en donde la información lateral varía en el tiempo.
- 12.** El método según cualquiera de las reivindicaciones 8 a 11, en donde el flujo de datos comprende, además, metadatos para el conjunto de objetos de audio formado sobre la base de los N objetos de audio que incluyen las posiciones espaciales del conjunto de objetos de audio formados sobre la base de los N objetos de audio, comprendiendo el método, además:
- 30 utilizar los metadatos para el conjunto de objetos de audio formado sobre la base de los N objetos de audio para la presentación del conjunto reconstruido de objetos de audio formado sobre la base de los N objetos de audio a los canales de salida de un sistema de reproducción.
- 13.** El método según cualquiera de las reivindicaciones 8 a 12, en donde el conjunto de objetos de audio formado sobre la base de los N objetos de audio es igual a los N objetos de audio; o
- 35 en donde el conjunto de objetos de audio formado sobre la base de los N objetos de audio comprende una pluralidad de objetos de audio que son combinaciones de los N objetos de audio y cuyo número es inferior a N.
- 14.** Un producto de programa informático que comprende un soporte legible por ordenador con instrucciones para realizar el método según cualquiera de las reivindicaciones 1 a 7, o con instrucciones para realizar el método según cualquiera de las reivindicaciones 8 a 13.
- 40 **15.** Un decodificador (200, 300, 400, 500) para decodificar un flujo de datos que incluye objetos de audio codificados, comprendiendo el decodificador un componente de recepción y un componente de reconstrucción, estando el decodificador configurado para realizar el método según cualquiera de las reivindicaciones 8 a 13.
- 45

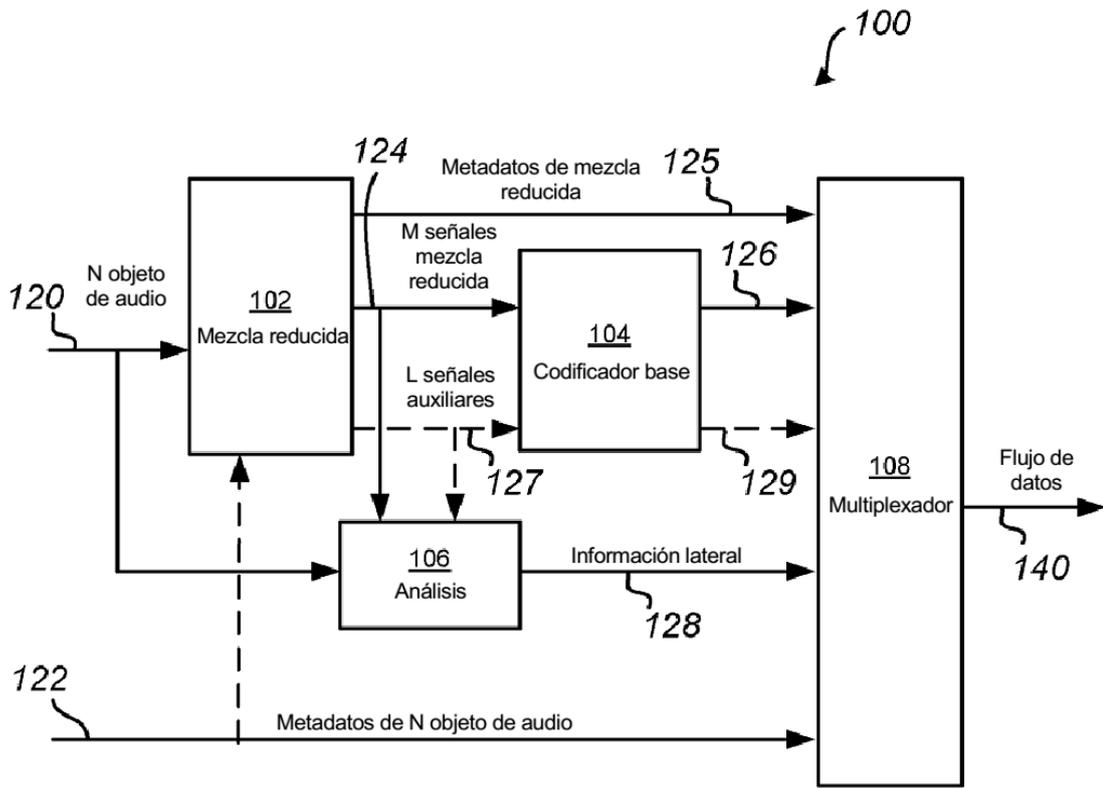


Fig. 1

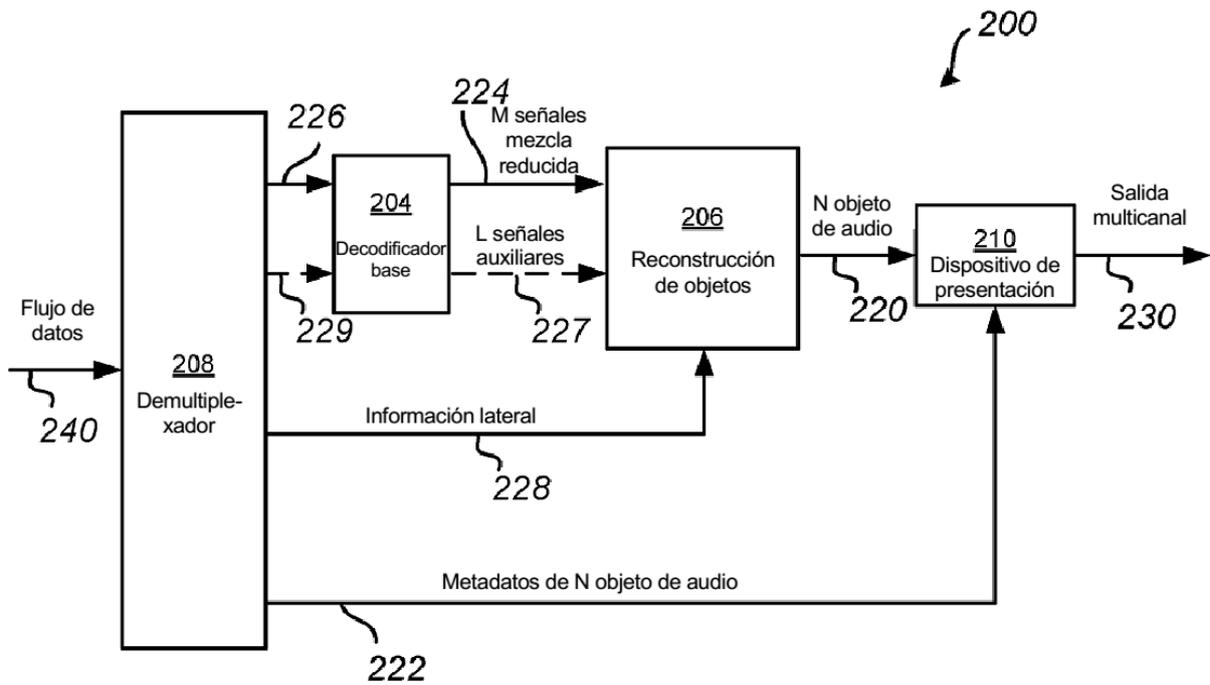


Fig. 2

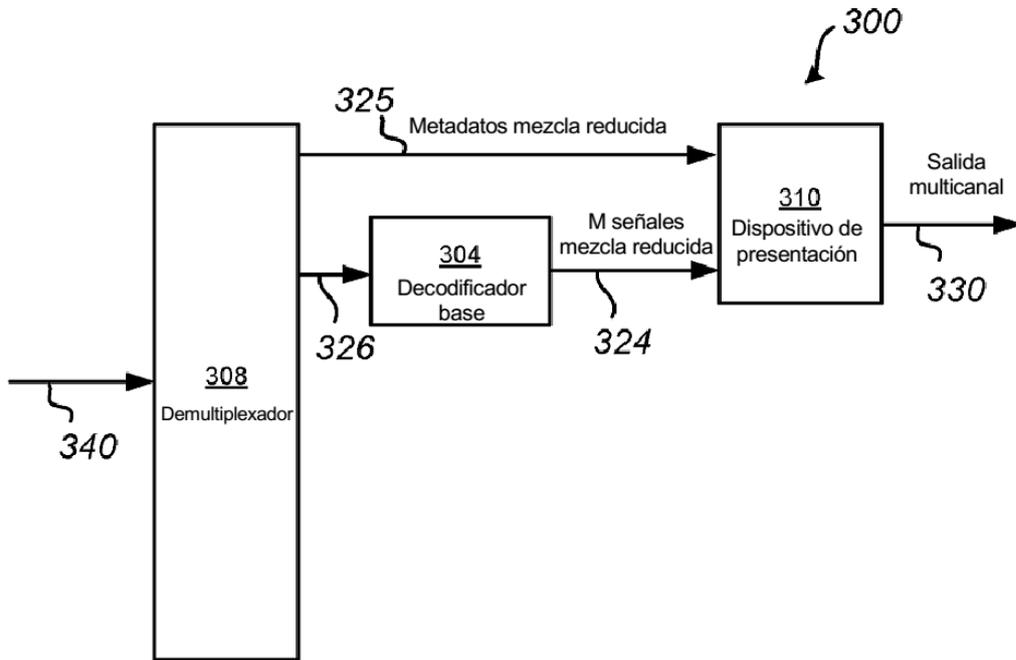


Fig. 3

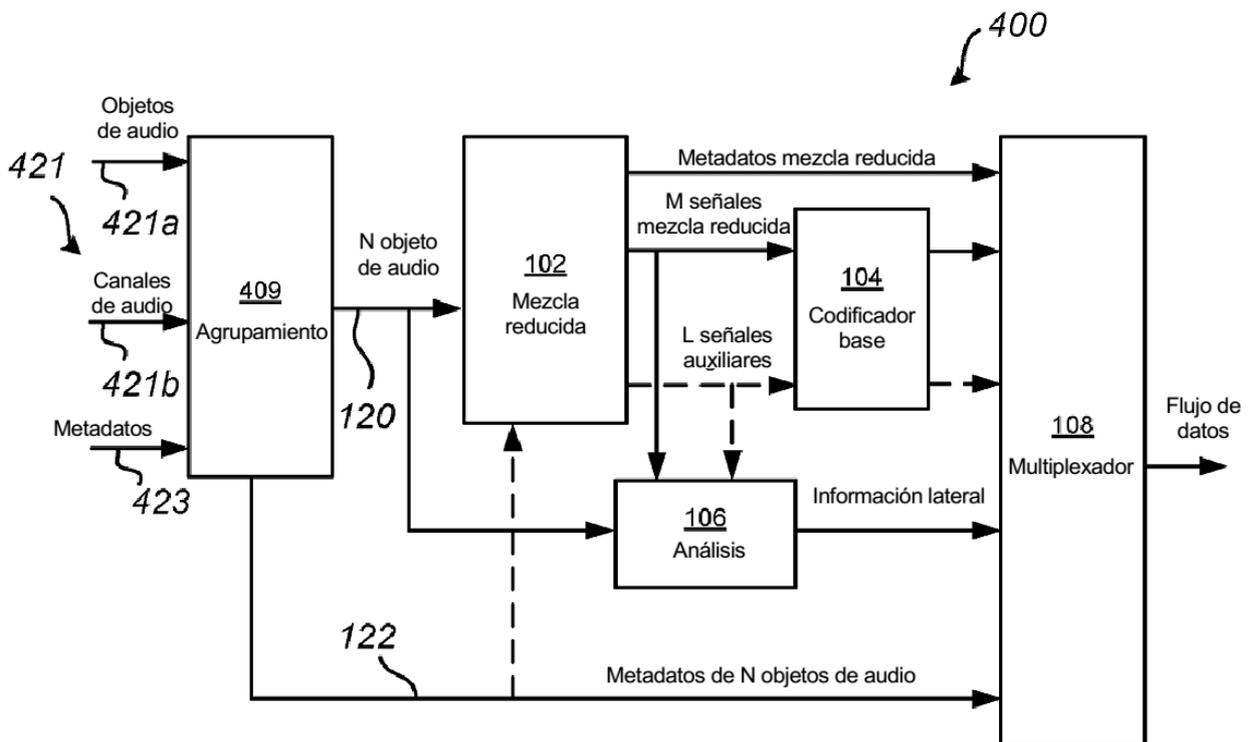


Fig. 4

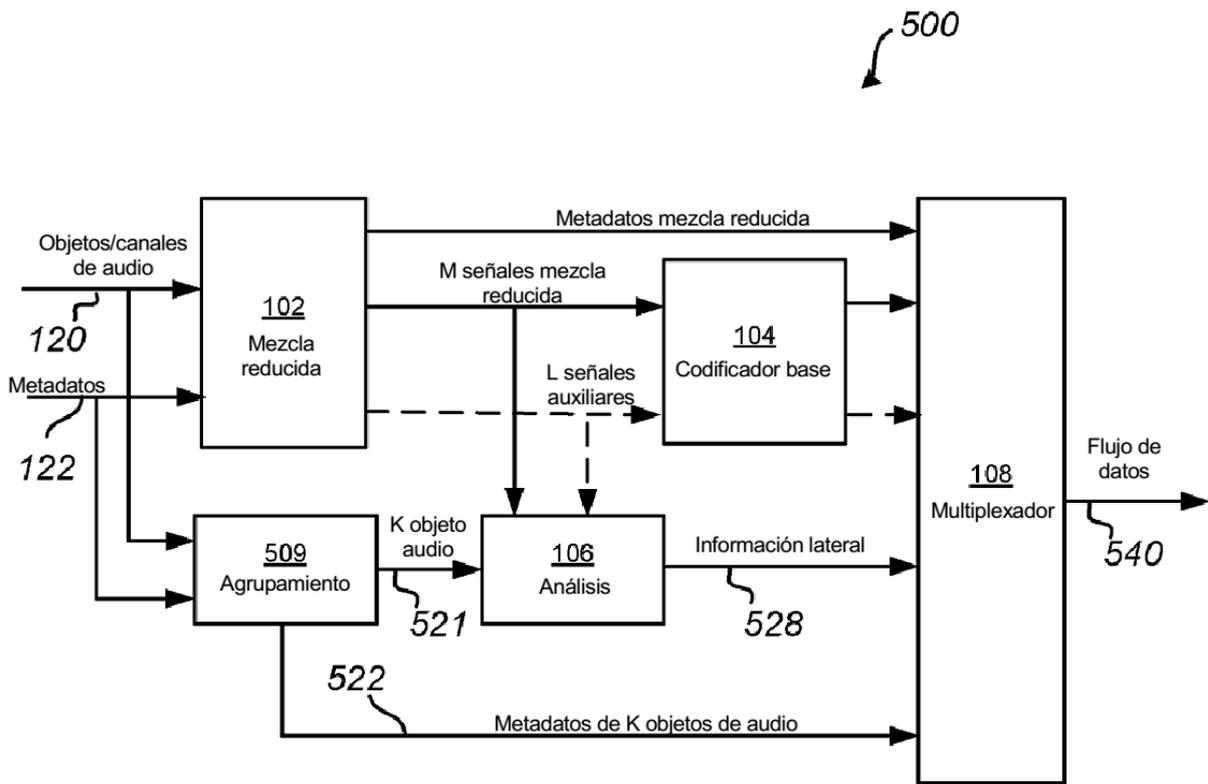


Fig. 5

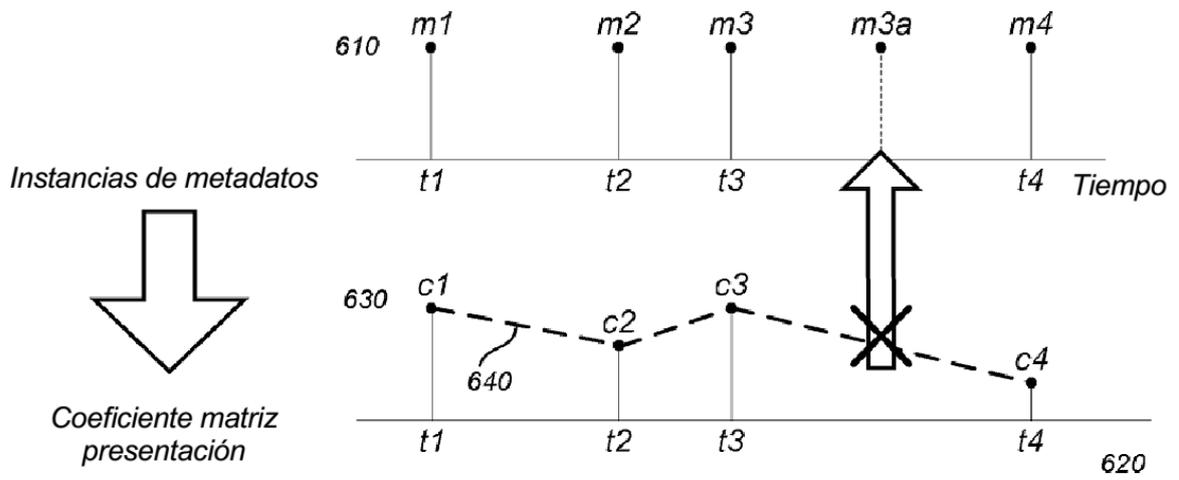


Fig. 6
(Técnica anterior)

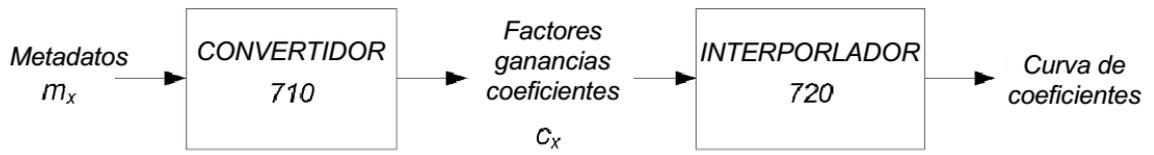


Fig. 7

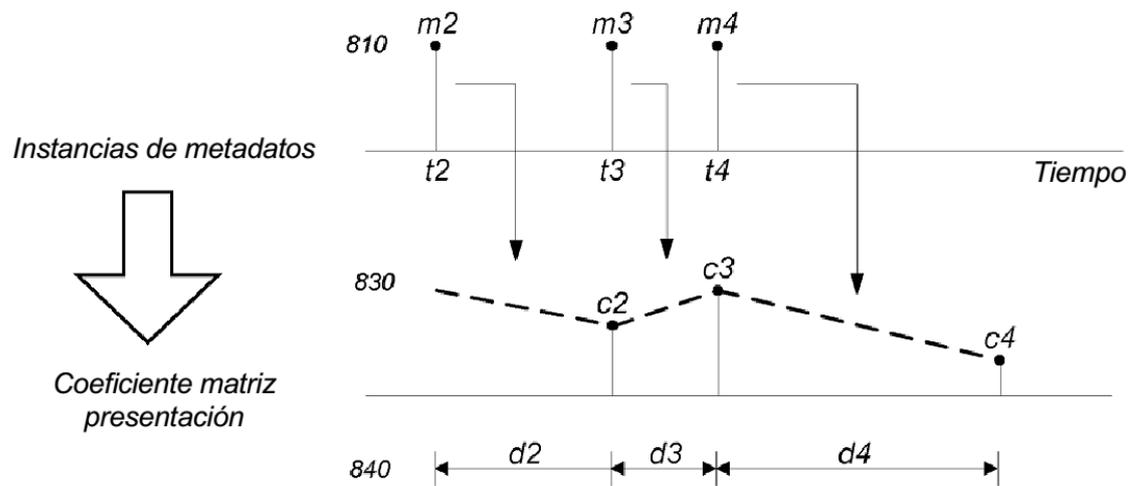


Fig. 8

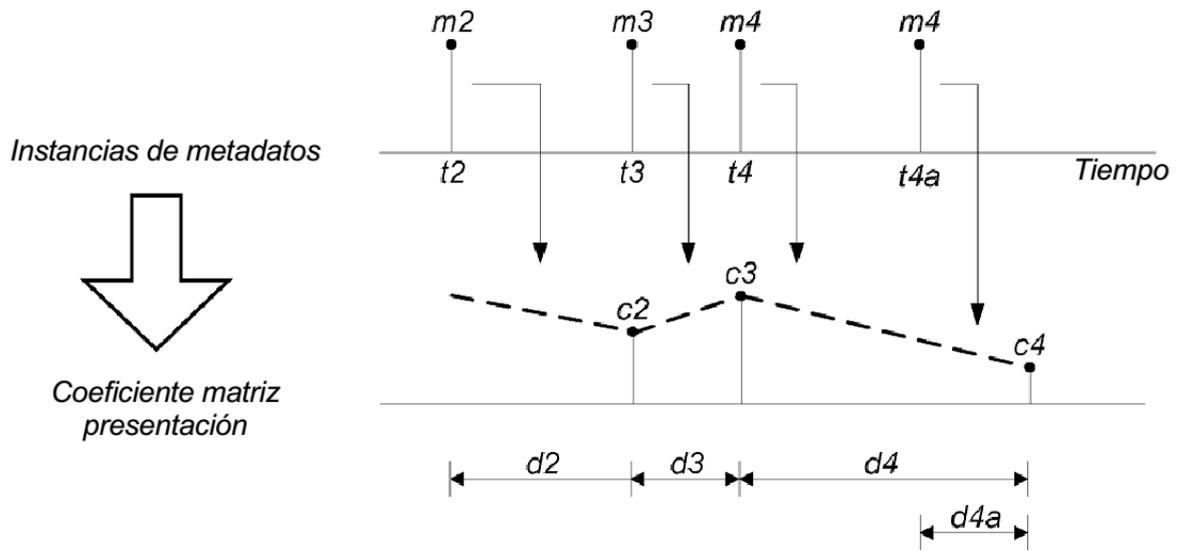


Fig. 9

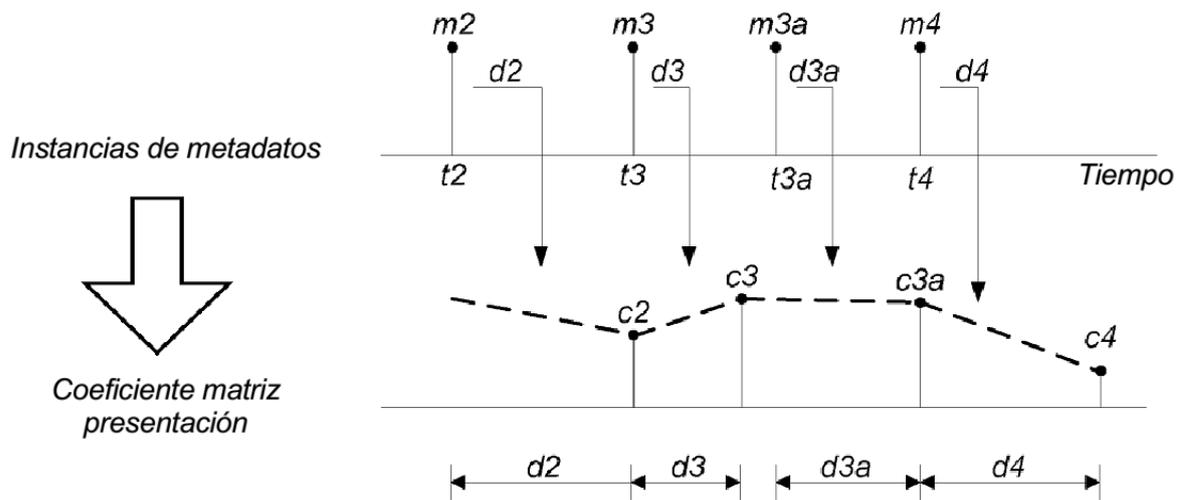


Fig. 10

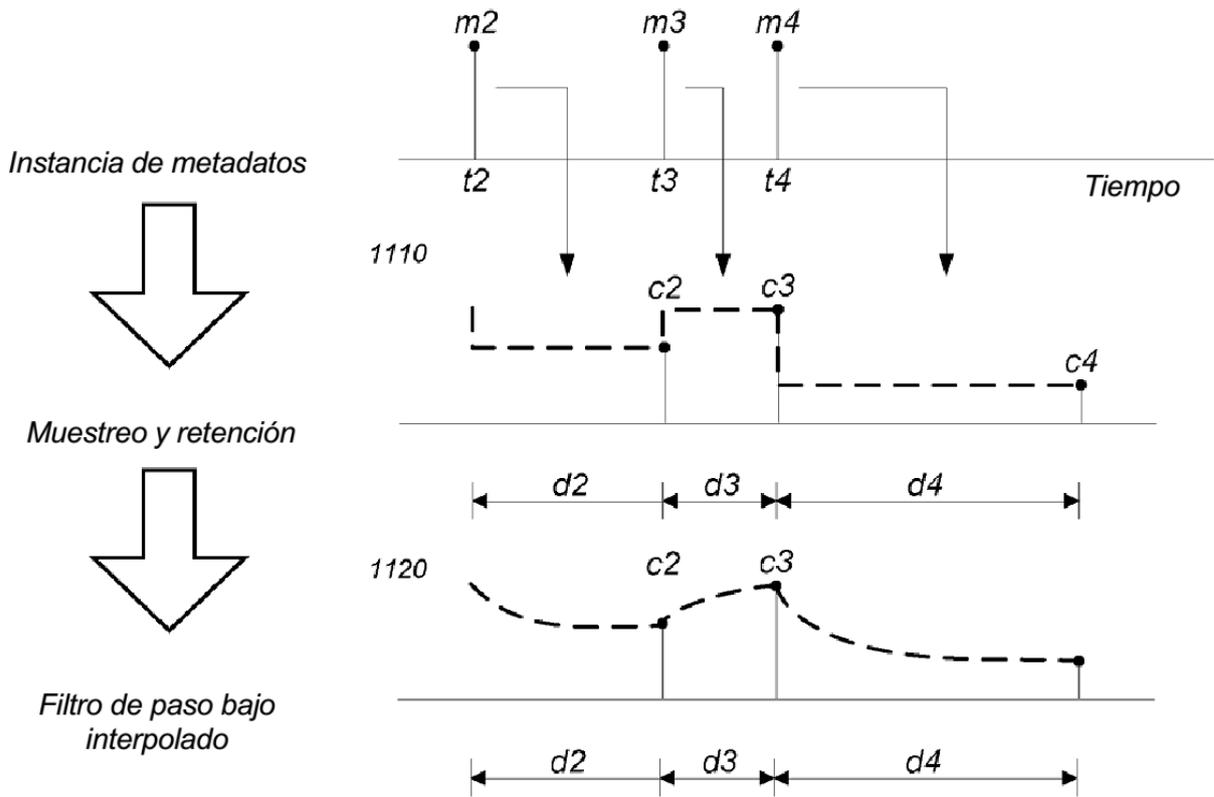


Fig. 11