

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 733 858**

51 Int. Cl.:

**H04N 21/845** (2011.01)  
**H04N 21/2343** (2011.01)  
**H04N 21/234** (2011.01)  
**H04N 21/233** (2011.01)  
**H04N 21/242** (2011.01)  
**G10L 25/57** (2013.01)  
**G10L 19/16** (2013.01)  
**G10L 21/055** (2013.01)

12

## TRADUCCIÓN DE PATENTE EUROPEA

T3

- 86 Fecha de presentación y número de la solicitud internacional: **08.03.2016 PCT/EP2016/054916**
- 87 Fecha y número de publicación internacional: **15.09.2016 WO16142380**
- 96 Fecha de presentación y número de la solicitud europea: **08.03.2016 E 16709348 (3)**
- 97 Fecha y número de publicación de la concesión europea: **24.04.2019 EP 3269147**

54 Título: **Codificación de audio alineada por fragmentos**

30 Prioridad:

**09.03.2015 EP 15158317**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**03.12.2019**

73 Titular/es:

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

72 Inventor/es:

**CZELHAN, BERND;  
FUCHS, HARALD;  
HOFMANN, INGO;  
THOMA, HERBERT y  
SCHREINER, STEPHAN**

74 Agente/Representante:

**SALVÀ FERRER, Joan**

ES 2 733 858 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 de audio alineada por fragmentos

- 5 **[0001]** La presente solicitud se refiere a un códec de audio adecuado, por ejemplo, para su uso en paralelo con vídeo codificado.
- [0002]** Cuando se entrega contenido de audio y vídeo a través de un canal de transmisión con una velocidad de bits fija o variable, una meta es asegurar la sincronización de audio y vídeo y la habilitación de casos de uso  
10 avanzados como el empalme.
- [0003]** La sincronización y alineación de audio y vídeo siempre ha sido una parte crucial cuando se construyen sistemas de audio y vídeo. Normalmente, los códecs de audio y vídeo no están usando la misma duración de trama. Debido a esta razón, los códecs de audio actuales no están alineados por trama. Como un  
15 ejemplo, esto también es cierto para la familia AAC ampliamente usada. El ejemplo se basa en el estándar de DVB, donde se utiliza un tamaño de trama de 1024 y una frecuencia de muestreo de 48 kHz. Esto conduce a tramas de
- $$\frac{1024 \text{ muestras}}{48000 \text{ Hz}} \approx 0.0213 \text{ seg}$$
- audio con una duración de . En contraste la velocidad de restauración de DVB común para vídeo es de 25 Hz o 50 Hz, lo cual conduce a duraciones de trama de vídeo de 0,02 segundos o 0,04 segundos respectivamente.  
20
- [0004]** Especialmente cuando se cambia la configuración del flujo de audio o se cambia el programa, es necesario que el vídeo y el audio sean alineados de nuevo. Los sistemas actuales cambiarán la configuración de audio ligeramente antes o después del vídeo correspondiente debido a que los seres humanos no son capaces de reconocer pequeñas diferencias en la sincronización de audio y vídeo.  
25
- [0005]** Desafortunadamente esto aumenta la complejidad de empalme donde una publicidad nacional tiene que ser reemplazada por una local, puesto que el flujo de vídeo reemplazado tiene que comenzar también con este pequeño desplazamiento. Además, los nuevos estándares están solicitando una sincronización de vídeo y audio más exacta para mejorar la experiencia total del usuario.  
30
- [0006]** Por lo tanto, los códecs de audio recientes pueden tratar con una amplia gama de tamaños de trama posibles para igualar el tamaño de la trama de vídeo. El problema aquí es que esto, además de resolver el problema de alineación, tiene un gran impacto sobre el rendimiento y la eficiencia de codificación.
- 35 **[0007]** El flujo unidireccional en ambientes de transmisión impone problemas especiales.
- [0008]** Desarrollos recientes han mostrado que el flujo unidireccional "adaptable" es considerado como una capa de transporte incluso para la transmisión lineal. Para unificar todos los requisitos que son ligeramente diferentes para la aplicación superior y para la aplicación aérea, ha sido optimizado el flujo unidireccional adaptable.  
40 Aquí nos centraremos en una tecnología de un flujo unidireccional adaptable concreta pero todos los ejemplos dados también funcionarán para otras tecnologías basadas en archivos como MMT.
- [0009]** La figura 7 muestra una propuesta para el estándar ATSC 3.0 el cual está actualmente en desarrollo. En esta propuesta, se considera una versión optimizada de MPEG-DASH que va a ser utilizada en un canal de  
45 transmisión de velocidad fija. Puesto que el DASH fue diseñado para una velocidad variable, el canal de unidifusión, como el LTE, 3G o Internet de Banda Ancha, fueron necesarios algunos ajustes los cuales son abarcados por la propuesta. La diferencia principal con el caso de uso del DASH regular es que el receptor de un canal de transmisión no tiene canal de respaldo y recibe una unidifusión. Normalmente el cliente puede extraer la ubicación del segmento de inicialización después de recibir y analizar sintácticamente el MPD. Después de eso el cliente puede decodificar  
50 un segmento después de otro y puede buscar hasta una marca temporal dada. Como se muestra en la figura anterior, en un ambiente de transmisión esta estrategia no es posible del todo. En su lugar el MPD y los segmentos de inicialización se repiten en una base regular. El receptor puede ser sintonizado entonces en cuanto reciba el MPD y todos los segmentos de inicialización necesarios.
- 55 **[0010]** Esto implica un intercambio entre el tiempo de sintonización corto y la sobrecarga total. Para un transmisor regular parece ser factible una longitud de segmento de aproximadamente 1 segundo. Esto significa que entre dos MPD existe un segmento de audio y uno de vídeo (si el programa contiene únicamente audio y vídeo) ambos con una longitud de aproximadamente un segundo.
- 60 **[0011]** Para una alineación de audio y vídeo el aspecto mencionado anteriormente también es cierto cuando se usa DASH. Además, los segmentos de audio tienen que ser ligeramente más largos o más cortos para mantener la alineación de audio y vídeo. Esto se muestra en la figura 8.
- [0012]** Si se activa un cambio de configuración de audio o vídeo. Este cambio tiene que ocurrir en un límite

de segmento, puesto que no existe otra manera de transmitir un segmento de inicialización actualizado. Para eso, el vídeo y el audio son rellenados (ya sea con tramas negras o silencios) para llenar un segmento completo. Pero esto no resuelve el problema de desalineación de vídeo y audio. Para empalmar y programar cambios puede existir una pequeña disparidad de audio y vídeo dependiendo de la duración del arrastre del segmento actual.

5 **[0013]** La "Especificación de flujo de transporte adaptativo de especificaciones de "OpenCable" (TM)", 14 de febrero de 2014, XP055188214, se refiere a la transmisión de televisión y proporciona consideraciones en la sección 9.6 con respecto a la alineación del flujo de audio y vídeo. En esta sección, se tratan los límites de fragmentos de audio implícitos y explícitos.

10 **[0014]** En la "nota técnica: directrices de implementación de AAC para DASH", 111. Reunión MPEG; ISO/IEC JTC1/SC29/WG11, N15072, directrices de implementación para la entrega de la codificación de audio avanzada MPEG-4 mediante el uso de DASH como se describe.

15 **[0015]** En el documento US 2013/0041672 A1, un concepto para proporcionar información sobre la validez de los datos de audio de manera que solo los datos válidos se compriman y descompriman, evitando así el procesamiento de la señal de audio adicional fuera del codificador y el decodificador.

20 **[0016]** El objetivo de la presente invención es proporcionar un códec de audio el cual hace las tareas de, por ejemplo, sincronización y alineación de audio y vídeo más efectivas, más fáciles de implementar sobre la base de las técnicas de compresión de audio existentes.

**[0017]** El objetivo es alcanzado por la materia objeto de las reivindicaciones independientes pendientes.

25 **[0018]** Una idea básica que subyace a la presente solicitud es que la sincronización y alineación de audio y vídeo o alineación de audio con algún otro reloj externo puede volverse más efectiva o más fácil cuando una red de fragmentos y red de tramas son tratadas como valores independientes, pero cuando, no obstante, por cada segmento la red de trama es alineada con el inicio de fragmento respectivo. Una pérdida de efectividad de compresión se puede mantener baja cuando se selecciona apropiadamente el tamaño de fragmento. Por otro lado, 30 la alineación de la red de trama con respecto a los inicios de fragmento permite una forma fácil y sincronizada por fragmento de manejar los fragmentos en relación con, por ejemplo, el flujo unidireccional de audio y vídeo paralelo, el flujo unidireccional adaptable por velocidad de bits o similar.

**[0019]** Las implementaciones ventajosas son la materia objeto de las reivindicaciones dependientes. Las 35 realizaciones preferidas de la presente solicitud son descritas más adelante con respecto a las figuras, entre las cuales

la figura 1 muestra un diagrama esquemático de un fragmento temporal que contiene vídeo y audio donde los fragmentos de vídeo y audio son alineados por tiempo de conformidad con una realización de la presente solicitud;

40 la figura 2 muestra un diagrama semi-esquemático y de bloques de un codificador, el contenido de audio codificado por éste y el flujo de datos codificado generado por éste de conformidad con una realización;

45 la figura 3 muestra un diagrama semi-esquemático y de bloques de un decodificador adaptado al codificador de la figura 2 de conformidad con una realización;

50 la figura 4 muestra un diagrama esquemático de ventanas, porciones del dominio del tiempo involucradas en el proceso de codificación/decodificación de conformidad con una realización según la cual se utiliza la codificación/decodificación basada en la transformación usada para codificar/decodificar tramas, es decir mediante la aplicación de una transformación superpuesta;

la figura 5 muestra un diagrama esquemático que ilustra la generación de información de finalización inmediata de conformidad con una realización;

55 la figura 6 muestra un diagrama esquemático que ilustra el caso de un cambio de configuración en el contenido de audio de conformidad con una realización que muestra que, por ejemplo, la información de finalización inmediata puede estar ausente en el caso de un cambio de configuración al inicio de un fragmento temporal respectivo, o donde la información de finalización inmediata de ese fragmento temporal codifique cero muestras en su lugar;

60 la figura 7 muestra un segmento de DASH empaquetado entregado a través de ROUTE (ruta) de conformidad con [1] con fines de comparación; y

la figura 8 muestra dos fragmentos consecutivos que contienen audio y vídeo de conformidad con un concepto de fragmentación actual según el cual la fragmentación de audio involucra diferentes duraciones fragmentadas.

65 **[0020]** Antes de describir varias realizaciones de la presente solicitud, se describen primero las ventajas

proporcionadas, y los pensamientos subyacentes a esas realizaciones. En particular, imagine que un contenido de audio va a ser codificado para acompañar una trama de vídeo compuesta de una secuencia de tramas de vídeo. El problema es como se ha indicado anteriormente en la porción introductoria de la presente solicitud: los códecs de audio actuales operan sobre una base de muestra y trama, la cual es una fracción no entera o múltiplos enteros de la velocidad de trama de vídeo. En consecuencia, las realizaciones descritas en lo sucesivo usan etapas de codificación/decodificación que operan en unidades de tramas "usuales" para lo cual están optimizadas. Por otro lado, el contenido de audio es sometido al códec de audio que subyace a esas etapas de codificación/decodificación en unidades de fragmentos temporales las cuales pueden ser de una o más, preferiblemente de una a cinco, o de manera incluso más preferible de una a dos tramas de vídeo de longitud. Para cada uno de esos fragmentos temporales, la red de trama es elegida de modo que no esté alineada con el inicio del fragmento temporal respectivo. En otras palabras, la idea que subyace a las realizaciones descritas posteriormente es producir fragmentos de audio los cuales son exactamente tan largos como la trama de vídeo correspondiente, con esta estrategia que tiene dos beneficios:

15 1) El codificador de audio puede funcionar incluso sobre una duración de trama optimizada/nativa y no tiene que dejar su red de trama hasta los límites de fragmento.

2) Cualquier retraso de audio puede ser compensado mediante el uso de información de finalización inmediata para las representaciones codificadas de los fragmentos temporales. El empalme puede ocurrir en cada límite de fragmento. Esto reduce la complejidad total del equipo de transmisión significativamente.

**[0021]** La figura 1 muestra un ejemplo para un fragmento de audio generado en una forma de conformidad con un ejemplo expuesto más adelante, cuyo fragmento de audio acompaña a un fragmento de vídeo correspondiente. Tanto el fragmento de audio como el fragmento de vídeo son ilustrados de una manera correspondiente a la figura B. Es decir que, en 2, es decir la fila superior de la figura 1 ilustra el fragmento de vídeo como si estuviese compuesto de un número  $N$  de tramas 4, es decir tramas de vídeo, en el que las tramas son mostradas como cuadrados dispuestos secuencialmente a lo largo de la fila de izquierda a derecha a lo largo de su orden de finalización temporal como es ilustrado por el eje temporal  $t$ . El borde izquierdo de la trama 0 y el borde derecho de la trama 59 son mostrados como si estuviesen registrados con el inicio y el fin del fragmento, lo que significa que la longitud temporal  $T_{\text{fragmento}}$  del fragmento es un entero múltiple de la longitud de la trama de vídeo, siendo aquí el entero múltiple  $N$  ejemplarmente 60. Temporalmente alineado al fragmento de vídeo 2, la figura 2 ilustra más adelante un fragmento de audio 10 que tiene codificado en él el contenido de audio que acompaña al fragmento de vídeo 2 en unidades de trama o unidades de acceso 12, aquí ilustradas como rectángulos que se extienden horizontalmente, es decir temporalmente, a un tono temporal el cual ilustrará su longitud de trama temporal y esta longitud de trama de audio es, desafortunadamente, tal que la longitud temporal  $T_{\text{fragmento}}$  del fragmento de audio 10 no es un entero múltiple de esta longitud de trama  $T_{\text{trama}}$ . Por ejemplo, la relación entre la longitud de trama  $T_{\text{trama}}$  y la longitud de trama correspondiente de las tramas de vídeo  $T_{\text{trama vídeo}}$  puede ser tal que la relación entre ellas es irracional o la relación entre ellas puede ser representada por una fracción apropiada, completamente reducida, donde el numerador es por ejemplo, mil veces mayor que el denominador, de modo que una longitud de fragmento que sería un múltiplo de ambas de la longitud de trama de vídeo  $T_{\text{trama vídeo}}$  y la longitud de trama de audio  $T_{\text{trama}}$  sería desventajosamente alta.

**[0022]** La figura 1 ilustra que, en consecuencia, una trama posterior o última, es decir la unidad de acceso 46, abarca temporalmente una porción temporal del contenido de audio que excede el extremo final 14 del fragmento de audio 10. Posteriormente, se mostrará que la porción 16 que excede o sigue al extremo posterior 14 puede ser truncada o pasada por alto en el lado del decodificador en la finalización, o que toda la trama posterior en realidad no es codificada en el decodificador, simplemente nivelando sus estados internos para llenar el "hueco temporal" de la porción de la trama posterior para superponerse con el fragmento temporal 10.

**[0023]** Con fines de ilustración, la figura 1 ilustra en la mitad inferior de la misma, es decir en 16, que la cantidad de bits disponibles para el fragmento temporal compuesto de vídeo y audio, es decir  $T_{\text{fragmento}} \cdot R$  con  $R$  siendo una velocidad de bits podría ser usada para llevar los datos de vídeo 18 en los cuales las tramas de vídeo 4 o el fragmento 2 son codificados, los datos de audio 20 en los cuales el contenido de audio del fragmento de audio son codificados, los datos del encabezado 22 y 24 de ambos, respectivamente, los datos de configuración 26 que indican, por ejemplo, la resolución espacial, resolución temporal y así sucesivamente en la cual las tramas de vídeo 4 son codificadas en datos 18 y la configuración como el número de canales en los cuales las tramas de audio 12 o fragmento 2 son codificados en datos 20 así como un manifiesto de descripción de presentación de medios aquí incluida ilustrativamente en los datos para los fragmentos coalineados 2 y 1 o para indicar, por ejemplo, la versión en la cual el vídeo y audio están disponibles, difiriendo las versiones en la velocidad de bits. Deberá comprenderse que el ejemplo de la figura 1 es meramente ilustrativo y que las realizaciones descritas aquí posteriormente no se restringen a ser usadas en relación con el flujo unidireccional adaptable por velocidad de bits y el envío de un manifiesto al cliente y así sucesivamente. La figura 1 será meramente ilustrativa del concepto común de las realizaciones explicadas más adelante de conformidad con las cuales la fragmentación de audio es alineada completamente con la fragmentación de vídeo alineando las tramas de audio 12 con el inicio 30 de los fragmentos 10, los cuales, a su vez, son elegidos para ser, por ejemplo, completamente alineados con las tramas de vídeo 4.

**[0024]** La figura 1 muestra de este modo un fragmento de audio y vídeo, siendo ambos alineados en la forma descrita. En el ejemplo de la figura 1, los fragmentos de vídeo y audio fueron elegidos de modo que tuvieran una

longitud temporal  $T_{\text{fragmento}}$  constante de  $\frac{1001}{1000} \text{ seg} = 1.001 \text{ seg}$  la cual es equivalente a 60 tramas de vídeo a la velocidad de trama NTSC de 59,94 Hz.

5

**[0025]** La última trama de audio de cada fragmento de audio, aquí AU 46, es por ejemplo truncada para igualar la duración de fragmento. En el ejemplo dado, la última trama de audio alcanza la muestra 47104 a 48127 donde ha sido elegida una numeración basada en cero, es decir que la primera muestra de audio en el fragmento es numerada como cero. Esto conduce a un tamaño de fragmento de un número de muestras que es ligeramente más grande de lo necesario, es decir 48128 en lugar de 48048. Por lo tanto, la última trama es cortada después de la 944<sup>ésima</sup> muestra. Esto se puede lograr mediante el uso, por ejemplo, de una lista de edición contenida por ejemplo en los datos del encabezado 24 o en los datos de configuración 26. La parte truncada 16 puede ser codificada con menos cantidad, por ejemplo. De manera alternativa, existiría la posibilidad de no transmitir todas las tramas de audio 12, sino dejar, por ejemplo, la codificación de la última trama aquí, ejemplarmente AU 46, puesto que el

10

15

**[0026]** En las realizaciones descritas más adelante, se mostrará que pueden ser tomadas medidas para contrarrestar el problema de que el decodificador que opera, por ejemplo, en una función de ventanas de superposición perderá su historia y no podrá producir una señal completa para la primera trama del siguiente fragmento. Por esta razón, la primera trama, en la figura 1 ejemplarmente AU0, es codificada como una trama IPF que permite la finalización inmediata (IPF = Trama de Finalización Inmediata). Ésta se coloca justo en el lado respectivo y cualquier fragmento de audio respectivamente. De igual modo, la primera trama de vídeo 4 puede ser una trama IDR (IDR = Restauración de Decodificación Instantánea).

20

Sobrecarga (el peor de los casos)	Sin sbr (1 seg.)	Sbr:1 (1 seg.)	Sin sbr (2 seg.)	Sbr 2:1 (2 seg.)	Sin sbr (0,5 seg.)	Sbr 2:1 (0,5 seg.)
Tamaño de fragmento (seg.)	1,001	1,001	2,002	2,002	0,5005	0,5005
Tamaño de trama (muestras):	1.024	2.048	1.024	2.048	1.024	2.048
Velocidad de muestreo	48.000	48.000	48.000	48.000	48.000	48.000
Preenrollado (aus):	5	3	5	3	5	3
Aus/fragmento normal	46,921875	23,4609375	93,84375	46,921875	23,4609375	11,73046875
Aus/fragmento alineado:	52	27	99	50	29	15
Sobrecarga	10,80%	15,10%	5,50%	6,60%	23,60%	27,90%

25

Tabla 1: Sobrecarga de velocidad de bits

**[0027]** La tabla anterior da un ejemplo para la sobrecarga de velocidad de bits esperada si no fuera aplicada la optimización. Puede observarse que la sobrecarga depende fuertemente de la duración de fragmento  $T_{\text{fragmento}}$  usada. Dependiendo de los requisitos del transmisor, es factible alinear solo cada segundo o tercer fragmento respectivo, es decir eligiendo que los fragmentos de audio sean más largos.

30

**[0028]** La figura 2 muestra un codificador para codificar contenido de audio en unidades de fragmentos temporales 10 o una red de fragmentos 32 en un flujo de datos codificado 34. El codificador se indica de manera general utilizando la señal de referencia 20 y comprende una etapa de codificación 36 y un proveedor de fragmentos 38. La etapa de codificación 36 está configurada para codificar contenido de audio en unidades de tramas 12 de una red de tramas y el proveedor de fragmentos 238 está configurado para proporcionar el contenido de audio 31 a la etapa de codificación 36 en unidades de fragmentos temporales 10 de modo que cada fragmento temporal sea codificado por la etapa de codificación 36 en una representación codificada 38 del fragmento temporal respectivo 10, donde el proveedor de fragmentos 38 está configurado para proporcionar el contenido de audio 31 a la etapa de codificación 36 en unidades de los fragmentos temporales 10 de modo que cada fragmento temporal de la red de tramas respectivas de tramas 12 sea alineado temporalmente con el inicio 30 del fragmento temporal respectivo 10 de modo que el inicio 30 coincida con un límite de trama 42 de la red de tramas respectiva de tramas 12. Es decir que, como se describe mejor en lo sucesivo, el proveedor de fragmentos 38 puede proporcionar la etapa de

35

40

codificación 36, el fragmento temporal 10 acertado, con una porción 34 del contenido de audio 31 que incluye el fragmento temporal 10 actualmente proporcionado y opcionalmente una porción 46 del contenido de audio 31 que precede temporalmente al fragmento temporal actual 10, y una porción 48 que sigue temporalmente al fragmento temporal actual 10. Para proporcionar la etapa de codificación 36 con una porción actual 44, el fragmento temporal actual 10 es alineado temporalmente 50 por el proveedor de fragmentos 38 de modo que los límites de trama 42 comprendan un límite de trama que coincida con el inicio 30 del fragmento temporal actual 10. Como se describió anteriormente con respecto a la figura 1, debido al hecho de que la longitud temporal del fragmento temporal 10 es un múltiplo no entero de la longitud temporal de las tramas 12, una trama posterior 12a abarca solo parcialmente o se superpone temporalmente con un fragmento temporal 10 con una porción 16 de la misma que abarca la porción 10 siguiente 48 del contenido de audio.

**[0029]** Antes de describir con detalle la funcionalidad del codificador de la figura 2, se hace referencia a la figura 3, la cual muestra un decodificador correspondiente de conformidad con una realización. El decodificador de la figura 3 se indica generalmente usando el signo de referencia 60 y se utiliza para decodificar el contenido de audio 15 31 en unidades de fragmentos temporales 10 del flujo de datos codificado 34. El decodificador 60 comprende una interfaz de entrada 62 la cual recibe representaciones codificadas de los fragmentos temporales. Como se muestra en la figura 3 mediante el uso de la incubación como ya se ha explicado con respecto a la figura 2, por cada fragmento temporal 10, está presente una representación codificada 40 del mismo en el flujo de datos 34. Cada representación codificada 40 tiene codificada por lo tanto su fragmento temporal asociado 10 en unidades de las 20 tramas mencionadas anteriormente 12 alineadas temporalmente con el inicio 30 del fragmento temporal respectivo 10 de modo que el inicio 30 coincida con un límite de trama 42 de la red de tramas.

**[0030]** El decodificador 60 comprende además una etapa de decodificación 64 configurada para decodificar versiones reconstruidas 66 de los fragmentos temporales 10 de las representaciones codificadas 40. Es decir que, la 25 etapa de decodificación 64 produce, por cada fragmento temporal 40, una versión reconstruida 66 del contenido de audio como la abarcada por el fragmento temporal 10 a la cual pertenece la representación codificada 40 respectiva.

**[0031]** El decodificador 60 comprende además un ensamblador 68 configurado para unir, o ensamblar, las versiones reconstruidas 66 de los fragmentos temporales 10 junto con, de manera inherente, la alineación de los 30 inicios de las versiones reconstruidas 66 de los fragmentos temporales para que coincida con los límites de fragmento de la red de fragmentos, es decir, en los inicios 30 de la red de fragmentos, cuando la redes de tramas individuales de los fragmentos 10 se registren con estos.

**[0032]** De este modo, el codificador 20 y el decodificador 60 de las figuras 2 y 3 operan del siguiente modo. 35 El codificador 20 codifica cada fragmento temporal 10 en una representación codificada correspondiente 40 de modo que la red de tramas de tramas 12 sea alineada con el inicio 30 del fragmento temporal correspondiente 10 de modo que una primera trama delantera 12<sub>b</sub> comience inmediatamente en el inicio 30, es decir que los inicios del fragmento temporal 10 y la primera trama 12<sub>b</sub> coincidan. El problema de cómo la etapa de codificación 36 trata la trama final 12<sub>a</sub>, la cual simplemente se superpone parcialmente con el fragmento temporal 10 puede ser resuelto de manera 40 diferente, como se expone más adelante. Además, cuando la etapa de codificación 36 realinea su red de tramas por cada fragmento temporal 10, la etapa de codificación 36 codifica los fragmentos temporales 10 en su representación codificada correspondiente 40 en una forma completamente independiente, es decir independiente de los otros fragmentos temporales. No obstante, la etapa de codificación 36 codifica los fragmentos temporales 10 en sus representaciones codificadas correspondientes 40 de modo que la finalización inmediata sea permitida en el lado de 45 decodificación por cada fragmento temporal. Los detalles de la posible implementación son expuestos más adelante. A su vez, el decodificador 60 reconstruye a partir de cada representación codificada 40 una versión reconstruida 66 del fragmento temporal correspondiente 10. La versión reconstruida 66 puede ser tan larga como el fragmento temporal correspondiente 10. Hasta este punto, como se describe más adelante, la etapa de codificación 64 puede efectuar la nivelación o limpieza para extender la longitud temporal de la versión reconstruida 66 hasta la longitud 50 temporal de los fragmentos temporales 10, o la etapa de decodificación 64 y el ensamblador 66 pueden, como se trata más adelante, cooperar para truncar o desestimar porciones temporales de la versión reconstruida 66, la cual en otras circunstancias excedería la longitud temporal de los fragmentos temporales. La etapa de decodificación 64, al efectuar la decodificación de las representaciones codificadas 40, también usa la red de tramas, es decir que efectúa la decodificación en unidades de las tramas 12 y efectúa sustancialmente una inversa del proceso de 55 codificación.

**[0033]** En lo sucesivo, se trata la posibilidad según la cual la etapa de codificación 36 atiende también la codificación de la trama final 12a en la representación codificada correspondiente 40, y que el decodificador atiende 60 una truncación de las porciones sobresalientes correspondientes de la versión reconstruida 66. En particular, de conformidad con este ejemplo, la etapa de codificación 36 y el proveedor de fragmentos 38 pueden cooperar de modo que, para un fragmento temporal actual 10, la codificación de este fragmento temporal 10 en la representación codificada 40 continúe más allá del extremo posterior 70 del fragmento temporal actual 10 hasta donde la trama final 12<sub>a</sub> esté relacionada. Es decir, que la etapa de codificación 36 también codifica la porción que sobresale 16 del contenido de audio en la representación codificada 40. Al hacer esto, sin embargo, la etapa de codificación 36 puede 65 modificar la velocidad de bits invertida para codificar esta trama final 12<sub>a</sub> en la representación codificada 40 de la porción que sobresale 16 a la porción restante de la trama final 12<sub>a</sub>, es decir la porción que se superpone

temporalmente con el fragmento temporal actual 10. Por ejemplo, la etapa de codificación 36 puede ser menor a la calidad a la cual la porción que sobresale 16 es codificada en la representación codificada 40 en comparación con la calidad a la cual la otra porción de la trama final 12<sub>a</sub> es codificada en la representación codificada 40, es decir una perteneciente al fragmento temporal actual 10. En ese caso, la etapa de decodificación 64 decodificaría en consecuencia de esta representación codificada 40 una versión reconstruida 66 del fragmento temporal correspondiente 10 que excede temporalmente la longitud temporal del fragmento temporal 10, es decir hasta donde la porción que sobresale 16 de la trama final 12<sub>a</sub> esté involucrada. El ensamblador 68, al alinear la versión reconstruida 66 con la red de fragmentación, es decir con los inicios de fragmento 30, truncaría la versión reconstruida 66 en la porción que sobresale 16. Es decir, que el ensamblador 68 desestimaría esta porción 16 de la versión reconstruida 66 al finalizar. El hecho de que esta porción 16 pueda haber sido codificada a una calidad menor como se ha explicado anteriormente, es en consecuencia transparente al escuchar el contenido de audio reconstruido 31', el cual es el resultado del ensamble de las versiones reconstruidas 66 en el ensamblador de salida 68, puesto que esta porción es reemplazada, al final, por el inicio de la versión reconstruida del siguiente fragmento temporal 10.

15

**[0034]** De manera alternativa, el codificador 20 puede operar para dejar la trama final 12<sub>a</sub> en codificación en un fragmento temporal actual 10. En su lugar, el decodificador puede intentar llenar la porción no codificada del fragmento temporal 10, es decir una con la cual la trama final 12<sub>a</sub> se superpone parcialmente, pero nivelando su estado interno como se describe de manera ejemplar más adelante. Es decir, que la etapa de codificación 36 y el proveedor de fragmentos 38 pueden cooperar de modo que, para un fragmento temporal actual 10, la codificación de este fragmento temporal en su representación codificada 40 sea atada en la trama 12 que precede inmediatamente a la trama final 12<sub>a</sub>. La etapa de codificación puede señalar dentro de la representación codificada 40 una señalización nivelada que indique al decodificador que llene la porción restante, de este modo no codificada del fragmento temporal 10, es decir la que se superpone con la trama final 12<sub>a</sub>, por medio de la nivelación de estados internos del codificador como se manifiesta en sí hasta la trama 12 inmediatamente anterior a la trama final 12<sub>a</sub>. En el lado del decodificador, la etapa de codificación 64 puede responder a esta señalización de nivelación para que, cuando se decodifique la representación codificada correspondiente 40, genere la versión reconstruida 66 del fragmento temporal 10 correspondiente a esta representación codificada 40 dentro de la porción en la cual el fragmento temporal 10 y la trama final 12<sub>a</sub> se superpongan nivelando sus estados internos de la etapa de decodificación 64 como se manifiesta en sí hasta la trama inmediatamente anterior 12 de la trama final 12<sub>a</sub>.

**[0035]** Para ilustrar el procedimiento de nivelación con mayor detalle, se hace referencia a la figura 4, la cual ilustra el caso de generación de una porción restante no codificada de la versión reconstruida 66 para el caso ejemplar de las etapas de codificación y decodificación que operan sobre la base de un códec de transformación. Por ejemplo, se puede utilizar una transformación superpuesta para codificar las tramas. Es decir, que la etapa de codificación 36 usa una ventana 72 de varias ventanas para ponderar los intervalos correspondientes 74 del contenido de audio descomponiendo espectralmente la porción de ventanas resultante mediante el uso de una transformación por descomposición de frecuencia como la MDCT o similar. La porción con ventanas 74 abarca y se extiende temporalmente más allá de los límites de trama 12' actuales. La figura 4, por ejemplo, ilustra que la ventana 72 o porción con ventanas 74 se superpone temporalmente con dos tramas 12 que preceden a la trama actual 12' en dos tramas que siguen a la trama actual 12'. De este modo la representación codificada 40 para un fragmento temporal actual 10 comprende la codificación de la transformación de la porción con ventanas 74 puesto que esta codificación 76 es la representación codificada de la trama 12'. La etapa de decodificación 64 efectúa la inversa para construir las tramas 12 de los fragmentos temporales 10: decodifica la transformación 76 por medio de, por ejemplo, decodificación por entropía, codifica la transformación inversa para dar como resultado una porción con ventana 74 que abarca la trama actual 12' a la cual pertenece la transformación 76, pero la etapa de decodificación 64 efectúa adicionalmente un proceso de superposición/adición entre porciones con ventanas consecutivas 74 para obtener la reconstrucción final del contenido de audio 31'. El proceso de superposición-adición puede ser efectuado por el ensamblador 68. Esto significa lo siguiente: la figura 4, por ejemplo, asume que una trama actual 12' es la penúltima trama que precede inmediatamente a la trama final 12<sub>a</sub> del fragmento temporal actual 10. La etapa de decodificación 64 reconstruye el contenido de audio cubierto por esta penúltima trama 12' efectuando, como ya se expuso, la transformación inversa sobre la transformación 76 para obtener una porción en el dominio del tiempo 76 dentro de la porción con ventanas 74. Como se ha explicado anteriormente, esta porción en el dominio del tiempo 76 se superpone temporalmente con la trama actual 12'. Otras porciones en el dominio del tiempo que han sido obtenidas por transformaciones codificadas por transformación inversa de tramas temporalmente vecinas de la trama actual 12' se superponen temporalmente, sin embargo, también con la trama actual 12'.

**[0036]** En la figura 4, esto se ilustra para porciones con ventanas pertenecientes a las dos tramas anteriores de la trama actual 12' indicadas con el signo de referencia 78 y 80. Una reconstrucción completa de la trama 12 se obtiene sin embargo por el proceso de superposición-adición que suma las porciones de todas las porciones en el dominio del tiempo 76, 78 y 80 resultantes de las transformaciones inversas aplicadas sobre la transformación codificada 76 de la trama 12' y las tramas vecinas de ésta, ya que superpone la trama actual 12' temporalmente. Para la trama última o final 12<sub>a</sub>, esto significa lo siguiente. Incluso si la etapa de codificación 36 no codifica las transformaciones de la porción con ventanas para esta trama final 12<sub>a</sub> en la representación codificada 40, el codificador puede obtener una estimación del contenido de audio dentro de esta trama final 12<sub>a</sub> sumando todas las porciones en el dominio del tiempo que se superpongan temporalmente a la trama final 12<sub>a</sub> de acuerdo a lo obtenido

por la transformación inversa de las transformaciones codificadas 76 de una o más tramas previas, es decir de la trama 12' y opcionalmente una o más tramas 12 que precedan a la penúltima trama 12' dependiendo del tamaño de la ventana, lo cual puede variar en comparación con la figura 4. Por ejemplo, el tamaño de ventana puede ser tal que la superposición temporal con las tramas temporalmente anteriores sea mayor que la superposición temporal con las siguientes tramas. Además, la superposición temporal puede involucrar a la trama inmediatamente anterior y/o inmediatamente siguiente de una trama actualmente codificada.

**[0037]** Existen diferentes posibilidades con respecto a la forma en la cual el decodificador 60 es informado del tamaño de la porción superpuesta 16. Por ejemplo, el decodificador 60 puede ser configurado para transmitir la información de truncación relacionada con este tamaño dentro del flujo de datos 34 por medio de la información de truncación que comprende un valor de longitud de trama y un valor de longitud de fragmento. El valor de longitud de trama podría indicar  $T_{trama}$  y el valor de longitud de fragmento  $T_{fragmento}$ . Otra posibilidad sería que el valor de longitud de truncación indique la longitud temporal de la porción superpuesta 16 en sí o la longitud temporal de la porción en la cual el fragmento temporal 10 y la trama final 12a se superponen temporalmente. Para permitir la finalización inmediata de la versión reconstruida 66 de cada fragmento temporal 10, la etapa de codificación 36 y el proveedor de fragmentos 38 pueden cooperar de modo que, por cada fragmento temporal 10, la representación codificada 40 también sea provista con la información de finalización inmediata que se refiere a la porción 40 que precede temporalmente al fragmento temporal respectivo 10. Por ejemplo, imagine que la transformación superpuesta referida en la figura 4 es una transformación superpuesta que introduce repliegue de espectro o aliasing, tal como un MDCT. En ese caso, sin una versión codificada transformada de la porción anterior 46, un decodificador no podría reconstruir un fragmento temporal actual 10 en su inicio, como dentro de las primeras o más tramas 12 del mismo sin repliegue de espectro o aliasing. En consecuencia, para efectuar la cancelación del repliegue de espectro o aliasing en el dominio del tiempo por medio del proceso de superposición-adición, la información de finalización inmediata transmitida dentro de la representación codificada 40 podría pertenecer a una versión codificada transformada de la porción anterior 46 con las etapas de codificación y decodificación mediante el uso del proceso de codificación por transformación superpuesta como ya se ha ilustrado con respecto a la figura 4.

**[0038]** Aunque no se ha sido tratado con mayor detalle anteriormente, cabe destacar que la etapa de codificación 36 y/o la etapa de decodificación 64 podrían estar compuestas de 2 o incluso más núcleos. Por ejemplo, la figura 2 ilustra que la etapa de codificación podría contener un primer núcleo de codificación 90 y un segundo núcleo de codificación 92 y, de igual forma o de manera adicional y alternativa, la figura 3 muestra que la etapa de decodificación 64 podría comprender un primer núcleo de codificación 94 y un segundo núcleo de decodificación 96. En lugar de codificar/decodificar secuencialmente los fragmentos temporales respectivos 10 y las representaciones codificadas 40, el procedimiento de codificación/decodificación efectuado con respecto a cada uno de esos pares de fragmentos temporales 10 y representaciones codificadas 40 podría ser efectuado de una forma sucesiva con los núcleos 94 y 96 (y 90 y 92) acoplándose alternativamente con una decodificación/codificación de la secuencia de fragmentos temporales 10 y las representaciones codificadas 40, respectivamente.

**[0039]** De este modo, de conformidad con la realización de la figura 2, el codificador de audio alinea la primera trama de audio 12<sub>b</sub> con el inicio 30 del fragmento temporal respectivo 10. Para permitir la finalización íntegra o inmediata de la versión reconstruida respectiva 66 de ese fragmento temporal 10 sin artefactos audibles en el lado de decodificación, el decodificador descrito anteriormente opera o trabaja sobre dos redes de tramas diferentes en los límites de fragmento. También se ha mencionado que para permitir una finalización inmediata de las versiones reconstruidas individuales de las versiones 66 al inicio del fragmento 30, dependiendo del códec de audio subyacente a las etapas de codificación/decodificación, la información de finalización inmediata puede ser transmitida dentro de las representaciones codificadas. Por ejemplo, la primera trama 12<sub>b</sub> de cada fragmento temporal puede ser codificada como una trama de finalización inmediata IPF. Si esa IPF es colocada al inicio de cada nuevo fragmento temporal puede, por ejemplo, abarcar todo el retraso del decodificador. Para ilustrar esto de nuevo, se hace referencia a la figura 5, la cual muestra una porción de un contenido de audio alrededor de un límite de fragmento entre dos fragmentos temporales 10a y 10b. Las tramas 12 en unidades de las cuales los fragmentos temporales 10<sub>a</sub> y 10<sub>b</sub> son codificados/decodificados se muestran también en la figura 5. En particular, la figura 5 revela que la trama final 12<sub>a</sub> del fragmento temporal 10<sub>a</sub> se superpone temporalmente a la primera trama 12<sub>b</sub> de las tramas de la red de tramas mediante el uso de la cual se codifica/decodifica el fragmento temporal 10<sub>b</sub>. En particular, es la porción 16 la cual se extiende más allá del extremo posterior 10<sub>a</sub> y el inicio 30 del fragmento temporal 10<sub>b</sub> de la trama final 12<sub>a</sub>, los que se superponen temporalmente con la primera trama 12<sub>b</sub> del fragmento temporal 10<sub>b</sub>. En la codificación de la primera trama 12<sub>b</sub>, la etapa de codificación codifica adicionalmente en la representación codificada 40 para el fragmento temporal 10<sub>b</sub> la información de finalización inmediata 48, es decir que codifica ejemplarmente aquí 100 de cinco tramas preenrolladas 12 de la red de tramas para codificar/decodificar el fragmento temporal 10<sub>b</sub> que precede a la primera trama 12<sub>b</sub>, estando las tramas preenrolladas indicadas por "AU -5" a "AU -1" en la figura 1. Esas tramas preenrolladas abarcan de este modo la porción anterior 46 anteriormente mencionada 46. Las codificaciones 100, como se ha mencionado anteriormente con respecto a la figura 4, se relacionan con la versión de codificación por transformación del contenido de audio de tramas preenrolladas para permitir que el lado del codificador efectúe la cancelación del dominio del tiempo mediante el uso de porciones en el dominio del tiempo que rodean esas tramas enrolladas usando la transformación inversa y usando las partes que se extienden hacia el fragmento temporal 10<sub>b</sub> para efectuar la cancelación en el dominio del tiempo en el proceso de superposición-adición.

**[0040]** El codificador es consciente de la duración exacta del fragmento. Como se ha explicado anteriormente, de conformidad con una realización, la parte de audio superpuesta 16 puede ser codificada dos veces con diferentes redes de trama.

5

**[0041]** Se efectúa una breve declaración con respecto a la “forma independiente” en la cual los fragmentos temporales individuales 10 son codificados en sus representaciones codificadas 40. Aunque esta forma independiente podría pertenecer también a datos de configuración como parámetros de codificación pertenecientes a datos más raramente cambiantes como un número de canales de audio codificados o similares, de modo que cada representación codificada 40 podría comprender estos datos de configuración, sería alternativamente posible que esos datos raramente cambiantes, es decir datos de configuración, sean transmitidos al lado de decodificación fuera de banda, no dentro de cada representación codificada 40 en lugar de ser incluidos en cada representación codificada 40. Si son incluidos en la representación codificada, los datos de configuración pueden ser transmitidos en otra capa de transporte. Por ejemplo, la configuración puede ser transmitida en el segmento de inicialización, y la trama de IPF 12b de cada fragmento temporal podría ser liberada llevando la información de datos de configuración.

**[0042]** En lo que se refiere al lado de decodificación, la descripción anterior de la figura 3 reveló que el decodificador está configurado para decodificar tramas preenrolladas, es decir tramas que preceden a la primera trama 12<sub>b</sub> por cada fragmento temporal. El decodificador puede atender a esta decodificación sin importar si la configuración cambia del fragmento temporal anterior a un fragmento temporal actual. Esto por supuesto tiene impacto sobre el rendimiento total del decodificador, pero de manera ventajosa, un decodificador ya pudo haber satisfecho un requisito de conformidad con el cual el decodificador puede decodificar una IPF sobre cada límite de fragmento como, por ejemplo, de conformidad con un caso de uso de flujo unidireccional adaptable en el peor de los casos, de modo que no se imponga un requisito adicional en el caso de esos casos. En lo que se refiere a la información de transacción mencionada anteriormente, debería destacarse que se puede efectuar una señalización de la misma a nivel de flujo de bits, o en alguna otra capa de transporte como con herramientas a nivel de sistema.

**[0043]** Finalmente, la figura 6 muestra un caso donde el contenido de audio 31 que será codificado muestra un cambio de configuración como un cambio en un número de canales de audio, en algún punto en el tiempo 110, es decir en un límite de fragmento entre dos fragmentos temporales 10. Por ejemplo, el instante de tiempo inmediatamente anterior 110, una primera configuración como estéreo se aplica, mientras que después del instante de tiempo 110, el contenido de audio 31 es por ejemplo una escena de audio de cinco canales. El flujo de datos de audio 34 comprende la información de datos de configuración. De este modo, está claro a partir del flujo de datos 34 que las representaciones codificadas del flujo de datos de los fragmentos temporales 10 que preceden al instante de tiempo 110 son codificados según la primera configuración, y que la segunda configuración se utiliza para codificar los fragmentos temporales 10 que siguen al instante de tiempo 110. La figura 6 muestra también la información de finalización inmediata 98 de la representación codificada 40. En el caso del fragmento temporal 10 que precede al instante de tiempo 110, la información de finalización inmediata 98 puede ser derivada como se ha descrito anteriormente con respecto a la figura 5, por ejemplo. Sin embargo, la situación es diferente para el fragmento temporal 10 que comienza inmediatamente en el instante de tiempo 110. Aquí, el contenido de audio 39 no permite formar la información de finalización inmediata 98 para la representación codificada 40 del fragmento temporal que comienza inmediatamente al instante de tiempo 110, puesto que el contenido de audio 39 en la segunda configuración no es elevable al tiempo antes del instante de tiempo 110. Una señal de cero puede ser codificada como la información de finalización inmediata 98 con respecto a este fragmento temporal 10 que comienza al instante de tiempo 110, en su lugar. Es decir que, en el caso de un cambio de configuración, el codificador puede codificar muestras de cero puesto que no existe señal de audio actual disponible para el pasado, como, por ejemplo, cuando se cambia de mono a 5.1 o similar. Una posible optimización sería generar esta trama de cero, es decir trama preenrollada de cero, en el lado del decodificador y transmitir únicamente la codificación de la primera trama 12<sub>b</sub> del primer fragmento temporal. Es decir que, en ese caso la información de finalización inmediata 98 podría dejarse fuera completamente.

**[0044]** De este modo, las realizaciones anteriores permiten la distribución de contenido de audio y vídeo sobre un canal de transmisión con una velocidad de bits fija o variable y permiten, en particular, la sincronización de audio y vídeo y permiten casos de uso avanzados como el empalme, como se mencionó anteriormente, el flujo de datos codificado como se codificó anteriormente, puede volverse más fácil de sincronizar con otros relojes tales como los relojes prescritos por otras señales de medios. Los codificadores descritos anteriormente permiten una adaptación de una longitud de trama de audio existente. La longitud de los fragmentos temporales puede ser establecida o fijada dependiendo de las necesidades de aplicación. Las realizaciones de codificador forman el flujo de datos codificado en tramos de la representación codificada de los fragmentos temporales que puede, por ejemplo, pero no exclusivamente, ser objeto del flujo unidireccional adaptable mediante el uso de esos fragmentos como los fragmentos de una representación de medios. Es decir que, el flujo de datos, compuesto de los fragmentos resultantes, puede ser ofrecido a un cliente por medio de un servidor a través de un protocolo de flujo unidireccional adaptable, y el cliente puede recuperar los fragmentos de flujo de datos con, quizá, una adición insertada en estos, a través del protocolo y enviar los mismos al decodificador para la decodificación. Pero esto no es obligatorio. En su lugar, el empalme puede ser efectuado de manera ventajosa por la formación del flujo de datos codificado de la invención incluso en otros escenarios de aplicación. Las realizaciones descritas anteriormente pueden ser

implementadas o usadas en relación con el códec de audio MPEG-H con las tramas de audio siendo las tramas de audio de MPEG-H, pero las realizaciones anteriores no se restringen al uso de este códec sino que pueden ser adaptadas a todos los códecs de audio (modernos).

5 **[0045]** Aunque algunos aspectos han sido descritos en el contexto de un aparato, está claro que esos aspectos también representan una descripción del método correspondiente, donde un bloque o dispositivo corresponde a una etapa del método o una característica de una etapa del método. De manera análoga, los aspectos descritos en el contexto de una etapa del método también representan una descripción de un bloque o elemento o característica correspondiente de un aparato correspondiente. Algunas o todas las etapas del método  
10 pueden ser ejecutadas por (o mediante el uso de) un aparato de hardware, como por ejemplo, un microprocesador, como un ordenador programable o un circuito electrónico. En algunas realizaciones, algunas de una o más de las etapas del método más importantes pueden ser ejecutadas por ese aparato.

**[0046]** Los flujos de datos de audio empalmados o que se pueden empalmar de la invención pueden ser  
15 almacenados en un medio de almacenamiento digital o pueden ser transmitidos sobre un medio de transmisión como un medio de transmisión inalámbrica o un medio de transmisión alámbrica tal como Internet.

**[0047]** Dependiendo de ciertos requisitos de implementación, las realizaciones de la invención pueden ser implementadas en hardware o software. La implementación puede ser efectuada mediante el uso de un medio de  
20 almacenamiento digital, por ejemplo un disco flexible, un DVD, un Blu-Ray, un CD, una ROM, una PROM, una EPROM, una EEPROM o una memoria FLASH, que tenga señales de control legibles electrónicamente almacenadas en ella, que cooperen (o sean capaces de cooperar) con un sistema informático programable de modo que sea efectuado el método respectivo. Por lo tanto, el modo de almacenamiento digital puede ser legible por ordenador.

25 **[0048]** Algunas realizaciones según la invención comprenden un soporte de datos que tiene señales de control legibles electrónicamente, las cuales pueden cooperar con un sistema informático programable, de modo que sea efectuado uno de los métodos descritos en esta invención.

30 **[0049]** Generalmente, las realizaciones de la presente invención pueden ser implementadas como un producto de programa informático con un código de programa, siendo operativo el código de programa para efectuar uno de los métodos cuando el producto de programa informático se ejecute en un ordenador. El código de programa puede ser almacenado por ejemplo en un soporte legible por máquina.

35 **[0050]** Otras realizaciones comprenden el programa informático para efectuar uno de los métodos descritos en esta invención, almacenado en un soporte legible por máquina.

**[0051]** En otras palabras, una realización del método de la invención es, por lo tanto, un programa informático que tiene un código de programa para efectuar uno de los métodos descritos en esta invención, cuando el programa  
40 informático se ejecute en un ordenador.

**[0052]** Una realización adicional de los métodos de la invención es, por lo tanto, un soporte de datos (o un medio de almacenamiento digital, o un medio legible por ordenador) que comprende, registrado en él, el programa informático para efectuar uno de los métodos descritos en esta invención. El soporte de datos, el medio de  
45 almacenamiento digital o el medio registrado son típicamente tangibles y/o no transitorios.

**[0053]** Una realización adicional del método de la invención es, por lo tanto, un flujo de datos o una secuencia de señales que representa el programa informático para efectuar uno de los métodos descritos en esta invención. El flujo de datos o secuencia de señales puede ser configurado por ejemplo para ser transferido a través de una  
50 conexión de comunicación de datos, por ejemplo a través de Internet.

**[0054]** Una realización adicional comprende medios de procesamiento, por ejemplo un ordenador, o un dispositivo lógico programable, configurado para o adaptado para efectuar uno de los métodos descritos en esta invención.

55 **[0055]** Una realización adicional comprende un ordenador que tiene instalado en él, el programa informático para efectuar uno de los métodos descritos en esta invención.

**[0056]** Una realización adicional de conformidad con la invención comprende un aparato o un sistema configurado para transferir (por ejemplo, electrónicamente u ópticamente) un programa informático para efectuar uno de los métodos descritos en esta invención a un receptor. El receptor puede, por ejemplo, ser un ordenador, un dispositivo móvil, un dispositivo de memoria o similar. El aparato o sistema puede, por ejemplo, comprender un servidor de archivos para transferir el programa informático al receptor.

65 **[0057]** En algunas realizaciones, se puede utilizar un dispositivo lógico programable (por ejemplo una matriz de puertas programable en el campo) para efectuar algunas o todas las funcionalidades de los métodos descritos en

esta invención. En algunas realizaciones, una matriz de puertas programable en el campo puede cooperar con un microprocesador para efectuar uno de los métodos descritos en esta invención. Generalmente, los métodos son efectuados preferiblemente por cualquier aparato de hardware.

5 **[0058]** El aparato descrito en esta invención puede ser implementado mediante el uso de un aparato de hardware, o mediante el uso de un ordenador, o mediante el uso de una combinación de un aparato de hardware y un ordenador.

10 **[0059]** Los métodos descritos en esta invención pueden ser efectuados mediante el uso de un aparato de hardware, o mediante el uso de un ordenador, o mediante el uso de una combinación de un aparato de hardware y un ordenador.

15 **[0060]** Las realizaciones descritas anteriormente son meramente ilustrativas de los principios de la presente invención. Debe comprenderse que las modificaciones y variaciones de las disposiciones y los detalles descritos en esta invención serán evidentes para otros expertos en la técnica. Se pretende, por lo tanto, estar limitado únicamente por el alcance de las reivindicaciones de patente inminentes y no por los detalles específicos presentados a modo de descripción y explicación de las realizaciones de esta invención.

Definiciones y abreviaturas

20

**[0061]**

<b>AAC</b>	Codificación de Audio Avanzada
<b>ATSC</b>	Comité de Sistemas de Televisión Avanzados
<b>AU</b>	Unidad de Acceso de Audio
<b>DASH</b>	Flujo Unidireccional Adaptable Dinámico sobre HTTP
<b>DVB</b>	Transmisión de Vídeo Digital
<b>IPF</b>	Trama de Finalización Instantánea
<b>MPD</b>	Descripción de Presentación de Medios
<b>MPEG</b>	Grupo de Expertos en Imágenes en Movimiento
<b>MMT</b>	Transporte de Medios MPEG
<b>NTSC</b>	Comité de Sistemas de Televisión Nacional
<b>PAL</b>	Línea de Fase Alternada –Verfahren

Referencias

25

**[0062]**

[1] "Delivery/Sync/FEC-Evaluation Criteria Report", ROUTE/DASH

30 [2] ISO/IEC 23008-3, "Information technology — High efficiency coding and media delivery in heterogeneous environments — Part 3: 3D audio"

[3] ISO/IEC 23009-1, "Information technology — Dynamic adaptive streaming over HTTP (DASH) — Part 1: Media presentation description and segment formats"

35

[4] ISO/IEC 23008-1, "Information technology -- High efficiency coding and media delivery in heterogeneous environments -- Part 1: MPEG media transport (MMT)"

## REIVINDICACIONES

1. Codificador (20) para la codificación de contenido de audio en un flujo de datos codificados (34), que comprende:
- 5 una etapa de codificación (36) configurada para codificar el contenido de audio en unidades de tramas de audio (12);  
y  
un proveedor de fragmentos (38) configurado para proporcionar el contenido de audio (31) a la etapa de codificación (36) en unidades de fragmentos temporales (10) de una red de fragmento proporcionando, para un fragmento temporal proporcionado actualmente (10), una porción del contenido de audio (31) a la etapa de codificación (36) que incluye el fragmento temporal proporcionado actualmente (10),  
10 en el que el codificador está configurado para codificar cada fragmento temporal (10) en una representación codificada (40) del fragmento temporal respectivo en unidades de tramas de audio (12), y el proveedor de fragmentos está configurado para proporcionar el contenido de audio (31) a la etapa de codificación (36) de modo que las tramas de audio (12) estén alineadas al fragmento temporal respectivo (10) de modo que para cada fragmento temporal (10) un inicio de una primera trama de audio y un inicio (30) del fragmento temporal respectivo coincidan, y  
15 en el que las representaciones codificadas de los fragmentos temporales (10) están comprendidas por el flujo de datos codificado, y una longitud temporal de los fragmentos temporales (10) es un múltiplo no entero de una longitud temporal de las tramas de audio,  
20 en el que el codificador está configurado para señalar dentro del flujo de datos codificado una información de truncación para identificar una porción de una trama de audio final de las tramas de audio en unidades de las cuales los fragmentos temporales son codificados, que excede un extremo final de los fragmentos temporales y se superpone temporalmente con un fragmento temporal inmediatamente posterior de la red de fragmentos, en el que la información de truncación comprende  
25 un valor de longitud de trama que indica la longitud temporal de las tramas de audio y un valor de longitud de fragmento que indica la longitud temporal de los fragmentos temporales y/o  
un valor de longitud de truncación que indica una longitud temporal de una porción de trama de audio final de las tramas de audio en unidades de las cuales son codificados los fragmentos temporales, que excede un extremo posterior de los fragmentos temporales y se superpone temporalmente con un fragmento temporal inmediatamente posterior, o la diferencia entre la longitud temporal de la porción de la trama de audio final y la longitud temporal de la trama de audio final.
2. Codificador (20) según la reivindicación 1, en el que la etapa de codificación (36) y el proveedor de fragmentos (38) cooperan de modo que, para un fragmento temporal predeterminado (10), la codificación del fragmento temporal predeterminado (10) en la representación codificada (40) del fragmento temporal respectivo se detiene en una trama de audio que precede inmediatamente a la trama de audio final de tramas de audio en unidades de las cuales el fragmento temporal predeterminado es codificado, que excede un extremo final del fragmento temporal predeterminado y se superpone temporalmente con un fragmento temporal inmediatamente posterior de la red de fragmentos.
3. Codificador (20) según la reivindicación 2, en el que la etapa de codificación (36) está configurada para señalar dentro de la representación codificada (40) del fragmento temporal predeterminado una señalización de nivelación que da instrucciones al decodificador para llenar una porción del fragmento temporal predeterminado abarcado por la trama de audio final sobre la base de los estados internos de la nivelación del codificador como se manifiestan en sí a la trama de audio que precede inmediatamente a la trama de audio final.
4. Codificador (20) según la reivindicación 1, en el que la etapa de codificación (36) y el proveedor de fragmentos (38) cooperan de forma que, para un fragmento temporal predeterminado (10), la codificación del fragmento temporal predeterminado en la representación codificada del fragmento temporal respectivo continúa más allá de un extremo final del fragmento temporal predeterminado dentro de la trama de audio final de las tramas de audio (12) en unidades de las cuales se codifica el fragmento temporal predeterminado que excede el extremo final del fragmento temporal predeterminado y se superpone de forma temporal con un fragmento temporal inmediatamente posterior de la red de fragmentos.
5. Codificador (20) según la reivindicación 4, en el que la etapa de codificación (36) está configurada para codificar el contenido de audio (31) dentro de una porción de la trama de audio final que excede el extremo final del fragmento temporal predeterminado y se superpone temporalmente con el fragmento temporal inmediatamente posterior, a una calidad menor que dentro del fragmento temporal predeterminado.
6. Codificador (20) según cualquiera de las reivindicaciones 1 a 5, en el que la etapa de codificación (36) y el proveedor de fragmentos (38) cooperan de modo que, por un fragmento temporal predeterminado (10) la codificación del fragmento temporal predeterminado en la representación codificada del fragmento temporal respectivo implica una derivación de información de finalización inmediata (98) de una o más tramas de audio preenrolladas de contenido de audio que precede inmediatamente la primera trama de audio de las tramas de audio en unidades de las cuales se codifica el fragmento temporal predeterminado (10) en la representación codificada

(40) del fragmento temporal respectivo y codifica la información de finalización inmediata en la representación codificada del fragmento temporal predeterminado.

7. Codificador (20) según la reivindicación 6, en el que la etapa de codificación (36) está configurada para efectuar la codificación mediante el uso de la codificación de transformación sobre la base de un repliegue de espectro o aliasing que introduce la transformación superpuesta y para derivar la información de finalización inmediata aplicando la codificación por transformación sobre la base del repliegue de espectro o aliasing que introduce la transformación superpuesta sobre el contenido de audio en una o más tramas de audio preenrolladas.
8. Codificador según cualquiera de las reivindicaciones 1 a 7, en el que la etapa de codificación (36) comprende un primer núcleo de codificación (90) y un segundo núcleo de codificación (92), en el que el proveedor de fragmentos (38) está configurado para acoplarse al primer núcleo de codificación (90) con la codificación de un primer fragmento temporal del contenido de audio, estando el primer núcleo de codificación (90) configurado para codificar el primer fragmento temporal del contenido de audio en unidades de trama de audio con alineación temporal de las tramas de audio con un inicio del primer fragmento temporal de modo que la primera trama de audio de las tramas de audio en unidades de las cuales se codifica el primer fragmento temporal del contenido de audio, que comienza inmediatamente al inicio del primer fragmento temporal para producir una representación codificada del primer fragmento temporal, y el proveedor de fragmentos (38) está configurado para acoplarse al segundo núcleo de codificación con la codificación de un segundo fragmento temporal del contenido de audio, inmediatamente posterior al primer fragmento temporal, estando el segundo núcleo de codificación (90) configurado para codificar el segundo fragmento temporal del contenido de audio en unidades de trama de audio con alineación temporal de las tramas de audio con un inicio del segundo fragmento temporal de modo que la primera trama de audio de las tramas de audio en unidades de las cuales se codifica el segundo fragmento temporal del contenido de audio, comenzando inmediatamente al inicio del segundo fragmento temporal para producir una representación codificada del segundo fragmento temporal, en el que el proveedor está configurado para acoplarse al primer núcleo de codificación (90) con la codificación también de un tercer fragmento temporal del contenido de audio.
9. Codificador (20) según la reivindicación 8, en el que el proveedor de fragmento (38) está configurado para acoplarse al primer y segundo núcleos de codificación (90, 92) de manera alternada con la codificación de los fragmentos temporales del contenido de audio.
10. Decodificador (60) para la decodificación de contenido de audio de un flujo de datos codificado, que comprende una interfaz de entrada (62) configurada para recibir las representaciones codificadas de fragmentos temporales (10) de una red de fragmentos (32) del contenido de audio, cada uno de los cuales tiene codificado en él un fragmento temporal respectivo en unidades de tramas de audio alineadas temporalmente con un inicio del fragmento temporal de modo que el inicio del fragmento temporal coincida con un inicio de una primera trama de audio de las tramas de audio; una etapa de decodificación (64) configurada para decodificar versiones reconstruidas (66) de los fragmentos temporales del contenido de audio de representaciones codificadas de los fragmentos temporales; y un ensamblador (68) configurado para ensamblar, por finalización, las versiones reconstruidas de los fragmentos temporales del contenido de audio juntos, en el que una longitud temporal entre los límites de fragmento de la red de fragmentos es un múltiplo no entero de una longitud temporal de las tramas de audio, en el que el ensamblador (68) está configurado para truncar la versión reconstruida (66) de un fragmento temporal predeterminado en una porción de una trama de audio final de las tramas de audio en unidades de las cuales se codifica el fragmento temporal predeterminado en la representación codificada del fragmento temporal predeterminado, que excede temporalmente un extremo final del fragmento temporal predeterminado y se superpone temporalmente con una versión reconstruida de un fragmento temporal inmediatamente posterior, en el que el decodificador está configurado para determinar la porción de la trama de audio final sobre la base de la información de truncación en el flujo de datos codificado, en el que la información de truncación comprende un valor de longitud de trama que indica una longitud temporal de las tramas de audio en unidades de las cuales se codifica el fragmento temporal predeterminado en la representación codificada del fragmento temporal predeterminado, y un valor de longitud de fragmento que indica una longitud temporal del fragmento temporal predeterminado desde el inicio de la versión reconstruida del fragmento predeterminado hasta el límite de fragmento con el cual el inicio de la versión reconstruida del fragmento temporal posterior coincide, y/o un valor de longitud de truncación que indica una longitud temporal de la porción de la trama de audio final o la diferencia entre la longitud temporal de la porción de la trama de audio final y la longitud temporal de la trama de audio final.
11. Decodificador (60) según la reivindicación 10, en el que la etapa de decodificación (64) está configurada para que, en la decodificación de un fragmento temporal predeterminado de la representación codificada del fragmento temporal predeterminado, genere la versión reconstruida del fragmento temporal predeterminado dentro de una porción de una trama de audio final de las tramas de audio en unidades de las cuales es codificado el

fragmento temporal predeterminado en la representación codificada del fragmento temporal predeterminado, que se extiende desde un extremo delantero de la trama de audio final hasta el límite de fragmento de una versión reconstruida de un fragmento temporal que sigue nivelando estados internos de la etapa de decodificación como se manifiesta en sí hasta una trama de audio que precede inmediatamente a la trama de audio final.

5 12. Decodificador (60) según la reivindicación 10 u 11, en el que la etapa de decodificación (64) está configurada para derivar información de finalización inmediata a partir de las representaciones codificadas de un fragmento temporal predeterminado, estando la información de finalización inmediata relacionada con el contenido de audio en una o más tramas de audio preenrolladas del contenido de audio que precede temporalmente a un inicio del fragmento temporal predeterminado y usa la información de finalización inmediata para reconstruir el contenido de audio en una o más tramas de audio del fragmento temporal predeterminado inmediatamente posterior al inicio del fragmento temporal.

15 13. Decodificador (60) según la reivindicación 12, en el que la etapa de decodificación (64) está configurada de modo que la información de finalización inmediata sea una reconstrucción del contenido de audio en una o más tramas de audio preenrolladas.

20 14. Decodificador (60) según la reivindicación 12 ó 13, en el que la etapa de decodificación (64) está configurada para usar la información de finalización inmediata en la reconstrucción del contenido de audio en una o más tramas de audio del fragmento temporal predeterminado inmediatamente posterior al inicio del fragmento temporal para la cancelación del repliegue o aliasing en el dominio del tiempo.

25 15. Decodificador según las reivindicaciones 10 a 14, en el que la etapa de decodificación (64) está configurada para decodificar las tramas de audio individualmente mediante el uso de una inversa de una transformación superpuesta que causa repliegue o aliasing e incurre en ventanas de transformación que se extiende más allá de los límites de las tramas.

30 16. Decodificador según cualquiera de las reivindicaciones 10 a 15, en el que la etapa de decodificación (64) comprende un primer núcleo de decodificación (94) configurado para decodificar una versión reconstruida de un primer fragmento temporal del contenido de audio en unidades de tramas de audio de una representación codificada del primer fragmento temporal de modo que la versión reconstruida del primer fragmento temporal comience en un extremo delantero de una primera trama de audio de las tramas de audio del primer fragmento temporal; un segundo núcleo de decodificación (96) configurado para decodificar una versión reconstruida de un segundo fragmento temporal del contenido de audio, inmediatamente posterior al primer fragmento temporal en unidades de tramas de audio, de una representación codificada del segundo fragmento temporal, de modo que la versión reconstruida del segundo fragmento temporal comience registrada en un extremo delantero de una primera trama de audio de las tramas de audio del segundo fragmento temporal, en el que el ensamblador (68) está configurado para ensamblar la versión reconstruida (66) del primer fragmento temporal y la versión reconstruida del segundo fragmento temporal juntos.

45 17. Decodificador según la reivindicación 16, en el que el primer núcleo de decodificación está configurado para decodificar también una versión reconstruida de un tercer fragmento temporal del contenido de audio del flujo de datos codificado.

18. Decodificador según la reivindicación 16 ó 17, en el que el primer y segundo núcleos de decodificación están configurados para atender alternativamente la a decodificación de versiones reconstruidas de los fragmentos temporales del contenido de audio de las representaciones codificadas de los fragmentos temporales.

50 19. Método (20) para la codificación de contenido de audio en un flujo de datos codificado (34), mediante el uso de una etapa de codificación (36) configurada para codificar el contenido de audio en unidades de tramas (12), comprendiendo el método el suministro del contenido de audio (31) a la etapa de codificación (36) en unidades de fragmentos temporales (10) de una red de fragmentos (32) proporcionando, para un fragmento temporal proporcionado actualmente (10), una porción del contenido de audio (31) a la etapa de codificación (36) que incluye el fragmento temporal proporcionado actualmente (10), la codificación, realizada por medio de la etapa de codificación, de cada fragmento temporal (10) en una representación codificada (40) del fragmento temporal respectivo en unidades de tramas de audio (12), en el que el contenido de audio (31) se proporciona a la etapa de codificación (36) de modo que las tramas de audio (12) estén alineadas con el fragmento temporal respectivo (10) de modo que para cada fragmento temporal (10) un inicio de la primera trama de audio de las tramas de audio en unidades en las cuales el fragmento temporal respectivo (10) sea codificado en la representación codificada (40) del fragmento temporal respectivo y un inicio (30) del fragmento temporal respectivo coincidan, en el que las representaciones codificadas de los fragmentos temporales (10) están comprendidas por el flujo de datos codificado, y una longitud temporal de los fragmentos temporales (10) es un múltiplo no entero de una longitud temporal de las tramas,

- en el que el método comprende la señalización dentro del flujo de datos codificado de una información de truncación para identificar una porción de una trama de audio final de las tramas de audio en unidades de las cuales son codificados los fragmentos temporales, que excede un extremo final de los fragmentos temporales y se superpone temporalmente con un fragmento temporal inmediatamente posterior de la red de fragmentos, en el que la información de truncación comprende
- 5 un valor de longitud de trama que indica la longitud temporal de las tramas de audio y un valor de longitud de fragmento que indica la longitud temporal de los fragmentos temporales y/o
- 10 un valor de longitud de truncación que indica una longitud temporal de una porción de una trama de audio final de las tramas de audio en unidades en las cuales son codificados los fragmentos temporales, que excede un extremo final de los fragmentos temporales y se superpone temporalmente con un fragmento temporal inmediatamente posterior de la red de fragmentos, o la diferencia entre la longitud temporal de la porción de la trama de audio final y la longitud temporal de la trama de audio final.
- 15 20. Método (60) para la decodificación de contenido de audio en unidades de fragmentos temporales de una red de fragmentos de un flujo de datos codificado, que comprende
- la recepción de representaciones codificadas de fragmentos temporales de una red de fragmentos del contenido de audio, cada uno de los cuales tiene codificado en él un fragmento temporal respectivo en unidades de tramas de audio alineadas temporalmente con un inicio del fragmento temporal respectivo de modo que el inicio del fragmento
- 20 temporal respectivo coincida con un inicio de una primera trama de audio de las tramas de audio;
- la decodificación de versiones reconstruidas (66) de los fragmentos temporales del contenido de audio de las representaciones codificadas de los fragmentos temporales; y
- el ensamble, por finalización, de las versiones reconstruidas de los fragmentos temporales del contenido de audio juntos,
- 25 en el que una longitud temporal entre límites de fragmento de la red de fragmentos es un múltiplo no entero de una longitud temporal de las tramas de audio,
- en el que el ensamble (68) comprende el truncado de la versión reconstruida (66) de un fragmento temporal predeterminado en una porción de una trama de audio final de las tramas de audio en unidades de las cuales se codifica el fragmento temporal predeterminado en la representación codificada del fragmento temporal
- 30 predeterminado, que excede temporalmente un extremo final del fragmento temporal predeterminado y se superpone temporalmente con una versión reconstruida de un fragmento temporal inmediatamente posterior,
- en el que el método comprende además la determinación de la porción de la trama de audio final sobre la base de la información de truncación en el flujo de datos codificado,
- en el que la información de truncación comprende
- 35 un valor de longitud de trama que indica una longitud temporal de las tramas de audio en unidades de las cuales se codifica el fragmento temporal predeterminado en la representación codificada del fragmento temporal predeterminado, y un valor de longitud de fragmento que indica una longitud temporal del fragmento temporal predeterminado desde el inicio de la versión reconstruida del fragmento predeterminado hasta el límite de fragmento con el cual el inicio de la versión reconstruida del fragmento temporal siguiente coincide, y/o
- 40 un valor de longitud de truncación que indica longitud temporal de la porción de la trama de audio final o la diferencia entre la longitud temporal de la porción de la trama de audio final y la longitud temporal de la trama de audio final.
21. Programa informático que tiene un código de programa para llevar a cabo, cuando se ejecuta en un ordenador, un método según la reivindicación 19 o 20.
- 45

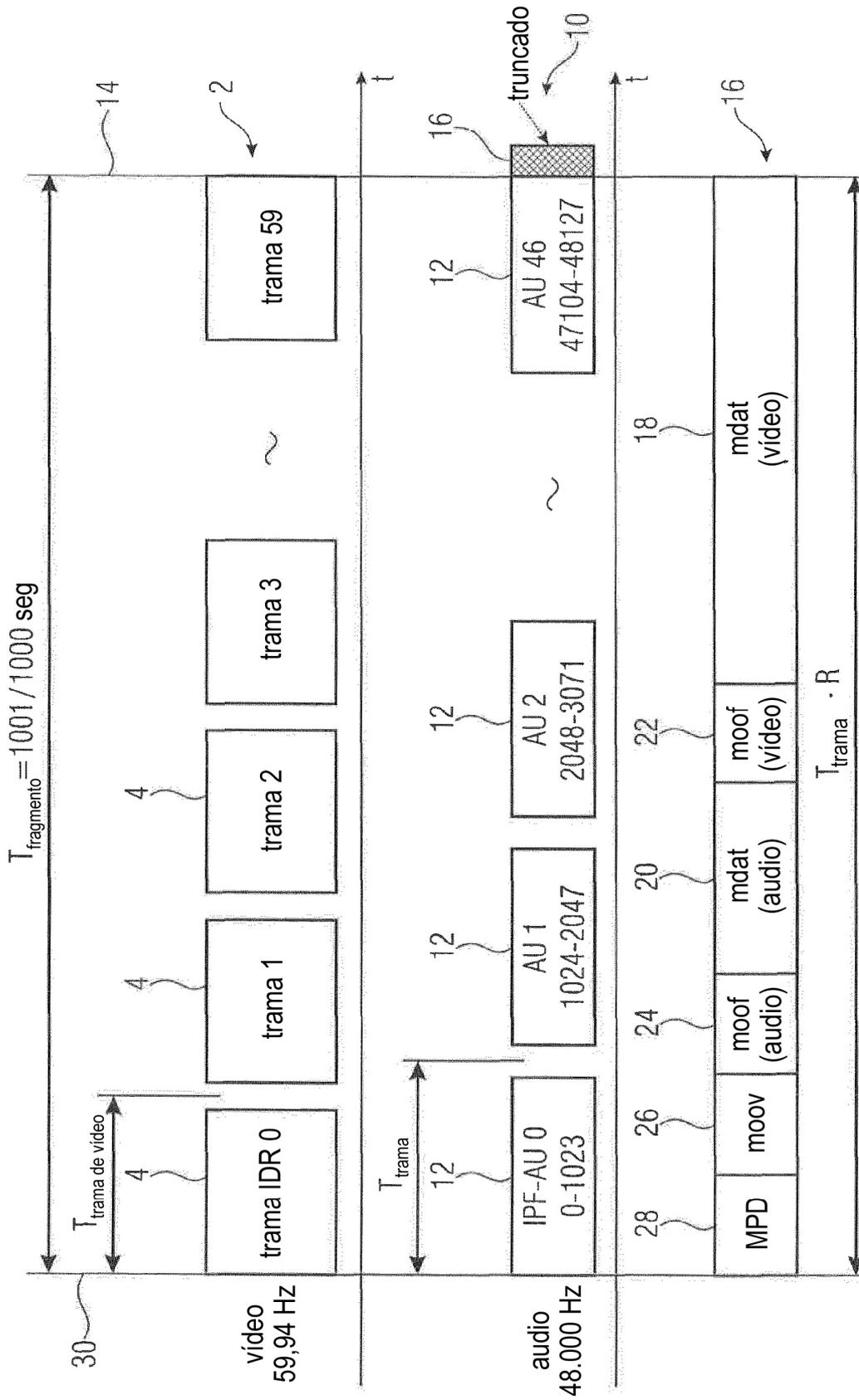


FIG 1

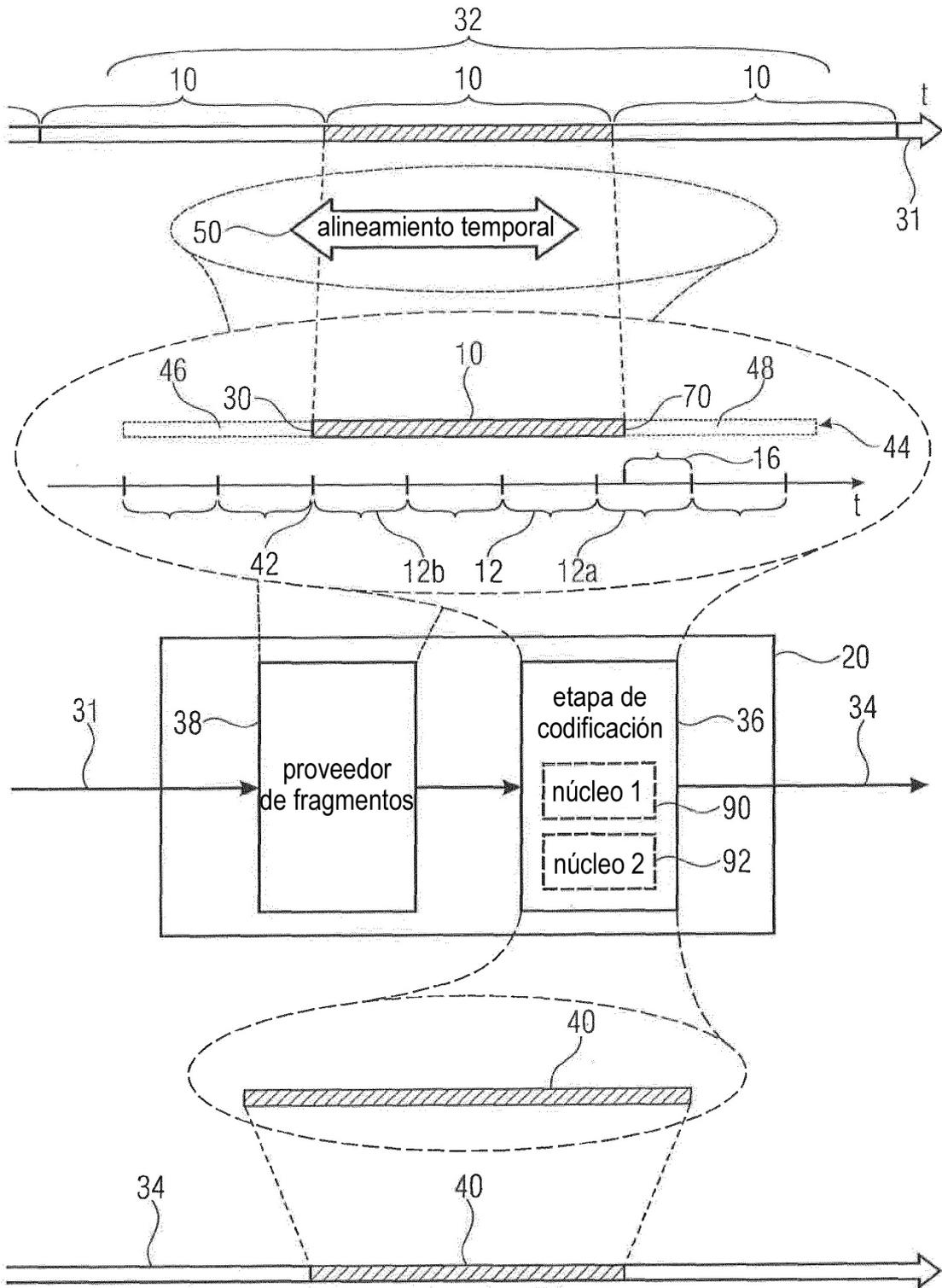


FIG 2

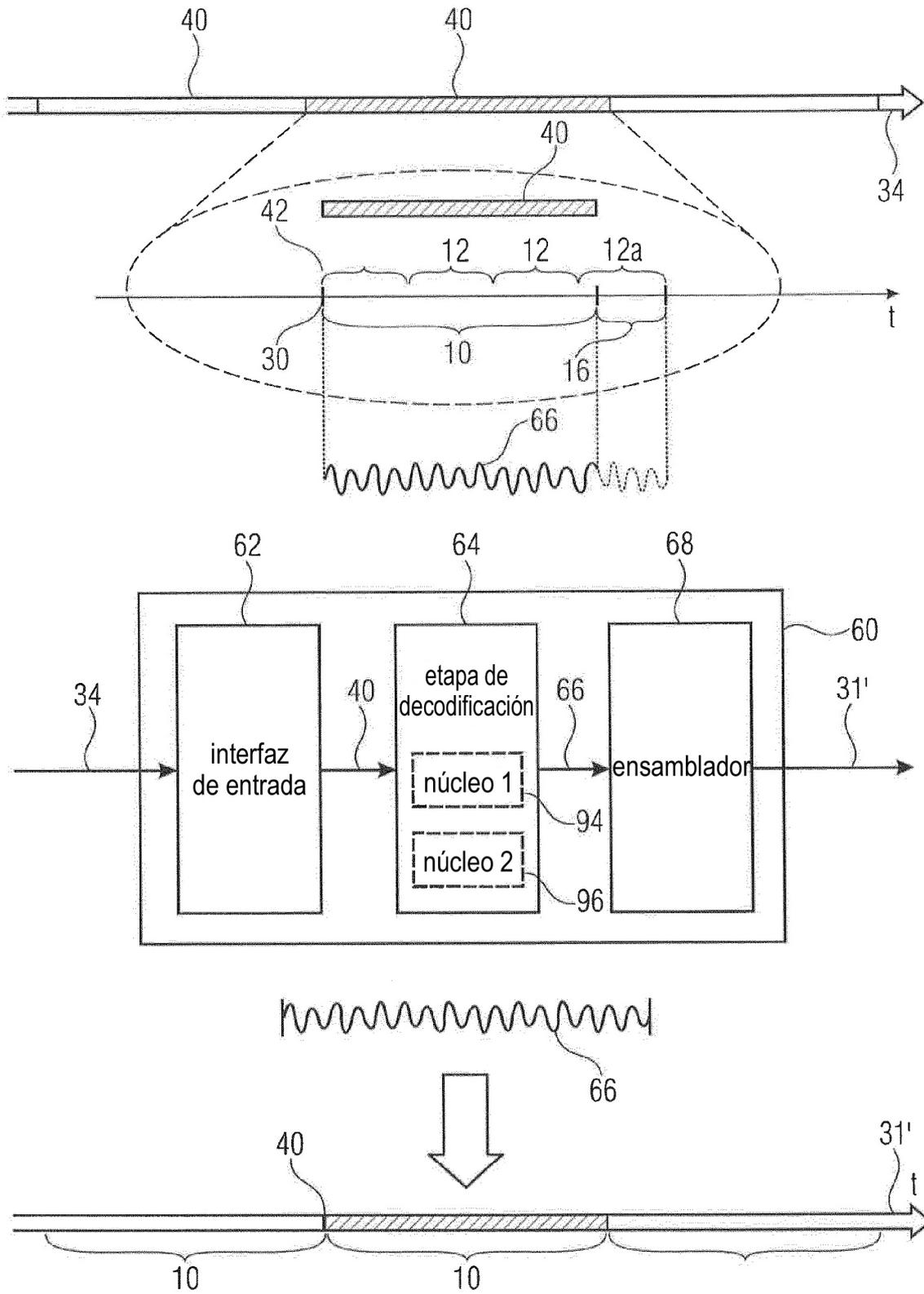


FIG 3

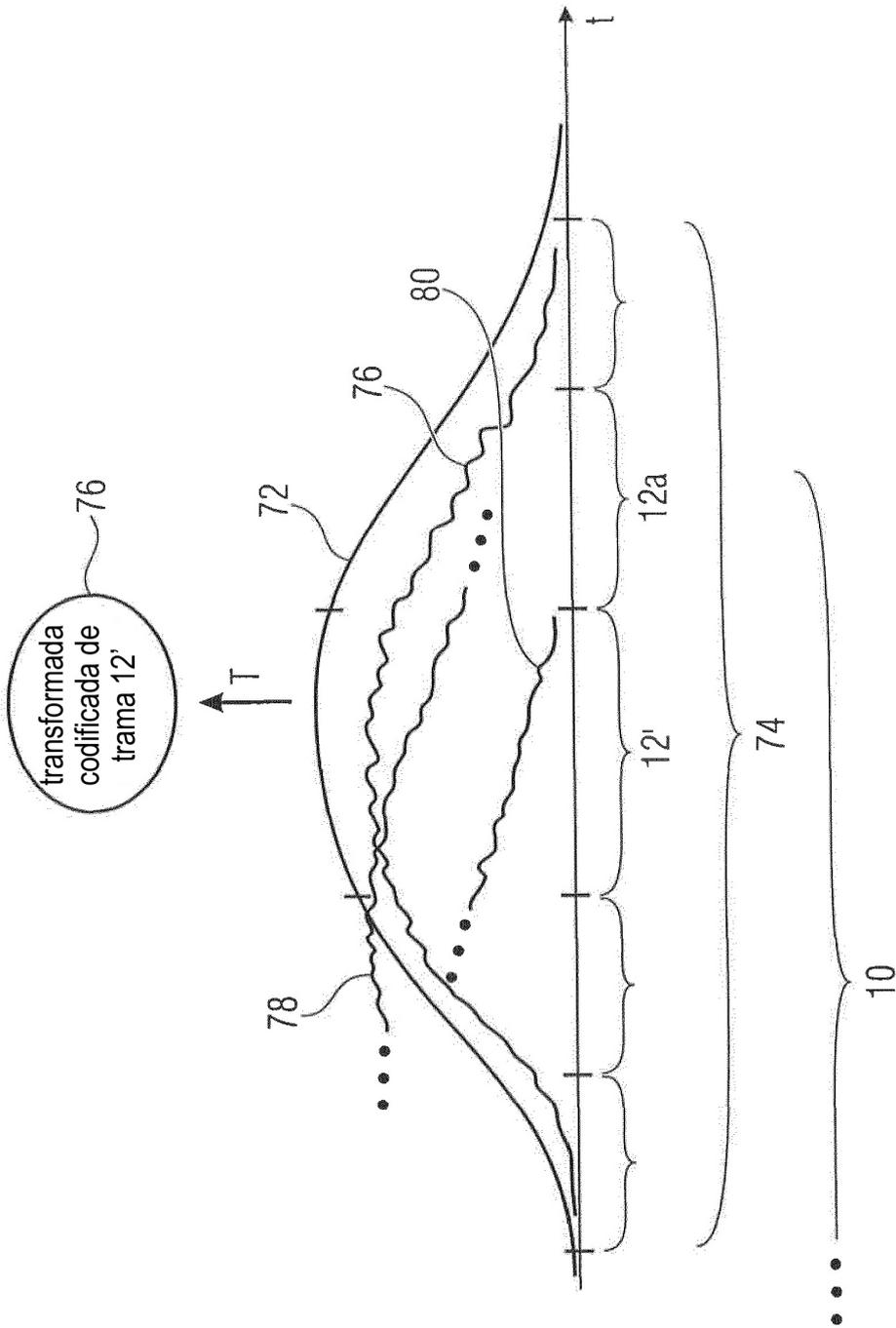


FIG 4

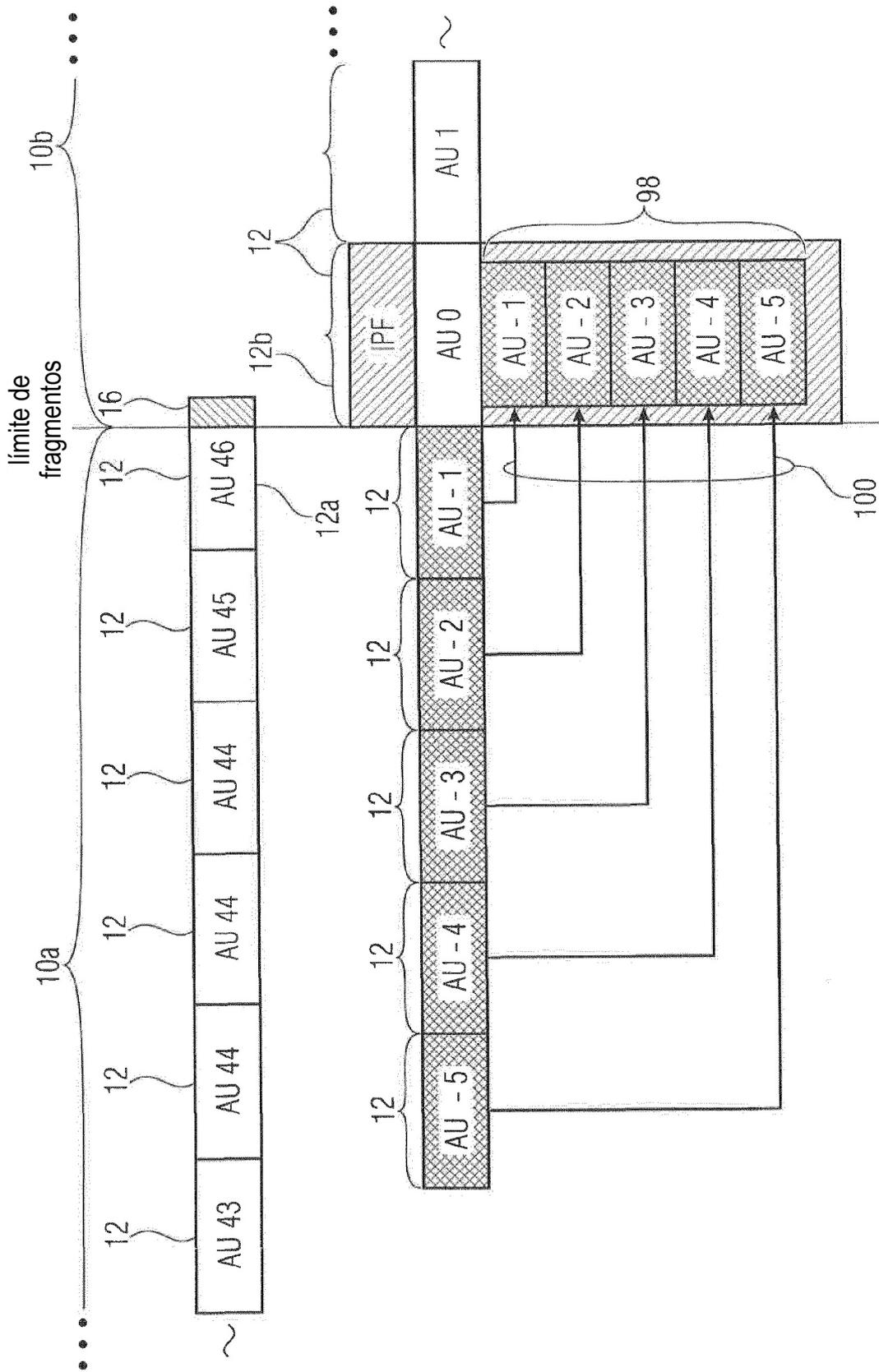


FIG 5

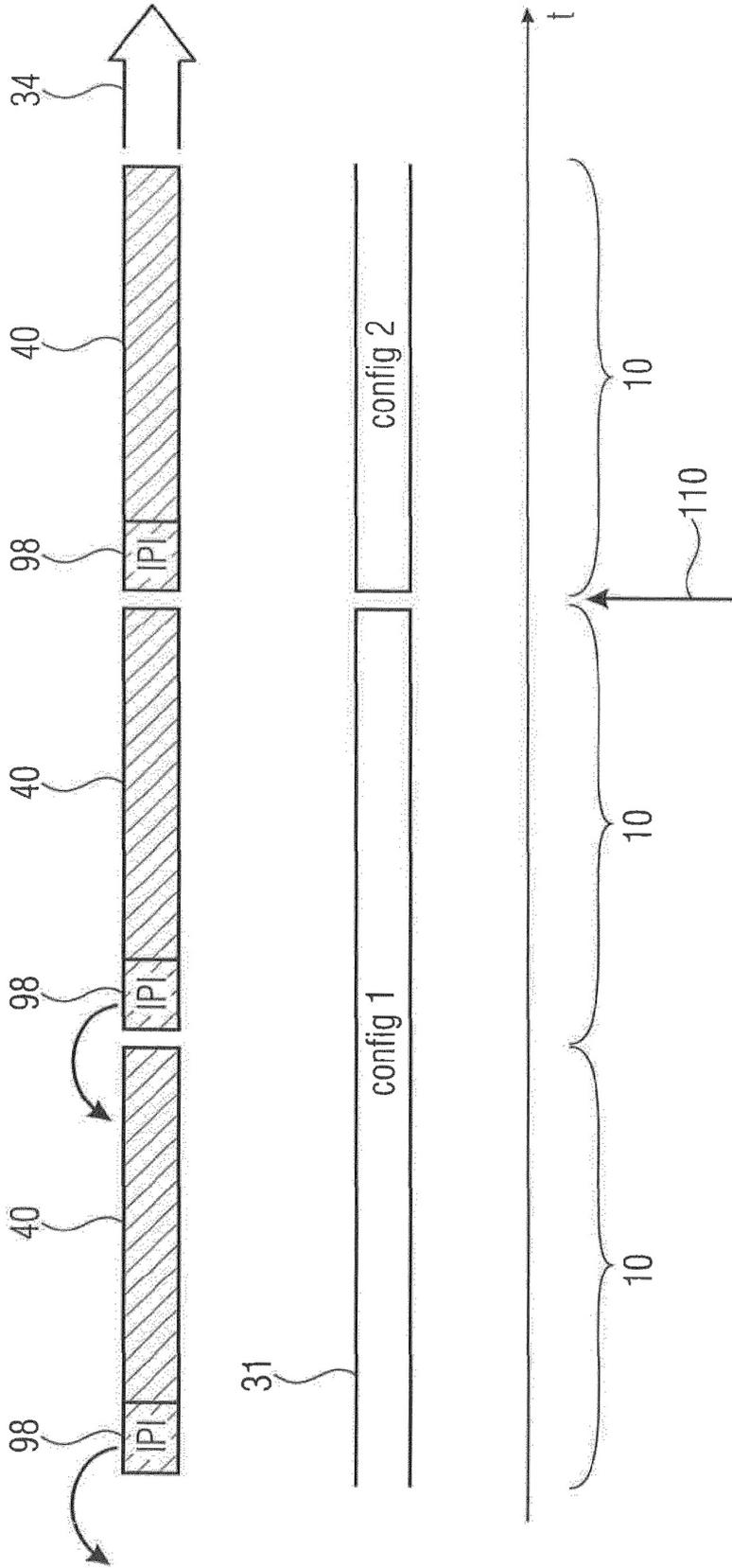


FIG 6



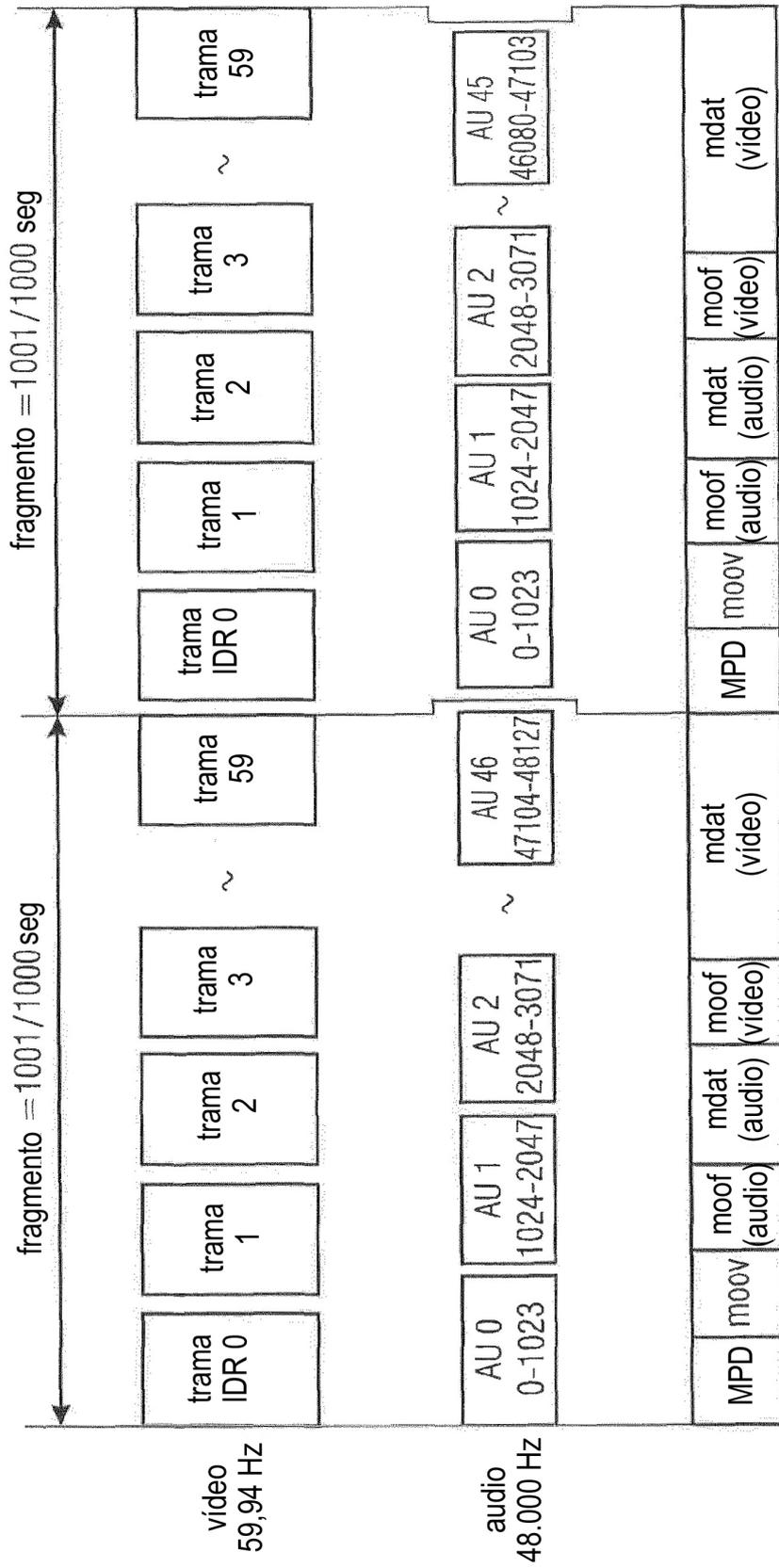


FIG 8