(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 709 274**

(51) Int. Cl.:

G10L 19/16 (2013.01)
G10L 19/022 (2013.01)
H04N 21/233 (2011.01)
H04N 21/234 (2011.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **23.10.2015 PCT/EP2015/074623**

(87) Fecha y número de publicación internacional: **28.04.2016 WO16062869**

(96) Fecha de presentación y número de la solicitud europea: **23.10.2015 E 15784396 (2)**

(97) Fecha y número de publicación de la concesión europea: **05.12.2018 EP 3210206**

(54) Título: **Codificación y decodificación de señales de audio**

(30) Prioridad:

24.10.2014 US 201462068187 P

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
15.04.2019

(73) Titular/es:

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

(72) Inventor/es:

KJOERLING, KRISTOFER;
GROESCHEL, ALEXANDER;
PURNHAGEN, HEIKO;
HOERICH, HOLGER y
KRAUSS, KURT

(74) Agente/Representante:

ELZABURU, S.L.P

ES 2 709 274 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 y decodificación de señales de audio

Referencia cruzada a solicitudes relacionadas

Esta solicitud reivindica prioridad a la Solicitud de Patente Provisional de EE.UU. No. 62/068.187, archivada el 24 de octubre del 2014.

Campo técnico

La invención descrita en este documento se relaciona de manera general con codificación y decodificación de señales de audio, y en particular con un formato de flujo de bits de audio con comportamiento de escalado ventajoso para altas tasas de tramas.

Antecedentes

Las tasas de tramas de audio y video (o frecuencias de tramas) usadas en la mayoría de aplicaciones comerciales disponibles hoy siguen estándares de industria establecidos separados manifestándose en grabar y reproducir productos de software, componentes de hardware así como formatos acordados para transmitir audio y video entre partes comunicantes. Las tasas de tramas de audio son típicamente específicas para diferentes algoritmos de codificación y asociadas con frecuencias de muestreo de audio específicas, tales como 44,1 y 48 kHz, que son tan notorias como las tasas de tramas de video 29,97 fps (NTSC) y 25 fps (PAL) en sus áreas geográficas respectivas; más tasas de tramas de video estándar incluyen 23,98, 24 y 30 fps, o en una forma más generalizada 24, 25, 30 fps y (24, 25, 30) x 1000/1001 fps. Intentos para unificar o armonizar tasas de tramas de audio todavía no han tenido éxito a pesar del cambio de distribución de analógico a digital, que implica que una trama de audio (por ejemplo, un paquete o una unidad de codificación adecuada para transmisión sobre una red) en general no se corresponde con un número entero de tramas de video en un flujo de datos audiovisuales.

La necesidad de sincronizar flujos de datos audiovisuales surge de manera repetida, como resultado de derivas de reloj o cuando varios flujos son recibidos desde diferentes fuentes para procesamiento común, edición o unión en un servidor, una situación encontrada de manera frecuente en estaciones de difusión. Un intento para mejorar la sincronización video-a-video entre dos flujos de datos audiovisuales mediante la duplicación o descarte de tramas de video en uno de los flujos (por ejemplo para preparar los flujos para unión) típicamente lleva a un retraso audio-a-video dentro de ese flujo de datos audiovisuales en el caso de que los tamaños de las tramas de audio y las tramas de video no coincidan. En general, un retraso persiste – al menos de una duración distinta a cero – aun si las tramas de audio correspondientes al video en edición son eliminadas o duplicadas.

Al coste de más procesamiento, un espacio más grande para maniobrar podría ser creado mediante la decodificación temporal del audio durante la sincronización en un formato de bajo nivel que sea independiente de la división en tramas, por ejemplo, formato de banda base, o modulación por codificación de pulsos (PCM) resuelta en la frecuencia de muestreo original. Tal decodificación sin embargo difumina el ancla exacta de los metadatos para especificar segmentos de audio y crea una pérdida de información que no puede ser remediada por la decodificación en un formato intermedio “perfecto”. Como un ejemplo, el control de rango dinámico (DRC) es normalmente dependiente del modo y dependiente del equipo, y puede por lo tanto ser consumido solo en el momento de reproducción real; una estructura de datos que gobierna las características del DRC a través de un paquete de audio es difícil de restaurar de manera fiel después de que la sincronización haya tenido lugar. Por lo tanto, la tarea de preservar los metadatos de este tipo pasada las consecutivas etapas de decodificación, sincronización y codificación no es una tarea simple si está sujeto a restricciones de complejidad.

Dificultades aun más serías pueden surgir en conexión con la infraestructura heredada que es diseñada para llevar señales PCM de dos canales y es por lo tanto capaz de manejar contenido multicanal solo en forma codificada.

Ciertamente es más conveniente codificar tramas de audio y video de manera síncrona en el sentido de que los datos en una trama dada se corresponden exactamente con el mismo segmento de tiempo en la señal audiovisual grabada y codificada. Esto preserva sincronización de audio-a-video bajo manipulación de trama de un flujo audiovisual, esto es, duplicación o rechazo de una o más unidades de codificación enteramente independientes en el flujo. Las longitudes de tramas disponibles en el formato de audio Dolby E™ coinciden con las longitudes de tramas de video. Con una tasa de bit típica de 448 kbps, sin embargo, este formato fue diseñado principalmente con el propósito de producción profesional, con medios duros tales como videocasetes digitales como su modalidad de almacenamiento preferida.

En la solicitud co-pendiente, solicitud todavía no publicada PCT/EP2014/056848, sistemas y métodos son propuestos que son compatibles con un formato de audio adecuado para propósitos de distribución como parte de un formato audiovisual de trama síncrona.

Existe la necesidad de un formato de audio alternativo adecuado para propósitos de distribución como parte de un formato audiovisual de trama síncrona, con comportamiento de escalado mejorado para altas tasas de tramas.

También existe la necesidad de equipo de codificación y decodificación adecuado par el uso con él.

El Reporte de Búsqueda Internacional emitido en conexión adjunto cita entre otros: LINK B., HAGER T. (DOLBY LABORATORIES), FLAKS J. (MICROSOFT CORPORATION), “RTP Payload Format for AC-3 Audio; rfc4184.txt”, 5. JCT-VC MEETING; 96. MPEG MEETING; 16-3-2011 – 23-3-2011, (en adelante, “Referencia D1”); KIM J ET AL, “FRAME SPLITTING SCHEME FOR ERROR-ROBUST AUDIO STREAMING OVER PACKET-SWITCHING NETWORKS”, IEICE TRANSACTIONS ON COMMUNICATIONS, COMUNICATIONS SOCIETY, TOKYO, JP, (200802), vol. E91B, no. 2, páginas 677 – 680, (en adelante, “Referencia D2”); Patente de Estados Unidos no. 5.936.925, (en adelante, “Referencia D3”); y la publicación de solicitud de patente Internacional no. WO 2014/161990 (en adelante, “Referencia D5”).

La Referencia D1 describe un formato de carga de RTP para transportar datos de audio mediante el uso del estándar de compresión de audio AC-3. El formato de carga de RTP en cuestión incluye soporte para fragmentación de datos.

La Referencia D2 presenta un esquema de división de trama para un flujo de audio robusto a errores sobre redes de conmutación de paquetes. Según el esquema, una trama de audio es dividida en varias subtramas en base a la configuración de la red tal que cada paquete pueda ser decodificado de manera independiente en el receptor.

La Referencia D3 describe un medio de grabación de información que incluye: información de grabación procesada que incluye una pluralidad de unidades de información; una información de control de búsqueda para buscar, fuera de la información de grabación procesada, una unidad de grabación de reproducción que es una unidad que incluye la información de grabación a ser reproducida. La información de control de búsqueda es multiplexada con la información de grabación procesada.

La Referencia D5 describe un sistema de procesamiento de audio que acepta un flujo de bits de audio que tiene una entre una pluralidad de tasas de tramas de audio predefinidas. El sistema comprende un componente de interfaz, que recibe un número variable de componentes espectrales cuantificadas, correspondientes a una trama de audio en cualquiera de las tasas de tramas de audio predefinidas, y realiza una cuantificación inversa según los niveles de cuantificación dependientes de la frecuencia predeterminados. Mediante su adaptabilidad de tasa de trama, el sistema puede configurarse para operar de manera síncrona a la trama en paralelo con un sistema de procesamiento de video que acepta tasas de tramas de video plurales.

Breve descripción de los dibujos

En lo que sigue, realizaciones ejemplares serán descritas en mayor detalle y con referencia a los dibujos que acompaña, en los cuales:

La Fig. 1 es un diagrama de bloques generalizado de un sistema de procesamiento de audio para representar una señal de audio como un flujo de bits de audio, según una realización ejemplar;

La Fig. 2 es un diagrama de flujo de bits de un método para representar una señal de audio como un flujo de audio, según una realización ejemplar;

Las Fig. 3 y 4 ilustran ejemplos de flujos de bits de audio proporcionados por el sistema de procesamiento de audio representado en la Fig. 1, según realizaciones ejemplares;

La Fig. 5 es un diagrama de bloques generalizado de un sistema de procesamiento de audio para reconstruir una señal de audio representada por un flujo de bits, según una realización ejemplar;

La Fig. 6 es un diagrama de flujo de un método para reconstruir una señal de audio representada por un flujo de bits, según una realización ejemplar; y

La Fig. 7 es un diagrama de bloques generalizado de un sistema de procesamiento de audio para transcodificar un flujo de bits de audio que representa una señal de audio, según una realización ejemplar.

Todas las figuras son esquemáticas y generalmente solo muestran partes que son necesarias para dilucidar la invención, mientras que otras partes pueden ser omitidas o meramente sugeridas.

Descripción de realizaciones ejemplares

Como se usa en este documento, una señal de audio puede ser una señal de audio independiente, una parte de audio de una señal audiovisual o señal multimedia o cualquiera de estas en combinación con metadatos.

I. Visión general – Lado del codificador

Según un primer aspecto, realizaciones ejemplares proponen sistemas de procesamiento de audio, métodos y productos de programas informáticos para representar una señal de audio como un flujo de bits de audio. Los sistemas, métodos y productos de programas informáticos propuestos, según el primer aspecto, pueden

generalmente compartir las mismas características y ventajas.

Según realizaciones ejemplares, se proporciona un método para representar una señal de audio como un flujo de bits de audio. El método comprende: codificar un segmento de la señal de audio como un conjunto decodificable de datos de audio mediante al menos la realización de análisis de señal, en el segmento de la señal de audio, con un paso, referido en este documento como un paso básico, correspondiente a un primer número de muestras de la señal de audio. El conjunto decodificable de datos de audio se corresponde a una primera tasa de tramas y a un primer número de muestras de la señal de audio por trama. El método comprende: romper el conjunto decodificable de datos de audio en N partes, donde N ≥ 2; y realizar N tramas del flujo de bits que llevan las partes respectivas. Las tramas del flujo de bits tienen una segunda tasa de tramas correspondiente a un segundo número de muestras de la señal de audio por trama del flujo de bits. El primer número de muestras es N veces el segundo número de muestras. El método comprende: sacar un flujo de bits segmentado en tramas del flujo de bits que incluyen las N tramas del flujo de bits.

En un flujo de datos audiovisuales, las tramas de audio y tramas de video pueden estar sincronizadas y pueden ser de igual duración, por ejemplo para facilitar el descarte de tramas o duplicación de tramas en conexión con la unión o compensación de deriva del reloj. Para mantener la sincronía audio-video en un flujo de datos audiovisuales para tasas de tramas de video más altas, la tasa de tramas de audio puede también ser aumentada. Sin embargo, mientras que la codificación predictiva es usada normalmente para reducir el coste de la tasa de bits de aumentar la tasa de tramas de video, la codificación predictiva puede ser menos efectiva para tramas de audio dado que el contenido de audio puede ser variable o en una escala de tiempo más corta, y puede asociarse con un grado más bajo de correlación entre tramas consecutivas que el contenido de video. Con el propósito de esta descripción, a menos que se indique lo contrario, una trama de video se corresponde con una imagen de pantalla completa (por ejemplo una imagen fija en una secuencia), mientras que una trama de audio puede en principio llevar datos de audio correspondientes a un segmento de señal de audio que tiene una duración.

La habilidad del presente método, para proporcionar N tramas del flujo de bits de la segunda tasa de tramas (más alta), junto a llevar un conjunto decodificable de datos de audio asociados con la primera tasa de tramas (más baja), permite mantener sincronización audiovisual para tasas de tramas de video más altas, pero sin un aumento correspondiente en consumo de tasa de bits. De manera más precisa, operar en tasa de tramas aumentada según el presente método normalmente resulta en una tasa de bits que es más baja que la requerida cuando se usan tramas de audio normales que tiene tales tasas de tramas más altas. El presente método puede por lo tanto por ejemplo facilitar la unión de flujos de datos audiovisuales y/o facilitar la compensación de la deriva del reloj.

En efecto, el conjunto decodificable de datos de audio puede corresponderse con la cantidad de datos llevada por una trama de audio normal de la primera tasa de tramas (más baja), y aun si las N tramas del flujo de bits pueden necesitar contener datos no de carga adicionales necesarios para cumplir con un formato de trama (vea a continuación), la cantidad total de datos transmitidos desde el lado del codificador al lado del decodificador puede reducirse al compararla con el empleo de tramas de audio normales que tienen la segunda tasa de tramas (más alta). En particular, realizar el análisis de señal con el paso básico, en vez de con un paso más corto (por ejemplo, correspondiente al segundo número de muestras de la señal de audio), reduce la cantidad de datos requerida para sintetizar la señal de audio otra vez en un lado del decodificador, y por lo tanto reduce la tasa de bits requerida para transmitir los datos al lado del decodificador.

Unir los flujos de bits de audio con otros flujos de bits puede por ejemplo ser realizado sin tener en cuenta los datos de audio llevado por las tramas del flujo de bits. En otras palabras, el dispositivo o unidad que realiza la unión no necesita ser consiente del hecho de que todas las N tramas del flujo de bits pueden requerir reconstruir el segmento de la señal de audio, y puede por ejemplo tratar las tramas del flujo de bits como si fueran decodificables de manera independiente. Perder tramas del flujo de bits de manera potencial en un flujo de bits unido puede por ejemplo ser tratado en el lado del decodificador mediante la ocultación de tales tramas del flujo de bits que pueden no permitir una decodificación exitosa.

Mediante un conjunto decodificable de datos de audio se entiende un conjunto de datos de audio que es suficiente para decodificar un segmento de la señal de audio. El conjunto decodificable de datos de audio puede estar completo en el sentido de que la decodificación de un segmento de la señal de audio puede realizarse sin datos adicionales relacionados al segmento de la señal de audio (sin datos de no carga tales como bits de sobrecarga, cabeceras o preámbulos pueden por ejemplo usarse para identificar el conjunto decodificable de datos de audio en el lado del decodificador).

Mediante la realización de análisis de señal con un paso básico correspondiente al primer número de muestras de la señal de audio se entiende que el análisis de señal es realizado dentro de una ventana de análisis de un cierto número de muestras de la señal de audio, y que, cuando el siguiente segmento de la señal de audio ha de ser codificado, la ventana de análisis es desplazada por el mismo número de muestras que el paso básico. El análisis de señal puede por ejemplo ser realizado con ventanas de análisis solapadas, en cuyo caso la ventana de análisis puede ser más larga que el paso básico. En otro ejemplo, la longitud de la ventana de análisis puede coincidir con el paso básico.

Se apreciará que si la señal de audio es una señal multicanal, entonces el paso básico se puede corresponder con el primer número de muestras de la señal de audio en una base por canal, más que una suma de muestras para los canales respectivos.

El paso de codificar un segmento de la señal de audio puede por ejemplo incluir múltiples subpasos, uno o más de los cuales puede incluir análisis de señal con el paso básico.

El conjunto decodificable de datos de audio puede representar un segmento de la señal de audio correspondiente al primer número de muestras de la señal de audio. El conjunto decodificable de datos de audio puede corresponderse a una trama que tiene la primera tasa de tramas.

Romper el conjunto decodificable de datos de audio puede por ejemplo incluir dividir el conjunto decodificable de datos en al menos N partes de tamaño aproximadamente igual, por ejemplo que comprende al menos aproximadamente el mismo número de bits.

Cada una de las N partes puede ser un conjunto incompleto de datos de audio en el sentido de que una parte puede ser insuficiente para decodificar un segmento (o un subsegmento) de la señal de audio, sin acceder a las otras partes.

Para cada una de las N tramas del flujo de bits, las N tramas del flujo de bits pueden por ejemplo ser la colección más pequeña de tramas del flujo de bits que incluyen esa trama del flujo de bits y desde la cual los datos de audio pueden unirse para decodificar un segmento de la señal de audio representado por los datos llevados por esa trama del flujo de bits. En otras palabras, las N tramas del flujo de bits pueden ser las que llevan datos contenidos originalmente en el mismo conjunto decodificable de datos de audio.

Las tramas del flujo de bits correspondientes a la segunda tasa de tramas (más alta) en el sentido de que las N tramas del flujo de bits juntas representan el mismo segmento de la señal de audio que el conjunto decodificable de datos de audio que se corresponde con la primera tasa de tramas (más baja).

De manera similar, las tramas del flujo de bits correspondientes al segundo número (más pequeño) de muestras por trama del flujo de bits en el sentido de que las N tramas del flujo de bits juntas representan el primer número (más alto) de muestras también representadas por el conjunto decodificable de datos de audio.

Se apreciará que las tramas del flujo de bits pueden por ejemplo llevar partes respectivas de una representación espectral del segmento de la señal de audio, y que puede no haber conexión entre una de las tramas del flujo de bits y un segundo número (más pequeño) de muestras de la señal de audio.

Las N tramas del flujo de bits pueden por ejemplo ser conformes a un formato de audio en el sentido de que las tramas del flujo de bits pueden llevar cargas y metadatos que, a un nivel de flujo elemental, están conformes con un formato de audio, por ejemplo como el proporcionado en los flujos elementales del Grupo de Expertos en Imágenes en Movimiento (MPEG). Se apreciará que, aunque conformes con un formato de audio en este sentido, la carga y al menos algunos de los metadatos llevados por las tramas del flujo de bits pueden por ejemplo ser de un tipo y/o formato diferente que en tramas de audio conocidas en la técnica.

Las N tramas del flujo de bits que llevan las N partes pueden por ejemplo ser sacadas como N tramas del flujo de bits consecutivas en el flujo de bits.

En una realización ejemplar, realizar el análisis de señal puede incluir realizar, con el paso básico: análisis espectral; análisis de energía; y/o análisis de entropía. El análisis espectral con el paso básico puede por ejemplo ser realizado para transformar el segmento de la señal de audio desde un dominio del tiempo a un dominio de la frecuencia. El análisis de energía con el paso básico puede por ejemplo ser realizado para codificar el segmento de la señal de audio con una técnica de codificación basada en la energía. El análisis de entropía con el paso básico puede por ejemplo ser realizado para codificar la señal de audio con una técnica de codificación basada en la entropía.

En una realización ejemplar, codificar un segmento de la señal de audio puede incluir: aplicar una transformada de ventana que tenga el paso básico como el paso de transformación; y/o calcular una señal de mezcla descendente y parámetros para reconstrucción paramétrica de la señal de audio a partir de la señal de mezcla descendente, donde los parámetros son calculados en base al análisis de señal.

La transformada de ventana puede por ejemplo ser una transformada armónica, tal como la Transformada de Coseno Discreta Modificada (MDCT), por ejemplo mediante el empleo de ventanas de transformación que se solapan.

La señal de audio puede por ejemplo ser una señal de audio multicanal y la señal de mezcla descendente puede ser una señal con menos canales que la señal multicanal, por ejemplo obtenidos como una combinación lineal de los canales de la señal multicanal. La señal de mezcla descendente puede por ejemplo ser una mezcla descendente mono o estéreo de la señal de audio multicanal.

En una realización ejemplar, el método puede comprender: incluir metadatos en al menos una de las N tramas del

flujo de bits que llevan las partes. Los metadatos pueden indicar que el conjunto decodificable completo de datos de audio es obtenible a partir de las partes llevadas por las N tramas del flujo de bits.

Cada una de las N tramas del flujo de bits por ejemplo llevan metadatos que las identifican como pertenecientes a un grupo de N tramas del flujo de bits a partir de las cuales el conjunto decodificable de datos de audio es obtenible. En otro ejemplo, una de las tramas del flujo de bits puede llevar metadatos que identifican todas las N tramas del flujo de bits, mientras que las otras N-1 tramas del flujo de bits del grupo no llevan necesariamente tales metadatos. El flujo de bits puede por ejemplo incluir otras tramas del flujo de bits que no llevan tales metadatos.

Los metadatos pueden permitir que las N tramas del flujo de bits sean ubicadas en posiciones no predeterminadas relacionadas entre ellas. Los metadatos pueden permitir para otras tramas del flujo de bits entre las N tramas del flujo de bits. Los metadatos pueden permitir detectar cuando una o más de las N tramas del flujo de bits faltan en el flujo de bits, por ejemplo debido a unión o a un descarte de trama.

En una realización ejemplar, el flujo de bits de audio puede estar asociado con un flujo de las tramas de video. El método puede además comprender: en respuesta al flujo de tramas de video que comprende una trama de video de un cierto tipo, codificar un segmento de la señal de audio temporalmente relacionada con la trama de video como un segundo conjunto decodificable de datos de audio mediante al menos la realización de análisis de señal, en el segmento de la señal de audio temporalmente relacionada con la trama de video, con un paso acortado correspondiente al segundo número de muestras de la señal de audio. El segundo conjunto decodificable de datos de audio puede corresponderse con la segunda tasa de tramas y con el segundo número de muestras de la señal de audio por trama. El método puede comprender: incluir una trama del flujo de bits que lleve el segundo conjunto decodificable de datos de audio en el flujo de bits.

Flujos de tramas de video pueden por ejemplo ser unidas en puntos adyacentes a tramas de un cierto tipo, tal como tramas de video codificado de manera independiente, para facilitar la decodificación de la secuencia unida de tramas de video en el lado del decodificador. El enfoque de codificar un segmento de la señal de audio temporalmente relacionada con la trama de video de dicho cierto tipo como un segundo conjunto decodificable de datos de audio correspondiente a la segunda tasa de tramas, y de incluir una trama del flujo de bits que lleve el segundo conjunto decodificable de datos de audio en el flujo de bits, permite la decodificación independiente de ese segmento de la señal de audio en el lado del decodificador. La presente realización ejemplar puede por lo tanto facilitar la decodificación de ese segmente de la señal de audio en el caso de que tramas del flujo de bits precedentes o sucesivas del flujo de bits de audio puedan faltar en el lado del decodificador, por ejemplo como resultado de la unión de un flujo audiovisual de datos que comprende el flujo de bits de audio y el flujo de tramas de video, con uno o más flujos de datos audiovisuales.

El segmento de la señal de audio temporalmente relacionado con la trama de video de cierto tipo puede por ejemplo corresponderse con un punto en el tiempo en el cual la trama de video de cierto tipo está destinada a ser reproducida en un elemento de presentación.

El flujo de tramas de video puede por ejemplo incluir tramas codificadas de manera independiente y tramas codificadas predichas (que tienen bien una dependencia unidireccional o bidireccional en tramas adyacentes), y el cierto tipo de trama de video puede por ejemplo ser una trama de video codificada de manera independiente.

El método puede por ejemplo comprender: detectar la presencia de cierto tipo de trama de video en el flujo de tramas de video. La presencia de cierto tipo de trama de video puede por ejemplo ser detectada a través de la señalización desde un codificador de video.

Realizar el análisis de señal con el paso acortado puede por ejemplo incluir realizar, con el paso acortado: análisis espectral; análisis de energía; y/o análisis de entropía.

Codificar un segmento de la señal de audio temporalmente relacionada con la trama de video de cierto tipo puede por ejemplo incluir: aplicar una transformada de ventana que tiene el paso acortado como el paso de transformación; y/o calcular una señal de mezcla descendente y parámetros para reconstrucción paramétrica de la señal de audio a partir de la señal de mezcla descendente, donde los parámetros son calculados en base al análisis de señal con el paso acortado.

En una realización ejemplar, el método puede comprender: en respuesta al flujo de tramas de video que comprenden una trama de video de cierto tipo, codificar N segmentos consecutivos de la señal de audio como conjuntos decodificables respectivos de datos de audio mediante, para cada uno de los N segmentos consecutivos, al menos aplicar análisis de señal con el paso acortado. El segmento temporalmente relacionado con la trama de video puede ser uno de los N segmentos consecutivos. El método puede incluir: incluir tramas del flujo de bits que lleven los conjuntos decodificables respectivos de datos de audio asociados con los N segmentos consecutivos en el flujo de bits.

El flujo de bits puede por ejemplo comprender grupos de N tramas del flujo de bits consecutivas que llevan partes respectivas de datos de audio que son decodificables juntas. Por lo tanto, en el lado del decodificador, N tramas del flujo de bits del flujo de bits pueden ser decodificadas a la vez. En la presente realización ejemplar, la estructura de

grupos de N tramas del flujo de bits puede ser preservada también cuando dicho cierto tipo de trama de video ocurra en el flujo asociado de tramas de video, por ejemplo independientemente de la posición en el flujo de tramas de video de la trama de video de dicho cierto tipo relacionado con las posiciones de los grupos de N tramas del flujo de bits consecutivas en el flujo de bits.

Según realizaciones ejemplares, se proporciona un sistema de procesamiento de audio para representar una señal de audio mediante un flujo de bits de audio. El sistema de procesamiento de audio comprende: una sección codificadora configurada para codificar un segmento de la señal de audio como un conjunto decodificable de datos de audio mediante al menos la realización de análisis de señal, en el segmento de audio de señal, con un paso básico correspondiente a un primer número de muestras de la señal de audio. El conjunto decodificable de datos de audio se corresponde con una primera tasa de tramas y con un primer número de muestras de la señal de audio por trama. El sistema de procesamiento de audio comprende una sección de reentramado configurada para: romper el conjunto decodificable de datos de audio en N partes, donde N ≥ 2; y formar N tramas del flujo de bits que llevan las partes respectivas. Las tramas del flujo de bits tienen una segunda tasa de tramas correspondiente a un segundo número de muestras de la señal de audio por trama del flujo de bits. El primer número de muestras es N veces el segundo número de muestras. La sección de reentramado está configurada para sacar un flujo de bits segmentado en tramas del flujo de bits que incluye las N tramas del flujo de bits formadas.

Según realizaciones ejemplares, se proporcionan productos de programas informáticos que comprenden un medio legible por un ordenador para realizar cualquiera de los métodos del primer aspecto.

Según realizaciones ejemplares, se puede mantener que N = 2 o N = 4, esto es las N tramas del flujo de bits pueden ser dos o cuatro tramas del flujo de bits.

II. Visión general – Lado del decodificador

Según un segundo aspecto, realizaciones ejemplares proponen sistemas de procesamiento de audio así como métodos y productos de programas informáticos para reconstruir una señal de audio representada por un flujo de bits. Los sistemas, métodos y productos de programas propuestos, según el segundo aspecto, pueden generalmente compartir las mismas características y ventajas. Además, las ventajas presentadas anteriormente para características de sistemas, métodos y productos de programas informáticos, según el primer aspecto, pueden generalmente ser válidas para las características de sistemas, métodos y productos de programas informáticos correspondientes según el segundo aspecto.

Según realizaciones ejemplares, se proporciona un método para reconstruir una señal de audio representada por un flujo de bits segmentado en tramas del flujo de bits. El método comprende: unir conjuntos de datos de audio llevados por N tramas del flujo de bits respectivas en un conjunto decodificable de datos de audio correspondiente a una primera tasa de tramas y a un primer número de muestras de la señal de audio por trama, donde N ≥ 2. Las tramas del flujo de bits tienen una segunda tasa de tramas correspondiente a un segundo número de muestras de la señal de audio por trama del flujo de bits. El primer número de muestras es N veces el segundo número de muestras. El método comprende: decodificar el conjunto decodificable de datos de audio en un segmento de la señal de audio mediante al menos el empleo de síntesis de señal, en base al conjunto decodificable de datos, con un paso, referido como en este documento un paso básico, correspondiente al primer número de muestras de la señal de audio.

En un flujo de datos audiovisuales, las tramas de audio y tramas de video pueden sincronizarse y pueden ser de igual duración, por ejemplo, para facilitar el descarte de tramas o duplicación de tramas en conexión con la unión o compensación de derivas del reloj. Para mantener sincronización de audio-video en un flujo de datos audiovisuales para tasas de tramas de video más altas, la tasa de tramas de video puede también ser aumentada. Sin embargo, mientras que la codificación predictiva es empleada normalmente para reducir el coste de la tasa de bits de aumentar la tasa de tramas de video, la codificación predictiva puede ser menos eficiente para tramas de audio ya que el contenido del audio puede ser variable en una escala de tiempos más corta, y puede estar asociado con un grado más bajo de correlación entre tramas consecutivas que el contenido del video. Una longitud de trama de audio demasiado corta debería ser evitada dado que puede limitar el paso de transformación, que a su vez establece un límite en la resolución de frecuencia.

La habilidad del presente método, para unir conjuntos de datos de audio llevados por N tramas del flujo de bits respectivas de la segunda tasa de tramas (más alta) en un conjunto decodificable de datos de audio asociado con la primera tasa de tramas (más baja), permite mantener sincronización audiovisual para tasas de tramas de video más altas, pero sin un aumento correspondiente en el consumo de la tasa de bits. De manera más precisa, la tasa de bits cuando se opera en la tasa de tramas aumentada según el presente método puede ser más baja que la requerida cuando se usan tramas de audio normales que tienen tales tasas de tramas más altas. El presente método puede por ejemplo facilitar la unión de tramas de datos de audio-video y/o facilitar la compensación de la deriva del reloj.

En particular, usar la síntesis de señal con el paso básico, en vez de una síntesis con un paso más corto (por ejemplo correspondiente al segundo número de muestras de la señal de audio), reduce la cantidad de datos requeridos para sintetizar la señal de audio, y de este modo reduce la tasa de bits requerida para transmitir los datos.

Cada uno de los conjuntos de datos unidos en el conjunto decodificable de datos de audio puede ser un conjunto

incompleto de datos de audio en el sentido de que uno de los conjuntos puede ser insuficiente para decodificar un segmento (o subsegmento) de la señal de audio, sin acceso a los otros conjuntos.

Para cada una de las N tramas del flujo de bits, las N tramas del flujo de bits pueden por ejemplo ser la colección más pequeña de tramas del flujo de bits que incluye esa trama del flujo de bits y a partir de la cual los datos de audio pueden ser unidos para decodificar un segmento de la señal de audio representada por los datos llevados por esa trama del flujo de bits.

Mediante un conjunto decodificable de datos de audio se entiende un conjunto de datos de audio que es suficiente para decodificar un segmento de la señal de audio. El conjunto decodificable de datos de audio puede estar completo en el sentido de que la decodificación de un segmento de la señal de audio puede ser realizada sin datos de audio adicionales.

Unir los conjuntos de datos de audio en el conjunto decodificable de datos de audio puede por ejemplo incluir concatenar los conjuntos de datos, por ejemplo mediante la disposición de los bits que representan los conjuntos respectivos de datos uno detrás del otro.

Mediante el uso de la síntesis de señal con un paso básico correspondiente al primer número de muestras de la señal de audio se entiende que la síntesis de señal se realiza para un segmento de la señal de audio correspondiente a un cierto número de muestras de la señal de audio, y que, cuando el siguiente segmento de la señal de audio va a ser reconstruido, el proceso de síntesis de señal produce una salida para un intervalo que ha sido desplazado por el mismo número de muestras que el paso básico.

La síntesis de señal con el paso básico puede por ejemplo usarse directamente en base al conjunto decodificado de datos de audio, o puede usarse indirectamente en base al conjunto decodificable de datos de audio, por ejemplo en base a datos de audio o señales obtenidas mediante el procesamiento del conjunto decodificable de datos de audio.

Se apreciará que si la señal de audio es una señal multicanal, entonces el paso básico puede corresponderse con el primer número de muestras de la señal de audio en una base por canal, más que una suma de muestras para los canales respectivos.

El paso de decodificar el conjunto decodificable de datos de audio puede por ejemplo incluir múltiples subpasos, uno o más de los cuales pueden incluir síntesis de señal con el paso básico.

Las N tramas del flujo de bits pueden por ejemplo ser conformes a un formato de audio en el sentido de que las tramas del flujo de bits pueden llevar cargas y metadatos que, a un nivel de flujo elemental, están conformes con un formato de audio, por ejemplo como el proporcionado en los flujos elementales del Grupo de Expertos en Imágenes en Movimiento (MPEG). Se apreciará que, aunque conformes con un formato de audio en este sentido, la carga y al menos algunos de los metadatos llevados por las tramas del flujo de bits pueden por ejemplo ser de un tipo y/o formato diferente que en tramas de audio conocidas en la técnica.

Un flujo de bits proporcionado por un codificador puede por ejemplo haber sido unido con otro flujo de bits antes de alcanzar el lado del decodificador. Por ejemplo, una o más de las N tramas del flujo de bits pueden por ejemplo faltar en el flujo de bits recibido en el lado del decodificador. En algunas realizaciones ejemplares, el método de procesamiento del audio puede por lo tanto incluir detectar si una o más de las N tramas del flujo de bits a partir de las cuales unir los conjuntos de datos de audio en el conjunto decodificable completo faltan en el flujo de bits. El método puede por ejemplo comprender: en respuesta a la detección de que una o más de las N tramas del flujo de bits faltan en el flujo de bits, aplicar ocultación del error. La ocultación del error puede por ejemplo incluir reemplazar datos de audio llevados por una o más de las tramas del flujo de bits recibidas por ceros, y de manera opcional, aplicar desaparición y/o reaparición.

En una realización ejemplar, decodificar el conjunto decodificable de datos de audio puede incluir: aplicar una transformada de ventana que tenga el paso básico como el paso de transformación; y/o realizar reconstrucción paramétrica, con el paso básico, del segmento de la señal de audio, en base a una señal de mezcla descendente y asociar parámetros obtenidos a partir del conjunto decodificable de datos de audio.

La transformada de ventana puede por ejemplo ser una transformada armónica, tal como la Transformada de Coseno Discreta Modificada (MDCT) inversa.

La señal de audio puede por ejemplo ser una señal de audio multicanal y la señal de mezcla descendente puede ser una señal con menos canales que la señal multicanal, por ejemplo obtenida como una combinación lineal de los canales de la señal multicanal. La señal de mezcla descendente puede por ejemplo ser una mezcla descendente mono o estéreo de la señal de audio multicanal. El conjunto decodificable de datos de audio puede por ejemplo comprender la señal de mezcla descendente y los parámetros asociados para la reconstrucción paramétrica del segmento de la señal de audio. De manera alternativa, el conjunto decodificable de datos de audio puede comprender datos que representan la señal de mezcla descendente y los parámetros asociados, por ejemplo, en forma cuantificada, a partir de los cuales la señal de mezcla descendente y los parámetros asociados pueden ser derivados.

En una realización ejemplar, las N tramas del flujo de bits, a partir de las cuales los conjuntos de datos de audio son unidos en el conjunto decodificable de datos de audio, pueden ser N tramas del flujo de bits consecutivas. Usar tramas consecutivas para llevar los conjuntos de datos de audio unibles en un conjunto decodificable de datos de audio puede facilitar la decodificación de la señal de audio y puede reducir la necesidad de metadatos para identificar las tramas del flujo de bits para las cuales los datos han de ser unidos en un conjunto decodificable de datos de audio. Usar tramas consecutivas para llevar los conjuntos de datos de audio unibles en el conjunto decodificable de datos de audio puede reducir la necesidad de almacenamiento intermedio de datos para realizar la decodificación.

En una realización ejemplar, el método puede además comprender: determinar, en base a los metadatos llevados por al menos algunas de las tramas del flujo de bits en el flujo de bits, un grupo de tramas del flujo de bits a partir de las cuales unir los conjuntos incompletos de datos de audio en el conjunto decodificable de datos de audio. Los metadatos pueden por ejemplo ser llevados por todas las tramas del flujo de bits, o por una o más tramas del flujo de bits por grupo de N tramas del flujo de bits, para identificar los grupos de N tramas del flujo de bits. Realizaciones pueden también ser concebidas en las cuales el flujo de bits comprende otras tramas que llevan metadatos que identifican los grupos de N tramas, mientras que las N tramas del flujo de bis mismas pueden no llevar tales metadatos.

En una realización ejemplar, el método puede además comprender: detectar si una trama del flujo de bits lleva un conjunto decodificable de datos de audio correspondiente a la segunda tasa de tramas; y decodificar el conjunto decodificable de datos de audio correspondiente a la segunda tasa de tramas en un segmento de la señal de audio mediante al menos el uso de síntesis de señal, en base al conjunto decodificable de datos de audio correspondiente a la segunda tasa de tramas, con un paso acortado correspondiente al segundo número de muestras.

Las tramas del flujo de bits que llevan conjuntos de datos de audio que son decodificables de manera independiente pueden usarse por ejemplo para facilitar la decodificación de flujos de bits tras unir y/o tras descartar/duplicar tramas. La habilidad el método en la presente realización ejemplar de decodificar mediante el uso del paso acortado puede hacerla compatible con un formato del flujo de bits que facilite la sincronización de tramas de audio y video.

Decodificar el conjunto decodificable de datos de audio correspondiente a la segunda tasa de tramas puede por ejemplo incluir: aplicar una transformación de ventana que tiene el paso acortado como paso de transformación; y/o realizar reconstrucción paramétrica, con el paso acortado, de un segmento de la señal de audio, en base a una señal de mezcla descendente y parámetros asociados obtenidos a partir del segundo conjunto decodificable de datos de audio.

La detección de si un flujo de bits lleva un conjunto decodificable de datos de audio correspondiente a la segunda tasa de tramas puede por ejemplo ser en base a los metadatos llevados por la trama del flujo de bits, o en base a una ausencia de un tipo particular de metadatos en la trama del flujo de bits.

En una realización ejemplar, decodificar un conjunto decodificable de datos de audio correspondiente a la segunda tasa de tramas puede incluir: proporcionar un retraso para que la decodificación de un grupo de N tramas del flujo de bits consecutivas en la segunda tasa de tramas se complete al mismo tiempo que si las tramas del flujo de bits del grupo de N tramas del flujo de bits hubieran cada una llevado conjuntos de datos de audio que requieren unirse en un conjunto decodificable de datos de audio. La presente realización ejemplar facilita una transición suave entre segmentos de la señal de audio reconstruidos mediante el uso del paso básico y segmentos de la señal de audio reconstruidos mediante el uso del paso acortado, y puede mejorar la calidad de reproducción percibida por un oyente.

En una realización ejemplar, el retraso puede ser proporcionado mediante el almacenamiento intermedio de al menos un conjunto decodificable de datos de audio correspondientes a la segunda tasa de tramas o mediante el almacenamiento intermedio de al menos un segmento de la señal de audio. Es decir, el retraso puede ser proporcionado antes de realizar la síntesis de señal, mediante el almacenamiento intermedio de uno o más conjuntos decodificables de datos de audio correspondientes a la segunda tasa de tramas, o tras realizar síntesis de señal, mediante el almacenamiento intermedio de uno o más segmentos de la señal de audio reconstruida a partir de uno o más conjuntos decodificables de datos de audio correspondientes a la segunda tasa de tramas.

En una realización ejemplar, el flujo de bits puede asociarse con un flujo de tramas de video que tiene una tasa de tramas que coincide con la segunda tasa de tramas. En la presente realización ejemplar, la tasa de tramas de las tramas del flujo de bits puede coincidir con la tasa de tramas de las tramas de video, lo que puede facilitar la unión y/o sincronización de un flujo de datos audiovisuales que comprende el flujo de bits y el flujo de tramas de video con otros flujos de datos audiovisuales.

En una realización ejemplar, decodificar un segmento de la señal de audio en base a un conjunto decodificable de datos de audio correspondientes a la primera tasa de tramas puede comprender: recibir coeficientes espectrales cuantificados correspondientes al conjunto decodificable de datos de audio correspondientes a la primera tasa de tramas; realizar cuantificación inversa seguida por una transformación frecuencia-a-tiempo, por lo cual una representación de una señal de audio intermedia es obtenida; realizar al menos un paso de procesamiento en el

dominio de la frecuencia de la señal de audio intermedia; y cambiar una tasa de muestreo de la señal de audio procesada en una frecuencia de muestreo objetivo, por lo cual una representación en el dominio del tiempo de una señal de audio reconstruida es obtenida.

La frecuencia de muestreo objetivo puede ser una cantidad predefinida, que es configurable por un usuario o diseñador del sistema de manera independiente a las propiedades (por ejemplo, la tasa de tramas) del flujo de bits entrante.

La cuantificación inversa puede realizarse con niveles de cuantificación predeterminados (o niveles de reconstrucción, o puntos de reconstrucción). Los niveles de cuantificación pueden haber sido escogidos en el lado del codificador en base a consideraciones psicoacústicas, por ejemplo, de tal manera que el ruido de cuantificación para una frecuencia dada (o banda de frecuencias) no excede el umbral de enmascaramiento. Porque el umbral de enmascaramiento es dependiente de la frecuencia, es preferido desde un punto de vista económico que el lado del codificador seleccione niveles de cuantificación que sean no uniformes con respecto a la frecuencia. Como una consecuencia, la cuantificación y descuantificación normalmente tienen lugar con una frecuencia de muestreo física en mente, a la cual la salida óptima es producida.

El al menos un paso de procesamiento puede por ejemplo estar asociado con replicación de banda espectral (SBR) y/o control de rango dinámico (DRC).

Dado que al menos un paso de procesamiento es realizado en el dominio de la frecuencia, el método puede incluir: realizar una transformación tiempo-a-frecuencia, por ejemplo realizada por un banco de filtros de análisis de Filtro de Espejo de Cuadratura (QMF), para obtener una representación de frecuencia de una señal de audio intermedia; y realizar una transformación frecuencia-a-tiempo adicional, por ejemplo realizada por un banco de filtros de síntesis QMF, para transformar la señal de audio procesada de vuelta al dominio del tiempo.

En una realización ejemplar, el método puede aceptar los flujos de bits asociados con al menos dos valores diferentes para la segunda tasa de tramas pero asociados con un valor común para el segundo número de muestras por trama. Los valores respectivos de la segunda tasa de tramas pueden diferir en como mucho el 5%. La transformación de frecuencia a tiempo se puede realizar en un componente funcional configurado para emplear una transformada de ventana que tenga un valor predefinido común para el paso básico como paso de transformada para al menos dos valores diferentes para la segunda tasa de tramas.

En un flujo audiovisual de datos, la tasa de tramas de audio se puede adaptar (por ejemplo puede coincidir con) la tasa de tramas de video, por ejemplo para facilitar la sincronización y/o la unión de audio y video. La capacidad del método en la presente realización ejemplar para aceptar flujos de audio con diferentes tasas de tramas puede facilitar por lo tanto la sincronización y/o la unión de audio y video de los flujos de datos audiovisuales.

En un sistema muestreado de manera crítica, la frecuencia de muestreo física corresponde a la relación de la duración física de una trama de audio con el número de coeficientes espectrales contenidos en la misma. El componente o los componentes espectrales que realizan la cuantificación inversa y la transformación de frecuencia a tiempo no necesitan ser conscientes de la duración física de los coeficientes en un conjunto decodificable de datos de audio, sólo de los coeficientes que pertenecen al mismo conjunto decodificable de datos de audio. Ya que los valores de la segunda tasa de tramas difieren en como mucho un 5%, la frecuencia de muestreo interna resultante cambiará bastante poco (en unidades físicas), y el factor de re muestreo usado en la conversión de la tasa de muestreo final estará cercano a uno. Por lo tanto, la no constancia de la frecuencia de muestreo interna no llevará normalmente a ninguna degradación perceptible de la señal de audio reconstruida. En otras palabras, el ligero sobre o sub muestreo de la señal de audio intermedia, que se produjo para ser óptima a una frecuencia de muestreo ligeramente diferente de la frecuencia de muestreo objetivo, no será psicoacústicamente significativo. En concreto, puede ser tolerable cierta cantidad de desajuste entre la frecuencia de muestreo física esperada y/o la transformación de frecuencia a tiempo, y las frecuencias de muestreo físicas a las que se sintonizan cualesquiera componentes por debajo de la misma, siempre que la desviación sea limitada.

Según las realizaciones ejemplares, se proporciona un sistema de procesamiento de audio para reconstruir una señal de audio representada por un flujo de bits segmentado en tramas del flujo de bits. El sistema de procesamiento de audio comprende: una memoria intermedia configurada para unir los conjuntos de datos de audio llevados por las N tramas respectivas del flujo de bits en un conjunto decodificable de datos de audio correspondiente a una primera trama y a un primer número de muestras de la señal de audio por trama, donde $N \geq 2$. Las tramas del flujo de bits tienen una segunda tasa de tramas correspondiente a un segundo número de muestras de la señal de audio por trama del flujo de bits. El primer número de muestras es N veces el segundo número de muestras. El sistema comprende una sección de decodificación configurada para decodificar el conjunto decodificable de datos de audio en un segmento de la señal de audio mediante al menos el empleo de síntesis de señal, en base al conjunto decodificable de datos de audio, con un paso básico correspondiente al primer número de muestras de la señal de audio.

Según las realizaciones ejemplares, se proporcionan productos de programas informáticos que comprenden un medio legible por ordenador para realizar cualquiera de los métodos del segundo aspecto.

Según las realizaciones de ejemplo, puede suceder que N = 2 o N = 4, esto es, las N tramas del flujo de bits pueden ser dos o cuatro tramas del flujo de bits.

III. Visión general – Transcodificación

Según un tercer aspecto, los ejemplos proponen sistemas de procesamiento de audio así como los métodos y los productos de programas informáticos para transcodificar un flujo de bits que representa una señal de audio. Los sistemas, métodos y productos de programas de audio propuestos, según el tercer aspecto, generalmente pueden compartir las mismas características y ventajas. Además, las ventajas presentadas anteriormente para las características de los sistemas, métodos y productos de programas informáticos, según el primer y/o aspecto, pueden de manera general ser válidos para las características correspondientes de los sistemas, métodos y productos de programas informáticos según el tercer aspecto.

Según los ejemplos, se proporciona un método de transcodificación de un flujo de bits de audio que representa una señal de audio. El flujo de bits comprende una secuencia de conjuntos decodificables de datos de audio correspondientes a una primera tasa de tramas y a un primer número de muestras de la señal de audio por trama. El método comprende: la extracción de un conjunto decodificable de datos de audio desde el flujo de bits, la partición del conjunto decodificable de datos de audio en N partes, donde N ≥ 2; y la creación de N tramas del flujo de bits que lleven las partes respectivas. Las tramas del flujo de bits tienen una segunda tasa de tramas correspondiente a un segundo número de muestras de la señal de audio por trama del flujo de bits. El primer número de muestras es N veces el segundo número de muestras. Tras esto, se emite un flujo de bits segmentado en tramas del flujo de bits que incluyen las N tramas del flujo de bits creadas. De manera opcional, se realiza un paso de procesamiento del conjunto decodificable de datos de audio antes del paso de partición del conjunto hasta en N partes. Dependiendo de la naturaleza del procesamiento, esto puede requerir la decodificación inicial de los datos de audio a una representación de transformada o de forma de onda.

La capacidad del presente método para proporcionar N tramas del flujo de bits de la segunda (superior) tasa de tramas, junto con llevar un conjunto decodificable de datos de audio asociados con la primera (inferior) tasa de trama, permite mantener la sincronización audiovisual para tasas de tramas de video superiores., pero sin el correspondiente aumento en el consumo de la tasa de bits. La tasa de bits cuando opera a tasas de tramas aumentadas según el presente método puede ser inferior que la requerida al usar tramas de audio normales que tengan dichas tasa de tramas superiores. El presente método puede por lo tanto facilitar por ejemplo la unión de los flujos de datos audiovisuales y/o facilitar la compensación de la deriva del reloj.

El método puede por ejemplo incluir la partición de una versión procesada del conjunto decodificable de datos de audio en N partes.

Según los ejemplos, se proporciona un sistema de procesamiento de audio para transcodificar un flujo de bits de audio que representa una señal de audio, en donde el flujo de bits comprende una secuencia de conjuntos decodificables de datos de audio correspondientes a la primera tasa de trama y a un primer número de muestras de la señal de audio por trama. El sistema de procesamiento de audio comprende; un sistema de recepción configurado para extraer un conjunto decodificable de datos de audio del flujo de bits; y una sección de procesamiento opcional configurada para procesar el conjunto decodificable de datos de audio. El sistema de procesamiento de audio comprende una sección de reentramado configurada para: partir el conjunto decodificable de datos de audio en N partes, donde N ≥ 2; y formar N tramas del flujo de bits que lleven las partes respectivas. Las tramas del flujo de bits tienen una segunda tasa de tramas correspondiente a un segundo número de muestras de la señal de audio por tasa del flujo de audio. El primer número de muestras es N veces el segundo número de muestras. La sección de reentramado se configura para emitir un flujo de bits segmentado en tramas del flujo de bits que incluyan las N tramas del flujo de bits creadas.

Según los ejemplos, se proporcionan productos de programa informático que comprenden un medio legible por ordenador para realizar cualquiera de los métodos del tercer aspecto.

Según los ejemplos, puede suceder que N = 2 o N = 4, esto es, las N tramas del flujo de bits pueden ser dos o cuatro tramas del flujo de bits.

III. Visión general – Medio legible por ordenador

Según un cuarto aspecto, las realizaciones ejemplares proponen medios legibles por ordenador que representan una señal de audio. Las ventajas presentadas anteriormente como características de los sistemas, métodos y productos de programas informáticos, según el primer, segundo y/o tercer aspecto, pueden de manera general ser válidos para las características correspondientes de los medios legibles por ordenador según el cuarto aspecto.

Según las realizaciones de ejemplo, se proporciona un medio legible por ordenador que represente una señal de audio segmentada en tramas del flujo de bits. En el medio legible por ordenador, N de las tramas del flujo de bits llevan los respectivos conjuntos de datos de audio que se pueden unir en un conjunto decodificable de datos de audio que corresponde a una primera tasa de tramas y a un primer número de muestras de la señal de audio por trama, donde N ≥ 2, El conjunto decodificable de datos de audio es decodificable en un segmento de la señal de

audio mediante al menos el empleo de síntesis de señal, en base al conjunto decodificable de datos de audio, con un paso básico correspondiente al primer número de muestras de la señal de audio. Las tramas del flujo de bits tienen una segunda tasa de tramas correspondiente a un segundo número de muestras de la señal de audio por trama del flujo de bits. El primer número de muestras es N veces el segundo número de muestras.

Las N tramas del flujo de bits de la segunda (superior) tasa de tramas, junto con llevar un conjunto decodificable de datos de audio asociado con la primera (inferior) tasa de tramas, permite mantener la sincronización audiovisual para tasas de tramas de video superiores, pero sin el correspondiente aumento en el consumo de la tasa de bits. Más precisamente, la tasa de bits al operar a tasas de tramas aumentadas según el presente medio legible por ordenador puede ser inferior que la requerida al usar tramas de audio normales que tengan dichas tasas de tramas superiores. El presente medio legible por ordenador puede por lo tanto por ejemplo facilitar la unión de flujos de datos audiovisuales y/o facilitar la compensación de la deriva del reloj.

Las N tramas del flujo de bits que llevan los respectivos conjuntos de datos de audio que se pueden unir en un conjunto decodificable de datos de audio pueden ser por ejemplo N tramas del flujo de bits consecutivas.

En una realización ejemplar, al menos una de las N tramas del flujo de bits puede llevar metadatos que indiquen un grupo de tramas del flujo de bits desde el cual unir los conjuntos de datos de audio en el conjunto decodificable de datos de audio.

En una realización ejemplar, el medio legible por ordenador puede comprender además una trama del flujo de bits que lleva un segundo conjunto de datos de audio decodificable en un segmento de la señal de audio mediante al menos el empleo de síntesis de señal, en base al segundo conjunto de datos de audio, con un paso acortado correspondiente al segundo número de muestras de la señal de audio.

Las tramas del flujo de bits que llevan los conjuntos de los datos de audio que son decodificables de manera independiente, según la presente realización ejemplar, se pueden emplear por ejemplo, para facilitar la decodificación de los flujos de bits después de unir y/o después de los descartes/duplicaciones de trama.

Según las realizaciones, puede suceder que N = 2 o N = 4, esto es, las N tramas del flujo de bits pueden ser dos o cuatro tramas del flujo de bits.

V. Realizaciones ejemplares

La Fig. 1 es un diagrama de bloques generalizado de un sistema 100 de procesamiento de audio para representar una señal X de audio como un flujo de bits B de audio, según una realización ejemplar.

El sistema 100 de procesamiento de audio comprende una sección 110 de codificación y una sección 120 de reentramado. La sección 110 de codificación codifica un segmento de la señal X de audio como un conjunto decodificable de datos D de audio mediante al menos la realización el análisis de la señal sobre el segmento de la señal X de audio con un paso básico correspondiente a un primer número de muestras de la señal X de audio.

Realizando el análisis de señal con un paso básico correspondiente al primer número de muestras de la señal X de audio se entiende que el análisis de señal se realiza dentro de una ventana de análisis de un cierto número de muestras de la señal X de audio, y cuando el siguiente segmento de la señal X de audio se ha de codificar la ventana de análisis se desplaza el mismo número de muestras que el paso básico. El análisis de señal puede por ejemplo ser realizado con ventanas de análisis de solapamiento, en cuyo caso la ventana de análisis puede ser mayor que el paso básico. En otro ejemplo, la longitud de la ventana de análisis puede coincidir con el paso básico.

La señal X de audio se ejemplifica en la presente memoria mediante una señal de audio multicanal. En la presente realización ejemplar, la sección 110 de codificación aplica una transformada de ventana, por ejemplo una Transformada de Coseno Discreta Modificada (MDCT), con el paso básico como el paso de transformada, al segmento de la señal X de audio para proporcionar una representación en el dominio de la frecuencia del segmento de la señal X de audio. En el dominio de la frecuencia, la sección 110 de codificación después calcula una señal de mezcla descendente (por ejemplo una mezcla descendente mono o estéreo) como una combinación lineal de los canales respectivos de la señal X de audio. La sección 110 de codificación también determina los parámetros para la reconstrucción paramétrica de la señal X de audio multicanal a partir de la señal de mezcla descendente. En la presente realización ejemplar, el conjunto decodificable de datos D de audio comprende la señal de mezcla descendente y los parámetros para la reconstrucción paramétrica.

Los parámetros por ejemplo pueden ser determinados en base al análisis de señal de la representación en el dominio de la frecuencia. Este análisis de señal puede emplear el paso básico, esto es puede emplear el mismo paso que la transformada de ventana. El análisis de señal puede por ejemplo incluir el cálculo de las energías y/o las covarianzas de los canales de la señal X de audio multicanal.

Se pueden prever también realizaciones en las que los parámetros para la reconstrucción paramétrica se determinen en base al análisis de señal con un paso diferente que el de la transformada de ventana. Se pueden prever realizaciones por ejemplo en las que la transformada de ventana emplee un paso de transformada más corto

que el paso básico, y en las que los parámetros para la reconstrucción paramétrica se determinen en base al análisis de señal con el paso básico.

El conjunto decodificable de datos D de audio corresponde a una primera tasa de tramas, por ejemplo, de 30 fps, y a un primer número de muestras de la señal de audio por trama. Es decir, el conjunto decodificable de los datos D representa el primer número de muestras de la señal de audio, y corresponde a una trama conforme a la primera tasa de trama.

La sección 120 de reentramado parte el conjunto decodificable de datos D de audio en N partes D_1, D_2, …, D_N por ejemplo dividiendo el conjunto decodificable de datos D de audio en N partes de al menos aproximadamente igual tamaño D_1, D_2, …, D_N. N puede ser por ejemplo 2 o 4 o puede ser cualquier número entero mayor o igual que 2.

En la presente realización ejemplar, el conjunto decodificable de datos D de audio es una representación en el dominio de la frecuencia del primer número de muestras. Por lo tanto, cuando el conjunto decodificable de datos D de audio se parte en las partes de igual tamaño D_1, D_2, …, D_N, estas partes D_1, D_2, …, D_N pueden comprender los subconjuntos respectivos de la representación en el dominio de la frecuencia, que no corresponden necesariamente con ningún subconjunto concreto del primer número de muestras de la señal de audio. Por lo tanto, las partes D_1, D_2, ..., D_N son conjuntos incompletos de datos de audio en el sentido de que ninguna de las partes D_1, D_2, ..., D_N se puede decodificar sin el acceso a todas las N partes D_1, D_2, ..., D_N.

La sección 120 de reentramado forma N tramas F_1, F_2, ..., F_N que llevan las respectivas partes D_1, D_2, ..., D_N. Ya que las N tramas F_1, F_2, ..., F_N representan un conjunto decodificable de datos D de audio, las tramas F_1, F_2, ..., F_N del flujo de bits tienen una segunda tasa de tramas que es N veces la tasa de tramas del conjunto decodificable de datos D de audio. De manera similar, aunque las tramas F_1, F_2, ..., F_N del flujo de bits no representan en sí ciertas muestras de la señal X de audio, las N tramas F_1, F_2, ..., F_N del flujo de bits representan el conjunto decodificable de datos D de audio y por lo tanto corresponden a un segundo número de muestras por trama, donde el primer número de muestras por trama es N veces el segundo número de muestras por trama.

La sección 120 de reentramado emite un flujo de bits B segmentado en tramas del flujo de bits que incluyen las N tramas F_1, F_2, ..., F_N del flujo de bits creadas como N tramas consecutivas del flujo de bits.

Además de las partes D_1, D_2, ..., D_N de los datos de audio, las tramas F_1, F_2, ..., F_N del flujo de bits también comprenden los respectivos metadatos μ_1, μ_2, ..., μ_N que indican que el conjunto decodificable de datos D de audio es obtenible a partir de las partes D_1, D_2, ..., D_N llevadas por las tramas F_1, F_2, ..., F_N del flujo de bits. Los metadatos μ_1, μ_2, ..., μ_N de cada una de las tramas F_1, F_2, ..., F_N del flujo de bits pueden indicar por ejemplo qué parte del conjunto decodificable de datos D de audio es llevada por la trama del flujo de bits, y, de manera opcional pueden indicar las tramas del flujo de bits que llevan las otras N – 1 partes del conjunto decodificable de datos D de audio.

Las Fig. 3 y 4 representan ejemplos de flujos de bits proporcionados por el sistema 100 de procesamiento de audio, descrito con referencia a la Fig. 1, según las realizaciones de ejemplo.

El flujo de bits B emitido por el sistema 100 de procesamiento de audio mostrado en la Fig. 1 se puede asociar con un flujo de tramas de video. El flujo de bits B se ejemplifica en la Fig. 3 mediante un flujo A1 de tramas del flujo de bits y un flujo V1 de tramas de video, donde la dirección hacia la derecha corresponde a un aumento de tiempo t.

El flujo V1 de tramas de video incluye tramas P de video codificadas predichas (que incluyen tramas que dependen de las tramas anteriores sólo y/o las así llamadas tramas bidireccionales, que dependen de tanto la trama anterior como la posterior) y una trama I de video codificada de manera independiente. El flujo A1 de las tramas del flujo de bits incluye las tramas del flujo de bits con la misma tasa de tramas y la misma duración que las tramas de video, para facilitar la unión y/o la sincronización con otros flujos de datos audiovisuales.

En la presente realización ejemplar, N = 4 y el sistema 100 de procesamiento de audio proporciona las tramas del flujo de bits en grupos 310 de cuatro tramas 311, 312, 313, 314 del flujo de bits que llevan las respectivas partes de un conjunto decodificable de datos de audio. Sin embargo, si el flujo V1 de tramas de video se ha de unir con otros flujos de tramas de video, la unión se puede realizar en un punto adyacente a la trama I de video codificada de manera independiente para facilitar la decodificación de las tramas de video después de la unión. Para mantener la sincronización de audio y video, el flujo A1 de tramas del flujo de bits se puede unir en el mismo punto de unión que el flujo V1 de tramas de video.

Para facilitar la decodificación de las tramas del flujo de bits después de la unión con otro flujo de tramas del flujo de bits, el sistema 100 de procesamiento de audio codifica un segmento de la señal X de audio de manera temporal en relación con la trama I de video codificada de manera independiente como un conjunto decodificable de datos de audio mediante la aplicación de análisis de señal con un paso de transformada acortado correspondiente al segundo número de muestras de la señal X de audio, que puede corresponder por ejemplo a la duración de la trama del video I codificado de manera independiente.

De manera análoga a la codificación en la que se emplea análisis de señal con el paso básico, la codificación que emplea análisis de señal con el paso acortado puede incluir la aplicación de una transformada de ventana, por

ejemplo una MDCT, con el paso acortado como paso de transformada, y determinar los parámetros para la reconstrucción paramétrica del segmento de la señal de audio a partir de la señal de mezcla descendente, donde los parámetros se determinan en base al análisis de señal con el paso acortado. El conjunto decodificable de datos de audio asociado con el paso acortado puede comprender la señal y los parámetros de la mezcla descendente.

El sistema 100 de procesamiento de audio incluye una trama 321 del flujo de bits que lleva el conjunto decodificable de datos de audio, que se puede decodificar de manera independiente sin acceder a los datos de audio llevados por las otras tramas del flujo de bits. En el flujo A1 de tramas del flujo de bits, la trama 321 del flujo de bits es seguida por otro grupo 330 de cuatro tramas 331, 332, 333, 334 de tramas de flujo de bits que llevan las partes respectivas de un conjunto decodificable de datos de audio.

El sistema 100 de procesamiento de audio puede comprender por ejemplo una sección de codificación adicional (no mostrada en la Fig. 1) configurada para codificar los segmentos de la señal X de audio aplicando análisis de señal con el paso acortado. De manera alternativa, la sección 110 de codificación puede ser operable para emplear el paso acortado, y la sección 120 de reentramado puede ser operable para incluir la trama 321 del flujo de bits, que lleva el conjunto decodificable de datos de audio asociado con el paso acortado, en el flujo de bits B.

En el ejemplo descrito con referencia a la Fig. 3, la presencia de una trama I de video codificado de manera independiente en ciertas posiciones puede ser manejada mediante la inclusión de una trama 321 del flujo de bits, que lleva un conjunto decodificable de datos de audio asociado con el paso acortado de la transformada, entre los grupos 310, 330 de cuatro tramas del flujo de bits. Sin embargo, en al menos algunos escenarios de ejemplo, las posiciones de las tramas I de video codificadas de manera independiente pueden ser desconocidas a priori, y/o las tramas I de video codificadas de manera independiente pueden ocurrir en posiciones que no coinciden con las posiciones entre los grupos de cuatro tramas del flujo de bits. Dicho escenario se ilustra en la Fig. 4.

El flujo de bits B y un flujo asociado de tramas de video se ejemplifican en la Fig. 4 mediante otro flujo de bit A2 de tramas del flujo de bits y otro flujo V2 de tramas de video, donde el tiempo t se propaga hacia la derecha.

De manera similar al escenario de ejemplo descrito con referencia a la Fig. 3, las tramas del flujo de bits se proporcionan mediante el sistema 100 de procesamiento de audio en los grupos 410, 430 de cuatro tramas del flujo de bits. Sin embargo, una vez que se detecta la trama I de video codificado de manera independiente en el flujo v2 de tramas de video, cuatro tramas 421, 422, 423, 424 del flujo de bits son codificadas por el sistema 100 de procesamiento de audio que emplea el paso acortado para cada una de ellas. Dependiendo de la posición de la trama I de video codificado de manera independiente en el flujo V2 de tramas de video, la trama I de video codificado de manera independiente puede corresponder con cualquiera de las cuatro tramas 421, 422, 423, 424 del flujo de bits proporcionadas usando el paso acortado de la transformada. En el escenario presente, una trama 423 del flujo de bits codificada de manera independiente se puede proporcionar en una posición en el flujo de bits A2 correspondiente al video I codificado de manera independiente, sin importar la posición de la trama I de video codificada de manera independiente en el flujo V2 de las tramas de video, en relación con cualesquiera grupos de cuatro tramas del flujo de bits en el flujo de bits A2 codificadas usando el paso básico. En el presente escenario, las tramas del flujo de bits se organizan en grupos de cuatro tramas del flujo de bits, sin importar la presencia de las tramas I de video codificadas de manera independiente en el flujo V2 de tramas de video.

La Fig. 2 es un diagrama de flujo de un método 200 de representación de una señal de audio mediante un flujo de bits de audio, según una realización ejemplar. El método 110 se ejemplifica en la presente memoria mediante un método realizado por el sistema 100 de codificación de audio, descrito con referencia a la Fig. 1.

El método 200 comprende la detección en 210 de si una trama actual del flujo V1 de tramas de video es codificada de manera independiente. Si la trama actual no se codifica de manera independiente, indicado mediante una N en el diagrama de flujo, el método 200 continua codificando en 220 un segmento de la señal X de audio como un conjunto decodificable de datos D de audio, mediante al menos el empleo de análisis de señal con el paso básico; partiendo en 230 el conjunto decodificable de datos D de audio en N partes D_1, D_2,..., D_N; creando en 240 N tramas F_1, F_2,..., F_N del flujo de bits que llevan las partes D_1, D_2,..., D_N respectivas y emitiendo en 250 las tramas F_1, F_2,..., F_N del flujo de bits como parte del flujo de bits B. El método 200 después vuelve a codificar otros segmentos de la señal X de audio.

Si, por otro lado, la trama actual del flujo V1 de tramas de video se codifica de manera independiente, indicado por una S en el diagrama de flujo, el método 200 en cambio continúa codificando en 260 un segmento de la señal X de audio como un conjunto decodificable de datos de audio mediante al menos el empleo de análisis de señal con el paso acortado; e incluyendo en 270 una trama del flujo de bits que lleva el segundo conjunto decodificable de datos de audio en el flujo de bits B. El método 200 entonces vuelve a decodificar otros segmentos de la señal X de audio.

La Fig. 5 es un diagrama de bloques generalizado de un sistema 500 de procesamiento de audio para reconstruir una señal de audio representada mediante un flujo de bits, según una realización ejemplar.

En la presente realización ejemplar, el flujo de bits es ejemplificado mediante el flujo de bits B emitido por el sistema 100 de procesamiento de audio, descrito con referencia a la Fig. 1. Más adelante se describen también realizaciones ejemplares en las que el sistema 500 de procesamiento de audio recibe los flujos de bits que se han modificado por

ejemplo mediante descartes de tramas y/o duplicaciones de tramas antes de ser recibidos por el sistema 500 de procesamiento de audio.

El sistema 500 de procesamiento de audio comprende una memoria intermedia 510 y una sección 520 de decodificación. La memoria intermedia 510 une los conjuntos de datos D_1, D_2, ..., D_N llevados mediante las tramas F_1, F_2, ..., F_N del flujo de bits en un conjunto de datos D de audio decodificable correspondiente a la primera tasa de tramas, por ejemplo de 30 fps, y al primer número de muestras de la señal de audio por trama. Como se describe con referencia a la Fig. 1, las tramas F_1, F_2, ..., F_N del flujo de bits tienen la segunda tasa de tramas correspondiente al segundo número de muestras de la señal de audio por trama del flujo de bits, donde el primer número de muestras es N veces el segundo número de muestras. La memoria intermedia 510 emplea los metadatos μ_1, μ_2, ..., μ_N. llevados por las tramas del flujo de bits para identificar las tramas F_1, F_2, ..., F_N que llevan los conjuntos de datos D_1, D_2, ..., D_N a ser unidos.

La sección 520 de decodificación decodifica el conjunto decodificable de datos D de audio en un segmento de la señal X de audio empleando síntesis de señal, en base al conjunto decodificable de datos D de audio, con el paso básico descrito con referencia a la Fig. 1, esto es el paso básico correspondiente al primer número de muestras de la señal X de audio. El sistema 500 de procesamiento de audio emite una versión $\tilde{X}$ reconstruida de la señal X de audio.

Como se describe con referencia a la Fig. 1, la señal X de audio es una señal de audio multicanal, y el conjunto decodificable de datos D de audio comprende la señal de mezcla descendente y los parámetros de mezcla ascendente asociados para la reconstrucción paramétrica de la señal X de audio. La sección 520 de decodificación realiza una reconstrucción paramétrica de una representación en el dominio de la frecuencia del segmento de la señal X de audio, empleando el paso básico. La sección 520 de decodificación después aplica una transformada de ventana, por ejemplo la MDCT inversa, que tiene el paso básico según el paso de la transformada, para obtener una representación en el dominio del tiempo del segmento de la señal X de audio.

Se pueden prever también realizaciones en las cuales la reconstrucción paramétrica se realiza con un paso diferente que el de la transformada de ventana. Se pueden prever realizaciones por ejemplo en la que la transformada de ventana emplea un paso de transformada más corto que el paso básico, y en las que la reconstrucción paramétrica se realiza con el paso básico.

Como se describe con referencia a las Fig. 3 y 4, el flujo de bits B puede comprender las tramas del flujo de bits que llevan conjuntos decodificables de datos de audio, esto es conjuntos de datos de audio que son decodificables de manera independiente de los demás mediante el empleo del paso acortado. El sistema 500 de procesamiento de audio puede por ejemplo comprender una sección de decodificación adicional (no mostrada en la Fig. 5) configurada para decodificar un conjunto decodificable de datos de audio que emplea el paso acortado. De manera alternativa, la sección 520 de decodificación puede ser operable para decodificar dicho conjunto de datos de audio usando el paso acortado, y la memoria intermedia 510 puede ser operable para transferir dicho conjunto decodificable de datos de audio a la sección 520 de decodificación sin unirlos con los datos de audio de otros flujos de bits.

Para permitir la conmutación suave entre los segmentos de la señal X de audio decodificada usando el paso acortado, y los segmentos de la señal X de audio decodificada usando el paso básico, el sistema 500 de procesamiento de audio puede proporcionar por ejemplo un retraso para que la decodificación de un grupo de N tramas del flujo de bits consecutivas a la segunda tasa de tramas, esto es que emplean el paso acortado, se complete al mismo tiempo que si las tramas del flujo de bits hubieran llevado cada una los conjuntos de datos de audio que requieren la unión en un conjunto decodificable de datos de audio, para decodificar. La memoria intermedia 510 puede proporcionar por ejemplo dicho retraso almacenando los conjuntos decodificables de datos de audio antes de transmitirlos a la sección 520 de decodificación. De manera alternativa, la sección 520 de decodificación puede proporcionar el retraso almacenando los segmentos reconstruidos de la señal X de audio, antes de proporcionarlos como salida.

Un flujo de bits B de audio emitido por el sistema 100 de procesamiento de audio, descrito con referencia a la Fig. 1, puede haber sido modificado, por ejemplo mediante la unión con otros flujos de bits, o mediante el descarte/duplicación de tramas antes de ser recibidas por el sistema 500 de procesamiento de audio descrito con referencia a la Fig. 5.

Como se describe con referencia a la Fig. 3, las tramas del flujo de bits pueden tener las mismas duraciones que las tramas de video correspondientes en un flujo V1 asociado de tramas de video. El empleo de dichos flujos A1 de audio y flujos V1 de video sincronizados en los flujos audiovisuales facilita la unión y/o la sincronización de los flujos audiovisuales.

Para un dispositivo o componente que realice la unión puede no ser necesario tomar en cuenta qué tipos de tramas del flujo de bits se disponen después de cada antes o después de la unión. En cambio, el sistema 500 de procesamiento de audio se puede adaptar para manejar una situación en la que alguna de las N tramas F_1, F_2,..., F_N del flujo de bits de un grupo que lleva las respectivas partes D_1, D_2,..., D_N de un conjunto decodificable de datos D de audio se pierden en el flujo de bits B recibido, por ejemplo, debido a la unión y/o el descarte/duplicación de tramas.

El sistema 500 de procesamiento de audio se puede configurar para detectar que las tramas del flujo de bits se pierden, por ejemplo, en base a los metadatos μ_1, μ_2, ..., μ_N llevados por las tramas F_1, F_2, ..., F_N respectivas del flujo de bits.

Una vez que se detecta que las tramas del flujo de bits requeridas para decodificar están perdidas, el sistema 500 de procesamiento de audio puede emplear por ejemplo una estrategia de ocultación de errores para continuar la decodificación de la señal X de audio. La estrategia de ocultación puede por ejemplo incluir el reemplazo de los datos de audio llevados por las tramas del flujo de bits en un grupo incompleto de tramas del flujo de bits, esto es un grupo del cual se pierden una o más tramas del flujo de bits en el flujo de bits recibido, por silencio (por ejemplo por ceros como coeficientes en el dominio de la frecuencia para la señal X de audio). La aparición y desaparición puede por ejemplo ser empleada por el sistema 500 de procesamiento de audio para proporcionar transiciones más suaves entre los segmentos decodificables de la señal X de audio, y el silencio reemplazando los segmentos no decodificables de la señal X de audio, según lo percibido por un oyente.

En algunas realizaciones ejemplares, el sistema 500 de procesamiento de audio se puede configurar para aceptar los flujos de bits asociados con al menos dos valores predefinidos diferentes para la segunda tasa de tramas, pero asociados con un valor común para el segundo número de muestras por trama. Esto se ejemplifica en la Tabla 1 mediante los valores 59,940 fps y 60,000 fps para la segunda tasa de tramas, y el valor común 768 para el segundo número de muestras por trama. Dichas tasas de tramas pueden ser útiles para los flujos de audio asociados con los flujos de video que tienen estas tasas de tramas.

Tabla 1

Tasa de tramas de video [fps]	**Segunda tasa de tramas [fps]**	**Segundo número de muestras por trama**	N	**Primera tasa de tramas [fps]**	**Primer número de muestras por trama**
29.970	29.970	1536	1	29.970	1536
30,000	30,000	1536	1	30,000	1536
59,940	59,940	768	2	29.970	1536
60,000	60,000	768	2	30,000	1536
119,880	119,880	384	4	29.970	1536
120,000	120,000	384	4	30.000	1536

En el presente ejemplo, los valores de la segunda tasa de tramas difieren por menos que el 5%. El sistema 500 de procesamiento de audio se puede adaptar para decodificar la señal X de audio que emplea el mismo valor para el paso básico para estos dos diferentes valores de la segunda tasa de tramas. Como se describe en la solicitud de patente pendiente del solicitante, aún no publicada, PCT/EP2014/056848 (véase en concreto la parte de la sección “II. Realizaciones de ejemplo”, que describe la Figura 1 y la Tabla 1 de ésta), la variación de la frecuencia de muestreo interna de la sección 520 de decodificación, provocada por la diferencia en la segunda tasa de tramas, normalmente puede ser tan pequeña que aún se puede proporcionar una calidad de reproducción aceptable de la señal X de audio reconstruida, según es percibida por el oyente, por el sistema 500 de procesamiento de audio. Otro ejemplo en la Tabla 1 de valores de la segunda tasa de tramas que difieren por menos del 5% está dado por los valores 119,880 fps y 120,000 fps para la segunda tasa de tramas, y el valor común 384 para el segundo número de muestras por trama.

Como se muestra en la Tabla 1, si la tasa de tramas de video es de 60,00 fps, se puede emplear N = 2 tramas del flujo de bits con la segunda tasa de tramas de 60,000 para representar un conjunto decodificable de datos de audio con la primera tasa de tramas de 30,000 fps. De manera similar, si la tasa de tramas de video es de 59,940 fps, se puede emplear N = 2 tramas del flujo de bits con la segunda tasa de tramas de 59,940 para representar un conjunto decodificable de datos de audio con la primera tasa de tramas de 29,970 fps. La Tabla 1 también muestra que si la tasa de tramas de video es de 120 fps, se puede emplear N = 4 tramas del flujo de bits con la segunda tasa de tramas de 120,000 para representar un conjunto decodificable de datos de audio con la primera tasa de tramas de 30,000 fps. De manera similar, si la tasa de tramas de video es de 119,880 fps, se puede emplear N = 4 tramas de flujo de bits con la segunda tasa de tramas de 119,880 para representar un conjunto decodificable de datos de audio con la primera tasa de tramas de 29,970 fps.

La Fig. 6 es un diagrama de flujo de un método 600 de procesamiento de audio para reconstruir una señal de audio representada mediante un flujo de bits, según una realización ejemplar. El método 600 se ejemplifica en la presente memoria mediante un método realizado por el sistema 500 de procesamiento de audio descrito con referencia a la Fig. 5.

El método 600 comprende la detección en 610 de si la trama del flujo de bits lleva un conjunto decodificable de datos de audio correspondiente a la segunda tasa de tramas.

Si no, indicado mediante una N en el diagrama de flujo, el método 600 continua uniendo en 620 los conjuntos de datos D_1, D_2,..., D_N de audio llevados por las N tramas F_1, F_2,..., F_N respectivas del flujo de bits en un conjunto decodificable de datos D de audio correspondiente a la primera tasa de tramas y al primer número de muestras de la señal de audio por trama; y decodificando en 630 el conjunto decodificable de datos D de audio en un segmento de la señal X de audio mediante al menos el empleo de síntesis de señal, en base al conjunto decodificable de datos D, con el paso básico correspondiente al primer número de muestras de la señal X de audio. El método 600 vuelve entonces al paso de detección 610 de si la siguiente trama del flujo de bits recibida lleva un conjunto decodificable de datos de audio.

Si es que si, indicado mediante una S en el diagrama de flujo, el método 600 en cambio continúa decodificando en 640 el conjunto decodificable de datos de audio correspondiente a la segunda tasa de tramas en un segmento de datos X de audio mediante al menos el empleo de el paso acortado correspondiente al segundo número de muestras de la señal X de audio. El método 600 entonces vuelve al paso de detección 610 de si la siguiente trama del flujo de bits lleva un conjunto decodificable de datos de audio.

La Fig. 7 es un diagrama de bloques generalizado de un sistema 700 de procesamiento de audio para transcodificar un flujo de bits de audio que representa una señal de audio, según una realización ejemplar.

El sistema 700 de procesamiento de audio comprende una sección 710 de recepción, una sección 720 de procesamiento opcional, y una sección 730 de reentramado. La sección 710 de recepción recibe un flujo de bits B1 que comprende una secuencia de conjuntos decodificables de datos D de audio correspondientes a una primera tasa de tramas y al primer número de muestras de la señal de audio por trama, por ejemplo descrita con referencia a la Fig. 1. La sección 710 de recepción extrae un conjunto decodificable de datos D de audio del flujo de bits B1.

La sección 720 de procesamiento (opcional) procesa el conjunto decodificable de datos D de audio. Dependiendo de la naturaleza del procesamiento, esto puede requerir la decodificación inicial de los datos de audio en una representación de transformada o de forma de onda; la sección 720 de procesamiento puede entonces realizar la síntesis de señal de secuencia, el procesamiento, el análisis de señal.

La sección 730 de reentramado parte el conjunto decodificable procesado de datos D de audio en N partes D_1, D_2, ..., D_N, y crea las N tramas F_1, F_2, ..., F_N del flujo de bits que llevan las respectivas partes D_1, D_2, ..., D_N. En la presente realización ejemplar, la sección 730 de reentramado realiza las mismas operaciones que la sección 120 de reentramado en el sistema 100 de procesamiento de audio, descrito con referencia a la Fig. 1. Por lo tanto, las tramas F_1, F_2, ..., F_N del flujo de bits tienen la segunda tasa de tramas correspondiente al segundo número de muestras de la señal de audio por trama del flujo de bits, y la sección 730 de reentramado emite un flujo de bits B2 segmentado en tramas del flujo de bits que incluyen las N tramas F_1, F_2, ..., F_N del flujo de bits creadas.

El flujo de bits B2 emitido por el sistema 700 de procesamiento de audio puede coincidir por ejemplo con el flujo de bits B emitido por el sistema 100 de procesamiento de audio, descrito con referencia a la Fig.1. El flujo de bits B1 recibido por el sistema 700 de procesamiento de audio puede por ejemplo ser un flujo de bits de 30 fps proporcionado por un codificador de audio conocido en la técnica.

Se apreciará que el flujo de bits B, descrito con referencia a las FIG. 1 y 5, y el flujo de bits A1 de tramas del flujo de bits, descrito con referencia a la Fig. 3, son ejemplos de un medio legible por ordenador que representa la señal X de audio y segmentado en tramas del flujo de bits, de acuerdo con las realizaciones ejemplares.

Se apreciará también que N puede ser un número entero mayor que 1.

VI. Equivalencias, extensiones, alternativas y miscelánea

Incluso aunque la presente descripción describa y represente realizaciones de ejemplo específicas, la invención no se restringe a estos ejemplos específicos. Se pueden hacer modificaciones y variaciones a las realizaciones de ejemplo anteriores sin salir del alcance de la invención, que se define sólo por las reivindicaciones adjuntas.

En las reivindicaciones, la palabra “comprendiendo” no excluye otros elementos o pasos, y el artículo indefinido “un” o “una” no excluye una pluralidad. El mero hecho de que ciertas mediciones se indiquen en reivindicaciones diferentes mutuamente dependientes no indica que no se pueda usar una combinación de estas mediciones como una ventaja. Cualquier signo de referencia que aparezca en las reivindicaciones no se debe entender como que limita su alcance.

Los dispositivos y métodos descritos anteriormente se pueden implementar como software, firmware, hardware o una combinación de los mismos. En una implementación de hardware, la división de tareas entre las unidades funcionales referidas en la anterior descripción no corresponde necesariamente a la división en unidades físicas; al contrario, un componente físico puede tener múltiples funcionalidades, y una tarea puede ser llevada a cabo de una forma distribuida, por varios componentes físicos en cooperación. Ciertos componentes o todos los componentes se

pueden implementar como un software ejecutado por un procesador digital, un procesador de señal o un microprocesador, o ser implementados como hardware o como un circuito integrado de aplicación específica. Dicho software se puede distribuir en medios legibles por un ordenador, que pueden comprender medios de almacenamiento informático (o medios no transitorios) y medios de comunicación (o medios transitorios). Como es bien sabido por una persona experta en la técnica, el término medios de almacenamiento informático incluye tanto los medios volátiles como los no volátiles, los extraíbles como los no extraíbles implementados en cualquier método o tecnología para el almacenamiento de información tal como las instrucciones legibles por un ordenador, las estructuras de datos, los módulos de programa u otros datos. Los medios de almacenamiento informático incluyen, pero no se limitan a, una RAM, una ROM, una EEPROM, una memoria flash u otra tecnología de memoria, un CD-ROM, un disco versátil digital (DVD) u otro almacenamiento de disco óptico, unos casetes magnéticos, una cinta magnética, un almacenamiento de disco magnético u otros dispositivos de almacenamiento magnéticos, o cualquier otro medio que se pueda usar para almacenar la información deseada y que pueda ser accedida por un ordenador. Además, es bien sabido por las personas expertas en la técnica que los medios de comunicación normalmente realizan instrucciones, estructuras de datos, módulos de programa u otros datos legibles por un ordenador en una señal de datos modulada como una onda portadora u otro mecanismo de transporte e incluye cualquier medio de entrega de información.

REIVINDICACIONES

1. Un método (200) de representación de una señal (X) de audio como un flujo de bits (B) de audio, comprendiendo el método:

codificar (220) un segmento de la señal de audio como un conjunto decodificable de datos (D) de audio mediante al menos la realización de análisis de señal, sobre el segmento de la señal de audio, con un paso básico correspondiente a un primer número de muestras de la señal de audio, el conjunto decodificable de datos de audio correspondiente a una primera tasa de tramas y a un primer número de muestras de la señal de audio por trama;

romper (230) el conjunto decodificable de datos de audio en N partes (D_1, D_2, ..., D_N) donde $N \geq 2$;

crear (240) N tramas (F_1, F_2,..., F_N) del flujo de bits que llevan las partes respectivas, en donde las N tramas del flujo de bits representan el conjunto decodificable de datos de audio y corresponden a un segundo número de muestras de la señal de audio por trama (F_1, F_2,..., F_N), en donde el primer número de muestras por trama es N veces el segundo número de muestras por trama, y en donde las N tramas del flujo de bits tienen una segunda tasa de tramas que es N veces la primera tasa de tramas; y

emitir (250) un flujo de bits segmentado en tramas del flujo de bits que incluyen las N tramas del flujo de bits anteriormente creadas,

en donde el método comprende además:

en respuesta al flujo de tramas de video que comprenden una trama de video de una cierto tipo (I), codificar (260) un segmento de la señal de audio relacionado temporalmente con dicha trama de video como un segundo conjunto decodificable de datos de audio mediante al menos la realización de análisis de señal, sobre el segmento de la señal de audio que se relaciona de manera temporal con dicha trama de video, con un paso acortado correspondiente al segundo número de muestras de la señal de audio, correspondiendo el segundo conjunto decodificable de datos de audio a dicha segunda tasa de tramas y dicho segundo número de muestras de la señal de audio por trama; e

incluir (270) una trama (321) del flujo de bits que lleva el segundo conjunto decodificable de datos de audio en el flujo de bits, siendo dicha trama del flujo de bits decodificable de manera independiente en un segmento, o en un subsegmento, de la señal de audio.

2. El método de la reivindicación 1, en donde realizar el análisis de señal incluye realizar, con el paso básico, al menos uno del grupo que comprende:

análisis espectral

análisis de energía,

análisis de entropía.

3. El método de la reivindicación 1 o 2, en donde la codificación de un segmento de la señal de audio incluye al menos uno del grupo que comprende:

aplicar una transformada de ventana que tenga el paso básico como paso de transformada;

calcular la señal de mezcla descendente y los parámetros para la reconstrucción paramétrica de la señal de audio a partir de la señal de mezcla descendente, en donde los parámetros se calculan en base a dicho análisis de señal.

4. El método de cualquiera de las reivindicaciones anteriores, comprendiendo además:

incluir los metadatos (μ_1, μ_2, ..., μ_N) en al menos una de las N tramas del flujo de bits que llevan dichas partes, indicando los metadatos que el conjunto decodificable completo de datos de audio es obtenible a partir de las partes llevadas por las N tramas del flujo de bits; y/o

en respuesta al flujo de las tramas de video que comprenden una trama de video de dicho tipo, codificar N segmentos consecutivos de la señal de audio como los respectivos conjuntos decodificables de datos de audio mediante, para cada uno de los N segmentos consecutivos, al menos aplicar el análisis de señal con el paso acortado, en donde dicho segmento temporalmente relacionado con dicha trama de video es uno de los N segmentos consecutivos; e

incluir unas tramas (421, 422, 423, 424) del flujo de bits que llevan los respectivos conjuntos decodificables de datos de audio asociados con los N segmentos consecutivos en el flujo de bits.

5. Un sistema (100) de procesamiento de audio para representar una señal (X) de audio mediante un flujo de bits (B) de audio, comprendiendo el sistema de procesamiento de audio:

una sección (110) de codificación configurada para codificar un segmento de la señal de audio como un conjunto

decodificable de datos (D) de audio mediante al menos la realización de análisis de señal, sobre el segmento de la señal de audio, con un paso básico correspondiente a un primer número de muestras de la señal de audio, correspondiendo el conjunto decodificable de datos de audio a una primera tasa de tramas y a un primer número de muestras de la señal de audio por trama;

una sección de codificación adicional configurada para, en respuesta al flujo de tramas de video que comprende una trama de video de un cierto tipo (I), codificar (260) un segmento de la señal de audio relacionado de manera temporal a dicha trama de video como un segundo conjunto decodificable de datos de audio mediante al menos la realización de análisis de señal, sobre el segmento de la señal de audio relacionada de manera temporal a dicha trama de video, con un paso acortado correspondiente al segundo número de muestras de la señal de audio, correspondiendo el segundo conjunto decodificable de datos de audio a dicha segunda tasa de tramas y a dicho segundo número de muestras de la señal de audio por trama.

una sección (120) de reentramado configurada para:

partir el conjunto decodificable de datos de audio en N partes (D_1, D_2, ..., D_N), donde $N \geq 2$;

crear N tramas (F_1, F_2,..., F_N) del flujo de bits que llevan las partes respectivas, en donde las N tramas del flujo de bits representan el conjunto decodificable de datos de audio y corresponden a un segundo número de muestras de la señal de audio por trama (F_1, F_2,..., F_N), en donde el primer número de muestras por trama es N veces el segundo número de muestras por trama, y en donde las tramas del flujo de bits que tienen una segunda tasa de tramas es N veces la primera tasa de tramas; y

emitir un flujo de bits segmentado en tramas del flujo de bits que incluyen las N tramas del flujo de bits anteriormente creadas, e incluir (270) una trama (321) del flujo de bits que lleve el segundo conjunto decodificable de datos de audio en el flujo de bits, siendo dicha trama del flujo de bits decodificable de manera independiente en un segmento, o en un subsegmento, de la señal de audio.

6. Un método (600) de reconstrucción de una señal (X) de audio representada mediante un flujo de bits (B) segmentado en tramas del flujo de bits, comprendiendo el método:

unir (620) los conjuntos de datos (D_1, D_2,..., D_N) de audio llevados por las N tramas (F_1, F_2,..., F_N) del flujo de bits respectivas en un conjunto decodificable de datos (D) de audio correspondiente a una primera tasa de tramas y a un primer número de muestras de la señal de audio por trama, donde $N \geq 2$, en donde las N tramas del flujo de bits representan el conjunto decodificable de datos de audio y corresponden a un segundo número de muestras de la señal de audio por trama (F_1, F_2,..., F_N), en donde el primer número de muestras por trama es N veces el segundo número de muestras por trama, y en donde las tramas del flujo de bits tienen una segunda tasa de tramas que es N veces la primera tasa de tramas; y

decodificar (630) el conjunto decodificable de datos de audio en un segmento de la señal de audio mediante al menos el empleo de síntesis de señal, en base al conjunto decodificable de datos, con un paso básico correspondiente al primer número de muestras de la señal de audio,

en donde el método comprende además:

detectar (610) si una trama del flujo de bits lleva un conjunto decodificable de datos de audio correspondiente a la segunda tasa de tramas; y

decodificar (640) el conjunto decodificable de datos de audio correspondiente a la segunda tasa de tramas en un segmento de la señal de audio mediante al menos el empleo de síntesis de señal, en base al conjunto decodificable de datos de audio correspondiente a la segunda tasa de tramas, con un paso acortado correspondiente a un segundo número de muestras, en donde el primer número de muestras es N veces el segundo número de muestras.

7. El método de la reivindicación 6, en donde decodificar el conjunto decodificable de datos de audio incluye al menos uno del grupo que comprende:

aplicar una transformada de ventana que tenga el paso básico como paso de transformada;

realizar la reconstrucción paramétrica, con dicho paso básico, del segmento de la señal de audio, en base a la señal de mezcla descendente y los parámetros asociados obtenidos a partir del conjunto decodificable de datos de audio.

8. El método de la reivindicación 6 o 7, en donde dichas N tramas del flujo de bits, a partir de las cuales los conjuntos de datos de audio se unen en el conjunto decodificable de datos de audio, son N tramas consecutivas del flujo de bits; y/o

en donde el flujo de bits se asocia con un flujo de tramas (V1, V2) de video que tengan una tasa de tramas que coincida con la segunda tasa de tramas.

9. El método de cualquiera de las reivindicaciones 6 – 8, que comprende además:

determinar, en base a los metadatos (μ_1, μ_2, ..., μ_N) llevados por al menos alguna de las tramas del flujo de bits en el flujo de bits, un grupo de tramas del flujo de bits a partir del cual unir los conjuntos incompletos de datos de audio en el conjunto decodificable de datos de audio.

10. El método de cualquiera de las reivindicaciones 6 – 9, en donde decodificar un conjunto decodificable de datos de audio correspondiente a la segunda tasa de tramas incluye proporcionar un retraso tal que la decodificación de un grupo de N tramas consecutivas del flujo de bits a la segunda tasa de tramas se completa en el mismo tiempo que si las tramas del flujo de bits de dicho grupo de N tramas del flujo de bits hubieran llevado cada una los conjuntos de datos de audio que requieren la unión en un conjunto decodificable de datos de audio, y de manera opcional, en donde dicho retraso se proporciona mediante el almacenamiento de al menos un conjunto decodificable de datos de audio correspondiente a la segunda tasa de tramas o mediante el almacenamiento de al menos un segmento de la señal de audio.

11. El método de cualquiera de las reivindicaciones 6 – 10, en donde decodificar un segmento de la señal de audio en base al conjunto decodificable de datos de audio correspondiente a la primera tasa de tramas comprende:

recibir los coeficientes espectrales cuantificados correspondientes al conjunto decodificable de datos de audio correspondiente a la primera tasa de tramas;

realizar la cuantificación inversa seguida por una transformación de frecuencia a tiempo, a través de la cual se obtiene una representación de una señal de audio intermedia;

realizar al menos un paso de procesamiento en el dominio de la frecuencia sobre la señal de audio intermedia; y

cambiar una tasa de muestreo de la señal de audio procesada a una frecuencia de muestreo objetivo, a través de la cual se obtiene una representación en el dominio del tiempo de la señal de audio reconstruida.

12. El método de la reivindicación 11, que acepta flujos de bits asociados con al menos dos valores diferentes para la segunda tasa de tramas pero asociados con un valor común para el primer número de muestras por trama, difiriendo los valores respectivos de la segunda tasa de tramas en como mucho un 5%, en donde la transformación de frecuencia la tiempo se realiza en un componente funcional configurado para emplear una transformada de ventana que tenga un valor predefinido común para el paso básico como paso de transformada para dichos al menos dos valores diferentes para la segunda tasa de tramas.

13. Un sistema (500) de procesamiento de audio para reconstruir una señal de audio (X) representada mediante un flujo de bits (B) segmentada en tramas del flujo de bits, comprendiendo el sistema de procesamiento de audio:

una memoria intermedia (510) configurada para unir los conjuntos de datos (D_1, D_2,..., D_N) de audio llevados por las respectivas N tramas (F_1, F_2,..., F_N) del flujo de bits en un conjunto decodificable de datos (D) de audio correspondiente a una primera tasa de tramas y a un primer número de muestras de la señal de audio por trama, donde $N \geq 2$, en donde las N tramas del flujo de bits representan el conjunto decodificable de datos de audio y corresponden a un segundo número de muestras de la señal de audio por trama (F_1, F_2,..., F_N), en donde el primer número de muestras por trama es N veces el segundo número de muestras por trama, y en donde las tramas del flujo de bits tienen una segunda tasa de tramas que es N veces la primera tasa de tramas; y

una sección (520) de decodificación configurada para decodificar el conjunto decodificable de datos de audio en un segmento de la señal de audio mediante al menos el empleo de síntesis de señal, en base al conjunto decodificable de datos de audio, con un paso básico correspondiente al primer número de muestras de la señal de audio,

en donde el sistema de procesamiento de audio se configura para detectar si una trama del flujo de bits lleva un conjunto decodificable de datos de audio correspondiente a la segunda tasa de tramas, y para decodificar el conjunto decodificable de datos de audio correspondiente a la segunda tasa de tramas en un segmento de la señal de audio mediante al menos el empleo de síntesis de señal, en base al conjunto decodificable de datos de audio correspondiente a la segunda tasa de tramas, con un paso acortado correspondiente a un segundo número de muestras, en donde el primer número de muestras en N veces el segundo número de muestras.

14. Un producto de programa informático que comprende un medio legible por un ordenador con instrucciones para realizar el método de cualquiera de las reivindicaciones 1 – 4 y 6 – 12.

15. El método de cualquiera de las reivindicaciones 1 – 4 y 6 – 12, el sistema de cualquiera de las reivindicaciones 5 y 13, o el producto de programa informático de la reivindicación 14, en donde N = 2 o N = 4.

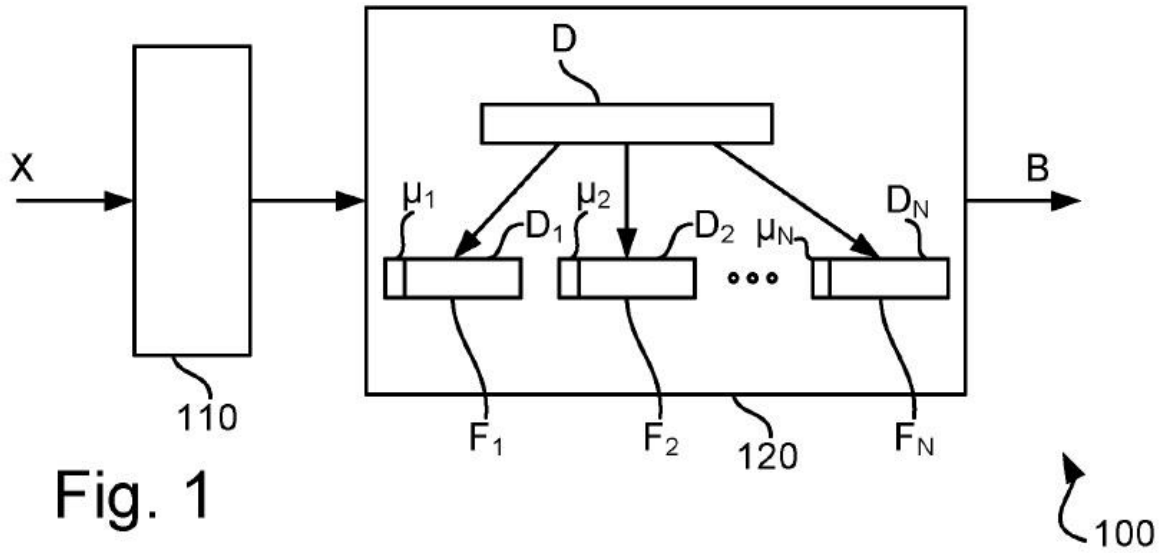
D
X
B
μ1
μ2
μN
D1
D2
DN
110
F1
F2
120
FN
100

Fig. 1

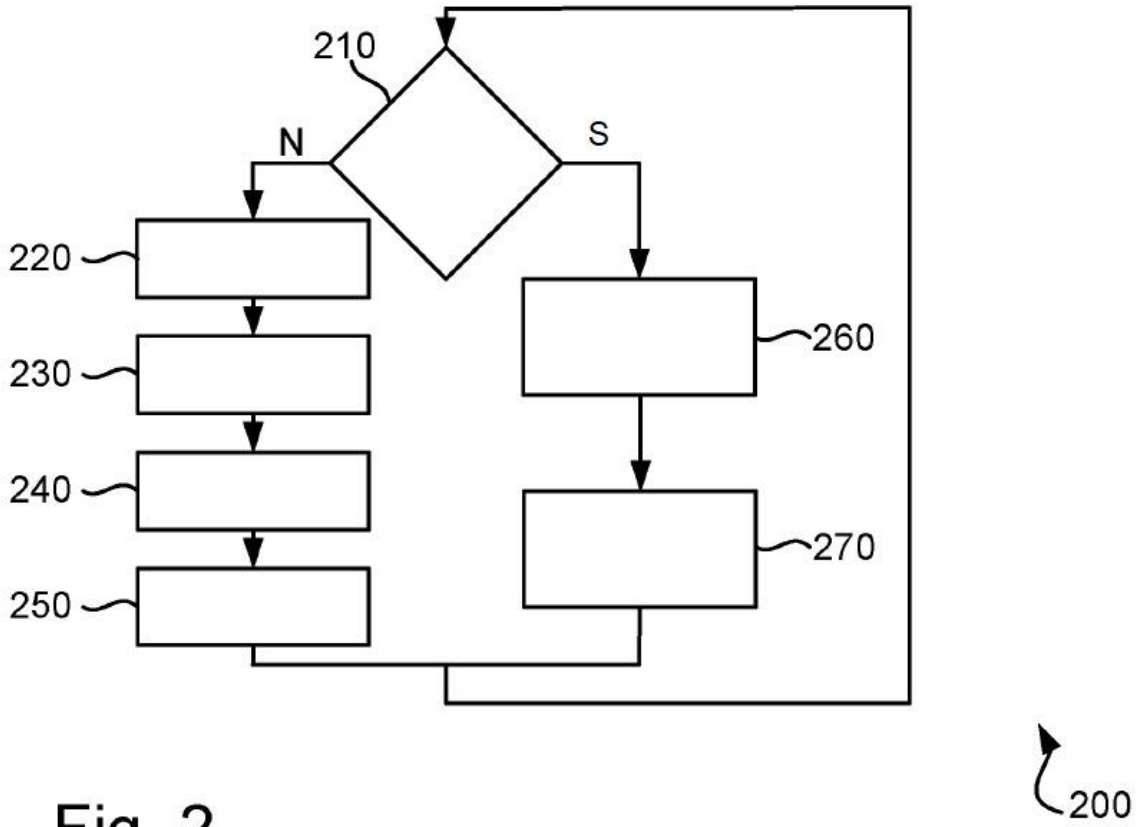
210
N
S
220
230
240
250
260
270
200

Fig. 2

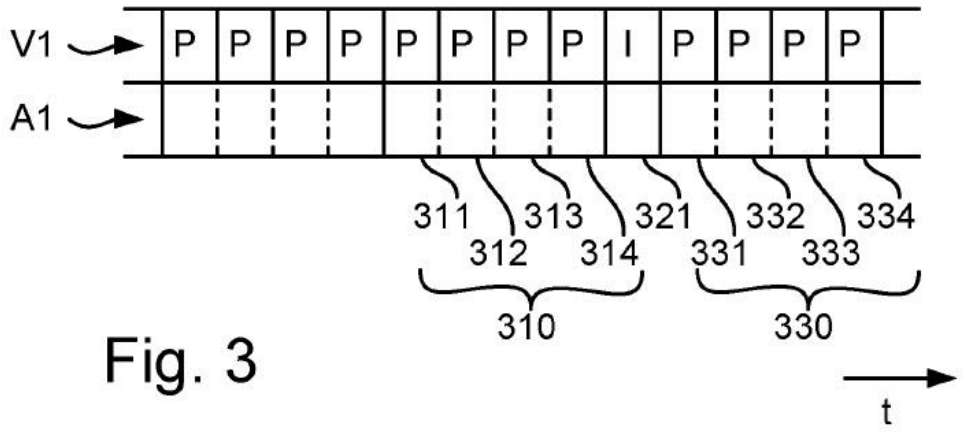

Fig. 3

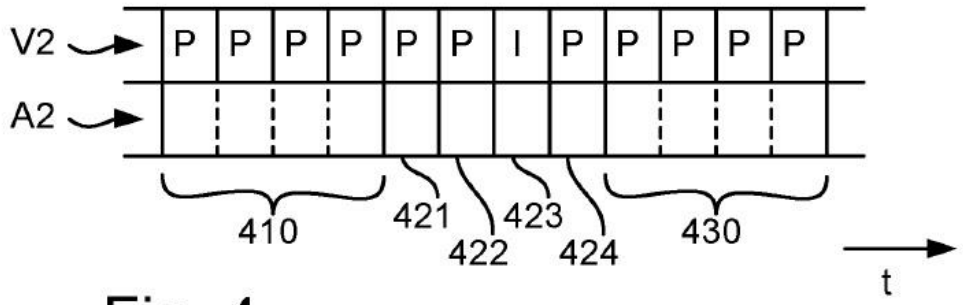

Fig. 4

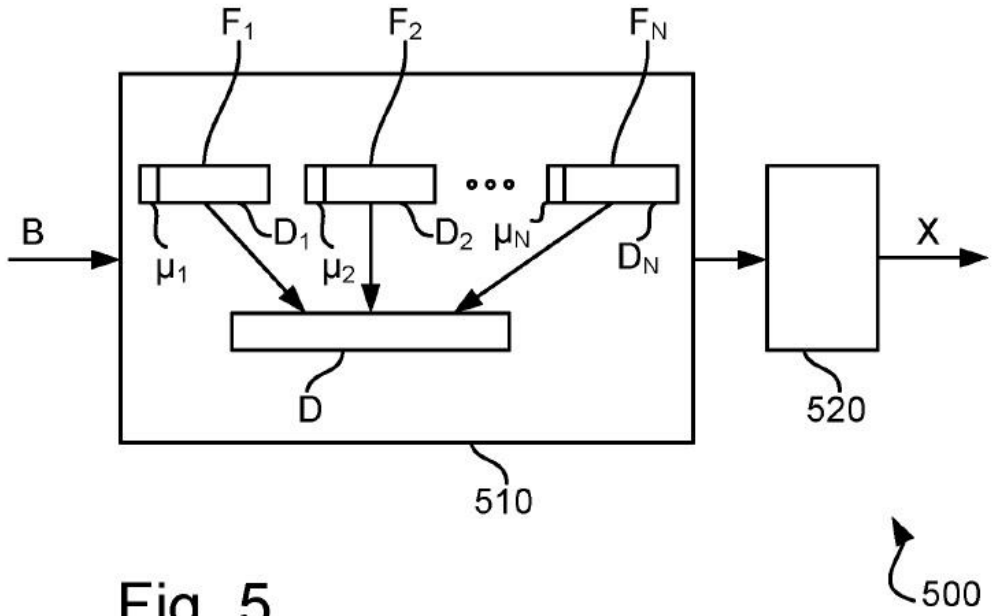

Fig. 5

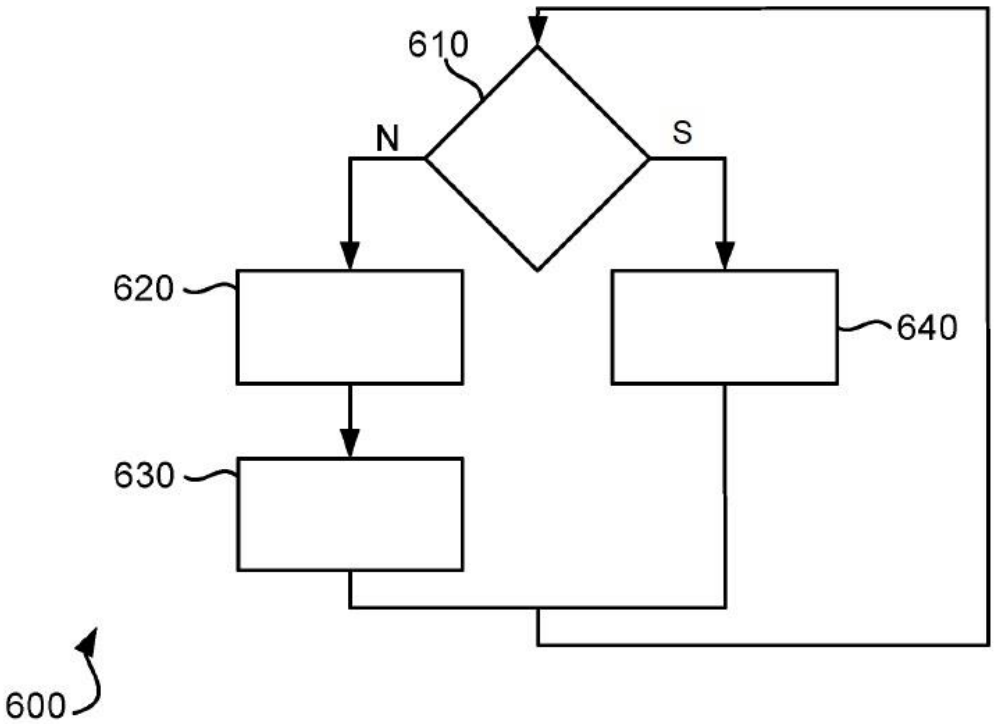
610
N
S
620
640
630
600

Fig. 6

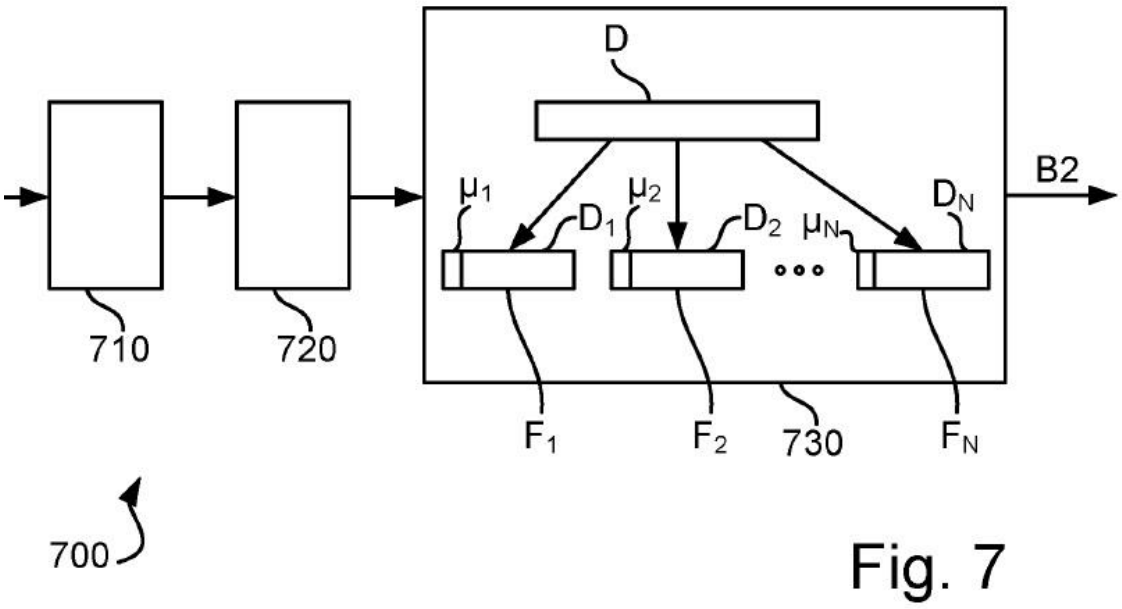
D
B2
μ1
μ2
μN
D1
D2
DN
710
720
F1
F2
730
FN
700

Fig. 7