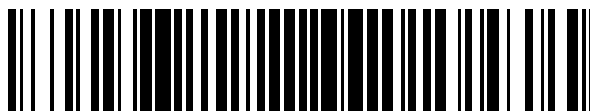


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 753 899**

51 Int. Cl.:

G10L 19/008 (2013.01)

G10L 21/038 (2013.01)

G10L 19/26 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **04.03.2009 PCT/EP2009/001534**

87 Fecha y número de publicación internacional: **11.09.2009 WO09109374**

96 Fecha de presentación y número de la solicitud europea: **04.03.2009 E 09716835 (5)**

97 Fecha y número de publicación de la concesión europea: **21.08.2019 EP 2260487**

54 Título: **Mezclado de trenes de datos de entrada y generación de un tren de datos de salida a partir de los mismos**

30 Prioridad:

04.03.2008 US 33590

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

14.04.2020

73 Titular/es:

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

72 Inventor/es:

**SCHNELL, MARKUS;
LUTZKY, MANFRED y
MULTRUS, MARKUS**

74 Agente/Representante:

SALVÀ FERRER, Joan

ES 2 753 899 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

Mezclado de trenes de datos de entrada y generación de un tren de datos de salida a partir de los mismos

- 5 **[0001]** Las realizaciones según la presente invención se refieren al mezclado de una pluralidad de trenes de datos de entrada para obtener un tren de datos de salida y a la generación de un tren de datos de salida mezclando los trenes de datos de entrada primero y segundo, respectivamente. El tren de datos de salida puede usarse, por ejemplo, en el campo de sistemas de conferencia que incluyen sistemas de videoconferencia y sistemas de teleconferencia.
- 10 **[0002]** En muchas aplicaciones, se deberá procesar más de una señal de audio de manera que a partir de la serie de señales de audio se generará una señal o al menos un número reducido de señales, lo que a menudo se refiere como "mezclado". El proceso de mezclado de señales de audio puede referirse por lo tanto a la agrupación de varias señales de audio individuales en una señal resultante. Este proceso se usa, por ejemplo, cuando se crean
- 15 piezas de música de un disco compacto ("adición de pistas de audio"). En este caso, en una canción se mezclan normalmente diferentes señales de audio de diferentes instrumentos junto con una o más señales de audio que comprenden ejecuciones vocales (canto).
- [0003]** Otros campos de aplicación, en los que el mezclado desempeña un papel importante, son sistemas de
- 20 videoconferencia y sistemas de teleconferencia. Este sistema normalmente es capaz de conectar a varios participantes distribuidos espacialmente en una conferencia empleando un servidor central, que mezcla de forma apropiada los datos de audio y vídeo de entrada de los participantes registrados y envía a cada uno de los participantes una señal resultante como retorno. Esta señal resultante o señal de salida comprende las señales de audio de todos los demás participantes de la conferencia.
- 25 **[0004]** En sistemas de conferencia digitales modernos concurren diversos objetivos y aspectos parcialmente contradictorios. Será preciso tener en cuenta la calidad de la señal de audio reconstruida, así como la aplicabilidad y utilidad de algunas técnicas de codificación y decodificación para diferentes tipos de señales de audio (por ejemplo, señales de habla comparadas con señales de audio en general y con señales musicales). Otros aspectos que pueden
- 30 tener que considerarse también cuando se diseñan e implementan sistemas de conferencia son las cuestiones relativas a la anchura de banda disponible y el retardo.
- [0005]** Por ejemplo, cuando se busca un equilibrio entre la calidad, por una parte, y la anchura de banda, por otra, en la mayoría de los casos es inevitable un compromiso. Sin embargo, pueden lograrse mejoras referentes a la
- 35 calidad implementando técnicas modernas de codificación y decodificación tales como la técnica de Códec de Audio Avanzado – Retardo Bajo Mejorado (AAC-ELD, AAC = Advanced Audio Codec; ELD = Enhanced Low Delay). Sin embargo, la calidad que puede alcanzarse puede influir de forma negativa en sistemas que emplean estas técnicas modernas por problemas y aspectos más fundamentales.
- 40 **[0006]** Por citar solo un reto que debe resolverse, todas las transmisiones de señal digital se enfrentan al problema de una cuantificación necesaria, que puede, al menos en principio, evitarse en circunstancias ideales en un sistema analógico sin ruido. Debido al proceso de cuantificación, se introduce inevitablemente una cierta cantidad de ruido de cuantificación en la señal para procesar. Para contrarrestar distorsiones posibles y audibles, puede resultar
- 45 tentador incrementar el número de niveles de cuantificación y por lo tanto incrementar en consecuencia la resolución de cuantificación. Esto, sin embargo, lleva a un número mayor de valores de señal para transmitir y por lo tanto a un aumento en la cantidad de datos para transmitir. Dicho de otro modo, mejorar la calidad reduciendo posibles distorsiones introducidas por ruido de cuantificación puede incrementar, en ciertas circunstancias, la cantidad de datos para transmitir y en su caso violar las restricciones de anchura de banda impuestas en un sistema de transmisión.
- 50 **[0007]** En el caso de los sistemas de conferencia, los retos de mejorar el compromiso entre calidad, anchura de banda disponible y otros parámetros pueden complicarse además por el hecho de que normalmente se procesará más de una señal de audio de entrada. Por lo tanto, podrá ser necesario considerar condiciones frontera impuestas por más de una señal de audio cuando se genera la señal de salida o se obtiene la señal producida por el sistema de conferencia.
- 55 **[0008]** En especial, dado el reto adicional de implementar sistemas de conferencia con un retardo suficientemente bajo para permitir una comunicación directa entre los participantes de una conferencia sin introducir retardos sustanciales que puedan considerarse inaceptables por los participantes, el desafío aumenta aún más.
- 60 **[0009]** En implementaciones de bajo retardo de sistemas de conferencia, las fuentes de retardo normalmente están restringidas en su número, lo que por otra parte puede llevar al reto de procesar los datos fuera del dominio del tiempo, en el que el mezclado de las señales de audio puede lograrse superponiendo o sumando las señales respectivas.
- 65 **[0010]** En términos generales, resulta favorable elegir el compromiso entre calidad, anchura de banda

disponible y otros parámetros adecuados para sistemas de conferencia de forma cuidadosa con el fin de afrontar el esfuerzo adicional de tratamiento para mezclar en tiempo real, reducir la cantidad de hardware requerida y contener los costes en términos de hardware y tara de transmisión razonables, sin comprometer la calidad de audio.

5 **[0011]** Para reducir la cantidad de datos transmitidos, los códecs de audio modernos usan a menudo herramientas altamente sofisticadas para describir información espectral referente a componentes espectrales de una señal de audio respectiva. Al usar estas herramientas, que se basan en fenómenos psicoacústicos y resultados de examen, puede lograrse un compromiso mejorado entre parámetros parcialmente contradictorios y condiciones frontera tales como la calidad de la señal de audio reconstruida a partir de datos transmitidos, la complejidad de
10 cálculo, la velocidad binaria y otros parámetros.

[0012] Algunos ejemplos de estas herramientas son, por ejemplo, sustitución de ruido perceptual (PNS = Perceptual Noise Substitution), conformado de ruido temporal (TNS = Temporal Noise Shaping) y replicación de banda espectral (SBR = Spectral Band Replication), por citar solos unos cuantos. Todas estas técnicas se basan en describir
15 al menos parte de la información espectral con un número reducido de bits, de manera que en comparación con un tren de datos basado en no usar estas herramientas, pueden asignarse más bits para partes espectralmente importantes del espectro. Como consecuencia, a la vez que se mantiene la velocidad binaria, puede mejorarse un nivel perceptible de calidad usando estas herramientas. Naturalmente, puede elegirse un compromiso diferente, en concreto reducir el número de bits transmitidos por trama de datos de audio que mantienen la impresión de audio total.
20 Pueden alcanzarse asimismo otros compromisos que se encuentran entre estos dos extremos.

[0013] Estas herramientas también pueden emplearse en aplicaciones de telecomunicaciones. Sin embargo, cuando están presentes más de dos participantes en esta situación de comunicaciones, puede ser muy ventajoso usar un sistema de conferencia para mezclar dos o más trenes de bits de más de dos participantes. Situaciones como estas
25 ocurren en tanto en situaciones de teleconferencia basadas puramente en audio como situaciones de videoconferencia.

[0014] Se describe, por ejemplo, un sistema de conferencia que opera en un dominio de la frecuencia en el documento US 2008/0097764 A1 que realiza el mezclado real en el dominio de la frecuencia y por lo tanto omite la
30 retransformación de las señales de audio de entrada de nuevo al dominio del tiempo.

[0015] Sin embargo, el sistema de conferencia aquí descrito no tiene en cuenta las posibilidades de herramientas tal como se han descrito anteriormente, que permiten una descripción de información espectral de al menos un componente espectral de una forma más condensada. Como resultado, este sistema de conferencia
35 necesita etapas de transformación adicionales para reconstruir las señales de audio que se proporcionan al sistema de conferencia al menos en un grado tal que las señales de audio respectivas estén presentes en el dominio de la frecuencia. Por otra parte, la señal de audio mezclada resultante también se necesita para la retransformación basada en las herramientas adicionales mencionadas anteriormente. Estas etapas de retransformación y transformación requieren, sin embargo, una aplicación de algoritmos complejos que pueden llevar a una mayor complejidad de cálculo
40 y, por ejemplo, en el caso de aplicaciones críticas en términos energéticos, portátiles, a un consumo incrementado de energía y por lo tanto a un tiempo operativo limitado.

[0016] El documento US 2006/173691 describe un aparato de tratamiento de mezclado de audio que incluye unidades de entrada configuradas para recibir una pluralidad de datos de audio, una unidad de mezclado configurada
45 para mezclar la pluralidad de datos de audio mezclados y una unidad de salida configurada para producir una forma codificada de los datos de audio mezclados con el exterior.

[0017] El documento US 2006/047523 describe un procedimiento para combinar señales codificadas en el dominio de la frecuencia de al menos dos fuentes de señales. Para permitir la combinación de señales sin decodificar
50 las señales por completo, el procedimiento proporciona la decodificación de las señales codificadas obteniendo componentes espectrales cuantificados, la cuantificación inversa del componente espectral cuantificado de las señales decodificadas que obtienen secuencias de ventana y la combinación de al menos dos señales cuantificadas inversas que obtienen una señal combinada.

55 **[0018]** El documento de la norma ISO/IEC 14496-3:2005/FDAM 9:2007(E) titulado AAC de bajo retardo mejorado describe procedimientos para mezclar objetos de audio de bajo retardo y mezclado en la parte SBR sin partes de decodificación de la carga útil de los trenes de objetos de entrada.

[0019] Por lo tanto un problema que deben resolver las realizaciones según la presente invención es facilitar
60 un compromiso mejorado entre calidad, anchura de banda disponible y otros parámetros adecuados para sistemas de conferencia, o permitir una reducción de complejidad de cálculo requerida en un sistema de conferencia tal como se ha descrito anteriormente.

[0020] Este objetivo se logra por un aparato según la reivindicación 1, un procedimiento para mezclar una
65 pluralidad de trenes de datos de entrada según la reivindicación 5 o un programa informático según la reivindicación

6.

[0021] Las realizaciones según la presente invención se basan en el hallazgo de que, cuando se mezcla una pluralidad de trenes de datos de entrada, se puede lograr un compromiso mejorado entre los parámetros y objetivos anteriormente mencionados, determinando un tren de datos de entrada basado en una comparación y copiando información al menos parcialmente espectral del tren de datos de entrada determinado en el tren de datos de salida. Al copiar información al menos parcialmente espectral de un tren de datos de entrada, puede omitirse una recuantificación y por lo tanto el ruido de recuantificación asociado. En caso de información espectral para la cual no puede determinarse un tren de entrada dominante, el mezclado de la información espectral correspondiente en el dominio de la frecuencia puede realizarse mediante una realización según la presente invención.

[0022] La comparación puede basarse, por ejemplo, en un modelo psicoacústico. La comparación puede relacionarse además con información espectral correspondiente a un componente espectral común (por ejemplo, una frecuencia o una banda de frecuencia) de al menos dos trenes de datos de entrada diferentes. Por lo tanto puede tratarse de una comparación entre canales. En el caso en que la comparación se basa en un modelo psicoacústico, la comparación puede describirse así de manera que considera un enmascaramiento entre canales.

[0023] Según un aspecto, las realizaciones según la presente invención se basan en el hallazgo de que puede reducirse la complejidad de operaciones durante el mezclado de un primer tren de datos de entrada y un segundo tren de datos de entrada para generar un tren de datos de salida si se tienen en cuenta valores de control asociados con los datos de carga útil del tren de datos de entrada respectivo, en el que los valores de control indican una forma en que los datos de carga útil representan al menos una parte de la información espectral correspondiente o dominio espectral de las señales de audio respectivas. En el caso en que los valores de control de los dos trenes de datos de entrada sean iguales, puede omitirse una nueva decisión sobre la forma en que el dominio espectral en la trama respectivo del tren de datos de salida y en su lugar la generación de trenes de salida puede basarse en la decisión ya determinada y de manera concordante por los codificadores del tren de datos de entrada, es decir, adoptar su valor de control a partir de ellos. Dependiendo de la forma indicada por los valores de control, puede ser incluso posible y preferido evitar la retransformación de los datos de carga útil respectiva de retorno en otra forma de representar el dominio espectral tal como la forma normal o simple con un valor espectral por muestra de tiempo/espectral. En este último caso, puede generarse un tratamiento directo del tren de carga útil para producir los datos de carga útil correspondiente del tren de datos de salida y siendo el valor de control igual a los valores de control de los trenes de datos de entrada primero y segundo, de manera que la "directividad" significa "sin cambiar la forma de dominio espectral representado" tal como mediante PNS o características de audio similares descritas con mayor detalle a continuación.

[0024] En realizaciones según una realización de la presente invención, los valores de control se refieren a al menos un solo componente espectral. Por otra parte, en realizaciones según la presente invención, estas operaciones pueden llevarse a cabo cuando tramas del primer tren de datos de entrada y del segundo tren de datos de entrada corresponden al índice de tiempo común con respecto a una secuencia de tramas apropiada de los dos trenes de datos de entrada.

[0025] En el caso en que los valores de control de los trenes de datos primero y segundo no sean iguales, las realizaciones según la presente invención pueden realizar la etapa consistente en transformar los datos de carga útil de una trama de uno de los trenes de datos de entrada primero y segundo, para obtener una representación de los datos de carga útil de una trama del otro tren de datos de entrada. Los datos de carga útil del tren de datos de salida pueden generarse entonces basándose en los datos de carga útil transformados y los datos de carga útil de los otros dos trenes. En algunos casos, las realizaciones según la presente invención que transforman los datos de carga útil de la trama del tren de datos de entrada a la representación de los datos de carga útil de la trama del otro tren de datos de entrada pueden realizarse directamente sin transformar la señal de audio respectiva de nuevo en el dominio de la frecuencia simple.

[0026] A continuación se describirán las realizaciones según la presente invención haciendo referencia a las siguientes figuras.

[0027] La invención se define en las reivindicaciones adjuntas. Todas las apariciones del término "realización" o "realizaciones", excepto las que corresponden a las reivindicaciones, se refieren a ejemplos útiles para entender la invención que se presentaron originalmente pero que no representan realizaciones de la invención reivindicada en la actualidad. Estos ejemplos se muestran únicamente con fines ilustrativos.

La Fig. 1 muestra un diagrama de bloques de un sistema de conferencia;

la Fig. 2 muestra un diagrama de bloques del sistema de conferencia basado en un códec de audio general;

la Fig. 3 muestra un diagrama de bloques de un sistema de conferencia que opera en un dominio de la frecuencia usando la tecnología de mezclado de trenes de bits;

la Fig. 4 muestra un dibujo esquemático de un tren de datos que comprende una pluralidad de tramas;

la Fig. 5 ilustra formas diferentes de componentes espectrales e información o datos espectrales;

5

la Fig. 6 ilustra un aparato para mezclar una pluralidad de trenes de datos de entrada según una realización de la presente invención con más detalle;

la Fig. 7 ilustra un modo de operación del aparato de la Fig. 6 según una realización de la presente invención;

10

la Fig. 8 muestra un diagrama de bloques de un aparato para mezclar una pluralidad de trenes de datos de entrada según una realización adicional de la presente invención en el contexto de un sistema de conferencia;

la Fig. 9 muestra un diagrama de bloques simplificado de un aparato para generar un tren de datos de salida según una realización de la presente invención;

15

la Fig. 10 muestra un diagrama de bloques más detallado de un aparato para generar un tren de datos de salida según una realización de la presente invención;

20 la Fig. 11 muestra un diagrama de bloques de un aparato para generar un tren de datos de salida a partir de una pluralidad de trenes de datos de entrada según una realización adicional de la presente invención en el contexto de un sistema de conferencia;

la Fig. 12a ilustra una operación de un aparato para generación de trenes de datos de salida según una realización de la presente invención para una implementación PNS;

25

la Fig. 12b ilustra una operación de un aparato para generación de trenes de datos de salida según una realización de la presente invención para una implementación SBR; y

30 la Fig. 12c ilustra el funcionamiento de un aparato para generación de trenes de datos de salida según una realización de la presente invención para una implementación M/S.

[0028] Con respecto a las Fig. 4 a 12C, se describirán diferentes realizaciones según la presente invención. Sin embargo, antes de describir estas realizaciones con más detalle, primero con respecto a las Fig. 1 a 3, se ofrecerá una breve descripción a la vista de los retos y demandas que pueden convertirse en importantes en el marco de sistemas de conferencia.

35

[0029] La Fig. 1 muestra un diagrama de bloques de un sistema de conferencia 100, que también puede referirse como una unidad de control de múltiples puntos (MCU = multi-point control unit). Como será evidente a partir de la descripción referente a su funcionalidad, el sistema de conferencia 100, tal como se muestra en la Fig. 1, es un sistema que opera en el dominio del tiempo.

40

[0030] El sistema de conferencia 100, tal como se muestra en la Fig. 1, se adapta para recibir una pluralidad de trenes de datos de entrada mediante un número apropiado de entradas 110-1, 110-2, 110-3, ... de los cuales en la Fig. 1 solo se muestran tres. Cada una de las entradas 110 se acopla a un decodificador respectivo 120. Para ser más preciso, la entrada 110-1 para el primer tren de datos de entrada se acopla a un primer decodificador 120-1, mientras que la segunda entrada 110-2 se acopla a un segundo decodificador 120-2 y la tercera entrada 110-3 se acopla a un tercer decodificador 120-3.

45

[0031] El sistema de conferencia 100 comprende además un número apropiado de sumadores 130-1, 130-2, 130-3, ... de los cuales de nuevo se ilustran tres en la Fig. 1. Cada uno de los sumadores se asocia con una de las entradas 110 del sistema de conferencia 100. Por ejemplo, el primer sumador 130-1 se asocia con la primera entrada 110-1 y el decodificador correspondiente 120-1.

50

[0032] Cada uno de los sumadores 130 se acopla a las salidas de todos los decodificadores 120, aparte del decodificador 120 al cual se acopla la entrada 110. Dicho de otro modo, el primer sumador 130-1 se acopla a todos los decodificadores 120, aparte del primer decodificador 120-1. En consecuencia, el segundo sumador 130-2 se acopla a todos los decodificadores 120, aparte del segundo decodificador 120-2.

55

[0033] Cada uno de los sumadores 130 comprende además una salida, cada una de las cuales se acopla a un codificador 140. Por lo tanto, el primer sumador 130-1 se acopla de manera de salida al primer codificador 140-1. En consecuencia, el segundo y el tercer sumador 130-2, 130-3 también se acoplan al segundo y el tercer codificador 140-2, 140-3, respectivamente.

60

[0034] A su vez, cada uno de los codificadores 140 se acopla a la salida respectiva 150. Dicho de otro modo,

65

el primer codificador, por ejemplo, se acopla a una primera salida 150-1. El segundo y el tercer codificador 140-2, 140-3 también se acoplan a las salidas segunda y tercera 150-2, 150-3, respectivamente.

5 **[0035]** Para poder describir el funcionamiento de un sistema de conferencia 100 tal como se muestra en la Fig. 1 con más detalle, la Fig. 1 también muestra una terminal de conferencia 160 de un primer participante. La terminal de conferencia 160 puede ser por ejemplo, un teléfono digital (por ejemplo, un teléfono de RDSI (red digital de servicios integrados)), un sistema que comprende una infraestructura de voz-sobre-IP o una terminal similar.

10 **[0036]** La terminal de conferencia 160 comprende un codificador 170 que se acopla a la primera entrada 110-1 del sistema de conferencia 100. La terminal de conferencia 160 también comprende un decodificador 180 que se acopla a la primera salida 150-1 del sistema de conferencia 100.

15 **[0037]** También pueden estar presentes terminales de conferencia 160 similares en los sitios de otros participantes. Estas terminales de conferencia no se muestran en la Fig. 1, exclusivamente por razones de sencillez. También debe observarse que no es ni mucho menos necesario que el sistema de conferencia 100 y las terminales de conferencia 160 estén físicamente en proximidad inmediata entre sí. Las terminales de conferencia 160 y el sistema de conferencia 100 pueden disponerse en sitios diferentes que, por ejemplo, pueden conectarse solo por medio de técnicas de redes extensas (WAN = Wide Area Networks).

20 **[0038]** Las terminales de conferencia 160 pueden comprender o estar conectadas además con componentes adicionales tales como micrófonos, amplificadores y altavoces o auriculares para permitir un intercambio de señales de audio con un usuario humano en una forma más comprensible. Estos elementos no se muestran en la Fig. 1 exclusivamente por razones de sencillez.

25 **[0039]** Como se indicó anteriormente, el sistema de conferencia 100 mostrado en la Fig. 1 es un sistema que funciona en el dominio del tiempo. Cuando, por ejemplo, el primer participante habla al micrófono (no mostrado en la Fig. 1), el codificador 170 de la terminal de conferencia 160 codifica la señal de audio respectiva en un tren de bits correspondiente y transmite el tren de bits a la primera entrada 110-1 del sistema de conferencia 100.

30 **[0040]** Dentro del sistema de conferencia 100, el tren de bits es decodificado por el primer decodificador 120-1 y se transforma de nuevo al dominio del tiempo. Dado que el primer decodificador 120-1 se acopla a los mezcladores segundo y tercero 130-1, 130-3, la señal de audio, tal como es generada por el primer participante, puede mezclarse en el dominio del tiempo simplemente sumando la señal de audio reconstruida con las señales de audio reconstruidas adicionales de los participantes segundo y tercero, respectivamente.

35 **[0041]** Así sucede también para las señales de audio que proporcionan los participantes segundo y tercero que son recibidas por las entradas segunda y tercera 110-2, 110-3 y son tratadas por los decodificadores segundo y tercero 120-2, 120-3, respectivamente. Estas señales de audio reconstruidas de los participantes segundo y tercero se proporcionan entonces al primer mezclador 130-1, que a su vez proporciona la señal de audio añadida en el dominio del tiempo al primer codificador 140-1. El codificador 140-1 recodifica la señal de audio sumada para formar un tren de bits y proporciona la misma en la primera entrada 150-1 a la terminal de conferencia del primer participante 160.

40 **[0042]** Similarmente, también los codificadores segundo y tercero 140-2, 140-3 codifican las señales de audio sumadas en el dominio del tiempo recibidas de los sumadores segundo y tercero 130-2, 130-3, respectivamente, y transmiten de nuevo los datos codificados a los participantes respectivos mediante las salidas segunda y tercera 150-2, 150-3, respectivamente.

45 **[0043]** Para llevar a cabo el mezclado real, las señales de audio se decodifican completamente y se suman de forma no comprimida. Posteriormente de forma opcional puede realizarse un ajuste de nivel comprimiendo las señales de salida respectivas para evitar efectos de recorte (es decir, sobrepasar un intervalo de valores permisibles). El recorte puede aparecer cuando los valores de muestras individuales ascienden por encima o descienden por debajo del intervalo de valores permitidos, de manera que los valores correspondientes son cortados (recortados). En el caso de una cuantificación de 16 bits, tal como se usa por ejemplo en el caso de los CD, se dispone de un intervalo de valores enteros entre -32.768 y 32.767 por valor de muestra.

50 **[0044]** Para contrarrestar una posible sobredimensión o subdimensión de la señal, se emplean algoritmos de compresión. Estos algoritmos limitan el desarrollo por encima o por debajo de un cierto valor umbral para mantener los valores de muestra dentro de un intervalo de valores permisible.

55 **[0045]** Cuando se codifican datos de audio en sistemas de conferencia tales como el sistema de conferencia 100, tal como se muestra en la Fig. 1, se aceptan algunos inconvenientes para realizar un mezclado en el estado no codificado de una forma que se alcance de manera más fácil. Por otra parte, las velocidades de datos de las señales de audio codificadas están limitadas adicionalmente a un intervalo menor de frecuencias transmitidas, ya que una anchura de banda menor permite una menor frecuencia de muestreo y, por lo tanto, en los datos, según el teorema de muestreo de Nyquist-Shannon. El teorema de muestreo de Nyquist-Shannon establece que la frecuencia de

muestreo depende de la anchura de banda de la señal muestreada y se necesita que sea (al menos) el doble de grande que la anchura de banda.

5 **[0046]** La Unión Internacional de Telecomunicaciones (UIT o, en inglés, ITU = International Telecommunication Union) y su sector de normalización de telecomunicaciones (ITU-T = telecommunication standardization sector) han desarrollado varias normas para sistemas de conferencia multimedia. H.320 es el protocolo de conferencia estándar para RDSI. H.323 define el sistema de conferencia estándar para una red basada en paquetes (TCP/IP). H.324 define sistemas de conferencia para redes de telefonía analógica y sistemas de radiotelecomunicaciones.

10 **[0047]** Dentro de estas normas, no solo se define la transmisión de las señales, sino también la codificación y el tratamiento de los datos de audio. El manejo de una conferencia se realiza por medio uno o más servidores, las denominadas unidades de control de múltiples puntos (MCU = multi-point control units) según la norma H.231. Las unidades de control de múltiples puntos también son responsables del tratamiento y la distribución de los datos de vídeo y de audio de los diversos participantes.

15 **[0048]** Para lograr lo anterior, la unidad de control de múltiples puntos envía a cada participante una salida mezclada o señal resultante que comprende los datos de audio de todos los otros participantes y proporciona la señal a los participantes respectivos. La Fig. 1 no solo muestra un diagrama de bloques de un sistema de conferencia 100, sino también un flujo de señal en esta situación de conferencia.

20 **[0049]** En el marco de las normas H.323 y H.320, los códecs de audio de la clase G.7xx se definen para funcionamiento en los sistemas de conferencia respectivos. La norma G.711 se usa para transmisiones RDSI, en sistemas de telefonía por cable. Para una frecuencia de muestreo de 8 kHz, la norma G.711 cubre una anchura de banda de audio entre 300 y 3.400 Hz, lo que requiere una velocidad binaria de 64 kbit/s a una profundidad
25 (cuantificación) de 8 bits. La codificación se forma por medio de una simple codificación logarítmica denominada Ley- μ o Ley-A que crea un retardo muy bajo de solo 0,125 ms.

[0050] La norma G.722 codifica una mayor anchura de banda de audio de 50 a 7.000 Hz a una frecuencia de muestreo de 16 kHz. Como consecuencia, el códec logra una mejor calidad cuando se compara con los códecs de
30 audio G.7xx de banda más estrecha a velocidades de bits de 48, 56 o 64 Kbit/s para un retardo de 1,5 ms. Por otra parte, existen otros dos desarrollos, G.722.1 y G.722.2, que proporcionan calidad de habla comparable a velocidades binarias incluso menores. La G.722.2 permite una selección de velocidad binaria entre 6,6 kbit/s y 23,85 kbit/s para un retardo de 25 ms.

35 **[0051]** La norma G.729 se usa normalmente en el caso de comunicaciones de telefonía-IP, que también se refiere como comunicaciones de voz sobre IP (VoIP). El códec se optimiza para habla y transmite un conjunto de parámetros de habla analizados para una síntesis posterior junto con una señal de error. Como resultado, G.729 logra una codificación significativamente mejor de aproximadamente 8 kbits/s a una velocidad de muestreo y una anchura de banda de audio comparables, cuando se comparan con la norma G.711. El algoritmo, más complejo, crea sin
40 embargo un retardo de aproximadamente 15 ms.

[0052] Como inconveniente, los códecs G.7.xx se optimizan para codificación de habla y muestran, aparte de una anchura de banda de frecuencia estrecha, problemas importantes cuando se codifica música junto con habla o
45 música pura.

[0053] Por lo tanto, aunque el sistema de conferencia 100, tal como se muestra en la Fig. 1, puede usarse para una calidad aceptable cuando se transmiten y procesan señales de habla, las señales de audio en general no se procesan satisfactoriamente cuando se emplean códecs de bajo retardo optimizados para habla.

50 **[0054]** Dicho de otro modo, el uso de códecs para codificar y decodificar una señal de habla con el fin de procesar señales de audio en general, incluyendo por ejemplo señales de audio con música, no lleva a un resultado satisfactorio en términos de calidad. Al usar códecs de audio para codificar y decodificar señales de audio en general en el marco del sistema de conferencia 100, tal como se muestra en la Fig. 1, la calidad es mejorable. Sin embargo, como se expondrá en el contexto de la Fig. 2 con más detalle, el uso de códecs de audio en general en este sistema
55 de conferencia puede llevar a efectos adicionales indeseados, tales como un mayor retardo, por citar solo uno.

[0055] Sin embargo, antes de describir la Fig. 2 con más detalle, debe observarse que en la presente descripción, los objetos se denotan con signos de referencia iguales o similares cuando los objetos respectivos aparecen más de una vez en una realización o una figura, o aparecen en varias realizaciones o figuras. A menos que
60 se indique lo contrario de forma explícita o implícita, los objetos denotados por signos de referencia iguales o similares pueden implementarse de una forma similar o igual, por ejemplo, en relación con sus circuitos, programación, características u otros parámetros. Por lo tanto, los objetos que aparecen en varias realizaciones de las figuras y que se denotan con signos de referencia iguales o similares pueden ser implementados de manera que tengan las mismas especificaciones, parámetros y características. Naturalmente, también pueden implementarse desviaciones y
65 adaptaciones, por ejemplo, cuando las condiciones frontera u otros parámetros cambian de una figura a otra, o de una

realización a otra.

- 5 **[0056]** Por otra parte, a continuación se usarán signos de referencia de resumen para denotar un grupo o clase de objetos, en vez de un objeto individual. En el marco de la Fig. 1, esto ya se ha hecho, por ejemplo, cuando se denota la primera entrada como entrada 110-1, la segunda entrada como entrada 110-2, y la tercera entrada como entrada 110-3, mientras que las entradas se han expuesto en términos del signo de referencia de resumen 110 solamente. Dicho de otro modo, a menos de que se indique explícitamente lo contrario, partes de la descripción que se refieren a objetos denotados con signos de referencia y resumen también pueden relacionarse con otros objetos que contienen los signos de referencia individuales correspondientes.
- 10 **[0057]** Dado que esto también se aplica a objetos denotados con signos de referencia iguales o similares, las dos medidas ayudan a recortar la descripción y a describir las realizaciones ilustradas en la presente memoria de una forma más clara y concisa.
- 15 **[0058]** La Fig. 2 muestra un diagrama de bloques de un sistema de conferencia tradicional 100 junto con una terminal de conferencia 160, los dos similares a los mostrados en la Fig. 1. El sistema de conferencia 100 ilustrado en la Fig. 2 también comprende entradas 110, decodificadores 120, sumadores 130, codificadores 140 y salidas 150, que están interconectados del mismo modo con el sistema de conferencia 100 tal como se ilustra en la Fig. 1. La terminal de conferencia 160 mostrada en la Fig. 2 también comprende de nuevo un codificador 170 y un decodificador 180.
- 20 **[0059]** Por lo tanto, se hace referencia a la descripción del sistema de conferencia 100 mostrado en la Fig. 1.
- 25 **[0060]** Sin embargo, el sistema de conferencia 100 ilustrado en la Fig. 2, así como la terminal de conferencia 160 mostrada en la Fig. 2 se adaptan para usar un códec (COdificador-DECodificador) de audio general. Como consecuencia, cada uno de los codificadores 140, 170, comprende una serie de conexiones de un convertidor de tiempo/frecuencia 190 acoplado antes de un cuantificador/codificador 200. El convertidor de tiempo/frecuencia 190 también se ilustra en la Fig. 2 como "T/F", mientras que el cuantificador/codificador 200 se refiere en la Fig. 2 como "Q/C".
- 30 **[0061]** Los decodificadores 120, 180 comprenden cada uno un decodificador/decuantificador 210, que se refiere en la Fig. 2 como "Q/C-1" conectado en serie con un convertidor de frecuencia/tiempo 220, que se refiere en la Fig. 2 como "T/F-1". Por razones de sencillez exclusivamente, el convertidor de tiempo/frecuencia 190, el cuantificador/codificador 200 y el decodificador/decuantificador 210, así como el convertidor de frecuencia/tiempo 220, se etiquetan como tales solo en el caso del codificador 140-3 y el decodificador 120-3. Sin embargo, la siguiente descripción también se refiere a los demás de estos elementos.
- 35 **[0062]** Empezando con un codificador como por ejemplo los codificadores 140, o el codificador 170, la señal de audio que se proporciona al convertidor de tiempo/frecuencia 190 es convertida desde el dominio del tiempo a un dominio de la frecuencia o un dominio relacionado con la frecuencia por el convertidor 190. Posteriormente, los datos de audio convertidos están, en una representación espectral generada por el convertidor de tiempo/frecuencia 190, cuantificados y codificados para formar un tren de bits, que después se proporciona, por ejemplo, a las salidas 150 del sistema de conferencia 100 en el caso del codificador 140.
- 40 **[0063]** En términos de los decodificadores tales como los decodificadores 120 o el decodificador 180, el tren de bits que se proporciona a los decodificadores primero se decodifica y recuantifica para formar la representación espectral de al menos una parte de una señal de audio, que después se convierte de nuevo al dominio del tiempo por los convertidores de frecuencia/tiempo 220.
- 45 **[0064]** Los convertidores de tiempo/frecuencia 190, así como los elementos inversos, los convertidores de frecuencia/tiempo 220 están adaptados por lo tanto para generar una representación espectral de al menos una pieza de una señal de audio que se le proporciona y para retransformar la parte representativa espectral en las partes correspondientes de la señal de audio en el dominio del tiempo, respectivamente.
- 50 **[0065]** En el proceso de convertir una señal de audio del dominio del tiempo al dominio de la frecuencia, y de nuevo del dominio de la frecuencia al dominio del tiempo, pueden producirse desviaciones de manera que la señal de audio reestablecida, reconstruida o decodificada puede diferir de la señal de audio original o fuente. Pueden añadirse otros artefactos en las etapas adicionales de cuantificación y decuantificación realizadas en el marco del codificador cuantificador 200 y el recodificador 210. Dicho de otro modo, la señal de audio original, así como la señal de audio reestablecida, pueden diferir entre sí.
- 55 **[0066]** Los convertidores de tiempo/frecuencia 190, así como los convertidores de frecuencia/tiempo 220 pueden implementarse, por ejemplo, basándose en una transformación coseno discreta modificada (MDCT = Modified Discrete Cosine Transformation), una transformación de seno discreta modificada (MDST = Modified Discrete Sine Transformation), un convertidor basado en transformada de Fourier rápida (FFT = Fast Fourier Transformation), u otro convertidor basado en Fourier. La cuantificación y la recuantificación en el marco del cuantificador/codificador 200 y
- 60 **[0066]** Los convertidores de tiempo/frecuencia 190, así como los convertidores de frecuencia/tiempo 220 pueden implementarse, por ejemplo, basándose en una transformación coseno discreta modificada (MDCT = Modified Discrete Cosine Transformation), una transformación de seno discreta modificada (MDST = Modified Discrete Sine Transformation), un convertidor basado en transformada de Fourier rápida (FFT = Fast Fourier Transformation), u otro convertidor basado en Fourier. La cuantificación y la recuantificación en el marco del cuantificador/codificador 200 y
- 65 **[0066]** Los convertidores de tiempo/frecuencia 190, así como los convertidores de frecuencia/tiempo 220 pueden implementarse, por ejemplo, basándose en una transformación coseno discreta modificada (MDCT = Modified Discrete Cosine Transformation), una transformación de seno discreta modificada (MDST = Modified Discrete Sine Transformation), un convertidor basado en transformada de Fourier rápida (FFT = Fast Fourier Transformation), u otro convertidor basado en Fourier. La cuantificación y la recuantificación en el marco del cuantificador/codificador 200 y

el decodificador/decuantificador 210 pueden implementarse, por ejemplo, basándose en una cuantificación lineal, una cuantificación logarítmica, u otro algoritmo de cuantificación más complejo, por ejemplo, teniendo en cuenta más específicamente las características de audición del ser humano. Las partes de codificador y decodificador del cuantificador/codificador 200 y el decodificador/decuantificador 210 pueden funcionar, por ejemplo, usando un esquema de codificación Huffman o decodificación Huffman.

[0067] Sin embargo, también pueden emplearse convertidores más complejos de tiempo/frecuencia y frecuencia/tiempo 190, 220, así como cuantificador/codificador y decodificador/decuantificador más complejos 200, 210 en diferentes realizaciones y sistemas tal como se describe en la presente memoria, siendo parte de o formando, por ejemplo, un codificador AAC-ELD como codificadores 140, 170, y un decodificador AAC-ELD como decodificadores 120, 180.

[0068] No es necesario decir que puede ser recomendable implementar codificadores 170, 140 y decodificadores 180, 120 idénticos o al menos compatibles, en el marco del sistema de conferencia 100 y la terminal de conferencia 160.

[0069] El sistema de conferencia 100, tal como se muestra en la Fig. 2, basado en un esquema de codificación y decodificación de señal de audio, también realiza en general el mezclado real de las señales de audio en el dominio del tiempo. Los sumadores 130 se proporcionan con las señales de audio reconstruidas en el dominio del tiempo para realizar una superposición y proporcionar las señales mixtas en el dominio del tiempo a los convertidores de tiempo/frecuencia 190 de los siguientes codificadores 140. Por lo tanto, el sistema de conferencia comprende de nuevo una conexión en serie de los decodificadores 120 y codificadores 140, que es la razón por la que un sistema de conferencia 100, tal como se muestra en las Fig. 1 y 2, normalmente se refiere como "sistemas de codificación en tándem".

[0070] Los sistemas de codificación en tándem a menudo muestran el inconveniente de una alta complejidad. La complejidad de un mezclado intenso depende de la complejidad de los decodificadores y codificadores empleados, y puede multiplicarse de forma importante en el caso de varias señales de entrada de audio y salida de audio. Por otra parte, debido al hecho de que la mayoría de los esquemas de codificación y decodificación no son sin pérdidas de datos, el esquema de codificación en tándem, tal como se usa en los sistemas de conferencia 100, mostrados en las Fig. 1 y 2, normalmente conduce a una influencia negativa en la calidad.

[0071] Como un inconveniente adicional, las etapas repetidas de decodificación y codificación también amplían el retardo total entre las entradas 110 y las salidas 150 del sistema de conferencia 100, que también se refiere como retardo de extremo a extremo. Dependiendo del retardo inicial de los decodificadores y codificadores empleados, el propio sistema de conferencia 100, puede incrementar el retardo hasta un nivel que hace no atractivo el uso en el marco del sistema de conferencia, cuando no perjudicial o incluso imposible. A menudo se considera que un retardo de aproximadamente 50 ms es el retardo máximo que pueden aceptar los participantes en las conversaciones.

[0072] Como fuentes principales para el retardo, los convertidores de tiempo/frecuencia 190, así como los convertidores de frecuencia/tiempo 220, son responsables del retardo de extremo a extremo del sistema de conferencia 100, y el retardo adicional impuesto por las terminales de conferencia 160. El retardo provocado por los elementos adicionales, es decir, los cuantificadores/codificadores 200 y los decodificadores/decuantificadores 210, es de menor importancia ya que estos componentes pueden operar a una frecuencia muy superior en comparación con los convertidores de tiempo/frecuencia y los convertidores de frecuencia/tiempo 190, 220. La mayoría de los convertidores de tiempo/frecuencias y los convertidores de frecuencia/tiempo 190, 220 operan por bloques u operan por tramas, lo que significa que en muchos casos debe tenerse en cuenta un retardo mínimo como una cantidad de tiempo, lo que equivale al tiempo requerido para llenar una memoria intermedia o una memoria que tiene la longitud de trama de un bloque. Este tiempo, sin embargo, se ve influido significativamente por la frecuencia de muestreo que normalmente está en el intervalo de unos kHz a varios 10 kHz, mientras que la velocidad operativa de los cuantificadores/codificadores 200, así como el decodificador/decuantificador 210, se determinan principalmente por medio de la frecuencia de reloj del sistema subyacente. Este valor es normalmente al menos 2, 3, 4 o más órdenes de magnitud mayores.

[0073] Por lo tanto, en sistemas de conferencia que emplean códecs de señal de audio generales se ha introducido la llamada tecnología de mezclado de trenes de bits. El procedimiento de mezclado de trenes de bits puede implementarse, por ejemplo, basándose en el códec MPEG-4 AAC-ELD, que ofrece la posibilidad de evitar al menos algunos de los inconvenientes mencionados anteriormente e introducidos por codificación en tándem.

[0074] Sin embargo debe observarse que, en principio, el sistema de conferencia 100 tal como se muestra en la Fig. 2, también puede implementarse basándose en el códec MPEG-4 AAC-ELD con una velocidad binaria similar y una anchura de banda de frecuencia significativamente mayor, en comparación con los códecs basados en habla mencionados anteriormente de la familia de códecs G.7xx. Esto implica también inmediatamente que puede lograrse una calidad de audio significativamente mejor para todos los tipos de señal al coste de una velocidad binaria significativamente incrementada. Aunque MPEG-4 AAC-ELD ofrece un retardo que está en el intervalo del

correspondiente al códec G.7xx, implementando el mismo en el marco de un sistema de conferencia tal como se muestra en la Fig. 2, puede no llevar a un sistema de conferencia práctico 100. A continuación, con respecto a la Fig. 3, se expondrá un sistema más práctico basado en el denominado mezclado de trenes de bits mencionado anteriormente.

5 **[0075]** Debe observarse que, por razones de sencillez exclusivamente, el foco se dirigirá principalmente al códec MPEG-4 AAC-ELD y sus trenes de datos y trenes de bits. Sin embargo, también pueden emplearse otros codificadores y decodificadores en el entorno de un sistema de conferencia 100 tal como se ilustra y muestra en la Fig. 3.

10 **[0076]** La Fig. 3 muestra un diagrama de bloques de un sistema de conferencia 100 que funciona según el mezclado de trenes de bits junto con una terminal de conferencia 160, tal como se describe en el contexto de la Fig. 2. El sistema de conferencia 100 en sí es una versión simplificada del sistema de conferencia 100 mostrado en la Fig. 2. Para ser más precisos, los decodificadores 120 del sistema de conferencia 100 en la Fig. 2 se han sustituido por
 15 decodificadores/decuantificadores 220-1, 220-2, 210-3, ... tal como se muestra en la Fig. 3. Dicho de otro modo, los convertidores de frecuencia/tiempo 120 de los decodificadores 120 se han retirado cuando se compara el sistema de conferencia 100 ilustrado en las Fig. 2 y 3. De manera similar, los codificadores 140 del sistema de conferencia 100 de la Fig. 2 se han sustituido por los cuantificadores/codificadores 200-1, 200-2, 200-3. Por lo tanto, los convertidores de tiempo/frecuencia 190 de los codificadores 140 se han retirado cuando se compara el sistema de conferencia 100
 20 mostrado en las Fig. 2 y 3.

[0077] Como resultado, los sumadores 130 no operan ya en el dominio del tiempo, sino, debido a la falta de los convertidores de frecuencia/tiempo 220 y los convertidores de tiempo/frecuencia 190, en la frecuencia o en un dominio relacionado con la frecuencia.

25 **[0078]** Por ejemplo, en el caso de los códecs MPEG-4 AAC-ELD, el convertidor de tiempo/frecuencia 190 y el convertidor de frecuencia/tiempo 220, que solo están presentes en las terminales de conferencia 160, se basan en una transformación MDCT. Por lo tanto, dentro del sistema de conferencia 100, los mezcladores 130 operan directamente en las contribuciones de las señales de audio en la representación de frecuencia MDCT.

30 **[0079]** Dado que los convertidores 190, 220 representan la fuente principal de retardo en el caso del sistema de conferencia 100 mostrado en la Fig. 2, el retardo se reduce significativamente al retirar estos convertidores 190, 220. Por otra parte, la complejidad introducida por los dos convertidores 190, 220 dentro del sistema de conferencia 100 también se reduce significativamente. Por ejemplo, en el caso de un decodificador MPEG-2 AAC, la
 35 transformación MDCT inversa que se lleva a cabo en el marco del convertidor de frecuencia/tiempo 220 es responsable en aproximadamente el 20% de la complejidad total. Dado que también el convertidor MPEG-4 se basa en una transformación similar, puede retirarse una contribución no irrelevante a la complejidad total al eliminar el convertidor de frecuencia/tiempo 220 solo del sistema de conferencia 100.

40 **[0080]** Es posible mezclar las señales de audio en el dominio MDCT, u otro dominio de la frecuencia, ya que en el caso de una transformación MDCT o en el caso de una transformación basada en Fourier similar, estas transformaciones son transformaciones lineales. Las transformaciones, por lo tanto, poseen la propiedad de la aditividad matemática, es decir,

45
$$f(x + y) = f(x) + f(y) \quad , \quad (1)$$

y la de homogeneidad matemática, es decir

50
$$f(a \cdot x) = a \cdot f(x) \quad , \quad (2)$$

en las que $f(x)$ es una función de transformación, x e y son sus argumentos convenientes y a es un valor real o una constante de valor complejo.

[0081] Las dos características de la transformación MDCT u otra transformación basada en Fourier permiten
 55 un mezclado en el dominio de la frecuencia respectivo similar al mezclado en el dominio del tiempo. Por lo tanto, todos los cálculos pueden llevarse a cabo igualmente basándose en valores espectrales. No se necesita una transformación de los datos en el dominio del tiempo.

[0082] En algunas circunstancias se puede requerir que se cumpla una condición adicional. Todos los datos
 60 espectrales relevantes deberán ser iguales en lo que respecta a sus índices de tiempo durante el proceso de mezclado para todos los componentes espectrales relevantes. Es posible que esto no suceda si, durante la transformación se usa la denominada técnica de conmutación de bloques de manera que el codificador de las terminales de conferencia 160 puede conmutar libremente entre diferentes longitudes de bloque, dependiendo de ciertas condiciones. La

conmutación de bloques puede poner en peligro la posibilidad de asignar de forma única valores espectrales individuales a muestras en el dominio del tiempo debido a la conmutación entre diferentes longitudes de bloques y longitudes de ventanas MDCT correspondientes, a menos que los datos al mezclarse hayan sido procesados con las mismas ventanas. Dado que, en un sistema general con terminales de conferencia distribuidas 160, esto puede no 5 garantizarse, pueden ser necesarias interpolaciones complejas que a su vez pueden crear retardo y complejidad adicionales. Como consecuencia, puede ser recomendable en su caso no implementar un proceso de mezclado de trenes de bits basándose en conmutación de longitudes o tramos de bloques.

[0083] En cambio, el códec AAC-ELD se basa en una longitud de un solo bloque y, por lo tanto, es capaz de 10 garantizar más fácilmente la asignación o sincronización de datos de frecuencia descrita anteriormente, de manera que pueda lograrse el mezclado más fácilmente. El sistema de conferencia 100 mostrada en la Fig. 3 es, dicho de otro modo, un sistema que es capaz de realizar el mezclado en el dominio de transformación o el dominio de la frecuencia.

[0084] Tal como se expuso anteriormente, con el fin de eliminar el retardo adicional introducido por los 15 convertidores 190, 200 en el sistema de conferencia 100 mostrado en la Fig. 2, los códec empleados en las terminales de conferencia 160 usan una ventana de longitud y forma fijas. Esto permite la implementación del proceso de mezclado descrito directamente sin transformar el tren de audio de nuevo al dominio del tiempo. Este enfoque es capaz de limitar la cantidad de retardo algorítmico introducido adicionalmente. Por otra parte, la complejidad se reduce debido a la ausencia de las etapas de transformación inversa en el decodificador y las etapas de transformación directa 20 en el codificador.

[0085] Sin embargo, también en el marco de un sistema de conferencia 100 tal como se muestra en la Fig. 3, puede ser necesario volver a cuantificar o recuantificar los datos de audio después del mezclado por los sumadores 130, que puede introducir ruido de cuantificación adicional. El ruido de cuantificación adicional, por ejemplo, puede 25 crearse debido a las diferentes etapas de cuantificación o diferentes señales de audio que se proporcionan al sistema de conferencia 100. Como consecuencia, por ejemplo, en el caso de muy bajas transmisiones de velocidad binaria en las que una serie de etapas de cuantificación ya están limitadas, el proceso de mezclar dos señales de audio en el dominio de la frecuencia o el dominio de transformación puede producir una cantidad adicional indeseable de ruido u otras distorsiones en la señal generada.

[0086] Antes de describir una primera realización según la presente invención en la forma de un aparato para 30 mezclar una pluralidad de trenes de datos de entrada, con respecto a la Fig. 4, se describirá brevemente un tren de datos o tren de bits, junto con los datos comprendidos en el mismo.

[0087] La Fig. 4 muestra esquemáticamente un tren de bits o tren de datos 250, que comprende al menos una, 35 o más a menudo, más de una trama 260 de datos de audio en un dominio espectral. De forma más precisa, la Fig. 4 muestra tres tramas 260-1, 260-2 y 260-3 de datos de audio en un dominio espectral. Por otra parte, el tren de datos 250 también puede comprender información adicional o bloques de información adicional 270, tales como valores de control que indican, por ejemplo, una forma en la que se codifican los datos de audio, otros valores de control de información referente a índices de tiempo u otros datos relevantes. Naturalmente, el tren de datos 250 tal como se 40 muestra en la Fig. 4 puede comprender además tramas adicionales o una trama 260 puede comprender datos de audio de más de un canal. Por ejemplo, en el caso de una señal de audio en estéreo, cada una de las tramas 260 puede comprender, por ejemplo, datos de audio de un canal izquierdo, un canal derecho, datos de audio derivados de ambos, los canales izquierdo y derecho o cualquier combinación de los datos mencionados anteriormente.

[0088] Por lo tanto, la Fig. 4 ilustra que un tren de datos 250 puede no solo comprender una trama de datos de 45 audio en un dominio espectral, sino también información de control adicional, valores de control, valores de estado, información de estado, valores relacionados a protocolo (por ejemplo, sumas de verificación), o similares.

[0089] Dependiendo de la implementación concreta del sistema de conferencia tal como se describe en el 50 contexto de las Fig. 1 a 3, o dependiendo de la implementación concreta de un aparato según una realización de la presente invención, como se describirá a continuación, en particular según los descritos con respecto a las Fig. 9 a 12C, los valores de control que indican una forma asociada a datos de carga útil de la trama y representan al menos una parte del dominio espectral o información espectral de una señal de audio pueden estar comprendidos igualmente 55 en las propias tramas 260, o en el bloque asociado 270 de información adicional. En el caso en que los valores de control se relacionan con componentes espectrales, los valores de control pueden estar codificados en las tramas 260 en sí. Sin embargo, si un valor de control se refiere a una trama entera, puede estar comprendido igualmente en los bloques 270 de información adicional. Sin embargo, en buena medida no se necesita que los sitios mencionados anteriormente para incluir los valores de control tal como se describió anteriormente estén comprendidos en las tramas 60 260 o el bloque 270 de los bloques adicionales. En el caso en que un valor de control se refiere solo a uno o a algunos componentes espectrales, también puede estar comprendido en el bloque 270. Por otra parte, un valor de control referente a una trama entero 260 también puede estar comprendido en las tramas 260.

[0090] La Fig. 5 ilustra esquemáticamente información (espectral) que se refiere a componentes espectrales, 65 por ejemplo, comprendidos en la trama 260 del tren de datos 250. Para ser más precisos, la Fig. 5 muestra un diagrama

de información simplificado en un dominio espectral de un solo canal de una trama 260. En el dominio espectral, una trama de datos de audio puede describirse, por ejemplo, en términos de sus valores de intensidad I como una función de la frecuencia f . En sistemas discretos como, por ejemplo, sistemas digitales, también la resolución de frecuencia es discreta, de manera que la información espectral normalmente solo está presente para ciertos componentes espectrales tales como frecuencias individuales o bandas o sub-bandas estrechas. Las frecuencias individuales o bandas estrechas, así como las sub-bandas, se refieren como componentes espectrales.

[0091] La Fig. 5 muestra esquemáticamente una distribución de intensidad para seis frecuencias individuales 300-1, ..., 300-6, así como una banda o sub-banda de frecuencias 310 que comprende, en el caso que se ilustra en la Fig. 5, cuatro frecuencias individuales. Ambas frecuencias individuales o bandas estrechas correspondientes 300, así como la sub-banda o banda de frecuencias 310, forman componentes espectrales con respecto a los cuales la trama comprende información referente a los datos de audio en el dominio espectral.

[0092] La información referente a la sub-banda 310, por ejemplo, puede ser una intensidad total, o un valor de intensidad promedio. Aparte de la intensidad u otros valores relacionados con la energía tales como la amplitud, la energía del propio componente espectral respectivo u otro valor derivado de la energía o la amplitud, la información de fase u otra información también pueden estar comprendidas en la trama y por lo tanto ser consideradas información referente a un componente espectral.

[0093] Después de haber descrito algunos de los problemas implicados y algún antecedente para sistemas de conferencia, se describen realizaciones según un primer aspecto de la presente invención según las cuales un tren de datos de entrada se determina basándose en una comparación con el fin de copiar al menos información espectral parcial del tren de datos de entrada determinado en el tren de datos de salida, permitiendo así omitir una recuantificación y, por lo tanto, el ruido de recuantificación asociado.

[0094] La Fig. 6 muestra un diagrama de bloques de un aparato 500 para mezclar una pluralidad de trenes de datos de entrada 510, de los cuales se ilustran dos 510-1, 510-2. El aparato 500 comprende una unidad de tratamiento 520 que se adapta para recibir los trenes de datos 510 y para generar un tren de datos de salida 530. Cada uno de los trenes de datos de entrada 510-1, 510-2 comprende una trama 540-1, 540-2, respectivamente, que es similar a la trama 260 mostrada en la Fig. 4 en el contexto de la Fig. 5, y comprende datos de audio en un dominio espectral. Este hecho se ilustra de nuevo mediante un sistema de coordenadas mostrado en la Fig. 6 del cual se muestra, como abscisa, la frecuencia f y, como ordenada, la intensidad I . El tren de datos de salida 530 también comprende una trama de salida 550 que comprende datos de audio en un dominio espectral, y también se ilustra mediante un sistema de coordenadas correspondiente.

[0095] La unidad de tratamiento 520 se adapta para comparar las tramas 540-1, 540-2 de una pluralidad de trenes de datos de entrada 510. Como se expondrá con más detalle a continuación, esta comparación puede basarse, por ejemplo, en un modelo psicoacústico, teniendo en cuenta efectos de enmascaramiento y otras propiedades de las características de audición humana. Basándose en este resultado de comparación, la unidad de tratamiento 520 se adapta además para determinar al menos un componente espectral, por ejemplo, los componentes espectrales 560 mostrados en la Fig. 6, que está presente en las dos tramas 540-1, 540-2, exactamente un tren de datos en la pluralidad de trenes de datos 510. Después, la unidad de tratamiento 520 puede adaptarse para generar el tren de datos de salida 530, que comprende la trama de salida 550, de manera que se copia una información referente al componente espectral 560 de la trama determinada 540 del tren de datos de entrada respectivo 510.

[0096] Para ser más precisos, la unidad de tratamiento 520 se adapta de manera que la comparación de la trama 540 de la pluralidad de trenes de datos de entrada 510 se basa en al menos dos elementos de información (los valores de intensidad son valores de energía relacionados) correspondientes al mismo componente espectral 560 de las tramas 540 de dos trenes de datos de entrada diferentes 510.

[0097] Para ilustrar esto de forma adicional, la Fig. 7 muestra esquemáticamente el caso en que el elemento de información (la intensidad I), correspondiente a los componentes espectrales 560, que se considera en este caso una frecuencia o una banda de frecuencia estrecha de la trama 540-1 de un primer tren de datos de entrada 510-1. Esto se compara con el valor de intensidad correspondiente I , que es el elemento de información referente al componente espectral 560 de la trama 540-2 de la segunda tren de datos de entrada 510-2. La comparación, por ejemplo, puede realizarse basándose en la evaluación de una proporción de energía entre la señal mixta en la que se incluyen solo algunos trenes de entrada y una señal mixta completa. Esto puede lograrse, por ejemplo, según

$$E_c = \sum_{n=1}^N E_n \quad (3)$$

60
y

$$E_{f^{(n)}} = \sum_{\substack{n=1 \\ n \neq 1}}^N E_i \quad (4)$$

y calculando la proporción $r(n)$ según

$$r(n) = 20 \cdot \log \frac{E_{f^{(n)}}}{E_c} \quad (5)$$

5

en la que n es un índice de un tren de datos de entrada y N es el número de todos los trenes de datos de entrada relevantes. Si la proporción $r(n)$ es lo suficientemente alta, los canales menos dominantes o las tramas menos dominantes de los trenes de datos de entrada 510 pueden verse como enmascarados por los dominantes. De esta manera, puede procesarse una reducción de irrelevancia, lo que significa que se incluyen solo aquellos componentes espectrales de un tren que son todos apreciables, mientras que los otros trenes se descartan.

[0098] Los valores de energía que se consideran en el marco de las ecuaciones (3) a (5), por ejemplo, pueden deducirse de los valores de intensidad tal como se muestra en la Fig. 6, calculando el cuadrado de los valores de intensidad respectivos. En el caso en que la información referente a los componentes espectrales pueda comprender otros valores, puede realizarse un cálculo similar dependiendo de la forma de información comprendida en la trama 510. Por ejemplo, en el caso de información de valor complejo, podrá tener que realizarse el cálculo del módulo de los componentes reales e imaginarios de los valores individuales que constituyen la información referente a los componentes espectrales.

20

[0099] Aparte de las frecuencias individuales, para la aplicación del módulo psicoacústico según las ecuaciones (3) a (5), la suma en las ecuaciones (3) y (4) puede comprender más de una frecuencia. Dicho de otro modo, en las ecuaciones (3) y (4) los valores de energía respectivos E_n pueden ser sustituidos por un valor de energía total que corresponde a una pluralidad de frecuencias individuales, una energía de una banda de frecuencia, o para expresarlo en términos más generales, por un único elemento de información espectral o una pluralidad de información espectral referente a uno o más componentes espectrales.

25

[0100] Por ejemplo, dado que AAC-ELD opera en líneas espectrales en forma de bandas, similar a los grupos de frecuencia en los cuales el sistema auditivo humano realiza el tratamiento al mismo tiempo, la estimación de irrelevancia o el modelo psicoacústico puede llevarse a cabo de manera similar. Al aplicar el modelo psicoacústico de esta manera es posible retirar o sustituir parte de una señal de solo una banda de frecuencia, si fuera necesario.

30

[0101] Como han mostrado los exámenes psicoacústicos, el enmascaramiento de una señal por otra señal depende de los tipos de señales respectivas. Como un umbral mínimo para una determinación de irrelevancia, puede aplicarse un escenario del peor caso. Por ejemplo, para enmascarar ruido mediante una senoide u otro sonido distinto y bien definido se necesita normalmente una diferencia de 21 a 28 dB. Las pruebas han demostrado que un valor umbral de aproximadamente 28,5 dB produce buenos resultados sustitutos. Este valor puede mejorarse, en su caso, teniendo en cuenta también las bandas de frecuencia reales.

35

[0102] Por lo tanto, los valores $r(n)$ según la ecuación (5) que son mayores que -28,5 dB pueden considerarse irrelevantes en términos de una evaluación psicoacústica o de una evaluación de irrelevancia basada en el componente espectral o los componentes espectrales en consideración. Para diferentes componentes espectrales pueden emplearse distintos valores. De esta manera, puede considerarse útil usar umbrales como indicadores para una irrelevancia psicoacústica de un tren de datos de entrada en términos de la trama en consideración de 10 dB a 40 dB, 20 dB a 30 dB o 25 dB a 30 dB.

45

[0103] En la situación mostrada en la Fig. 7, esto significa que con respecto al componente espectral 560, se determina el primer tren de datos de entrada 510-1, mientras que se descarta el segundo tren de datos de entrada 510-2 con respecto al componente espectral 560. Como consecuencia, el elemento de información referente al componente espectral 560 se copia al menos parcialmente de la trama 540-1 del primer tren de datos de entrada 510-1 a la trama de salida 550 del tren de datos de salida 530. Así se ilustra en la Fig. 7 mediante una flecha 570. Al mismo tiempo, los elementos de información referentes a los componentes espectrales 560 de la trama 540 de los otros trenes de datos de entrada 510 (es decir, en la Fig. 7, trama 540-2 del tren de datos de entrada 510-2) se descartan como se ilustra mediante la línea discontinua 580.

55

[0104] Todavía dicho de otro modo, el aparato 500 que, por ejemplo, puede emplearse como una MCU o un sistema de conferencia 100, se adapta de manera que se genera el tren de datos de salida 530 junto con su trama de salida 550, de manera que la información del componente espectral correspondiente se copia solo de la trama 540-1 del tren de datos de entrada determinada 510-1 que describe el componente espectral 560 del tren de salida 550 del

tren de datos de salida 530. Naturalmente, el aparato 500 también puede adaptarse de manera que la información referente a más de un componente espectral pueda copiarse desde un tren de datos de entrada, descartando los otros trenes de datos de entrada, al menos con respecto a estos componentes espectrales. Además es posible que un aparato 500, o su unidad de tratamiento 520, se adapte de tal forma que para diferentes componentes espectrales se determinan diferentes trenes de datos de entrada 510. La misma trama de salida 550 del tren de datos de salida 530 puede comprender información espectral copiada referente a diferentes componentes espectrales de diferentes trenes de datos de entrada 510.

[0105] Naturalmente, puede ser recomendable implementar el aparato 500 de manera que en el caso de una secuencia de tramas 540 en un tren de datos de entrada 510, solo se considerarán las tramas 540 durante la comparación y la determinación, lo que corresponde a un índice de tiempo similar o igual.

[0106] Dicho de otro modo, la Fig. 7 ilustra los principios operativos de un aparato para mezclar una pluralidad de trenes de datos de entrada como se describió anteriormente según una realización. Como se expuso anteriormente, el mezclado no se realiza de una forma directa en el sentido de que todos los trenes de entrada se decodifican, lo que incluye una transformación inversa al dominio del tiempo, mezclado y de nuevo recodificación de las señales.

[0107] Las realizaciones de las Fig. 6 a 8 se basan en el mezclado realizado en el dominio de la frecuencia del códec respectivo. Un códec posible puede ser el códec AAC-ELD, o cualquier otro códec con una ventana de transformación uniforme. En este caso, no se necesita transformación de tiempo/frecuencia para poder mezclar los datos respectivos. Las realizaciones según una realización de la presente invención hacen uso del hecho de que es posible acceder a todos los parámetros del tren de bits, tales como el tamaño de la etapa de cuantificación y otros parámetros, y que estos parámetros pueden emplearse para generar un tren de bits de salida mixto.

[0108] Las realizaciones de las Fig. 6 a 8 hacen uso del hecho de que el mezclado de las líneas espectrales o información espectral referente a componentes espectrales puede llevarse a cabo mediante una suma ponderada de las líneas espectrales fuente o información espectral. Los factores de ponderación pueden ser cero o uno, o en principio, cualquier valor intermedio. Un valor cero significa que las fuentes se tratan como irrelevantes y en la práctica no se usarán. Grupos de líneas, tales como bandas o bandas de factor de escala, pueden usar el mismo factor de ponderación. Sin embargo, como se ilustró anteriormente, los factores de ponderación (por ejemplo, una distribución de ceros y unos) pueden modificarse para los componentes espectrales de una sola trama 540 de un solo tren de datos de entrada 510. Por otra parte, no es necesario usar exclusivamente los factores de ponderación cero o uno cuando se mezcla información espectral. Puede suceder que, en ciertas circunstancias, no para uno solo de una pluralidad de información espectral total de una trama 540 de un tren de datos de entrada 510, los factores de ponderación respectivos pueden ser diferentes de cero o uno.

[0109] Un caso particular es aquel en que todas las bandas de componentes espectrales de una fuente (tren de datos de entrada 510) se ajustan a un factor de uno y todos los factores de las otras fuentes se ajustan a cero. En este caso, el tren de bits de entrada completa de un participante se copia de forma idéntica como un tren de bits mezclado final. Los factores de ponderación pueden calcularse trama por trama, pero también pueden calcularse o determinarse basándose en grupos o secuencias de tramas más largos. Naturalmente, incluso dentro de esta secuencia de tramas o dentro de tramas individuales, los factores de ponderación pueden diferir para distintos componentes espectrales, como se expuso anteriormente. Los factores de ponderación pueden calcularse o determinarse según los resultados del modelo psicoacústico.

[0110] Un ejemplo de un modelo psicoacústico ya se ha descrito anteriormente en el contexto con las ecuaciones (3), (4) y (5). El modelo psicoacústico o un módulo respectivo calcula la proporción de energía $r(n)$ entre una señal mixta en la que se incluyen solo algunos trenes de entrada para llevar a un valor de energía E_f y la señal mixta completa que tiene un valor de energía E_c . La proporción de energía $r(n)$ se calcula posteriormente según la ecuación (5) como 20 veces el valor logarítmico de E_f dividido por E_c .

[0111] Si la proporción es suficientemente alta, los canales menos dominantes pueden considerarse enmascarados por los dominantes. De esta manera se procesa una reducción de irrelevancia, lo que significa que se incluyen solo aquellos trenes que no son perceptibles en modo alguno, a los cuales se atribuye un factor de ponderación, mientras que en todos los demás trenes se descarta al menos una información espectral de un componente espectral. Dicho de otro modo, a estos se les atribuye un factor de ponderación de cero.

[0112] La ventaja reside en que pueden introducirse menos efectos de codificación en tándem, o ninguno, debido la reducción del número de etapas de recuantificación. Dado que cada etapa de cuantificación conlleva un riesgo importante de reducir ruido de cuantificación adicional, la calidad total de la señal de audio puede mejorarse usando cualquiera de las realizaciones mencionadas anteriormente para mezclar una pluralidad de trenes de datos de entrada. Así puede suceder cuando la unidad de tratamiento 520 del aparato 500, tal como se muestra por ejemplo en la Fig. 6, se adapta de manera que el tren de datos de salida 530 se genera de tal modo que se mantiene una distribución de niveles de cuantificación comparado con una distribución de niveles de cuantificación de la trama del tren de entrada determinado o de sus partes. Dicho de otro modo, al copiar y, por lo tanto, al reutilizar los datos

respectivos sin recodificar la información espectral puede omitirse una introducción de ruido de cuantificación adicional.

- [0113]** Por otra parte, el sistema de conferencia, por ejemplo, un sistema de tele/videoconferencia con más de un participante que usa cualquier realización descrita anteriormente en relación con las Fig. 6 a 8, puede ofrecer la ventaja de una menor complejidad en comparación con un mezclado en el dominio del tiempo, ya que pueden omitirse las etapas de transformación de tiempo-frecuencia y las etapas de recodificación. Por otra parte, no se provoca mayor retardo a causa de estos componentes en comparación con el mezclado en el dominio del tiempo, debido a la ausencia del retardo de banco de filtros.
- 10 **[0114]** Para resumir, las realizaciones descritas anteriormente, por ejemplo, pueden adaptarse de manera que no se recuantifiquen bandas de información espectral que corresponden a los componentes espectrales, que se toman completamente de una fuente. Por lo tanto, solo se recuantifican bandas o información espectral que se mezclan, lo que reduce el ruido de cuantificación adicional.
- 15 **[0115]** Sin embargo, las realizaciones descritas anteriormente también pueden emplearse en diferentes aplicaciones, tales como sustitución de ruido perceptual (PNS = Perceptual Noise Substitution), conformado de ruido temporal (TNS = Temporal Noise Shaping), replicación de banda espectral (SBR = Spectral Band Replication) y modos de codificación de estéreo. Antes de describir la operación de un aparato capaz de procesar al menos uno de los parámetros PNS, los parámetros TNS, los parámetros SBR o los parámetros de codificación estéreo, se describirá una realización con más detalle con referencia a la Fig. 8.
- 25 **[0116]** La Fig. 8 muestra un diagrama de bloques esquemático de un aparato 500 para mezclar una pluralidad de trenes de datos de entrada que comprenden una unidad de tratamiento 520. Para ser más preciso, la Fig. 8 muestra un aparato altamente flexible 500 que es capaz de procesar señales de audio altamente diferentes codificadas en trenes de datos de entrada (trenes de bits). Algunos de los componentes que se describirán a continuación son, por lo tanto, componentes opcionales y no se requiere implementarlos en todas las circunstancias.
- 30 **[0117]** La unidad de tratamiento 520 comprende un decodificador de trenes de bits 700 para cada uno de los trenes de datos de entrada o trenes de bits de audio codificados para su tratamiento por la unidad de tratamiento 520. Por razones exclusivamente de sencillez, la Fig. 8 muestra solo dos decodificadores de trenes de bits 700-1, 700-2. Naturalmente, dependiendo del número de trenes de datos de entrada para procesar, puede implementarse un número superior de decodificadores de trenes de bits 700 o un número inferior si, por ejemplo, un decodificador de trenes de bits 700 es capaz de procesar secuencialmente más de uno de los trenes de datos de entrada.
- 35 **[0118]** El decodificador de trenes de bits 700-1, así como los otros decodificadores de trenes de bits 700-2, ... comprenden cada uno un lector de trenes de bits 710 que se adapta para recibir y procesar las señales recibidas, y para aislar y extraer datos comprendidos en el tren de bits. Por ejemplo, el lector del tren de bits 710 puede adaptarse para sincronizar los datos de entrada con un reloj interno y además puede adaptarse para separar el tren de bits de entrada en las tramas apropiadas.
- 40 **[0119]** El decodificador de trenes de bits 700 comprende además un decodificador Huffman 720 acoplado a la salida del lector de trenes de bits 710 para recibir los datos aislados del lector de trenes de bits 710. Una salida del decodificador Huffman 720 se acopla a un decuantificador 730, que también se refiere como cuantificador inverso. El decuantificador 730 acoplado tras el decodificador Huffman 720 es seguido por un dimensionador 740. El decodificador Huffman 720, el decuantificador 730 y el dimensionador 740 forman una primera unidad 750 a la salida de la cual al menos una parte de la señal de audio del tren de datos de entrada respectiva está disponible en el dominio de la frecuencia o el dominio relacionado con la frecuencia en el que opera el codificador del participante (no mostrado en la Fig. 8).
- 45 **[0120]** El decodificador de trenes de bits 700 comprende además una segunda unidad 760 que está acoplada en forma de datos después de la primera unidad 750. La segunda unidad 760 comprende un decodificador estéreo 770 (módulo M/S) detrás del cual se acopla un decodificador PNS. Al decodificador PNS 780 le sigue en forma de datos un decodificador TNS 790 que, junto con el decodificador PNS 780 en el decodificador de estéreo 770, forma la segunda unidad 760.
- 50 **[0121]** Aparte del flujo descrito de datos de audio, el decodificador de trenes de bits 700 comprende además una pluralidad de conexiones entre diferentes módulos referentes a datos de control. Para ser más preciso, el lector de trenes de bits 710 también se acopla al decodificador Huffman 720 para recibir datos de control apropiados. Por otra parte, el decodificador Huffman 720 se acopla directamente al dimensionador 740 para transmitir información de dimensión o de escala al dimensionador 740. El decodificador de estéreo 770, el decodificador PNS 780 y el decodificador TNS 790 se acoplan también cada uno al lector de trenes de bits 710 para recibir datos de control apropiados.
- 60 **[0122]** La unidad de tratamiento 520 comprende además una unidad de mezclado 800 que a su vez comprende un mezclador espectral 810 que se acopla en forma de entrada con los decodificadores de trenes de bits 700. El

mezclador espectral 810, por ejemplo, puede comprender uno o más sumadores para realizar el mezclado real en el dominio de la frecuencia. Por otra parte, el mezclador espectral 810 puede comprender además multiplicadores para permitir una combinación lineal arbitraria de la información espectral que es proporcionada por los decodificadores de trenes de bits 700.

5

[0123] La unidad de mezclado 800 comprende además un módulo de optimización 820 que está acoplado en forma de datos con una salida del mezclador espectral 810. Sin embargo, el módulo de optimización 820 también se acopla al mezclador espectral 810 para proporcionar al mezclador espectral 810 información de control. En forma de datos, el módulo de optimización 820 representa una salida de la unidad de mezclado 800.

10

[0124] La unidad de mezclado 800 comprende además un mezclador SBR 830 que se acopla directamente a una salida de lector de trenes de bits 710 de los diferentes decodificadores de trenes de bits 700. Una salida del mezclador SBR 830 forma otra salida de la unidad de mezclado 800.

15

[0125] La unidad de tratamiento 520 comprende además un codificador de trenes de bits 850 que se acopla a la unidad de mezclado 800. El codificador de trenes de bits 850 comprende una tercera unidad 860 que comprende un codificador TNS 870, un codificador PNS 880 y un codificador de estéreo 890, que se acoplan en serie en el orden descrito. La tercera unidad 860 forma por lo tanto una unidad inversa de la primera unidad 750 del decodificador de trenes de bits 700.

20

[0126] El codificador de trenes de bits 850 comprende además una cuarta unidad 900 que comprende un dimensionador 910, un cuantificador 920 y un codificador Huffman 930, que forman una conexión en serie entre una entrada de la cuarta unidad y una salida de la misma. La cuarta unidad 900, por lo tanto forma un módulo inverso de la primera unidad 750. En consecuencia, el dimensionador 910 también se acopla directamente al codificador Huffman 930 para proporcionar al codificador Huffman 930 los datos de control respectivos.

25

[0127] El codificador del tren de bits 850 también comprende un dispositivo de escritura de trenes de bits 940, que se acopla a la salida del codificador Huffman 930. Además, el dispositivo de escritura de trenes de bits 940 también se acopla al codificador TNS 870, al codificador PNS 880, al codificador estéreo 890 y al codificador Huffman 930 para recibir datos de control e información de estos módulos. Una salida del dispositivo de escritura de trenes de bits 940 forma una salida de la unidad de tratamiento 520 y del aparato 500.

30

[0128] El codificador de trenes de bits 850 también comprende un módulo psicoacústico 950, que también se acopla a la salida de la unidad de mezclado 800. El codificador de trenes de bits 850 se adapta para proporcionar los módulos de la tercera unidad 860 con información de control apropiada indicando por ejemplo, que puede emplearse para codificar la salida de señal de audio por la unidad de mezclado 800 en el marco de las unidades de la tercera unidad 860.

35

[0129] En principio, a las salidas de la segunda unidad 760 hasta la entrada de la tercera unidad 860 es posible, por lo tanto, un tratamiento de la señal de audio en el dominio espectral, como se define por el codificador empleado en el lado de remitente. Sin embargo, como se indicó anteriormente, en su caso podría no ser necesario proceder a una completa decodificación, decuantificación, disminución en dimensiones y otras etapas de tratamiento y, por ejemplo, es dominante la información espectral de una trama de uno de los trenes de datos de entrada. Al menos se copia entonces una parte de la información espectral de los componentes espectrales respectivos en el componente espectral de la trama respectiva del tren de datos de salida.

45

[0130] Para permitir este tratamiento, el aparato 500 y la unidad de tratamiento 520 comprenden líneas de señal adicionales para un intercambio de datos optimizado. Para permitir este tratamiento en la realización mostrada en la Fig. 8, se acopla una salida del decodificador Huffman 720, así como salidas del dimensionador 740, el decodificador estéreo 770 y el decodificador PNS 780, junto con los componentes respectivos de los otros lectores de trenes de bits 710, al módulo de optimización 820 de la unidad de mezclado 800 para un tratamiento respectivo.

50

[0131] Para facilitar, después de un tratamiento respectivo, un flujo de datos correspondiente dentro del codificador de trenes de bits 850, también se implementan líneas de datos correspondientes para un flujo de datos optimizado. Para ser más precisos, una salida del módulo de optimización 820 se acopla a una entrada del codificador PNS 780, el codificador estéreo 890, una entrada de la cuarta unidad 900 y el dimensionador 910, así como una entrada en el codificador Huffman 930. Por otra parte, la salida del módulo de optimización 820 también se acopla directamente al dispositivo de escritura de trenes de bits 940.

55

[0132] Como se indicó anteriormente, casi todos los módulos que se describieron anteriormente son módulos opcionales, que no es preciso implementar. Por ejemplo, en el caso de los trenes de datos de audio que comprenden solo un canal individual, pueden omitirse las unidades de codificación y decodificación en estéreo 770, 890. En consecuencia, en el caso en que no se vayan a procesar señales basadas en PNS, el decodificador PNS y codificador PNS correspondientes 780, 880 también pueden omitirse. Los módulos TNS 790, 870 también pueden omitirse en el caso en que la señal para tratamiento y la señal enviada de salida no se basen en los datos TNS. Dentro de las

60

65

unidades primera y cuarta 750, 900 el cuantificador inverso 730, el dimensionador 740, el cuantificador 920, así como el dimensionador 910 también pueden omitirse en su caso. El decodificador Huffman 720 y el codificador Huffman 930 pueden ser implementados de forma diferente, usando otro algoritmo u omitirse por completo.

5 **[0133]** El mezclador SBR 830 también puede omitirse en su caso si, por ejemplo, no están presentes parámetros SBR de datos. Además, el mezclador espectral 810 puede ser implementado de forma diferente, por ejemplo, en cooperación con el módulo de optimización 820 y el módulo psicoacústico 860. Por lo tanto, también estos módulos habrán de considerarse componentes opcionales.

10 **[0134]** Con respecto al modo de operación del aparato 500 junto con la unidad de tratamiento 520 que comprende, primero se lee un tren de datos de entrada de entrada y es separado en elementos de información apropiados por el lector de trenes de bits 710. Después de decodificación Huffman, la información espectral resultante puede ser recuantificada en su caso por el decuantificador 730 y ajustarse en escala de forma apropiada por medio del dimensionador en sentido inverso 740.

15

[0135] Posteriormente, dependiendo de la información de control comprendida en el tren de datos de entrada, la señal de audio codificada por el tren de datos de entrada puede descomponerse en señales de audio para dos o más canales en el marco del decodificador de estéreo 770. Si, por ejemplo, la señal de audio comprende un canal central (M) y un canal lateral (S), los datos de canal izquierdo y de canal derecho correspondientes pueden obtenerse sumando y restando los datos de los canales central y lateral entre sí. En muchas implementaciones, el canal central es proporcional a la suma de datos de audio del canal izquierdo y el canal derecho, mientras que el canal lateral es proporcional a una diferencia entre el canal izquierdo (L) y el canal derecho (R). Dependiendo de la implementación, los canales referidos anteriormente pueden sumarse y/o restarse teniendo en cuenta un factor 1/2 para evitar efectos de recorte. En términos generales, los diferentes canales pueden procesarse mediante combinaciones lineales para producir los canales correspondientes.

[0136] Dicho de otro modo, después del decodificador de estéreo 770, los datos de audio pueden descomponerse, si fuera apropiado, en dos canales individuales. Naturalmente, también puede realizarse una decodificación inversa por medio del decodificador de estéreo 770. Si, por ejemplo, la señal de audio tal como se recibe por el lector de trenes de bits 710 comprende un canal izquierdo y un canal derecho, el decodificador de estéreo 770 puede igualmente calcular o determinar datos de canal central y lateral apropiados.

[0137] Dependiendo de la implementación no solo del aparato 500, sino también dependiendo de la implementación del codificador del participante que proporciona el tren de datos de entrada respectivo, el tren de datos respectivo puede comprender parámetros de sustitución de ruido perceptual, PNS (PNS = Perceptual Noise Substitution). PNS se basa en el hecho de que el oído humano muy probablemente no es capaz de distinguir sonidos de tipo ruido en un intervalo de frecuencia limitado o componente espectral tal como una banda o una frecuencia individual, de un ruido generado de forma sintética. Por lo tanto, PNS sustituye la contribución de tipo ruido real de la señal de audio por un valor de energía indicando la introducción de un nivel de ruido de forma sintética en el componente espectral respectivo y despreciando la señal de audio real. Dicho de otro modo, el decodificador PNS 780 puede regenerar en uno o más componentes espectrales la contribución de señal de audio de tipo ruido real basándose en un parámetro PNS comprendido en el tren de datos de entrada.

[0138] En términos del decodificador TNS 790 y el codificador TNS 870, las señales de audio respectivas pueden tener que volverse a transformar en una versión no modificada con respecto a un módulo TNS que opera en el lado del remitente. El conformado de ruido temporal (TNS = Temporal Noise Shaping) es un medio para reducir artefactos pre-eco provocados por ruido de cuantificación, que pueden estar presentes en el caso de una señal de tipo transitorio en una trama de la señal de audio. Para contrarrestar este transitorio, se aplica al menos un filtro de predicción adaptativa a la información espectral partiendo del lado bajo del espectro, el lado alto del espectro o ambos lados del espectro. Las longitudes de los filtros de predicción pueden ser adaptadas, así como los intervalos de frecuencia a los cuales se aplican los filtros respectivos.

[0139] Dicho de otro modo, la operación de un módulo TNS se basa en el cálculo de uno o más filtros adaptativos filtros de respuesta de impulso infinita (IIR = Infinite Impulse Response), y codificando y transmitiendo una señal de error que describe la diferencia entre la señal de audio pronosticada y real junto con los coeficientes de filtro de los filtros de predicción. Como consecuencia, puede ser posible incrementar la calidad de audio mientras que se mantiene la velocidad binaria del tren de datos del transmisor al afrontar las señales de tipo transitorio aplicando un filtro de predicción en el dominio de la frecuencia para reducir la amplitud de la señal de error restante, que puede entonces codificarse usando menos etapas de cuantificación en comparación con la codificación directa de la señal de audio de tipo transitorio con un ruido de cuantificación similar.

[0140] En términos de una aplicación TNS, en ciertas circunstancias puede ser recomendable usar la función del decodificador TNS 760 para decodificar las partes TNS del tren de datos de entrada para llegar a una representación "pura" en el dominio espectral determinada por el códec empleado. Esta aplicación de la funcionalidad de los decodificadores TNS 790 puede ser útil si una estimación del modelo psicoacústico (por ejemplo, aplicado en

el módulo psicoacústico 950) no puede haber sido ya estimada basándose en los coeficientes de filtro de los filtros de predicción comprendidos en los parámetros TNS. Este hecho puede ser especialmente importante en el caso en que al menos un tren de datos de entrada usa TNS, mientras que otro no lo hace.

5 **[0141]** Cuando la unidad de tratamiento determina, basándose en la comparación de las tramas de los trenes de datos de entrada, que se va a usar la información espectral de una trama de un tren de datos de entrada que usa TNS, pueden emplearse los parámetros TNS para la trama de datos de salida. Si, por ejemplo, por razones de incompatibilidad el receptor del tren de datos de salida no es capaz de decodificar datos TNS, puede ser útil no copiar los datos espectrales respectivos de la señal de error y los parámetros TNS adicionales, sino procesar los datos
10 reconstruidos de los datos relacionados con TNS para obtener la información en el dominio espectral y no usar el codificador TNS 870. Este hecho ilustra de nuevo que no es preciso implementar partes de los componentes o módulos ilustrados en la Fig. 8, sino que opcionalmente pueden descartarse.

[0142] En el caso de al menos un tren de entrada de audio que compara datos PNS, puede aplicarse una
15 estrategia similar. Si en la comparación de las tramas para un componente espectral de los trenes de datos de entrada se revela que un tren de datos de entrada está en términos de su trama actual y el componente espectral respectivo o los componentes espectrales que lo dominan, los parámetros PNS respectivos (es decir, los valores de energía respectivos) también pueden copiarse directamente al componente espectral respectivo de la trama de salida. Sin embargo si el receptor no es capaz de aceptar los parámetros PNS, la información espectral puede reconstruirse a
20 partir del parámetro PNS para los componentes espectrales respectivos al generar ruido con el nivel de energía apropiado tal como se indica mediante el valor de energía respectivo. Después, los datos de ruido pueden ser procesados de forma consiguiente en el dominio espectral.

[0143] Como se expuso anteriormente, los datos transmitidos también pueden comprender datos SBR que
25 pueden ser procesados en el mezclador SBR 830. La replicación de banda espectral (SBR = Spectral Band Replication) es una técnica para replicar una parte de un espectro de una señal de audio basándose en las contribuciones y la parte inferior del mismo espectro. Como consecuencia, no es necesario transmitir la parte superior del espectro, aparte de los parámetros SBR que describen valores de energía de una forma dependiente de frecuencia y dependiente del tiempo usando una cuadrícula de tiempo/frecuencia apropiada. Como consecuencia, en la práctica
30 no es necesario transmitir la parte superior del espectro. Para poder mejorar adicionalmente la calidad de la señal reconstruida, pueden añadirse contribuciones de ruido y contribuciones sinusoidales adicionales a la parte superior del espectro.

[0144] Para ser un poco más específico, para frecuencias por encima de una frecuencia de cruce f_x , la señal
35 de audio se analiza en términos de un banco de filtros de espejo de cuadratura (QMF = Quadrature Mirror Filter) que crea un número específico de señales de sub-banda (por ejemplo, 32 señales de sub-banda) que tienen una resolución de tiempo que se reduce en un factor igual o proporcional al número de sub-bandas de banco de filtros QMF (por ejemplo, 32 o 64). Como consecuencia, puede determinarse una cuadrícula de tiempo/frecuencia que comprende, en el eje del tiempo, dos o más envolventes y, por cada envolvente, normalmente de 7 a 16 valores de energía que
40 describen la parte superior respectiva del espectro.

[0145] De forma adicional, los parámetros SBR pueden comprender información referente a ruido y sinusoides
adicionales que después se atenúan o se determinan con respecto a su fuerza por medio de la cuadrícula de
45 tiempo/frecuencia anteriormente mencionada.

[0146] En el caso de un tren de datos de entrada basado en SBR que es el tren de datos de entrada dominante
con respecto a la presente trama, puede realizarse un copiado de los parámetros SBR respectivos junto con los
componentes espectrales. Si, de nuevo, el receptor no es capaz de decodificar señales basadas en SBR, puede
realizarse una reconstrucción respectiva en el dominio de la frecuencia seguida por codificación de la señal
50 reconstruida según los requerimientos del receptor.

[0147] Dado que SBR permite dos canales de codificación de estéreo, la codificación del canal izquierdo y el
canal derecho separados, así como la codificación de los mismos en términos de un canal de acoplamiento (C), según
una realización de la presente invención, el copiado de los parámetros SBR respectivos o al menos sus partes, puede
55 comprender el copiado de los elementos C de los parámetros SBR a los elementos izquierdo y derecho del parámetro
SBR para determinar y transmitir. o a la inversa, dependiendo de los resultados de la comparación y de los resultados
de la determinación.

[0148] Por otra parte, dado que en diferentes realizaciones de la presente invención los trenes de datos de
60 entrada pueden comprender las dos señales de audio mono y estéreo que comprenden uno y dos canales individuales,
respectivamente, puede realizarse adicionalmente una premezcla multipistas de mono a estéreo o una submezcla a
dos pistas de estéreo a mono en el marco del copiado de al menos parte de la información cuando se genera al menos
parte de información de un componente espectral correspondiente de la trama del tren de datos de salida.

65 **[0149]** Como ha mostrado la descripción precedente, el grado de copiado de información espectral y/o

parámetros respectivos referentes a componentes espectrales e información espectral (por ejemplo, parámetros TNS, parámetros SBR, parámetros PNS) puede basarse en diferentes números de datos para copiar y puede determinar si es preciso copiar también la información espectral subyacente o elementos de la misma. Por ejemplo, si se copian datos SBR, puede ser recomendable copiar toda la trama del tren de datos respectivo para evitar información espectral de mezclado complicada para componentes espectrales diferentes. Este mezclado puede requerir una recuantificación que, en la práctica, puede reducir el ruido de cuantificación.

[0150] En términos de los parámetros TNS, también puede ser recomendable copiar los parámetros TNS respectivos junto con la información espectral de toda la trama del tren de datos de entrada dominante al tren de datos de salida para evitar una recuantificación.

[0151] En caso de información espectral basada en PNS, una forma viable puede ser copiar valores de energía individuales sin copiar los componentes espectrales subyacentes. Además, en este caso el copiado de solo el parámetro PNS respectivo del componente espectral dominante de las tramas de las pluralidades de trenes de datos de entrada al componente espectral correspondiente de la trama de salida del tren de datos de salida tiene lugar sin introducir ruido de cuantificación adicional. Debe observarse que, también al recuantificar un valor de energía en la forma de un parámetro PNS, puede introducirse ruido de cuantificación adicional.

[0152] Tal como se ha indicado anteriormente, la realización expuesta anteriormente también puede realizarse simplemente copiando una información espectral referente a un componente espectral después de comparar las tramas de la pluralidad de trenes de datos de entrada y después de determinar, basándose en la comparación, para un componente espectral de una trama de salida del tren de datos de salida exactamente un tren de datos como fuente de la información espectral.

[0153] El algoritmo de sustitución realizado en la trama del módulo psicoacústico 950 examina cada información espectral referente a los componentes espectrales subyacentes (por ejemplo, bandas de frecuencia) de la señal resultante para identificar componentes espectrales con un único componente activo. Para estas bandas, los valores cuantificados del tren de datos de entrada respectivo del tren de bits de entrada pueden copiarse desde el codificador sin volver a codificar o recuantificar los datos espectrales respectivos para el componente espectral específico. En ciertas circunstancias, todos los datos cuantificados pueden tomarse de una sola señal de entrada activa para formar el tren de bits de salida o tren de datos de salida de manera que, en términos del aparato 500, se logre una codificación sin pérdidas del tren de datos de entrada.

[0154] Además, puede hacerse posible omitir etapas de tratamiento tales como el análisis psicoacústico dentro del codificador. Así se permite acortar el proceso de codificación y con ello reducir la complejidad de cálculo, ya que en principio en ciertas circunstancias solo debe realizarse el copiado de datos de un tren de bits en otro tren de bits.

[0155] Por ejemplo, en el caso de PNS, puede realizarse una sustitución ya que los factores de ruido de la banda codificada PNS pueden copiarse de uno de los trenes de datos de salida al tren de datos de salida. La sustitución de componentes espectrales individuales con parámetros PNS apropiados es posible ya que los parámetros PNS son específicos del componente espectral, o dicho de otro modo, en una muy buena aproximación independiente entre sí.

[0156] Sin embargo, puede ocurrir que una aplicación demasiado agresiva del algoritmo descrito produzca una experiencia de audición degradada o una reducción de calidad indeseable. Por lo tanto puede ser recomendable limitar la sustitución a tramas individuales en vez de información espectral referente a componentes espectrales individuales. En este modo de operación, la estimación de relevancia o la determinación de relevancia, así como el análisis de sustitución, pueden llevarse a cabo sin cambios. Sin embargo, en este modo de operación una sustitución solo llevarse a cabo cuando se sustituye la totalidad o al menos un número importante de componentes espectrales dentro de la trama activa.

[0157] Aunque esto puede llevar a un menor número de sustituciones, en algunas situaciones puede mejorarse la fuerza interior de la información espectral para conseguir una calidad ligeramente mejorada.

[0158] A continuación se describen realizaciones según un segundo aspecto de la presente invención según lo cual se tienen en cuenta los valores de control asociados con datos de carga útil de las tramas de datos de entrada respectivas, de manera que los valores de control indican una forma en que los datos de carga útil representan al menos una parte de la información espectral correspondiente o dominio espectral de las señales de audio respectivas, en la que en el caso en que los valores de control de las dos tramas de datos de entrada son iguales, se evita una nueva decisión en la forma en que el dominio espectral de la trama respectivo del tren de datos de salida y por el contrario la generación de tren de salida se basa en la decisión ya determinada por los codificadores del tren de datos de entrada. Según algunas realizaciones descritas a continuación, se evita la retransformación de los datos de carga útil respectivos de nuevo en otra forma de representar el dominio espectral como la ruta normal o simple sin un valor espectral por muestra de tiempo/espectral.

[0159] Tal como se ha indicado anteriormente, las realizaciones según la presente invención se basan en

realizar un mezclado, que no se lleva a cabo de una forma directa en el sentido de que todos los trenes de entrada se decodifican, lo que incluye transformación inversa al dominio del tiempo, mezclado y de nuevo recodificación de las señales. Las realizaciones según la presente invención se basan en el mezclado efectuado en el dominio de la frecuencia del códec respectivo. Un códec posible puede ser el códec AAC-ELD, o cualquier otro códec con una
5 ventana de transformación uniforme. En este caso, no se necesita transformación de tiempo/frecuencia para poder mezclar los datos respectivos. Además, es posible el acceso a todos los parámetros de trenes de bits tales como el tamaño de la etapa de cuantificación y otros parámetros, y estos parámetros pueden emplearse para generar un tren de bits de salida mixta.

10 **[0160]** Adicionalmente, el mezclado de las líneas espectrales o información espectral referente a componentes espectrales puede llevarse a cabo mediante una suma ponderada de las líneas espectrales fuente o información espectral. Los factores de ponderación pueden ser cero o uno o en principio cualquier valor intermedio. Un valor de
15 cero significa que las fuentes que se tratan son irrelevantes y, en la práctica, no se usarán. Grupos de líneas tales como bandas o bandas de factor de escala pueden usar el mismo factor de ponderación. Los factores de ponderación (por ejemplo, una distribución de ceros y unos) pueden variarse para los componentes espectrales de una sola trama de un único tren de datos de entrada. En las realizaciones descritas a continuación no es ni mucho menos necesario usar exclusivamente los factores de ponderación de cero o uno cuando se mezcla información espectral. En ciertas circunstancias puede suceder que no solo en una, sino en una pluralidad de información espectral total de una trama de un tren de datos de entrada, los factores de ponderación respectivos pueden ser diferentes de cero o uno.

20 **[0161]** Un caso particular es aquel en que todas las bandas o el componente espectral de una fuente (tren de datos de entrada) se ajustan a un factor de uno y todos los factores de las otras fuentes se ajustan a cero. En este caso, el tren de bits de entrada completa de un participante puede copiarse de forma idéntica como un tren de bits mezclado final. Los factores de ponderación pueden calcularse trama por trama, pero también pueden calcularse o
25 determinarse basándose en grupos o secuencias de tramas más largos. Naturalmente, incluso dentro de esta secuencia de tramas o dentro de tramas individuales, los factores de ponderación pueden diferir para componentes espectrales diferentes, tal como se expuso anteriormente. En algunas realizaciones, los factores de ponderación pueden calcularse o determinarse según los resultados del modelo psicoacústico.

30 **[0162]** Esta comparación puede realizarse, por ejemplo, basándose en la evaluación de una proporción de energía entre la señal mixta en la que se incluyen solo algunos trenes de entrada y una señal mixta completa. Esto puede lograrse, por ejemplo, como se describió anteriormente con respecto a las ecuaciones (3) a (5). Dicho de otro modo, el modelo psicoacústico puede calcular la proporción de energía $r(n)$ entre una señal mixta en la que se incluyen solo algunos trenes de entrada para llevar a un valor de energía E_f y la señal mixta completa que tiene un valor de
35 energía E_c . La proporción de energía $r(n)$ se calcula entonces según la ecuación (5) como 20 veces el valor logarítmico de E_f dividido por E_c .

[0163] Como consecuencia, de forma similar a la descripción anterior de realizaciones con respecto a las Fig. 6 a 8, si la proporción es suficientemente alta, los canales menos dominantes pueden considerarse enmascarados por
40 los dominantes. De esta manera, se procesa una reducción de irrelevancia lo que significa que solo se incluyen aquellos trenes que, en la práctica, no son en modo alguno perceptibles, a los cuales se atribuye un factor de ponderación de uno, mientras que todos los demás trenes (al menos una información espectral de un componente espectral) se descartan. Dicho de otro modo, a estos se les atribuye un factor de ponderación de cero.

45 **[0164]** Esta situación puede llevar a la ventaja adicional de que se producen menos efectos de codificación en tándem, o ninguno, debido a la reducción del número de etapas de recuantificación. Dado que cada etapa de recuantificación conlleva un riesgo importante de reducir el ruido de cuantificación adicional, puede mejorarse por lo tanto la calidad total de la señal de audio.

50 **[0165]** De forma similar a las realizaciones descritas anteriormente de las Fig. 6 a 8, las realizaciones descritas a continuación pueden emplearse con un sistema de conferencia que, por ejemplo, puede ser un sistema de tele/videoconferencia con más de dos participantes y puede ofrecer la ventaja de una menor complejidad en comparación con un mezclado en el dominio del tiempo, ya que pueden omitirse las etapas de transformación de tiempo-frecuencia y las etapas de recodificación. Por otra parte, no se provoca mayor retardo por estos componentes
55 en comparación con el mezclado en el dominio del tiempo, debido a la ausencia de retardo del banco de filtros.

[0166] La Fig. 9 muestra un diagrama de bloques simplificado de un aparato 500 para mezclar trenes de datos de entrada según una realización de la presente invención. La mayoría de los signos de referencia se han adoptado para las realizaciones de las Fig. 6 a 8 con el fin de facilitar la comprensión y evitar descripciones duplicadas. Otros
60 signos de referencia se han incrementado en 1000 para denotar que la funcionalidad de los mismos se define de forma diferente en comparación con las realizaciones anteriores de las Fig. 6 a 8, en cualquier funcionalidad adicional o funcionalidad alterna, pero de manera que la función general del elemento respectivo es comparable.

[0167] Basándose en el primer tren de datos de entrada 510-1 y en un segundo tren de datos de entrada 510-2, se adapta una unidad de tratamiento 1520 comprendida en el aparato 1500 para generar un tren de datos de salida

1530. Los trenes de datos de entrada primero y segundo 510 comprenden cada uno una trama 540-1, 540-2, respectivamente, comprendiendo cada uno un valor de control 1545-1, 1545-2, respectivamente, que indica una forma en que los datos de carga útil de las tramas 540 representan al menos una parte del dominio espectral o información espectral de una señal de audio.

5

[0168] El tren de datos de salida 530 también comprende una trama de salida 1550 con un valor de control 555, que indica de manera similar una forma en la que los datos de carga útil de la trama de salida 550 representan información espectral en el dominio espectral de la señal de audio codificada en el tren de datos de salida 530.

10 **[0169]** La unidad de tratamiento 1520 del aparato 1500 está adaptada para comparar los valores de control 1545-1 de la trama 540-1 del primer tren de datos de entrada 510-1 y el valor de control 1545-2 de una trama 540-2 del segundo tren de datos de entrada 510-2 para dar un resultado de comparación. Basándose en este resultado de comparación, la unidad de tratamiento 1520 está adaptada además para generar el tren de datos de salida 530 que comprende la trama de salida 550, de manera que cuando el resultado de comparación indica que los valores de control 1545 de las tramas 540 de los trenes de datos de entrada primero y segundo 510 son idénticos o iguales, la trama de salida 550 comprende como valor de control 1550 un valor igual al de los valores de control 1545 de las tramas 540 de los dos trenes de datos de entrada 510. Los datos de carga útil comprendidos en la trama de salida 550 se deducen a partir de los datos de carga útil correspondientes de las tramas 540 con respecto a los valores de control idénticos 1545 de las tramas 540 mediante el tratamiento en el dominio espectral, es decir, sin visitar el dominio del tiempo.

25 **[0170]** Si, por ejemplo, los valores de control 1545 indican una codificación especializada de información espectral de uno o más componentes espectrales (por ejemplo, datos PNS) y los valores de control respectivos 1545 de los dos trenes de datos de entrada son idénticos, entonces la información espectral correspondiente de la trama de salida 550 correspondiente al mismo componente espectral o componentes espectrales puede obtenerse mediante el tratamiento de los datos de carga útil correspondientes en el dominio espectral incluso directamente, esto es, no dejando el tipo de representación del dominio espectral. Como se expondrá a continuación, en el caso de una representación espectral basada en PNS, esto puede lograrse mediante la suma de los datos PNS respectivos, acompañado opcionalmente por un proceso de normalización. Es decir, los datos PNS de cualquier tren de datos de entrada se convierten de nuevo en representación simple con un valor por muestra espectral.

35 **[0171]** La Fig. 10 muestra un diagrama más detallado de un aparato 1500 y se diferencia de la Fig. 9, principalmente en lo que respecta a una estructura interior de la unidad de tratamiento 1520. Para ser más concretos, la unidad de tratamiento 1520 comprende un comparador 1560, que se acopla a entradas apropiadas para los trenes de datos de entrada primero y segundo 510 y que se adapta para comparar los valores de control 1545 de sus tramas respectivas 540. Los trenes de datos de entrada se proporcionan además a un transformador opcional 1570-1, 1570-2, para cada uno de los dos trenes de datos de entrada 510. El comparador 1560 también se acopla a los transformadores opcionales 1570 para proporcionar los mismos que con el resultado de comparación.

40 **[0172]** La unidad de tratamiento 1520 comprende además un mezclador 1580, que se acopla en forma de entrada con los transformadores opcionales 1570, o en el caso de que uno o más de los transformadores 1570 no se implementan, a las entradas correspondientes para los trenes de datos de entrada 510. El mezclador 1580 se acopla con una salida de un normalizador opcional 1590 que a su vez se acopla, si se implementa, con una salida de la unidad de tratamiento 1520 y la del aparato 1500 para proporcionar el tren de datos de salida 530.

45

[0173] Como se expuso anteriormente, la comparación 1560 se adapta para comparar los valores de control de las tramas 540 de los dos trenes de datos de entrada 510. El comparador 1560 proporciona, si se implementa, a los transformadores 1570 una señal que indica si los valores de control 1545 de las tramas respectivas 540 son idénticos, o no. Si la señal que representa el resultado de comparación indica que los dos valores de control 1545 son, al menos con respecto a un componente espectral, idénticos o iguales, los transformadores 1570 no transforman los datos de carga útil respectivos comprendidos en las tramas 540.

55 **[0174]** Los datos de carga útil comprendidos en las tramas 540 del tren de datos de entrada 510 se mezclarán entonces por medio del mezclador 1580 y enviarán una salida al normalizador 1590, si se implementan, para llevar a cabo una etapa de normalización con el fin de asegurar que los valores resultantes no superen ni estén por debajo de un intervalo de valores permisibles. A continuación se expondrán con más detalle ejemplos de mezclado de datos de carga útil en el contexto de las Fig. 12a a 12c.

60 **[0175]** El normalizador 1590 puede implementarse como un cuantificador adaptado para recuantificar los datos de carga útil según sus valores respectivos, alternativamente, el normalizador 1590 también puede adaptarse para modificar únicamente un factor de escala indicando una distribución de etapas de cuantificación o un valor absoluto de un nivel de cuantificación mínimo o máximo dependiendo de su implementación concreta.

[0176] En el caso en que el comparador 1560 indique que los valores de control 1545 son diferentes al menos con respecto a uno o más componentes espectrales, el comparador 1560 puede proporcionar a uno o a los dos

transformadores 1570 una señal de control respectiva indicando los transformadores respectivos 1570 para transformar los datos de carga útil de al menos uno de los trenes de datos de entrada 510 con el otro de los trenes de datos de entrada. En este caso, el transformador puede adaptarse para cambiar simultáneamente el valor de control de la trama transformado de manera que el mezclador 1580 sea capaz de generar la trama de salida 550 del tren de datos de salida 530 con un valor de control 1555 que es igual al de una trama 540 de los dos trenes de datos de entrada, que no se transforma o con un valor común de datos de carga útil de las dos tramas 540.

[0177] A continuación se describirán ejemplos más detallados en el contexto de las Fig. 12a a 12c para diferentes aplicaciones tales como implementaciones PNS, implementaciones SBR e implementaciones M/S, respectivamente.

[0178] Debe observarse que las realizaciones de las Fig. 9 a 12C no están ni mucho menos limitadas a dos trenes de datos de entrada 1510-1, 1510-2 tal como se muestra en las Fig. 9, 10 y la Fig. 11 que se mostrará seguidamente. Al contrario, puede adaptarse para tratar una pluralidad de trenes de datos de entrada que comprenden más de dos trenes de datos de entrada 510. En este caso, el comparador 1560, por ejemplo, puede adaptarse para comparar una cantidad apropiada de trenes de datos de entrada 510 y las tramas 540 comprendidas en ellos. Por otra parte, dependiendo de la implementación concreta, puede implementarse también una serie apropiada de transformadores 1570. El mezclador 1580 junto con el normalizador opcional 1590 puede adaptarse en su caso al mayor número de trenes de datos para procesar.

[0179] En el caso de más de solo dos trenes de datos de entrada 510, el comparador 1560 puede adaptarse para comparar todos los valores de control relevantes 1545 de los trenes de datos de entrada 510 con el fin de decidir si se va a realizar una etapa de transformación mediante uno o más de los transformadores 1570 implementados opcionalmente. De forma alternativa o adicional, el comparador 1560 también puede adaptarse para determinar un conjunto de trenes de datos de entrada que serán transformados por los transformadores 1570, cuando el resultado de comparación indica que puede conseguirse una transformación a una forma de representación común de los datos de carga útil. Por ejemplo, a menos de que la representación diferente de datos de carga útil implicados requiera una cierta representación, el comparador 1560 puede adaptarse, por ejemplo, para activar los transformadores 1570 de manera que se reduzca al mínimo la complejidad total. Esto puede lograrse, por ejemplo, basándose en estimaciones predeterminadas de valores de complejidad almacenados dentro del comparador 1560 o disponibles para el comparador 1560 de una forma diferente.

[0180] Por otra parte, debe observarse que el transformador 1570 puede omitirse, en su caso, cuando, por ejemplo, la transformación en el dominio de la frecuencia puede llevarse a cabo opcionalmente por el mezclador 1580 a demanda. De forma alternativa o adicional, la funcionalidad de los transformadores 1570 también puede incorporarse en el mezclador 1580.

[0181] Además, debe observarse que las tramas 540 pueden comprender más de un valor de control, tal como sustitución de ruido perceptual (PNS), conformado de ruido temporal (TNS) y modos de codificación de estéreo. Antes de describir el funcionamiento de un aparato capaz de procesar al menos uno de los parámetros PNS, parámetros TNS o parámetros de codificación de estéreo, se hace referencia a la Fig. 11 que es igual a la Fig. 8 aunque, no obstante, se usan los signos de referencia 1500 y 1520 en lugar de 500 y 520, respectivamente, para ilustrar que la Fig. 8 ya muestra una realización para generar un tren de datos de salida a partir de los trenes de datos de entrada primero y segundo en los que las unidades de tratamiento 520 y 1520, respectivamente, también pueden adaptarse para llevar a cabo la funcionalidad descrita con respecto a las Fig. 9 y 10. En particular, dentro de la unidad de tratamiento 1520, la unidad de mezclado 800 que comprende el mezclador espectral 810, el módulo de optimización 820 y el mezclador SBR 830 realiza las funciones descritas anteriormente que se exponen con respecto a las Fig. 9 y 10. Como se indicó anteriormente, los valores de control comprendidos en las tramas de los trenes de datos de entrada pueden ser igualmente parámetros PNS, parámetros SBR o datos de control referentes a codificación de estéreo, es decir, parámetros M/S. En el caso en que los valores de control respectivos sean iguales o idénticos, la unidad de mezclado 800 puede procesar los datos de carga útil para generar los datos de carga útil correspondientes que serán procesados adicionalmente comprendidos en la trama de salida del tren de datos de salida. A este respecto, como ya se ha expuesto anteriormente, como SBR permite dos canales de codificación de estéreo, la codificación del canal izquierdo y el canal derecho por separado, así como la codificación de los mismos en términos de un canal de acoplamiento (C), según una realización de la presente invención, el tratamiento de los parámetros SBR respectivos o al menos de sus partes puede comprender el tratamiento de los elementos C de los parámetros SBR para obtener los dos, los elementos izquierdo y derecho del parámetro SBR o a la inversa, dependiendo de los resultados de la comparación y del resultado de la determinación. De forma similar, el grado de tratamiento de información espectral y/o los parámetros respectivos referentes a componentes espectrales e información espectral (por ejemplo, parámetros TNS, parámetros SBR, parámetros PNS) puede basarse en diferentes números de datos para procesar y puede determinar si también se necesita decodificar la información espectral subyacente o partes de la misma. Por ejemplo, en el caso de copiado de los datos SBR, puede ser recomendable procesar toda la trama del tren de datos respectivo para evitar un mezclado complicado de información espectral para diferentes componentes espectrales. El mezclado de los mismos puede obligar a una recuantificación que, en la práctica, puede reducir el ruido de cuantificación. En términos de parámetros TNS también puede ser recomendable descomponer los parámetros TNS

respectivos junto con la información espectral de toda la trama a partir del tren de datos de entrada dominante en el tren de datos de salida para evitar una recuantificación. En el caso de información espectral basada en PNS, el tratamiento de valores de energía individuales sin copiar los componentes espectrales subyacentes puede ser una forma viable. Además, en este caso el tratamiento de solo el parámetro PNS respectivo a partir del componente espectral dominante desde las tramas de la pluralidad de trenes de datos de entrada al componente espectral correspondiente de la trama de salida del tren de datos de salida tiene lugar sin introducir ruido de cuantificación adicional. Debe observarse además que al recuantificar un valor de energía en la forma de un parámetro PSN, puede introducirse ruido de cuantificación adicional.

10 **[0182]** Con respecto a las Fig. 12A a 12C se describirán con más detalle tres modos diferentes de mezclado de datos de carga útil basándose en una comparación de valores de control respectivos. La Fig. 12a muestra un ejemplo de una implementación basada en PNS de un aparato 500 según una realización de la presente invención, mientras que la Fig. 12b muestra una implementación SBR similar y la Fig. 12c muestra una implementación M/S de la misma.

15 **[0183]** La Fig. 12a muestra un ejemplo con un primer y un segundo tren de datos de entrada 510-1, 510-2, respectivamente, con tramas de entrada apropiadas 540-1, 540-2 y valores de control respectivos 545-1, 545-2. Como se indica mediante las flechas en la Fig. 11a, los valores de control 1545 de las tramas 540 de los trenes de datos de entrada 510 indican que un componente espectral no se describe en términos de información espectral indirectamente, sino en términos de un valor de energía de una fuente de ruido, o dicho de otro modo, por medio de un parámetro PNS apropiado. Más en concreto, la Fig. 12a muestra un primer parámetro PNS 2000-1 y la trama 540-2 del segundo tren de datos de entrada 510-2 que comprende un parámetro PNS 2000-2.

25 **[0184]** Dado que, como se considera con respecto a la Fig. 12a, los valores de control 1545 de las dos tramas 540 de los dos trenes de datos de entrada 510 indican que el componente espectral específico se va a sustituir por el parámetro PNS respectivo 2000, la unidad de tratamiento 1520 y el aparato 1500, como se describió anteriormente, son capaces de mezclar los dos parámetros PNS 2000-1, 2000-2 para llegar al parámetro PNS 2000-3 de la trama de salida 550 que se incluirá en el tren de datos de salida 530. El valor de control respectivo 1555 de la trama de salida 550 indica también esencialmente que el componente espectral respectivo se va a sustituir por el parámetro PNS 2000-3 mixto. Este proceso de mezclado se ilustra en la Fig. 12a mostrando el parámetro PNS 2000-3 como los parámetros PNS combinados 2000-1, 2000-2 de las tramas respectivas 540-1, 540-2.

30 **[0185]** Sin embargo, la determinación del parámetro PNS 2000-3, que también se refiere como un parámetro de salida PNS, también puede lograrse basándose en una combinación lineal según

35

$$PNS = \sum_{i=1}^N a_i \cdot PNS(i) \quad , \quad (6)$$

en la que PNS(i) es el parámetro PNS respectivo del tren de datos de entrada i, N es el número de trenes de datos de entrada que se mezclarán y a_i es el factor de ponderación apropiado. Dependiendo de la implementación concreta, los factores de ponderación a_i pueden seleccionarse de manera que sean iguales

40

$$a_1 = \dots = a_N \quad . \quad (7)$$

45 **[0186]** Una implementación simplificada que se ilustra en la Fig. 12a puede consistir en que todos los parámetros de ponderación a_i son iguales a 1, dicho de otro modo,

$$a_1 = \dots = a_N = 1 \quad . \quad (8)$$

[0187] En el caso en que se omite un normalizador 1590 tal como se muestra en la Fig. 10, los factores de ponderación también pueden definirse como iguales a $1/N$ de manera que la ecuación

50

$$a_1 = \dots = a_N = \frac{1}{N} \quad (9)$$

sigue cumpliéndose.

55

[0188] El parámetro N en este caso es el número de trenes de datos de entrada que se mezclarán, y el número de trenes de datos de entrada que se proporciona al aparato 1500 es un número similar. Por razones de sencillez, debe observarse que también pueden implementarse normalizaciones diferentes en términos de factores de ponderación a_i .

5

[0189] Dicho de otro modo, en el caso de una herramienta PNS activada en el lado del participante, el factor de energía de ruido sustituye a un factor de escala apropiado junto con los datos cuantificados en un componente espectral (por ejemplo, una banda espectral). Aparte de este factor, no se proporcionarán datos adicionales en el tren de datos de salida por medio de la herramienta PNS. En el caso de que se mezclen componentes espectrales PNS, pueden producirse dos situaciones distintas.

10

[0190] Como se describió anteriormente, los componentes espectrales respectivos de todas las tramas 540 de los trenes de datos de entrada relevantes se expresan cada uno en términos de parámetros PNS. Dado que los datos de frecuencia de una descripción relacionada con PNS de un componente de frecuencia (por ejemplo, banda de frecuencia) se deducen directamente del factor de energía de ruido (parámetro PNS), los factores apropiados pueden mezclarse simplemente sumando los valores respectivos. El parámetro PNS mixto generará entonces dentro del decodificador PNS en el lado del receptor una resolución de frecuencia equivalente al mezclado con los valores espectrales puros de otros componentes espectrales. En el caso en que se utilice un proceso de normalización durante el mezclado, puede servir de ayuda implementar un factor de normalización similar en términos de los factores de ponderación a_i . Por ejemplo, cuando se normaliza con un factor proporcional a $1/N$, los factores de ponderación a_i pueden seleccionarse según la ecuación (9).

15

20

[0191] En el caso en que los valores del control 1545 de al menos un tren de datos de entrada 510 difieren con respecto a un componente espectral, y si los trenes de datos de entrada respectivos no se descartan debido a un bajo nivel de energía, puede ser recomendable que el decodificador PNS tal como se muestra en la Fig. 11 genere la información espectral o los datos espectrales basándose en los parámetros PNS y mezcle los datos respectivos en el marco del mezclador espectral 810 de la unidad de mezclado en lugar de mezclar parámetros PNS en el marco del módulo de optimización 820.

25

[0192] Debido a la independencia de los componentes espectrales PNS entre sí, y con respecto a parámetros globalmente definidos del tren de datos de salida, así como los trenes de datos de entrada, puede adaptarse una selección del procedimiento de mezclado sobre una base de bandas. En el caso en que este mezclado basado en PNS no sea posible, puede ser recomendable considerar la recodificación del componente espectral respectivo por el codificador PNS 1880 después de un mezclado en el dominio espectral.

30

35

[0193] La Fig. 12b muestra un ejemplo adicional de un principio operativo de una realización según una realización de la presente invención. Para ser más precisos, la Fig. 12b muestra el caso de dos trenes de datos de entrada 510-1, 510-2 con tramas apropiadas 540-1, 540-2 y sus valores de control 1545-1, 1545-2. Las tramas 540 comprenden datos SBR para componentes espectrales sobre una denominada frecuencia de cruce f_x . El valor de control 1545 comprende información sobre si se usan los parámetros SBR así como la información referente a la cuadrícula de trama real o la cuadrícula de tiempo/frecuencia.

40

[0194] Como se expuso anteriormente, la herramienta SBR replica en una banda espectral superior sobre las frecuencias de cruce f_x partes del espectro al replicar una parte inferior de un espectro que se codifica de manera diferente. La herramienta SBR determina una serie de intervalos de tiempo por cada trama SBR que son iguales a las tramas 540 del tren de datos de entrada 510 que comprenden también información espectral adicional. Los intervalos de tiempo separan el intervalo de frecuencia de la herramienta SBR en bandas de frecuencia o componentes espectrales con igual separación. El número de estas bandas de frecuencia en una trama SBR será determinado por el remitente o la herramienta SBR antes de la codificación. En caso de MPEG-4 AAC-ELD, el número de intervalos de tiempo se fija en 16.

50

[0195] Los intervalos de tiempo se incluyen a continuación en las denominadas envolventes de manera que cada envolvente comprende al menos dos o más intervalos de tiempo que forman un grupo respectivo. Cada envolvente se atribuye a una serie de datos de frecuencia SBR. En la cuadrícula de trama o cuadrícula de tiempo/frecuencia se almacena el número y la longitud en unidades de intervalos de tiempo de las envolventes individuales.

55

[0196] La resolución de frecuencia de las envolventes individuales determina cuántos datos de energía SBR se calculan para una envolvente y se almacenan con respecto a la misma. La herramienta SBR difiere solo entre una alta y una baja resolución, de manera que una envolvente que comprende una alta resolución comprende el doble de valores que una envolvente con una baja resolución. El número de valores de frecuencia o componentes espectrales para envolventes que comprenden una alta o una baja resolución depende de parámetros adicionales del codificador tales como la velocidad binaria, la frecuencia de muestreo y así sucesivamente.

60

[0197] En el contexto de MPEG-4 AAC ELD, la herramienta SBR a menudo usa 16 o 14 valores con respecto

65

a la envolvente que tiene alta resolución.

[0198] Debido a la división dinámica de la trama 540 con un número apropiado de valores de energía con respecto a la frecuencia, puede considerarse un transitorio. En el caso en que exista un transitorio en una trama, el codificador SBR divide la trama respectiva en una cantidad apropiada de envolventes. Esta distribución está normalizada en el caso de la herramienta SBR empleada con el códec AAC ELD y depende de la posición de la transposición del transitorio en unidades de intervalos de tiempo. En muchos casos, la trama de cuadrícula o cuadrícula de tiempo/frecuencia resultante comprende tres envolventes cuando existe un transitorio. Una primera envolvente, la envolvente de inicio, comprende el comienzo de una trama hasta el intervalo de tiempo que recibe el transitorio que tiene los índices de intervalo de tiempo cero a transposición-1. La segunda envolvente comprende una longitud de dos intervalos de tiempo que circunscriben la transitoria desde la transposición de índice de intervalo de tiempo a transposición+2. La tercera envolvente comprende todos los intervalos de tiempo restantes con los índices de transposición+3 a 16.

[0199] Sin embargo, la longitud mínima de una envolvente es de dos intervalos de tiempo. Como consecuencia, tramas que comprenden una transitoria cerca de las fronteras de trama puede eventualmente comprender solo dos envolventes. En caso de que no hay presente transitoria en el intervalo, los intervalos de tiempo se distribuyen sobre envolventes de la misma longitud.

[0200] La Fig. 12b ilustra esta cuadrícula de tiempo/frecuencia o cuadrícula de trama dentro de las tramas 540. En el caso en que los valores de control 1545 indican que las mismas cuadrículas de tiempo SBR o cuadrículas de tiempo/frecuencia están presentes en las dos tramas 540-1, 540-2, los datos SBR respectivos pueden copiarse de manera similar al procedimiento descrito en el marco de las ecuaciones (6) a (9) anteriores. Dicho de otro modo, en dicho caso la herramienta de mezclado SBR o el mezclador SBR 830, tal como se muestra en la Fig. 11, puede copiar la cuadrícula de tiempo/frecuencia o cuadrícula de trama de las tramas de entrada respectivas en la trama de salida 550 y calcular los valores de energía respectivos similares a las ecuaciones (6) a (9).

[0201] Dicho de otro modo todavía, los datos de energía SBR de la cuadrícula de trama pueden mezclarse simplemente sumando los datos respectivos y opcionalmente normalizando los datos respectivos.

[0202] La Fig. 12c muestra un ejemplo adicional de un modo de funcionamiento de una realización según la presente invención. Para ser más precisos, la Fig. 12c muestra una implementación M/S. De nuevo, la Fig. 12c muestra dos trenes de datos de entrada 510 junto con dos tramas 540 y valores de control asociados 545 que indican una forma en que se representan las tramas de datos de carga útil 540, al menos con respecto a un componente espectral de los mismos como mínimo.

[0203] Las tramas 540 comprenden cada una datos de audio o información espectral de dos canales, un primer canal 2020 y un segundo canal 2030. Dependiendo del valor de control 1545 de la trama respectiva 540, el primer canal 2020 puede ser, por ejemplo, un canal izquierdo o un canal central, mientras que el segundo canal 2030 puede ser un canal derecho de un canal de estéreo o un canal lateral. El primero de los modos de codificación se refiere a menudo como modo LR, mientras que el segundo modo a menudo se refiere como modo M/S.

[0204] En el modo M/S, que en ocasiones también se refiere como estéreo conjunto, el canal central (M) se definirá como proporcional a una suma del canal izquierdo (L) y del canal derecho (R). A menudo, en la definición se incluye un factor adicional de $\frac{1}{2}$, de manera que el canal central comprende, tanto en el dominio del tiempo como en el dominio de la frecuencia, un valor promedio de los dos canales de estéreo.

[0205] El canal lateral normalmente se define como proporcional a una diferencia de los dos canales de estéreo, es decir, proporcional a una diferencia entre el canal izquierdo canal (L) y el canal derecho (R). En ocasiones, también se incluye un factor adicional de $\frac{1}{2}$ de manera que el canal lateral representa realmente la mitad del valor de desviación entre los dos canales de la señal de estéreo, o la desviación del canal central. En consecuencia, el canal izquierdo puede reconstruirse sumando el canal central y el canal lateral, mientras que el canal derecho puede obtenerse restando el canal lateral del canal central.

[0206] En el caso en que para las tramas 540-1 y 540-2 se use la misma codificación de estéreo (L/R o M/S), puede omitirse la retransformación de los canales comprendidos en la trama, lo que permite un mezclado directo en el dominio de codificación L/R o M/S respectivo.

[0207] En este caso, el mezclado puede llevarse a cabo de nuevo directamente en el dominio de la frecuencia que conduce a una trama 550 comprendida en un tren de datos de salida 530 que tiene el valor de control respectivo 1555 con un valor igual a los valores de control 1545-1, 1545-2 de las dos tramas 540. De forma correspondiente, la trama de salida 550 comprende dos canales 2020-3, 2030-3 derivados del primer y el segundo canal de las tramas del tren de datos de entrada.

[0208] En el caso en que los valores de control 1545-1, 1545-2 de las dos tramas 540 no sean iguales, puede

ser recomendable transformar una de las tramas en la otra representación basándose en el proceso descrito anteriormente. El valor de control 1555 de la trama de salida 550 puede ajustarse de forma consiguiente al valor indicativo de la trama transformada.

5 **[0209]** Según realizaciones de la presente invención, puede ser posible que los valores de control 1545, 1555 indiquen una representación de toda la trama 540, 550, respectivamente, o que los valores de control respectivos sean específicos del componente de frecuencia. Mientras que, en el primer caso, los canales 2020, 2030 se codifican sobre solo la trama por medio de uno de los procedimientos específicos, en el segundo caso, en principio, cada información espectral con respecto a un componente espectral puede codificarse de forma diferente. Naturalmente, también
10 pueden describirse subgrupos de componentes espectrales mediante uno de los valores de control 1545.

[0210] Adicionalmente, puede ejecutarse un algoritmo de sustitución en el marco del módulo psicoacústico 950 para examinar cada uno de los elementos de información espectral referentes a los componentes espectrales subyacentes (por ejemplo, bandas de frecuencia) de la señal con el fin de identificar componentes espectrales con
15 solo un componente activo individual. Para estas bandas, los valores cuantificados del tren de datos de entrada respectivo del tren de bits de entrada pueden copiarse a partir del codificador sin recodificación o recuantificación de los datos espectrales respectivos para el componente espectral específico. En ciertas circunstancias todos los datos cuantificados pueden tomarse de una única señal de entrada activa para formar el tren de bits de salida o el tren de datos de salida, de manera que, en términos del aparato 1500, se consiga una codificación sin pérdidas del tren de
20 datos de entrada.

[0211] Además, puede hacerse posible omitir etapas del tratamiento tales como el análisis psicoacústico dentro del codificador. Así se permite acortar el proceso de codificación y, de esta manera, reducir la complejidad de cálculo ya que en principio solo debe realizarse en ciertas circunstancias el copiado de datos de un tren de bits a otro tren de
25 bits.

[0212] Por ejemplo, en el caso de PNS, puede llevarse a cabo una sustitución ya que pueden copiarse factores de ruido en la banda codificada PNS de uno de los trenes de datos de salida al tren de datos de salida. Es posible sustituir componentes espectrales individuales con parámetros PNS apropiados, ya que los parámetros PNS son
30 específicos de la componente espectral o, dicho de otro modo, en una muy buena aproximación, son independientes entre sí.

[0213] Sin embargo, puede ocurrir que una aplicación demasiado agresiva del algoritmo descrito pueda producir una experiencia de audición degradada o una reducción indeseada de la calidad. Por tanto puede ser
35 recomendable limitar la sustitución a tramas individuales, en vez de a la información espectral, en referencia a componentes espectrales individuales. En este modo de operación la estimación de irrelevancia o la determinación de irrelevancia, así como el análisis de sustitución, pueden llevarse a cabo sin cambios. Sin embargo, en este modo de funcionamiento, la sustitución solo puede llevarse a cabo cuando es posible sustituir la totalidad o al menos un número importante de los componentes espectrales dentro de la trama activa.

40 **[0214]** Aunque esto puede traducirse en un número menor de sustituciones, en algunas situaciones puede mejorarse la fuerza interior de la información espectral para alcanzar una calidad incluso ligeramente mejorada.

[0215] Las realizaciones expuestas anteriormente pueden diferir, naturalmente, con respecto a sus
45 implementaciones. Aunque en las realizaciones precedentes se ha descrito la decodificación y codificación Huffman como único esquema de codificación de entropía, también pueden emplearse otros esquemas de codificación de entropía. Por otra parte, la implementación de un codificador de entropía o un decodificador de entropía dista de ser necesaria. En consecuencia, aunque la descripción de las realizaciones anteriores se ha enfocado principalmente en el códec ACC-ELD, pueden emplearse también otros códecs para proporcionar los trenes de datos de entrada y para
50 decodificar el tren de datos de salida en el lado del participante. Por ejemplo, puede emplearse cualquier códec que se base, por ejemplo, en una sola ventana sin conmutación de longitud de bloque.

[0216] Como la descripción precedente de las realizaciones mostró en las Fig. 8 y 11, por ejemplo, ha demostrado también que los módulos descritos en las mismas no son obligatorios. Por ejemplo, un aparato según una
55 realización de la presente invención puede lograrse simplemente trabajando en la información espectral de las tramas.

[0217] Debe observarse que las realizaciones descritas anteriormente con respecto a las Fig. 6 a 12C pueden conseguirse de formas muy diferentes. Por ejemplo, puede conseguirse un aparato 500/1500 para mezclar una pluralidad de trenes de datos de entrada y su unidad de tratamiento 520/1520 basándose en dispositivos eléctricos y
60 electrónicos discretos tales como resistores, transistores, inductores y similares. Además, las realizaciones según la presente invención también pueden lograrse basándose solo en circuitos integrados, por ejemplo, en forma de sistemas sobre chip, SOC (SOC = System On Chip), procesadores tales como UPC, unidades centrales de tratamiento (CPU = Central Processing Unit), unidades de procesamiento, GPU (GPU = Graphic Processing Unit) y otros circuitos integrados (CI), tales como circuitos integrados específicos de aplicaciones (ASIC).

65

[0218] También debe observarse que pueden emplearse dispositivos eléctricos que forman parte de la implementación discreta o son parte de un circuito integrado para diferentes propósitos y distintas funciones a través de la implementación de un aparato según una realización de la presente invención. Naturalmente, también puede emplearse una combinación de circuitos basándose en circuitos integrados y circuitos discretos para implementar una
5 realización según la presente invención.

[0219] Basándose en un procesador, las realizaciones según la presente invención también pueden implementarse basándose en un programa informático, un programa de software o un programa que se ejecuta en un
10 procesador.

[0220] Dicho de otro modo, dependiendo de ciertos requisitos de implementación de realizaciones de los procedimientos de la invención, las realizaciones de los procedimientos de la invención pueden implementarse en hardware o en software. La implementación puede realizarse usando un medio de almacenamiento digital, en particular un disco, un CD o un DVD que tiene señales legibles electrónicamente almacenadas que cooperan con un procesador
15 u ordenador programable, de manera que se lleve a cabo una realización de la invención. En general, una realización de la presente invención es por lo tanto un producto de programa informático con un código de programa almacenado en un soporte legible por máquina, siendo el código de programa operativo para llevar a cabo una realización del procedimiento de la invención cuando el producto del programa informático se ejecuta en un ordenador o un
20 procesador. Expresado en otros términos, las realizaciones de los procedimientos de la invención son por lo tanto un programa informático que tiene un código de programa para llevar a cabo al menos una de las realizaciones de los procedimientos de la invención, cuando el programa informático se ejecuta en un ordenador o procesador. Un procesador puede estar formado por un ordenador, una tarjeta de circuito integrado, una tarjeta inteligente, un circuito integrado específico de aplicación, un sistema sobre chip (SOC) o un circuito integrado CI.

25 Lista de signos de referencia

[0221]

100	Sistema de conferencia
30 110	Entrada
120	Decodificador
130	Sumador
140	Codificador
150	Salida
35 160	Terminal de conferencia
170	Codificador
180	Decodificador
190	Convertidor de tiempo/frecuencia
40 200	Cuantificador/codificador
210	Decodificador/decuantificador
220	Convertidor frecuencia/tiempo
250	Tren de datos
260	Trama
45 270	Bloques de información adicional
300	Frecuencia
310	Banda de frecuencia
50 500	Aparato
510	Tren de datos de entrada
520	Unidad de tratamiento
530	Tren de datos de salida
540	Trama
55 550	Trama de salida
560	Componente espectral
570	Flecha
580	Línea discontinua
60 700	Decodificador de trenes de bits
710	Lector de trenes de bits
720	Codificador Huffman
730	Decuantificador
740	Dimensionador
65 750	Primera unidad

	760	Segunda unidad
	770	Decodificador estéreo
	780	Decodificador PNS
	790	Decodificador TNS
5		
	800	Unidad de mezclado
	810	Mezclador espectral
	820	Módulo de optimización
	830	Mezclador SBR
10	850	Codificador de trenes de bits
	860	Tercera unidad
	870	Codificador TNS
	880	Codificador PNS
	890	Codificador estéreo
15		
	900	Cuarta unidad
	910	Dimensionador
	920	Cuantificador
	930	Codificador Huffman
20	940	Elemento de escritura de trenes de bits
	950	Módulo psicoacústico
	1500	Aparato
	1520	Unidad de tratamiento
25	1545	Valor de control
	1550	Trama de salida
	1555	Valor de control

REIVINDICACIONES

1. Un aparato (1500) para la generación de un tren de datos de salida (530) a partir de un primer tren de datos de entrada (510-1) y un segundo tren de datos de entrada (510-2), en el que los trenes de datos de entrada primero y segundo (510) comprenden cada uno una trama (540), en el que las tramas (540) comprenden cada una un valor de control (1545) y datos de carga útil asociados, de manera que el valor de control indica una forma en que los datos de carga útil representan al menos una parte de un dominio espectral de una señal de audio, que comprende:
- una unidad de procesador (1520) adaptada para comparar el valor de control (1545) de la trama (540) del primer tren de datos de entrada (510-1) y el valor de control (1545) de la trama (540) del segundo tren de datos de entrada (510-2) para producir un resultado de comparación, en el que la unidad de procesador (1520) está adaptada además para, si el resultado de comparación indica que los valores de control de las tramas de los trenes de datos de entrada primero y segundo son idénticos, generar el tren de datos de salida (530) que comprende una trama de salida (550) de manera que la trama de salida comprende un valor de control (1555) igual al de la trama de los trenes de datos de entrada primero y segundo y los datos de carga útil deducidos a partir de los datos de carga útil de las tramas (540) de los trenes de datos de entrada primero y segundo (510) mediante el tratamiento de los datos de audio en el dominio espectral, en el que la unidad de procesador está adaptada además para generar el tren de datos de salida deduciendo los datos de carga útil del tren de datos de salida a partir de los datos de carga útil de las tramas de los trenes de datos de entrada primero y segundo permaneciendo dentro de la forma de representación del dominio espectral, tal como se indica mediante los valores de control con uno de los valores de control (1545) de las tramas (540) de los trenes de datos de entrada primero y segundo (510) que indican si la al menos una parte del dominio espectral se describe en términos de información espectral o si debe ser sustituida por un parámetro PNS (2000-1, 2000-2) respectivo formado por la trama respectiva del tren de datos de entrada respectivo, con la unidad de procesador (1520) configurada para, si los valores de control (1545) de las tramas (540) de los trenes de datos de entrada primero y segundo (510) indican que la al menos una parte del dominio espectral debe sustituirse por el parámetro PNS (2000-1, 2000-2) respectivo formado por la trama respectiva del tren de datos de entrada respectivo, deducir los datos de carga útil del tren de datos de salida a partir de los datos de carga útil de las tramas de los trenes de datos de entrada primero y segundo permaneciendo dentro de la forma de representación del dominio espectral mediante la determinación de un parámetro PNS (2000-3) de la trama de salida (550) basándose en una combinación lineal de los parámetros PNS (2000-1, 2000-2) de los trenes de datos de entrada, los valores de control (1545) de las tramas (540) de los trenes de datos de entrada primero y segundo (510) indican cuadrículas de tiempo SBR presentes en las tramas (540) de los trenes de datos de entrada primero y segundo (510), con la unidad de procesador (1520) configurada para, si los valores de control (1545) de las tramas (540) de los trenes de datos de entrada primero y segundo (510) indican que están presentes las mismas cuadrículas de tiempo SBR en las tramas (540) de los trenes de datos de entrada primero y segundo (510), deducir los datos de carga útil del tren de datos de salida a partir de los datos de carga útil de las tramas de los trenes de datos de entrada primero y segundo permaneciendo dentro de la forma de representación del dominio espectral mediante el copiado de la cuadrícula de tiempo SBR en la trama de salida (550); y
- los valores de control (1545) de las tramas (540) de los trenes de datos de entrada primero y segundo (510) indican si los canales de entrada primero y segundo 2020 del tren de datos de entrada (510) respectivo están codificados en modo LR de codificación estéreo o en modo M/S de codificación estéreo, con la unidad de procesador (1520) configurada para, si los valores de control (1545) de las tramas (540) de los trenes de datos de entrada primero y segundo (510) indican que los canales de entrada primero y segundo 2020 de los trenes de datos de entrada primero y segundo (510) están codificados en el mismo de entre el modo LR de codificación estéreo y el modo M/S de codificación estéreo, deducir los datos de carga útil del tren de datos de salida a partir de los datos de carga útil de las tramas de los trenes de datos de entrada primero y segundo permaneciendo dentro de la forma de representación del dominio espectral mediante el mezclado directo en el modo respectivo de entre el modo L/R o M/S de codificación estéreo.
2. El aparato (1500) según la reivindicación 1, en el que la unidad de procesador (1520) está adaptada además para transformar los datos de carga útil de la trama (540) de uno de los trenes de datos de entrada primero y segundo (510) a una representación de los datos de carga útil de la trama del otro de los trenes de datos de entrada primero y segundo (510), cuando el resultado de comparación indica que los valores de control (1545) de los trenes de datos de entrada primero y segundo (510) no son idénticos antes de generar la trama de salida (550) que comprende un valor de control (555) igual al de la trama (540) del otro de los trenes de datos de entrada primero y segundo (510) y los datos de carga útil deducidos a partir de los datos de carga útil de las tramas del tren de datos de entrada y la representación transformada del otro tren de datos de entrada mediante el tratamiento de los datos de audio en el dominio espectral.
3. El aparato (1500) según la reivindicación 1 o 2, en el que la unidad de procesador (1520) está adaptada para generar la trama de salida de manera que se mantiene una distribución de niveles de cuantificación con respecto a al menos una parte de al menos una de las tramas de los trenes de datos de entrada primero y segundo.
4. El aparato (1500) según una cualquiera de las reivindicaciones 1 a 3, en el que el aparato (500) está

adaptado para el tratamiento de una pluralidad de trenes de datos de entrada (510) que comprende más de dos trenes de datos de entrada (510), comprendiendo la pluralidad de trenes de datos de entrada (510) los trenes de datos de entrada primero y segundo.

5 5. Un procedimiento para la generación de un tren de datos de salida (530) a partir de un primer tren de datos de entrada (510) y un segundo tren de datos de entrada (510), en el que los trenes de datos de entrada primero y segundo (510) comprenden cada uno una trama (540), en el que la trama (540) comprende el valor de control (1545) y los datos de carga útil asociados, indicando el valor de control (1545) una forma en que los datos de carga útil representan al menos a una parte de un dominio espectral de una señal de audio, que comprende:

10

la comparación del valor de control (1545) de la trama (540) del primer tren de datos de entrada (510-1) y el valor de control (1545) de la trama (540) del segundo tren de datos de entrada (510-2) para producir un resultado de comparación; y

15

si el resultado de comparación indica que los valores de control de las tramas de los trenes de datos de entrada primero y segundo son idénticos, la generación del tren de datos de salida (530) que comprende una trama de salida (550), de manera que la trama de salida (550) comprende un valor de control (1555) igual al de la trama (540) de los trenes de datos de entrada primero y segundo (510) y los datos de carga útil deducidos a partir de los datos de carga útil de las tramas de los trenes de datos de entrada primero y segundo mediante el tratamiento de los datos de audio en el dominio espectral,

20

en el que la generación del tren de datos de salida comprende la deducción de los datos de carga útil del tren de datos de salida a partir de los datos de carga útil de las tramas de los trenes de datos de entrada primero y segundo permaneciendo dentro de la forma de representación del dominio espectral, tal como se indica mediante los valores de control con uno de entre

25

los valores de control (1545) de las tramas (540) de los trenes de datos de entrada primero y segundo (510) indican si la al menos una parte del dominio espectral se describe en términos de información espectral o debe sustituirse por un parámetro PNS (2000-1, 2000-2) respectivo formado por la trama respectiva del tren de datos de entrada respectivo, con deducción, si los valores de control (1545) de las tramas (540) de los trenes de datos de entrada primero y segundo (510) indican que la al menos una parte del dominio espectral debe sustituirse por el parámetro PNS (2000-1, 2000-2) respectivo formado por la trama respectiva del tren de datos de entrada respectivo, de los datos de carga útil del tren de datos de salida a partir de los datos de carga útil de las tramas de los trenes de datos de entrada primero y segundo permaneciendo dentro de la forma de representación del dominio espectral que comprende la determinación de un parámetro PNS (2000-3) de la trama de salida (550) basándose en una combinación lineal de los parámetros PNS (2000-1, 2000-2) de los trenes de datos de entrada,

35

los valores de control (1545) de las tramas (540) de los trenes de datos de entrada primero y segundo (510) indican cuadrículas de tiempo SBR presentes en las tramas (540) de los trenes de datos de entrada primero y segundo (510), con la deducción, si los valores de control (1545) de las tramas (540) de los trenes de datos de entrada primero y segundo (510) indican que están presentes las mismas cuadrículas de tiempo SBR en las tramas (540) de los trenes de datos de entrada primero y segundo (510), de los datos de carga útil del tren de datos de salida a partir de los datos de carga útil de las tramas de los trenes de datos de entrada primero y segundo permaneciendo dentro de la forma de representación del dominio espectral que comprende el copiado de la cuadrícula de tiempo SBR en la trama de salida (550); y

40

los valores de control (1545) de las tramas (540) de los trenes de datos de entrada primero y segundo (510) indican si los canales de entrada primero y segundo 2020 del tren de datos de entrada (510) respectivo están codificados en modo LR de codificación estéreo o en modo M/S de codificación estéreo, con la deducción, si los valores de control (1545) de las tramas (540) de los trenes de datos de entrada primero y segundo (510) indican que los canales de entrada primero y segundo 2020 de los trenes de datos de entrada primero y segundo (510) están codificados en el mismo de entre el modo LR de codificación estéreo y el modo M/S de codificación estéreo, de los datos de carga útil del tren de datos de salida a partir de los datos de carga útil de las tramas de los trenes de datos de entrada primero y segundo permaneciendo dentro de la forma de representación del dominio espectral

45

que comprende el mezclado directo en el modo respectivo de entre el modo L/R o M/S de codificación estéreo.

50

6. Programa para llevar a cabo, cuando se ejecuta en un procesador, un procedimiento para la generación de un tren de datos de salida según la reivindicación 5.

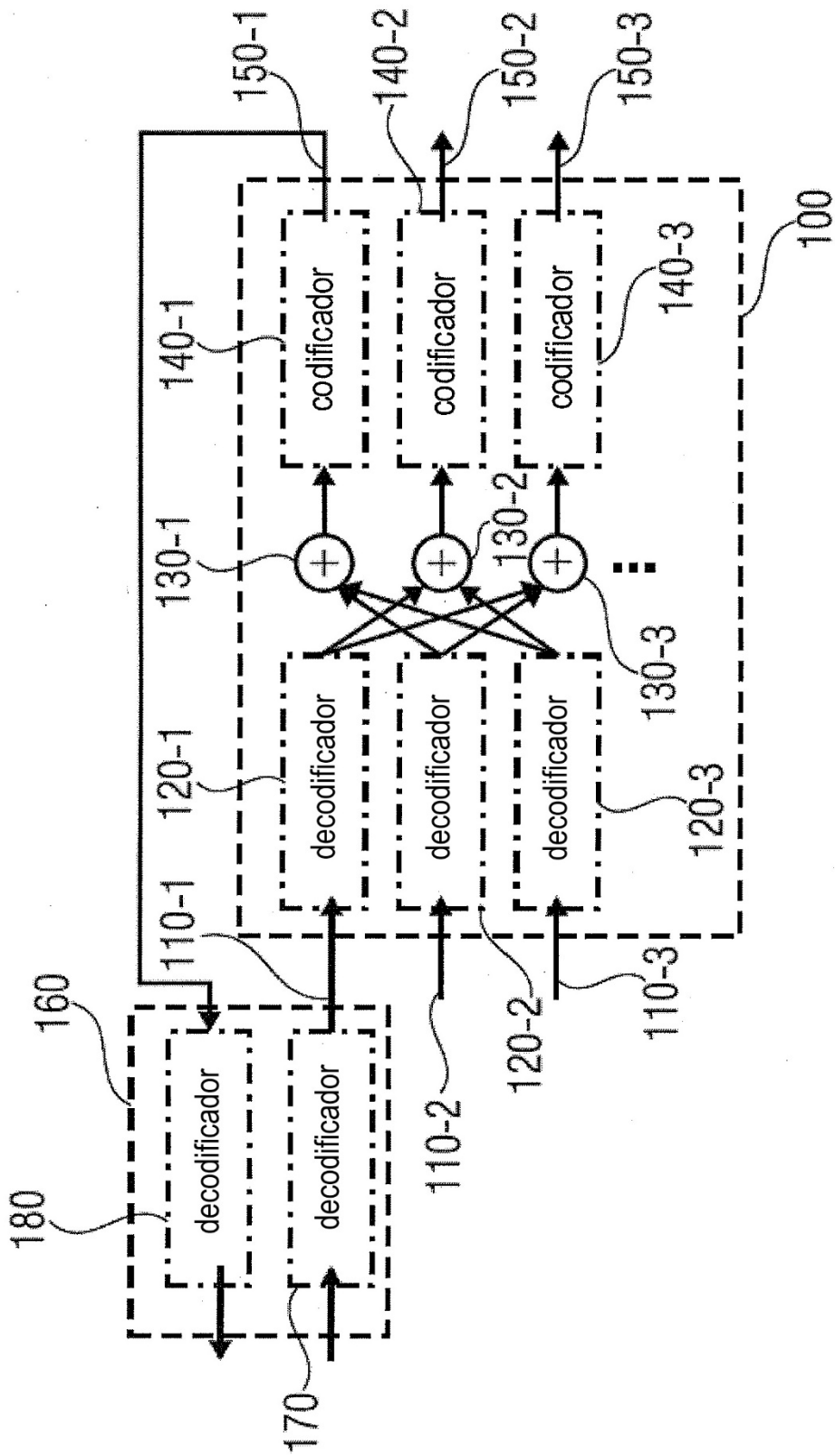


FIG 1

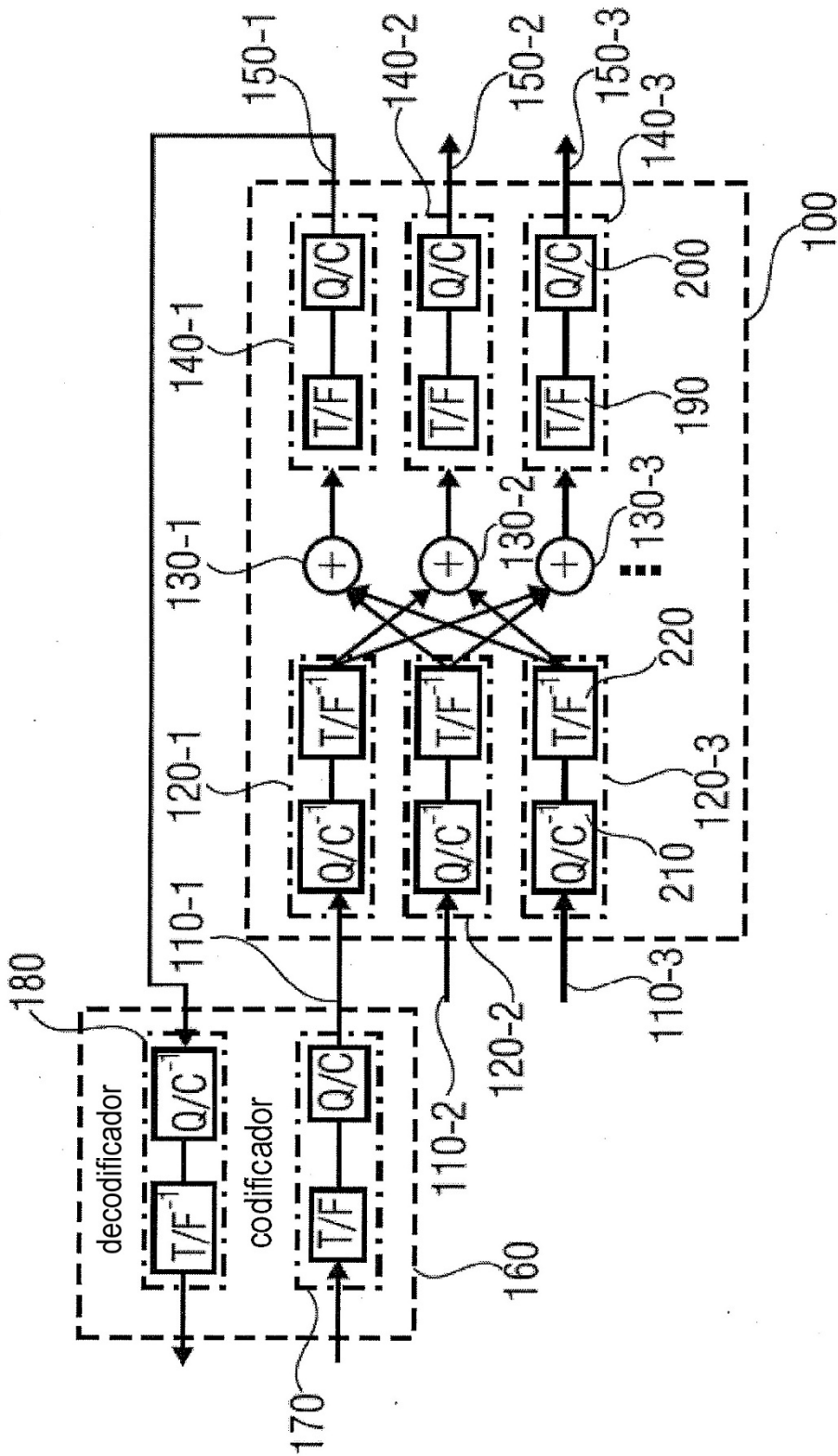


FIG 2

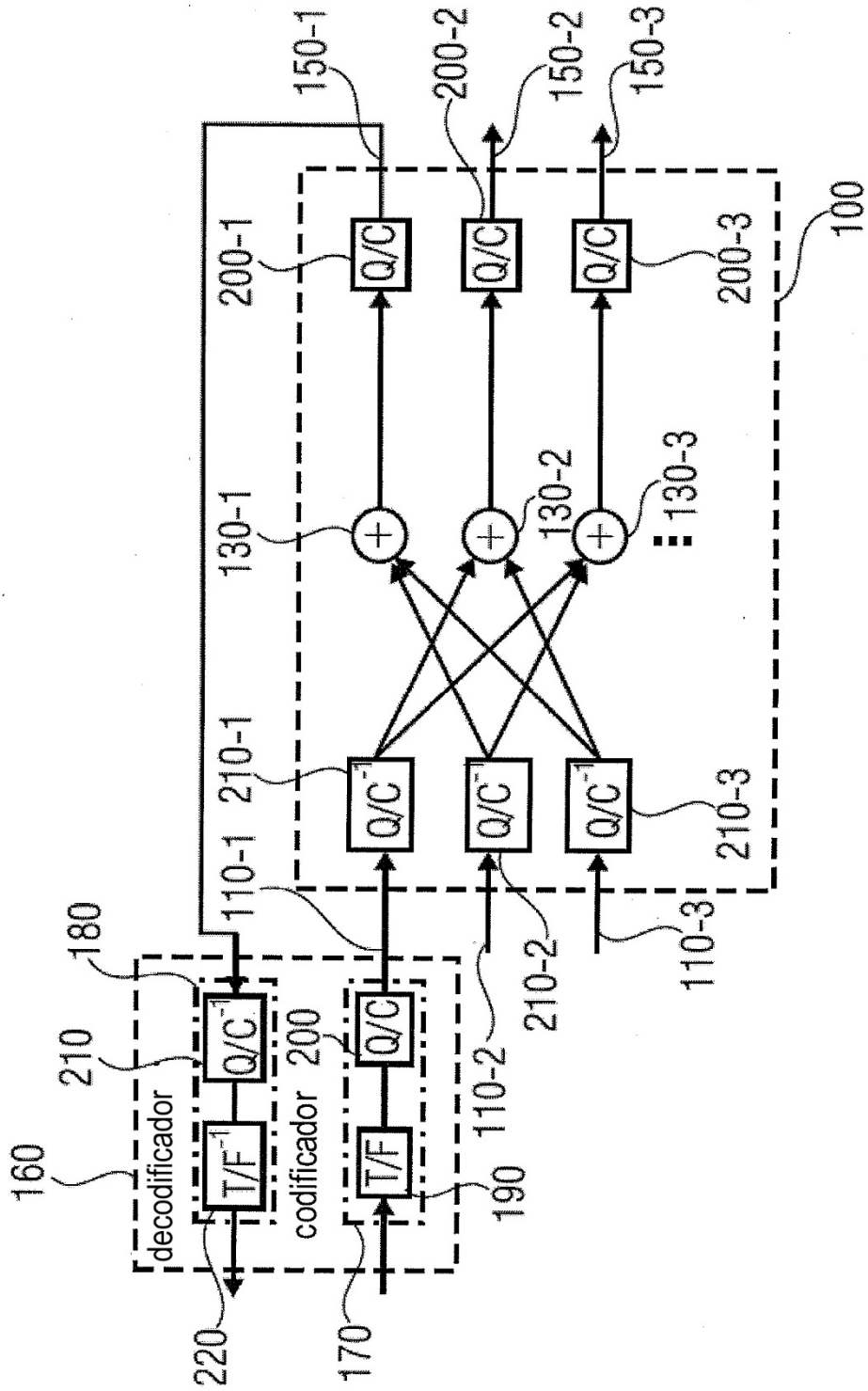


FIG 3

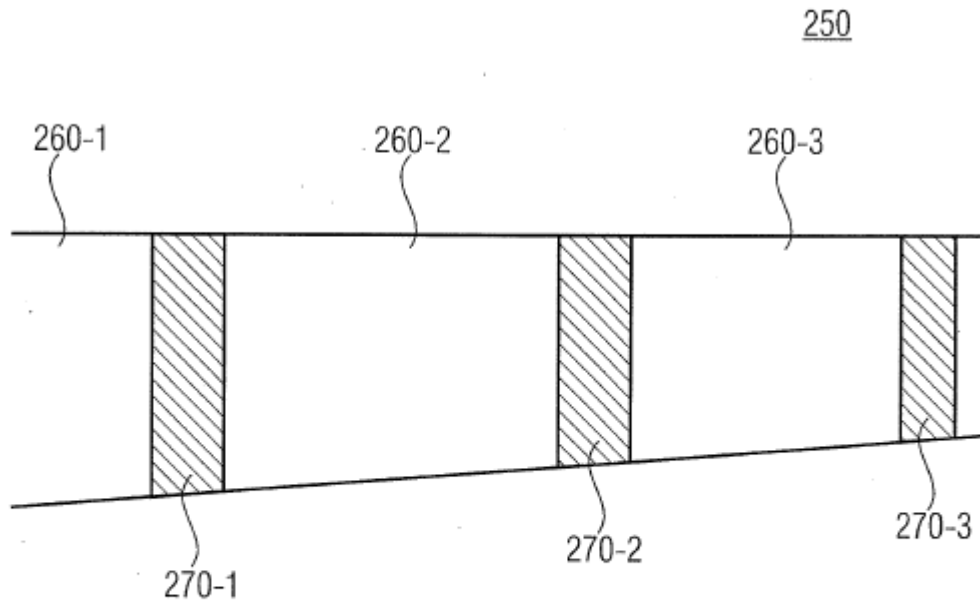


FIG 4

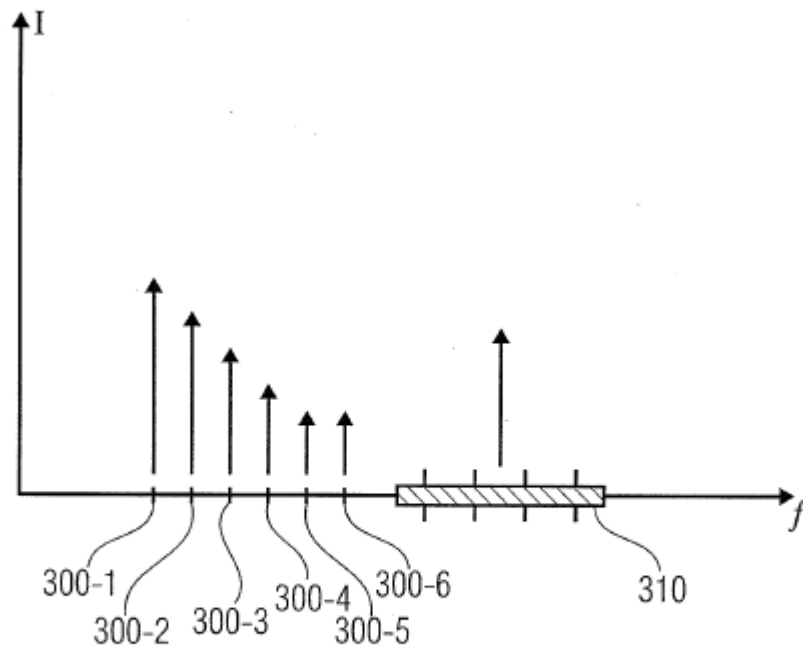


FIG 5

500

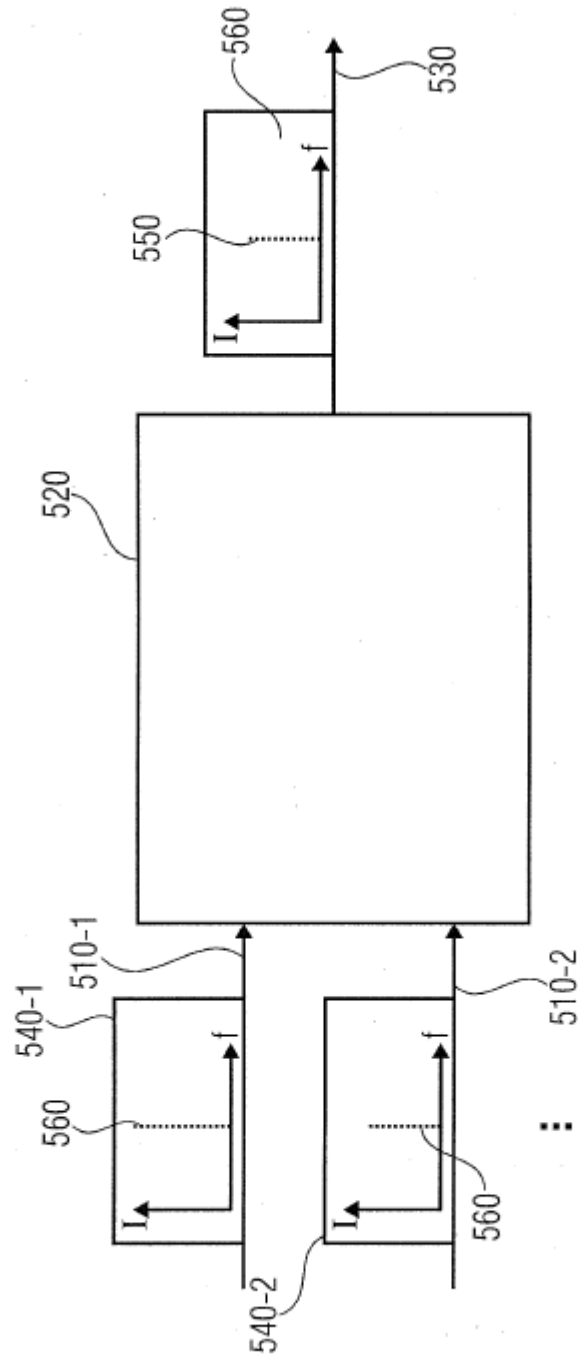


FIG 6

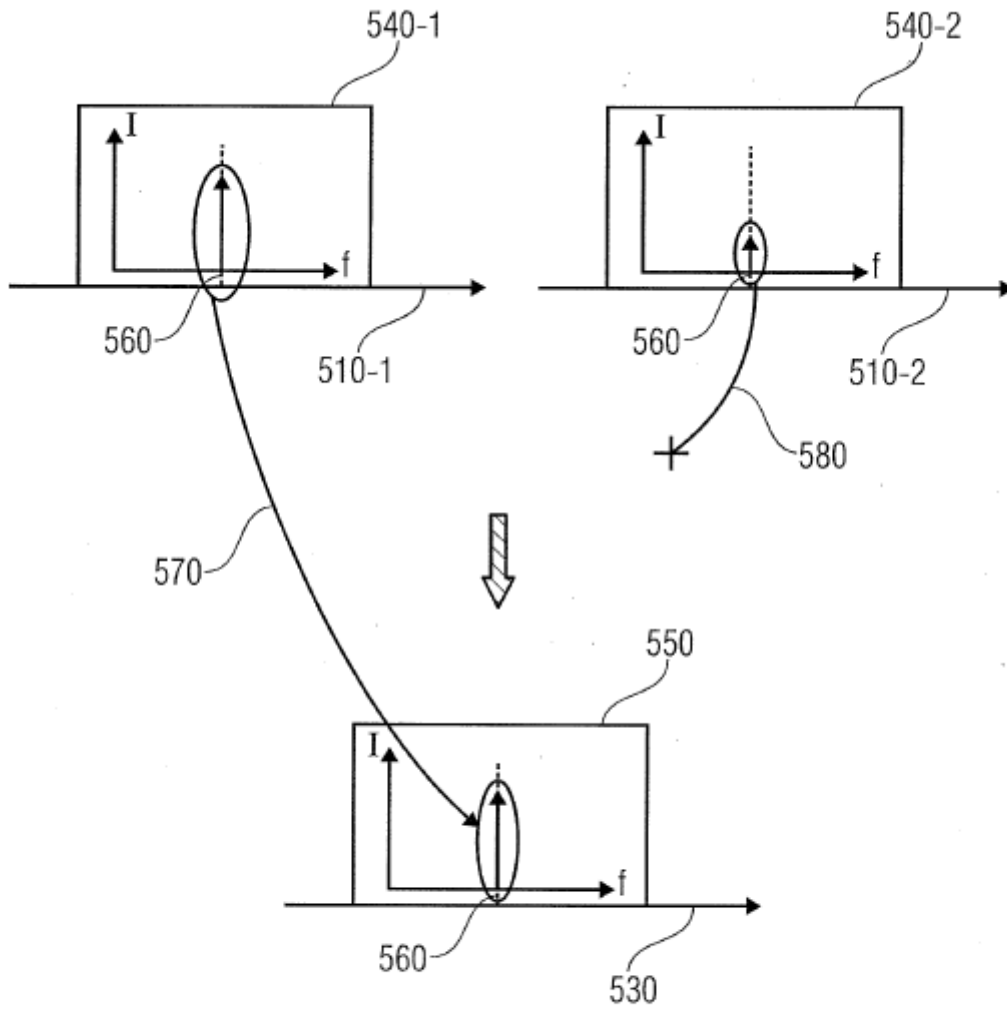


FIG 7

500

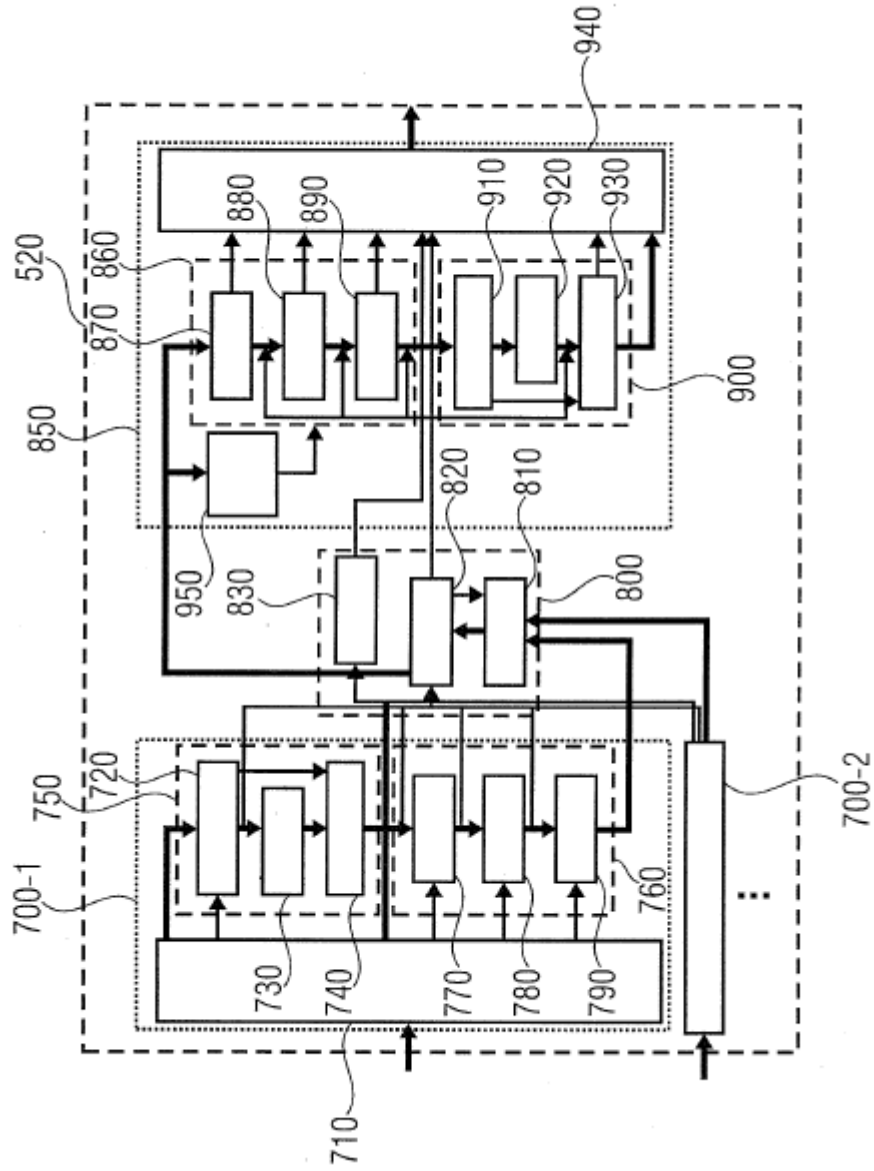


FIG 8

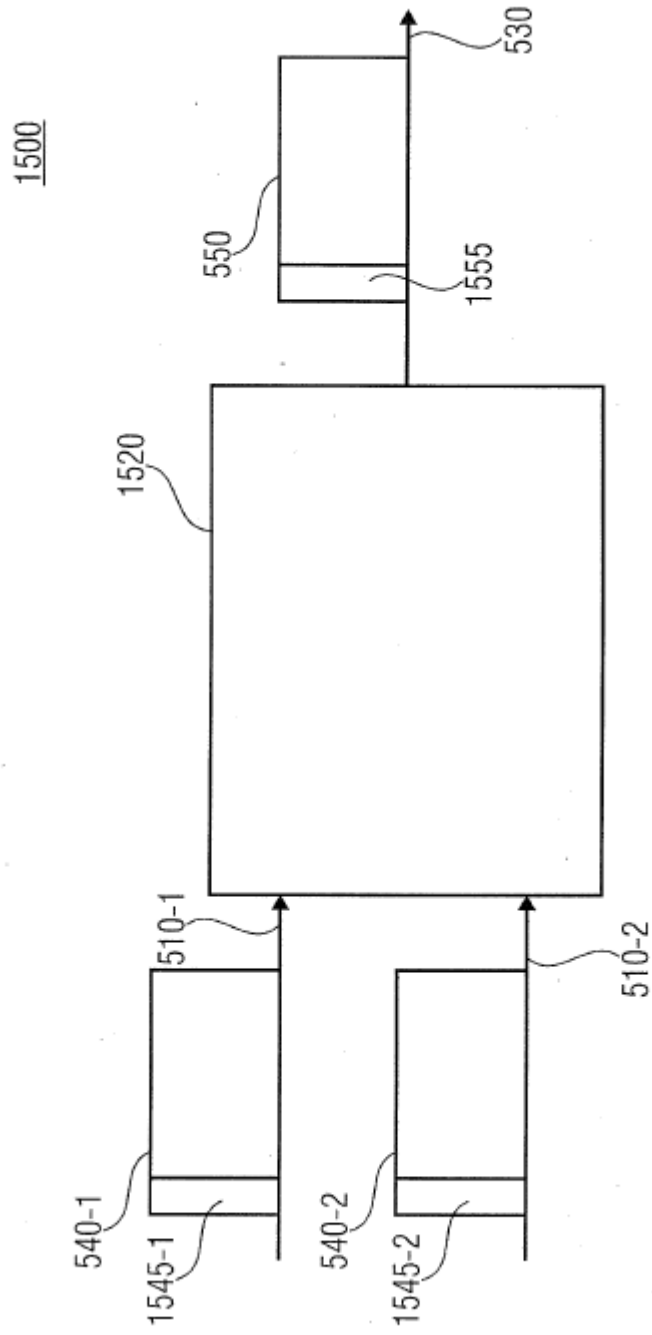


FIG 9

1500

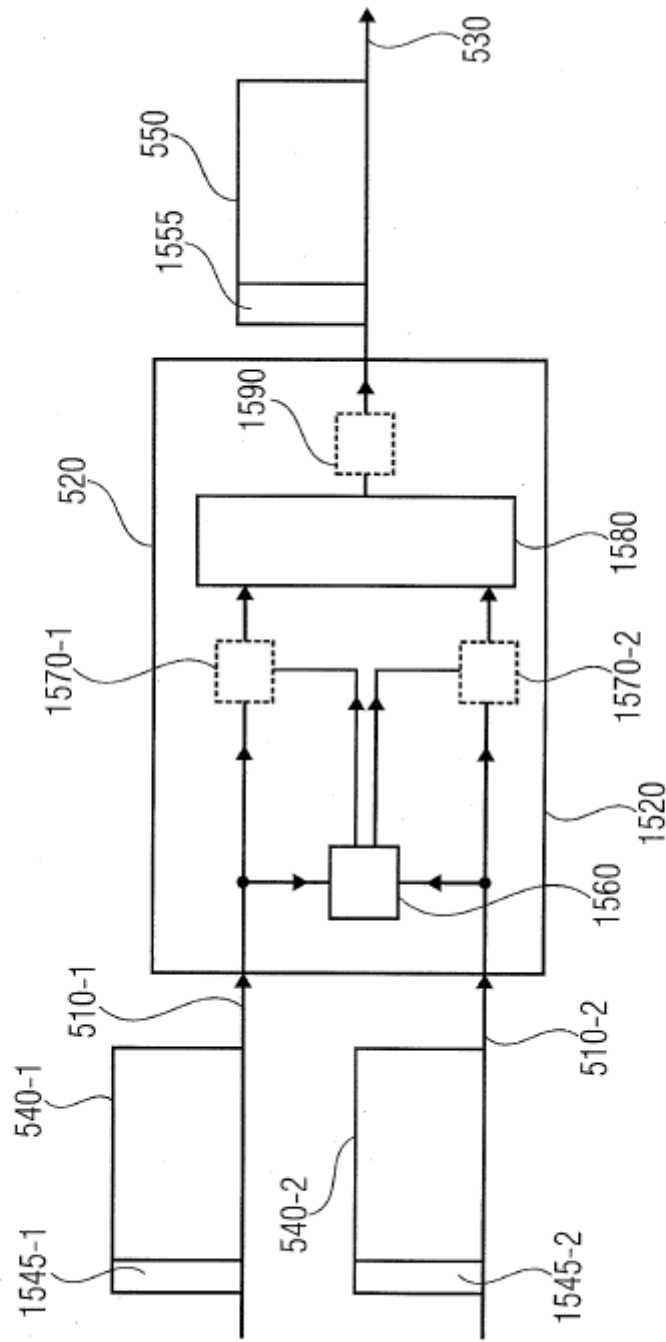


FIG 10

1500

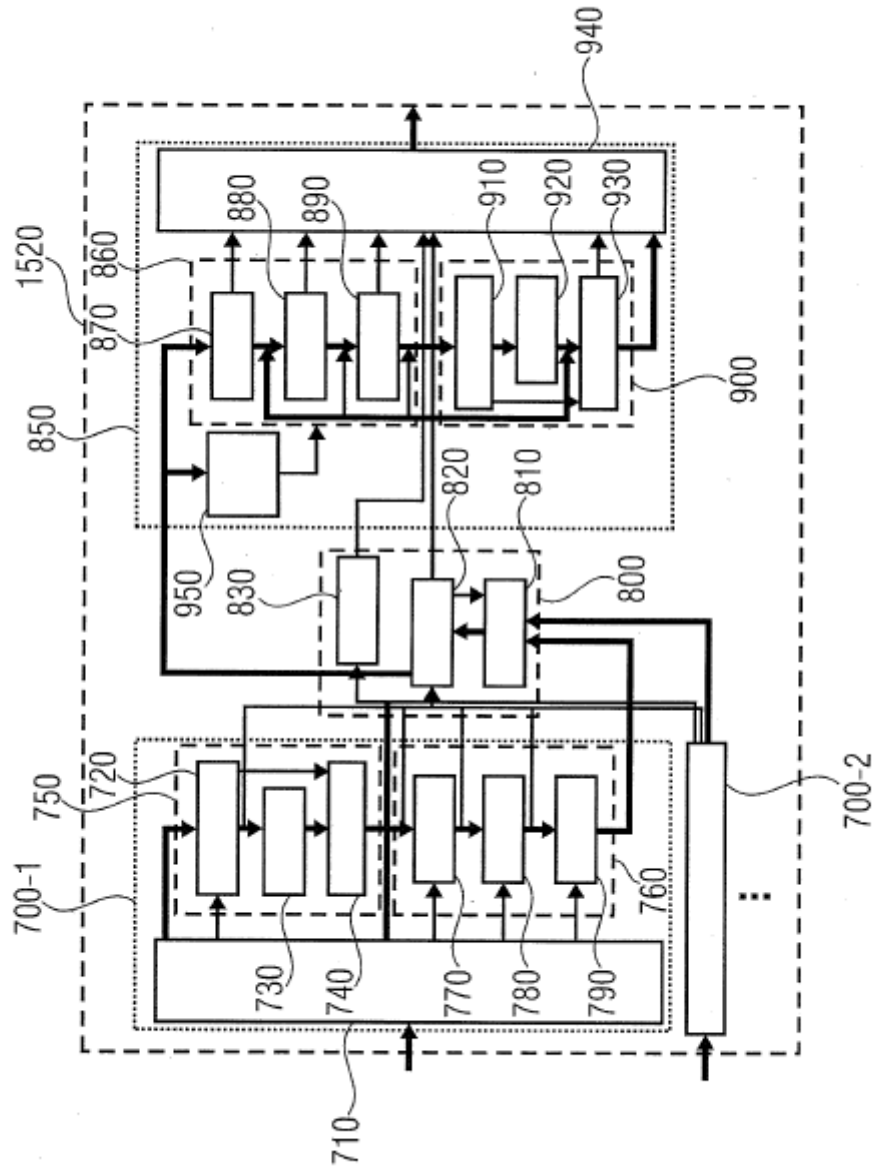


FIG 11

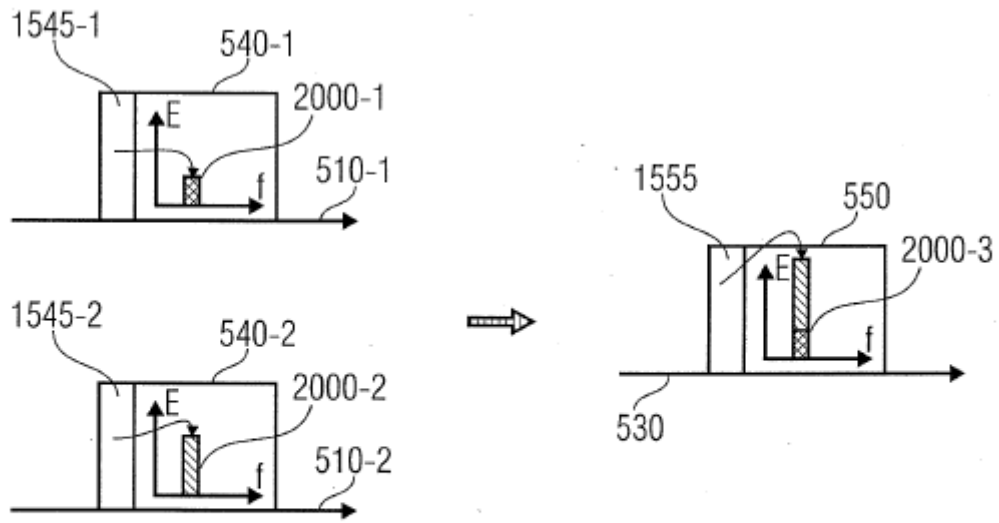


FIG 12A

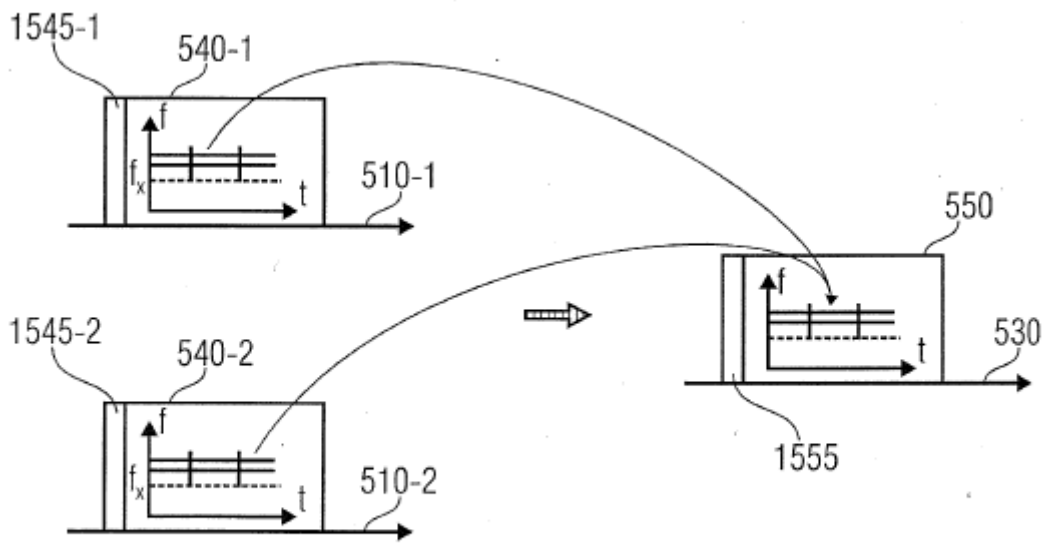


FIG 12B

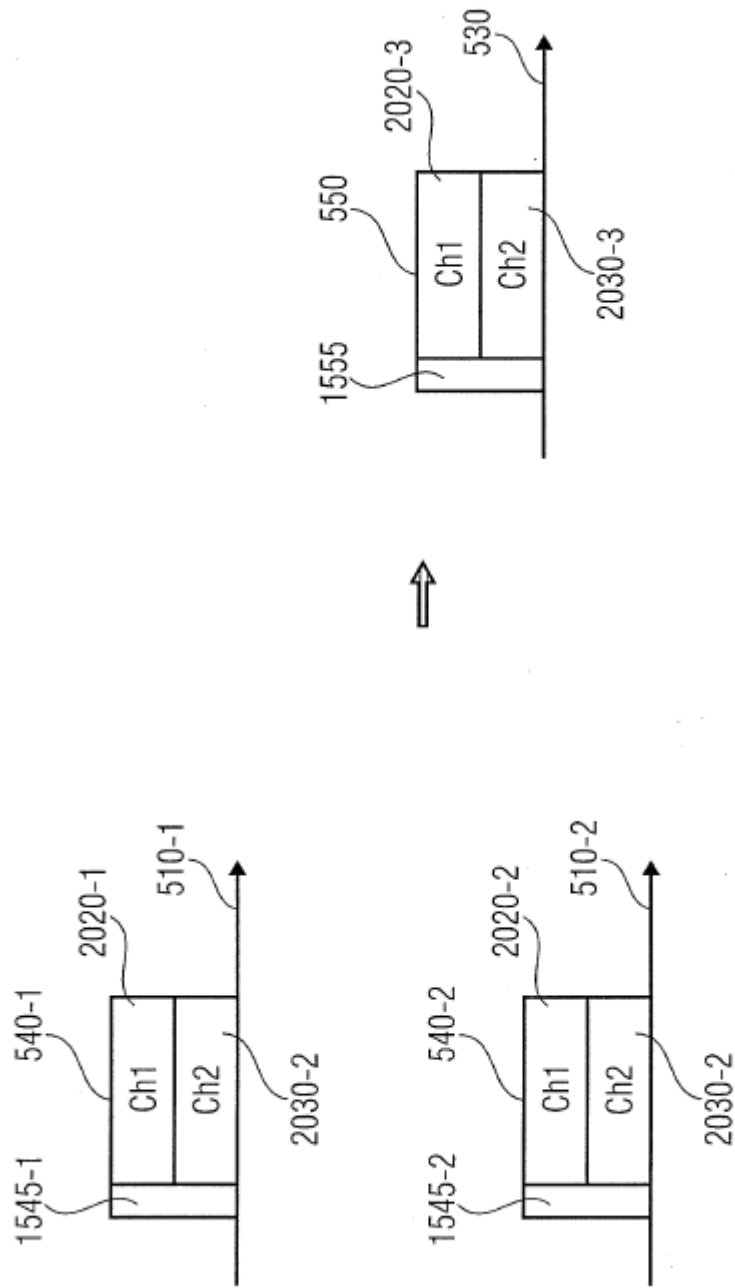


FIG 12C