

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 665 766**

51 Int. Cl.:

G10L 19/08 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **04.03.2009** E 11162197 (5)

97 Fecha y número de publicación de la concesión europea: **24.01.2018** EP 2378518

54 Título: **Mezclado de corrientes de datos de entrada y generación a partir de ahí de una corriente de datos de salida**

30 Prioridad:

04.03.2008 US 33590 P

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

27.04.2018

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:

SALVA FERRER, Joan

ES 2 665 766 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 corrientes de datos de entrada y generación a partir de ahí de una corriente de datos de salida

5 **[0001]** Las realizaciones según la presente invención se refieren al mezclado de una pluralidad de corrientes de datos de entrada para obtener una corriente de datos de salida y generar una corriente de datos de salida al mezclar las primeras y segundas corrientes de datos de entrada, respectivamente. La corriente de datos de salida por ejemplo puede emplearse en el campo de sistemas de conferencia incluyendo sistemas de vídeo conferencia y sistemas de teleconferencia.

10 **[0002]** En muchas aplicaciones, más de una señal de audio se va a procesar de tal manera que del número de señales de audio, una señal o al menos un número reducido de señales se va a generar, que al menos se refiere como "mezclado". El proceso de mezclar señales de audio, por lo tanto, puede referirse como formar en haces varias señales de audio individuales en una señal resultante. Este proceso se emplea por ejemplo cuando se crean
15 piezas de música de un disco compacto ("doblaje"). En este caso, diferentes señales de audio de diferentes instrumentos junto con una o más señales de audio que comprenden desempeños vocales (canto) se mezclan típicamente en una canción.

20 **[0003]** Campos adicionales de aplicación, en los que el mezclado juega un papel importante, son sistemas de conferencia de vídeo y sistemas de teleconferencia. Este sistema típicamente es capaz de conectar a varios participantes distribuidos espacialmente en una conferencia al emplear 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 en retorno. Esta señal resultante o señal de salida comprende las señales de audio de todos los otros participantes de la conferencia.

25 **[0004]** En sistemas de conferencias digitales modernos una cantidad de metas y aspectos parcialmente contradictorios compiten entre sí. 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 señales musicales), se deben tomar en consideración.
30 Aspectos adicionales que pueden tener que considerarse también cuando se diseñan e implementan sistemas de conferencias son el ancho de banda disponible y las cuestiones de retraso.

[0005] Por ejemplo, cuando se balancea la calidad por una parte y el ancho de banda por otra parte, es inevitable en la mayoría de los casos un compromiso. Sin embargo, mejoras referentes a la calidad pueden lograrse
35 al implementar técnicas modernas de codificación y decodificación tales como la técnica de Codec de Audio Avanzado – Retraso Bajo Mejorado (AAC-ELD, AAC = Advanced Audio Codec; ELD = Enhanced Low Delay). Sin embargo, la calidad alcanzable puede ser afectada de forma negativa en sistemas que emplean estas técnicas modernas por problemas y aspectos más fundamentales.

40 **[0006]** Por nombrar solo un reto a cumplir, todas las transmisiones de señal digital enfrentan el problema de una cuantificación necesaria, que puede, al menos en principio, ser evitada bajo 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 que se va a procesar. Para contra-atacar distorsiones posibles y audibles, se puede estar tentado a incrementar el número de niveles de cuantificación y por lo tanto incrementar por consiguiente
45 la resolución de cuantificación. Esto, sin embargo, lleva a un número mayor de valores de señal que se van a transmitir y por lo tanto a un aumento en la cantidad de datos que se van a transmitir. En otras palabras, el mejorar la calidad al reducir posibles distorsiones introducidas por ruido de cuantificación puede, bajo ciertas circunstancias, incrementar la cantidad de datos que se van a transmitir y eventualmente violar las restricciones de ancho de banda impuestas en un sistema de transmisión.

50 **[0007]** En el caso de los sistemas de conferencia, los retos de mejorar una compensación entre calidad, ancho de banda disponible, y otros parámetros, incluso pueden ser más complicados por el hecho de que típicamente más de una señal de audio de entrada se va a procesar. Por lo tanto, las condiciones frontera impuestas por más de una señal de audio tendrán que considerarse cuando se genera la señal de salida o la señal resultante
55 producida por el sistema de conferencia.

[0008] En especial, en vista del reto adicional de implementar sistemas de conferencias con un retraso suficientemente bajo para permitir una comunicación directa entre los participantes de una conferencia sin introducir retrasos sustanciales que puedan considerarse inaceptables por los participantes, aumenta adicionalmente el reto.

[0009] En implementaciones de bajo retraso de sistemas de conferencias, fuentes de retraso típicamente se restringen en términos de su número, que por otra parte pueden llevar al reto de procesar los datos fuera del dominio de tiempo, en donde el mezclado de las señales de audio puede lograrse al superponer o agregar las 5 señales respectivas.

[0010] Hablando de manera general, es favorable seleccionar una compensación entre calidad, ancho de banda disponible y otros parámetros adecuados para sistemas de conferencia de forma cuidadosa para enfrentar el gasto general de procesamiento para mezclar en tiempo real, reducir la cantidad de equipo físico requerida y 10 mantener los costes en términos de equipo físico y gastos generales de transmisión razonables, sin comprometer la calidad de audio.

[0011] Para reducir una cantidad de datos transmitidos, codecs de audio modernos a menudo utilizan herramientas altamente sofisticadas para describir información espectral referente a componentes espectrales de 15 una señal de audio respectiva. Al utilizar estas herramientas, que se basan en fenómenos psico-acústicos y resultados de examen, puede lograrse una compensación mejorada entre parámetros parcialmente contradictorios y condiciones frontera tales como la calidad de la señal de audio reconstruida a partir de datos transmitidos, complejidad de computación, velocidad de bits y parámetros adicionales.

[0012] Ejemplos de tales herramientas son por ejemplo sustituciones de interferencia perceptual (PNS = Perceptual Noise Substitution), conformado de interferencia temporal (TNS = Temporal Noise Shaping), y replicación de banda espectral (SBR = Spectral Band Replication), por nombrar solos unos cuantos. Todas esas técnicas se basan en describir al menos parte de la información espectral con un número reducido de bits, de tal manera que, en comparación con una corriente de datos con base en no utilizar estas herramientas, pueden asignarse más bits para 25 partes espectralmente importantes del espectro. Como consecuencia, mientras que se mantiene la velocidad de bits, un nivel perceptible de calidad puede mejorarse al utilizar estas herramientas. De forma natural, puede seleccionarse una compensación diferente, es decir reducir el número de bits transmitidos por cuadro de datos de audio que mantienen la impresión de audio total. Diferentes compensaciones que se basan entre estos dos extremos también pueden ser igualmente bien realizadas.

[0013] Estas herramientas también pueden emplearse en aplicaciones de telecomunicaciones. Sin embargo, cuando más de dos participantes en esta situación de comunicaciones están presentes, puede ser muy ventajoso emplear un sistema de conferencias para mezclar dos o más corrientes de bits de más de dos participantes. Situaciones como estas ocurren en ambas, situaciones de teleconferencia y basadas puramente en audio, así como 35 situaciones de vídeo conferencia.

[0014] Un sistema de conferencias que opera en un dominio de frecuencia por ejemplo se describe en US 2008/0097764 A1 que realiza el mezclado actual en el dominio de frecuencia y por lo tanto omite la retransformación de las señales de audio de entrada de regreso al dominio de tiempo. 40

[0015] Sin embargo, el sistema de conferencias aquí descrito no toma en cuenta las posibilidades de herramientas como se ha 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 conferencias requiere etapas de transformación adicionales para reconstruir las señales de audio que se proporcionan al sistema de conferencias al menos a un grado tal que las señales de audio respectivas estén presentes en el dominio de 45 frecuencia. Además, la señal de audio mezclada resultante también se requiere para retransformar con base en las herramientas adicionales anteriormente mencionadas. Estas etapas de retransformación y transformación requieren, sin embargo, una aplicación de algoritmos complejos que pueden llevar a una complejidad computacional incrementada y, por ejemplo, en el caso de aplicaciones críticas energéticamente, portátiles, a un consumo 50 incrementado de energía y por lo tanto a un tiempo operacional limitado.

[0016] El documento US 5,717,764 describe un sistema de compresión de audio de percepción de cinco canales que codifica cinco canales de matriz intercambiando entre 14 modos de codificación, utilizando cada uno un conjunto diferente respectivo de canales de matriz. Seis modos son para la codificación de los tres canales frontales 55 y ocho de los modos son para la codificación de los dos canales posteriores. El índice de codificación de percepción del canal de matriz se controla mediante el ajuste de los umbrales de ruido individuales como una función de un umbral de enmascaramiento global.

[0017] Es por lo tanto un problema que debe ser resuelto por las realizaciones según la presente invención

para permitir una compensación mejorada entre calidad, ancho de banda disponible y otros parámetros adecuados para sistemas de conferencias, o permitir una reducción de complejidad computacional requerida en un sistema de conferencias como se ha descrito anteriormente.

- 5 **[0018]** Este objetivo se logra por un aparato según la reivindicación 1, un procedimiento según la reivindicación 10 y un programa informático para llevar a cabo un procedimiento según la reivindicación 11. Las realizaciones según la presente invención se describirán de aquí en adelante haciendo referencia a las siguientes figuras.
- 10 La Figura 1 muestra un diagrama de bloques de un sistema de conferencia;
La Figura 2 muestra un diagrama de bloques del sistema de conferencias con base en un codec de audio general;
La Figura 3 muestra un diagrama de bloques de un sistema de conferencias que opera en un dominio de frecuencia mediante el uso de la tecnología de mezclado de corriente de bits;
La Figura 4 muestra un dibujo esquemático de corriente de datos que comprende una pluralidad de cuadros;
- 15 La Figura 5 ilustra formas diferentes de componentes espectrales y datos espectrales o información;
La Figura 6 ilustra un aparato para mezclar una pluralidad de corrientes de datos de entrada según una realización de la presente invención, con más detalle;
La Figura 7 ilustra un modo de operación del aparato de la Figura 6 según una realización de la presente invención;
La Figura 8 muestra un diagrama de bloques de un aparato para mezclar una pluralidad de corrientes de datos de entrada según una realización adicional de la presente invención en el contexto de un sistema de conferencia;
- 20 La Figura 9 muestra un diagrama de bloques simplificado de un aparato para generar una corriente de datos de salida según una realización de la presente invención;
La Figura 10 muestra un diagrama de bloques más detallado de un aparato para generar una corriente de datos de salida según una realización de la presente invención;
- 25 La Figura 11 muestra un diagrama de bloques de un aparato para generar una corriente de datos de salida a partir de una pluralidad de corrientes de datos de entrada según una realización adicional de la presente invención en el contexto de un sistema de conferencia;
La Figura 12a ilustra una operación de un aparato para generación de corriente de datos de salida según una realización de la presente invención para una implementación PNS;
- 30 La Figura 12b ilustra una operación de un aparato para generación de corriente de datos de salida según una realización de la presente invención para una implementación SBR; y
La Figura 12c ilustra una operación de un aparato para generación de corriente de datos de salida según una realización de la presente invención para una implementación M/S.
- 35 **[0019]** Con respecto a las Figuras 4 a 12C, diferentes realizaciones según la presente invención se describirán con más detalle. Sin embargo, antes de describir estas realizaciones con más detalle, primero con respecto a las Figuras 1 a 3, se dará una breve introducción en vista de los retos y demandas que pueden volverse importantes en el marco de sistemas de conferencias.
- 40 **[0020]** La Figura 1 muestra un diagrama de bloques de un sistema de conferencias 100, que también puede ser referido como una unidad de control de múltiples puntos (MCU = multi-point control unit). Como será aparente a partir de la descripción referente a su funcionalidad, el sistema de conferencias 100, como se muestra en la Figura 1, es un sistema que opera en el dominio de tiempo.
- 45 **[0021]** El sistema de conferencias 100, como se muestra en la Figura 1, está adaptado para recibir una pluralidad de corrientes de datos de entrada mediante un número apropiado de entradas 110-1, 110-2, 110-3, ... de los cuales en la Figura 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 la primera corriente 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
- 50 la tercera entrada 110-3 se acopla a un tercer decodificador 120-3.
- [0022]** El sistema de conferencias 100 comprende además un número apropiado de sumadores 130-1, 130-2, 130-3, ... de los cuales de nuevo se ilustran tres en la Figura 1. Cada uno de los sumadores se asocia con una de las alimentaciones 110 del sistema de conferencias 100. Por ejemplo, el primer sumador 130-1 se asocia con la
- 55 primera entrada 110-1 y el decodificador correspondiente 120-1.
- [0023]** 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. En otras palabras, el primer sumador 130-1 se acopla a todos los decodificadores 120, aparte del primer decodificador 120-1. Según esto, el segundo sumador 130-2 se acopla a

todos los decodificadores 120, aparte del segundo decodificador 120-2.

[0024] Cada uno de los sumadores 130 comprende además una salida que se acopla cada una a un codificador 140. Por lo tanto, el primer sumador 130-1 se acopla a manera de salida al primer codificador 140-1. Según esto, el segundo y tercer sumadores 130-2, 130-3 también se acoplan al segundo y tercer codificadores 140-2, 140-3, respectivamente.

[0025] A su vez, cada uno de los codificadores 140 se acopla a la salida respectiva 150. En otras palabras, el primer codificador, se acopla por ejemplo a una primera salida 150-1. El segundo y tercer codificadores 140-2, 140-3 también se acoplan a segundas y terceras salidas 150-2, 150-3, respectivamente.

[0026] Para poder describir la operación de un sistema de conferencias 100 como se muestra en la Figura 1 con más detalle, la Figura 1 también muestra un terminal de conferencias 160 de un primer participante. El terminal de conferencias 160 puede ser por ejemplo un teléfono digital (por ejemplo, un teléfono de red digital de servicios integrados (ISDN = Integrated Services Digital Network)), un sistema que comprende una infraestructura de voz-sobre-IP o un terminal similar.

[0027] El terminal de conferencias 160 comprende un codificador 170 que se acopla a la primera entrada 110-1 del sistema de conferencias 100. El terminal de conferencias 160 comprende también un decodificador 180 que se acopla a la primera salida 150-1 del sistema de conferencias 100.

[0028] Similares terminales de conferencias 160 también pueden estar presentes en los sitios de participantes adicionales. Estos terminales de conferencias no se muestran en la Figura 1, solamente por razones de simplicidad. También cabe destacar que no se requiere con diferencia que el sistema de conferencias 100 y los terminales de conferencias 160 estén presentes físicamente en la vecindad inmediata entre sí. Los terminales de conferencias 160 y el sistema de conferencias 100 pueden disponerse en diferentes sitios, que, por ejemplo, pueden conectarse solo por medio de técnicas de redes de área amplia WAN (WAN = Wide Area Networks).

[0029] Los terminales de conferencias 160 pueden comprender además o estar conectados con componentes adicionales tales como micrófonos, amplificadores y bocinas o audífonos para permitir un intercambio de señales de audio con un usuario humano de una forma más comprensible. Estas no se muestran en la Figura 1 solamente por razones de simplicidad.

[0030] Como se ha indicado anteriormente, el sistema de conferencias 100 mostrado en la Figura 1 es un sistema que opera en el dominio de tiempo. Cuando, por ejemplo, el primer participante habla al micrófono (no mostrado en la Figura 1), el codificador 170 del terminal de conferencias 160 codifica la señal de audio respectiva en una corriente de bits correspondiente y transmite la corriente de bits a la primera entrada 110-1 del sistema de conferencias 100.

[0031] Dentro del sistema de conferencias 100, la corriente de bits se decodifica por el primer decodificador 120-1 y transforma de nuevo al dominio de tiempo. Ya que el primer decodificador 120-1 se acopla al segundo y tercer mezcladores 130-1, 130-3, la señal de audio, como se genera por el primer participante puede mezclarse en el dominio de tiempo al agregar simplemente la señal de audio reconstruidas con las señales de audio reconstruidas adicionales del segundo y tercer participantes, respectivamente.

[0032] Esto es también cierto para las señales de audio que se proporcionan por el segundo y tercer participantes que se reciben por la segunda y tercera alimentaciones 110-2, 110-3 y procesan por el segundo y tercer decodificadores 120-2, 120-3, respectivamente. Estas señales de audio reconstruidas del segundo y tercer participante se proporcionan entonces al primer mezclador 130-1, que a su vez proporciona la señal de audio agregada en el dominio de tiempo al primer codificador 140-1. El codificador 140-1 re-codifica la señal de audio sumada para formar una corriente de bits y proporciona la misma en la primera entrada 150-1 al terminal de conferencias del primer participante 160.

[0033] Similarmente, también el segundo y tercer codificadores 140-2, 140-3 codifican las señales de audio sumadas en el dominio de tiempo recibidas del segundo y tercer sumadores 130-2, 130-3, respectivamente, y transmite de vuelta los datos codificados a los participantes respectivos mediante la segunda y tercera salidas 150-2, 150-3, respectivamente.

[0034] Para realizar el mezclado actual, las señales de audio se decodifican completamente y agregan de

una forma no comprimida. Posteriormente, pueden realizarse de forma opcional un ajuste de nivel al comprimir 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 sencillos ascienden sobre o caen por debajo del intervalo de valores permitidos, de tal manera que los valores correspondientes son cortados (recortados). En el caso de una cuantificación de 16-bits, como por ejemplo se emplea en el caso de CDs, está disponible un intervalo de valores enteros entre -32768 y 32767 por valor de muestra.

[0035] Para contra-atacar una posible sobre o sub-dirección 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.

[0036] Cuando se codifican datos de audio en sistemas de conferencias tales como sistema de conferencias 100, como se muestra en la Figura 1, se aceptan algunas desventajas para realizar un mezclado en el estado no codificado de una forma que se alcance de manera más fácil. Además, las velocidades de datos de las señales de audio codificadas están limitadas adicionalmente a un intervalo menor de frecuencias transmitidas, ya que un ancho 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 del ancho de banda de la señal muestreada y se requiere que sea (al menos) el doble de grande que el ancho de banda.

[0037] La Unión de Telecomunicación Internacional (ITU = International Telecommunication Union) y su sector de estandarización de telecomunicaciones (ITU-T = telecommunication standardization sector) han desarrollado varias normas para sistemas de conferencias de multimedia. El H.320 es el protocolo de conferencia estándar para ISDN. H.323 define el sistema de conferencias estándar para una red basada en paquetes (TCP/IP). H.324 define sistemas de conferencias para redes de telefonía analógica y sistemas de radio telecomunicaciones.

[0038] Dentro de estas normas, no solo se define transmitir las señales, sino también codificar y procesar los datos de audio. El manejo o gestión de una conferencia se realiza por uno o más servidores, las así 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 por el procesamiento y distribución de los datos de vídeo y de audio de los diversos participantes.

[0039] Para lograr esto, 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 Figura 1 no solo muestra un diagrama de bloques de un sistema de conferencias 100, sino también un flujo de señal en esta situación de conferencia.

[0040] En el marco de las normas H.323 y H.320, los codecs de audio de la clase G.7xx se definen para operación en los sistemas de conferencia respectivos. La norma G.711 se utiliza para transmisiones ISDN, en sistemas de telefonía ligados o limitados por cable. A una frecuencia de muestreo de 8 kHz, la norma G.711 cubre un ancho de banda de audio entre 300 y 3400 Hz, requiriendo una velocidad de bits de 64 kbit/s a una profundidad (cuantificación) de 8-bits. La codificación se forma por una simple codificación logarítmica denominada Ley- μ o Ley-A que crea un retraso muy bajo de solo 0,125 ms.

[0041] La norma G.722 codifica un mayor ancho de banda de audio de 50 a 7000 Hz a una frecuencia de muestreo de 16 kHz. Como consecuencia, el codec logra una mejor calidad cuando se compara con los codecs de audio G.7xx de banda más estrecha a velocidades de bits de 48, 56 ó 64 Kbit/s a un retraso de 1,5 ms. Además, existen dos desarrollos adicionales, G.722.1 y G.722.2, que proporcionan calidad de habla comparable a velocidades de bits incluso menores. La G722.2 permite una selección de velocidad de bits entre 6,6 kbit/s y 23,85 kbit/s a un retraso de 25 ms.

[0042] La norma G.729 se emplea típicamente en el caso de comunicaciones de telefonía-IP, que también se refiere como comunicaciones de voz-sobre-IP (VoIP). El codec se optimiza para hablar 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 ancho de banda de audio comparable, cuando se comparan con la norma G.711. El algoritmo más complejo, sin embargo, crea un retraso de aproximadamente 15 ms.

[0043] Como una desventaja, los codecs G.7.xx se optimizan para codificación de habla y muestran, aparte

de un ancho de banda de frecuencia estrecha, problemas significantes cuando se codifica música junto con habla o música pura.

5 **[0044]** Por lo tanto, aunque el sistema de conferencias 100, como se muestra en la Figura 1, puede emplearse para una calidad aceptable cuando se transmiten y procesan señales de habla, señales de audio en general no se procesan satisfactoriamente cuando se emplean codecs de bajo retraso optimizados para habla.

10 **[0045]** En otras palabras, el emplear codecs para codificar y decodificar una señal de habla para procesar señales de audio en general, incluyendo por ejemplo señales de audio con música, no lleva a un resultado satisfactorio en términos de la calidad. Al emplear codecs de audio para codificar y decodificar señales de audio en general en el marco del sistema de conferencias 100, como se muestra en la Figura 1, es improbable la calidad. Sin embargo, como se establecerá en el contexto con la Figura 2 con más detalle, el emplear codecs de audio en general en este sistema de conferencias puede llevar a efectos adicionales, indeseados, tales como un retraso incrementado, por nombrar solo uno.

15

20 **[0046]** Sin embargo, antes de describir la Figura 2 con más detalle, cabe destacar que, en la presente descripción, se denotan objetos 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 se denote de otra forma en forma explícita o implícita, objetos denotados por los mismos o similares signos de referencia pueden implementarse de una forma similar o igual, por ejemplo, en términos de sus circuitos, programación, características u otros parámetros. Por lo tanto, objetos que aparecen en varias realizaciones de las figuras y que se denotan con signos de referencia iguales o similares pueden ser implementados que tienen las mismas especificaciones, parámetros y características. De forma natural, desviaciones y adaptaciones pueden ser implementadas también, por ejemplo, cuando las condiciones frontera u otros parámetros cambian de figura a figura, 25 o de realización a realización.

30 **[0047]** Además, resumiendo a continuación los signos de referencia se emplearán para denotar un grupo o clase de objetos, en vez de un objeto individual. En el marco de la Figura 1, esto ya se ha hecho, por ejemplo, cuando se denota la primera entrada como la entrada 110-1, la segunda entrada como entrada 110-2, y la tercera entrada como la entrada 110-3, mientras que las entradas se han discutido en términos del signo de referencia en resumen 110 solamente. En otras palabras, a menos que se anote explícitamente de otra forma, partes de la descripción que se refieren a objetos denotados con signos de referencia y resumen también pueden relacionarse a otros objetos que contienen los signos de referencia individuales correspondientes.

35 **[0048]** Ya que esto también es cierto para objetos denotados con los mismos o similares signos de referencia, ambas medidas ayudan a recortar la descripción y para describir las realizaciones aquí ilustradas de una forma más clara y concisa.

40 **[0049]** La Figura 2 muestra un diagrama de bloques de un sistema de conferencias tradicional 100 junto con un terminal de conferencias 160, ambos similares a aquellos mostrados en la Figura 1. El sistema de conferencias 100 mostrado en la Figura 2 también comprende alimentaciones 110, decodificadores 120, sumadores 130, codificadores 140 y salidas 150, que igualmente se interconectan en comparación con el sistema de conferencias 100 como se muestra en la Figura 1. El terminal de conferencias 160 mostrado en la Figura 2 comprende también de nuevo un codificador 170 y un decodificador 180.

45

[0050] Por lo tanto, se hace referencia a la descripción del sistema de conferencias 100 mostrado en la Figura 1.

50 **[0051]** Sin embargo, el sistema de conferencias 100 ilustrado en la Figura 2, así como el terminal de conferencias 160 mostrado en la Figura 2 se adaptan para utilizar un codec de audio general (COder - DECOder). 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 Figura 2 como "T/F", mientras que el cuantificador/codificadores 200 se etiquetan en la Figura 2 con "Q/C".

55

[0052] Los decodificadores 120, 180 comprenden cada uno a un decodificador/descuantificador 210, que se refiere en la Figura 2 como "Q/C-1" conectado en serie con un convertidor de frecuencia/tiempo 220, que se refiere en la Figura 2 como "T/F-1". Por razones de simplicidad solamente, el convertidor de tiempo/frecuencia 190, el cuantificador/codificador 200 y el decodificador/descuantificador 210, así como el convertidor de frecuencia/tiempo

220 se etiquetan como tal solo en el caso del codificador 140-3 y el decodificador 120-3. Sin embargo, la siguiente descripción también se refiere a los otros de estos elementos.

5 **[0053]** Empezando con un codificador tal como los codificadores 140 o el codificador 170, la señal de audio que se proporciona al convertidor de tiempo/frecuencia 190 se convierte desde el dominio de tiempo en un dominio de frecuencia o un dominio relacionado a 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 una corriente de bits, que después se proporciona, por ejemplo, a las salidas 150 del sistema de conferencias 100 en el caso del codificador 140.

10 **[0054]** En términos de los decodificadores tales como los decodificadores 120 o el decodificador 180, la corriente de bits que se proporciona a los decodificadores primero se decodifica y re-cuantifica para formar la representación espectral de al menos una parte de una señal de audio, que después se convierte de vuelta al dominio de tiempo por los convertidores de frecuencia/tiempos 220.

15 **[0055]** Los convertidores de tiempo/frecuencias 190, así como los elementos inversos, los convertidores de frecuencia/tiempos 220 por lo tanto se adaptan para generar una representación espectral de al menos una pieza de una señal de audio que se le proporciona y a re-transformar la representativa espectral en las partes correspondientes de la señal de audio en el dominio de tiempo, respectivamente.

20 **[0056]** En el proceso de convertir una señal de audio del dominio de tiempo al dominio de frecuencia, y de vuelta del dominio de frecuencia al dominio de tiempo, pueden ocurrir desviaciones de tal manera que la señal de audio re-establecida, reconstruida o decodificada puede diferir de la señal de audio original o fuente. Artefactos adicionales pueden ser agregados por las etapas adicionales de cuantificación y descuantificación realizadas en el marco del codificador cuantificador 200 y el re-codificador 210. En otras palabras, la señal de audio original, así como la señal de audio re-establecida, pueden diferir entre sí.

25 **[0057]** Los convertidores de tiempo/frecuencia 190, así como los convertidores de frecuencia/tiempo 220 pueden, por ejemplo, ser implementados con base en una transformación coseno discreta modificada (MDCT = Modified Discreet Cosine Transformation), una transformación de seno discreta modificada (MDST = Modified Discrete Sine Transformation), un convertidor basado en transformada fourier rápida (FFT = Fast Fourier Transformation), u otro convertidor basado en Fourier. La cuantificación y la re-cuantificación en el marco del cuantificador/codificador 200 y el decodificador/descuantificador 210 pueden, por ejemplo, implementarse con base en una cuantificación lineal, una cuantificación logarítmica, u otro algoritmo de cuantificación más complejo, por ejemplo, tomando en cuenta más específicamente las características de audición del humano. Las partes de codificador y decodificador del cuantificador/codificador 200 y el decodificador/descuantificador 210 pueden, por ejemplo, funcionar al emplear un esquema de codificación Huffman o decodificación Huffman.

30 **[0058]** Sin embargo, también convertidores más complejos de tiempo/frecuencia y frecuencia/tiempo 190, 220, así como cuantificador/codificador y decodificador/descuantificador más complejos 200, 210 pueden emplearse en diferentes realizaciones y sistemas como se describe aquí, siendo parte de o que forman, por ejemplo, un codificador AAC-ELD como codificadores 140, 170, y un decodificador AAC-ELD como decodificadores 120, 180.

35 **[0059]** No es necesario decir que puede ser recomendable el implementar codificadores idénticos o al menos compatibles, 170, 140 y decodificadores 180, 120, en el marco del sistema de conferencias 100 y el terminal de conferencias 160.

40 **[0060]** El sistema de conferencias 100, como se muestra en la Figura 2, basado en un esquema de codificación y decodificación de señal de audio en general también realiza el mezclado actual de las señales de audio en el dominio de tiempo. Los sumadores 130 se proporcionan con las señales de audio reconstruidas en el dominio de tiempo para realizar una súper-posición y proporcionar las señales mixtas en el dominio de tiempo a los convertidores de tiempo/frecuencia 190 de los siguientes codificadores 140. Por lo tanto, el sistema de conferencias comprende de nuevo una conexión en serie de los decodificadores 120 y codificadores 140, que es la razón por la que un sistema de conferencias 100, como se muestra en las Figuras 1 y 2, típicamente se refieren como "sistemas de codificación en tándem".

45 **[0061]** Los sistemas de codificación en tándem a menudo muestran la desventaja de una alta complejidad. La complejidad de mezclar fuertemente depende de la complejidad de los decodificadores y codificadores empleados, y puede multiplicar de forma significativa en el caso de varias señales de entrada de audio y salida de audio. Además,

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, como se emplea en los sistemas de conferencias 100, mostrados en las Figuras 1 y 2, típicamente llevan a una influencia negativa en la calidad.

5 **[0062]** Como una desventaja adicional, las etapas repetidas de decodificación y codificación también amplían el retraso total entre las entradas 110 y las salidas 150 del sistema de conferencias 100, que también se refiere como el retraso de extremo-a-extremo. Dependiendo de un retraso inicial de los decodificadores y codificadores empleados, el propio sistema de conferencias 100, puede incrementar el retraso hasta un nivel que hace no atractivo el uso en el marco del sistema de conferencias, si no es perjudicial, o incluso imposible. A menudo un retraso de 10 aproximadamente 50 ms se considera como el retraso máximo que pueden aceptar los participantes en las conversaciones.

[0063] Como fuentes principales para el retraso, los convertidores de tiempo/frecuencia 190, así como los convertidores de frecuencia/tiempo 220 son responsables por el retraso de extremo-a-extremo del sistema de 15 conferencias 100, y el retraso adicional impuesto por los terminales de conferencias 160. El retraso provocado por los elementos adicionales, es decir los cuantificadores/codificadores 200 y los decodificadores/descuantificadores 210 es de menor importancia ya que estos componentes pueden operarse 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 se operan por 20 bloques u operan por cuadros, lo que significa que en muchos casos un retraso mínimo como una cantidad de tiempo, debe tomarse en cuenta, lo que es igual al tiempo requerido para llenar un amortiguador o una memoria que tiene la duración de cuadro de un bloque. Este tiempo, sin embargo, se influencia significativamente por la frecuencia de muestreo que típicamente está en el intervalo de unos cuantos kHz a unos cuantos 10 kHz, mientras que la velocidad operacional de los cuantificadores/codificadores 200, así como el 25 decodificador/descuantificador 210 primordialmente se determina por la frecuencia de reloj del sistema subyacente. Esto típicamente es al menos 2, 3, 4 o más ordenes de magnitud mayores.

[0064] Por lo tanto, en sistemas de conferencias que emplean codecs de señal de audio generales, la así llamada tecnología de mezclado de corriente de bits se ha introducido. El procedimiento de mezclado de corriente de 30 bits puede, por ejemplo, ser implementado con base en el codec MPEG-4 AAC-ELD, que ofrece la posibilidad de evitar al menos algunas de las desventajas mencionadas anteriormente e introducidas por codificación en tándem.

[0065] Sin embargo, cabe destacar que, en principio, el sistema de conferencias 100 como se muestra en la Figura 2, también puede implementarse con base en el codec MPEG-4 AAC-ELD son de una velocidad de bits 35 similar y a un ancho de banda de frecuencia significativamente mayor, en comparación con los codecs basados en habla previamente mencionados de la familia de codecs G.7xx. Esto implica también inmediatamente que una calidad de audio significativamente mejor para todos los tipos de señal puede lograrse al coste de una velocidad de bits significativamente incrementada. Aunque MPEG-4 AAC-ELD ofrece un retraso que está en el intervalo de aquel del codec G.7xx, implementado el mismo en el marco de un sistema de conferencias como se muestra en la Figura 40 2, puede no llevar a un sistema de conferencias práctico 100. A continuación, con respecto a la Figura 3, será resaltado un sistema más práctico con base en el así denominado mezclado de corriente de bits previamente mencionado.

[0066] Cabe destacar que por razones de simplicidad solamente, el foco se aplicará primordialmente en el 45 codec MPEG-4 AAC-ELD y sus corrientes de datos y corrientes de bits. Sin embargo, también otros codificadores y decodificadores pueden emplearse en el entorno de un sistema de conferencias 100 como se ilustra y muestra en la Figura 3.

[0067] La Figura 3 muestra un diagrama de bloques de un sistema de conferencias 100 que trabaja según el 50 mezclado de corriente de bits junto con un terminal de conferencias 160, como se describe en el contexto de la Figura 2. El sistema de conferencias 100 mismo es una versión simplificada del sistema de conferencias 100 mostrado en la Figura 2. Para ser más precisos, los decodificadores 120 del sistema de conferencias 100 en la Figura 2 se han reemplazado por decodificadores/descuantificadores 220-1, 220-2, 210-3, ... como se muestra en la Figura 3. En otras palabras, los convertidores de frecuencia/tiempo 120 de los decodificadores 120 se han retirado 55 cuando se compara el sistema de conferencias 100 ilustrado en las Figuras 2 y 3. De manera similar, los codificadores 140 del sistema de conferencias 100 de la Figura 2 se han reemplazado 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 conferencias 100 mostrado en las Figuras 2 y 3.

[0068] Como resultado, los sumadores 130 no operan más en el dominio de tiempo, pero 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 a frecuencia.

5 **[0069]** Por ejemplo, en el caso de los codecs MPEG-4 AAC-ELD, el convertidor de tiempo/frecuencia 190 y el convertidor de frecuencia/tiempo 220, que solo están presentes en los terminales de conferencias 160, se basan en una transformación MDCT. Por lo tanto, dentro del sistema de conferencias 100, los mezcladores 130 operan directamente en las contribuciones de las señales de audio en la representación de frecuencia MDCT.

10 **[0070]** Ya que los convertidores 190, 220 representan la fuente principal de retraso en el caso del sistema de conferencias 100 mostrado en la Figura 2, el retraso se reduce significativamente al retirar estos convertidores 190, 220. Además, la complejidad introducida por los dos convertidores 190, 220 dentro del sistema de conferencias 100 también se reduce significativamente. Por ejemplo, en el caso de un decodificador MPEG-2 AAC, la transformación MDCT inversa que se lleva a cabo en el marco del convertidor de frecuencia/tiempo 220 es responsable de
15 aproximadamente el 20% de la complejidad total. Ya que también el convertidor MPEG-4 se basa en una transformación similar, una contribución no irrelevante a la complejidad total puede retirarse al eliminar el convertidor de frecuencia/tiempo 220 solo del sistema de conferencias 100.

[0071] Es posible mezclar las señales de audio en el dominio MDCT, u otro dominio de 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,

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

25 y la de homogeneidad matemática, es decir

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

30 en la 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.

[0072] Ambas características de la transformación MDCT u otra transformación basada en Fourier permiten el mezclado en el dominio de frecuencia respectivo similar al mezclado en el dominio de tiempo. Por lo tanto, todos los cálculos pueden ser igualmente llevados a cabo con base en valores espectrales. Una transformación de los
35 datos en el dominio de tiempo no se requiere.

[0073] Bajo algunas circunstancias, se puede requerir que se cumpla una condición adicional. Todos los datos espectrales relevantes deberán ser iguales respecto a sus índices de tiempo durante el proceso de mezclado para todos los componentes espectrales relevantes. Esto puede eventualmente no ser el caso si, durante la
40 transformación la así denominada técnica de conmutación de bloques, se emplea de tal manera que el codificador de los terminales de conferencias 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 de tiempo debido a la conmutación entre
45 diferentes longitudes de bloques y longitudes de ventanas MDCT correspondientes, a menos que los datos al mezclarse hayan sido procesados con las mismas ventanas. Ya que en un sistema general con terminales de conferencias distribuidas 160, esto puede no ser garantizado eventualmente, interpolaciones complejas pueden ser necesarias que a su vez pueden crear retraso y complejidad adicionales. Como una consecuencia, puede eventualmente ser recomendable el no implementar un proceso de mezclado de corriente de bits con base en
50 conmutación de longitudes o tramos de bloques.

[0074] En contraste, el codec AAC-ELD se basa en una longitud de un solo bloque y, por lo tanto, es capaz de garantizar más fácilmente la asignación previamente descrita o sincronización de datos de frecuencia, de tal manera que pueda lograrse más fácilmente un mezclado. El sistema de conferencia es 100 mostrado en la Figura 3, en otras palabras, es un sistema que es capaz de realizar el mezclado en el dominio de transformación o el dominio
55 de frecuencia.

- [0075]** Como se ha establecido anteriormente, a fin de eliminar el retraso adicional introducido por los convertidores 190, 200 en el sistema de conferencia 100 mostrado en la Figura 2, los codecs empleados en los terminales de conferencia 160 utilizan una ventana de longitud y forma fijas. Esto permite la implementación del proceso de mezclado descrito directamente sin transformar la corriente de audio de vuelta al dominio de tiempo. Esta estrategia es capaz de limitar la cantidad de retraso algorítmico adicionalmente introducido. Además, la complejidad se disminuye debido a la ausencia de las etapas de transformación inversa en el decodificador y las etapas de transformación directa en el codificador.
- 10 **[0076]** Sin embargo, también en el marco de un sistema de conferencias 100 como se muestra en la Figura 3, puede ser necesario el volver a cuantificar o re-cuantificar 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 crearse debido a las diferentes etapas de cuantificación o diferentes señales de audio que se proporcionan al sistema de conferencia 100. Como un resultado, por ejemplo, en el caso de muy bajas transmisiones de velocidad de bits en donde una cantidad de etapas de cuantificación ya están limitadas, el proceso de mezclar dos señales de audio en el dominio de frecuencia o el dominio de transformación, puede resultar en una cantidad adicional indeseable de ruido u otras distorsiones en la señal generada.
- 15 **[0077]** Antes de describir una primera realización según la presente invención en la forma de un aparato para mezclar una pluralidad de corrientes de datos de entrada, con respecto a la Figura 4, se describirá brevemente, una corriente de datos o corriente de bits, junto con los datos ahí comprendidos.
- [0078]** La Figura 4 muestra esquemáticamente una corriente de bits o corriente de datos 250, que comprende al menos uno, o más a menudo, más de un cuadro 260 de datos de audio en un dominio espectral. De forma más precisa, la Figura 4 muestra tres cuadros 260-1, 260-2, y 260-3 de datos de audio en un dominio espectral. Además, la corriente 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 que se codifican los datos de audio, otros valores de control de información referente a índices de tiempo u otros datos relevantes. Naturalmente, la corriente de datos 250 como se muestra en la Figura 4 además puede comprender cuadros adicionales o un cuadro 260 puede comprender datos de audio de más de un canal. Por ejemplo, en el caso de una señal de audio en estéreo, cada uno de los cuadros 260 puede, por ejemplo, comprender 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 previamente mencionados.
- 25 **[0079]** Por lo tanto, la Figura 4 ilustra que una corriente de datos 250 puede no solo comprender un cuadro de datos de 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 semejantes.
- 30 **[0080]** Dependiendo de la implementación concreta del sistema de conferencias como se describe en el contexto de las Figuras 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 aquellos descritos con respecto a las Figuras 9 a 12C, los valores de control indican una forma asociada a datos de carga útil del cuadro representa al menos una parte del dominio espectral o información espectral de una señal de audio puede igualmente estar comprendida en los propios cuadros 260, o en el bloque asociado 270 de información adicional. En el caso de que valores de control se relacionen a componentes espectrales, los valores de control pueden ser codificados en los cuadros 260 mismos. Sin embargo, si un valor de control se refiere a un cuadro entero, puede igualmente estar comprendido en los bloques 270 de información adicional. Sin embargo, los sitios previamente mencionados para incluir los valores de control son como se ha descrito anteriormente, en mucho no se requiere que estén comprendidos en los cuadros 260 o el bloque 270 de los bloques adicionales. En el caso de que un valor de control se refiere solo a un solo o unos cuantos componentes espectrales, igualmente puede estar bien comprendido en el bloque 270. Por otra parte, un valor de control referente a un cuadro entero 260 también puede estar comprendido en los cuadros 260.
- 35 **[0081]** La Figura 5 ilustra esquemáticamente información (espectral) que se refiere a componentes espectrales, tal como por ejemplo comprendidos en el cuadro 260 de la corriente de datos 250. Para ser más precisos, la Figura 5 muestra un diagrama de información simplificado en un dominio espectral de un solo canal de un cuadro 260. En el dominio espectral, un cuadro de datos de audio puede, por ejemplo, describirse en términos de sus valores de intensidad I como una función de la frecuencia f . En sistemas discretos, tales como por ejemplo
- 40
- 45
- 50
- 55

sistemas digitales, también la resolución de frecuencia es discreta, de tal manera que la información espectral típicamente solo está presente para ciertos componentes espectrales tales como frecuencias individuales o bandas o sub-bandas estrechas. Frecuencias individuales o bandas estrechas, así como sub-bandas se refieren como componentes espectrales.

5

[0082] La Figura 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 Figura 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 respecto a los cuales el cuadro comprende información referente a los datos de audio en el dominio espectral.

10

[0083] 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 a energía tales como la amplitud, la energía del propio componente espectral respectiva u otro valor derivado de la energía o la amplitud, información de fase u otra información también pueden estar comprendidos en el cuadro y por lo tanto considerados como información referente a un componente espectral.

15

[0084] Después de haber descrito algunos de los problemas involucrados y algún antecedente para sistemas de conferencias, las realizaciones según un primer aspecto de la presente invención se describen de acuerdo a lo cual una corriente de datos de entrada se determina con base en una comparación a fin de copiar al menos información espectral parcial de la corriente de datos de entrada determinada a la corriente de datos de salida, permitiendo de esta manera omitir una re-cuantificación y por lo tanto, el ruido de re-cuantificación asociado.

20

[0085] La Figura 6 muestra un diagrama de bloques de un aparato 500 para mezclar una pluralidad de corrientes de datos de entrada 510, de las cuales dos se ilustran 510-1, 510-2. El aparato 500 comprende una unidad de procesamiento 520 que se adapta para recibir las corrientes de datos 510 y para generar una corriente de datos de salida 530. Cada una de las corrientes de datos de entrada 510-1, 510-2 comprende un cuadro 540-1, 540-2, respectivamente, que es similar al cuadro 260 mostrado en la Figura 4 en el contexto de la Figura 5, comprende datos de audio en un dominio espectral. Esto se ilustra de nuevo por un sistema de coordenadas mostrado en la Figura 6 en la abscisa, del cual se muestran la frecuencia f y la ordenada del cual la intensidad I . La corriente de datos de salida 530 también comprende un cuadro de salida 550 que comprende datos de audio en un dominio espectral, y también ilustrado por un sistema de coordenadas correspondiente.

30

[0086] La unidad de procesamiento 520 se adapta para comparar los cuadros 540-1, 540-2 de una pluralidad de corrientes de datos de entrada 510. Como se establecerá con más detalle a continuación, esta comparación por ejemplo puede basarse en un modelo psico-acústico, tomando en consideración efectos de enmascarado y otras propiedades de las características de audición humana. Con base en este resultado de comparación, la unidad de procesamiento 520 además se adapta para determinar al menos un componente espectral, por ejemplo, los componentes espectrales 560 mostrados en la Figura 6, que está presente en ambos cuadros 540-1, 540-2, exactamente una corriente de datos en la pluralidad de corrientes de datos 510. Después, la unidad de procesamiento 520 puede ser adaptada para generar la corriente de datos de salida 530, que comprende el cuadro de salida 550, de tal manera que una información referente al componente espectral 560 se copia del cuadro determinado 540 de la corriente de datos de entrada respectiva 510.

35

[0087] Para ser más precisos, la unidad de procesamiento 520 se adapta de tal manera que comparar el cuadro 540 de la pluralidad de corrientes de datos de entrada 510 se basa en al menos dos piezas de información – los valores de intensidad son valores de energía relacionados – correspondientes al mismo componente espectral 560 de los cuadros 540 de dos corrientes de datos de entrada diferentes 510.

45

[0088] Para ilustrar adicionalmente esto, la Figura 7 muestra esquemáticamente el caso en el que la pieza de información (la intensidad I), correspondiente a los componentes espectrales 560, que se considera aquí como una frecuencia o una banda de frecuencia estrecha del cuadro 540-1 de una primera corriente de datos de entrada 510-1. Esto se compara con valor de intensidad correspondiente I , que es la pieza de información referente al componente espectral 560 del cuadro 540-2 de la segunda corriente de datos de entrada 510-2. La comparación, por ejemplo, puede realizarse con base en la evaluación de una proporción de energía entre la señal mixta en la que solo algunas corrientes de entrada se incluyen y una señal mixta completa. Esto por ejemplo puede lograrse según

50

55

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

y

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

5

y calcular la proporción $r(n)$ según

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

10

en el que n es un índice de una corriente de datos de entrada y N es el número de todas las corrientes de datos de entrada relevantes. Si la proporción $r(n)$ es lo suficientemente alta, los canales menos dominantes o los cuadros menos dominantes de las corrientes 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 solo aquellos componentes espectrales totalmente notables de una corriente se incluyen, mientras que las otras corrientes se descartan.

[0089] Los valores de energía que se van a considerar en el marco de las ecuaciones (3) a (5), por ejemplo, pueden derivarse de los valores de intensidad como se muestra en la Figura 6, al calcular el cuadrado de los valores de intensidad respectivos. En el caso de que la información referente a los componentes espectrales puede comprender otros valores, puede llevarse a cabo un cálculo similar dependiendo de la forma de la información comprendida en el cuadro 510. Por ejemplo, en el caso de la información de valor complejo, podrá tener que realizarse el calcular el módulo de los componentes reales e imaginarios de los valores individuales que constituyen la información referente a los componentes espectrales.

25

[0090] Aparte de las frecuencias individuales, para la aplicación del módulo psico-acústico según las ecuaciones (3) a (5), la suma en las ecuaciones (3) y (4) puede comprender más de una frecuencia. En otras palabras, en las ecuaciones (3) y (4) los valores de energía respectivos E_n pueden ser reemplazados por un valor de energía total que corresponde a una pluralidad de frecuencias individuales, una energía de una banda de frecuencia, o para ponerlo en términos más generales, por una sola pieza de información espectral o una pluralidad de información espectral referente a uno o más componentes espectrales.

30

[0091] Por ejemplo, ya que AAC-ELD opera en líneas espectrales de una forma a manera de banda, similar a los grupos de frecuencia en los que el sistema auditivo humano trata al mismo tiempo, la estimación de irrelevancia o el modelo psico-acústico puede llevarse a cabo de manera similar. Al aplicar el modelo psico-acústico de esta manera, es posible retirar o sustituir parte de una señal de solo una banda de frecuencia, de ser necesario.

35

[0092] Como han mostrado los exámenes psico-acústicos, el enmascarar 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 por una sinusoide u otro sonido distinto y bien definido, se requiere una diferencia de 21 a 28 dB típicamente. Pruebas han mostrado que un valor umbral de aproximadamente 28,5 dB produce buenos resultados sustitutos. Este valor puede mejorarse eventualmente, también tomando en cuenta bajo consideración las bandas de frecuencia actuales.

40

[0093] Por lo tanto, valores $r(n)$ según la ecuación (5) que son mayores que -28,5 dB, pueden considerarse como irrelevantes en términos de una evaluación psico-acústica o evaluación de irrelevancia con base en el componente espectral o los componentes espectrales bajo consideración. Para diferentes componentes espectrales, pueden emplearse diferentes valores. De esta manera, utilizando umbrales como indicadores para una irrelevancia

45

psico-acústica de una corriente de datos de entrada en términos del cuadro bajo consideración de 10 dB a 40 dB, 20 dB a 30 dB, o 25 dB a 30 dB pueden ser considerados útiles.

[0094] En la situación mostrada en la Figura 7, esto significa que con respecto al componente espectral 560, la primera corriente de datos de entrada 510-1 se determina, mientras que la segunda corriente de datos de entrada 510-2 se descarta con respecto al componente espectral 560. Como resultado, la pieza de información referente al componente espectral 560 es al menos copiada parcialmente del cuadro 540-1 de la primera corriente de datos de entrada 510-1 al cuadro de salida 550 de la corriente de datos de salida 530. Esto se ilustra en la Figura 7 por una flecha 570. Al mismo tiempo, las piezas de información referentes a los componentes espectrales 560 del cuadro 540 de las otras corrientes de datos de entrada 510 (es decir, en la Figura 7, cuadro 540-2 de la corriente de datos de entrada 510-2) se descarta como se ilustra por la línea punteada 580.

[0095] Todavía en otras palabras, el aparato 500, que por ejemplo puede emplearse como una MCU o un sistema de conferencias 100, se adapta de tal manera que la corriente de datos de salida 530 junto con su cuadro de salida 550 se genera, de tal manera que la información del componente espectral correspondiente se copia solo del cuadro 540-1 de la corriente de datos de entrada determinada 510-1 que describe el componente espectral 560 de la corriente de salida 550 de la corriente de datos de salida 530. Naturalmente, el aparato 500 también puede adaptarse de tal manera que la información referente a más de un componente espectral puede ser copiada desde una corriente de datos de entrada, descartando las otras corrientes de datos de entrada, al menos con respecto a estos componentes espectrales. Además, es posible que un aparato 500, o su unidad de procesamiento 520, se adapte de tal forma que, para diferentes componentes espectrales, se determinan diferentes corrientes de datos de entrada 510. El mismo cuadro de salida 550 de la corriente de datos de salida 530 puede comprender información espectral copiada referente a diferentes componentes espectrales de diferentes corrientes de datos de entrada 510.

[0096] Naturalmente, puede ser recomendable el implementar el aparato 500 de tal manera que en el caso de una secuencia de cuadros 540 en una corriente de datos de entrada 510, solo los cuadros 540 se considerarán durante la comparación y determinación, que corresponde a un índice de tiempo similar o igual.

[0097] En otras palabras, la Figura 7 ilustra los principios operacionales de un aparato para mezclar una pluralidad de corrientes de datos de entrada como se describió anteriormente según una realización. Como se ha establecido anteriormente, el mezclado no se realiza de una forma directa en el sentido de que todas las corrientes de entrada se decodifican, lo que incluye una transformación inversa al dominio de tiempo, mezclado y de nuevo re-codificación de las señales.

[0098] Las realizaciones de las Figuras 6 a 8 se basan en el mezclado realizado en el dominio de frecuencia del codec respectivo. Un codec posible puede ser el codec AAC-ELD, o cualquier otro codec con una ventana de transformación uniforme. En tal caso, no se requiere transformación de tiempo/frecuencia para poder mezclar los datos respectivos. Realizaciones según una realización de la presente invención hacen uso del hecho de que el acceso a todos los parámetros de la corriente de bits, tales como el tamaño de la etapa de cuantificación y otros parámetros, es posible y que estos parámetros pueden emplearse para generar una corriente de bits de salida mixta.

[0099] Las realizaciones de las Figuras 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 por 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 cero significa que se tratan fuentes como irrelevantes y no se utilizarán de hecho. Grupos de líneas, tales como bandas o bandas de factor de escala pueden emplear el mismo factor de ponderación. Sin embargo, como se ha ilustrado anteriormente, los factores de ponderación (por ejemplo, una distribución de ceros y unos) puede variarse para los componentes espectrales de un solo cuadro 540 de una sola corriente de datos de entrada 510. Además, no es necesario utilizar exclusivamente los factores de ponderación cero o uno cuando se mezcla información espectral. Puede ser el caso de que, bajo ciertas circunstancias, no para uno solo, una pluralidad de información espectral total de un cuadro 540 de una corriente de datos de entrada 510, los factores de ponderación respectivos pueden ser diferentes de cero o uno.

[0100] Un caso particular es que todas las bandas de componentes espectrales de una fuente (corriente 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, la corriente de bits de entrada completa de un participante se copia de forma idéntica como una corriente de bits mezclada final. Los factores de ponderación pueden calcularse en una base cuadro-a-cuadro, pero también pueden calcularse o determinarse con base en grupos más largos o secuencias de cuadros. De forma natural,

incluso dentro de esta secuencia de cuadros o dentro de cuadros sencillos, los factores de ponderación pueden diferir para diferentes componentes espectrales, como se establece anteriormente. Los factores de ponderación pueden calcularse o determinarse según resultados del modelo psico-acústico.

5 **[0101]** Un ejemplo de un modelo psico-acústico ya se ha descrito anteriormente en el contexto con las ecuaciones (3), (4), y (5). El modelo psico-acústico o un módulo respectivo calcula la proporción de energía $r(n)$ entre una señal mixta en la que solo algunas corrientes de entrada se incluyen llevando a un valor de energía E_f y la señal mixta completa que tiene un valor de energía E_c . La proporción de energía $r(n)$ después se calcula según la ecuación (5) como 20 veces el logarítmico de E_f dividido por E_c .

10

[0102] Si la proporción es suficientemente alta, los canales menos dominantes pueden considerarse como enmascarados por los dominantes. De esta manera, una reducción de irrelevancia se procesa, lo que significa que solo aquellas corrientes se incluyen que no son por completo notables, a lo cual se atribuye un factor de ponderación, mientras que en todas las otras corrientes – al menos una información espectral de un componente
15 espectral – se descarta. En otras palabras, a estos se atribuye un factor de ponderación de cero.

[0103] Puede introducirse la ventaja de que menos o ningún efecto de codificación en tándem ocurre debido a un número reducido de etapas de re-cuantificación. Ya que cada etapa de cuantificación involucra un riesgo significativo de reducir ruido de cuantificación adicional, la calidad total de la señal de audio puede mejorarse al
20 emplear cualquiera de las realizaciones anteriormente mencionadas para mezclar una pluralidad de corrientes de datos de entrada. Esto puede ser el caso cuando la unidad de procesamiento 520 del aparato 500, tal como por ejemplo como se muestra en la Figura 6, se adapta de tal manera que la corriente de datos de salida 530 se genera de tal modo que una distribución de niveles de cuantificación comparado con una distribución de niveles de cuantificación del cuadro de la corriente de entrada determinada o sus partes, se mantiene. En otras palabras, al
25 copiar y, por lo tanto, al reutilizar los datos respectivos sin re-codificar la información espectral, puede omitirse una introducción de ruido de cuantificación adicional.

[0104] Además, el sistema de conferencias, por ejemplo, un sistema de tele/vídeo conferencia con más de un participante que emplea cualquier realización anteriormente descrita respecto a las Figuras 6 a 8, puede ofrecer la
30 ventaja de una menor complejidad en comparación con un mezclado en dominio de tiempo, ya que las etapas de transformación de tiempo-frecuencia y las etapas de re-codificación pueden ser omitidas. Además, no se provoca mayor retraso por estos componentes en comparación con mezclado en dominio de tiempo, debido a la ausencia del retraso de banco de filtros.

35 **[0105]** Para resumir, las realizaciones anteriormente descritas, por ejemplo, pueden adaptarse de tal manera que bandas de información espectral que corresponden a los componentes espectrales, que se toman completamente de una fuente, no se re-cuantifiquen. Por lo tanto, solo bandas o información espectral que se mezclan son re-cuantificadas, lo que reduce el ruido de cuantificación adicional.

40 **[0106]** Sin embargo, las realizaciones anteriormente descritas 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, parámetros TNS, parámetros SBR o parámetros de codificación estéreo, una realización se
45 describirá con más detalle con referencia a la Figura 8.

[0107] La Figura 8 muestra un diagrama de bloques esquemático de un aparato 500 para mezclar una pluralidad de corrientes de datos de entrada que comprenden una unidad de procesamiento 520. Para ser más preciso, la Figura 8 muestra un aparato altamente flexible 500 que es capaz de procesar señales de audio altamente
50 diferentes codificadas en corrientes de datos de entrada (corrientes de bits). Algunos de los componentes que se describirán a continuación, por lo tanto, son componentes opcionales que no se requieren implementados bajo todas las circunstancias.

[0108] La unidad de procesamiento 520 comprende un decodificador de corriente de bits 700 para cada una de las corrientes de datos de entrada o corrientes de bits de audio codificadas para procesarse por la unidad de procesamiento 520. Por razones de simplicidad solamente, la Figura 8 muestra solo dos decodificadores de corriente de bits 700-1, 700-2. De forma natural, dependiendo del número de corrientes de datos de entrada que se va a procesar, un número superior de decodificadores de corriente de bits 700, o un número inferior puede implementarse, si por ejemplo un decodificador de corriente de bits 700 es capaz de procesar secuencialmente más

de una de las corrientes de datos de entrada.

[0109] El decodificador de corriente de bits 700-1, así como los otros decodificadores de corrientes de bits 700-2, ... comprende cada uno un lector de corriente de bits 710 que se adapta para recibir y procesar las señales recibidas, y para aislar y extraer datos comprendidos en la corriente de bits. Por ejemplo, el lector de la corriente de bits 710 puede ser adaptado para sincronizar los datos de entrada con un reloj interno y además puede adaptarse para separar la corriente de bits de entrada en los cuadros apropiados.

[0110] El decodificador de corriente de bits 700 comprende además un decodificador Huffman 720 acoplado a la salida del lector de corriente de bits 710 para recibir los datos aislados del lector de corriente de bits 710. Una salida del decodificador Huffman 720 se acopla a un descuantificador 730, que también se refiere como un cuantificador inverso. El descuantificador 730 acoplado tras el decodificador Huffman 720 es seguido por un dimensionador o factor multiplicador 740. El decodificador Huffman 720, el descuantificador 730 y el dimensionador 740 forman una primera unidad 750 a la salida del cual al menos una parte de la señal de audio de la corriente de datos de entrada respectiva está disponible en el dominio de frecuencia o el dominio relacionado a frecuencia en el que opera el codificador del participante (no mostrado en la Figura 8).

[0111] El decodificador de corriente de bits 700 comprende además una segunda unidad 760 que es acoplada a manera de datos después de la primera unidad 750. La segunda unidad 760 comprende un decodificador de estéreo 770 (módulo M/S) tras el cual se acopla un decodificador PNS. El decodificador PNS 780 es seguido a manera de datos por un decodificador TNS 790, que junto con el decodificador PNS 780 en el decodificador de estéreo 770 forma la segunda unidad 760.

[0112] Aparte del flujo descrito de datos de audio, el decodificador de corriente de bits 700 comprende además una pluralidad de conexiones entre diferentes módulos referentes a datos de control. Para ser más preciso, el lector de corriente de bits 710 también se acopla al decodificador Huffman 720 para recibir datos de control apropiados. Además, 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 acopla también cada uno al lector de corriente de bits 710 para recibir datos de control apropiados.

[0113] La unidad de procesamiento 520 comprende además una unidad de mezclado 800 que a su vez comprende un mezclador espectral 810 que es acoplado a modo de entrada con los decodificadores de corriente de bits 700. El mezclador espectral 810, por ejemplo, puede comprender uno o más sumadores para realizar el mezclado actual en el dominio de frecuencia. Además, el mezclador espectral 810 puede comprender además multiplicadores para permitir una combinación lineal arbitraria de la información espectral que se proporciona por los decodificadores de corriente de bits 700.

[0114] La unidad de mezclado 800 comprende además un módulo de optimización 820 que es acoplado a manera de datos con una salida del mezclador espectral 810. El módulo de optimización 820, sin embargo, también se acopla al mezclador espectral 810 para proporcionar el mezclador espectral 810 con información de control. A modo de datos, el módulo de optimización 820 representa una salida de la unidad de mezclado 800.

[0115] La unidad de mezclado 800 comprende además un mezclador SBR 830 que se acopla directamente a una salida de lector de corriente de bits 710 de los diferentes decodificadores de corriente de bits 700. Una salida del mezclador SBR 830 forma otra salida de la unidad de mezclado 800.

[0116] La unidad de procesamiento 520 comprende además un codificador de corriente de bits 850 que se acopla a la unidad de mezclado 800. El codificador de corriente de bits 850 comprende una tercera unidad 860 que comprende un codificador TNS 870, codificador PNS 880, y un codificador de estéreo 890, que se acoplan en serie en el orden descrito. La tercera unidad 860, por lo tanto, forma una unidad inversa de la primera unidad 750 del decodificador de corriente de bits 700.

[0117] El codificador de corriente 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. Según esto, el dimensionador 910 también se acopla directamente al codificador Huffman 930 para proporcionar el codificador Huffman 930 con los datos de control respectivos.

[0118] El codificador de la corriente de bits 850 comprende también un escritor de corriente de bits 940, que se acopla a la salida del codificador Huffman 930. Además, el escritor de corriente de bits 940 se acopla también al codificador TNS 870, el codificador PNS 880, el codificador estéreo 890, y el codificador Huffman 930 para recibir datos de control e información de estos módulos. Una salida del escritor de corriente de bits 940 forma una salida de la unidad de procesamiento 520 y del aparato 500.

[0119] El codificador de corriente de bits 850 comprende también un módulo psico-acústico 950, que también se acopla a la salida de la unidad de mezclado 800. El codificador de corriente 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.

[0120] En principio, a las salidas de la segunda unidad 760 hasta la entrada de la tercera unidad 860, un procesamiento de la señal de audio en el dominio espectral, como se define por el codificador empleado en el lado de remitente, es por lo tanto posible. Sin embargo, como se ha indicado anteriormente, una completa decodificación, descuantificación, des-dimensionado o disminución en dimensiones y adicionales etapas de procesamiento pueden eventualmente no ser necesarias y, por ejemplo, es dominante información espectral de un cuadro de una de las corrientes de datos de entrada. Al menos una parte de la información espectral de los componentes espectrales respectivos, se copia entonces al componente espectral del cuadro respectivo de la corriente de datos de salida.

[0121] Para permitir este procesamiento, el aparato 500 y la unidad de procesamiento 520 comprenden líneas de señal adicionales para un intercambio de datos optimizados. Para permitir este procesamiento en la realización mostrada en la Figura 8, una salida del decodificador Huffman 720, así como salidas del dimensionador 740, el decodificador estéreo 770, y el decodificador PNS 780 son, junto con los componentes respectivos de los otros lectores de corriente de bits 710, acoplados al módulo de optimización 820 de la unidad de mezclado 800 para un procesamiento respectivo.

[0122] Para facilitar, después de un procesamiento respectivo, un flujo de datos correspondientes dentro del codificador de corriente de bits 850, correspondientes líneas de datos para un flujo de datos optimizado también se implementan. 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. Además, la salida del módulo de optimización 820 también se acopla directamente al escritor de corriente de bits 940.

[0123] Como se ha indicado anteriormente, casi todos los módulos que se han descrito anteriormente son módulos opcionales, que no se requiere que se implementen. Por ejemplo, en el caso de las corrientes de datos de audio que comprenden solo un canal sencillo, las unidades de codificación y decodificación en estéreo 770, 890, pueden ser omitidas. Según esto, en el caso en que no se van a procesar señales basadas en PNS, el decodificador PNS y codificador PNS correspondientes 780, 880 también pueden ser omitidos. Los módulos TNS 790, 870 pueden ser omitidos también en el caso de la señal que se va a procesar y la señal enviada de salida si no está basada en los datos TNS. Dentro de la primera y cuarta unidades 750, 900 el cuantificador inverso 730, el dimensionador 740, el cuantificador 920, así como el dimensionador 910 pueden ser omitidos también eventualmente. El decodificador Huffman 720 y el codificador Huffman 930 pueden ser implementados de forma diferente, utilizando otro algoritmo u omitidos por completo.

[0124] El mezclador SBR 830 puede ser omitido también eventualmente 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 psico-acústico 860. Por lo tanto, también estos módulos habrán de considerarse como componentes opcionales.

[0125] Con respecto al modo de operación del aparato 500 junto con la unidad de procesamiento 520 ahí comprendida, primero se lee una corriente de datos de entrada de introducción y se separa en piezas de información apropiadas por el lector de corriente de bits 710. Después de decodificación Huffman, la información espectral resultante puede re-cuantificarse eventualmente por el descuantificador 730 y ajustarse en escala o dimensionarse de forma apropiada por el dimensionador en sentido inverso 740.

[0126] Posteriormente, dependiendo de la información de control comprendida en la corriente de datos de entrada, la señal de audio codificada por la corriente 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 medio (M) y un canal lateral (S), los datos de canal izquierdo y canal derecho correspondientes pueden obtenerse al agregar y sustraer los datos de canal medio y lateral entre sí. En muchas implementaciones, el canal medio es proporcional a la suma de datos de audio de canal izquierdo y 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 anteriormente referidos pueden agregarse y/o sustraerse tomando un factor 1/2 en cuenta para evitar efectos de recorte. Generalmente hablando, los diferentes canales pueden procesarse por combinaciones lineales para dar como resultado los canales correspondientes.

10 **[0127]** En otras palabras, después del decodificador de estéreo 770, los datos de audio pueden, si es apropiado, descomponerse en dos canales individuales. De forma natural, también una decodificación inversa puede realizarse por el decodificador de estéreo 770. Si, por ejemplo, la señal de audio como se recibe por el lector de corriente de bits 710 comprende un canal izquierdo y un canal derecho, el decodificador de estéreo 770 puede igualmente bien calcular o determinar datos de canal medio y lateral apropiados.

15 **[0128]** Dependiendo de la implementación no sólo del aparato 500, sino también dependiendo de la implementación del codificador del participante que proporciona la corriente de datos de entrada respectiva, la corriente de datos respectiva 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 tipo ruido en un intervalo de frecuencia limitado o componente espectral tal como una banda o
20 una frecuencia individual, de un ruido generado de forma sintética. PNS por lo tanto sustituye la contribución tipo ruido actual de la señal de audio con un valor de energía indicando un nivel de ruido para introducirse de forma sintética en el componente espectral respectivo y despreciando la señal de audio actual. En otras palabras, el decodificador PNS 780 puede regenerar en uno o más componentes espectrales la contribución de señal de audio tipo ruido actual con base en un parámetro PNS comprendido en la corriente de datos de entrada.

25 **[0129]** En términos del decodificador TNS 790 y el codificador TNS 870, 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 tipo transitoria
30 en un cuadro de la señal de audio. Para compensar esta transitoria, al menos un filtro de predicción adaptativa se aplica 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 al igual que los intervalos de frecuencia a los cuales se aplican los filtros respectivos.

35 **[0130]** En otras palabras, la operación de un módulo TNS se basa en el cómputo de uno o más filtros adaptativos filtros de respuesta de impulso infinita (IIR = Infinite Impulse Response), y al codificar y transmitir una señal de error que describe la diferencia entre la señal de audio pronosticada y actual junto con los coeficientes filtro de los filtros de predicción. Como una consecuencia, puede ser posible el incrementar la calidad de audio mientras que se mantiene la velocidad de bits de la corriente de datos del transmisor al enfrentar las señales tipo transitorias
40 aplicando un filtro de predicción en el dominio de frecuencia para reducir la amplitud de la señal de error restante, que puede entonces codificarse utilizando menos etapas de cuantificación en comparación con codificar directamente la señal de audio tipo transitoria con un ruido de cuantificación similar.

45 **[0131]** En términos de una aplicación TNS, puede ser recomendable bajo ciertas circunstancias el emplear la función del decodificador TNS 760 para decodificar las partes TNS de la corriente de datos de entrada para llegar a una representación "pura" en el dominio espectral determinada por el codec empleado. Esta aplicación de la funcionalidad de los decodificadores TNS 790 puede ser útil si un estimado del modelo psico-acústico (por ejemplo, aplicado en el módulo psico-acústico 950) no puede haber sido estimado ya con base en los coeficientes filtro de los filtros de predicción comprendidos en los parámetros TNS. Esto puede ser especialmente importante en el caso
50 cuando al menos una corriente de datos de entrada utiliza TNS, mientras que otra no.

[0132] Cuando la unidad de procesamiento determina, con base en la comparación de los cuadros de las corrientes de datos de entrada que la información espectral de un cuadro de una corriente de datos de entrada que utiliza TNS se va a utilizar, los parámetros TNS pueden emplearse para el cuadro de datos de salida. Si, por ejemplo, por razones de incompatibilidad el recipiente de la corriente de datos de salida no es capaz de decodificar
55 datos TNS, puede ser útil el no copiar los datos espectrales respectivos de la señal de error y los adicionales parámetros TNS, sino procesar los datos reconstruidos de los datos relacionados con TNS para obtener la información en el dominio espectral y no utilizar el codificador TNS 870. Esto ilustra de nuevo que partes de los componentes o módulos ilustrados en la Figura 8 no tiene que ser implementados, sino que pueden ser

opcionalmente descartados.

[0133] En el caso de al menos una corriente de entrada de audio que compara datos PNS, puede aplicarse una estrategia similar. Si en la comparación de los cuadros para un componente espectral de las corrientes de datos de entrada revela que una corriente de datos de entrada está en términos de su cuadro presente 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 ser copiados directamente al componente espectral respectivo del cuadro de salida. Sin embargo, si el recipiente no es capaz de aceptar los parámetros PNS, la información espectral puede reconstruirse a partir del parámetro PNS para los componentes espectrales respectivos al generar ruido con el nivel de energía apropiado como se indica por el valor de energía respectivos. Después, los datos de ruido pueden ser procesados de conformidad en el dominio espectral.

[0134] Como se ha establecido anteriormente, los datos transmitidos también pueden comprender datos SBR que pueden ser procesados en el mezclador SBR 830. Replicación de banda espectral (SBR = Spectral Band Replication) es una técnica para replicar una parte de un espectro de una señal de audio con base en las contribuciones y la parte inferior del mismo espectro. Como una consecuencia, la parte superior del espectro no requiere ser transmitida, aparte de los parámetros SBR que describen valores de energía de una forma dependiente de frecuencia y dependiente de tiempo al emplear una rejilla de tiempo/frecuencia apropiada. Como una consecuencia, la parte superior del espectro no se requiere que sea transmitida de hecho. Para poder mejorar adicionalmente la calidad de la señal reconstruida, contribuciones de ruido y contribuciones sinusoidales adicionales pueden agregarse a la parte superior del espectro.

[0135] Para ser ligeramente más específico, para frecuencias sobre una frecuencia de cruce f_x , la señal 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 por un factor igual a, o proporcional al número de sub-bandas de banco de filtros QMF (por ejemplo 32 ó 64). Como consecuencia, puede determinarse una rejilla de tiempo/frecuencia que comprende en el eje del tiempo dos o más así denominados envolventes y, por cada envolvente, típicamente 7 a 16 valores de energía que describen la parte superior respectiva del espectro.

[0136] De forma adicional, los parámetros SBR pueden comprender información referente a adicionales ruido y sinusoides que después se atenúan o determinan con respecto a su fuerza por la rejilla de tiempo/frecuencia anteriormente mencionada.

[0137] En el caso de una corriente de datos de entrada basada en SBR que es la corriente de datos de entrada dominante con respecto al presente cuadro, puede realizarse copiado de los parámetros SBR respectivos junto con los componentes espectrales. Si de nuevo, el recipiente no es capaz de decodificar señales basadas en SBR, una reconstrucción respectiva en el dominio de frecuencia puede realizarse seguida por codificación de la señal reconstruida según los requisitos del recipiente.

[0138] Ya que SBR permite dos canales de codificación de estéreo, la codificación del canal izquierdo y el canal derecho separadas así como codificar los mismos en términos de un canal de acoplamiento (C), según una realización de la presente invención, copiando los parámetros SBR respectivos o al menos sus partes, puede comprender el copiar los elementos C de los parámetros SBR a ambos, los elementos izquierdo y derecho del parámetro SBR para determinarse y transmitirse o viceversa, dependiendo de los resultados de la comparación y los resultados de la determinación.

[0139] Además, ya que en diferentes realizaciones de la presente invención corrientes de datos de entrada pueden comprender ambas señales de audio mono y estéreo que comprenden uno y dos canales individuales, respectivamente una premezcla multipistas de mono a estéreo o un remezclado a dos pistas estéreo a mono pueden realizarse adicionalmente en el marco de copiar al menos parte de la información cuando se genera al menos parte de información de un componente espectral correspondiente del cuadro de la corriente de datos de salida.

[0140] Como ha mostrado la descripción anterior, el grado de copiar 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 que se van a copiar y puede determinar si la información espectral subyacente o piezas de la misma también se requiere que se copien. Por ejemplo, en el caso de copiar datos SBR, puede ser recomendable el copiar todo el cuadro de la corriente de datos respectiva para evitar información espectral de mezclado complicada para componentes espectrales diferentes. El

mezclar estos puede requerir una re-cuantificación que de hecho puede reducir el ruido de cuantificación.

[0141] En términos de los parámetros TNS, también puede ser recomendable copiar los parámetros TNS respectivos junto con la información espectral de todo el cuadro de la corriente de datos de entrada dominante a la corriente de datos de salida para evitar una re-cuantificación.

[0142] En caso de información espectral basada en PNS, copiar valores de energía individuales sin copiar los componentes espectrales subyacentes puede ser una forma viable. Además, en este caso al copiar sólo el parámetro PNS respectivo del componente espectral dominante de los cuadros de las pluralidades de corrientes de datos de entrada al componente espectral correspondiente del cuadro de salida de la corriente de datos de salida, ocurre sin introducir ruido de cuantificación adicional. Cabe destacar que también al re-cuantificar un valor de energía en la forma de un parámetro PNS, ruido de cuantificación adicional puede ser introducido.

[0143] Como se ha establecido anteriormente, la realización anteriormente resaltada también puede realizarse al copiar simplemente una información espectral referente a un componente espectral después de comparar los cuadros de la pluralidad de corrientes de datos de entrada y después de determinar, con base en la comparación para un componente espectral de un cuadro de salida de la corriente de datos de salida exactamente una corriente de datos para hacer la fuente de la información espectral.

[0144] El algoritmo de reemplazo realizado en el cuadro del módulo psico-acústico 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 sólo un componente activo sencillo. Para estas bandas, los valores cuantificados de la corriente de datos de entrada respectiva de la corriente de bits de entrada pueden copiarse del codificador sin volver a codificar o re-cuantificar los datos espectrales respectivos para el componente espectral específico. Bajo ciertas circunstancias, todos los datos cuantificados pueden tomarse de una sola señal de entrada activa para formar la corriente de bits de salida o corriente de datos de salida de tal manera que - en términos del aparato 500 - se logre una codificación sin pérdidas de la corriente de datos de limitación.

[0145] Además, puede ser posible omitir etapas de procesamiento tales como el análisis psico-acústico dentro del codificador. Esto permite acortar el proceso de codificación y de esta manera reducir la complejidad computacional ya que en principio sólo copiar datos de una corriente de bits en otra corriente de bits debe realizarse bajo ciertas circunstancias.

[0146] Por ejemplo, en el caso de PNS, puede llevarse a cabo un reemplazo ya que los factores de ruido de la banda codificada PNS pueden copiarse de una de las corrientes de datos de salida a la corriente de datos de salida. El reemplazo de componentes espectrales individuales con parámetros PNS apropiados es posible ya que los parámetros PNS son específicos de componente espectral, o, en otras palabras, a una muy buena aproximación independiente entre sí.

[0147] Sin embargo, puede ocurrir que una aplicación muy agresiva del algoritmo descrito puede producir una experiencia de audición degradada o una reducción indeseable en calidad. Por lo tanto, puede ser recomendable limitar el reemplazo a cuadros individuales en vez de información espectral, referente a componentes espectrales individuales. En tal modo de operación, la estimación de relevancia o determinación de relevancia, así como análisis de reemplazo pueden llevarse a cabo sin cambio. Sin embargo, un reemplazo puede en este modo de operación, llevarse a cabo solo cuando todo o al menos un número significativo de componentes espectrales dentro del cuadro activo se reemplazan.

[0148] Aunque esto puede llevar a un menor número de reemplazos, una fuerza interior de la información espectral puede en algunas situaciones mejorarse llevando a una calidad ligeramente mejorada.

[0149] A continuación, realizaciones según un segundo aspecto de la presente inversión se describen según lo cual valores de control asociados con datos de carga útil de las corrientes de datos de entrada respectivas se toman en cuenta, 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 donde en el caso en que valores de control de las dos corrientes de datos de entrada sean iguales, una nueva decisión en la forma en que el dominio espectral del cuadro respectivo de la corriente de datos de salida se evita y por el contrario la generación de corriente de salida se basa en la decisión ya determinada por los codificadores de la corriente de datos de entrada. Según algunas realizaciones descritas a continuación, se evita la re-transformación de los datos de carga útil respectivos de vuelta en otra forma de representar el dominio espectral tal como la ruta

normal o simple sin un valor espectral por muestra de tiempo/espectral.

[0150] Como se ha establecido anteriormente, realizaciones según la presente invención se basan en realizar un mezclado, que no se realiza de una forma directa en el sentido de que todas las corrientes de entrada se decodifican, lo que incluye una transformación inversa al dominio de tiempo, mezclado y de nuevo re-codificación de las señales. Realizaciones según la presente invención se basan en el mezclado efectuado en el dominio de frecuencia del codec respectivo. Un codec posible puede ser el codec AAC-ELD, o cualquier otro codec con una ventana de transformación uniforme. En tal caso, no se requiere transformación de tiempo/frecuencia para poder mezclar los datos respectivos. Además, acceso a todos los parámetros de corriente de bits tales como el tamaño de la etapa de cuantificación y otros parámetros, es posible y estos parámetros pueden emplearse para generar una corriente de bits de salida mixta.

[0151] Adicionalmente, el mezclado de las líneas espectrales o información espectral referente a componentes espectrales, puede llevarse a cabo por 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 cero significa que las fuentes que se tratan son irrelevantes y no se utilizarán de hecho. Grupos de líneas tales como bandas o bandas de factor de dimensionado o de escala pueden emplear el mismo factor de ponderación. Los factores de ponderación (por ejemplo, una distribución de ceros y unos) pueden variarse para los componentes espectrales de un solo cuadro de una corriente de datos de entrada sencilla. No se requiere con diferencia que las realizaciones descritas a continuación utilicen exclusivamente los factores de ponderación de cero o uno cuando se mezcla información espectral. Puede ser el caso que, bajo ciertas circunstancias, no por una sola, una, una pluralidad de información espectral total de un cuadro de una corriente de datos de entrada, los factores de ponderación respectivos pueden ser diferentes de cero a uno.

[0152] Un caso particular es que todas las bandas o componente espectral de una fuente (corriente 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, la corriente de bits de entrada completa de un participante puede copiarse de forma idéntica como una corriente de bits mezclada final. Los factores de ponderación pueden calcularse en una base de cuadro-a-cuadro, pero también pueden calcularse o determinarse con base en grupos o secuencias de cuadros más largos. De forma natural, incluso dentro de esta secuencia de cuadros o dentro de cuadros sencillos, los factores de ponderación pueden diferir para componentes espectrales diferentes, como se ha establecido anteriormente. Los factores de ponderación pueden en algunas realizaciones, ser calculados o determinados según resultados del modelo psico-acústico.

[0153] Tal comparación puede realizarse por ejemplo con base en la evaluación de una proporción de energía entre la señal mixta en la que solo algunas corrientes de entrada se incluyen y una señal mixta completa. Esto puede lograrse por ejemplo como se ha descrito anteriormente con respecto a las ecuaciones (3) a (5). En otras palabras, el modelo psico-acústico puede calcular la proporción de energía $r(n)$ entre una señal mixta en la que solo algunas corrientes de entrada se incluyen llevando a un valor de energía E_i y la señal mixta completa que tiene un valor de energía E_c . La proporción de energía $r(n)$ se calcula entonces según la ecuación (5) como 20 veces la logarítmica de E_i dividida por E_c .

[0154] Según esto, similar a la descripción anterior de realizaciones con respecto a las Figuras 6 a 8, si la proporción es suficientemente alta, los canales menos dominantes pueden considerarse como enmascarados por los dominantes. De esta manera, se procesa una reducción de irrelevancia, lo que significa que solo aquellas corrientes se incluyen que de hecho no son notables todas, a lo cual se atribuye un factor de ponderación de uno, mientras que todas las otras corrientes - al menos una información espectral de un componente espectral - se descarta. En otras palabras, para esto se atribuye un factor de ponderación de cero.

[0155] Esto puede llevar a una ventaja adicional que menos o ningún efecto de codificación en tándem ocurre debido a un número reducido de etapas de re-cuantificación. Ya que cada etapa de re-cuantificación tiene un peligro significativo de reducir ruido de cuantificación adicional, la calidad total de la señal de audio puede mejorarse, por lo tanto.

[0156] Similar a las realizaciones anteriormente descritas de las Figuras 6 a 8, las realizaciones descritas a continuación pueden emplearse con un sistema de conferencias que por ejemplo puede ser un sistema de conferencias de tele/vídeo con más de dos participantes y puede ofrecer la ventaja de una menor complejidad en comparación con un mezclado en dominio de tiempo ya que las etapas de transformación del tiempo-frecuencia y las etapas de recodificación pueden ser omitidas. Además, no se provoca mayor retraso por estos componentes en comparación con mezclado en el dominio de tiempo, debido a la ausencia del retraso de banco de filtros.

- [0157]** La Figura 9 muestra un diagrama de bloques simplificado de un aparato 500 para mezclar corrientes 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 Figuras 6 a 8 a fin de facilitar la comprensión y evitar descripciones duplicadas. Otros signos de referencia se han incrementado por 1000 con el fin de denotar que la funcionalidad de los mismos se define de forma diferente en comparación con las realizaciones anteriores de las Figuras 6 a 8 – tanto en funcionalidades adicionales como funcionalidad alterna, pero con la función general del elemento respectivo que es comparable.
- 10 **[0158]** Con base en la primera corriente de datos de entrada 510-1, y una segunda corriente de datos de entrada 510-2, una unidad de procesamiento 1520 comprendida en el aparato 1500 se adapta para generar una corriente de datos de salida 1530. La primera y segunda corrientes de datos de entrada 510 comprenden cada una un cuadro 540-1, 540-2, respectivamente, que comprende cada uno un valor de control 1545-1, 1545-2, respectivamente que indica una forma en que los datos de carga útil de los cuadros 540 representan al menos una parte del dominio espectral o información espectral de una señal de audio.
- 15 **[0159]** La corriente de datos de salida 530 comprende también un cuadro de salida 1550 con un valor de control 555, indicando de manera similar, una forma en la que los datos de carga útil del cuadro de salida 550 representan información espectral en el dominio espectral de la señal de audio codificada en la corriente de datos de salida 530.
- 20 **[0160]** La unidad procesadora 1520 del aparato 1500 se adapta para comparar los valores de control 1545-1 del cuadro 540-1 de la primera corriente de datos de entrada 510-1 y el valor de control 1545-2 de un cuadro 540-2 de la segunda corriente de datos de entrada 510-2 para dar un resultado de comparación. Con base en este resultado de comparación, la unidad procesadora 1520 se adapta además para generar la corriente de datos de salida 530 que comprende el cuadro de salida 550, de tal manera que cuando el resultado de comparación indica que los valores de control 1545 de los cuadros 540 de la primera y segunda corrientes de datos de entrada 510 son idénticas o iguales, el cuadro de salida 550 comprende como el valor de control 1550, un valor igual al de los valores de control 1545 de los cuadros 540 de las dos corrientes de datos de entrada 510. Los datos de carga útil comprendidos en el cuadro de salida 550 se derivan de los datos de carga útil correspondientes de los cuadros 540 con respecto a los valores de control idénticos 1545 de los cuadros 540 al procesar en el dominio espectral, es decir sin visitar el dominio de tiempo.
- 25 **[0161]** 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 las dos corrientes de datos de entrada son idénticas, entonces la información espectral correspondiente del cuadro de salida 550 correspondiente al mismo componente espectral o componentes espectrales, puede obtenerse al procesar los datos de carga útil correspondientes en el dominio espectral incluso directamente, es decir, al no dejar el tipo de representación del dominio espectral. Como se establecerá a continuación, en el caso de una representación espectral basada en PNS, esto puede lograrse al sumar los datos PNS respectivos, acompañados opcionalmente por un proceso de normalización. Es decir, los datos PNS de cualquier corriente de datos de entrada, se convierten de vuelta en representación simple con un valor por muestra espectral.
- 30 **[0162]** La Figura 10 muestra un diagrama más detallado de un aparato 1500 y difiere de la Figura 9, principalmente con respecto a una estructura interior de la unidad de procesamiento 1520. Para ser más específicos, la unidad de procesamiento 1520 comprende un comparador 1560, que se acopla a alimentaciones apropiadas para primeras y segundas corrientes de datos de entrada 510 y que se adapta para comparar los valores de control 1545 de sus cuadros respectivos 540. Las corrientes de datos de entrada se proporcionan además a un transformador opcional 1570-1, 1570-2, para cada una de las dos corrientes de datos de entrada 510. El comparador 1560 se acopla también a los transformadores opcionales 1570 para proporcionar los mismos con el resultado de comparación.
- 35 **[0163]** La unidad de procesamiento 1520 comprende además un mezclador 1580, que se acopla a modo 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 alimentaciones correspondientes para las corrientes 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 procesamiento 1520 y la del aparato 1500 para proporcionar la corriente de datos de salida 530.
- 40
45
50
55

[0164] Como se ha establecido anteriormente, la comparación 1560 se adapta para comparar los valores de control de los cuadros 1540 de las dos corrientes 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 los cuadros respectivos 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 los cuadros 540.

[0165] Los datos de carga útil comprendidos en los cuadros 540 de la corriente de datos de entrada 510 se mezclarán entonces por el mezclador 1580 y enviarán de salida al normalizador 1590, si se implementan, para realizar una etapa de normalización con el fin de asegurar que los valores resultantes no rebasen o estén por debajo de un intervalo de valores permisibles. Ejemplos de mezclado de datos de carga útil se establecerán con más detalle a continuación en el contexto de las Figuras 12a a 12c.

[0166] El normalizador 1590 puede implementarse como un cuantificador adaptado para re-cuantificar los datos de carga útil según sus valores respectivos, de forma alterna, el normalizador 1590 puede adaptarse también para alterar solo 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.

[0167] En el caso de que el comparador 1560 indique que los valores de control 1545 son al menos con respecto a uno o más componentes espectrales diferentes, el comparador 1560 puede proporcionar uno o ambos de los transformadores 1570 con una señal de control respectiva indicando los transformadores respectivos 1570 para transformar los datos de carga útil de al menos una de las corrientes de datos de entrada 510 con la otra de las corrientes de datos de entrada. En este caso, el transformador puede adaptarse para cambiar simultáneamente el valor de control del cuadro transformado de tal manera que el mezclador 1580 sea capaz de generar el cuadro de salida 550 de la corriente de datos de salida 530 con un valor de control 1555 que es igual al de un cuadro 540 de las dos corrientes de datos de entrada, que no se transforma o con un valor común de datos de carga útil de ambos cuadros 540.

[0168] Ejemplos más detallados se describirán a continuación en contexto con las Figuras 12a a 12c para diferentes aplicaciones tales como implementaciones PNS, implementaciones SBR e implementaciones M/S, respectivamente.

[0169] Cabe señalar que las realizaciones de las Figuras 9 a 12C no están limitadas con diferencia a dos corrientes de datos de entrada 1510-1, 1510-2 como se muestra en las Figuras 9, 10 y la siguiente Figura 11. Más bien, lo mismo puede adaptarse para procesar una pluralidad de corrientes de datos de entrada que comprenden más de dos corrientes de datos de entrada 510. En este caso, el comparador 1560, puede adaptarse por ejemplo para comparar una cantidad apropiada de corrientes de datos de entrada 510 y los cuadros 540 ahí comprendidos. Además, dependiendo de la implementación concreta, también puede implementarse un número apropiado de transformadores 1570. El mezclador 1580 junto con el normalizador opcional 1590 puede adaptarse eventualmente al número incrementado de corrientes de datos que se va a procesar.

[0170] En el caso de más de solo dos corrientes de datos de entrada 510, el comparador 1560 puede adaptarse para comparar todos los valores de control relevantes 1545 de las corrientes de datos de entrada 510 para decidir si una etapa de transformación se va a realizar por uno o más de los transformadores opcionalmente implementados 1570. De forma alterna o adicional, el comparador 1560 también puede adaptarse para determinar un conjunto de corrientes de datos de entrada para transformarse por los transformadores 1570, cuando el resultado de comparación indica que una transformación a una forma de representación común de los datos de carga útil es alcanzable. Por ejemplo, a menos que la representación diferente de datos de carga útil involucrados requiera una cierta representación, el comparador 1560 puede adaptarse, por ejemplo, para activar los transformadores 1570 de tal manera que reduzca al mínimo la complejidad total. Esto por ejemplo puede lograrse con base en estimaciones predeterminadas de valores de complejidad almacenados dentro del comparador 1560 o disponibles al comparador 1560 de una forma diferente.

[0171] Además, cabe destacar que el transformador 1570 puede ser omitido eventualmente cuando, por ejemplo, una transformación en el dominio de frecuencia puede llevarse a cabo opcionalmente por el mezclador 1580 ante demanda. De forma alternativa o adicionalmente, la funcionalidad de los transformadores 1570 puede incorporarse también en el mezclador 1580.

[0172] Además, cabe destacar que los cuadros 540 pueden comprender más de un valor de control, tal como

la sustitución de ruido perceptual (PNS), conformado de ruido temporal (TNS) y modos de codificación de estéreo. Antes de describir la operación 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 Figura 11 que es igual a la Figura 8 sin embargo, con los signos de referencia 1500 y 1520 que se utilizan en lugar de 500 y 520, respectivamente, para mostrar que la Figura 8 ya muestra una realización para generar una corriente de datos de salida de la primera y segunda corrientes de datos de entrada en las que las unidades de procesamiento 520 y 1520, respectivamente, pueden adaptarse también para llevar a cabo la funcionalidad descrita con respecto a las Figuras 9 y 10. En particular, con la unidad de procesamiento 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 anteriormente descritas que se establecen con respecto a las Figuras 9 y 10. Como se ha indicado anteriormente, los valores de control comprendidos en los cuadros de las corrientes de datos de entrada, igualmente pueden ser parámetros PNS, parámetros SBR o datos de control referentes a codificación de estéreo, en otras palabras, 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 para ser adicionalmente procesados comprendidos en el cuadro de salida de la corriente de datos de salida. En este aspecto, como ya se ha establecido anteriormente, ya que SBR permite dos canales de codificación de estéreo, la codificación del canal izquierdo y el canal derecho separadamente, así como la codificación de los mismos en términos de un canal de acoplamiento (C), según una realización de la presente invención, el procesamiento de los parámetros SBR respectivos o al menos sus partes, puede comprender el procesamiento de los elementos C de los parámetros SBR para obtener ambos, los elementos izquierdo y derecho del parámetro SBR o vice-versa, dependiendo de los resultados de la comparación y el resultado de la determinación. De forma similar, el grado de procesamiento 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 que se van a procesar y puede determinar si también se requiere que la información espectral subyacente o partes de la misma se decodifiquen. Por ejemplo, en el caso de copiar datos SBR, puede ser recomendable el procesamiento de todo el cuadro de la corriente de datos respectiva para evitar un mezclado complicado de información espectral para diferentes componentes espectrales. El mezclado de estos puede requerir una re-cuantificación que, de hecho, 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 todo el cuadro a partir de la corriente de datos de entrada dominante a la corriente de datos de salida para evitar una re-cuantificación. En el caso de información espectral basada en PNS, el procesamiento de valores de energía individuales sin copiar los componentes espectrales subyacentes puede ser una forma viable. Además, en este caso al procesar solo el parámetro PNS respectivo a partir del componente espectral dominante a partir de los cuadros de la pluralidad de corrientes de datos de entrada al componente espectral correspondiente del cuadro de salida de la corriente de datos de salida, ocurre sin introducir ruido de cuantificación adicional. Cabe destacar que también al re-cuantificar un valor de energía en la forma de un parámetro PSN, puede introducirse ruido de cuantificación adicional.

[0173] Con respecto a las Figuras 12A a 12C, tres modos diferentes de mezclado de datos de carga útil en base a una comparación de valores de control respectivos, se describirán con más detalle. La Figura 12a muestra un ejemplo de una implementación basada en PNS de un aparato 500 según una realización de la presente invención, mientras que la Figura 12b muestra una implementación SBR similar y la Figura 12c muestra una implementación M/S de la misma.

[0174] La Figura 12a muestra un ejemplo con una primera y una segunda corrientes de datos de entrada 510-1, 510-2, respectivamente, con cuadros de entrada apropiados 540-1, 540-2 y valores de control respectivos 545-1, 545-2. Como se indica por las flechas en la Figura 11a, los valores de control 1545 de los cuadros 540 de las corrientes 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, en otras palabras, por un parámetro PNS apropiado. Más específicamente, la Figura 12a muestra un primer parámetro PNS 2000-1 y el cuadro 540-2 de la segunda corriente de datos de entrada 510-2 que comprende un parámetro PNS 2000-2.

[0175] Ya que, como se considera con respecto a la Figura 12a, los valores de control 1545 de los dos cuadros 540 de las dos corrientes de datos de entrada 510 indican que el componente espectral específico se va a reemplazar por el parámetro PNS respectivo 2000, la unidad de procesamiento 1520 y el aparato 1500, como se ha descrito anteriormente, es capaz de mezclar los dos parámetros PNS 2000-1, 2000-2 para llegar al parámetro PNS 2000-3 del cuadro de salida 550 para incluirse en la corriente de datos de salida 530. El valor de control respectivo 1555 del cuadro de salida 550 esencialmente indica también que el componente espectral respectivo se va a reemplazar por el parámetro PNS 2000-3 mixto. Este proceso de mezclado se ilustra en la Figura 12a al mostrar el parámetro PNS 2000-3 como los parámetros PNS combinados 2000-1, 2000-2 de los cuadros respectivos 540-1,

540-2.

[0176] Sin embargo, la determinación del parámetro PNS 2000-3, que también se refiere como un parámetro de salida PNS, también puede lograrse con base en una combinación lineal según

5

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

en el que PNS(i) es el parámetro PNS respectivo de la corriente de datos de entrada i, N es el número de corrientes de datos de entrada que se va a mezclar y a_i es el factor de ponderación apropiado. Dependiendo de la
10 implementación concreta, los factores de ponderación a_i pueden seleccionarse para ser iguales

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

[0177] Una implementación directa que se ilustra en la Figura 12a puede ser que cuando todos los
15 parámetros de ponderación a_i son iguales a 1, en otras palabras,

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

[0178] En el caso de que un normalizador 1590 como se muestra en la Figura 10 se vaya a omitir, los
20 factores de ponderación pueden definirse como iguales a $1/N$ de tal manera que la ecuación

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

se mantiene.

25

[0179] El parámetro N aquí es el número de corrientes de datos de entrada que se va a mezclar, y el número de corrientes de datos de entrada que se proporciona al aparato 1500, es un número similar. Por razones de simplicidad, cabe destacar que pueden implementarse también normalizaciones diferentes en términos de factores de ponderación a_i .

30

[0180] En otras palabras, en el caso de una herramienta PNS activada en el lado del participante, el factor de energía de interferencia reemplaza 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 la corriente de datos de salida por la herramienta PNS. En el caso de mezclar componentes
35 espectrales PNS, puede llegar a dos casos distintos.

[0181] Como se ha descrito anteriormente, cuando los componentes espectrales respectivos de todos los cuadros 540 de las corrientes de datos de entrada relevantes cada uno se expresa en términos de parámetros PNS. Ya que los datos de frecuencia de una descripción relacionada a PNS de un componente de frecuencia (por ejemplo, banda de frecuencia) se derivan directamente del factor de energía de interferencia (parámetro PNS), los factores apropiados pueden mezclarse al agregar simplemente los valores respectivos. El parámetro PNS mixto se generará entonces dentro del decodificador PNS en el lado de recipiente una resolución de frecuencia equivalente a mezclarse con los valores espectrales puros de otros componentes espectrales. En el caso de que se utilice un proceso de normalización durante el mezclado, puede ayudar la implementación de un factor de normalización
40 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).

[0182] En el caso en que los valores del control 1545 de al menos una corriente de datos de entrada 510 difieran con respecto a un componente espectral, y si las corrientes de datos de entrada respectivas no se descartan debido a un bajo nivel de energía, puede ser recomendable que el decodificador PNS como se muestra en la Figura
50 11 genere la información espectral o datos espectrales con base 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.

- [0183]** Debido a la independencia de los componentes espectrales PNS entre sí, y con respecto a parámetros globalmente definidos de la corriente de datos de salida, así como las corrientes de datos de entrada, una selección del procedimiento de mezclado puede adaptarse en una base a lo ancho de la banda. En el caso de que este mezclado basado en PNS no sea posible, puede ser recomendable considerar la re-codificación del componente espectral respectivo por el codificador PNS 1880 después de mezclar en el dominio espectral.
- 10 **[0184]** La Figura 12b muestra un ejemplo adicional de un principio operacional de una realización según una realización de la presente invención. Para ser más precisos, la Figura 12b muestra el caso de dos corrientes de datos de entrada 510-1, 510-2 con cuadros apropiados 540-1, 540-2 y sus valores de control 1545-1, 1545-2. Los cuadros 540 comprenden datos SBR para componentes espectrales sobre una frecuencia de cruce f_x así denominada. El valor de control 1545 comprende información de si los parámetros SBR se utilizan de hecho, e información referente a la rejilla de cuadro actual o la rejilla de tiempo/frecuencia.
- 15 **[0185]** Como se ha establecido 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 un número de ranuras de tiempo por cada cuadro SBR que es igual a los cuadros 540 de la corriente de datos de entrada 510 que comprende también información espectral adicional. Las ranuras de tiempo separan el intervalo de frecuencia de la herramienta SBR en bandas de frecuencia o componentes espectrales igualmente espaciados. El número de estas bandas de frecuencia en un cuadro SBR se determinará por el remitente o la herramienta SBR antes de codificar. En el caso de un MPEG-4 AAC-ELD, el número de ranuras de tiempo se fija en 16.
- 20 **[0186]** Las ranuras de tiempo se incluyen ahora en los así denominados envolventes que cada envolvente comprende al menos dos o más ranuras de tiempo que forman un grupo respectivo. Cada envolvente se atribuye a un número de datos de frecuencia SBR. En la rejilla de cuadro o rejilla de tiempo/frecuencia, se almacenan el número y longitud en unidades de ranuras de tiempo de las envolventes individuales.
- 25 **[0187]** La resolución de frecuencia de las envolventes individuales determina que tantos datos de energía SBR se calculan para una envolvente y almacenan con respecto a la misma. La herramienta SBR difiere solo entre una alta y una baja resolución, en la que una envolvente comprende una alta resolución comprende dos veces valores de una envolvente con una baja resolución. El número de valores de frecuencia o componentes espectrales para envolventes que comprenden una alta o baja resolución, depende de parámetros adicionales del codificador tales como velocidad de bits, frecuencia de muestreo y así sucesivamente.
- 30 **[0188]** En el contexto de MPEG-4 AAC ELD, la herramienta SBR a menudo utiliza de 16 a 14 valores con respecto a la envolvente que tiene una alta resolución.
- 35 **[0189]** Debido a la división dinámica del cuadro 540 con un número apropiado de valores de energía con respecto a frecuencia, puede considerarse una transitoria. En el caso de que una transitoria esté presente en un cuadro, el codificador SBR divide el cuadro respectivo en una cantidad apropiada de envolventes. Esta distribución se estandariza en el caso de la herramienta SBR empleada con el codec AAC ELD y depende de la posición de la transposición transitoria en unidades de la ranura de tiempo. En muchos casos, el cuadro de rejilla o rejilla de tiempo/frecuencia resultante comprende tres envolventes cuando está presente una transitoria. Una primera envolvente, la envolvente de inicio comprende el inicio de un cuadro hasta la ranura de tiempo que recibe la transitoria que tiene los índices de ranura de tiempo cero a transposición-1. La segunda envolvente comprende una longitud de dos ranuras de tiempo que circunscriben la transitoria desde la transposición de índice de ranura de tiempo a transposición+2. La tercera envolvente comprende todas las ranuras de tiempo restantes con los índices de transposición+3 a 16.
- 40 **[0190]** Sin embargo, la longitud mínima de una envolvente es de dos ranuras de tiempo. Como consecuencia, cuadros que comprenden una transitoria cerca de las fronteras de cuadro pueden comprender eventualmente solo dos envolventes. En caso de que no haya presente transitoria en el cuadro, las ranuras de tiempo se distribuyen sobre envolventes igualmente largas.
- 45 **[0191]** La Figura 12b ilustra tal rejilla de tiempo/frecuencia o rejilla de cuadro dentro de los cuadros 540. En el caso en que los valores de control 1545 indican que las mismas rejillas de tiempo SBR o rejillas de tiempo/frecuencia
- 50

están presentes en los dos cuadros 540-1, 540-2, los datos SBR respectivos pueden copiarse similar al procedimiento descrito en contexto con las ecuaciones (6) a (9) anteriores. En otras palabras, en dicho caso la herramienta de mezclado SBR o el mezclador SBR 830, como se muestra en la Figura 11, puede copiar la rejilla de tiempo/frecuencia o rejilla de cuadro de los cuadros de entrada respectivos al cuadro de salida 550 y calcular los valores de energía respectivos similares a las ecuaciones (6) a (9).

[0192] Incluso en otras palabras, los datos de energía SBR de la rejilla de cuadro pueden mezclarse simplemente al sumar los datos respectivos y opcionalmente al normalizar los datos respectivos.

10 **[0193]** La Figura 12c muestra un ejemplo adicional de un modo de operación de una realización según la presente invención. Para ser más precisos, la Figura 12c muestra una implementación M/S. De nuevo, la Figura 12c muestra dos corrientes de datos de entrada 510 junto con dos cuadros 540 y valores de control asociados 545 que indican una forma en que se representan los cuadros de datos de carga útil 540, al menos con respecto a un componente espectral de los mismos como mínimo.

15

[0194] Los cuadros 540 comprenden cada uno datos de audio o información espectral de dos canales, un primer canal 2020, y un segundo canal 2030. Dependiendo del valor de control 1545 del cuadro respectivo 540, el primer canal 2020 puede ser, por ejemplo, un canal izquierdo o un canal medio, 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 a menudo se refiere como un modo LR, mientras que el segundo modo a menudo se refiere como un modo M/S.

20

[0195] En el modo M/S, que en ocasiones también se refiere como estéreo conjunto, el canal medio (M) se va a definir como proporcional a una suma del canal izquierdo (L) y de canal derecho (R). A menudo, un factor adicional de $\frac{1}{2}$ se incluye en la definición, de tal manera que el canal medio comprende en ambos el dominio de tiempo y el dominio de frecuencia, un valor promedio de los dos canales de estéreo.

25

[0196] El canal lateral se define típicamente como proporcional a una diferencia de los dos canales de estéreo, es decir, proporcional a una diferencia del canal izquierdo canal (L) y el canal derecho (R). En ocasiones, también un factor adicional de $\frac{1}{2}$ se incluye de tal manera que el canal lateral actualmente represente la mitad del valor de desviación entre los dos canales de la señal de estéreo, o la desviación del canal medio. Según esto, el canal izquierdo puede ser reconstruido al sumar el canal medio y el canal lateral, mientras que el canal derecho puede obtenerse al sustraer el canal lateral del canal medio.

30

[0197] En el caso de que para los cuadros 540-1 y 540-2 se utilice la misma codificación de estéreo (L/R o M/S), una retransformación de los canales comprendidos en el cuadro puede omitirse, permitiendo un mezclado directo en el dominio de codificación L/R- o M/S- respectivo.

35

[0198] En este caso, el mezclado puede ser llevado a cabo de nuevo directamente en el dominio de frecuencia que lleva a un cuadro 550 comprendido en una corriente 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 los dos cuadros 540. El cuadro de salida 550 comprende de manera correspondiente dos canales 2020-3, 2030-3 derivados del primer y segundo canales de los cuadros de la corriente de datos de entrada.

40

[0199] En el caso en que los valores de control 1545-1, 1545-2 de los dos cuadros 540 no sean iguales, puede ser recomendable transformar uno de los cuadros en la otra representación con base en el proceso descrito anteriormente. El valor de control 1555 del cuadro de salida 550 puede ajustarse de conformidad al valor indicativo del cuadro transformado.

45

[0200] Según realizaciones de la presente invención, puede ser posible que los valores de control 1545, 1555 que indiquen una representación de todo el cuadro 540, 550, respectivamente o los valores de control respectivos puedan ser específicos del componente de frecuencia. Mientras que, en el primer caso, los canales 2020, 2030 se codifican sobre todo el cuadro por uno de los procedimientos específicos, en el segundo caso, en principio, cada información espectral con respecto a un componente espectral puede ser codificada de forma diferente. De manera natural, también sub-grupos de componentes espectrales pueden describirse por uno de los valores de control 1545.

55

[0201] Adicionalmente, puede realizarse un algoritmo de reemplazo en el marco del módulo psico-acústico 950 para examinar cada una de las piezas de información espectral referentes a los componentes espectrales subyacentes (por ejemplo, bandas de frecuencia) de la señal para identificar componentes espectrales con solo un componente activo sencillo. Para estas bandas, los valores cuantificados de la corriente de datos de entrada

respectiva de la corriente de bits de entrada pueden copiarse del codificador sin re-codificación o recuantificación de los datos espectrales respectivos para el componente espectral específico. Bajo ciertas circunstancias todos los datos cuantificados pueden tomarse de una sola señal de entrada activa para formar la corriente de bits de salida o la corriente de datos de salida, de tal manera que - en términos del aparato 1500 - se alcance una codificación sin pérdida de la corriente de datos de entrada.

[0202] Además, puede ser posible omitir etapas del procesamiento tal como el análisis psico-acústico dentro del codificador. Esto permite acortar el proceso de codificación y de esta manera, reducir la complejidad computacional ya que en principio solo la copia de datos de una corriente de bits en otra corriente de bits debe realizarse bajo ciertas circunstancias.

[0203] Por ejemplo, en el caso de PNS, puede llevarse a cabo un reemplazo ya que factores de ruido en la banda codificada PNS pueden copiarse de una de las corrientes de datos de salida a la corriente de datos de salida. El reemplazo de componentes espectrales individuales con parámetros PNS apropiados es posible, ya que los parámetros PNS son específicos de componente espectral o, en otras palabras, en una muy buena aproximación, independientes entre sí.

[0204] Sin embargo, puede ocurrir que una aplicación muy agresiva del algoritmo descrito pueda dar por resultado una experiencia de audición degradada o una reducción indeseada en calidad. Por tanto, puede ser recomendable limitar el reemplazo a cuadros individuales, en vez de información espectral, referente a componentes espectrales individuales. En este modo de operación la estimación de irrelevancia o determinación de irrelevancia, así como el análisis de reemplazo pueden llevarse a cabo sin cambio. Sin embargo, un reemplazo puede en este modo de operación, llevarse a cabo solo cuando todo o al menos un número significativo de componentes espectrales dentro del cuadro activo son reemplazables.

[0205] Aunque esto puede llevar a un número menor de reemplazos, una fuerza interior de la información espectral en algunas situaciones puede mejorarse llevando a una calidad incluso ligeramente mejorada.

[0206] Las realizaciones anteriormente establecidas pueden de forma natural, diferir con respecto a sus implementaciones. Aunque en las realizaciones anteriores, se ha descrito una decodificación y codificación Huffman como un solo esquema de codificación de entropía, también pueden emplearse otros esquemas de codificación de entropía. Además, la implementación de un codificador de entropía o un decodificador de entropía no se requiere con diferencia. Según esto, aunque la descripción de las realizaciones anteriores se ha enfocado primordialmente en el codec ACC-ELD, otros codecs pueden emplearse también para proporcionar las corrientes de datos de entrada y para decodificar la corriente de datos de salida en el lado del participante. Por ejemplo, cualquier codec que se basa por ejemplo en una sola ventana sin conmutación de longitud de bloque, puede emplearse.

[0207] Como la descripción anterior de las realizaciones mostró en las Figuras 8 y 11, por ejemplo, también ha mostrado que los módulos descritos ahí no son obligatorios. Por ejemplo, un aparato según una realización de la presente invención puede lograrse simplemente al operar en la información espectral de los cuadros.

[0208] Cabe destacar que las realizaciones anteriormente descritas con respecto a las Figuras 6 a 12C pueden lograrse formas muy diferentes. Por ejemplo, un aparato 500/1500 para mezclar una pluralidad de corrientes de datos de entrada y su unidad de procesamiento 520/1520, puede lograrse en base a dispositivos eléctricos y electrónicos discretos tales como resistores, transistores, inductores y semejantes. Además, realizaciones según la presente invención también pueden lograrse con base solo en circuitos integrados, por ejemplo, en la forma de sistema en chip, SOC (SOC = System On Chip), procesadores tales como UPCs, unidad de procesamiento central (CPU = Central Processing Unit), unidad de procesamiento gráfico, GPU (GPU = Graphic Processing Unit), y otros circuitos integrados (IC), tales como circuitos integrados específicos de aplicación (ASIC).

[0209] Se debería destacar también que dispositivos eléctricos que son parte de la implementación discreta o son parte de un circuito integrado, pueden emplearse para diferentes propósitos y diferentes funciones a través de la implementación de un aparato según una realización de la presente invención. Naturalmente, también una combinación de circuitos con base en circuitos integrados y circuitos discretos, puede emplearse para implementar una realización según la presente invención.

[0210] Con base en un procesador, realizaciones según la presente invención pueden implementarse también con base en un programa informático, un programa de soporte lógico o un programa que se ejecuta en un procesador.

[0211] En otras palabras, dependiendo de ciertos requisitos de implementación de realizaciones de los procedimientos de la invención, realizaciones de los procedimientos de la invención pueden implementarse en equipo físico o en soporte lógico. La implementación puede realizarse utilizando un medio de almacenamiento digital, en particular un disco, un CD o un DVD que tiene señales legibles electrónicamente almacenadas que cooperan con un ordenador programable o procesador, de tal manera que se realice una realización de la invención. En general, una realización de la presente invención por lo tanto es un producto de programa informático con un código de programa almacenado en un soporte legible por máquina, el código de programa es operativo para realizar una realización del procedimiento de la invención cuando el producto del programa informático se ejecuta en un ordenador o procesador. Incluso en otras palabras, realizaciones de los procedimientos de la invención por lo tanto son un programa informático que tiene un código de programa para realizar 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 formarse por un ordenador, una tarjeta de circuito integrado, una tarjeta inteligente, un circuito integrado específico de aplicación, un sistema en chip (SOC), o un circuito integrado CI (IC = Integrated Circuit).

15

Lista de signos de referencia

[0212]

- 20 100 Sistema de conferencia
 - 110 Entrada
 - 120 Decodificador
 - 130 Sumador
 - 140 Codificador
- 25 150 Salida
 - 160 Terminal de conferencia
 - 170 Codificador
 - 180 Decodificador
 - 190 Convertidor de tiempo/frecuencia
- 30 200 Cuantificador/codificador
 - 210 Decodificador/descuantificador
 - 220 Convertidor de frecuencia/tiempo
 - 250 Corriente de datos
 - 260 Cuadro
- 35 270 Bloques de información adicional
 - 300 Frecuencia
 - 310 Banda de frecuencia
 - 500 Aparato
 - 510 Corriente de datos de entrada
- 40 520 Unidad de procesamiento
 - 530 Corriente de datos de salida
 - 540 Cuadro
 - 550 Cuadro de salida
 - 560 Componente espectral
- 45 570 Flecha
 - 580 Línea punteada
 - 700 Decodificador de corriente de bits
 - 710 Lector de corriente de bits
 - 720 Codificador Huffman
- 50 730 Descuantificador
 - 740 Dimensionador
 - 750 Primera unidad
 - 760 Segunda unidad
 - 770 Decodificador estéreo
- 55 780 Decodificador PNS
 - 790 Decodificador TNS
 - 800 Unidad de mezclado
 - 810 Mezclador espectral
 - 820 Módulo de optimización

- 830 Mezclador SBR
- 850 Codificador de corriente de bits
- 860 Tercera unidad
- 870 Codificador TNS
- 5 880 Codificador PNS
- 890 Codificador estéreo
- 900 Cuarta unidad
- 910 Dimensionador
- 920 Cuantificador
- 10 930 Codificador Huffman
- 940 Escritor de corriente de bits
- 950 Módulo psico-acústico
- 1500 Aparato
- 1520 Unidad de procesamiento
- 15 1545 Valor de control
- 1550 Cuadro de salida
- 1555 Valor de control

REIVINDICACIONES

1. Un aparato (500) para mezclar una pluralidad de corrientes de datos de entrada de participantes de un sistema de conferencia, en el que las corrientes de datos de entrada (510) comprenden cada una un cuadro de datos de audio en un dominio espectral, un cuadro (540) de una corriente de datos de entrada (510) que comprende información espectral para una pluralidad de componentes espectrales, comprendiendo el aparato (500) :
 5 una unidad de procesamiento (520) adaptada para comparar los cuadros de la pluralidad de corrientes de datos de entrada (510) con base en un modelo psico-acústico, considerando un enmascarado inter-canales,
 10 en el que la unidad de procesamiento (520) está adaptada además para determinar, con base en la comparación, para un componente espectral de un cuadro de salida (550) de una corriente de datos de salida (530), exactamente una corriente de datos de entrada (510) de la pluralidad de corrientes de datos de entrada (510) que como una corriente de datos de entrada dominante enmascara las otras corrientes de datos de entrada (510) de la pluralidad
 15 de corrientes de datos de entrada (510); y
- en el que la unidad de procesamiento (520) está adaptada además para generar la corriente de datos de salida al copiar el componente espectral del cuadro de salida (550) de la corriente de datos de salida (530) de al menos una parte de información de un componente espectral correspondiente del cuadro (540) de la corriente de datos de
 20 entrada determinada (510), sin recodificar o re-cuantificar el componente espectral correspondiente, con descarte de piezas de información del componente espectral correspondiente del cuadro de la otra corriente de datos de entrada (510).
2. El aparato según la reivindicación 1, en el que la unidad de procesamiento (520) está adaptada de tal
 25 manera que comparar los cuadros de la pluralidad de corrientes de datos de entrada (510) se basa en al menos dos piezas de información espectral correspondientes al mismo componente espectral de cuadros (540) de dos corrientes de datos de entrada diferentes (510).
3. El aparato según la reivindicación 1 ó 2, en el que el aparato (500) está adaptado de tal manera que
 30 un componente espectral de una pluralidad de componentes espectrales corresponde a una frecuencia o una banda de frecuencias.
4. Aparato (500) según cualquiera de las reivindicaciones 1 a 3, en el que la unidad de procesamiento
 35 (520) está adaptada de tal manera que generar la corriente de datos de salida comprende copiar al menos parte de la información del componente espectral correspondiente solo del cuadro de la corriente de datos de entrada determinado (510) para describir el componente espectral del cuadro de salida (550) de la corriente de datos de salida (530).
5. El aparato (500) según cualquiera de las reivindicaciones 1 a 4, en el que la unidad de procesamiento
 40 (520) está adaptada de tal manera que generar la corriente de datos de salida comprende copiar datos de audio en el dominio espectral, que corresponden al componente espectral del cuadro de la corriente de datos de entrada determinada (510).
6. El aparato (500) según cualquiera de las reivindicaciones 1 a 5, en el que las corrientes de datos de
 45 entrada (510) de la pluralidad de corrientes de datos de entrada (510) comprenden cada una, con respecto al tiempo, una secuencia de cuadros de datos de audio en el dominio espectral y en el que la unidad de procesamiento (520) está adaptada de tal manera que comparar los cuadros (540) se basa en cuadros que solo corresponden a un índice de tiempo común de la secuencia de cuadros.
- 50 7. El aparato (500) según cualquiera de las reivindicaciones 1 a 6, en el que la unidad de procesamiento (520) está adaptada de tal manera que generar la corriente de datos de salida (530) mantiene una distribución de niveles de cuantificación en comparación con una distribución de niveles de cuantificación de al menos parte de la información del componente espectral correspondiente del cuadro de la corriente de entrada determinada (510).
- 55 8. El aparato (500) según cualquiera de las reivindicaciones 1 a 7, en el que al menos parte de la información del componente espectral correspondiente comprende información referente a niveles de cuantificación, un parámetro de sustitución de ruido perceptual (PNS), un parámetro de sustitución de ruido temporal (TNS) o un parámetro de replicación de banda espectral (SBR).

9. El aparato (500) según cualquiera de las reivindicaciones 1 a 8, en el que la unidad de procesamiento (520) está adaptada además para realizar la determinación con base en la comparación para determinar exactamente una corriente de datos de entrada (510) de la pluralidad de corrientes de datos de entrada (510) para cada uno de los componentes espectrales diferentes y
- 5 en el que la unidad de procesamiento (520) está adaptada además para generar la corriente de datos de salida al copiar al menos la parte de información del componente espectral respectivo del cuadro (540) de la corriente de datos de entrada determinada (510) para cada uno de los diferentes componentes espectrales para describir el componente espectral respectivo del cuadro de salida (550) de la corriente de datos de salida (530) de tal manera
- 10 que el cuadro de salida de la corriente de datos de salida (530) ha copiado la parte de información como mínimo de los componentes espectrales respectivos de los diferentes de la pluralidad de corrientes de datos de entrada,
- o en el que
- 15 la unidad de procesamiento (520) está adaptada además para realizar la determinación con base en la comparación para determinar exactamente una corriente de datos de entrada (510) de la pluralidad de corrientes de datos de entrada (510) para primeros componentes espectrales y determinar que no hay corriente de datos de entrada dominante para un segundo componente espectral, y
- 20 en el que la unidad de procesamiento (520) está adaptada además para generar la corriente de datos de salida al copiar al menos la parte de información del componente espectral respectivo del cuadro (540) de la corriente de datos de entrada determinada (510) para los primeros componentes espectrales, para describir el primer componente espectral del cuadro de salida (550) de la corriente de datos de salida (530) de tal manera que el cuadro de salida de la corriente de datos de salida (530) ha copiado en él al menos parte de la información del
- 25 primer componente espectral a partir de la corriente de datos de entrada determinada y al mezclar el segundo componente espectral de la pluralidad de corrientes de entrada de datos en el dominio espectral con el fin de describir el segundo componente espectral del cuadro de salida (550) de la corriente de datos de salida (530).
10. Un procedimiento para mezclar una pluralidad de corrientes de datos de entrada (510) de participantes
- 30 de un sistema de conferencia, en el que las corrientes de datos de entrada (510) comprenden cada una un cuadro (540) de datos de audio en un dominio espectral, un cuadro (540) de una corriente de datos de entrada (510) que comprende una pluralidad de componentes espectrales, comprendiendo el procedimiento:
- comparar los cuadros (540) de la pluralidad de corrientes de datos de entrada (510), con base en un modelo psico-
- 35 acústico, considerando un enmascarado inter-canales;
- determinar con base en la comparación para un componente espectral de un cuadro de salida (550) de una corriente de datos de salida (530) exactamente una corriente de datos de entrada (510) de la pluralidad de corrientes de datos de entrada (510) que como una corriente de datos de entrada dominante enmascara las otras corrientes de datos de
- 40 entrada (510) de la pluralidad de corrientes de datos de entrada (510); y generar la corriente de datos de salida (530) al copiar al menos una parte de una pieza de información del componente espectral del cuadro de salida (550) de la corriente de datos de salida (530) a partir de un componente espectral correspondiente del cuadro de la corriente de datos de entrada determinada (510), sin recodificar o re-cuantificar el componente espectral correspondiente, descartando piezas de información del componente espectral correspondiente del cuadro de la otra corriente de
- 45 datos de entrada (510).
11. Un programa informático para realizar, cuando se ejecuta en un procesador, un procedimiento para mezclar una pluralidad de corrientes de datos de entrada (510) según la reivindicación 10.

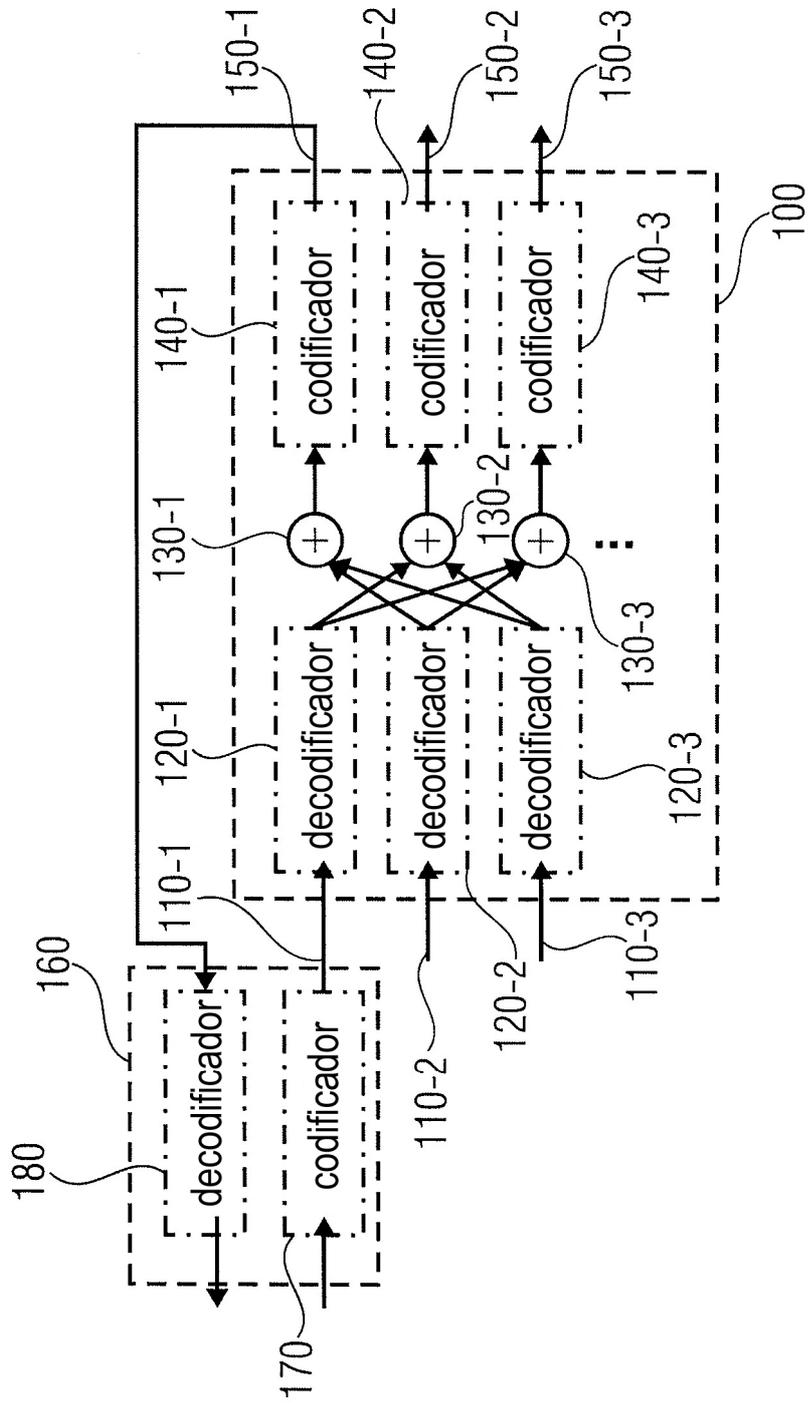


FIG 1

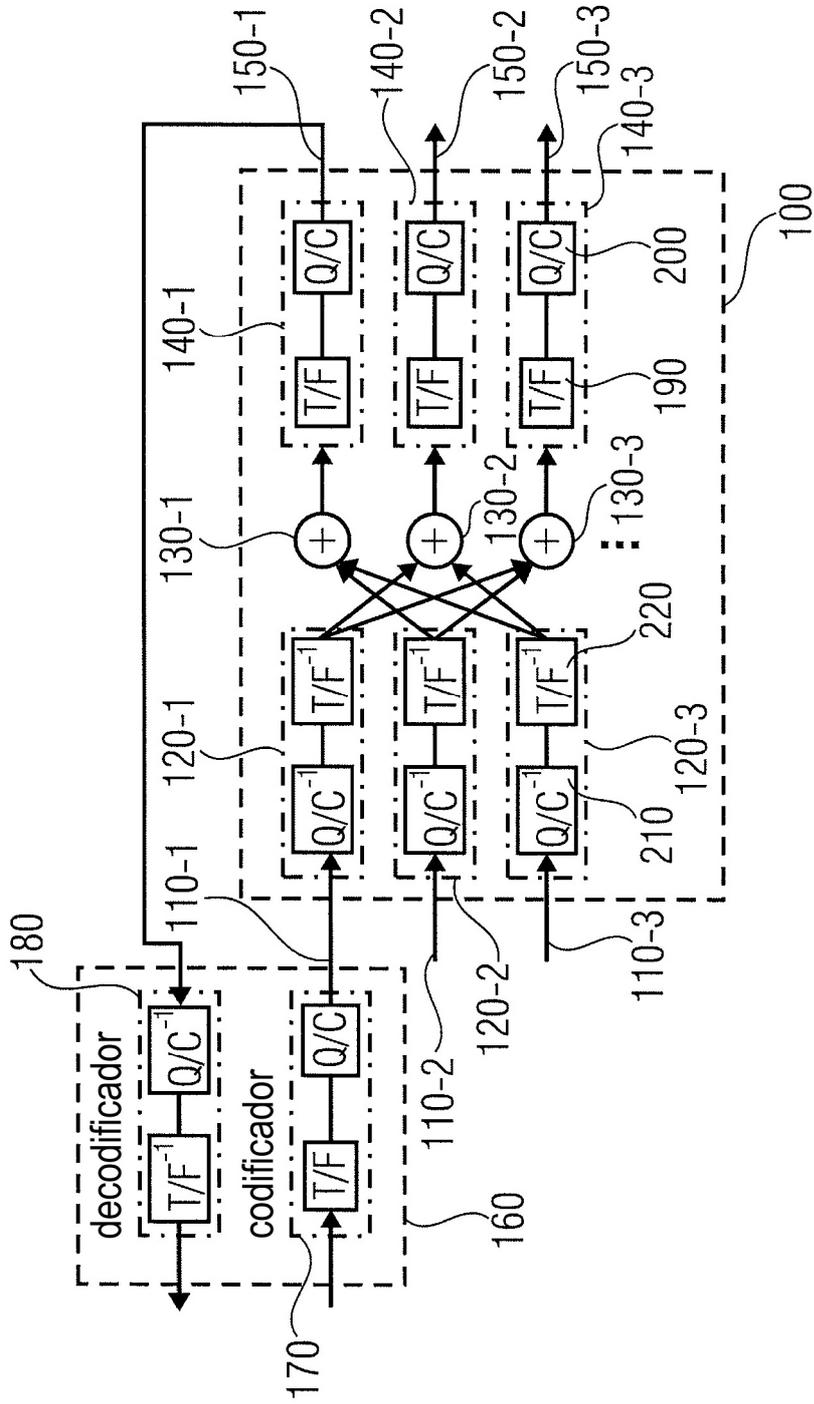


FIG 2

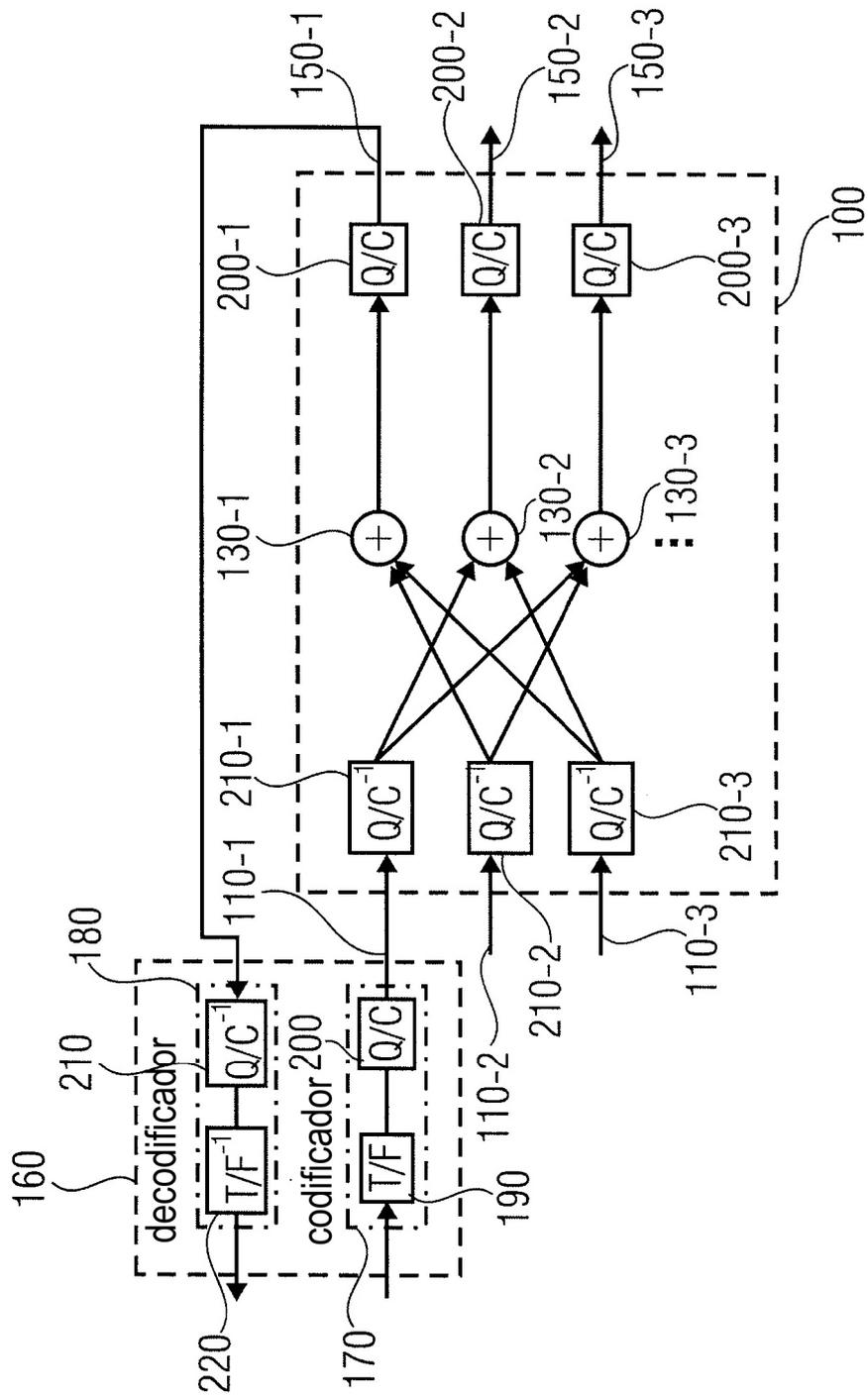


FIG 3

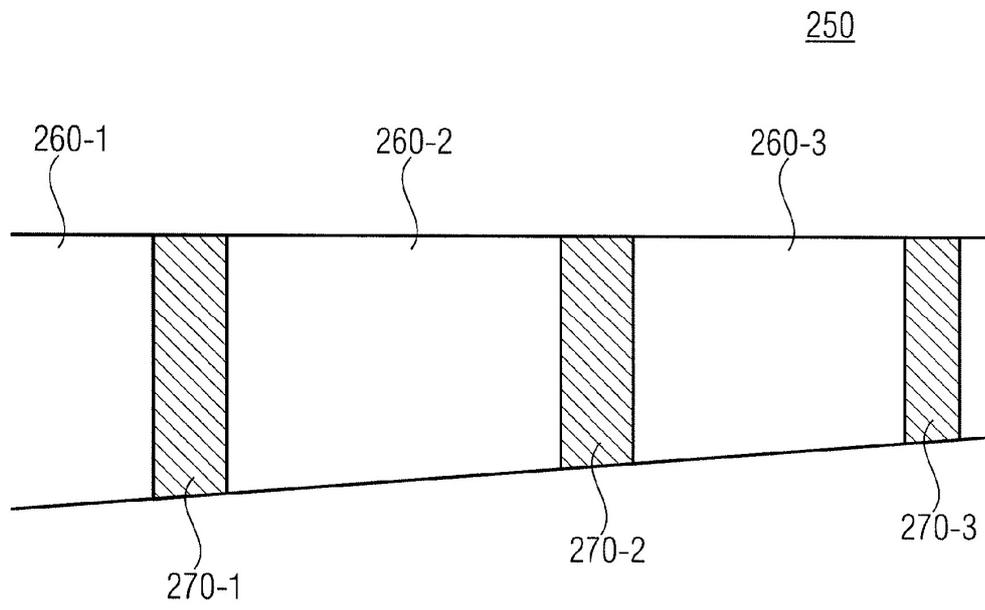


FIG 4

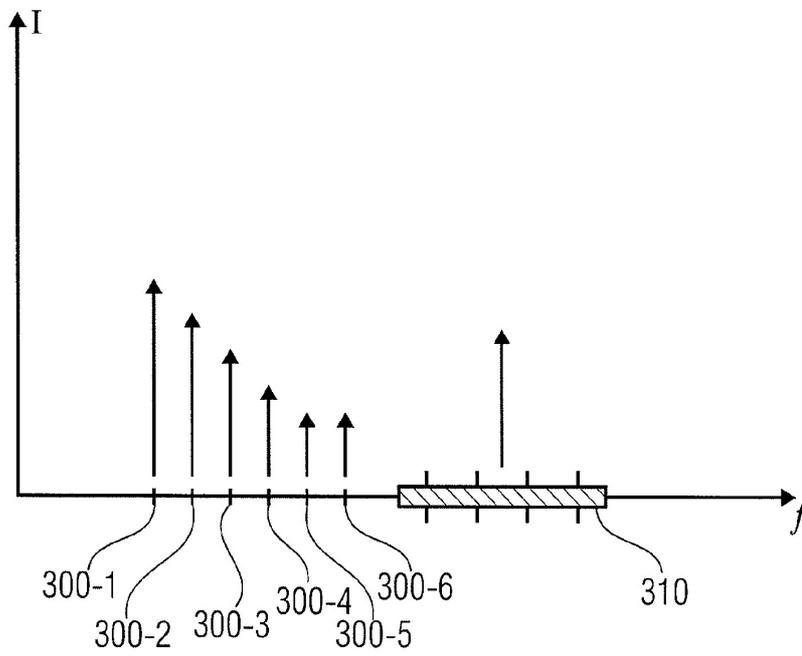


FIG 5

500

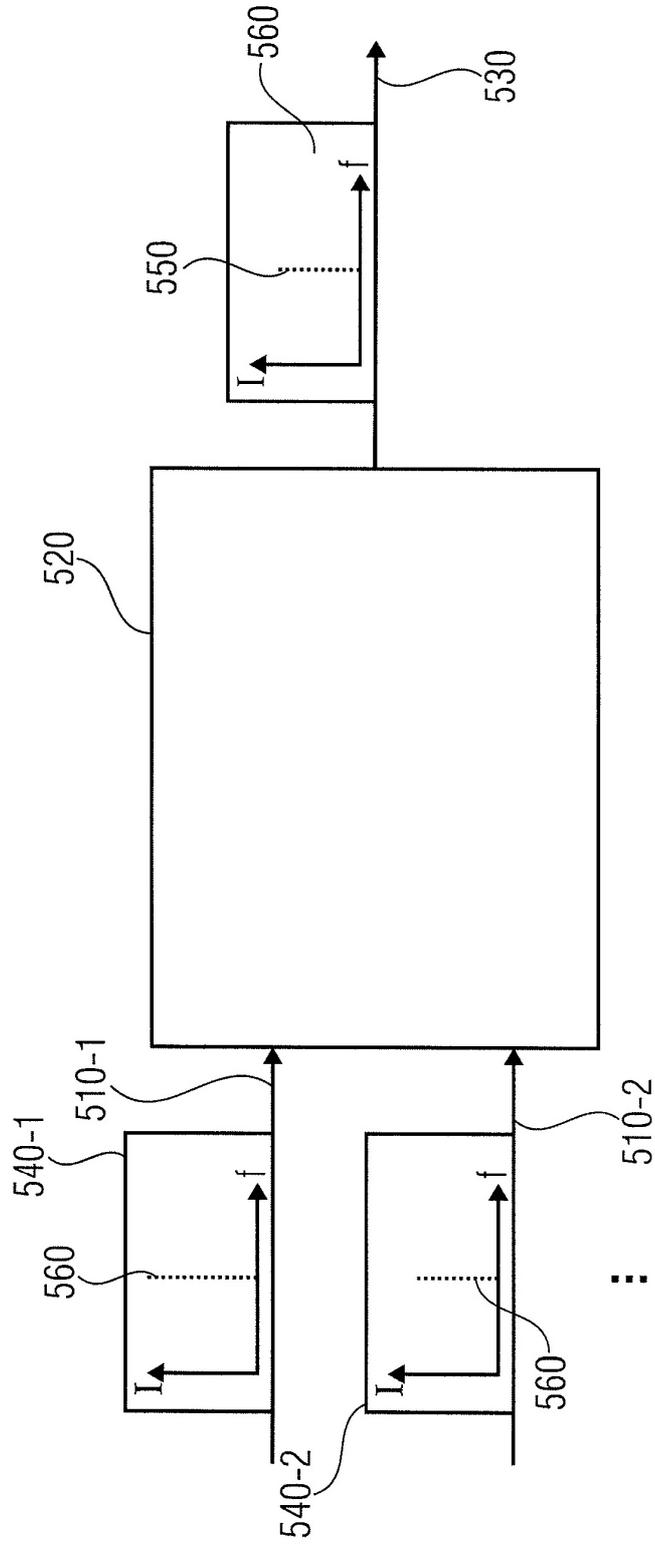


FIG 6

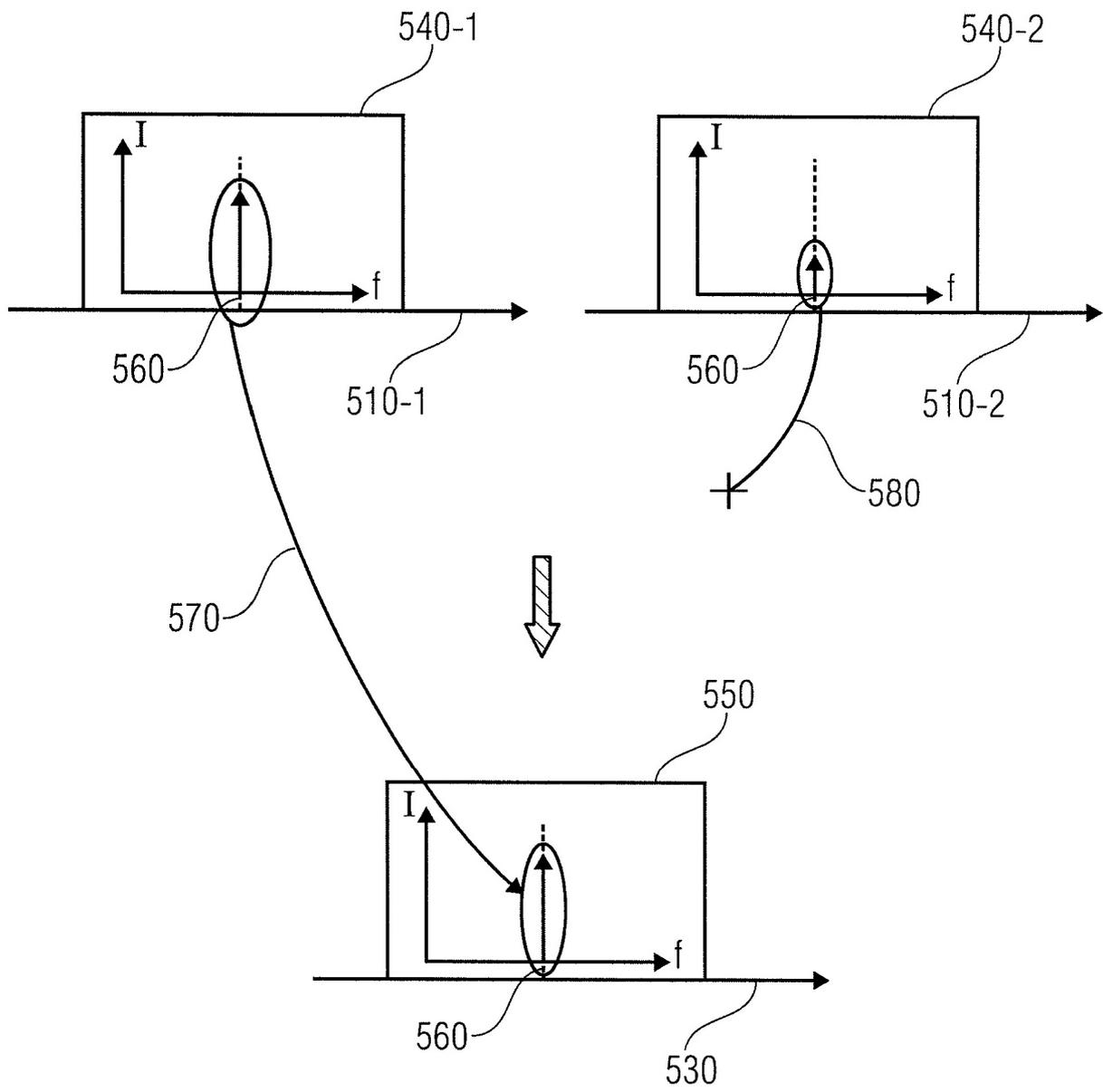


FIG 7

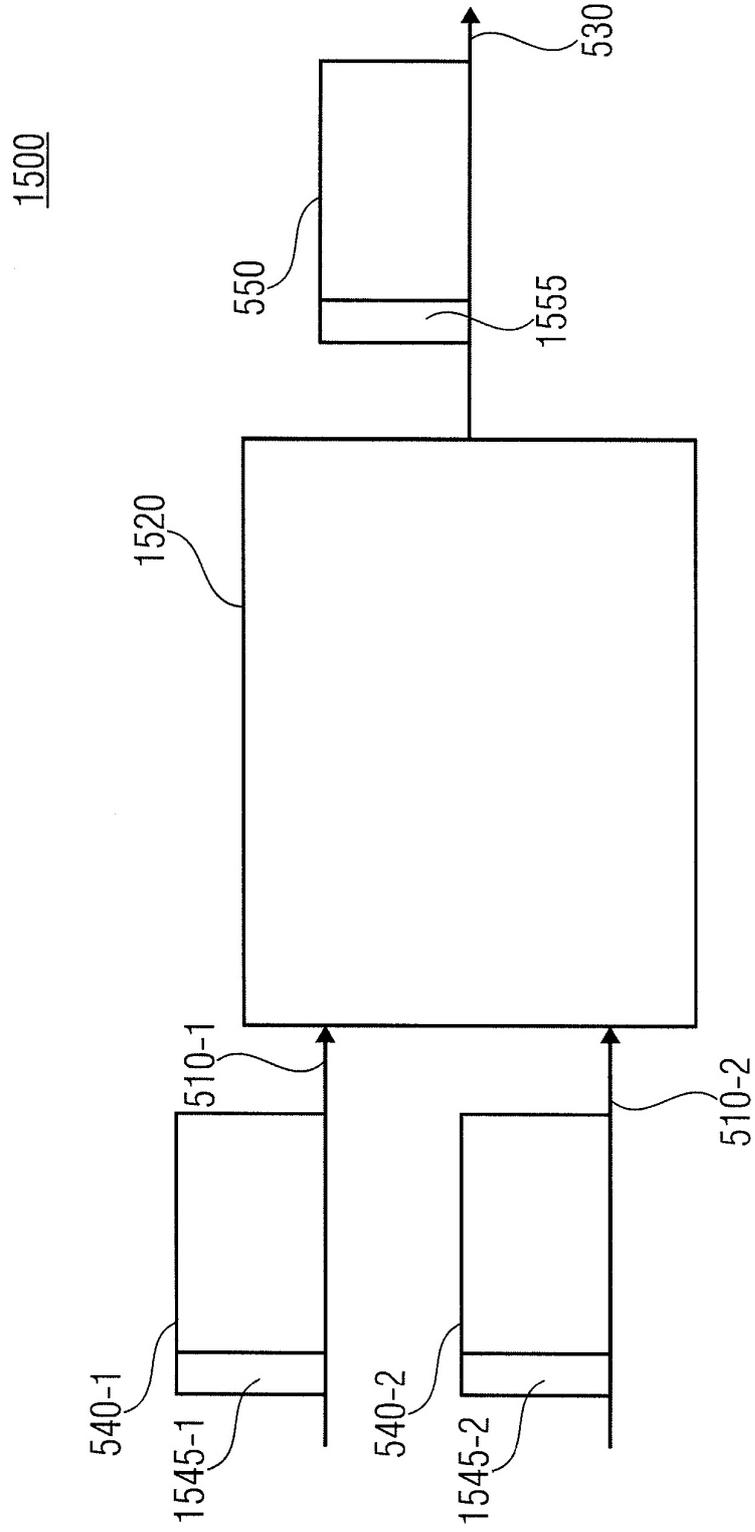


FIG 9

1500

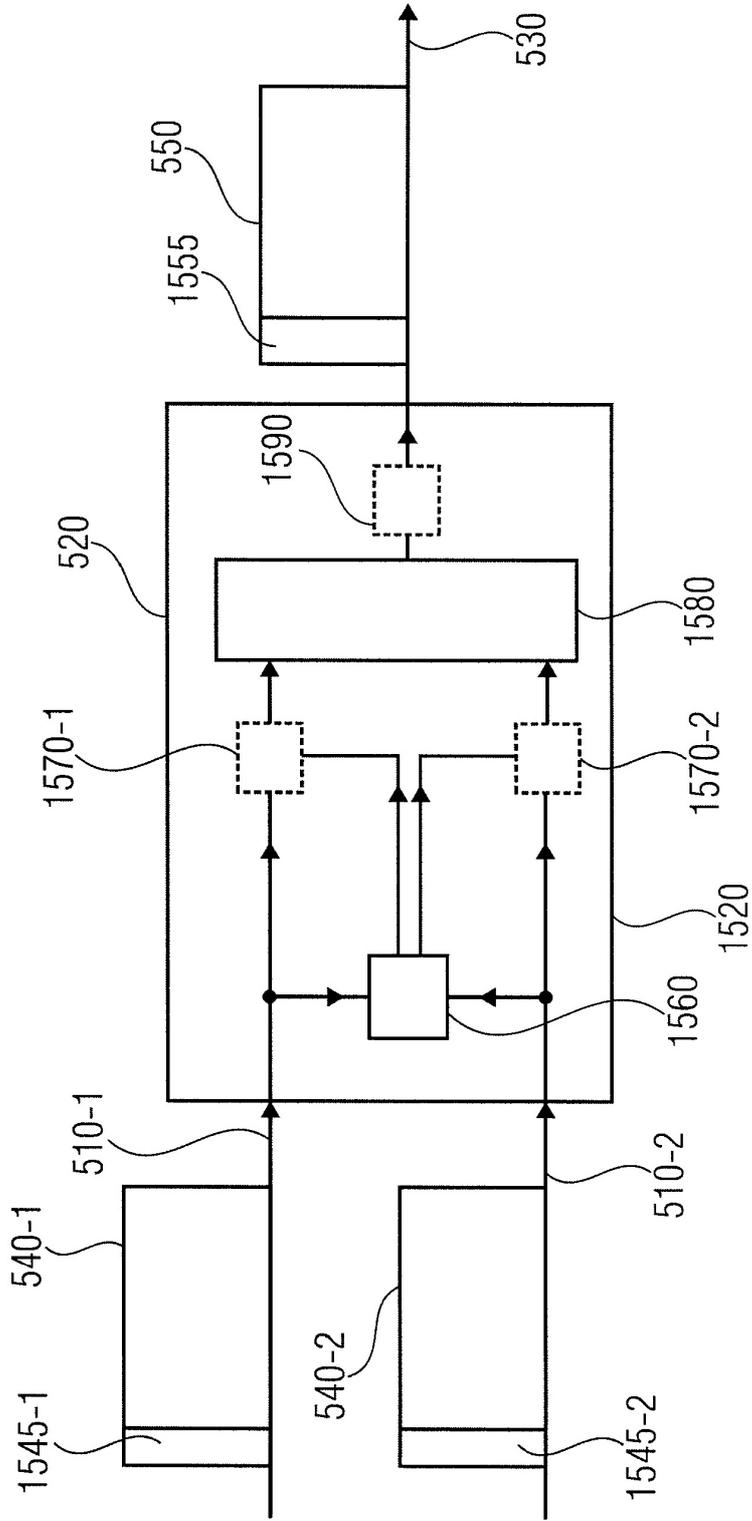


FIG 10

1500

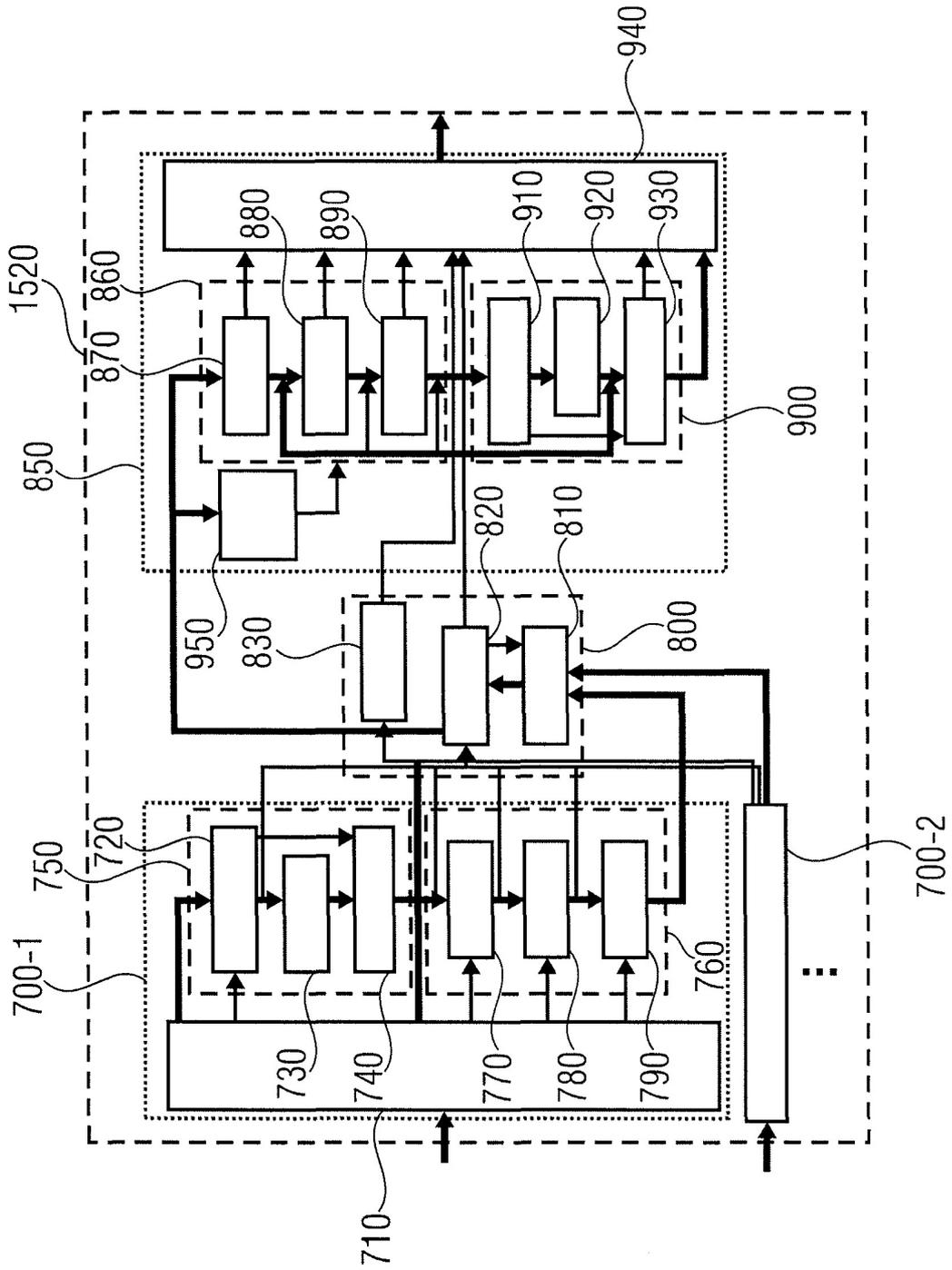


FIG 11

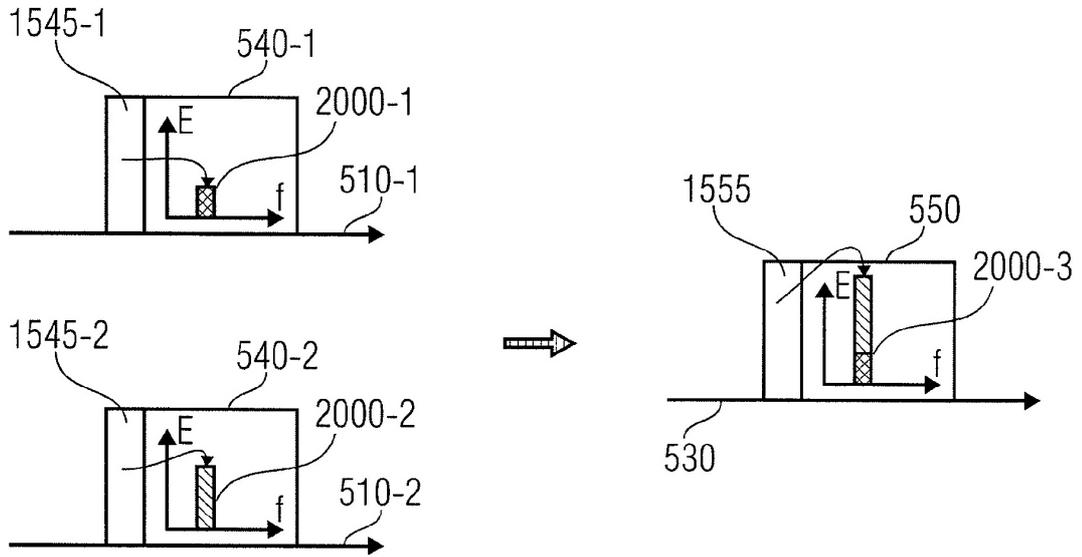


FIG 12A

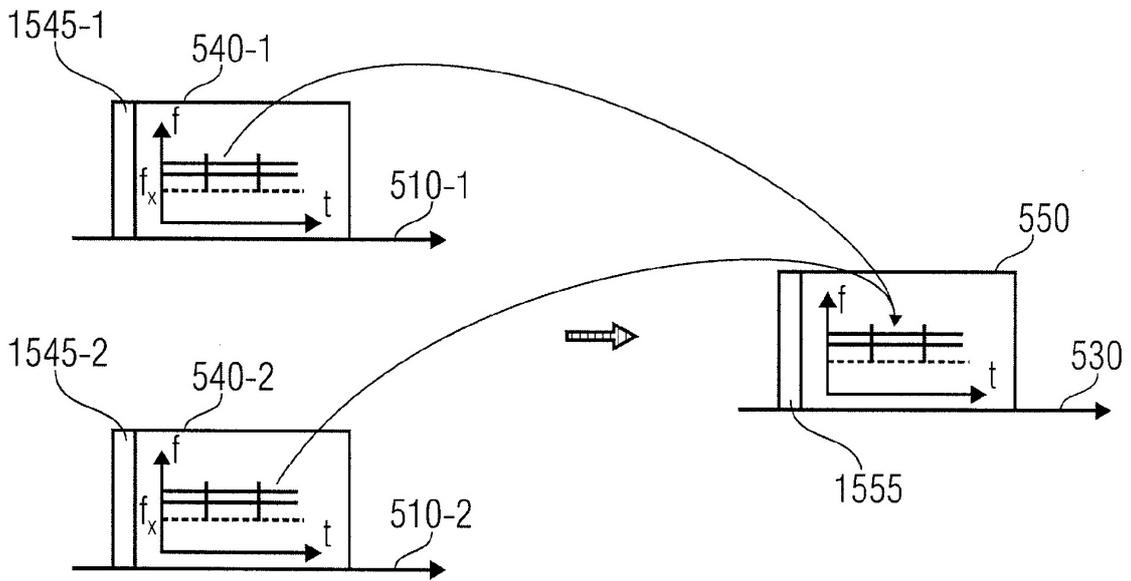


FIG 12B

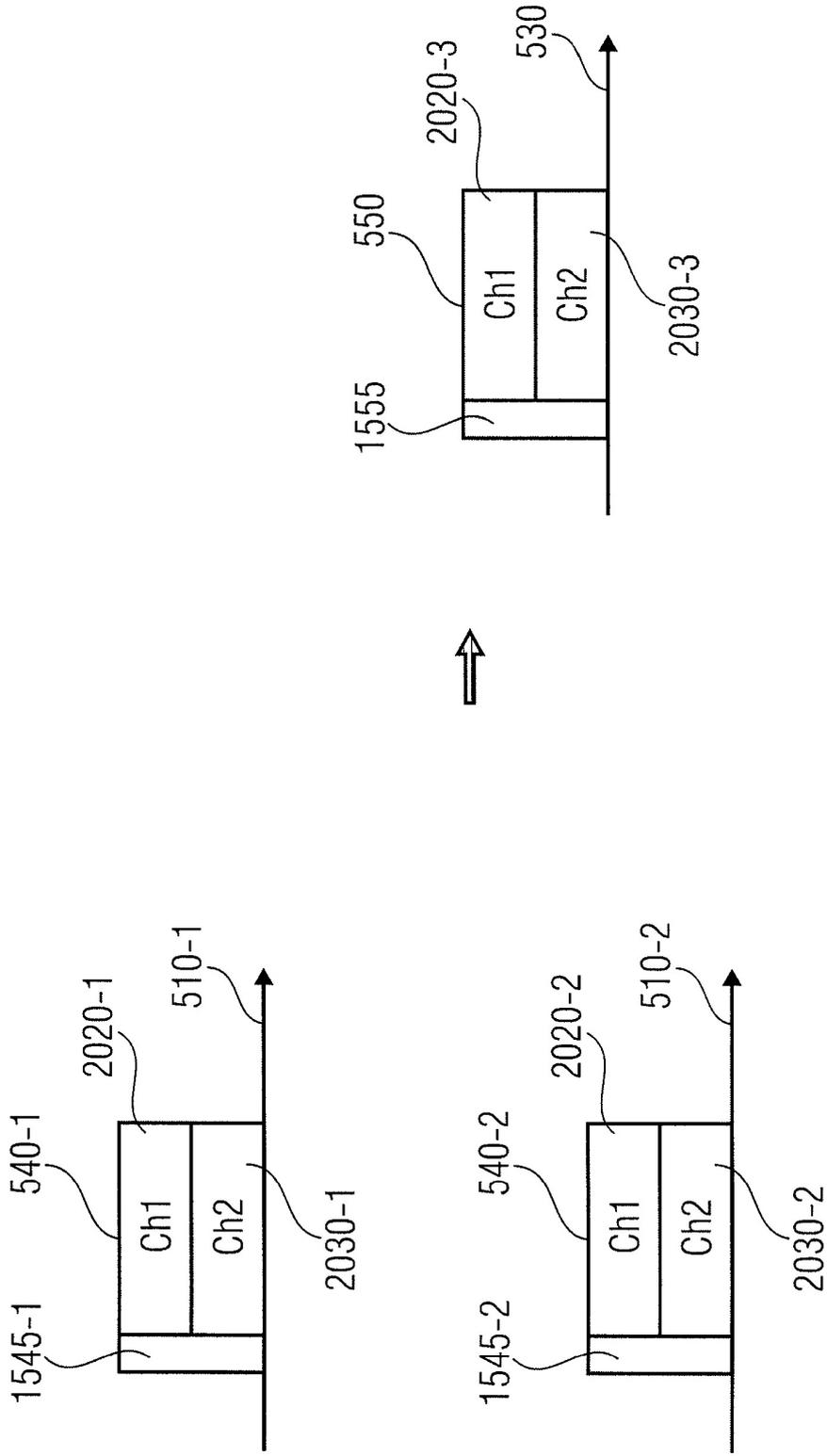


FIG 12C