

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 374 496**

51 Int. Cl.:  
**G10L 19/00** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 96 Número de solicitud europea: **09716202 .8**  
96 Fecha de presentación: **04.03.2009**  
97 Número de publicación de la solicitud: **2250641**  
97 Fecha de publicación de la solicitud: **17.11.2010**

54 Título: **APARATO PARA MEZCLAR UNA PLURALIDAD DE FLUJOS DE DATOS DE ENTRADA.**

30 Prioridad:  
**04.03.2008 US 33590**

45 Fecha de publicación de la mención BOPI:  
**17.02.2012**

45 Fecha de la publicación del folleto de la patente:  
**17.02.2012**

73 Titular/es:  
**Fraunhofer-Gesellschaft zur Förderung der  
angewandten Forschung e.V.  
Hansastraße 27c  
80686 München, DE**

72 Inventor/es:  
**SCHNELL, Markus;  
LUTZKY, Manfred y  
MULTRUS, Markus**

74 Agente: **Arizti Acha, Monica**

**ES 2 374 496 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

Aparato para mezclar una pluralidad de flujos de datos de entrada

5 Las realizaciones según la presente invención se refieren a aparatos para mezclar una pluralidad de flujos de datos de entrada para obtener un flujo de datos de salida, que pueden usarse, por ejemplo, en el campo de los sistemas de conferencia que incluyen sistemas de videoconferencia y sistemas de teleconferencia.

10 En muchas aplicaciones deben procesarse más de una señal de audio de tal manera que, a partir del número de señales de audio, se genera una señal o al menos un número reducido de señales, lo que se denomina frecuentemente como "mezcla". El proceso de mezcla de señales de audio, por tanto, puede denominarse como agrupamiento de varias señales de audio individuales en una señal resultante. Este proceso se usa, por ejemplo, cuando se crean piezas de música para un disco compacto ("dubbing"). En este caso, en una canción normalmente se mezclan diferentes señales de audio de diferentes instrumentos junto con una o más señales de audio que comprenden ejecuciones vocales (canto).

15 Campos adicionales de aplicación en los cuales la mezcla desempeña un papel importante son los sistemas de videoconferencia y los sistemas de teleconferencia. Un sistema de este tipo normalmente puede conectar varios participantes distribuidos espacialmente en una conferencia empleando un servidor central, que mezcla apropiadamente los datos de vídeo y audio entrantes de los participantes registrados y ena través de a cada uno de los participantes una señal resultante en respuesta. Esta señal resultante o señal de salida comprende las señales de audio de todos los demás participantes de la conferencia.

20 En los sistemas de conferencia digitales modernos, varios objetivos y aspectos en parte contradictorios compiten entre sí. La calidad de una 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 en comparación con señales de audio generales y señales musicales) tienen que tomarse en consideración. Aspectos adicionales que pueden tener que considerarse también cuando se diseñan e implementan sistemas de conferencia son las cuestiones de ancho de banda disponible y retardo.

25 Por ejemplo, cuando se equilibra la calidad por una parte y el ancho de banda por otra parte, un compromiso es en la mayoría de los casos inevitable. Sin embargo, mejoras concernientes a la calidad pueden obtenerse implementando técnicas de codificación y decodificación modernas tales como la técnica de AAC-ELD (AAC = *Advanced Audio C*ódec, códec de audio avanzado; ELD = *Enhanced Low Delay*, bajo retardo mejorado). Sin embargo, la calidad que puede lograrse puede verse afectada negativamente en sistemas que emplean tales técnicas modernas por problemas y aspectos más fundamentales.

30 Por nombrar sólo un desafío que ha de lograrse, todas las transmisiones de señales digitales se enfrentan al problema de una cuantificación necesaria, que puede, al menos en principio, ser evitable en circunstancias ideales en un sistema analógico sin ruido. Debido al proceso de cuantificación se introduce eventualmente una cierta cantidad de ruido de cuantificación a la señal que va a procesarse. Para contrarrestar distorsiones posibles y audibles, podría pensarse en incrementar el número de niveles de cuantificación y, por tanto, incrementar la resolución de cuantificación de manera correspondiente. Esto, sin embargo, conduce a un mayor número de valores de señal que deben transmitirse y, por tanto, a un incremento en la cantidad de datos que deben transmitirse. En otras palabras, la mejora de la calidad reduciendo posibles distorsiones introducidas por el ruido de cuantificación podría en determinadas circunstancias incrementar la cantidad de datos que deben transmitirse y puede eventualmente violar las restricciones de ancho de banda impuestas en un sistema de transmisión.

35 En el caso de los sistemas de conferencia, los desafíos de mejorar un compromiso entre calidad, ancho de banda disponible y otros parámetros pueden ser aún más complicados por el hecho de que normalmente deben procesarse más de una señal de audio de entrada. Por tanto, las condiciones de frontera impuestas por más de una señal de audio pueden tener que tomarse en consideración cuando se genera la señal de salida o señal resultante producida por el sistema de conferencia.

40 Especialmente en vista del desafío adicional de implementar sistemas de conferencia con un retardo suficientemente bajo, permitir una comunicación directa entre los participantes de una conferencia sin introducir retardos sustanciales que puedan considerarse inaceptables por los participantes incrementa adicionalmente el desafío.

45 En implementaciones con bajo retardo de sistemas de conferencia, las fuentes de retardo están restringidas normalmente en cuanto a su número, lo que por otra parte podría conducir al desafío de procesar los datos fuera del dominio de tiempo, en el cual la mezcla de las señales de audio puede obtenerse superponiendo o sumando las señales respectivas.

50 Para mejorar el compromiso entre calidad y tasa de transmisión de bits en el caso de señales de audio generales, existe un número significativo de técnicas que pueden mejorar adicionalmente un compromiso entre tales parámetros contradictorios, tales como la calidad de una señal reconstruida, la tasa de transmisión de bits, el retardo, la complejidad computacional y parámetros adicionales.

Una herramienta sumamente flexible para mejorar el compromiso mencionado anteriormente es la denominada herramienta de representación de banda espectral (SBR). El módulo de SBR normalmente no se implementa para formar parte de un codificador central, tal como el codificador AAC de MPEG-4, sino que es más bien un codificador y decodificador adicional. SBR utiliza una correlación entre las frecuencias más altas y más bajas dentro de una señal de audio. SBR se basa en la suposición de que las frecuencias más altas de una señal son solamente múltiplos enteros de una oscilación base, de tal manera que las frecuencias más altas pueden replicarse basándose en el espectro inferior. Puesto que la resolución audible del oído humano en el caso de las frecuencias más altas logarítmicamente, las pequeñas diferencias concernientes a intervalos de frecuencia más altos sólo pueden percibirse además por oyentes muy experimentados de tal manera que las inexactitudes introducidas por el codificador de SBR lo más probable es que pasen desapercibidas para la gran mayoría de oyentes.

El codificador de SBR procesa previamente la señal de audio proporcionada al codificador de MPEG-4 y separa la señal de entrada en intervalos de frecuencia. El intervalo de frecuencia más bajo o banda de frecuencia más baja se separa de una banda de frecuencia o intervalo de frecuencia superior mediante una denominada frecuencia de cruce, que puede establecerse de manera variable, dependiendo de la tasa de transmisión de bits disponible y parámetros adicionales. El codificador de SBR utiliza un banco de filtros para analizar la frecuencia, que se implementa normalmente para que sea una banda de filtros espejo en cuadratura (QMF).

El codificador de SBR extrae a partir de la representación de frecuencias del intervalo de frecuencia superior valores de energía que se usarán más tarde para reconstruir este intervalo de frecuencia basándose en la banda de frecuencia inferior.

El codificador de SBR, por tanto, proporciona datos de SBR o parámetros de SBR junto con una señal de audio filtrada o datos de audio filtrados a un codificador de núcleo o codificador central, que se aplica a la banda de frecuencia inferior basándose en la mitad de la frecuencia de muestreo de la señal de audio original. Esto proporciona la oportunidad de procesar significativamente menos valores de muestra, de tal manera que los niveles de cuantificación individuales pueden establecerse con mayor exactitud. Los datos adicionales proporcionados por el codificador de SBR, concretamente los parámetros de SBR, se almacenarán en un flujo de bits resultante por el codificador de MPEG-4 o cualquier otro codificador como información lateral. Esto puede lograrse usando un multiplexor de bits apropiado.

En el lado del decodificador, los flujos de bits entrantes se demultiplexan en primer lugar por un demultiplexor de bits, que separa al menos los datos de SBR y los proporciona a un decodificador de SBR. Sin embargo, antes de que el decodificador de SBR procese los parámetros de SBR, la banda de frecuencia inferior se decodificará primero por un decodificador central o decodificador de núcleo para reconstruir la señal de audio de la banda de frecuencia inferior. El propio decodificador de SBR calcula, basándose en los niveles de energía de SBR (parámetros de SBR) y la información espectral del intervalo de frecuencia inferior, la parte superior del espectro de la señal de audio. En otras palabras, el decodificador de SBR replica la banda espectral superior de la señal de audio basándose en la banda inferior así como en los parámetros de SBR transmitidos en el flujo de bits descrito anteriormente. Aparte de la posibilidad descrita anteriormente del módulo de SBR, para mejorar la percepción de audio global de la señal de audio reconstruida, SBR ofrece además la posibilidad de codificar fuentes de ruido adicionales así como sinusoides individuales.

Por tanto, SBR representa una herramienta muy flexible para mejorar el compromiso entre calidad y tasa de transmisión de bits, lo que también hace que SBR sea un candidato interesante para aplicaciones en el campo de los sistemas de conferencia. Sin embargo, debido a la complejidad y al gran número de posibilidades y opciones, las señales de audio codificadas por SBR solamente se han mezclado hasta ahora en el dominio de tiempo decodificando completamente las señales de audio respectivas en señales de dominio de tiempo para efectuar el proceso de mezcla real en este dominio y, después de esto, recodificar la señal mezclada en una señal codificada por SBR. Aparte del retardo adicional introducido debido a la codificación de las señales en el dominio de tiempo, también la reconstrucción de la información espectral de la señal de audio codificada puede requerir una complejidad computacional significativa que puede, por ejemplo, no ser atractiva en el caso de aplicaciones portátiles u otras aplicaciones eficientes en cuanto a energía o eficientes en cuanto a complejidad computacional.

El documento US 2005/102137 da a conocer (las referencias entre paréntesis se aplican a este documento) un puente de conferencia en dominio comprimido, que permite a un usuario oír a una pluralidad de hablantes a través de un único canal. El puente comprende varios transcodificadores paralelos que conectan los canales de entrada a los canales de salida, y que convierte los parámetros de las normas de codificación de entrada en parámetros de las normas de codificación de salida y que mezcla tramas correspondientes a un primer canal de entrada con tramas correspondientes a un segundo canal de entrada en una trama correspondiente a un canal de salida. Los procesos de transcodificación usados para la conversión entre las diferentes normas de codificación son LPC, MELP y TDVC (véase la figura 2 y 12, y las secciones 37 a 41).

Por consiguiente, es un objeto de la presente invención reducir la complejidad computacional involucrada cuando se mezclan señales de audio codificadas por SBR.

Este objeto se resuelve mediante un aparato según la reivindicación 1 o 3, un método según la reivindicación 15 o un programa según la reivindicación 16.

5 Las realizaciones según la presente invención se basan en el descubrimiento de que la complejidad computacional puede reducirse efectuando la mezcla para una frecuencia por debajo de un mínimo de las frecuencias de cruce involucradas mezclando la información espectral en el dominio espectral, para una frecuencia por encima de una frecuencia de cruce máxima en el dominio de SBR, y para una frecuencia en una región entre el valor mínimo y el valor máximo estimando al menos un valor de SBR y generando un valor de SBR correspondiente basándose en el al menos un valor de SBR estimado o estimando un valor espectral o una información espectral basándose en los datos de SBR respectivos y generando un valor espectral de una información espectral basándose en este valor espectral estimado o información espectral.

10 En otras palabras, las realizaciones según la presente invención se basan en el descubrimiento de que para una frecuencia por encima de una frecuencia de cruce máxima, la mezcla puede efectuarse en el dominio de SBR, mientras que para una frecuencia por debajo de un mínimo de las frecuencias de cruce, la mezcla puede efectuarse en el dominio espectral procesando directamente valores espectrales correspondientes. Además, un aparato según una realización de la presente invención, puede, para una frecuencia intermedia entre el valor máximo y el valor mínimo, efectuar la mezcla en el dominio de SBR o en el dominio espectral estimando a partir de un valor de SBR correspondiente, un valor espectral, o estimando a partir de un valor espectral, un valor de SBR, y efectuar la mezcla real basándose en el valor estimado en el dominio de SBR o en el dominio espectral. En este contexto, ha de observarse que una frecuencia de cruce de salida puede ser cualquiera de las frecuencias de cruce de los flujos de datos de entrada u otro valor.

20 Como consecuencia, el número de etapas que deben efectuarse por un aparato y, por tanto, la complejidad computacional involucrada se reduce, puesto que la mezcla real por encima y por debajo de todas las frecuencias de cruce relevantes se efectúa basándose en una mezcla directa en los dominios respectivos, mientras que una estimación ha de efectuarse solamente en una región intermedia entre el valor mínimo de todas las frecuencias de cruce y un máximo de todas las frecuencias de cruce involucradas. Basándose en esta estimación, entonces se calcula o se determina el valor de SBR real o el valor espectral real. Por tanto, en muchos casos, incluso en esa región de frecuencia intermedia, la complejidad computacional se reduce puesto que normalmente no se requiere llevar a cabo una estimación y un procesamiento para todos los flujos de datos de entrada involucrados.

30 En realizaciones según una realización de la presente invención, la frecuencia de cruce de salida puede ser igual a una de las frecuencias de cruce de los flujos de datos de entrada o puede escogerse independientemente, por ejemplo, teniendo en cuenta el resultado de una estimación psicoacústica. Además, en realizaciones según la presente invención, los datos de SBR generados o los valores espectrales generados pueden aplicarse de manera diferente para suavizar o alterar los datos de SBR o valores espectrales en el intervalo de frecuencia intermedio.

Realizaciones según la presente invención se describirán a continuación en el presente documento haciendo referencia a las siguientes figuras.

La figura 1 muestra un diagrama de bloques de un sistema de conferencia;

35 la figura 2 muestra un diagrama de bloques del sistema de conferencia basándose en un códec de audio general;

la figura 3 muestra un diagrama de bloques de un sistema de conferencia que opera en un dominio de frecuencia utilizando la tecnología de mezcla de flujos de bits;

la figura 4 muestra un dibujo esquemático de flujo de datos que comprende una pluralidad de tramas;

la figura 5 ilustra formas diferentes de componentes espectrales y datos espectrales o información espectral;

40 la figura 6a muestra un diagrama de bloques simplificado de un aparato para mezclar una primera trama de un primer flujo de datos de entrada y una segunda trama de un segundo flujo de datos de entrada según una realización de la presente invención;

la figura 6b muestra un diagrama de bloques de una resolución de rejilla de tiempo/frecuencia de una trama de un flujo de datos;

45 la figura 7 muestra un diagrama de bloques más detallado de un aparato según una realización de la presente invención;

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

50 las figuras 9a y 9b muestran una primera trama y una segunda trama de un primer y un segundo flujo de datos de entrada tal como se proporcionan a un aparato según una realización de la presente invención, respectivamente;

la figura 9c muestra una situación de superposición en las tramas de entrada mostradas en las figuras 9a y 9b;

la figura 9d muestra una trama de salida, tal como se genera por un aparato según una realización de la presente invención con una frecuencia de cruce de salida que es la menor de las dos frecuencias de cruce de las tramas de entrada;

- 5 la figura 9e muestra una trama de salida, tal como se genera por un aparato según una realización de la presente invención con una frecuencia de cruce de salida que es la mayor de las frecuencias de cruce de las tramas de entrada y la figura 10 ilustra la correspondencia entre resoluciones de rejilla de baja y alta frecuencia.

10 Con respecto a las figuras 4 a 10, diferentes realizaciones según la presente invención se describirán en más detalle. Sin embargo, antes de describir estas realizaciones en más detalle, primero con respecto a las figuras 1 a 3, se dará una breve introducción en vista de los desafíos y demandas que pueden resultar importantes en el marco de los sistemas de conferencia.

La figura 1 muestra un diagrama de bloques de un sistema 100 de conferencia, que puede también denominarse como una unidad de control de multipunto (MCU). Como resultará evidente a partir de la descripción concerniente con su funcionalidad, el sistema 100 de conferencia, tal como se muestra en la figura 1, es un sistema que opera en el dominio de tiempo.

- 15 El sistema 100 de conferencia, tal como se muestra en la figura 1, está adaptado para recibir una pluralidad de flujos de datos de entrada a través de un número apropiado de entradas 110-1, 110-2, 110-3,... de las cuales en la figura 1 solamente se muestran tres. Cada una de las entradas 110 está acoplada a un decodificador 120 respectivo. Para ser más precisos, la entrada 110-1 para el primer flujo de datos de entrada está acoplada a un primer decodificador 120-1, mientras que la segunda entrada 110-2 está acoplada a un segundo decodificador 120-2 y la tercera entrada 110-3 está acoplada a un tercer decodificador 120-3.

El sistema 100 de conferencia además comprende un número apropiado de sumadores 130-1, 130-2, 130-3,... de los cuales una vez más se muestran tres en la figura 1. Cada uno de los sumadores está asociado con una de las entradas 110 del sistema 100 de conferencia. Por ejemplo, el primer sumador 130-1 está asociado con la primera entrada 100-1 y el decodificador 120-1 correspondiente.

- 25 Cada uno de los sumadores 130 está acoplado a las salida de todos los decodificadores 120, aparte del decodificador 120 al cual está acoplada la entrada 110. En otras palabras, el primer sumador 130-1 está acoplado a todos los decodificadores 120, aparte del primer decodificador 120-1. Por consiguiente, el segundo sumador 130-2 está acoplado a todos los decodificadores 120, aparte del segundo decodificador 120-2.

- 30 Cada uno de los sumadores 130 comprende además una salida que está acoplada, cada una, a un codificador 140. Por tanto, el primer sumador 130-1 está acoplado por la salida al primer decodificador 140-1. Por consiguiente, los sumadores 130-2, 130-3 segundo y tercero también están acoplados a los codificadores 140-2, 140-3 segundo y tercero, respectivamente.

- 35 A su vez, cada uno de los decodificadores 140 está acoplado a la salida 150 respectiva. En otras palabras, el primer codificador está, por ejemplo, acoplado a una primera salida 150-1. Los codificadores 140-2, 140-3 segundo y tercero también están acoplados a las salidas 150-2, 150-3 segunda y tercera, respectivamente.

Para poder describir el funcionamiento de un sistema 100 de conferencia tal como se muestra en la figura 1 en más detalle, la figura 1 también muestra un terminal 160 de conferencia de un primer participante. El terminal 160 de conferencia puede ser, por ejemplo, un teléfono digital (por ejemplo, un teléfono RDSI (RDSI = red digital de servicios integrados), un sistema que comprende una infraestructura de voz sobre IP o un terminal similar.

- 40 El terminal 160 de conferencia comprende un codificador 170 que está acoplado a la primera entrada 110-1 del sistema 100 de conferencia. El terminal 160 de conferencia también comprende un decodificador 180 que está acoplado a la primera salida 150-1 del sistema 100 de conferencia.

- 45 Terminales 160 de conferencia similares pueden también estar presentes en los emplazamientos de participantes adicionales. Estos terminales de conferencia no se muestran en la figura 1, meramente por motivos de simplicidad. También ha de observarse que no se requiere ni mucho menos que el sistema 100 de conferencia y los terminales 160 de conferencia estén físicamente presentes unos cerca de otros. Los terminales 160 de conferencia y el sistema 100 de conferencia pueden estar dispuestos en sitios diferentes, que pueden por ejemplo estar conectados solamente por medio de técnicas WAN (WAN = *wide area network*, red de área amplia).

- 50 Los terminales 160 de conferencia pueden comprender además o estar conectados a componentes adicionales tales como micrófonos, amplificadores y altavoces o auriculares para permitir un intercambio de señales de audio con un usuario humano de manera más comprensible. Estos no se muestran en la figura 1 por motivos de simplicidad solamente.

Como se indicó anteriormente, el sistema 100 de conferencia mostrado en la figura 1 es un sistema que opera en el dominio de tiempo. Cuando, por ejemplo, el primer participante habla en el micrófono (no mostrado en la figura 1), el

codificador 170 del terminal 160 de conferencia codifica la señal de audio respectiva en un flujo de bits correspondiente y transmite el flujo de bits a la primera entrada 110-1 del sistema 100 de conferencia.

5 Dentro del sistema 100 de conferencia, el flujo de bits se decodifica por el primer decodificador 120-1 y se transforma de regreso al dominio de tiempo. Puesto que el primer decodificador 120-1 está acoplado a los mezcladores 130-1, 130-3 segundo y tercero, la señal de audio, tal como se genera por el primer participante, puede mezclarse en el dominio de tiempo simplemente sumando la señal de audio reconstruida con señales de audio reconstruidas adicionales del segundo y tercer participante, respectivamente.

10 Esto es también cierto para las señales de audio proporcionadas por el segundo y tercer participante recibidas por las entradas 110-2, 110-3 segunda y tercera y procesadas por los decodificadores 120-2, 120-3 segundo y tercero, respectivamente. Estas señales de audio reconstruidas de los participantes segundo y tercero se proporcionan entonces al primer mezclador 130-1, que a su vez proporciona la señal de audio sumada en el dominio de tiempo al primer codificador 140-1. El codificador 140-1 recodifica la señal de audio sumada para formar un flujo de bits y lo proporciona en la primera salida 150-1 al terminal 160 de conferencia del primer participante.

15 De forma similar, también los codificadores 140-2, 140-3 segundo y tercero codifican las señales de audio sumadas en el dominio de tiempo recibidas de los sumadores 130-2, 130-3 segundo y tercero, respectivamente y transmiten los datos codificados de regreso a los participantes respectivos a través de las salidas 150-2, 150-3 segunda y tercera, respectivamente.

20 Para efectuar la mezcla real, las señales de audio se decodifican completamente y se suman en una forma sin comprimir. Después de esto, opcionalmente puede efectuarse un ajuste de nivel comprimiendo las señales de salida respectivas para impedir efectos de recorte (esto es, exceder un intervalo admisible de valores). El recorte puede aparecer cuando valores de muestra individuales se elevan por encima o caen por debajo de un intervalo admitido de valores de tal manera que los valores correspondientes se cortan (recortan). En el caso de una cuantificación de 16 bits, tal como se emplea por ejemplo en el caso de los CD, está disponible un intervalo de valores enteros entre -32768 y 32767 por valor de muestra.

25 Para contrarrestar una posible orientación por exceso o por defecto de la señal, se emplean algoritmos de compresión. Estos algoritmos limitan el desarrollo por encima o por debajo de un determinado valor umbral para mantener los valores de muestra dentro de un intervalo de valores admisible.

30 Cuando se codifican datos de audio en sistemas de conferencia tales como el sistema 100 de conferencia, tal como se muestra en la figura 1, se aceptan algunos inconvenientes con el fin de efectuar una mezcla en el estado sin codificar de la manera más fácilmente alcanzable. Además, las tasas de transmisión de datos de las señales de audio codificadas están adicionalmente limitadas a un intervalo menor de frecuencias transmitidas, puesto que un ancho de banda menor permite una frecuencia de muestreo más baja y por tanto, menos datos, según el teorema de muestreo de Nyquist-Shannon. El teorema de muestreo de Nyquist-Shannon afirma que la frecuencia de muestreo depende del ancho de banda de la señal muestreada y se requiere que sea (al menos) dos veces tan grande como el ancho de banda.

35 La International Telecommunication Union (ITU) y su área de normalización de las telecomunicaciones (ITU-T) han desarrollado diversas normas para sistemas de conferencia multimedia. H.320 es el protocolo de conferencia estándar para RDSI. H.323 define el sistema de conferencia estándar para una red basada en paquetes (TCP/IP). H.324 define sistemas de conferencia para redes de telefonía analógicas y sistemas de radiotelecomunicación.

40 Dentro de estas normas, no solamente se define la transmisión de las señales, sino también la codificación y el procesamiento de los datos de audio. La gestión de una conferencia la asumen uno o más servidores, las denominadas unidades de control multipunto (MCU) según la norma H.231. Las unidades de control multipunto también son responsables del procesamiento y la distribución de datos de vídeo y audio de los diversos participantes.

45 Para conseguir esto, la unidad de control multipunto envía a cada participante una señal resultante o de demás participantes y proporciona la señal a los participantes respectivos. La figura 1 no solamente muestra un diagrama de bloques de un sistema 100 de conferencia sino también un flujo de señal en tal situación de conferencia.

50 En el marco de las normas H.323 y H.320, los códecs de audio de la clase G.7xx se definen en cuanto a su funcionamiento en los respectivos sistemas de conferencia. La norma G.711 se usa para transmisiones de RDSI en sistemas telefónicos limitados a cable. A una frecuencia de muestreo de 8 KHz, la norma G.711 cubre un ancho de banda de audio de entre 300 y 3400 Hz, requiriendo una tasa de transmisión de bits de 64 Kbits/segundo a una profundidad (de cuantificación) de 8 bits. La codificación se forma mediante una codificación logarítmica simple denominada Ley  $\mu$  o Ley A que crea un retardo muy bajo de solamente 0,125 ms.

55 La norma G.722 codifica un ancho de banda de audio mayor desde 50 hasta 7000 Hz a una frecuencia de muestreo de 16 KHz. Como consecuencia, el códec consigue una mejor calidad cuando se compara con los códecs de audio G.7xx de banda más estrecha a tasas de transmisión de bits de 48, 56 ó 64 Kbit/s, con un retardo de 1,5 ms. Además, existen dos desarrollos adicionales, G.722.1 y G.722.2, que proporcionan calidad de habla comparable a tasas de transmisión de bits aún más bajas. G.722.2 permite una elección de tasa de transmisión de bits entre 6,6 Kbit/s y 23,85 Kbit/s con un retardo de 25 ms.

5 La norma G.729 se emplea normalmente en el caso de comunicación telefónica sobre IP, que también se denomina como comunicaciones de voz sobre IP (VoIP). El códec está optimizado para habla y transmite un conjunto de parámetros de habla analizados para una síntesis posterior junto con una señal de error. Como resultado, G.729 consigue una codificación significativamente mejor, de aproximadamente 8 Kbit/s a una velocidad de muestreo comparable y ancho de banda de audio comparable, cuando se compara con la norma G.711. El algoritmo más complejo, sin embargo, crea un retardo de aproximadamente 15 ms.

Como inconveniente, los códecs de G.7.xx están optimizados para comunicación de habla y muestran, aparte de un ancho de banda de frecuencia estrecha, problemas significativos cuando se codifica música junto con habla o música pura.

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

15 En otras palabras, emplear códecs para codificación y decodificación de señales de habla para procesar señales de audio generales, en las que se incluyen por ejemplo, señales de audio con música, no conduce a un resultado satisfactorio en cuanto a la calidad. Al emplear códecs de audio para codificación y decodificación de señales de audio generales en el marco del sistema 100 de conferencia tal como se muestra en la figura 1, la calidad puede mejorarse. Sin embargo, como se expondrá en el contexto de la figura 2 en más detalle, el empleo de códecs de audio general en un sistema de conferencia de este tipo puede conducir a efectos indeseables adicionales tales como un retardo incrementado por nombrar sólo uno.

20 Sin embargo, antes de describir la figura 2 en más detalle, ha de observarse que en la presente descripción, los objetos se indican mediante los mismos o similares símbolos de referencia 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 indique explícita o implícitamente de otra manera, los objetos indicados por los mismos o similares símbolos de referencia pueden implementarse de una manera similar o igual, por ejemplo, en cuanto a sus circuitos, programación, características u otros parámetros. Por tanto, los objetos que aparecen en varias realizaciones de las figuras y que se indican por los mismos o similares símbolos de referencia pueden implementarse con las mismas especificaciones, parámetros y características. Naturalmente, también pueden implementarse desviaciones y adaptaciones, por ejemplo, cuando condiciones de frontera u otros parámetros cambian de una figura a otra o de una realización a otra.

30 Además, en lo sucesivo se usarán símbolos de referencia resumen para indicar un grupo o clase de objetos, en lugar de un objeto individual. En el marco de la figura 1, esto ya se ha hecho, por ejemplo, cuando se indica la primera entrada como entrada 110-1, la segunda entrada como entrada 110-2 y la tercera entrada como entrada 110-3, mientras que las entradas se han explicado en términos del símbolo de referencia resumen 110 solamente. En otras palabras, a menos que se indique de otra manera explícitamente, las partes de la descripción que se refieren a objetos indicados con símbolos de referencia resumen también pueden referirse a otros objetos que llevan los correspondientes símbolos de referencia individuales.

35 Puesto que esto es también cierto para objetos indicados con los mismos o similares símbolos de referencia, ambas medidas ayudan a abreviar la descripción y a describir las realizaciones dadas a conocer en el presente documento de manera más clara y concisa.

40 La figura 2 muestra un diagrama de bloques de un sistema 100 de conferencia adicional junto con un terminal 160 de conferencia, que son ambos similares a los mostrados en la figura 1. El sistema 100 de conferencia mostrado en la figura 2 también comprende entradas 110, decodificadores 120, sumadores 130, codificadores 140 y salidas 150 que están interconectados igualmente en comparación con el sistema 100 de conferencia mostrado en la figura 1. El terminal 160 de conferencia mostrado en la figura 2 también comprende otra vez un codificador 170 y un decodificador 180. Por lo tanto, se hace referencia a la descripción del sistema 100 de conferencia mostrado en la figura 1.

45 Sin embargo, el sistema 100 de conferencia mostrado en la figura 2, así como el terminal 160 de conferencia mostrado en la figura 2 están adaptados para usar un códec de audio general (codificador-decodificador). Como consecuencia, cada uno de los codificadores 140, 170 comprende una conexión en serie de un convertidor 190 de tiempo/frecuencia acoplado antes de un cuantificador/codificador 200. El convertidor 190 de tiempo/frecuencia también se ilustra en la figura 2 como "T/F", mientras que los cuantificadores/codificadores 200 están marcados en la figura 2 con "Q/C".

50 Cada uno de los decodificadores 120, 180 comprende un decodificador/descuantificador 210, que se denomina en la figura 2 como "Q/C<sup>-1</sup>" conectado en serie con un convertidor 220 de frecuencia/tiempo, que se denomina en la figura 2 como "T/F<sup>-1</sup>". Por motivos de simplicidad solamente, el convertidor 190 de tiempo/frecuencia, el cuantificador/codificador 200 y el decodificador/descuantificador 210, así como el convertidor 220 de frecuencia/tiempo están marcados como tales solamente en el caso del codificador 140-3 y el decodificador 120-3. Sin embargo, la siguiente descripción también se refiere al resto de tales elementos.

Empezando con un codificador tal como los codificadores 140, o el codificador 170, la señal de audio proporcionada al convertidor 190 de tiempo/frecuencia se convierte del dominio de tiempo a un dominio de frecuencia o un dominio relacionado con la frecuencia mediante el convertidor 190. Después de eso, los datos de audio convertidos, en una

representación espectral generada por el convertidor 190 de tiempo/frecuencia, se cuantifican y codifican para formar un flujo de bits, que se proporciona entonces, por ejemplo, a las salidas 150 del sistema 100 de conferencia en el caso del codificador 140.

5 En cuanto a los decodificadores tales como los decodificadores 120 o el decodificador 180, el flujo de bits proporcionado a los decodificadores primero se decodifica y recuantifica para formar la representación espectral de al menos parte de una señal de audio, que se convierte entonces de regreso al dominio de tiempo mediante los convertidores 220 de frecuencia/tiempo.

10 Los convertidores 190 de tiempo/frecuencia, así como los elementos inversos, los convertidores 220 de frecuencia/tiempo, están adaptados por tanto para generar una representación espectral de al menos una pieza de una señal de audio proporcionada a los mismos y para volver a transformar la representación espectral en las partes correspondientes de la señal de audio en el dominio de tiempo, respectivamente.

15 En el proceso de convertir una señal de audio del dominio de tiempo al dominio de frecuencia y de regreso del dominio de frecuencia al dominio de tiempo, pueden producirse desviaciones de tal manera que la señal de audio restablecida, reconstruida o decodificada puede diferir de la señal de audio original o fuente. Pueden añadirse artefactos adicionales por las etapas adicionales de cuantificación y descuantificación efectuadas en el marco del cuantificador/codificador 200 y el decodificador 210. En otras palabras, la señal de audio original, así como la señal de audio restablecida, pueden diferir entre sí.

20 Los convertidores 190 de tiempo/frecuencia, así como los convertidores 220 de frecuencia/tiempo, pueden por ejemplo implementarse basándose en una MDCT (transformada de coseno discreta modificada), una MDST (transformada de seno discreta modificada), un convertidor basado en FFT (FFT = *Fast Fourier Transformation*, transformada de Fourier rápida) u otro convertidor basado en Fourier. La cuantificación y la recuantificación en el marco del cuantificador/codificador 200 y el decodificador/descuantificador 210 puede por ejemplo implementarse basándose en una cuantificación lineal, una cuantificación logarítmica u otro algoritmo de cuantificación más complejo, por ejemplo teniendo en cuenta más específicamente las características de audición del ser humano. Las partes del codificador y decodificador del cuantificador/codificador 200 y el decodificador/descuantificador 210, pueden por ejemplo funcionar empleando un esquema de codificación de Huffman o esquema de decodificación de Huffman.

25 Sin embargo, también pueden emplearse convertidores 190 de tiempo/frecuencia y 220 de frecuencia/tiempo más complejos, así como cuantificador/codificador 200 y decodificador/descuantificador 210 más complejos en diferentes realizaciones y sistemas según se describe en el presente documento, siendo parte de o formando, por ejemplo, un codificador de AAC-ELD como los codificadores 140, 170 y un decodificador de AAC-ELD como los decodificadores 120, 180.

30 No es necesario decir que podría ser recomendable implementar codificadores 170, 140 y decodificadores 180, 120 idénticos o al menos compatibles en el marco del sistema 100 de conferencia y los terminales 160 de conferencia.

35 El sistema 100 de conferencia, tal como se muestra en la figura 2, basándose en un esquema de codificación y decodificación de señal de audio general también efectúa la mezcla real de las señales de audio en el dominio de tiempo. Se proporcionan a los sumadores 130 las señales de audio reconstruidas en el dominio de tiempo para realizar una superposición y proporcionar las señales mezcladas en el dominio de tiempo a los convertidores 190 de tiempo/frecuencia de los siguientes codificadores 140. Por tanto, el sistema de conferencia un vez más comprende una conexión en serie de decodificadores 120 y codificadores 140, que es la razón por la cual un sistema 100 de conferencia tal como se muestra en las figuras 1 y 2, se denominan normalmente como "sistemas de codificación en tándem".

40 Los sistemas de codificación en tándem frecuentemente muestran el inconveniente de una alta complejidad. La complejidad de mezcla depende fuertemente de la complejidad de los decodificadores y codificadores empleados y puede multiplicarse significativamente en el caso de varias señales de entrada de audio y de salida de audio. Además, debido al hecho de que la mayoría de los esquemas de codificación y decodificación no son sin pérdidas, el esquema de codificación en tándem, según se emplea en los sistemas 100 de conferencia mostrados en las figuras 1 y 2, conduce normalmente a una influencia negativa sobre la calidad.

45 Como un inconveniente adicional, las etapas repetidas de decodificación y codificación también amplían el retardo global entre las entradas 110 y la salida 150 del sistema 100 de conferencia, lo que también se denomina retardo de extremo a extremo. Dependiendo del retardo inicial de los decodificadores y codificadores usados, el sistema 100 de conferencia en sí mismo puede incrementar el retardo hasta un nivel que hace que su uso en el marco del sistema de conferencia no sea atractivo, si no perturbador, o incluso imposible. Frecuentemente, se considera que un retardo de aproximadamente 50 ms es el retardo máximo que los participantes pueden aceptar en conversaciones.

50 Como fuentes principales para el retardo, los convertidores 190 de tiempo/frecuencia, así como los convertidores 220 de frecuencia/tiempo son responsables del retardo de extremo a extremo del sistema 100 de conferencia y el retardo adicional impuesto por los terminales 160 de conferencia. El retardo provocado por los elementos adicionales, concretamente los cuantificadores/codificadores 200 y los decodificadores/descuantificadores 210 es de menos importancia puesto que estos componentes pueden hacerse funcionar a una frecuencia mucho más alta en comparación con los convertidores 190 de tiempo/frecuencia y los convertidores 220 de frecuencia/tiempo. La mayoría

de los convertidores 190 de tiempo/frecuencia y convertidores 220 de frecuencia/tiempo se hacen funcionar por bloques o se hacen funcionar por tramas, lo que significa que en muchos casos se tiene que tener en cuenta un retardo mínimo tal como una cantidad de tiempo, que es igual al tiempo necesario para llenar una memoria intermedia o una memoria que tiene la longitud de trama de un bloque. Este tiempo, sin embargo, se ve influenciado significativamente por la frecuencia de muestreo que está normalmente en el intervalo de unos pocos kilohercios a unos pocos 10 kHz, mientras que la velocidad operacional de los cuantificadores/codificadores 200, así como el decodificador/descuantificador 210 se determina principalmente por la frecuencia de reloj del sistema subyacente. Esto es normalmente al menos 2, 3, 4 ó más órdenes de magnitud mayor.

Por tanto, en sistemas de conferencia que emplean códecs de señal de audio general, se ha introducido la denominada tecnología de mezcla de flujos de bits. El método de mezcla de flujos de bits puede por ejemplo implementarse basándose en el códec AAC-ELD de MPEG-4, que ofrece la posibilidad de evitar al menos algunas de las desventajas mencionadas anteriormente e introducidas por la codificación en tándem.

Ha de observarse sin embargo que, en principio, el sistema 100 de conferencia tal como se muestra en la figura 2, puede también implementarse basándose en el códec AAC-ELD de MPEG-4 con una tasa de transmisión de bits similar y un ancho de banda de frecuencia significativamente mayor, en comparación con los códecs basados en habla mencionados anteriormente de la familia de códecs G.7xx. Esto inmediatamente también implica que una calidad de audio significativamente mejor para todos los tipos de señal puede conseguirse a costa de una tasa de transmisión de bits significativamente incrementada. Aunque el AAC-ELD de MPEG-4 ofrece un retardo que está en el intervalo de aquél del códec G.7xx, la implementación del mismo en el marco de un sistema de conferencia tal como se muestra en la figura 2, puede no conducir a un sistema 100 de conferencia práctico. En lo sucesivo, con respecto a la figura 3, se expondrá un sistema más práctico basándose en la denominada mezcla de flujos de bits mencionada anteriormente.

Ha de observarse que meramente por motivos de simplicidad, la atención principalmente se pondrá en el códec AAC-ELD de MPEG-4 y sus flujos de datos y flujos de bits. Sin embargo, también otros codificadores y decodificadores pueden ser empleados en el entorno de un sistema 100 de conferencia tal como se ilustra y se muestra en la figura 3.

La figura 3 muestra un diagrama de bloques de un sistema 100 de conferencia que funciona según el principio de mezcla de flujos de bits junto con un terminal 160 de conferencia, tal como se describe en el contexto de la figura 2. El sistema 100 de conferencia es en sí mismo es una versión simplificada del sistema 100 de conferencia mostrado en la figura 2. Para ser más precisos, los decodificadores 120 del sistema 100 de conferencia en la figura 2 se han reemplazado por decodificadores/descuantificadores 220-1, 220-2, 210-3,... tal como se muestra en la figura 3. En otras palabras, los convertidores 120 de frecuencia/tiempo de los decodificadores 120 se han eliminado cuando se compara el sistema 100 de conferencia mostrado en las figuras 2 y 3. De forma similar, los codificadores 140 del sistema 100 de conferencia de la figura 2 se han reemplazado por cuantificadores/codificadores 200-1, 200-2, 200-3. Por tanto, los convertidores 190 de tiempo/frecuencia de los codificadores 140 se han eliminado cuando se compara el sistema 100 de conferencia mostrado en las figuras 2 y 3.

Como resultado, los sumadores 130 ya no funcionan en el dominio de tiempo, sino, debido a la ausencia de los convertidores 220 de frecuencia/tiempo y los convertidores 190 de tiempo/frecuencia, en el dominio de frecuencia o en un dominio relacionado con la frecuencia.

Por ejemplo, en el caso de los códec de AAC-ELD de MPEG-4, el convertidor 190 de tiempo/frecuencia y el convertidor 220 de frecuencia/tiempo, que están solamente presentes en los terminales 160 de conferencia, se basan en una transformada de MDCT. Por consiguiente, dentro del sistema 100 de conferencia, los mezcladores 130 directamente en las contribuciones de las señales de audio en la representación de frecuencia por MDCT.

Puesto que los convertidores 190, 220 representan la fuente principal de retardo en el caso del sistema 100 de conferencia mostrado en la figura 2, el retardo se reduce significativamente al eliminar estos convertidores 190, 220. Además, la complejidad introducida por los dos convertidores 190, 220 dentro del sistema 100 de conferencia también se reduce significativamente. Por ejemplo, en el caso de un decodificador AAC de MPEG-2, la transformada de MDCT inversa llevada a cabo en el marco del convertidor 220 de frecuencia/tiempo es responsable de aproximadamente el 20% de la complejidad global. Puesto que también el convertidor de MPEG-4 está basado en una transformada similar, una contribución no irrelevante a la complejidad global puede eliminarse al eliminar el convertidor 220 de frecuencia/tiempo por sí solo del sistema 100 de conferencia.

La mezcla de señales de audio en el dominio de MDCT u otro dominio de frecuencia es posible, puesto que en el caso de una transformada de MDCT o en el caso de una transformada basada en Fourier similar, estas transformaciones son transformaciones lineales. Las transformaciones, por consiguiente, tienen la propiedad aditiva matemática, es decir

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

y la propiedad homogénea matemática, es decir

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

donde  $f(x)$  es una función de transformada,  $x$  e  $y$  argumentos apropiados de las mismas y  $a$  una constante de valor real o de valor complejo.

5 Ambos elementos de la transformada de MDCT u otra transformada basada en Fourier permiten una mezcla en el dominio de frecuencia respectivo similar a la mezcla en el dominio de tiempo. Por tanto, todos los cálculos pueden llevarse a cabo igual de bien basándose en valores espectrales. No se requiere una transformación de los datos al dominio de tiempo.

10 En algunas circunstancias, una condición adicional podría tener que satisfacerse. Todos los datos espectrales relevantes deberían ser iguales con respecto a sus índices de tiempo durante el proceso de mezcla para todas las componentes espectrales relevantes. Éste puede eventualmente no ser el caso si, durante la transformación se emplea la denominada técnica de conmutación por bloques de tal manera que el codificador de los terminales 160 de conferencia puede conmutar libremente entre diferentes longitudes de bloque, dependiendo de ciertas condiciones. La conmutación por bloques puede poner en peligro la posibilidad de asignar de manera única valores espectrales individuales a muestras en el dominio de tiempo debido a la conmutación entre diferentes longitudes de bloque y correspondientes longitudes de ventana de MDCT, a menos que los datos que van a mezclarse hayan sido procesados con las mismas ventanas. Puesto que en un sistema general con terminales 160 de conferencia distribuidos, esto eventualmente puede no estar garantizado, podrían resultar necesarias interpolaciones complejas que a su vez pueden crear un retardo y una complejidad adicionales. Como consecuencia, puede eventualmente ser recomendable no implementar un proceso de mezcla de flujos de bits basándose en conmutación de longitudes de bloque.

20 En cambio, el códec de AAC-ELD está basado en una sola longitud de bloque y por consiguiente, puede garantizar más fácilmente que la asignación o sincronización descrita anteriormente de datos de frecuencia de tal manera que una mezcla puede realizarse más fácilmente. El sistema 100 de conferencia mostrado en la figura 3 es, en otras palabras, un sistema que puede efectuar la mezcla en el dominio de transformada o dominio de frecuencia.

25 Como se resumió anteriormente, con el fin de eliminar el retardo adicional introducido por los convertidores 190, 200 en el sistema 100 de conferencia mostrado en la figura 2, los códecs utilizados en los terminales 160 de conferencia utilizan una ventana de longitud y forma fijas. Esto permite la implementación del proceso de mezcla descrito directamente sin transformación del flujo de audio de regreso al dominio de tiempo. Este enfoque puede limitar la cantidad de retardo algorítmico introducido adicionalmente. Además, se disminuye la complejidad debido a la ausencia de las etapas de transformada inversa en el decodificador y las etapas de transformada directa en el codificador.

30 Sin embargo, también en el marco de un sistema 100 de conferencia como se muestran en la figura 3, puede hacerse necesario recuantificar los datos de audio después de la mezcla por los sumadores 130, lo que puede introducir ruido de cuantificación adicional. El ruido de cuantificación adicional puede crearse por ejemplo debido a las diferentes etapas de cuantificación de diferentes señales de audio proporcionadas al sistema 100 de conferencia. Como resultado, por ejemplo en el caso de transmisiones de muy bajas tasas de transmisión de bits en las cuales un número de etapas de cuantificación ya están limitadas, el proceso de mezclar dos señales de audio en el dominio de frecuencia o dominio de transformación puede dar como resultado una cantidad adicional indeseable de ruido u otras distorsiones en la señal generada.

Antes de describir una primera realización según la presente invención en la forma de un aparato para mezclar una pluralidad de flujos de datos de entrada, con respecto a la figura 4, se describirá brevemente un flujo de datos o flujo de bits junto con datos comprendidos en los mismos.

40 La figura 4 muestra esquemáticamente un flujo de bits o flujo 250 de datos que comprende al menos una o más frecuentemente, más de una trama 260 de datos de audio en un dominio espectral. Más precisamente, la figura 4 muestra tres tramas 260-1, 260-2 y 260-3 de datos de audio en un dominio espectral. Además, el flujo 250 de datos puede también comprender información o bloques 270 de información adicional, tal como valores de control que indican, por ejemplo, cómo se codifican los datos de audio, otros valores de control o información concerniente a los índices de tiempo u otros datos relevantes. Naturalmente, el flujo 250 de datos, como se muestra en la figura 4, puede comprender además tramas adicionales o una trama 260 puede comprender datos de audio de más de un canal. Por ejemplo, en el caso de una señal de audio estéreo, cada uno de las tramas 260 puede comprender por ejemplo datos de audio de un canal izquierdo, un canal derecho, datos de audio derivados de ambos, los canales izquierdo y derecho o cualquier combinación de los datos mencionados anteriormente.

50 Por tanto, la figura 4 ilustra que un flujo 250 de datos puede no solamente comprender una trama 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 con el protocolo (por ejemplo, sumas de control) o similares.

55 La figura 5 ilustra esquemáticamente información (espectral) concerniente a componentes espectrales tal y como están comprendidas por ejemplo en la trama 260 del flujo 250 de datos. Para ser más precisos, la figura 5 muestra un diagrama simplificado de información en un dominio espectral de un sólo canal de una trama 260. En el dominio espectral, una trama o datos de audio puede por ejemplo describirse en términos de sus valores de intensidad  $I$  como función de la frecuencia  $f$ . En sistemas discretos, tales como por ejemplo sistemas digitales, la resolución de frecuencia también es discreta, de tal manera que la información espectral está normalmente presente sólo para ciertas

componentes espectrales tales como frecuencias individuales o bandas estrechas o subbandas. Se denominan las frecuencias individuales o bandas estrechas, así como subbandas, como componentes espectrales.

5 La figura 5 muestra esquemáticamente una distribución de intensidad para seis frecuencias 300-1, ..., 300-6 individuales, así como una banda 310 o subbanda de frecuencia, que comprende, en el caso como se ilustra en la figura 5, cuatro frecuencias individuales. Tanto las frecuencias individuales o bandas 300 estrechas correspondientes, como la subbanda o banda 310 de frecuencia, forman componentes espectrales con respecto a los cuales la trama comprende información concerniente a los datos de audio en el dominio espectral.

10 La información concerniente a la subbanda 310 puede por ejemplo ser una intensidad global o un valor de intensidad promedio. Aparte de la intensidad u otros valores relacionados con la energía, tales como la amplitud, la energía de la componente espectral respectiva en sí mismo u otro valor derivado de la energía o la amplitud, la información de fase y otra información puede también estar comprendida en la trama y por tanto, considerarse como información concerniente a una componente espectral.

15 Los principios operacionales de una realización según la presente invención no son de tal manera que la mezcla se realice de manera directa en el sentido de que todos los flujos entrantes son decodificados, lo que incluye una transformación inversa al dominio de tiempo, la mezcla y otra vez recodificación de las señales.

20 Las realizaciones según la presente invención se basan en la mezcla efectuada en el dominio de frecuencia del códec respectivo. Un códec posible podría ser el códec de AAC-ELD o cualquier otro códec con una ventana de transformada uniforme. En tal caso, no es necesaria ninguna transformación de tiempo/frecuencia para poder mezclar los datos respectivos. Las realizaciones según una realización de la presente invención hacen uso del hecho de que el acceso a todos los parámetros de flujo de bits, tales como el tamaño de etapa de cuantificación y otros parámetros, es posible y que esos parámetros pueden utilizarse para generar un flujo de bits de salida mezclada.

25 Las realizaciones según una realización de la presente invención hacen uso del hecho de que la mezcla de líneas espectrales o información espectral concerniente 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 entre los mismos. Un valor de cero significa que se tratan a las fuentes como irrelevantes y no se usarán. Los grupos de líneas, tales como bandas o bandas de factor de escala, pueden usar el mismo factor de ponderación en el caso de realizaciones según la presente invención. Sin embargo, como se ilustró anteriormente, los factores de ponderación (por ejemplo, una distribución de ceros y unos) pueden variarse para las componentes espectrales de una sólo trama de un sólo flujo de datos de entrada. Además, las realizaciones según una realización de la presente invención hasta ahora no se requiere que usen exclusivamente los factores de ponderación cero o uno cuando se mezcla información espectral. Puede ser el caso que bajo algunas circunstancias, no para una sólo, una pluralidad de información espectral global de una trama de un flujo de datos de entrada, los factores de ponderación respectivos pueden ser diferentes de cero o uno.

35 Un caso particular es aquel que se establecen todas las bandas o la componente espectral de una fuente (flujo 510 de datos de entrada) en un factor de uno y se establecen todos los factores de las otras fuentes en cero. En este caso, se copia el flujo de bits de entrada completo de un participante idénticamente como un flujo de bits mezclado final. Pueden calcularse los factores de ponderación en una base de trama en trama, pero también pueden calcularse o determinarse basándose en grupos o secuencias más grandes de tramas. Naturalmente, incluído dentro de tal secuencia de trama o dentro de tramas individuales, pueden diferirse los factores de ponderación para diferentes componentes espectrales, como se resumió anteriormente. Los factores de ponderación pueden, en algunas realizaciones según una realización de la presente invención, calcularse o determinarse según los resultados del modelo psicoacústico.

45 Un modelo psicoacústico o un módulo respectivo puede calcular la proporción  $r(n)$  de energía entre una señal mezclada, en la que se incluyen solamente algunos flujos de entrada conduciendo a un valor  $E_f$  de energía y la señal mezclada completa que tiene un valor  $E_c$  de energía. Entonces se calcula la proporción  $r(n)$  de energía como 20 veces el logaritmo de  $E_f$  dividido por  $E_c$ .

50 Si la proporción es lo suficientemente alta, los canales menos dominantes pueden considerarse como enmascarados por los dominantes. Así, se procesa una reducción de irrelevancia lo que significa que solamente se incluyen aquellos flujos que no se notan, a los cuales se les atribuye un factor de ponderación de uno, mientras que se descartan todos los otros flujos – al menos una información espectral de una componente espectral. En otras palabras, a estos se les atribuye un factor de ponderación de cero.

Para ser más precisos, esto puede obtenerse por ejemplo según

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

55 y

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

5 y calcular la proporción según

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

10 en el que n es un índice de un flujo de datos de entrada y N es el número de todos los flujos de datos de entrada o los flujos de datos de entrada relevantes. Si la proporción r(n) es lo suficientemente alta, los canales menos dominantes o las tramas menos dominantes de los flujos 510 de datos de entrada pueden verse como enmascarados por los dominantes. Así, una reducción de irrelevancia puede procesarse, lo que significa que solamente se incluyen aquellas componentes espectrales de un flujo que se notan, mientras que se descartan los otros flujos.

15 Los valores de energía que van a considerarse en el marco de las ecuaciones (3) a (5) pueden, por ejemplo, derivarse de los valores de intensidad al calcular el cuadrado de los valores de intensidad respectivos. En el caso de información concerniente a las componentes espectrales que comprenden otros valores, puede llevarse a cabo un cálculo similar dependiendo de la forma de la información comprendida en la trama. Por ejemplo, en el caso de información de valor complejo, puede ser necesario llevarse a cabo el cálculo del módulo de las componentes real e imaginaria de los valores individuales que componen la información concerniente a las componentes espectrales.

20 Aparte de las frecuencias individuales, para la aplicación del módulo psicoacústico según las ecuaciones (3) a (5), pueden comprender las sumas en las ecuaciones (3) y (4) más de una frecuencia. En otras palabras, en las ecuaciones (3) y (4) los valores E<sub>n</sub> de energía respectivos pueden reemplazarse por un valor de energía global correspondiente a una pluralidad de frecuencias individuales, una energía de una banda de frecuencia o para expresarlo en términos más generales, por una sola pieza de información espectral o una pluralidad de información espectral concerniente a una o más componentes espectrales.

25 Por ejemplo, puesto que el AAC-ELD opera sobre líneas espectrales de manera de banda en banda, similar a los grupos de frecuencia en los cuales el sistema auditivo humano trata al mismo tiempo, la estimación de irrelevancia o el modelo psicoacústico puede llevarse a cabo de manera similar. Al aplicar el modelo psicoacústico de esta manera, es posible eliminar o sustituir parte de una señal de solamente una banda de frecuencia, si es necesario.

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

35 Por tanto, los valores r(n) según la ecuación (5) que son mayores de -28,5 dB pueden considerarse irrelevantes en términos de una evaluación psicoacústica o evaluación de irrelevancia basándose en la componente espectral o las componentes espectrales bajo consideración. Para diferentes componentes espectrales, pueden usarse valores diferentes. Así, el uso de umbrales como indicadores para una irrelevancia psicoacústica de un flujo de datos de entrada en términos de la trama bajo consideración de 10 dB a 40 dB, 20 dB a 30 dB o 25 dB a 30 dB puede considerarse útil.

40 Puede surgir una ventaja de que menos o ningún efecto de codificación en tándem se presenta debido a un número reducido de etapas de recuantificación. Puesto que cada etapa de cuantificación conlleva un peligro significativo de reducir ruido de cuantificación adicional, la calidad global de la señal de audio puede mejorarse al emplear una realización según la presente invención en forma de un aparato para mezclar una pluralidad de flujos de datos de entrada. Esto puede ser el caso cuando el flujo de datos de salida se genera de tal manera que se mantiene una distribución de niveles de cuantificación en comparación con una distribución de niveles de cuantificación de la trama del flujo de entrada determinada o partes de la misma.

45 La figura 6a muestra un diagrama de bloques simplificado de un aparato 500 para mezclar tramas de un primer flujo 510-1 de datos de entrada y un segundo flujo 510-2 de datos de entrada. El aparato 500 comprende una unidad 520 de procesamiento que está adaptada para generar un flujo 530 de datos de salida. Para ser ligeramente más precisos, el aparato 500 y la unidad 520 de procesamiento están adaptados para generar, basándose en una primera trama 540-1 y una segunda trama 540-2 de los flujos 510-1, 510-2 de datos de entrada primero y segundo, respectivamente, una trama 550 de salida comprendido en el flujo 530 de datos de salida.

50 Tanto la primera trama 540-1 como la segunda trama 540-2 cada una comprenden información espectral concerniente a una primera y segunda señal de audio, respectivamente. La información espectral se separa en una parte inferior de un espectro y en una parte superior del espectro respectivo, en el que la parte superior del espectro se describe por datos de SBR en términos de energía o valores relacionados con la energía en una resolución de rejilla de tiempo/frecuencia.

5 La parte inferior y la parte superior del espectro se separan entre sí en una denominada frecuencia de cruce, que es uno de los parámetros de SBR. Las partes inferiores del espectro se describen en términos de valores espectrales dentro de las tramas 540 respectivas. Esto se ilustra esquemáticamente en la figura 6a mediante una representación esquemática de la información espectral 560. La información espectral 560 se describirá en más detalle en el contexto de la figura 6b a continuación.

Naturalmente, puede ser recomendable implementar una realización según la presente invención en forma de un aparato 500, de tal manera que en el caso de una secuencia de tramas 540 en un flujo 510 de datos de entrada, solamente las tramas 540 se considerarán durante la comparación y determinación, que corresponde a un índice de tiempo similar o al mismo índice de tiempo.

10 La trama 550 de salida también comprende la representación 560 de información espectral similar, que también se muestra esquemáticamente en la figura 6a. Así, también la trama 550 de salida comprende una representación 560 de información espectral similar con una parte superior de un espectro de salida y una parte inferior de un espectro de salida que se tocan entre sí en la frecuencia de cruce de salida. Similar a las tramas 540 de los flujos 510 de datos de entrada, también se describe la parte inferior del espectro de salida de la trama 550 de salida en términos de valores espectrales de salida, mientras que se describe la parte superior del espectro (la parte más alta) en términos de datos de SBR que comprenden valores de energía en una resolución de rejilla de tiempo/frecuencia de salida.

15 Como se indicó anteriormente, la unidad 520 de procesamiento está adaptada para generar y emitir la trama de salida tal como se describió anteriormente. Ha de observarse que en casos generales, la primera frecuencia de cruce de la primera trama 540-1 y la segunda frecuencia de cruce de la segunda trama 540-2 son diferentes. Como consecuencia, la unidad de procesamiento está adaptada de tal manera que los datos espectrales de salida correspondientes a frecuencias debajo de un valor mínimo de una primera frecuencia de cruce, la segunda frecuencia de cruce y la frecuencia de cruce de salida se generan directamente en un dominio espectral basándose en primeros y segundos datos espectrales. Esto puede obtenerse por ejemplo al agregar o combinar linealmente la información espectral respectiva correspondiente a las mismas componentes espectrales.

20 Además, la unidad 520 de procesamiento está adaptada adicionalmente para generar los datos de SBR de salida que describen la parte superior del espectro de salida de la trama 550 de salida al procesar los respectivos primeros y segundos datos de SBR de las tramas 540-1, 540-2 primera y segunda directamente en el dominio de SBR. Esto se explicará en más detalle con respecto a las figuras 9a a 9e.

25 Como también se explicará en más detalles a continuación, la unidad 520 de procesamiento puede adaptarse de tal manera que para una región de frecuencia entre el valor mínimo y el valor máximo, tal como se definió anteriormente, al menos un valor de SBR de al menos uno de primeros y segundos datos espectrales se estima y un valor de SBR correspondiente de los datos de SBR de salida se genera basándose en al menos aquel valor de SBR estimado. Esto puede ser el caso, por ejemplo, cuando la frecuencia y la consideración de una componente espectral bajo consideración es más baja que la frecuencia de cruce máxima involucrada, pero más alta que el valor mínimo de la misma.

30 En tal situación, puede ocurrir que al menos una de las tramas 540 de entrada comprenda valores espectrales como parte de la parte inferior del espectro respectivo, mientras que la trama de salida espera datos de SBR, puesto que la componente espectral respectiva cae por encima de la frecuencia de cruce de salida. En otras palabras, en esta región de frecuencia intermedia entre el valor mínimo de las frecuencias de cruce involucradas y el valor máximo de los valores de frecuencia de cruce involucradas, puede ocurrir que basándose en datos espectrales de una parte inferior de uno de los espectros correspondientes datos de SBR tienen que estimarse. Los datos de SBR de salida correspondientes a la componente espectral bajo consideración entonces están basados en al menos los datos de SBR estimados. Una descripción más detallada de cómo esto puede llevarse a cabo según una realización de la presente invención se presentará en el contexto de las figuras 9a a 9e a continuación.

35 Por otra parte, puede ocurrir que para una componente espectral o una frecuencia involucrada, que caen en la región de frecuencia intermedia definida anteriormente, la trama 550 de salida espera valores espectrales, puesto que la componente espectral respectiva pertenece a la parte inferior del espectro de salida. Sin embargo, una de las tramas 540 de entrada puede comprender solamente datos de SBR para la componente espectral relevante. En este caso, puede ser recomendable estimar la información espectral correspondiente basándose en los datos de SBR y opcionalmente, basándose en la información espectral o al menos partes de la misma, de la parte inferior del espectro de la trama de entrada bajo consideración. En otras palabras, también una estimación de datos espectrales basándose en datos de SBR puede ser necesaria bajo algunas circunstancias. Basándose en el valor espectral estimado, el valor espectral correspondiente de la componente espectral respectiva entonces puede determinarse u obtenerse al procesar directamente el mismo en el dominio espectral.

40 Sin embargo, para facilitar un mejor entendimiento de los procesos y operaciones de un aparato 500 según una realización de la presente invención y SBR en general, la figura 6b muestra una representación 560 más detallada de información espectral que emplea datos de SBR.

Como se resumió en las partes introductorias de la memoria, la herramienta de SBR o módulo de SBR opera normalmente como un codificador o decodificador separado próximo a los codificadores o decodificadores de MPEG-4 básicos. La herramienta de SBR está basada en emplear un banco de filtros espejo en cuadratura (QMF) que también representa una transformación lineal.

5 La herramienta de SBR almacena, dentro del flujo de datos o flujo de bits del codificador de MPEG, sus propias informaciones y datos (parámetros de SBR) para facilitar la decodificación correcta de los datos de frecuencia descritos. Las informaciones se describirán en términos de la herramienta de SBR como rejilla de tramas o resolución de rejilla de tiempo/frecuencia. La rejilla de tiempo/frecuencia comprende datos sólo con respecto a la trama 540, 550 actual.

10 La figura 6b muestra esquemáticamente tal rejilla de tiempo/frecuencia para una sola trama 540, 550. Mientras la abscisa es un eje de tiempo, la ordenada es un eje de frecuencia.

15 El espectro mostrado en términos de su frecuencia  $f$  se separa, como se ilustró anteriormente, por la frecuencia ( $f_x$ ) 570 de cruce definida anteriormente en una parte 580 inferior y una parte 590 superior o más alta. Mientras que la parte 580 inferior del espectro se extiende normalmente desde la frecuencia accesible más baja (por ejemplo 0 Hz) hasta la frecuencia 570 de cruce, la parte 590 superior del espectro comienza en la frecuencia 570 de cruce y termina normalmente en dos veces la frecuencia ( $2f_x$ ) de cruce, como se indica en la figura 6b por la línea 600.

20 La parte 580 inferior del espectro se describe normalmente por datos 610 espectrales o valores espectrales como un área sombreada puesto que en muchos códecs basados en tramas y sus convertidores de tiempo/frecuencia, la trama respectiva de datos de audio se transfiere completamente al dominio de frecuencia, de tal manera que los datos 610 espectrales normalmente no comprenden una dependencia de tiempo interna de trama explícita. Como consecuencia, en términos de la parte 580 inferior del espectro, los datos 610 espectrales pueden no ser mostrados correctamente de manera completa en tal sistema de coordenadas de tiempo/frecuencia mostrado en la figura 6b.

25 Sin embargo, como se resumió anteriormente, la herramienta de SBR opera basándose en una conversión de tiempo/frecuencia de QMF que separa al menos la parte 590 superior del espectro en una pluralidad de subbandas, en la que cada una de las señales de subbanda comprende una dependencia de tiempo o resolución de tiempo. En otras palabras, la conversión al dominio de subbanda tal como se efectúa por la herramienta de SBR crea una "representación del tiempo y frecuencia mezclada".

30 Como se resumió en las partes introductorias de la memoria, basándose en la suposición de que la parte 590 superior del espectro conlleva un parecido significativo a la parte 580 inferior y afirma por tanto una correlación significativa, la herramienta de SBR puede derivar valores relacionados con la energía o de energía para describir, en términos de la frecuencia la manipulación de la amplitud de los datos espectrales de la parte 580 inferior del espectro copiados a las frecuencias en las componentes espectrales de la parte 590 superior. Por consiguiente, al copiar la información espectral de la parte 580 inferior a las frecuencias de la parte 590 superior y al modificar sus amplitudes respectivas, la parte 590 superior de los datos espectrales es replicada, tal como sugiere el nombre de la herramienta.

35 Mientras que la resolución en el tiempo de la parte 580 inferior del espectro está inherentemente presente, por ejemplo, al incluir información de fase u otros parámetros, la descripción de subbanda de la parte 590 superior de un espectro permite un acceso directo a la resolución en el tiempo.

40 La herramienta de SBR genera los parámetros de SBR que comprenden un número de ranuras de tiempo para cada trama de SBR, que es idéntico a las tramas 540, 550, en el caso de que las longitudes de trama de SBR y las longitudes de trama de codificador subyacentes sean compatibles y ni la herramienta de SBR ni el codificador o decodificador subyacente usan una técnica de reconmutación por bloques. Esta condición frontera es, por ejemplo, satisfecha por el códec de AAC-ELD de MPEG-4.

45 Las ranuras de tiempo dividen el acceso de tiempo de la trama 540, 550 del módulo de SBR en pequeñas regiones de tiempo espaciadas igualmente. El número de estas regiones de tiempo en cada trama de SBR se determina antes de la codificación de la trama respectiva. La herramienta de SBR usada en contexto con el códec de AAC-ELD de MPEG-4 se establece en 16 ranuras de tiempo.

Entonces se combinan estas ranuras de tiempo para formar uno o más envolventes. Un envoltente comprende al menos 2 o más ranuras de tiempo, formadas en un grupo. Cada uno de los envolventes tiene un número específico de datos de frecuencia de SBR con los cuales está asociado. En la rejilla de trama, el número y la longitud en términos de ranuras de tiempo se almacenarán con cada envoltente.

50 La representación simplificada de la información 560 espectral mostrada en la figura 6B muestra un primero y un segundo envoltente 620-1, 620-2. Aunque en principio, el envoltente 620 puede definirse libremente, incluso teniendo una longitud de menos de 2 ranuras de tiempo, en el marco del códec de AAC-ELD de MPEG-4, las tramas de SBR pertenecen a cualquiera de las dos clases, la clase FIXFIX y la clase LD\_TRAN solamente. Como consecuencia, aunque en principio es posible cualquier distribución de las ranuras de tiempo en términos de los envolventes, en lo siguiente se hará referencia principalmente al códec de AAC-ELD de MPEG-4 de tal manera que las implementaciones del mismo se describirán principalmente.

5 La clase FIXFIX divide las 16 ranuras de tiempo disponibles en un número de envolventes igualmente largos (por ejemplo 1, 2, 4, que comprenden 16, 6, 4 ranuras de tiempo cada uno, respectivamente), mientras que la clase LD\_TRAN comprende dos o tres envolventes de los cuales uno exactamente comprende dos ranuras. El envoltente que comprende exactamente dos ranuras de tiempo comprende un transitorio en la señal de audio o en otras palabras, el cambio abrupto de la señal de audio tal como un sonido muy fuerte y repentino. Las ranuras de tiempo antes y después de este transitorio pueden estar comprendidas en hasta dos envolventes adicionales siempre que los envolventes respectivos sean suficientemente largos.

10 En otras palabras, puesto que el módulo de SBR permite una división dinámica de las tramas en envolventes, es posible reaccionar ante transitorios en la señal de audio con una resolución de frecuencia más exacta. En el caso de que un transitorio esté presente en la trama actual, el codificador de SBR divide la trama en una estructura de envoltente apropiada. Como se resumió anteriormente, la división de trama está normalizada en el caso de AAC-ELD junto con SBR y depende de la posición del transitorio en términos de las ranuras de tiempo tal como se caracterizan por la variable TRANPOS.

15 La clase de trama de SBR escogida por el codificador de SBR en el caso de que un transitorio esté presente, la clase LD\_TRAN comprende normalmente tres envolventes. El envoltente inicial comprende el comienzo de trama hasta la posición del transitorio con índices de ranura de tiempo de cero a TRANPOS-1, el transitorio se encierra mediante un envoltente que comprende exactamente dos ranuras de tiempo con índices de ranura de tiempo de TRANPOS a TRANPOS+2. El tercer envoltente comprende todas las siguientes ranuras de tiempo con índices TRANPOS +3 a TRANPOS +16. Sin embargo, la longitud mínima de un envoltente en el códec de AAC-ELD junto con SBR se limita a dos ranuras de tiempo de tal manera que las tramas con un transitorio cercano a una frontera de trama sólo se dividirán en dos envolventes.

20 En la figura 6b se muestra una situación en la que los dos envolventes 620-1, 620-2 son igualmente largos pertenecientes a la clase de trama de SBR FIXFIX con un número de dos envolventes. Así, cada uno de los envolventes comprende una longitud de 8 ranuras de tiempo.

25 La resolución de frecuencia atribuida a cada uno de los envolventes determina el número de valores de energía o valores de energía de SBR a calcular para cada envoltente y a almacenar con respecto a la misma. La herramienta de SBR en contexto con el códec de AAC-ELD puede conmutarse entre una alta resolución y una baja resolución. En el caso de un envoltente de alta resolución, cuando se compara con un envoltente de baja resolución. Se usarán dos veces más los valores de energía para permitir una resolución de frecuencia más precisa para este envoltente en el caso de un envoltente de alta resolución cuando se compara con un envoltente de baja resolución. El número de valores de frecuencia para un envoltente de alta resolución o de baja resolución depende de los parámetros del codificador tales como la tasa de transmisión de bits, la frecuencia de muestreo y otros parámetros. En el caso del códec de AAC-ELD de MPEG-4, la herramienta de SBR muy frecuentemente usa de 16 a 14 valores en envolventes de alta resolución. Así, en envolventes de baja resolución, el número de valores de energía está frecuentemente en el intervalo de entre 7 y 8 por envoltente.

30 La figura 6b muestra para cada uno de los dos envolventes 620-1, 620-2, 6 regiones 630-1a, ..., 630-1f, 630-2a, ..., 630-2f de tiempo/frecuencia, representando cada una de las regiones de tiempo/frecuencia un valor de energía o de SBR relacionado con la energía. Meramente por motivos de simplicidad, tres de las regiones 630 de tiempo/frecuencia para cada uno de los dos envolventes 620-1, 620-2 se han indicado como tal. Además, por la misma razón, la distribución de frecuencia de la región 630 de tiempo/frecuencia para los dos envolventes 620-1, 620-2 se ha escogido de manera idéntica. Naturalmente, esto representa solamente una posibilidad entre un número significativo de posibilidades. Para ser más precisos, las regiones 630 de tiempo/frecuencia pueden distribuirse individualmente para cada uno de los envolventes 620. Por consiguiente, hasta ahora no se requiere ni mucho menos dividir el espectro o su parte 590 superior con la misma distribución cuando se cambia entre envolventes 620. Ha de observarse también que el número de regiones 630 de tiempo/frecuencia puede igualmente depender del envoltente 620 bajo consideración como se indicó anteriormente.

35 Además, como datos de SBR adicionales, valores de energía relacionados con el ruido y valores de energía relacionados con la senoide pueden también estar comprendidos en cada uno de los envolventes 620. Estos valores adicionales no se han mostrado meramente por motivos de simplicidad. Mientras que los valores relacionados con el ruido describen un valor de energía con respecto al valor de energía de la región 630 de tiempo/frecuencia respectiva de una fuente de ruido predefinida, los valores de energía de senoide se relacionan con oscilaciones de seno con frecuencias predefinidas y un valor de energía igual a aquel de la región de tiempo/frecuencia respectiva. Normalmente, dos a tres de los valores relacionados con el ruido o relacionados con la senoide pueden incluirse por envoltente 620. Sin embargo, también puede incluirse un número más pequeño o más grande.

40 La figura 7 muestra un diagrama de bloque más detallado adicional de un aparato 500 según una realización de la presente invención que está basada en la figura 6a. Por consiguiente, se hace referencia a la descripción de la figura 6a.

45 Como la discusión previa de información espectral y representación 560 en la figura 6b ha mostrado, podría ser recomendable para realizaciones según la presente invención analizar primero las rejillas de trama con el fin de generar

una nueva rejilla de trama para la trama 550 de salida. Como consecuencia, la unidad 520 de procesamiento comprende un analizador 640 al cual se proporcionan los dos flujos de datos de entrada 510-1, 510-2. La unidad 520 de procesamiento comprende además un mezclador 650 espectral, al cual se acoplan los flujos 510 de datos de entrada o las salidas del analizador 640. Además, la unidad 520 de procesamiento también comprende un mezclador 660 de SBR, que también está acoplado al flujo 510 de datos de entrada o a la salida del analizador 640. La unidad 520 de procesamiento comprende además un estimador 670, que también está acoplado a los dos flujos 510 de datos de entrada y/o el analizador 640 para recibir los datos analizados y/o los flujos de datos de entrada con las tramas 540 comprendidas en los mismos. Dependiendo de la implementación concreta, el estimador 670 puede estar acoplado a al menos uno de los mezcladores 650 espectrales o el mezclador 660 de SBR para proporcionar al menos uno de ellos con un valor de SBR estimado o un valor espectral estimado para frecuencias en la región intermedia definida anteriormente entre el valor máximo de las frecuencias de cruce involucradas y los valores mínimos de las mismas.

El mezclador 660 de SBR, así como el mezclador 650 espectral, está acoplado a un mezclador 680 que genera y emite el flujo 530 de datos de salida que comprende la trama 550 de salida.

Con respecto al modo de operación, el analizador 640 está adaptado para analizar las tramas 540 para determinar las rejillas de trama comprendidas en la misma y para generar una nueva rejilla de tramas que incluye, por ejemplo una frecuencia de cruce. Mientras que el mezclador 650 espectral está adaptado para mezclar en el dominio espectral, los valores espectrales o la información espectral de las tramas 540 para frecuencias o componentes espectrales debajo del mínimo de las frecuencias de cruce involucradas, el mezclador 660 de SBR está adaptado de forma similar para mezclar los respectivos datos de SBR en el dominio de SBR. El estimador 670 proporciona para la región de frecuencia intermedia entre los valores máximo y mínimo mencionados anteriormente de la misma, cualquiera de los dos mezcladores 650, 660, con datos apropiados en el dominio espectral o de SBR para permitir que estos mezcladores también operan en este dominio de frecuencia intermedia, si es necesario. El mezclador 680 entonces recopila los datos espectrales y de SBR recibidos de los dos mezcladores 650, 660 para formar y generar la trama 550 de salida.

Las realizaciones según la presente invención pueden por ejemplo emplearse en el marco de sistemas de conferencia, por ejemplo, un sistema de tele/vídeo conferencia con más de dos participantes. Tales sistemas de conferencia pueden ofrecer la ventaja de una complejidad menor en comparación con una mezcla de dominio de tiempo, puesto que las etapas de transformación de tiempo-frecuencia y etapas de recodificación pueden omitirse. Además, no se provoca ningún retardo adicional por estos componentes en comparación con la mezcla en el dominio de tiempo, debido a la ausencia del retardo del banco de filtros.

Sin embargo, las realizaciones según la presente invención pueden emplearse también en aplicaciones más complejas, que comprenden módulos tales como sustitución de ruido perceptivo (PNS), conformación de ruido temporal (TNS) y diferentes modos de codificación estéreo. Tal realización se describirá en más detalle con referencia a la figura 8.

La figura 8 muestra un diagrama de bloques esquemático de un aparato 500 para mezclar una pluralidad de flujos de datos de entrada que comprenden una unidad 520 de procesamiento. Para ser más precisos, la figura 8 muestra un aparato 500 altamente flexible que puede procesar señales de audio altamente diferentes codificadas en flujos de datos de entrada (flujos de bits). Algunos de los componentes que se describirán a continuación son por tanto componentes opcionales que no se requieren implementarse bajo todas las circunstancias y en el marco de todas las realizaciones según la presente invención.

La unidad 520 de procesamiento comprende un decodificador 700 de flujo de bits para cada uno de los flujos de datos de entrada o flujos de bits de audio codificados a procesarse por la unidad 520 de procesamiento. Meramente por motivos de simplicidad, la figura 8 muestra solamente dos decodificadores 700-1, 700-2 de flujo de bits. Naturalmente, dependiendo del número de flujos de datos de entrada a procesarse, puede implementarse un número más alto de decodificadores 700 de flujo de bits o un número más bajo, si por ejemplo un decodificador 700 de flujo de bits puede procesar secuencialmente más de uno de los flujos de datos de entrada.

El decodificador 700-1 de flujo de bits, así como los otros decodificadores 700-2, ... de flujo de bits, cada uno comprenden un lector 710 de flujo de bits que está adaptado para recibir y procesar las señales recibidas y aislar y extraer datos comprendidos en el flujo de bits. Por ejemplo, el lector 710 de flujo de bits puede estar adaptado para sincronizar los datos entrantes con un reloj interno y puede además estar adaptado para separar el flujo de bits entrante en las tramas apropiadas.

El decodificador 700 de flujo bits comprende además un decodificador 720 de Huffman acoplado a la salida del lector 710 de flujo de bits para recibir los datos aislados del lector 710 de flujo de bits. La salida del decodificador 720 de Huffman está acoplada a un descuantificador 730, que también se denomina como cuantificador inverso. El descuantificador 730 que está acoplado detrás del decodificador 720 de Huffman es seguido de un escalador 740. El decodificador 720 de Huffman, el descuantificador 730 y el escalador 740 forman una primera unidad 750 en cuya salida al menos parte de la señal de audio del flujo de datos de entrada respectivo está disponible en el dominio de frecuencia o el dominio relacionado con la frecuencia en el que el codificador del participante (no mostrado en la figura 8) opera.

El decodificador 700 de flujos de bits comprende además una segunda unidad 760 que está acoplada de dato en dato después de la primera unidad 750. La segunda unidad 760 comprende un decodificador estéreo (módulo M/S) detrás

del cual un decodificador de PNS está acoplado. El decodificador 780 de PNS es seguido de dato en dato por un decodificador 790 de TNS, que junto con el decodificador 780 de PNS en el decodificador 770 estéreo forman la segunda unidad 760.

- 5 Aparte del flujo descrito de datos de audio, el decodificador 700 de flujo de bits comprende además una pluralidad de conexiones entre diferentes módulos concernientes a datos de control. Para ser más precisos, el lector 710 de flujo de bits también está acoplado al decodificador 720 de Huffman para recibir datos de control apropiados. Además, el decodificador 720 de Huffman está acoplado directamente al escalador 740 para transmitir información de escalamiento al escalador 740. El decodificador 770 estéreo, el decodificador 780 de PNS y el decodificador 790 de TNS están también cada uno acoplados al lector 710 de flujo de bits para recibir datos de control apropiados.
- 10 La unidad 520 de procesamiento comprende además una unidad 800 mezcladora que a su vez comprende un mezclador 810 espectral que está acoplado de entrada en entrada a los decodificadores 700 de flujo de bits. El mezclador 810 espectral puede comprender por ejemplo, uno o más sumadores para efectuar la mezcla real en el dominio de frecuencia. Además, el mezclador 810 espectral puede comprender además multiplicadores para permitir una combinación lineal arbitraria de la información espectral proporcionada por los decodificadores 700 de flujo de bits.
- 15 La unidad 800 mezcladora comprende además un módulo 820 de optimización que está acoplado de dato en dato a una salida del mezclador 810 espectral. El módulo 820 de optimización está, sin embargo, también acoplado al mezclador 810 espectral para proporcionar al mezclador 810 espectral información de control. De dato en dato, el módulo 820 de optimización representa una salida de la unidad 800 mezcladora.
- 20 La unidad 800 mezcladora comprende además un mezclador 830 de SBR que está acoplado directamente a una salida del lector 710 de flujo de bits de los diferentes decodificadores 700 de flujo de bits. Una salida del mezclador 830 de SBR forma otra salida de la unidad 800 mezcladora.
- 25 La unidad 520 de procesamiento comprende además un codificador 850 de flujo de bits que está acoplado a la unidad 800 mezcladora. El codificador 850 de flujo de bits comprende una tercera unidad 860 que comprende un codificador 870 de TNS, un codificador 880 de PNS y un codificador 890 estéreo, que están acoplados en serie en el orden descrito. La tercera unidad 860, por tanto, forma una unidad inversa de la primera unidad 750 del decodificador 700 de flujo de bits.
- 30 El codificador 850 de flujo de bits comprende además una cuarta unidad 900 que comprende un escalador 910, un cuantificador 920 y un codificador 930 de Huffman que forman una conexión en serie entre una entrada de la cuarta unidad y una salida de la misma. La cuarta unidad 900, por tanto, forma un módulo inverso de la primera unidad 750. Así, el escalador 910 está también acoplado directamente al codificador 930 de Huffman para proporcionar al codificador 930 de Huffman datos de control respectivos.
- 35 El codificador 850 de flujo de bits también comprende un escritor 940 de flujo de bits que está acoplado a la salida del codificador 930 de Huffman. Además, el escritor 940 de flujo de bits está también acoplado al codificador 870 de TNS, al codificador 880 de PNS, al codificador 890 estéreo y al codificador 930 de Huffman para recibir datos de control e información de estos módulos. Una salida del escritor 940 de flujo de bits forma una salida de la unidad 520 de procesamiento y del aparato 500.
- 40 El codificador 850 de flujo de bits también comprende un módulo 950 psicoacústico que está también acoplado a la salida de la unidad 800 mezcladora. El codificador 850 de flujo de bits está adaptado para proporcionar los módulos de la tercera unidad 860 con información de control apropiada que indica, por ejemplo, cuáles pueden emplearse para codificar las salidas de la señal de audio por la unidad 800 mezcladora en el marco de las unidades de la tercera unidad 860.
- 45 En principio, en las salidas de la segunda unidad 760 hasta la entrada de a tercera unidad 860, un procesamiento de la señal de audio en el dominio espectral, tal como se definió por el codificador usado en el lado del remitente, es por tanto posible. Sin embargo, como se indicó anteriormente, una decodificación, descuantificación desescalamiento completos y etapas de procesamiento adicionales pueden eventualmente no ser necesarias si, por ejemplo, la información espectral de una trama de una de los flujos de datos de entrada es dominante. Según una realización de la presente invención, al menos parte de la información espectral de las componentes espectrales respectivas entonces se copia a la componente espectral de la trama respectiva del flujo de datos de salida.
- 50 Para permitir tal procesamiento, el aparato 500 y la unidad 520 de procesamiento comprenden líneas de señal adicionales para un intercambio de datos optimizados. Para permitir tal procesamiento en la realización mostrada en la figura 8, una salida del decodificador 720 de Huffman, así como salidas del escalador 740, del decodificador estéreo 760 y del decodificador 780 de PNS, están, junto con los componentes respectivos de otros lectores de flujo de bits 710, acopladas al módulo 820 de optimización de la unidad 800 mezcladora para un procesamiento respectivo.
- 55 Para facilitar, después de un procesamiento respectivo, un flujo de datos correspondiente dentro del codificador 850 de flujo de bits, también se implementan líneas de datos correspondientes para un flujo de datos optimizado. Para ser más precisos, una salida del módulo 820 de optimización está acoplada a una entrada del codificador 780 de PNS, el codificador 890 estéreo, una salida de la cuarta unidad 900 y el escalador 910, así como una entrada al codificador 930

de Huffman. Además, la salida del módulo 820 de optimización está también acoplada directamente al escritor 940 de flujo de bits.

Como se indicó anteriormente, casi todos los módulos tal como se describieron anteriormente son módulos opcionales, que no se requieren implementarse en realizaciones según la presente invención. Por ejemplo, en el caso de los flujos de datos de audio que comprenden sólo un único canal, las unidades 770, 890 de codificación y de decodificación estéreo, pueden omitirse. Así, en el caso de que ninguna señal basada en PNS vaya a procesarse, el decodificador de PNS y codificador de PNS correspondientes 780, 880, pueden también omitirse. Los módulos 790, 870 de TNS pueden también omitirse en el caso de que la señal a procesar y la señal a omitir no estén basadas en datos de TNS. Dentro de las unidades 750, 900 primera y cuarta, el cuantificador 730 inverso, el mezclador 740, el cuantificador 920, así como el escalador 910 pueden eventualmente también omitirse. Por consiguiente, estos módulos también se considerarán componentes opcionales.

El decodificador 720 de Huffman y el codificador 930 de Huffman pueden implementarse de manera diferente, utilizando otro algoritmo u omitirse completamente.

Con respecto al modo de operación del aparato 500 junto con la unidad 520 de procesamiento comprendida en el mismo, primero se lee un flujo de datos de entrada entrante y se separa en informaciones separadas por el lector 710 de flujo de bits. Después de la decodificación de Huffman, la información espectral resultante puede eventualmente recuantificarse por el descuantificador 730 y escalarse de manera apropiada por el desescalador 740.

Después, dependiendo de la información de control comprendida en el flujo de datos de entrada, la señal de audio codificada en el flujo de datos de entrada puede descomponerse en señales de audio para dos o más canales en el marco del decodificador 770 estéreo. Por ejemplo, si la señal de audio comprende un canal central (M) y un canal lateral (S), los datos del canal izquierdo y del canal derecho correspondientes pueden obtenerse al sumar y restar entre sí los datos del canal central y del canal lateral. En muchas implementaciones, el canal central es proporcional a la suma de los datos de audio del canal izquierdo y del canal derecho, mientras que el canal lateral es proporcional a la diferencia entre el canal izquierdo (L) y el canal derecho (R). Dependiendo de la implementación, los canales a los que se hace referencia pueden sumarse y/o restarse teniendo en cuenta un factor de 1/2 para impedir efectos de recorte. En general, los diferentes canales pueden procesarse mediante combinaciones lineales para producir los canales correspondientes.

En otras palabras, después de que el decodificador 770 estéreo, los datos de audio pueden, si es apropiado, descomponerse en dos canales individuales. Naturalmente, también una decodificación inversa puede efectuarse por el decodificador 770 estéreo. Por ejemplo, si la señal de audio tal como es recibida por el lector 710 de flujo de bits comprende un canal izquierdo y un canal derecho, el decodificador 770 estéreo puede igualmente bien calcular o determinar datos de canal medio y de canal lateral apropiados.

Dependiendo de la implementación no solamente del aparato 500, sino también dependiendo de la implementación del codificador del participante que proporciona el flujo de datos de entrada respectivo, el flujo de datos respectivo puede comprender parámetros de PNS (PNS = sustitución de ruido perceptivo). La PNS se basa en el hecho de que es más probable que el oído humano no puede distinguir sonidos de tipo ruido en un intervalo de frecuencia limitado o en una componente espectral tal como una banda o una frecuencia individual de un ruido generado sintéticamente. Por consiguiente, PNS sustituye la contribución de tipo ruido real de la señal de audio con un valor de energía que indica un nivel de ruido a introducirse sintéticamente en la componente espectral respectiva y que omite la señal de audio real. En otras palabras, el decodificador 780 de PNS puede regenerar en una o más componentes espectrales la contribución de señal de audio de tipo ruido real basándose en un parámetro de PNS comprendido en el flujo de datos de entrada.

En términos del decodificador 790 de TNS y el codificador 870 de TNS, es posible que sea necesario retransformar señales de audio respectivas en una versión sin modificar con respecto a un módulo de TNS que opera en el lado del remitente. La conformación de ruido temporal (TNS) es un medio para reducir artefactos de preeco provocados por el ruido de cuantificación, que puede estar presente en el caso de una señal de tipo transitorio en una trama de la señal de audio. Para contrarrestar este transitorio, al menos un filtro de predicción adaptable se aplica a la información espectral comenzando desde el lado inferior del espectro, el lado superior del espectro o ambos lados del espectro. Las longitudes de los filtros de predicción así como los intervalos de frecuencia a los cuales se aplican los filtros respectivos pueden estar adaptadas.

En otras palabras, la operación de un módulo de TNS está basada en el cálculo de uno o más filtros de IIR adaptables (IIR = respuesta de impulsos infinita) y la codificación y transmisión de una señal de error que describe la diferencia entre la señal de audio predicha y la real junto con los coeficientes de filtro en los filtros de predicción. Como consecuencia, puede ser posible incrementar la calidad de audio mientras que se mantiene la tasa de transmisión de bits del flujo de datos del transmisor al hacer frente a las señales de tipo transitorio mediante la aplicación de un filtro de predicción en el dominio de frecuencia para reducir la amplitud de la señal de error restante, que podría entonces codificarse utilizando menos etapas de cuantificación en comparación con la codificación directa de la señal de audio de tipo transitorio con un ruido de cuantificación similar.

En términos de una aplicación de TNS, puede ser recomendable bajo algunas circunstancias emplear la función del decodificador 760 de TNS para decodificar la parte TNS del flujo de datos de entrada para llegar a una representación

5 “pura” en el dominio espectral determinada por el códec utilizado. Esta aplicación de la funcionalidad de los decodificadores 790 de TNS puede ser útil si una estimación del modelo psicoacústico (por ejemplo, aplicado en el módulo 950 psicoacústico) no puede ya estimarse basándose en los coeficientes de filtro de los filtros de predicción comprendidos en los parámetros de PNS. Esto puede especialmente ser importante en el caso cuando al menos un flujo de datos de entrada usa TNS, mientras que otro no.

10 Cuando la unidad de procesamiento determina que, basándose en la comparación de las tramas de los flujos de datos de entrada, la información espectral de una trama de un flujo de datos de entrada que utiliza TNS debe usarse, los parámetros de TNS pueden usarse para la trama de los datos de salida. Si, por ejemplo, por razones de incompatibilidad, el receptor del flujo de datos de salida no puede decodificar los datos de TNS, podría ser útil no copiar los datos espectrales respectivos de la señal de error y los parámetros de TNS adicionales, sino procesar los datos reconstruidos de los datos relacionados con TNS para obtener la información en el dominio espectral y no usar el codificador 870 de TNS. Una vez más, esto ilustra que partes de los componentes o módulos mostrados en la figura 8 no no se requieren implementarse en diferentes realizaciones según la presente invención.

15 En el caso de al menos un flujo de entrada de audio que compara datos de PNS, una estrategia similar puede aplicarse. Si en la comparación de las tramas para una componente espectral del flujo de datos de entrada se muestra que un flujo de datos de entrada es, en términos de su trama actual y la componente espectral respectiva o las componentes espectrales dominantes, los parámetros de PNS respectivos (esto es, los valores de energía respectivos) pueden también copiarse directamente a la componente espectral respectiva de la trama de salida. Sin embargo, si el receptor no puede aceptar los parámetros de PNS, la información espectral puede reconstruirse a partir del parámetro de PNS para las componentes espectrales respectivas al generar ruido con el nivel de energía apropiado tal como se indica por el valor de energía respectivo. Entonces, los datos de ruido pueden así procesarse en el dominio espectral.

20 Como se resumió anteriormente, los datos transmitidos también comprenden datos de SBR, que entonces son procesados por el mezclador 830 de SBR que efectúa la funcionalidad descrita anteriormente.

25 Puesto que SBR permite dos canales estéreo de codificación, la codificación del canal izquierdo y del canal derecho por separado, así como la codificación de los mismos en términos de un canal (C) de acoplamiento, según una realización de la presente invención, el procesamiento de los parámetros de SBR respectivos o al menos partes de los mismos, puede comprender copiar los elementos C de los parámetros de SBR a ambos de los elementos izquierdo y derecho del parámetro de SBR a determinarse y transmitirse o viceversa.

30 Además, puesto que en diferentes realizaciones según una realización de la presente invención, los flujos de datos de entrada pueden comprender ambas señales de audio mono y estéreo que comprenden uno y dos canales individuales, respectivamente, una mezcla ascendente mono a estéreo o una mezcla descendente estéreo a mono puede adicionalmente efectuarse en el marco del procesamiento de las tramas de los flujos de datos de entrada y de la generación de la trama de salida del flujo de datos de salida.

35 Como se ha mostrado en la descripción precedente, en términos de parámetros de TNS puede ser recomendable procesar los parámetros de TNS respectivos junto con la información espectral de toda la trama desde el flujo de datos de entrada dominante hasta el flujo de datos de salida para impedir una recuantificación.

40 En el caso de una información espectral basada en PNS, el procesamiento de valores de energía individuales sin decodificación de las componentes espectrales subyacentes puede ser una manera viable. Además, en este caso al procesar solamente el parámetro de PNS respectivo desde un componente espectral dominante de las tramas de las pluralidades de flujos de datos de entrada hasta la componente espectral correspondiente de la trama de salida del flujo de datos de salida ocurre sin introducir ruido de cuantificación adicional.

45 Como se resumio anteriormente, una realización según la presente invención puede también comprender simplemente copiar una información espectral concerniente a una componente espectral después de comparar las tramas de la pluralidad de flujos de datos de entrada y después determinar, basándose en la comparación, para una componente espectral de una trama de salida del flujo de datos de salida exactamente un flujo de datos a ser la fuente de la información espectral.

50 El algoritmo de reemplazo efectuado en el marco del módulo 950 psicoacústico examina cada pieza de la información espectral concerniente a las componentes espectrales subyacentes (por ejemplo, bandas de frecuencia) de la señal resultante para identificar componentes espectrales con solamente una componente activa. Para estas bandas, los valores cuantificados del flujo de datos de entrada respectivo del flujo de bits de entrada pueden copiarse del codificador sin recodificación o recuantificación de los datos espectrales respectivos para la componente espectral específica. Bajo algunas circunstancias, todos los datos cuantificados pueden tomarse de una sola señal de entrada activa para formar el flujo de bits de salida o flujo de datos de salida, de tal manera que – en términos del aparato 500- puede obtenerse una codificación sin pérdidas del flujo de datos de entrada.

55 Además, puede hacerse posible omitir unas etapas de procesamiento, como el análisis psicoacústico dentro del codificador. Esto permite el acortamiento del proceso de codificación y mediante esto, una reducción de la complejidad de cálculo puesto que, en principio, sólo es necesario copiar los datos de un flujo de bits a otro flujo de bits bajo las ciertas circunstancias.

Por ejemplo, en el caso de PNS, puede llevarse a cabo un reemplazo puesto que los factores de ruido de la banda PNS-codificada pueden copiarse de uno de los flujos de datos de salida al flujo de datos de salida. El reemplazo de componentes espectrales individuales con parámetros de PNS apropiados es posible puesto que los parámetros de PNS son específicos de la componente espectral o en otras palabras, a una aproximación muy buena independientes entre sí.

Sin embargo, puede ocurrir que una aplicación demasiado agresiva del algoritmo descrito puede producir una experiencia de escucha degradada o una reducción de calidad indeseable. Por tanto, puede ser recomendable limitar el reemplazo a tramas individuales, en lugar de información espectral, concerniente a componentes espectrales individuales. En tal modo de operación, la estimación de irrelevancia o determinación de irrelevancia, así como el análisis de reemplazo, pueden llevarse a cabo sin cambios. Sin embargo, un reemplazo puede, en este modo de operación, solamente llevarse a cabo cuando todas o al menos un número significativo de las componentes espectrales dentro de la trama activa son reemplazables.

Aunque esto podría conducir a un número menor de reemplazos, una intensidad interna de la información espectral puede en algunas situaciones mejorarse, conduciendo a una calidad aun ligeramente mejorada.

Volviendo a la mezcla de SBR según una realización de la presente invención, al dejar fuera componentes adicionales y opcionales del aparato 500 mostrado en la figura 8, los principios de operación de SBR y mezclas de los datos de SBR se describirán en más detalle a continuación.

Como se resumió anteriormente, la herramienta de SBR utiliza un QMF (banco de filtros espejo en cuadratura) que representa una transformación lineal. Como consecuencia, no sólo es posible procesar los datos 610 espectrales (véanse la figura 6b) directamente en el dominio espectral, sino también procesar los valores de energía asociados con cada una de las regiones 630 de tiempo/frecuencia en la parte 590 superior del espectro (véanse la figura 6b). Sin embargo, como se indicó anteriormente, podría ser recomendable, y en algunos casos incluso necesario, ajustar primero las rejillas de tiempo/frecuencia involucradas antes de la mezcla.

Aunque en principio es posible generar una rejilla de tiempo/frecuencia completamente nueva, en lo siguiente se describirá una situación en la que una rejilla de tiempo/frecuencia que ocurre en una fuente se usará como la rejilla de tiempo/frecuencia de la trama 550 de salida. La decisión de cuál de las rejillas de tiempo/frecuencia puede usarse puede estar basada, por ejemplo, en una consideración psicoacústica. Por ejemplo, cuando una de las rejillas comprende transitorios, podría ser recomendable utilizar una rejilla de tiempo/frecuencia que comprende este transitorio o que es compatible con este transitorio, puesto que debido a los efectos de enmascaramiento del sistema auditivo humano, artefactos audibles pueden eventualmente introducirse cuando se desvía de esta rejilla específica. En el caso de que, por ejemplo, dos o más tramas con transitorios vayan a procesarse por el aparato 500 según una realización de la presente invención, puede ser recomendable escoger la rejilla de tiempo/frecuencia compatible con el más anterior de estos transitorios. Una vez más, debido a los efectos de enmascaramiento, la elección para la rejilla que contiene el ataque más anterior puede ser, basada en consideraciones psicoacústicas, una elección preferente.

Sin embargo, debe indicarse que aún bajo estas circunstancias, otras rejillas de tiempo/frecuencia pueden también calcularse o puede escogerse una diferente.

Cuando se mezclan las rejillas de trama de SBR, es por consiguiente en algunos casos recomendable analizar y determinar la presencia y posición de uno o más transitorios comprendidos en las tramas 540. Adicional o alternativamente, esto también puede lograrse al evaluar las rejillas de tramas de los datos de SBR de una trama 540 respectiva y verificar si las rejillas de tramas por sí mismos son compatibles con o si indican la presencia de un transitorio respectivo. Por ejemplo, el uso de la clase de trama LD\_TRAN, en el caso del códec de AAC ELD puede indicar que un transitorio está presente. Puesto que esta clase también comprende la variable TRANSPOSE, también la posición del transitorio en términos de las ranuras de tiempo es conocida por el analizador 640, tal como se muestra en la figura 7.

Sin embargo, puesto que la otra clase de trama de SBR FIXFIX puede usarse, pueden ocurrir diferentes constelaciones cuando se genera la rejilla de tiempo/frecuencia de la trama 550 de salida.

Por ejemplo, las tramas sin transitorios o con posiciones de transitorios iguales pueden ocurrir. Si las tramas no comprenden transitorios, puede incluso ser posible usar una estructura envolvente con un solo envolvente que solamente expande toda la trama. También, en el caso de que el número de envoltentes sea idéntico, la estructura de trama básica puede copiarse. En el caso de que el número de envoltentes comprendido en una trama sea un número entero de aquel de la otra trama, la distribución de envolvente más fina también puede usarse.

De forma similar, cuando todas las tramas 540 comprenden transitorios en la misma posición, la rejilla de tiempo/frecuencia puede copiarse ya sea de una u otra de las dos rejillas.

Cuando se mezclan tramas sin transitorios con un solo envolvente y una trama con un transitorio, la estructura de trama de la trama que comprende transitorio puede copiarse. En este caso, se puede suponer de manera segura que ningún nuevo transitorio se tendrá como resultado cuando se mezclan los datos respectivos. Es más probable que sólo el transitorio ya presente podría amplificarse o amortiguarse.

En el caso de que tramas con diferentes posiciones de transitorios estén involucradas, cada una de las tramas comprende un transitorio en diferentes posiciones con respecto a las ranuras de tiempo subyacentes. En este caso, una distribución adecuada basándose en las posiciones de transitorios es deseable. En muchas situaciones, la posición del primer transitorio es relevante puesto que efectos de preeco y otros problemas más probablemente serán enmascarados por las repercusiones del primer transitorio. Podría ser adecuado en esta situación adaptar la rejilla de tramas en consecuencia a la posición del primer transitorio.

Después de determinar la distribución de envolventes con respecto a las tramas, puede determinarse la resolución de frecuencia de los envolventes individuales. Como una resolución de la nueva envolvente, normalmente se usará la resolución más alta de los envolventes de entrada. Si por ejemplo, la resolución de uno de los envolventes analizados es alta, la trama de salida también comprende un envolvente con alta resolución en términos de su frecuencia.

Para ilustrar esta situación con más detalle, especialmente en el caso de que las tramas 540-1, 540-2 de entrada de las dos flujos 510-1, 510-2 de datos de entrada comprenda diferentes frecuencias de cruce, la figura 9a y la figura 9b ilustran representaciones respectivas tal como se muestra en la figura 6a para las dos 510-1, 540-2 tramas de entrada, respectivamente. Debido a la descripción muy detallada de la figura 6b, la descripción de las figuras 9a y 9b puede resumirse aquí. Además, la trama 540-1, tal como se muestra en la figura 9, es idéntica a aquella mostrado en la figura 6b. Como se describió anteriormente, comprende dos envolventes 620-1, 620-2 igualmente largos, con una pluralidad de regiones 630 de tiempo/frecuencia por encima de la frecuencia 570 de cruce.

La segunda trama 540-2, tal como se muestra esquemáticamente en la figura 9b, con respecto a unos pocos aspectos difiere de la trama mostrada en la figura 9a. Aparte del hecho de que la rejilla de trama comprende tres envolventes 620-1, 620-2 y 620-3 que no son igualmente largos, también la resolución de frecuencia con respecto a la región 630 de tiempo/frecuencia y la frecuencia 570 de cruce difiere de aquella mostrada en la figura 9a. En el ejemplo mostrado en la figura 9b, la frecuencia 570 de cruce es más grande que aquella de la trama 540-1 de la figura 9a. Como consecuencia, una parte superior del espectro 590 es así más grande que aquella de la trama 540-1 mostrado en la figura 9a.

El hecho de que, basándose en una suposición de un códec de AAC ELD haya proporcionado las tramas 540 tal como se muestran en las figuras 9a y 9b, la rejilla de trama de la trama 540-2 comprende tres envolventes 620 no igualmente largos da lugar a la conclusión de que el segundo de los tres envolventes 620 comprende un transitorio. Así, la rejilla de trama de la segunda trama 540-2 es, al menos con respecto a su distribución en el tiempo, la resolución a escoger para la trama 550 de salida.

Sin embargo, tal como se muestra en la figura 9c, surge un desafío adicional del hecho de que diferentes frecuencias 570 de cruce se emplean en la presente. Para ser más precisos, la figura 9c muestra una situación superpuesta en la que las dos tramas 540-1, 540-2, en términos de sus representaciones 560 de información espectral, se han mostrado conjuntamente. Al considerar sólo las frecuencias 570-1 de cruce de la primera trama 540, tal como se muestra en la figura 9a (frecuencia  $f_{x1}$  de cruce) y la frecuencia 570-2 de cruce más alta de la segunda trama 540-2 tal como se muestra en la figura 9b (frecuencia  $f_{x2}$  de cruce), un intervalo 1000 de frecuencia intermedio para el cual solamente los datos de SBR de la primera trama 540-1 y para el cual solamente los datos 610 espectrales de la segunda trama 540-1 están disponibles. En otras palabras, para componentes espectrales de frecuencias dentro del intervalo 1000 de frecuencia intermedio, el procedimiento de mezcla depende de los valores de SBR estimados o de datos espectrales estimados, tal como se proporciona por el estimador 670 mostrado en la figura 7.

En la situación ilustrada en la figura 9c, el intervalo 1000 de frecuencia intermedio, encerrado en términos de la frecuencia por las dos frecuencias 570-1, 570-2 de cruce representa el intervalo de frecuencia en el que el estimador 670 y la unidad 520 de procesamiento operan. En este intervalo 1000 de frecuencia, los datos de SBR están disponibles solamente de la primera trama 540-1, mientras que de la segunda trama 540-2 en aquel intervalo de frecuencia solamente la información espectral o los valores espectrales están disponibles. Como consecuencia, dependiendo de si una frecuencia o componente espectral del intervalo 1000 de frecuencia intermedio está por encima o por debajo de la frecuencia de cruce de salida, o bien un valor de SBR o bien un valor espectral debe evaluarse antes de mezclar el valor estimado con el valor original de una de las tramas 540-1, 540-2 en el dominio de SBR están en el dominio espectral.

La figura 9d ilustra la situación en la que la frecuencia de cruce de la trama de salida es igual a la más baja de las dos frecuencias 570-1, 570-2 de cruce. Como consecuencia, la frecuencia 570-3 ( $f_{xo}$ ) de cruce de salida es igual a la primera frecuencia 570-1 ( $f_{x1}$ ) de cruce, lo que también limita la parte superior del espectro codificado para que sea dos veces las frecuencias de cruce mencionadas.

Al copiar o redeterminar la resolución de frecuencia de la rejilla de tiempo/frecuencia basándose en la resolución de tiempo determinada anteriormente o la distribución de envolvente de la misma, se determinan los datos de SBR de salida en el intervalo 1000 de frecuencia intermedio (véanse la figura 9c) al estimar datos de SBR correspondientes a partir de los datos 610 espectrales de la segunda trama 540-2 para estas frecuencias.

Esta estimación puede llevarse a cabo basándose en los datos 610 espectrales de la segunda trama 540-2 en aquel intervalo de frecuencia teniendo en cuenta los datos de SBR para frecuencias por encima de la segunda frecuencia 570-2 de cruce. Esto está basado en la suposición de que en términos de la resolución de tiempo o las frecuencias de distribución de envolvente alrededor de la segunda frecuencia 570-2 de cruce son más probablemente influenciadas de

manera equivalente. Por consiguiente, la estimación de los datos de SBR en el intervalo 1000 de frecuencia intermedio puede lograrse, por ejemplo, mediante el cálculo sobre la resolución de tiempo y de frecuencia más fina descrita por los datos de SBR los valores de energía respectivos basándose en la información espectral para cada componente espectral y mediante la atenuación o amplificación de cada uno basándose en el desarrollo en el tiempo de la amplitud como se indicó por los envolventes de los datos de SBR de la segunda trama 540-2.

Después, mediante la aplicación de un filtro de suavizado u otra etapa de filtrado, los valores de energía estimados se mapean sobre las regiones 630 de tiempo/frecuencia de la rejilla de tiempo/frecuencia determinada para la trama 550 de salida. Esta solución, tal como se ilustra en la figura 9d, puede ser, por ejemplo, interesante para tasas de transmisión de bits más bajas. La frecuencia de cruce de SBR más baja de todas los flujos entrantes se usará como la frecuencia de cruce de SBR para la trama de salida y se estiman los valores de energía de SBR para la región 1000 de frecuencia en el espacio entre el decodificador central (que opera hasta la frecuencia de cruce) y el codificador de SBR (que opera por encima de la frecuencia de cruce) a partir de la información espectral o coeficientes espectrales. La estimación puede llevarse a cabo basándose en una amplia variedad de información espectral, por ejemplo, derivable de coeficientes espectrales de MDCT (transformada de coseno discreta modificada) o de LDFB (banco de filtros de bajo retardo). Adicionalmente, filtros de suavizado pueden aplicarse para cerrar el espacio entre el codificador central y la parte de SBR.

También ha de observarse que esta solución puede también usarse para reducir un flujo de una tasa de transmisión de bits alta, por ejemplo que comprende 64 kilobits/segundo, a un flujo de una tasa de transmisión de bits más baja, que comprende, por ejemplo, solamente 32 kilobits/segundo. Una situación en la que podría ser recomendable implementar tal solución es, por ejemplo proporcionar flujos de bits para participantes con conexiones de baja velocidad de datos a la unidad mezcladora que, por ejemplo, se establecen mediante la marcación por modem en conexiones o similares.

Otro caso de diferentes frecuencias de cruce se ilustra en la figura 9e.

La figura 9e muestra el caso en el que se usa la más alta de las dos frecuencias 570-1, 570-2 de cruce como la frecuencia 570-3 de cruce de salida. Por tanto, la trama 550 de salida comprende hasta la información 610 espectral de la frecuencia de cruce de salida y por encima de los datos de SBR correspondientes de la frecuencia de cruce de salida hasta una frecuencia de normalmente dos veces la frecuencia 570-3 de cruce. Esta situación, sin embargo, hace surgir la pregunta sobre cómo restablecer los datos espectrales en el intervalo 1000 de frecuencia intermedio (véanse la la figura 9c). Después de determinar la resolución de tiempo o la distribución de envolvente de la rejilla de tiempo/frecuencia y después de copiar o determinar al menos parcialmente la resolución de frecuencia de la rejilla de tiempo/frecuencia para frecuencias por encima de la frecuencia 570-3 de cruce de salida, basándose en los datos de SBR de la primera trama 540-1 en el intervalo 1000 de frecuencia intermedio, deben estimarse los datos espectrales mediante la unidad 520 de procesamiento y el estimador 670. Esto puede lograrse al reconstruir parcialmente la información espectral basándose en los datos de SBR para aquel intervalo 1000 de frecuencia de la primera trama 540-1, aunque teniendo en cuenta opcionalmente algo de o toda la información 610 espectral debajo de la primera frecuencia 570-1 de cruce (véanse la la 9a). En otras palabras, la estimación de la información espectral faltante puede lograrse al replicar de manera espectral la información espectral a partir de los datos de SBR y la información espectral correspondiente de la parte 580 inferior del espectro al aplicar el algoritmo de reconstrucción del decodificador de SBR al menos parcialmente a las frecuencias del intervalo 1000 de frecuencia intermedio.

Después de estimar la información espectral del intervalo de frecuencia intermedio mediante, por ejemplo, la aplicación de una decodificación de SBR o reconstrucción parcial al dominio de frecuencia, la información espectral estimada resultante puede mezclarse directamente con la información espectral de la segunda trama 540-2 en el dominio espectral mediante, por ejemplo, la aplicación de una combinación lineal.

La reconstrucción o replicación de información espectral para frecuencias o componentes especiales por encima de la frecuencia de cruce es también denominada como filtrado inversa. En este contexto, ha de observarse que armónicas adicionales y valores de energía de ruido adicionales pueden tenerse en consideración cuando se estima la información espectral respectiva para frecuencias o componentes en el intervalo 1000 de frecuencia intermedio.

Esta solución puede ser interesante, por ejemplo, para participantes que están conectados al aparato 500 o una unidad mezcladora que tiene disponibles tasas de transmisión de bits más altas. Un algoritmo de parche o copia puede aplicarse a la información espectral del dominio espectral, por ejemplo, a los coeficientes espectrales de MDCT o LDFB, para copiar éstos de la banda más baja a las bandas más altas para cerrar el espacio entre el codificador central y la parte de SBR que están separados por la frecuencia de cruce respectiva. Estos coeficientes de copia se atenúan según los parámetros de energía almacenados en la carga útil de SBR.

En ambos escenarios tal como se describen en las figuras 9d y 9e, la información espectral debajo de las frecuencias de cruce más bajas puede procesarse directamente en el dominio espectral, mientras que los datos de SBR que están por encima de la frecuencia de cruce más alta pueden procesarse directamente en el dominio de SBR. Para frecuencias muy altas por encima de la más baja de las frecuencias más altas como se describe por los datos de SBR, normalmente por encima de dos veces el valor mínimo de las frecuencias de cruce involucradas, dependiendo de la frecuencia de cruce de la trama 550 de salida pueden aplicarse diferentes enfoques. En principio, cuando se usa la más alta de las frecuencias de cruce involucradas como la frecuencia 570-3 de cruce de salida tal como se ilustra en la figura 9e, los

datos de SBR para la frecuencia más alta están basados principalmente en los datos de SBR de la segunda trama 540-2 solamente. Como opción adicional, estos valores pueden atenuarse mediante un factor de normalización o factor de amortiguación aplicado en el marco de la combinación lineal de los valores de energía de SBR para las frecuencias debajo de aquella frecuencia de cruce. En la situación tal como se ilustra en la figura 9d, cuando se utiliza la más baja de las frecuencias de cruce disponibles como la frecuencia de cruce de salida, los datos de SBR respectivos de la segunda trama 540-2 pueden descartarse.

Naturalmente, ha de observarse que las realizaciones según la presente invención no están, ni mucho menos, limitadas a solamente dos flujos de datos de entrada que pueden extenderse fácilmente a una pluralidad de flujos de datos de entrada que comprenden más de dos flujos de datos de entrada. En tal caso, los enfoques descritos pueden aplicarse fácilmente a diferentes flujos de datos de entrada dependiendo de la frecuencia de cruce real usada en vista de aquel flujo de datos de entrada. Cuando, por ejemplo, la frecuencia de cruce de este flujo de datos de entrada son de una trama comprendida en aquel flujo de datos de entrada es más alta que la frecuencia de cruce de salida de la trama 550 de salida, pueden aplicarse los algoritmos tal como se describe en el contexto de la figura 9d. Por el contrario, cuando la frecuencia de cruce correspondiente es más baja, los algoritmos y procesos descritos en contexto con la figura 9e pueden aplicarse a este flujo de datos de entrada. La mezcla real de los datos de SBR o la información espectral en el sentido de que se suman más de dos de los datos respectivos.

Además, ha de observarse que la frecuencia 570-3 de cruce de salida puede escogerse arbitrariamente. No se requiere, ni mucho menos, que sea idéntica a cualquiera de las frecuencias de cruce de los flujos de datos de entrada. Por ejemplo, en la situación tal como se describe en contexto con las figuras 9d y 9e, la frecuencia de cruce podría también caer entre, por debajo o por encima de ambas frecuencias 570-1, 570-2 de cruce de los flujos 510 de datos de entrada. En caso de que la frecuencia de cruce de la trama 550 de salida pueda escogerse libremente, puede ser recomendable implementar todos los algoritmos descritos anteriormente en términos de estimar datos espectrales así como datos de SBR.

Por otra parte, algunas realizaciones según la presente invención pueden implementarse de tal manera que se utiliza siempre la frecuencia de cruce más baja o siempre la frecuencia de cruce más alta. En tal caso, podría no ser necesario implementar la plena funcionalidad tal como se describió anteriormente. Por ejemplo, en caso de que siempre se emplee la frecuencia de cruce más baja, normalmente no se requiere que el estimador 670 pueda estimar información espectral, sino solamente datos de SBR. Por tanto, la funcionalidad de estimar datos espectrales puede eventualmente evitarse en el presente. Por el contrario, en el caso de que una realización según la presente invención se implemente de tal manera que siempre se emplee la frecuencia de cruce de salida más alta, la funcionalidad del estimador 670 de poder estimar los datos de SBR podría no requerirse y por tanto omitirse.

Las realizaciones según la presente invención pueden comprender además componentes de mezcla descendente de multicanal o componentes de mezcla ascendente de multicanal, por ejemplo, componentes de mezcla descendente estéreo o componentes de mezcla ascendente estéreo en el caso de que algunos participantes puedan enviar flujos estéreo u otros flujos de multicanal y algunos flujos mono solamente. En este caso, puede ser recomendable implementar una mezcla ascendente o mezcla descendente correspondiente en términos del número de canales comprendidos en los flujos de datos de entrada. Puede ser recomendable procesar algunos de los flujos mediante la mezcla ascendente o la mezcla descendente para proporcionar flujos de bits mezclados que coinciden con los parámetros de los flujos entrantes. Esto puede significar que el participante que envía un flujo mono puede también desear recibir un flujo mono en respuesta. Como consecuencia, los datos de audio estéreo u otros datos de audio multicanal de otros participantes pueden tener que convertirse a un flujo mono o al revés.

Dependiendo de las restricciones de implementación y otras condiciones de frontera esto puede lograrse, por ejemplo, al implementar una pluralidad de aparatos según una realización de la presente invención o procesar todas los flujos de datos de entrada basándose en un solo aparato, en el que los flujos de datos entrantes se mezclan de manera descendente o se mezclan de manera ascendente antes del procesamiento por el aparato y se mezclan de manera descendente o se mezclan de manera ascendente después del procesamiento para coincidir con los requisitos del terminal del participante.

SBR también permite dos modos de codificación de canales estéreo. Un modo de operación trata los canales izquierdo y derecho (LR) separadamente, mientras que un segundo modo de operación opera en un canal acoplado (C). Para mezclar un elemento LR-codificado y un elemento C-codificado, o bien tiene que enviarse el elemento LR-codificado a un elemento C o viceversa. La decisión real de qué método va a usarse puede establecerse anteriormente o efectuarse al tener en cuenta condiciones tales como consumo de energía, cálculo y complejidad y similares o puede basarse en una estimación psicoacústica en términos de la relevancia de un tratamiento separado.

Como se indicó anteriormente, la mezcla de los datos relacionados con la energía de SBR real se puede llevar a cabo en el dominio de SBR por una combinación lineal de los valores de energía respectivos. Esto se puede obtener según la ecuación

$$E(n) = \sum_{k=1}^N a_k \cdot E_k(n) \quad (6)$$

en el que  $a_k$  es un factor de ponderación,  $E_k(n)$  es el valor de energía del flujo de datos de entrada  $k$ , correspondiente a una posición en la rejilla de tiempo/frecuencia indicada por  $n$ .  $E(n)$  es el valor de energía SBR correspondiente al mismo índice  $n$ .  $N$  es el número de flujos de datos de entrada y en el ejemplo mostrado en la figura 9a y 9e igual a dos.

5 Los coeficientes  $a_k$  pueden usarse para efectuar una normalización, así como una ponderación con respecto a cada región 630 de tiempo/frecuencia de la trama 550 de salida y las regiones 630 de tiempo/frecuencia correspondientes de la trama 450 de entrada respectiva superpuesta. Por ejemplo, en el caso de que las dos regiones 630 de tiempo/frecuencia de la trama 550 de salida y la trama 540 de entrada respectiva tengan una superposición entre sí hasta una extensión del 50% en el sentido de que el 50% de la región 630 de tiempo/frecuencia bajo consideración de la trama 550 de salida está compuesta por la región 630 de tiempo/frecuencia correspondiente de la trama 540 de entrada, el valor de 0.5 (= 50%) puede multiplicarse por un factor de ganancia global que indica la relevancia del flujo de entrada de audio respectivo y la trama 540 de entrada comprendida en la misma.

Para ponerlo en términos más generales, cada uno de los coeficientes  $a_k$  puede definirse según

15 (7) 
$$a_k = \sum_{i=1}^M r_{ik} \cdot g$$

en el que  $r_{ik}$  es el valor que indica la región de superposición de las dos regiones 630 de tiempo/frecuencia  $i$  y  $k$  de la trama 540 de entrada y la trama 550 de salida, respectivamente.  $M$  es el número de todas las regiones 630 de tiempo/frecuencia de la trama 540 de entrada y  $g$  un factor de normalización global, que puede ser por ejemplo, igual a  $1/N$  para impedir que el resultado del proceso de mezcla se sobreimpulse o se subimpulse a un intervalo permisible de valores. Los coeficientes  $r_{ik}$  pueden estar en el intervalo de entre 0 y 1, en el que 0 indica que las dos regiones 630 de tiempo/frecuencia no se superponen y un valor de 1 indica que la región 630 de tiempo/frecuencia de la trama 540 de entrada está completamente comprendida en la respectiva región 630 de tiempo/frecuencia de la trama 550 de salida.

25 Sin embargo, puede también ocurrir que las rejillas de trama de las tramas 540 de entrada sean iguales. En este caso, las rejillas de tramas pueden copiarse de una de las tramas 540 de entrada a la trama 550 de salida. Así, la mezcla de los valores de energía de SBR relevantes puede efectuarse muy fácilmente. Los valores de frecuencia correspondientes pueden agregarse en este caso de forma similar a la mezcla de información espectral correspondiente (por ejemplo, valores de MDCT) al agregar y normalizar los valores de salida.

30 Sin embargo, puesto que el número de regiones 630 de tiempo/frecuencia en términos de la frecuencia puede cambiar dependiendo de la resolución del envolvente respectiva, puede ser recomendable implementar un mapeo de una envolvente baja a una envolvente alta y viceversa.

35 La figura 10 ilustra esto para el ejemplo de ocho regiones 630-1 de tiempo/frecuencia y una envolvente alta que comprende 16 regiones 630-h de tiempo/frecuencia correspondientes. Como se define anteriormente, una envolvente de baja resolución comprende normalmente sólo la mitad del número de datos de frecuencia cuando se compara con una envolvente de alta resolución, se puede establecer una coincidencia simple como se ilustra en la figura 10. Cuando se mapea el envolvente de baja resolución a una envolvente de alta resolución, cada una de las regiones 630-1 de tiempo/frecuencia del envolvente de baja resolución se mapean a dos regiones 630-h de tiempo/frecuencia correspondientes de una envolvente de alta resolución.

40 Dependiendo de la situación concreta, por ejemplo, en términos de normalización, el uso de un factor adicional de 0.5 puede ser recomendable para impedir el sobreimpulso de los valores de energía de SBR mezclados. En el caso de que el mapeo se haga al contrario, dos regiones 630-h de tiempo/frecuencia vecinas pueden ser promediadas al determinar el valor medio aritmético para obtener una región 630-1 tiempo/frecuencia de una envolvente de baja resolución.

45 Dicho de otro modo, en la primera situación con respecto a la ecuación (7), los factores  $r_{ik}$  son 0 o 1, mientras que el factor  $g$  es igual a 0.5, en el segundo caso el factor  $g$  puede establecerse en 1 mientras que el factor  $r_{ik}$  puede ser 0 ó 0.5.

50 Sin embargo, el factor  $g$  puede tener que modificarse adicionalmente al incluir un factor de normalización adicional teniendo en cuenta el número de flujos de datos de entrada a mezclarse. Para mezclar los valores de energía de todas las señales de entrada, las mismas se suman y opcionalmente se multiplican con un factor de normalización aplicado durante el procedimiento de mezcla espectral. Este factor de normalización adicional puede eventualmente también tener que tenerse en cuenta, cuando se determina el factor  $g$  en la ecuación (7). Como consecuencia, esto puede asegurar eventualmente que los factores de escala de los coeficientes espectrales del códec base coinciden con el intervalo permisible de valores de los valores de energía de SBR.

55 Las realizaciones según la presente invención pueden naturalmente diferir con respecto a sus implementaciones. Aunque en las realizaciones anteriores, se ha descrito una decodificación y codificación de Huffman como un solo esquema de codificación de entropía, también pueden usarse otros esquemas de codificación de entropía. Además, no se requiere en absoluto la implementación de un codificador de entropía o un decodificador de entropía. Así, aunque la descripción de las realizaciones previas se ha enfocado principalmente en el códec AAP-ELD, también otros códecs

pueden usarse para proporcionar los flujos de datos de entrada y para decodificar el flujo de datos de salida en el lado del participante. Por ejemplo, cualquier códec que basado en, por ejemplo, puede usarse una sola ventana con conmutación de longitud de bloque.

5 Como ha mostrado también la descripción precedente de la realización mostrada en la figura 8, los módulos descritos en la misma no son determinantes. Por ejemplo, un aparato según una realización de la presente invención puede realizarse simplemente al operar sobre la información espectral de las tramas.

10 Ha de observarse adicionalmente que las realizaciones según la presente invención pueden realizarse de maneras muy diferentes. Por ejemplo, un aparato 500 para mezclar una pluralidad de flujos de datos de entrada y su unidad 520 de procesamiento puede realizarse basándose en dispositivos eléctricos y electrónicos discretos tales como resistores, transistores, inductores y similares. Además, las realizaciones según la presente invención pueden realizarse también basándose en circuitos integrados solamente, por ejemplo, en forma de SOC (SOC = sistema sobre chip), procesadores tales como CPU (CPU = unidad de procesamiento central), GPU (GPU = unidad de procesamiento gráfica) y otros circuitos integrados tales como circuitos integrados específicos de la aplicación (ASIC).

15 También ha de observarse que los dispositivos eléctricos que son parte de la implementación discreta o que son parte de un circuito integrado pueden usarse para funciones propuestas diferentes y diferentes funciones en toda la implementación de un aparato según una realización de la presente invención. Naturalmente, también una combinación de circuitos basada en circuitos integrados y circuitos discretos puede usarse para implementar una realización según la presente invención.

20 Basándose en un procesador, las realizaciones según la presente invención pueden también ser implementadas basándose en un programa de ordenador, un programa de elementos de programación o un programa que se ejecuta en un procesador.

25 En otras palabras, dependiendo de ciertos requerimientos de implementación de realizaciones de métodos de la invención, las realizaciones de los métodos de la invención pueden implementarse en elementos físicos o en elementos de programación. La implementación puede efectuarse utilizando un medio de almacenamiento digital, en particular un disco, un CD o un DVD que tiene señales que se pueden leer electrónicamente guardadas en el mismo que cooperan con un ordenador o procesador programable de tal manera que se efectúa una realización del método de la invención. En general, una realización de la presente invención es, por consiguiente, un producto de programa de ordenador con códigos de programa guardados en un portador que se puede leer por la máquina, los códigos de programa son operativos para efectuar una realización del método de la invención cuando el producto de programas de ordenador se ejecuta en un ordenador o procesador. Dicho aún de otro modo, las realizaciones de los métodos de la invención son, por consiguiente, un programa de ordenador que tiene códigos de programa para efectuar al menos una de las realizaciones de los métodos de la invención, cuando el programa de computadora se ejecuta en un ordenador o procesador. Un procesador puede estar formado por un ordenador, una tarjeta de chip, una tarjeta inteligente, un circuito integrado específico de la aplicación, un sistema sobre chip (SOC) o un circuito integrado (IC).

35 Lista de símbolos de referencia

- 100 Sistema de conferencia
- 110 Entrada
- 120 Decodificador
- 130 Sumador
- 40 140 Codificador
- 150 Salida
- 160 Terminal de conferencia
- 170 Codificador
- 180 Decodificador
- 45 190 Convertidor de tiempo/frecuencia
- 200 Cuantificador/codificador
- 210 Decodificador/descuantificador
- 220 Convertidor de frecuencia/tiempo
- 50 250 Flujo de datos
- 260 Trama
- 270 Bloques de información adicional
- 300 Frecuencia
- 55 310 Banda de frecuencia
- 500 Aparato
- 510 Flujo de datos de entrada
- 520 Unidad de procesamiento
- 60 530 Flujo de datos de salida

	540 Trama
	550 Trama de salida
	560 Representación de información espectral
	570 Frecuencia de cruce
5	580 Parte inferior del espectro
	590 Parte superior del espectro
	600 Línea
	610 Datos espectrales
10	620 Envoltente
	630 Región de tiempo/frecuencia
	640 Analizador
	650 Mezclador espectral
	660 Mezclador de SBR
15	670 Estimador
	680 Mezclador
	700 Decodificador de flujo de bits
	710 Lector de flujo de bits
20	720 Codificador de Huffman
	730 Descuantificador
	740 Escalador
	750 Primera unidad
	760 Segunda unidad
25	770 Decodificador estéreo
	780 Decodificador de PNS
	790 Decodificador de TNS
	800 Unidad mezcladora
30	810 Mezclador espectral
	820 Módulo de optimización
	830 Mezclador SBR
	850 Codificador de flujo de bits
	860 Tercera unidad
35	870 Codificador de TNS
	880 Codificador de PNS
	890 Codificador estéreo
	900 Cuarta unidad
40	910 Escalador
	920 Cuantificador
	930 Codificador de Huffman
	940 Grabador de flujo de bits
45	950 Módulo psicoacústico
	1000 Intervalo de frecuencia intermedia

**REIVINDICACIONES**

1. Aparato (500) para mezclar una primera trama (540-1) de un primer flujo (510-1) de datos de entrada y una segunda trama (540-2) de un segundo flujo (510-2) de datos de entrada para obtener una trama (550) de salida de un flujo (530) de datos de salida, en el que la primera trama (540-1) comprende primeros datos espectrales que describen una parte (580) inferior de un primer espectro de una primera señal de audio hasta una primera frecuencia (570) de cruce y primeros datos de replicación de banda espectral (SBR) que describen una parte (590) superior del primer espectro partiendo de la primera frecuencia (570) de cruce, en el que la segunda trama (540-2) comprende segundos datos espectrales que describen una parte (580) inferior de un segundo espectro de una segunda señal de audio hasta una segunda frecuencia (570) de cruce y segundos datos de SBR que describen una parte (590) superior del segundo espectro partiendo de la segunda frecuencia (570) de cruce, en el que los primeros y segundos datos de SBR describen las respectivas partes (590) superiores del primer y el segundo espectro por medio de valores relacionados con la energía en resoluciones de rejilla de tiempo/frecuencia y en el que la primera frecuencia (570) de cruce es diferente de la segunda frecuencia (570) de cruce, comprendiendo el aparato (500):
- una unidad (520) de procesamiento adaptada para generar la trama (550) de salida, comprendiendo la trama (550) de salida datos espectrales de salida que describen una parte (580) inferior de un espectro de salida hasta una frecuencia (570) de cruce de salida y comprendiendo la trama (550) de salida además datos de SBR de salida que describen una parte (590) superior del espectro de salida por encima de la frecuencia (570) de cruce de salida por medio de valores relacionados con la energía en una resolución de rejilla de tiempo/frecuencia de salida,
- en el que la unidad (520) de procesamiento está adaptada de tal manera que los datos espectrales de salida correspondientes a la frecuencia por debajo de un valor mínimo de la primera frecuencia (570) de cruce, la segunda frecuencia (570) de cruce y la frecuencia (570) de cruce de salida se generan en un dominio espectral basándose en los primeros y segundos datos espectrales;
- en el que la unidad (520) de procesamiento está adaptada además de tal manera que los datos de SBR de salida correspondientes a las frecuencias por encima de un valor máximo de la primera frecuencia (570) de cruce, la segunda frecuencia (570) de cruce y la frecuencia (570) de cruce de salida se procesan en un dominio de SBR basándose en los primeros y segundos datos de SBR; y
- en el que la unidad (520) de procesamiento está adaptada además de tal manera que para una región de frecuencia entre el valor mínimo y el valor máximo, se estima al menos un valor de SBR a partir de al menos uno de los primeros y segundos datos espectrales y se genera un valor de SBR correspondiente de los datos de SBR de salida, basándose en al menos el valor de SBR estimado.
2. Aparato (500) según la reivindicación 1, en el que la unidad (520) de procesamiento está adaptada para estimar el al menos un valor de SBR basándose en un valor espectral correspondiente a una componente de frecuencia correspondiente al valor de SBR que va a estimarse.
3. Aparato (500) para mezclar una primera trama (540-1) de un primer flujo (510-1) de datos de entrada y una segunda trama (540-2) de un segundo flujo (510-2) de datos de entrada para obtener una trama (550) de salida de un flujo (530) de datos de salida, en el que la primera trama (540-1) comprende primeros datos espectrales que describen una parte (580) inferior de un primer espectro de una primera señal de audio hasta una primera frecuencia (570) de cruce y primeros datos de replicación de banda espectral (SBR) que describen una parte (590) superior del primer espectro partiendo de la primera frecuencia (570) de cruce, en el que la segunda trama (540-2) comprende segundos datos espectrales que describen una parte (580) inferior de un segundo espectro de una segunda señal de audio hasta una segunda frecuencia (570) de cruce y segundos datos de SBR que describen una parte (590) superior del segundo espectro partiendo de la segunda frecuencia (570) de cruce, en el que los primeros y segundos datos de SBR describen las partes (590) superiores respectivas del primer y el segundo espectro por medio de valores relacionados con la energía en resoluciones de rejilla de tiempo/frecuencia y en el que la primera frecuencia (570) de cruce es diferente de la segunda frecuencia (570) de cruce, comprendiendo el aparato (500):
- una unidad (520) de procesamiento adaptada para generar la trama (550) de salida, comprendiendo la trama (550) de salida datos espectrales de salida que describen una parte (580) inferior de un espectro de salida hasta una frecuencia (570) de cruce de salida y comprendiendo la trama (550) de salida además datos de SBR de salida que describen una parte (590) superior del espectro de salida por encima de la frecuencia (570) de cruce de salida por medio de valores relacionados con la energía en una resolución de rejilla de tiempo/frecuencia de salida,
- en el que la unidad (520) de procesamiento está adaptada de tal manera que los datos espectrales de salida correspondientes a las frecuencias por debajo de un valor mínimo de la primera frecuencia (570) de cruce, la segunda frecuencia (570) de cruce y la frecuencia (570) de cruce de salida se generan en un dominio espectral basándose en los primeros y segundos datos espectrales;
- en el que la unidad (520) de procesamiento está adaptada además de tal manera que los datos de SBR de salida correspondientes a las frecuencias por encima de un valor máximo de la primera frecuencia (570) de cruce, la segunda frecuencia (570) de cruce y la frecuencia (570) de cruce de salida se procesan en un dominio de SBR basándose en los primeros y segundos datos de SBR; y

en el que el aparato (500) está adaptado además de tal manera que para una región de frecuencia entre el valor mínimo y el valor máximo, se estima al menos un valor espectral a partir de al menos una de las tramas primera y segunda basándose en los datos de SBR de la trama respectiva y se genera un valor espectral correspondiente de los datos espectrales de salida basándose en al menos el valor espectral estimado procesándolo en el dominio espectral.

- 5 4. Aparato según la reivindicación 3, en el que la unidad de procesamiento está adaptada para estimar el al menos un valor espectral basándose en la reconstrucción de al menos un valor espectral para una componente espectral basándose en los datos de SBR y los datos espectrales de la parte inferior del espectro respectivo de la trama respectiva.
- 10 5. Aparato (500) según cualquiera de las reivindicaciones anteriores, en el que la unidad (520) de procesamiento está adaptada para determinar la frecuencia (570) de cruce de salida para que sea la primera frecuencia de cruce o la segunda frecuencia de cruce.
- 15 6. Aparato (500) según cualquiera de las reivindicaciones anteriores, en el que la unidad (520) de procesamiento está adaptada para establecer la frecuencia de cruce de salida a la frecuencia de cruce más baja de una primera y una segunda frecuencia de cruce, o para establecer la frecuencia de cruce de salida a la más alta de las frecuencias de cruce primera y segunda.
- 20 7. Aparato (500) según cualquiera de las reivindicaciones anteriores, en el que la unidad (520) de procesamiento está adaptada para determinar la resolución de rejilla de tiempo/frecuencia de salida para que sea compatible con una posición transitoria de un transitorio indicado por la resolución de rejilla de tiempo/frecuencia de la primera o segunda trama.
- 25 8. Aparato (500) según la reivindicación 7, en el que la unidad (520) de procesamiento está adaptada para establecer la resolución de rejilla de tiempo/frecuencia para que sea compatible con un transitorio anterior indicado por las resoluciones de rejilla de tiempo/frecuencia de las tramas primera y segunda, cuando las resoluciones de rejilla de tiempo/frecuencia de las tramas primera y segunda indican una presencia de más de un transitorio.
- 30 9. Aparato (500) según cualquiera de las reivindicaciones anteriores, en el que la unidad (520) de procesamiento está adaptada para emitir datos espectrales o para emitir datos de SBR basándose en una combinación lineal en el dominio de frecuencia de SBR o en el dominio de SBR.
- 35 10. Aparato (500) según cualquiera de las reivindicaciones anteriores, en el que la unidad (520) de procesamiento está adaptada para generar los datos de SBR de salida que comprenden datos de SBR relacionados con la sinusoide basándose en una combinación lineal de datos de SBR relacionados con la sinusoide de las tramas primera y segunda.
- 40 11. Aparato (500) según cualquiera de las reivindicaciones anteriores, caracterizado porque la unidad (520) de procesamiento está adaptada para generar los datos de SBR de salida que comprenden datos de SBR relacionados con el ruido basándose en una combinación lineal de datos de SBR relacionados con el ruido de las tramas primera y segunda.
- 45 12. Aparato (500) según cualquiera de las reivindicaciones 10 u 11, en el que la unidad (520) de procesamiento está adaptada para incluir los datos de SBR relacionados con la sinusoide o relacionados con el ruido en una estimación psicoacústica de la relevancia de datos de SBR respectivos de las tramas primera y segunda.
- 50 13. Aparato (500) según cualquiera de las reivindicaciones anteriores, en el que la unidad (520) de procesamiento esta adaptada para generar los datos de SBR de salida basándose en un filtrado de suavizado.
- 55 14. Aparato (500) según cualquiera de las reivindicaciones anteriores, en el que el aparato está adaptado para procesar una pluralidad de flujos (510) de datos de entrada, comprendiendo la pluralidad de flujos de datos de entrada más de dos flujos de datos de entrada, en el que la pluralidad de flujos de datos de entrada comprende los flujos (510-1, 510-2) de datos de entrada primero y segundo.
15. Método para mezclar una primera trama (540-1) de un primer flujo (510-1) de datos de entrada y una segunda trama (540-2) de un segundo flujo (510-2) de datos de entrada para obtener una trama (550) de salida de un flujo (530) de datos de salida, en el que la primera trama comprende primeros datos espectrales que describen una parte (580) inferior de un espectro de una primera señal de audio hasta una primera frecuencia (570) de cruce y primeros datos de replicación de banda espectral (SBR) que describen una parte (590) superior del espectro partiendo de la primera frecuencia de cruce, en el que la segunda trama comprende segundos datos espectrales que describen una parte inferior de un segundo espectro de una segunda señal de audio hasta una segunda frecuencia de cruce y segundos datos de SBR que describen una parte superior de un segundo espectro partiendo de la segunda frecuencia de cruce, en el que los primeros y segundos datos de SBR describen las partes superiores respectivas de los espectros respectivos por medio de valores relacionados con la energía en resoluciones de rejilla de tiempo/frecuencia y en el que la primera frecuencia de cruce es diferente de la segunda frecuencia de cruce, que comprende:
- generar la trama de salida que comprende datos espectrales de salida que describen una parte inferior de un espectro de salida hasta una frecuencia de cruce de salida y comprendiendo además la trama de salida datos de SBR de salida

que describen una parte superior del espectro de salida por encima de la frecuencia de cruce de salida por medio de valores relacionados con la energía en una resolución de rejilla de tiempo/frecuencia de salida;

5 generar datos espectrales correspondientes a frecuencias por debajo de un valor mínimo de la primera frecuencia de cruce, la segunda frecuencia de cruce y una frecuencia de cruce de salida en un dominio espectral basándose en los primeros y segundos datos espectrales;

generar datos de SBR de salida correspondientes a frecuencias por encima de un valor máximo de la primera frecuencia de cruce, la segunda frecuencia de cruce y la frecuencia de cruce de salida en un dominio de SBR basándose en los primeros y segundos datos de SBR; y

10 estimar al menos un valor de SBR a partir de al menos uno de primeros y segundos datos espectrales para una frecuencia en una región de frecuencia entre el valor mínimo y el valor máximo y generar un valor de SBR correspondiente para los datos de SBR de salida, basándose en al menos el valor de SBR estimado; o

15 estimar al menos un valor espectral a partir de al menos una de las tramas primera y segunda basándose en los datos de SBR de la trama respectiva para una frecuencia en una región de frecuencia entre el valor mínimo y el valor máximo y generar un valor espectral de los datos espectrales de salida basándose en al menos el valor espectral estimado procesándolo en el dominio espectral.

16. Programa para efectuar, cuando se ejecuta en un procesador, un método para mezclar una primera trama de un primer flujo de datos de entrada y una segunda trama de un segundo flujo de datos de entrada según la reivindicación 15.

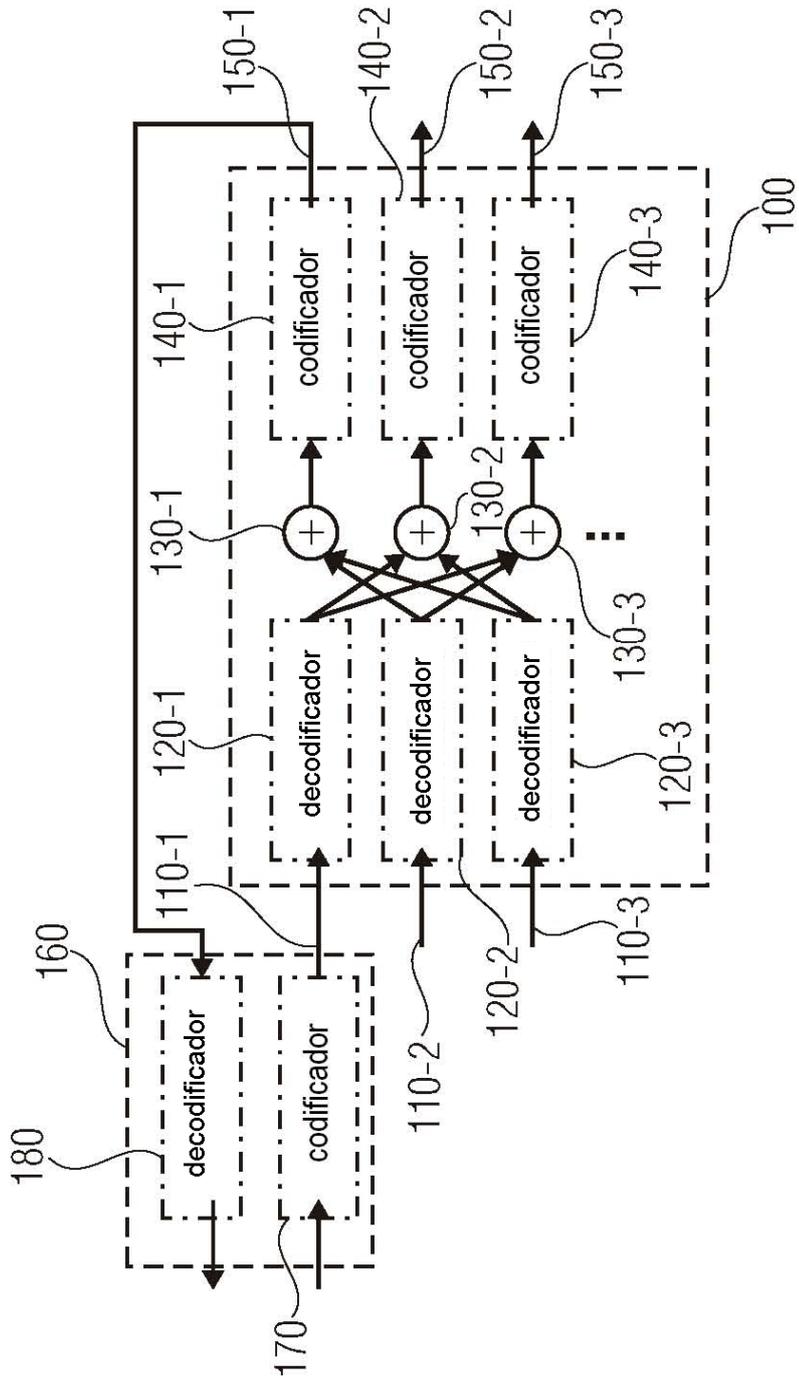


FIG 1

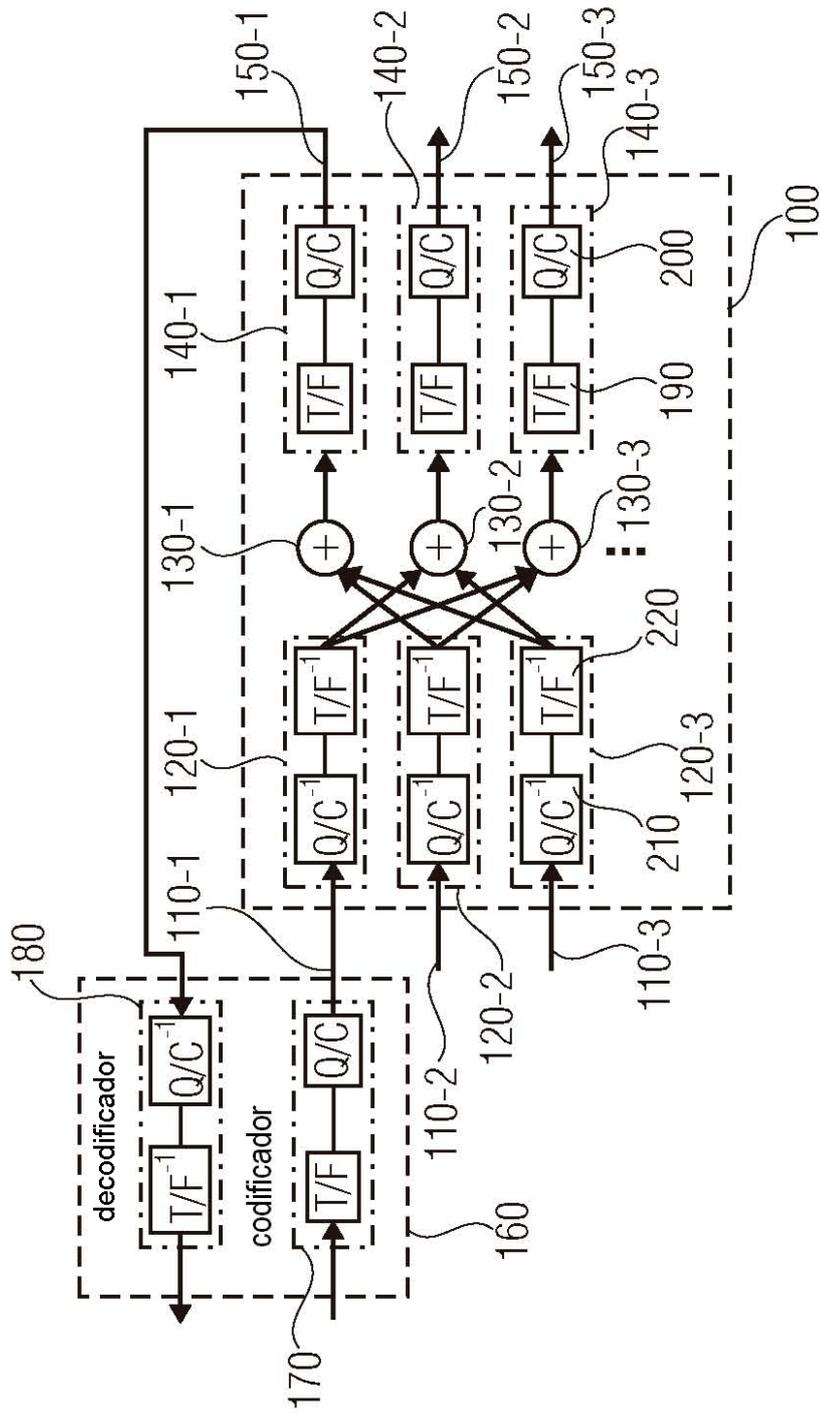


FIG 2



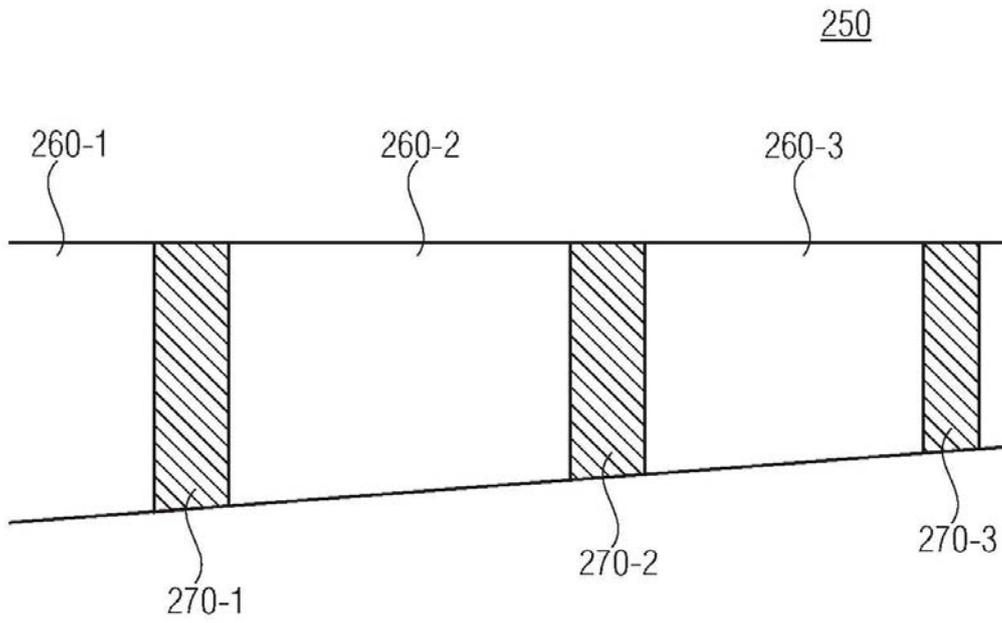


FIG 4

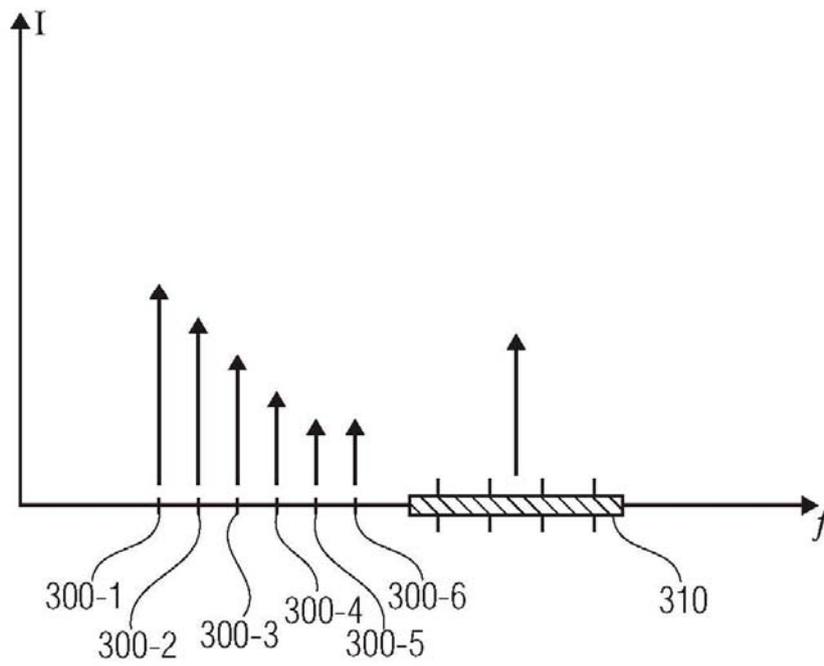


FIG 5

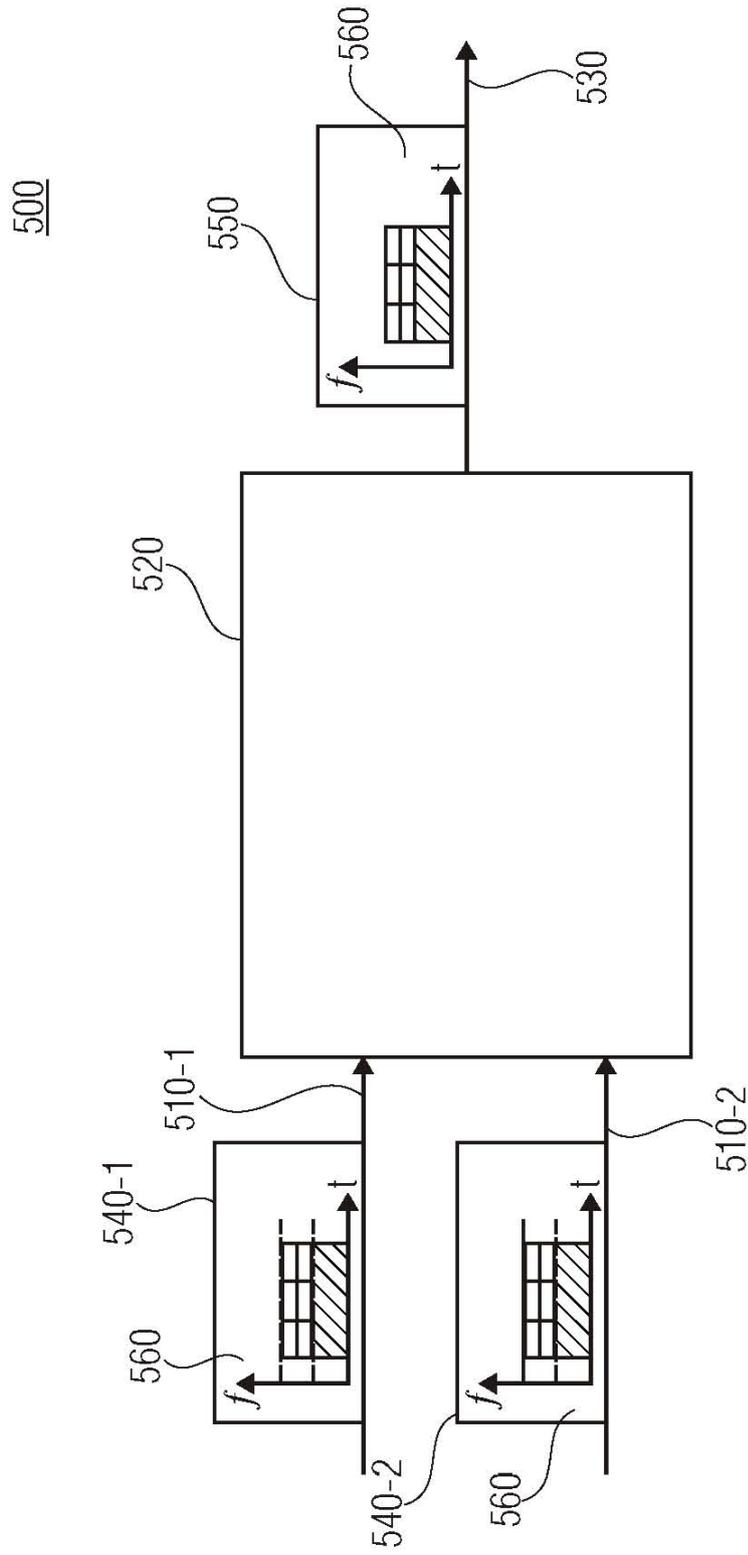


FIG 6A

560

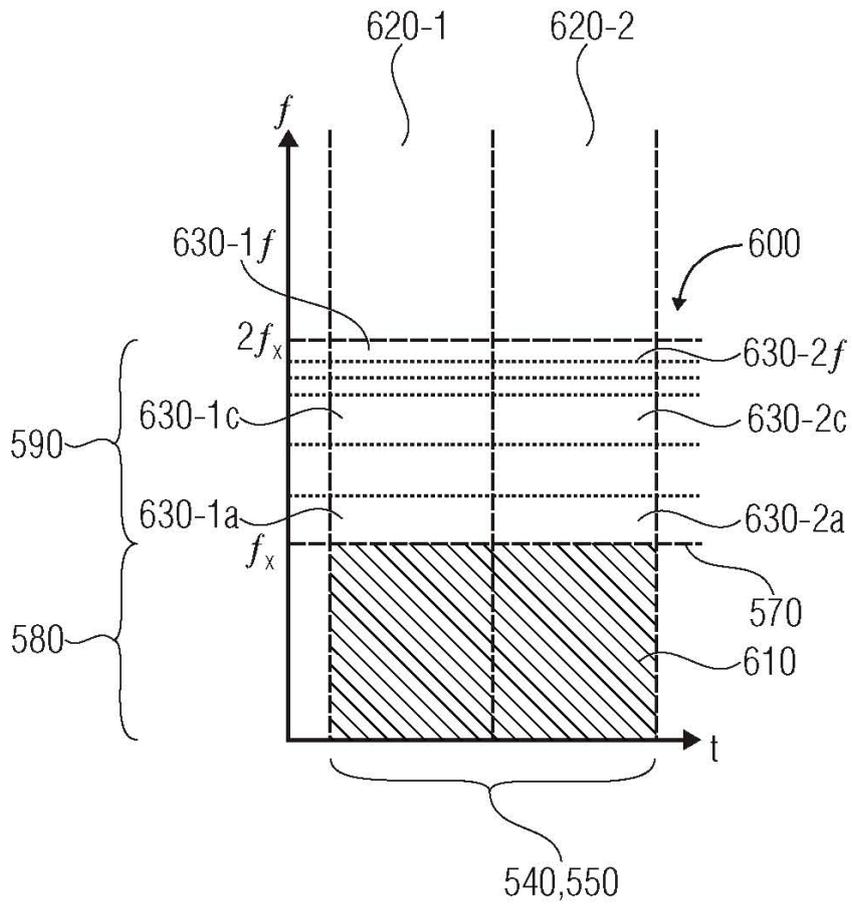


FIG 6B

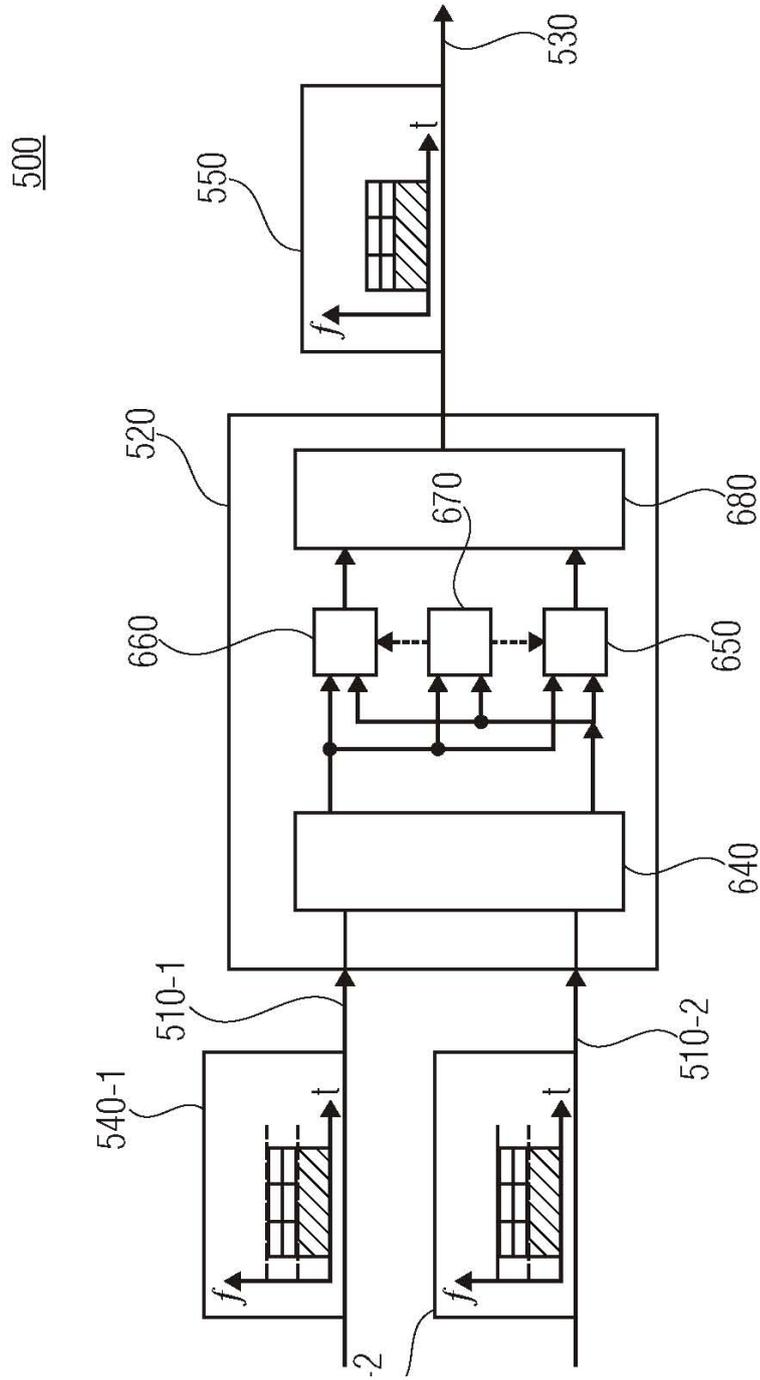


FIG 7

500

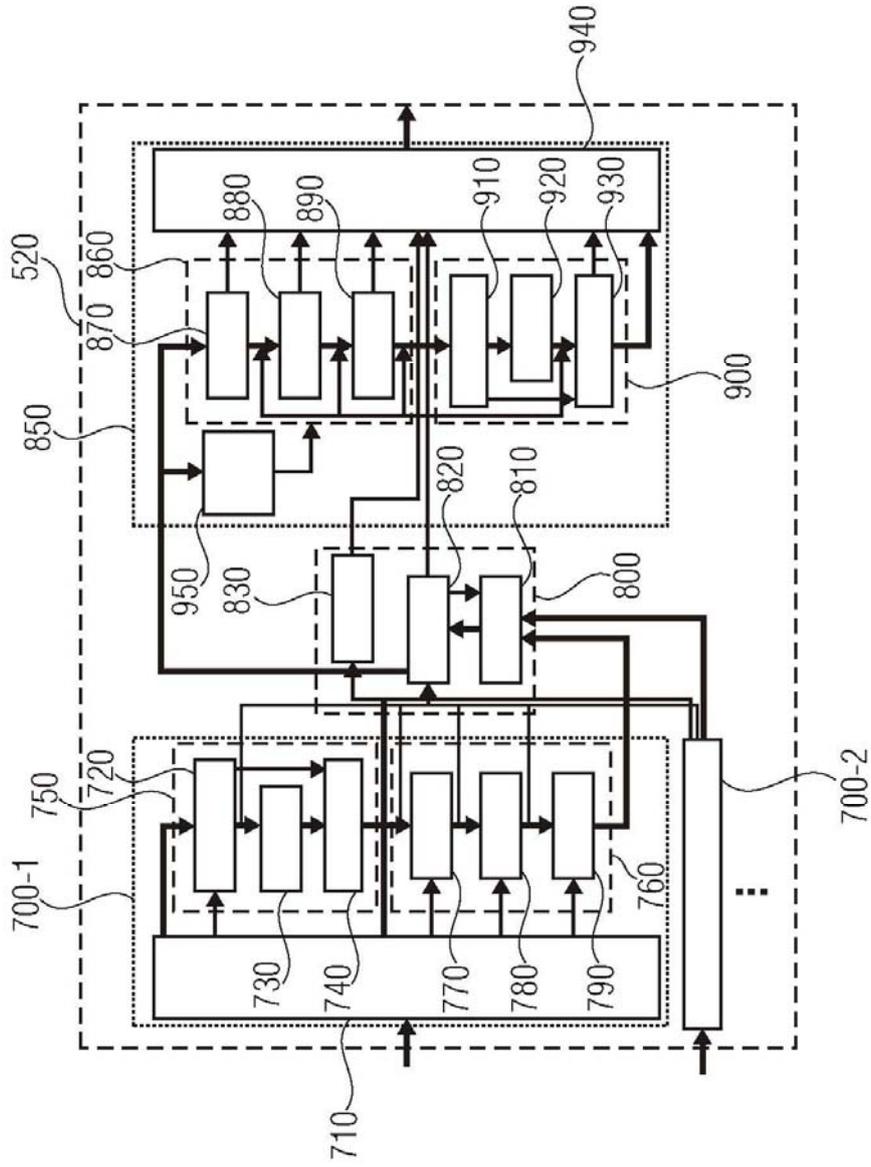


FIG 8

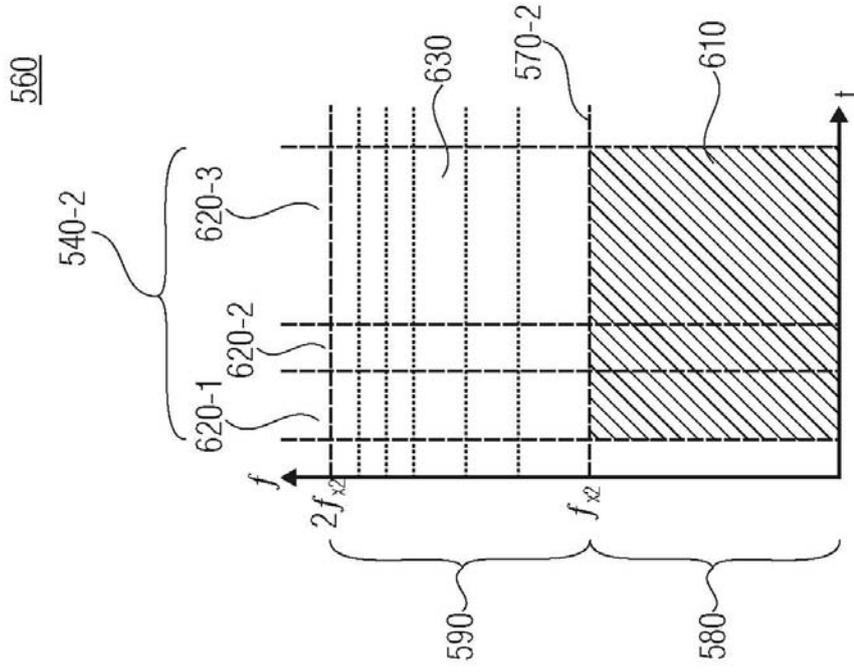


FIG 9A

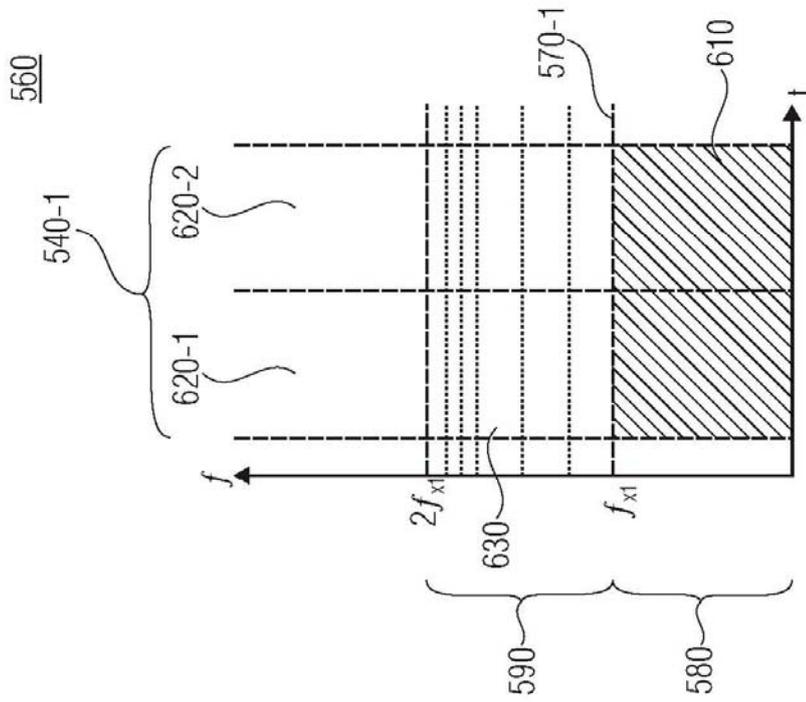


FIG 9B

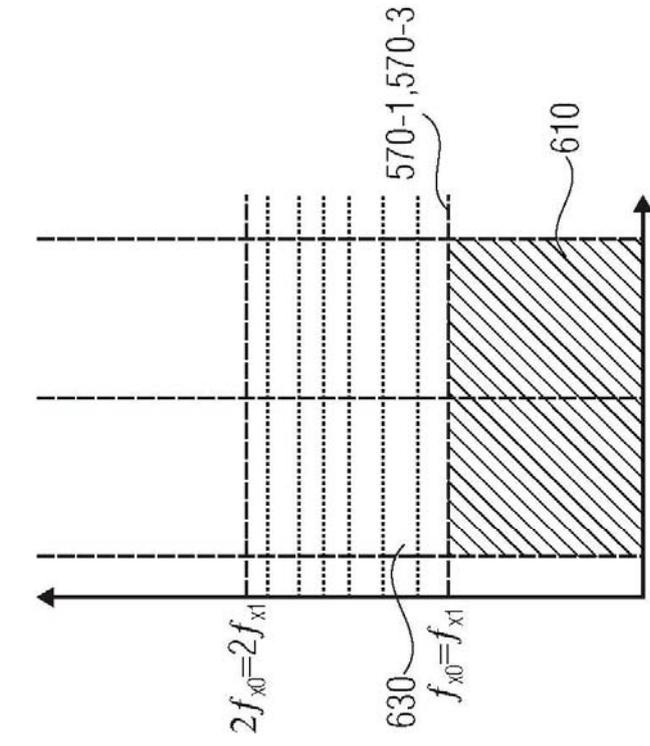


FIG 9D

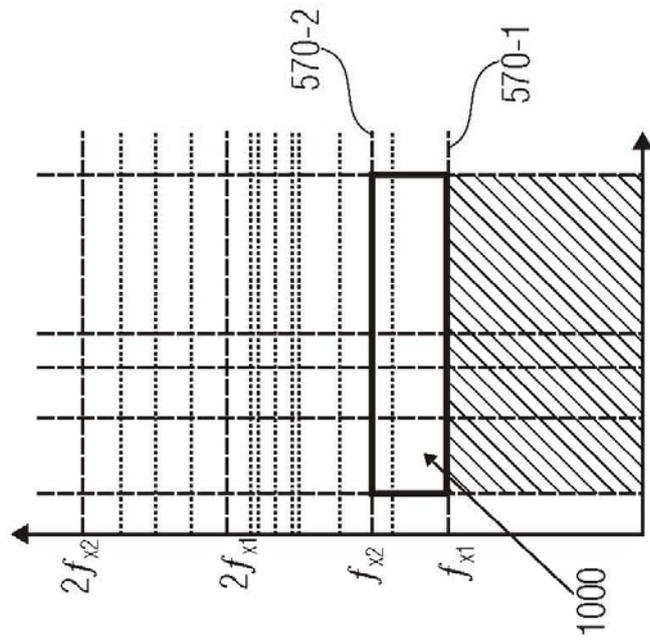


FIG 9C

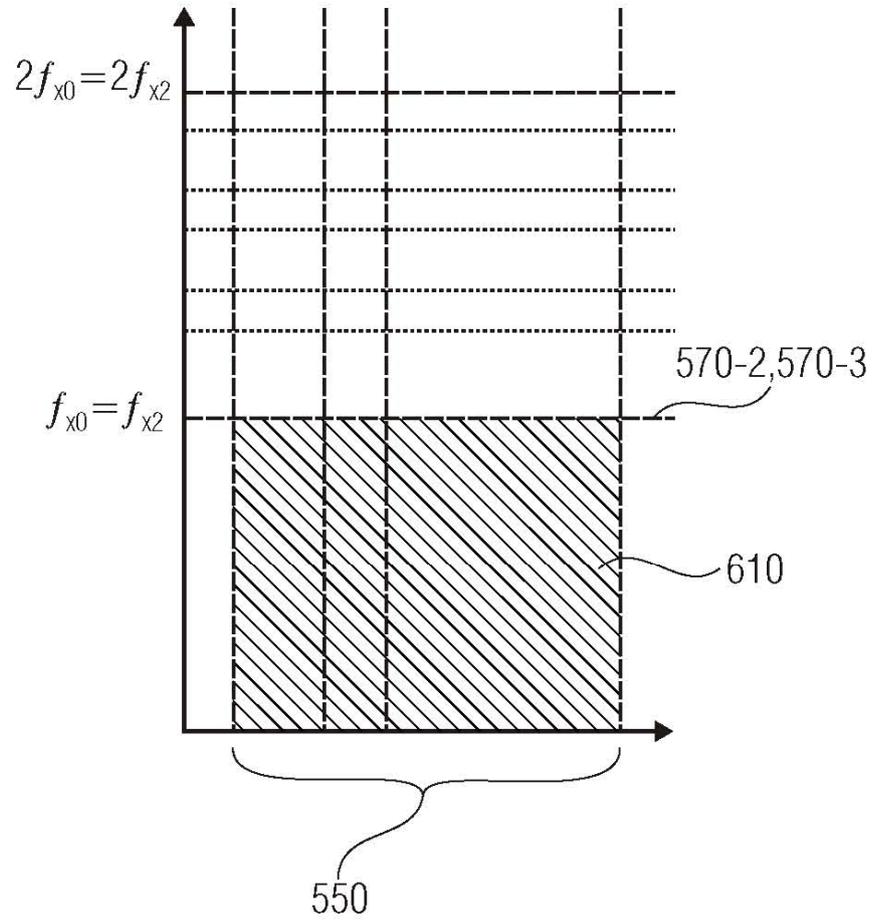


FIG 9E

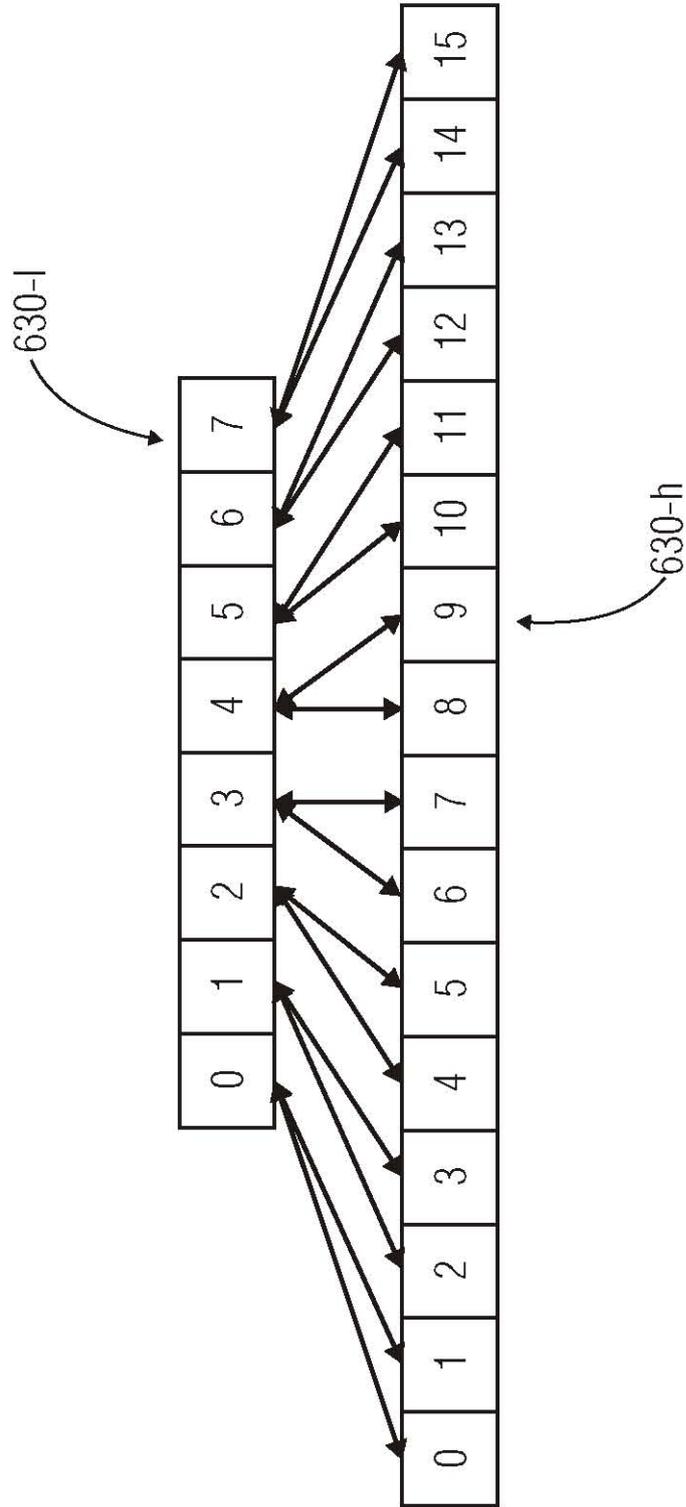


FIG 10