

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **3 028 541**

51 Int. Cl.:

G10L 19/012 (2013.01)

G10L 19/008 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **30.06.2021 PCT/EP2021/068079**

87 Fecha y número de publicación internacional: **03.03.2022 WO22042908**

96 Fecha de presentación y número de la solicitud europea: **30.06.2021 E 21739085 (5)**

97 Fecha y número de publicación de la concesión europea: **23.04.2025 EP 4205107**

54 Título: **Generador de señales multicanal, codificador de audio y procedimientos relacionados que se basan en una señal de ruido de mezcla**

30 Prioridad:

31.08.2020 EP 20193716

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

19.06.2025

73 Titular/es:

**FRAUNHOFER-GESELLSCHAFT ZUR
FÖRDERUNG DER ANGEWANDTEN
FORSCHUNG E.V. (100.00%)
Hansastraße 27c
80686 München, DE**

72 Inventor/es:

**KIENE, JAN FREDERIK;
FUCHS, GUILLAUME;
KORSE, SRIKANTH;
MULTRUS, MARKUS y
FOTOPOULOU, ELENI**

74 Agente/Representante:

PONTI & PARTNERS, S.L.P.

ES 3 028 541 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

Generador de señales multicanal, codificador de audio y procedimientos relacionados que se basan en una señal de ruido de mezcla

5

Campo de la Invención

[0001] La presente invención se refiere, entre otros, a la Generación de Ruido de Confort (CNG) para permitir Transmisión Discontinua (DTX) en Códecs Estéreo. La invención también se refiere a un generador de señales multi-
10 canal, un codificador de audio y procedimientos relacionados, por ejemplo, que se basan en una señal de ruido de mezcla. La invención se puede implementar en un dispositivo, un aparato, un sistema, en un procedimiento, en una unidad de almacenamiento no transitorio que almacena instrucciones que, cuando se ejecutan por un ordenador (pro-
cesador, controlador) hacen que el ordenador (procesador, controlador) realice un procedimiento particular, y en una señal de audio multicanal codificada.

15

Antecedentes de la Invención**Introducción**

[0002] Los generadores de ruido de confort se usan habitualmente en la transmisión discontinua (DTX) de señales de audio, en particular de señales de audio que contienen voz. De este modo, la señal de audio se clasifica primero en tramas activas e inactivas mediante un detector de actividad de voz (VAD). Con base en el resultado VAD, solo las tramas de voz activas se codifican y transmiten a la velocidad de bits nominal. Durante pausas largas, donde solo está presente el ruido de fondo, la velocidad de bits se reduce o pone a cero y el ruido de fondo se codifica
25 paraméricamente usando tramas de descriptor de inserción de silencio (tramas SID). La velocidad de bits promedio se reduce entonces significativamente.

[0003] El ruido se genera durante las tramas inactivas en el lado de decodificador por un generador de ruido de confort (CNG). El tamaño de una trama SID es muy limitado en la práctica. Por lo tanto, el número de parámetros que describen el ruido de fondo se debe mantener lo más pequeño posible. Con este fin, la estimación de ruido no se aplica directamente en la salida de las transformadas espectrales. En su lugar, se aplica a una resolución espectral más baja al promediar el espectro de potencia de entrada entre grupos de bandas, por ejemplo, siguiendo la escala de Bark. El promediado se puede lograr ya sea por medios aritméticos o geométricos. Desafortunadamente, el número limitado de parámetros transmitidos en las tramas SID no permite capturar la estructura espectral fina del ruido de
30 fondo. Por lo tanto, solo la envolvente espectral suave del ruido se puede reproducir por el CNG. Cuando el VAD dispara una trama CNG, la discrepancia entre el espectro suave del ruido de confort reconstruido y el espectro del ruido de fondo real se puede hacer muy audible en las transiciones entre tramas activas (que implican codificación y decodificación regulares de una porción de voz ruidosa de la señal) y tramas CNG.

[0004] Algunas tecnologías típicas CNG se pueden encontrar en las Recomendaciones ITU-T G.729B [1], G.729.1C [2], G.718 [3], o en las Especificaciones 3GPP para AMR [4] y AMR-WB [5]. Todas estas tecnologías generan ruido de confort (CN) al usar la estrategia de análisis/síntesis que hace uso de predicción lineal (LP).

[0005] Para reducir aún más la velocidad de transmisión, el códec de telecomunicaciones 3GPP para los Ser-
45 vicios de Voz Mejorados (EVS) de LTE [6] está equipado con un modo de Transmisión Discontinua (DTX) que aplica Generación de Ruido de Confort (CNG) para tramas inactivas, es decir, tramas que se determina que consisten solo en ruido de fondo. Para estas tramas, una representación paramétrica de baja velocidad de la señal se transporta por tramas de Descriptor de Inserción de Silencio (SID) como máximo cada 8 tramas (160 ms). Esto permite que el CNG en el decodificador produzca una señal de ruido artificial que se asemeja al ruido de fondo real. En EVS, el CNG se puede lograr utilizando un esquema predictivo lineal (LP-CNG) o un esquema de dominio de frecuencia (FD-CNG),
50 dependiendo de las características espectrales del ruido de fondo.

[0006] La estrategia de LP-CNG en EVS [7] opera sobre una base de banda dividida con la codificación que consiste tanto en una etapa de codificación de análisis/síntesis de banda baja como de banda alta. En contraste con
55 la codificación de banda baja, no se realiza ningún modelado de parámetros del espectro de ruido de banda alta para la señal de banda alta. Solo la energía de la señal de banda alta se codifica y transmite al decodificador y el espectro de ruido de banda alta se genera puramente en el lado de decodificador. Tanto el CN de banda baja como de banda alta se sintetizan filtrando una excitación a través de un filtro de síntesis. La excitación de banda baja se deriva de la energía de excitación de banda baja recibida y la envolvente de frecuencia de excitación de banda baja. El filtro de síntesis de banda baja se deriva de los parámetros LP recibidos en forma de coeficientes de frecuencia espectral de línea (LSF). La excitación de banda alta se obtiene usando energía que se extrapola a partir de la energía de banda baja y el filtro de síntesis de banda alta se deriva de una interpolación LSF de lado decodificador. La síntesis de banda alta se invierte espectralmente y se añade a la síntesis de banda baja para formar la señal CN final.

60

[0007] La estrategia de FD-CNG [8] [9], hace uso de un algoritmo de estimación de ruido en el dominio de la frecuencia seguido de una cuantificación vectorial de la envolvente espectral suavizada del ruido de fondo. La envolvente decodificada se refina en el decodificador al ejecutar un segundo estimador de ruido en el dominio de la frecuencia. Como se usa una representación puramente paramétrica durante las tramas inactivas, la señal de ruido no está disponible en el decodificador en este caso. En FD-CNG, la estimación de ruido se realiza en cada trama (activa e inactiva) en los lados de codificador y decodificador con base en el algoritmo estadístico mínimo.

[0008] Un procedimiento para generar ruido de confort en el caso de dos (o más) canales se describe en [10]. En [10], se describe un sistema para DTX estéreo y CNG que combina un SID mono con una medición de coherencia de banda calculada en los dos canales estéreo de entrada en el codificador. En el decodificador, la información CNG mono y los valores de coherencia se decodifican a partir del flujo de bits y se sintetiza la coherencia objetivo en varias bandas de frecuencia. Para reducir la velocidad de bits de la trama SID estéreo resultante, los valores de coherencia se codifican usando un esquema predictivo seguido de una codificación entrópica con velocidad de bits variable. El ruido de confort se genera para cada canal con los procedimientos descritos en los párrafos anteriores y, a continuación, las dos CN se mezclan en banda utilizando una fórmula con ponderación basada en los valores de coherencia de banda transmitidos incluidos en la trama SID.

Motivación/Inconvenientes de la Técnica Anterior

[0009] En un sistema estéreo, generar el ruido de fondo por separado conduce a un ruido completamente no correlacionado que suena desagradable y es muy diferente del ruido de fondo real que causa transiciones audibles abruptas cuando cambiamos a/desde el fondo de modo activo a los fondos de modo DTX. Además, no es posible preservar la imagen estéreo del fondo utilizando solo dos fuentes de ruido completamente no correlacionadas. Por último, si hay una fuente de ruido de fondo y el hablante se está moviendo con un dispositivo de mano alrededor de la fuente, la imagen espacial del ruido de fondo cambiará con el tiempo, algo que no se podría replicar al reconstruir el ruido de fondo para cada canal de forma independiente. Por lo tanto, es necesario desarrollar una nueva estrategia para adaptarse al problema de las señales estereofónicas.

[0010] Esto también se aborda en [10], sin embargo, en realizaciones, la inserción de una fuente de ruido común para los dos canales para imitar el ruido correlacionado para generar el ruido de confort final juega un papel importante en la imitación de grabación de ruido de fondo estereofónico.

[0011] Los códecs de voz de comunicación actuales típicamente solo codifican señales mono. Por lo tanto, la mayoría de los sistemas DTX existentes están diseñados para CNG mono. Simplemente aplicar la operación DTX independientemente en ambos canales de una señal estéreo parece sencillo, pero incluye varios problemas. En primer lugar, este enfoque requiere la transmisión de dos conjuntos de parámetros que describen las dos señales de ruido de fondo en los dos canales. Esto incrementaría la velocidad de transmisión de datos necesaria para la transmisión de tramas SID, lo que disminuye el beneficio de la reducción de carga en la red. Otro aspecto problemático radica en la decisión VAD, que se debe sincronizar entre los canales para evitar rarezas y distorsiones de la imagen espacial de la señal estéreo y también para optimizar la reducción de la velocidad de bits del sistema. Además, cuando se aplica CNG en el lado de receptor independientemente en ambos canales, los dos algoritmos CNG independientes producirán típicamente dos señales de ruido aleatorias con coherencia cero o muy baja. Esto dará como resultado una imagen estéreo muy amplia en el ruido de confort generado. Por otro lado, solo aplicar en el generador de ruido y usar la misma señal de ruido de confort en ambos canales conduce a una coherencia muy alta y una imagen estéreo muy estrecha. Para la mayoría de las señales estéreo, sin embargo, la imagen estéreo y su impresión espacial estarán en algún lugar entre estos dos extremos. Por lo tanto, la conmutación a o desde tramas activas al modo DTX introduciría transiciones audibles abruptas. También, si hay una fuente de ruido de fondo y el hablante se está moviendo con un dispositivo de mano alrededor de la fuente, la imagen espacial del ruido de fondo cambiará con el tiempo, algo que no se podría replicar al reconstruir el ruido de fondo para cada canal de forma independiente. Por lo tanto, es necesaria una nueva estrategia para adaptarse al problema de las señales estereofónicas.

[0012] El sistema descrito en [10] abordó estos problemas al transmitir información para CNG mono junto con valores de parámetros que se utilizan para volver a sintetizar la imagen estéreo del ruido de fondo en el decodificador. Este tipo de sistema DTX se ajusta bien para codificadores estéreo paramétricos que aplican una mezcla descendente a los dos canales de entrada antes de la codificación y transmisión a partir de la cual se pueden derivar los parámetros mono CNG. Sin embargo, en un esquema de codificación estéreo discreto normalmente todavía se codifican dos canales de manera conjunta y normalmente no se derivan parámetros de mezcla ascendente como una medición de coherencia refinada. Por lo tanto, para este tipo de codificadores estéreo, se necesita una estrategia diferente.

[0013] El documento US 2016/027447 A1 describe una técnica para generar ruido de confort.

[0014] DANIELE MIRABILII y col., "Simulación de ruido de viento multicanal basado en el modelo Corcos", ARXIV.ORG, 40 CORNELL UNIVERSITY LIBRARY, 201 OLIN LIBRARY CORNELL UNIVERSITY ITHACA, NY 14853, 24 de mayo de 2018, el documento XP081555728, describe un generador de ruido de viento artificial multicanal

basado en un modelo de dinámica de fluidos.

[0015] El documento US 2017/047072 A1 describe un procedimiento para generar ruido de confort.

5 [0016] El documento WO 2019/193156 A1 describe una técnica para generar ruido de confort.

Aspectos de la Presente Invención

[0017] Los presentes ejemplos proporcionan una transmisión eficiente de señales de voz estéreo. La transmisión de una señal estéreo puede mejorar la experiencia del usuario y la inteligibilidad del voz a través de la transmisión de un solo canal de audio (mono), especialmente en situaciones con ruido de fondo impuesto u otros sonidos. Las señales estéreo se pueden codificar de una manera paramétrica donde se aplica una mezcla descendente mono de los dos canales estéreo y este único canal de mezcla descendente se codifica y transmite al receptor junto con información adicional que se utiliza para aproximar la señal estéreo original en el decodificador. Otra estrategia es emplear codificación estéreo discreta que tiene como objetivo eliminar redundancia entre los canales para lograr una representación de dos canales más compacta de la señal original por medio de algún pre-procesamiento de señal. Los dos canales procesados entonces se codifican y transmiten. En el decodificador, se aplica un procesamiento inverso. Aun así, la información adicional relevante para el procesamiento estéreo se puede transmitir a lo largo de los dos canales. Por lo tanto, la principal diferencia entre los procedimientos de codificación estéreo paramétricos y discretos está en el número de canales transmitidos.

[0018] Por lo general, en una conversación hay períodos en los que no todos los oradores están hablando activamente. Por lo tanto, la señal de entrada a un codificador de voz en estos períodos consiste principalmente en ruido de fondo o (casi) silencio. Para ahorrar velocidad de datos y reducir la carga en la red de transmisión, los codificadores de voz intentan distinguir entre tramas que contienen voz (tramas activas) y tramas que contienen principalmente ruido de fondo o silencio (tramas inactivas). Para tramas inactivas, la velocidad de datos se puede reducir significativamente al no codificar la señal de audio como en las tramas activas, sino derivando una descripción paramétrica de baja velocidad de bits del ruido de fondo actual en forma de una trama de Descriptor de Inserción de Silencio (SID). Esta trama SID se transmite periódicamente al decodificador para actualizar los parámetros que describen el ruido de fondo, en tanto que para las tramas inactivas entre la velocidad de bits se reduce o incluso no se transmite información. En el decodificador, el ruido de fondo se remodela usando los parámetros transmitidos en la trama SID por un algoritmo de Generación de Ruido de Confort (CNG). De esta manera, la velocidad de transmisión se puede reducir o incluso poner a cero para tramas inactivas sin que el usuario lo interprete como una interrupción o final de la conexión.

[0019] Describimos un sistema DTX para señales estéreo codificadas discretamente que consiste en un SID estéreo y un procedimiento para CNG que genera un ruido de confort estéreo al modelar las características espectrales del ruido de fondo en ambos canales, así como el grado de correlación entre ellos, en tanto que se mantiene la velocidad de bits de bits promedio comparable a las aplicaciones mono.

Resumen

[0020] La invención se define en las reivindicaciones independientes.

Figuras

[0021]

- La Fig. 1 muestra un ejemplo en un codificador, en particular para clasificar una trama como activa o inactiva.
- La Fig. 2 muestra un ejemplo de un codificador y un decodificador.
- Las Fig. 3a-3f muestran ejemplos de generadores de señal multicanal, que se pueden usar en un decodificador.
- La Fig. 4 muestra un ejemplo de un codificador y un decodificador.
- La Fig. 5 muestra un ejemplo de una Etapa de Cuantificación de Parámetros de Ruido.
- La Fig. 6 muestra un ejemplo de una Etapa de Descuantificación de Parámetros de Ruido.

Algunos aspectos que se pueden implementar en los ejemplos

[0022] En el presente documento, se describe, entre otras cosas, una nueva técnica, por ejemplo, para DTX y CNG para señales estéreo codificadas discretamente. En lugar de operar en una mezcla descendente mono de la señal estéreo, los parámetros de ruido para ambos canales se derivan, codifican y transmiten conjuntamente. En el decodificador (o más en general en un generador multicanal), se pueden mezclar tres señales de ruido de confort independientes con base en un solo valor de coherencia intercanal de banda ancha que se transmite, por ejemplo, a lo largo de los dos conjuntos de parámetros de ruido. Algunos de los aspectos de los ejemplos pueden cubrir, en algunos ejemplos, al menos uno de los siguientes aspectos:

- CNG en el decodificador al mezclar, por ejemplo, tres señales de ruido independientes. Después de decodificar el SID estéreo y reconstruir los parámetros de ruido para el canal izquierdo y derecho, se pueden generar dos señales de ruido, por ejemplo, como una mezcla de ruido correlacionado y no correlacionado. Para esto, se pueden mezclar conjuntamente una fuente de ruido común para ambos canales (que sirve como la fuente de ruido correlacionada) y dos fuentes de ruido individuales (que proporcionan ruido no correlacionado). El proceso de mezcla se puede controlar mediante el valor de coherencia entre canales transmitido en el SID estéreo. Después de la mezcla, las dos señales de ruido mezcladas se forman espectralmente usando los parámetros de ruido reconstruidos para los canales izquierdo y derecho, respectivamente.
- La codificación conjunta de los parámetros de ruido se puede derivar de los dos canales de una señal estéreo. Para mantener baja la velocidad de bits del SID estéreo, los parámetros de ruido pueden comprimirse adicionalmente antes de codificarlos en el SID estéreo. Esto se puede lograr, por ejemplo, convirtiendo la representación del canal izquierdo/derecho de los parámetros de ruido en una representación media/lateral y codificando los parámetros de ruido lateral con un número menor de bits que los parámetros de ruido medio.
- Un SID para DTX de dos canales (SID estéreo). Este SID puede contener parámetros de ruido para ambos canales de una señal estéreo junto con un solo valor de coherencia intercanal de banda ancha y una bandera que indica parámetros de ruido iguales para ambos canales.

[0023] Se mostrará que los ejemplos más adelante se pueden implementar en dispositivos, aparatos, sistemas, procedimientos, controladores y unidades de almacenamiento no transitorio que almacenan instrucciones que, cuando se ejecutan por un procesador, hacen que el procesador lleve a cabo las técnicas descritas (por ejemplo, procedimientos, como secuencias de operaciones).

[0024] En particular, al menos uno de los bloques más adelante se puede controlar mediante un controlador.

25 **Ejemplos**

[0025] Antes de analizar en detalle los aspectos de los presentes ejemplos, se proporciona una descripción general rápida de algunos de los más importantes:

1) Las Figs. 3a-3f muestran ejemplos de generadores de señales multicanal (por ejemplo, formados por al menos una primera señal, o canal, y una segunda señal de audio, o canal), que generan una señal de audio multicanal (por ejemplo, en un decodificador). La señal de audio multicanal (originalmente en forma de múltiples canales decorrelacionados) se puede ver influenciada (por ejemplo, modificada en escala) por uno o más elementos de amplitud. La cantidad de influencia se puede basar en unos datos de coherencia entre la primera y segunda señales de audio como se estima en el codificador. La primera y segunda señales de audio se pueden someter a mezcla con una señal de mezcla común (que también se puede decorrelacionar e influenciar, por ejemplo, modificada en escala, por los datos de coherencia). La cantidad de influencia para la señal de mezcla puede ser de modo que la primera y la segunda señales de audio se modifiquen en escala por un alto factor de ponderación (por ejemplo, 1 o menos de, pero por ejemplo, cerca de, 1) cuando la señal de mezcla se modifica en escala por un bajo factor de ponderación (por ejemplo, 0 o más de, pero por ejemplo, cerca de, 0), y viceversa. La cantidad de influencia para la señal de mezcla puede ser de modo que una alta coherencia como se mide en el codificador hace que la primera y segunda señales de audio se modifiquen en escala por un bajo factor de ponderación (por ejemplo, 0 o más de, pero por ejemplo, cerca de, 0), y una alta coherencia como se mide en el codificador hace que la primera y segunda señales de audio se modifiquen en escala por un alto factor de ponderación (por ejemplo, 1 o menos de, pero por ejemplo, cerca de, 1). Las técnicas de Las Figs. 3A-3F se puede utilizar para implementar un generador de ruido de confort (CNG).

1) Las Figs. 1, 2 y 4 muestran ejemplos de codificadores. Un codificador puede clasificar una trama de audio como activa o inactiva. Si la trama de audio está inactiva, entonces solo algunos datos de ruido paramétricos se codifican en el flujo de bits (por ejemplo, para proporcionar una forma de ruido paramétrica, que da una representación paramétrica de la forma del ruido, sin la necesidad de proporcionar la propia señal de ruido), y también se pueden proporcionar datos de coherencia entre los dos canales.

2) Las Figs. 2 y 4 muestran ejemplos de decodificadores. Un decodificador puede generar una señal de audio (ruido de confort), por ejemplo:

- a. utilizando una de las técnicas mostradas en las Figs. 3A-3F (punto 1) anterior) (en particular teniendo en cuenta el valor de coherencia proporcionado por el codificador y aplicándolo como factor de ponderación en el elemento o elementos de amplitud); y
- b. conformar la señal de audio generada (ruido de confort) utilizando los datos de ruido paramétrico codificados en el flujo de bits.

[0026] En particular, no es necesario que el codificador proporcione la señal de audio completa para la trama inactiva, sino solo el valor de coherencia y la representación paramétrica de la forma de ruido, reduciendo así la cantidad de bits que se van a codificar en el flujo de bits.

Generador de señales (por ejemplo, lado de decodificador), CNG

- 5 **[0027]** Las Figs. 3a-3f muestran ejemplos de un CNG, o más en general un generador de señales multicanal 200, para generar una señal multicanal 204 que tiene un primer canal 201 y un segundo canal 203. (En la presente descripción, las señales de audio generadas 221 y 223 se consideran ruido, pero también son posibles diferentes tipos de señales que no son ruido). Se hace referencia inicialmente a la Fig. 3f, que es general, en tanto que Las Figs. 3a-3e muestran ejemplos particulares.
- 10 **[0028]** Una primera fuente de audio 211 puede ser una primera fuente de ruido y se puede indicar aquí para generar la primera señal de audio 221, que puede ser una primera señal de ruido. La fuente de ruido de mezcla 212 puede generar una señal de ruido de mezcla 222. La segunda fuente de audio 213 puede generar una segunda señal de audio 223 que puede ser una segunda señal de ruido. El generador de señales multicanal 200 puede mezclar la primera señal de audio (primera señal de ruido) 221 con la señal de ruido de mezcla 222 y la segunda señal de audio (segunda señal de ruido) 223 con la señal de ruido de mezcla 222. (Además o como alternativa, la primera señal de audio 221 se puede mezclar con una versión 221a de la señal de ruido de mezcla 222, y la segunda señal de audio 223 se puede mezclar con una versión 221b de la señal de ruido de mezcla 222, donde las versiones 221a y 221b pueden diferir, por ejemplo, en un 20% entre sí; cada una de las versiones 221a y 221b puede ser, por ejemplo, una versión aumentada y/o reducida de una señal común 222). Por consiguiente, se puede obtener un primer canal 201 de la señal multicanal 204 a partir de la primera señal de audio (primera señal de ruido) 221 y la señal de ruido de mezcla 222. De manera análoga, el segundo canal 203 de la señal multicanal 204 se puede obtener de la segunda señal de audio 223 mezclada con la señal de ruido de mezcla 222. También se observa que las señales pueden estar aquí en el dominio de frecuencia, y k se refiere al índice o coeficiente particular (asociado con un intervalo de frecuencia particular).
- 25 **[0029]** Como se puede ver en las Figs 3a-3f, la primera señal de audio 221, la señal de ruido de mezcla 222 y la segunda señal de audio 223 pueden estar decorrelacionadas entre sí. Esto se puede obtener, por ejemplo, al decorrelacionar la misma señal (por ejemplo, en un decorrelacionador) y/o al generar independientemente ruido (se proporcionan ejemplos más adelante).
- 30 **[0030]** Se puede implementar un mezclador 206 para mezclar la primera señal de audio 221 y la segunda señal de audio 223 con la señal de ruido de mezcla 222. La mezcla puede ser del tipo de suma de señales (por ejemplo, en las etapas de sumador 206-1 y 206-3) después de que la primera señal de audio 221, la señal de ruido de mezcla 222 y la segunda señal de audio 223 se hayan ponderado mediante modificación de escala (por ejemplo, en los elementos de amplitud 208-1, 208-2, 208-3). La mezcla es del tipo "suma después de ponderación". Las Figs. 3a-3f muestran el procesamiento de señal real que se aplica para generar las señales de ruido $N_i[k]$ y $N_r[k]$ con el elemento de adición (+) que indica la adición por muestra de dos señales (k es el índice del intervalo de frecuencia).
- 40 **[0031]** Los elementos de amplitud (o elementos de ponderación o elementos de modificación de escala) 208-1, 208-2 y 208-3 se pueden obtener, por ejemplo, al modificar la escala de la primera señal de audio 221, la señal de ruido de mezcla 222 y la segunda señal de audio 223 mediante coeficientes adecuados, y pueden emitir una versión ponderada 221' de la primera señal de audio 221, una versión ponderada 222' de la señal de ruido de mezcla 222 y una versión ponderada 223' de la segunda señal de audio 223. Los coeficientes adecuados pueden ser $\sqrt{\text{coh}}$ y $\sqrt{1-\text{coh}}$ y se pueden obtener, por ejemplo, a partir de información de coherencia codificada en la señalización de una trama de descriptor particular (ver también más adelante) ($\sqrt{}$ se refiere aquí a la operación de raíz cuadrada).
- 45 La coherencia "coh" se analiza más adelante en detalle, y puede ser, por ejemplo, la indicada con "c" o "c_{ind}" o "c_q" más adelante, por ejemplo, codificada en una información de coherencia 404 de un flujo de bits 232 (ver más adelante, en combinación con Las Figs. 2 y 4). Particularmente, la señal de ruido de mezcla 222 se puede someter, por ejemplo, a una modificación de escala por un factor de ponderación que es una raíz cuadrada de un valor de coherencia, en tanto que la primera señal de audio 221 y la segunda señal de audio 222 se pueden modificar en escala por un factor de ponderación que es la raíz cuadrada del valor complementario a uno de la coherencia coh. No obstante, la señal de ruido de mezcla 222 se puede considerar como una señal de modo común, una porción de la cual se mezcla con la versión ponderada 221' de la primera señal de audio 221 y la versión ponderada 223' de la segunda señal de audio 223 para obtener el primer canal 201 de la señal multicanal 204 y el segundo canal 203 de la señal multicanal 204, respectivamente. En algunos casos, la primera fuente de ruido 211 o la segunda fuente de ruido 213 se puede configurar para generar la primera señal de ruido 221 o la segunda señal de ruido 223 de modo que la primera señal de ruido 221 y/o la segunda señal de ruido 223 se decorrelaciona de la señal de ruido de mezcla 222 (ver más adelante con referencia a Las Figs. 3b-3e).
- 50 **[0032]** Al menos una (o cada una de) la primera fuente de audio 211, la segunda fuente de audio 213 y la fuente de ruido de mezcla 212) puede ser una fuente de ruido gaussiano.
- 60 **[0033]** En el ejemplo de la Fig. 3a, la primera fuente de audio 211 (aquí indicada con 211a) puede comprender o estar conectada a un primer generador de ruido, y la segunda fuente de audio 213 (213a) puede comprender o estar conectada a un segundo generador de ruido. La fuente de ruido de mezcla 212 (212a) puede comprender o conectarse

a un tercer generador de ruido. El primer generador de ruido 211 (211a), el segundo generador de ruido 213 (213a) y el tercer generador de ruido 212 (212a) pueden generar señales de ruido mutuamente decorrelacionadas.

5 **[0034]** En ejemplos, al menos una de la primera fuente de audio 211 (211a), la segunda fuente de audio 213 (213a) y la fuente de ruido de mezcla 212 (212a) puede operar usando una tabla de ruido prealmacenada, que, por lo tanto, puede proporcionar una secuencia aleatoria.

10 **[0035]** En algunos ejemplos, al menos una de la primera fuente de audio 211, la segunda fuente de audio 213 y la fuente de ruido de mezcla 212 puede generar un espectro complejo para una trama usando un primer valor de ruido para una parte real y un segundo valor de ruido para una parte imaginaria. Opcionalmente, el al menos un generador de ruido puede generar un valor espectral de ruido complejo (por ejemplo, coeficiente) para un intervalo de frecuencia k usando para una de la parte real y la parte imaginaria, un primer valor aleatorio en un índice k y usando, para la otra de la parte real y la parte imaginaria, un segundo valor aleatorio en un índice $(k+M)$. El primer valor de ruido y el segundo valor de ruido se pueden incluir en una matriz de ruido, por ejemplo, derivada de un generador de secuencia de números aleatorios o una tabla de ruido o un proceso de ruido, que varía de un índice de inicio a un índice de fin, el índice de inicio es menor que M , y el índice de fin es igual o menor que $2 \times M$ (que es el doble de M). M y k pueden ser números enteros (k que es el índice del intervalo de frecuencia de bit particular en la representación de dominio de frecuencia de la señal).

20 **[0036]** Cada fuente de audio 211, 212, 213 puede incluir al menos un generador de fuente de audio (generador de ruido) que genera el ruido, por ejemplo, en términos de $N_1[k]$, $N_2[k]$, $N_3[k]$.

25 **[0037]** El generador de señales multicanal 200 de Las Figs. 3A-3F se puede utilizar, por ejemplo, para un decodificador 200a, 200B (200'). En particular, el generador de señales multicanal 200 se puede ver como parte del generador de ruido de confort (CNG) 220 en la Fig. 4. El decodificador 200 se puede usar en general para decodificar señales que se han codificado por un codificador, o al generar señales que se van a conformar mediante información de energía obtenida de un flujo de bits, para generar una señal de audio que corresponde a una entrada de señal de audio de entrada original al codificador. En algunos ejemplos, hay una clasificación entre las tramas con voz (o en general señales de audio no vacías) y tramas de descriptor de inserción de silencio. Como se explicó anteriormente y 30 más adelante, las tramas de descriptor de inserción de silencio (SID) (las denominadas "tramas inactivas 308", que se pueden codificar como tramas SID 241 y/o 243, por ejemplo) se proporcionan en general por debajo de la información de velocidad de bits y, por lo tanto, se proporcionan con menos frecuencia que las tramas de voz normales (las denominadas "tramas activas 306", ver también más adelante). Además, la información que está presente en las tramas de descriptor de inserción de silencio (SID, tramas inactivas 308) es en general limitada (y puede corresponder sustancialmente a información de energía sobre la señal).

35 **[0038]** No obstante, se ha entendido que es posible complementar el contenido de las tramas SID con el ruido multicanal 204 generado por el generador de señales multicanal. Básicamente, las fuentes de audio 211, 212, 213 pueden procesar señales (por ejemplo, ruido) que pueden ser independientes y no correlacionadas entre sí. La primera señal de audio 221, la señal de ruido de mezcla 222 y la segunda señal de audio 223 se pueden, no obstante, modificar en escala mediante información de coherencia proporcionada por el codificador e insertada en el flujo de bits. Como se puede ver en la Fig. 3A-3F, el valor de coherencia puede ser el mismo que el de la señal de ruido de mezcla 222 que proporciona una señal de modo común tanto a la primera señal de audio 221 como a la segunda señal de audio 223, lo que permite obtener el primer canal 201 y el segundo canal 203 de la señal multicanal 204. En general, la señal 45 de coherencia es un valor entre 0 y 1:

- Coherencia igual a 0 significa que el primer canal de audio original (por ejemplo, L, 301) y el segundo canal de audio (por ejemplo, R, 303) no están totalmente correlacionados entre sí, y el elemento de amplitud 208-2 de la señal de ruido de mezcla 222 modificará en escala por 0 la señal de ruido de mezcla 222, lo que hará que la primera señal de audio 221 y la segunda señal de audio 223 no se mezclen con ninguna señal de modo común (al mezclarse con la señal que es constantemente 0), y los canales de salida 201, 203 serán sustancialmente los mismos que la primera señal de ruido 221 y la segunda señal de ruido 223 de la señal multicanal 204.

50 - Coherencia igual a 1 significa que el primer canal de audio original (por ejemplo, L, 301) y el segundo canal de audio (por ejemplo, R, 303) serán iguales, y los elementos de amplitud 208-1 y 208-3 modificarán en escala por 0 las señales de entrada, y el primer y segundo canales son entonces iguales a la señal de ruido de mezcla 222 (que se escala en 1 en el elemento de amplitud 208-2).

- Coherencias intermedias entre 0 y 1 causarán mezclas intermedias entre las dos situaciones anteriores.

60 **[0039]** A continuación se analizan algunos aspectos y variantes del mezclador 206 y/o el CNG 220.

[0040] La primera fuente de audio (211) puede ser una primera fuente de ruido y la primera señal de audio (221) puede ser una primera señal de ruido, o la segunda fuente de audio (213) es una segunda fuente de ruido y la segunda señal de audio (223) es una segunda señal de ruido. La primera fuente de ruido (211) o la segunda fuente de ruido (213) se puede configurar para generar la primera señal de ruido (221) o la segunda señal de ruido (223), de

modo que la primera señal de ruido (221) o la segunda señal de ruido (223) se decorrelaciona de la señal de ruido de mezcla (222).

El mezclador (206) se puede configurar para generar el primer canal (201) y el segundo canal (203) de modo que la cantidad de la señal de ruido de mezcla (222) en el primer canal (201) es igual a la cantidad de la señal de ruido de mezcla (222) en el segundo canal (203), o está dentro de un intervalo de 80 por ciento a 120 por ciento de la cantidad de la señal de ruido de mezcla (222) en el segundo canal (203) (por ejemplo, sus porciones 221a y 221b son diferentes dentro de un intervalo de 80 por ciento a 120 por ciento entre sí y de la señal de ruido de mezcla original 222).

[0041] En algunos casos,

10

la cantidad de influencia realizada por el primer elemento de amplitud (208-1) y la cantidad de influencia realizada por el segundo elemento de amplitud (208-3) son iguales entre sí (por ejemplo, cuando no hay distinción entre las porciones 221a y 221b), o

15

la cantidad de influencia realizada por el segundo elemento de amplitud (208-3) es diferente en menos del 20 por ciento de la cantidad realizada por el primer elemento de amplitud (208-1) (por ejemplo, cuando la diferencia entre las porciones 221a y 221b es menor del 20%).

[0042] El mezclador (206) y/o el CNG 220 pueden comprender una entrada de control para recibir un parámetro de control (404, c). Por lo tanto, el mezclador (206) se puede configurar para controlar la cantidad de la señal de ruido de mezcla (222) en el primer canal (201) y el segundo canal (203) en respuesta al parámetro de control (404, c).

20

[0043] En Las Figs. 3a-3f, se muestra que la señal de ruido de mezcla 222 se somete a un coeficiente $\sqrt{\text{coh}}$, y la primera y segunda señales de audio 221, 223 se someten a un coeficiente $\sqrt{1-\text{coh}}$.

25

[0044] Como se explicó anteriormente, la Fig. 3a muestra un CNG 220a en el que la primera fuente 211a (211), la segunda fuente 213a (213) y la fuente de ruido de mezcla 212a (212) comprenden diferentes generadores. Esto no es estrictamente necesario, y son posibles varias variantes.

[0045] Más en general:

30

1. 1ª variante CNG 220b, (figura 3b):

a. la primera fuente de audio 211b (211) puede comprender un primer generador de ruido para generar la primera señal de audio (221) como una primera señal de ruido,

35

b. la segunda fuente de audio 213b (213) puede comprender un decorrelacionador para decorrelacionar la primera señal de ruido (221) para generar la segunda señal de audio (213) como una segunda señal de ruido (por ejemplo, la segunda señal de audio que se obtiene a partir de la primera señal de audio después de una decorrelación), y

40

c. la fuente de ruido de mezcla 212b (212) puede comprender un segundo generador de ruido (que no está correlacionado de forma nativa con el primer generador de ruido);

2. 2ª variante CNG 220c (figura 3c):

a. la primera fuente de audio 211c (211) puede comprender un primer generador de ruido para generar la primera señal de audio (221) como una primera señal de ruido,

45

b. la segunda fuente de audio 213c (213) puede comprender un segundo generador de ruido para generar la segunda señal de audio (223) como una segunda señal de ruido (por ejemplo, el segundo generador de ruido que no está correlacionado de forma nativa con el primer generador de ruido), y

50

c. la fuente de ruido de mezcla 212c (212) puede comprender un decorrelacionador para decorrelacionar la primera señal de ruido (221) o la segunda señal de ruido (223) para generar la señal de ruido de mezcla (222);

3. 3ª variante CNG 220d (figuras 3d y 3e):

a. una de la primera fuente de audio 211d o 211e (211), la segunda fuente de audio 213d o 213e (213), y la fuente de ruido de mezcla 212d o 212e (212) puede comprender un generador de ruido para generar una señal de ruido,

55

b. otra de la primera fuente de audio 211d o 211e (211), la segunda fuente de audio 213d o 213e (213) y la fuente de ruido de mezcla 212d o 212e (212) puede comprender un primer decorrelacionador para decorrelacionar la señal de ruido, y

60

c. una más de la primera fuente de audio 211d o 211e (211), la segunda fuente de audio 213d o 213e (213) y la fuente de ruido de mezcla 212d o 212e (212) puede comprender un segundo decorrelacionador para decorrelacionar la señal de ruido,

d. el primer decorrelacionador y el segundo decorrelacionador pueden ser diferentes entre sí, de modo que las señales de salida del primer decorrelacionador y el segundo decorrelacionador se decorrelacionan entre sí;

4. 4ª variante CNG 220 (figura 3a):

- a. la primera fuente de audio 211a (211) comprende un primer generador de ruido,
- b. la segunda fuente de audio 213a (213) comprende un segundo generador de ruido,
- c. la fuente de ruido de mezcla 212a (212) comprende un tercer generador de ruido,
- d. el primer generador de ruido, el segundo generador de ruido y el tercer generador de ruido se pueden generar señales de ruido mutuamente decorrelacionadas (por ejemplo, los generadores de árbol que no están correlacionados de forma nativa entre sí).

5. 5ª variante:

- a. de la primera fuente de audio (211), la segunda fuente de audio (213) y la fuente de ruido de mezcla (212) pueden comprender un generador de secuencia de números pseudoaleatorios para generar una secuencia de números pseudoaleatorios en respuesta a una semilla,
- b. al menos dos de la primera fuente de audio (211), la segunda fuente de audio (213) y la fuente de ruido de mezcla (212) pueden inicializar el generador de secuencia de números pseudoaleatorios usando diferentes semillas.

6. 6ª variante:

- a. al menos una de la primera fuente de audio (211), la segunda fuente de audio (213) y la fuente de ruido de mezcla (212) puede operar usando una tabla de ruido prealmacenada,
- b. opcionalmente, al menos una de la primera fuente de audio (211), la segunda fuente de audio (213) y la fuente de ruido de mezcla (212) puede generar un espectro complejo para una trama usando un primer valor de ruido para una parte real y un segundo valor de ruido para una parte imaginaria
- c. opcionalmente, al menos un generador de ruido puede generar un valor espectral de ruido complejo para un intervalo de frecuencia k usando para una de la parte real y la parte imaginaria, un primer valor aleatorio en un índice k y usando, para la otra de la parte real y la parte imaginaria, un segundo valor aleatorio en un índice $(k+M)$ (el primer valor de ruido y el segundo valor de ruido se incluyen en una matriz de ruido, por ejemplo, derivada de un generador de secuencia de números aleatorios o una tabla de ruido o un proceso de ruido, que varía de un índice de inicio a un índice de fin, el índice de inicio que es menor que M , y el índice de fin que es igual o menor que $2 \times M$, M y k que son números enteros)

[0046] Como se puede ver en la Fig. 4, el decodificador 200' (200a, 200b) puede incluir, además del CNG 220 de la Fig. 3, también una interfaz de entrada 210 para recibir datos de audio codificados en una secuencia de tramas comprendiendo una trama activa y una trama inactiva después de la trama activa; y un decodificador de audio para decodificar datos de audio codificados para la trama activa para generar una señal multicanal decodificada para la trama activa, donde la primera fuente 211 de audio, la segunda fuente 213 de audio, la fuente 212 de ruido de mezcla y el mezclador 206 están activos en la trama inactiva para generar la señal multicanal para la trama inactiva.

[0047] Particularmente, las tramas activas son aquellas que el codificador clasifica como que tienen voz (o cualquier otro tipo de sonido sin ruido) y las tramas inactivas son aquellas que se clasifican como que tienen silencio o solo ruido.

[0048] Cualquiera de los ejemplos del CNG 220 (220a-220e) se puede controlar mediante un controlador adecuado.

Codificador

[0049] Ahora se analiza un codificador. El codificador puede codificar tramas activas y tramas inactivas. Para las tramas inactivas, el codificador puede codificar datos de ruido paramétricos (por ejemplo, forma de ruido y/o valor de coherencia) sin codificar la señal de audio por completo. Se observa que la codificación de las tramas de audio inactivas se puede reducir con respecto a las tramas de audio activas, para reducir la cantidad de información que se va a codificar en el flujo de bits. Además, los datos de ruido paramétricos (por ejemplo, forma de ruido) para las tramas inactivas pueden tener menos información para cada banda de frecuencia y/o pueden tener menos intervalos que los codificados en las tramas activas. Los datos de ruido paramétricos se pueden proporcionar en el dominio izquierdo/derecho o en otro dominio (por ejemplo, dominio medio/lateral), por ejemplo, proporcionando una primera combinación lineal entre datos de ruido paramétricos del primer y segundo canales y una segunda combinación lineal entre datos de ruido paramétricos del primer y segundo canales (en algunos casos, también es posible proporcionar información de ganancia que no está asociada a la primera y segunda combinaciones lineales, pero se proporcionan en el dominio izquierdo/derecho). La primera y segunda combinaciones lineales son en general linealmente independientes entre sí.

[0050] El codificador puede incluir un detector de actividad que clasifica si una trama está activa o inactiva.

[0051] Las Figs. 1, 2 y 4 muestran ejemplos de codificadores 300a y 300b (que también se denominan 300 cuando no es necesario distinguir entre el codificador 300a del codificador 300b). Cada codificador 300 de audio puede generar una señal 232 de audio multicanal codificada para una secuencia de tramas de una señal 304 de entrada. La señal de entrada 304 se considera aquí dividida entre un primer canal 301 (también indicado como canal izquierdo o "l", donde "l" es la letra cuya versión en mayúscula es "L" y es la primera letra de "izquierda" en inglés) y un segundo canal 303 (o "r", donde "r" es la letra cuya versión en mayúscula es "R" y es la primera letra de "derecha" en inglés).

[0052] La señal de audio multicanal codificada 232 se puede definir en una secuencia de tramas, que puede estar, por ejemplo, en el dominio de tiempo (por ejemplo, cada muestra "n" se puede referir a un instante de tiempo particular y las muestras de una trama pueden formar una secuencia, por ejemplo, una secuencia de muestreo de una señal de audio de entrada o una secuencia después de haber filtrado una señal de audio de entrada).

[0053] El codificador 300 (300a, 300b) puede incluir un detector de actividad 380, que no se muestra en Las Figs. 2 y 4 (a pesar de estar en algunos ejemplos implementados en el mismo), pero se muestra en la Fig. 1. La Fig. 1 muestra que cada trama de la señal de entrada 304 se puede clasificar como una "trama activa 306" o una "trama inactiva 308". Una trama inactiva 308 es de modo que la señal se considera que es silencio (y, por ejemplo, solo hay silencio o ruido), en tanto que la trama activa 306 puede tener alguna detección de señal de audio sin ruido (por ejemplo, voz, música, etc.).

[0054] En la señal de audio múltiple codificada 232 codificada (por ejemplo, flujo de bits) por el codificador 300, la información sobre si la trama es una trama activa 306 o una trama de silencio 308 se puede señalar, por ejemplo, en la denominada "información adicional de generación de ruido de confort" 402 (p_frame), también denominada "información adicional".

[0055] La Fig. 1 muestra una etapa de pre-procesamiento 360 que puede determinar (por ejemplo, clasificar) si una trama es una trama activa 306 o una trama silenciosa 308. Se observa aquí que los canales 301 y 303 de la señal de entrada 304 se indican con letras mayúsculas, como L (301, canal izquierdo) y R (303, canal derecho) para indicar que están en el dominio de frecuencia. Como se puede ver en la Fig. 1, se puede aplicar una etapa de paso de análisis espectral 370 (un primer análisis espectral 370-1 al primer canal 301, L; y una segunda etapa 370-3 para el segundo canal 303, R). La etapa de análisis espectral 370 se puede realizar para cada trama de la señal de entrada 304 y se puede basar, por ejemplo, en mediciones de armonicidad. En particular, en algunos ejemplos, el análisis espectral se realiza por la etapa 370 en el primer canal 301 se puede realizar por separado del análisis espectral realizado en el segundo canal 303 de la misma trama. En algunos casos, la etapa de análisis espectral 370 puede incluir el cálculo de parámetros relacionados con la energía, tal como la energía promedio para un intervalo de bandas de frecuencia predefinidas y la energía promedio total.

[0056] Se puede aplicar una etapa de detección de actividad 380 (que se puede considerar una detección de actividad de voz en el caso de que se busque la voz). Una primera etapa de detección de actividad 380-1 se puede aplicar al primer canal 301 (y en particular a las mediciones realizadas en el primer canal), y la segunda etapa de detección de actividad 380-3 se puede aplicar al segundo canal 303 (y en particular a las mediciones realizadas en el segundo canal). En ejemplos, la etapa de detección de actividad 380 puede estimar la energía del ruido de fondo en la señal de entrada 304 y usar esa estimación para calcular una relación señal-ruido, que se compara con un umbral de relación señal-ruido para determinar si la trama se clasifica como activa o inactiva (es decir, la relación señal-ruido calculada que está por encima del umbral de relación señal-ruido que implica que la trama se clasifica como activa; y la relación señal-ruido calculada que está por debajo del umbral de relación señal-ruido que implica que la trama se clasifica como inactiva). En ejemplos, la etapa 380 puede comparar la armonicidad obtenida por las etapas de análisis espectral 370-1 y 370-3, respectivamente, con uno o dos umbrales de armonicidad (por ejemplo, un primer umbral para el primer canal 301 y un segundo umbral para el segundo canal 303). En ambos casos, puede ser posible clasificar no solo cada trama, sino también cada canal de cada trama como que es cualquiera de un canal activo o un canal inactivo.

[0057] Se puede realizar una decisión 381, y con base en la misma, es posible decidir (como se identifica por el conmutador 381') si realizar un procesamiento estéreo discreto 306a o un procesamiento de transmisión discontinua estéreo (DTX estéreo) 306b. En particular, en caso de trama activa (y procesamiento estéreo discreto 306a), la codificación se puede realizar según cualquier estrategia o norma de procesamiento o proceso, y por lo tanto aquí no se analiza adicionalmente en detalle. La mayor parte del análisis más adelante se referirá a la DTX estéreo 306b.

[0058] Particularmente, en ejemplos, una trama se clasifica (en la etapa 381) como trama inactiva solo si ambos canales 301 y 303 se clasifican como inactivos por las etapas 380-1 y 380-3, respectivamente. Por lo tanto, se evitan problemas en la decisión de detección de actividad como se analizó anteriormente. En particular, no es necesario señalar la clasificación de activa/inactiva para cada canal para cada trama (reduciendo así la señalización), y se obtiene inherentemente una sincronización entre los canales. Además, donde el decodificador es como se analiza en el presente documento, es posible hacer uso de la coherencia entre el primer y segundo canales 301 y 303 y generar

algunas señales de ruido, que se correlacionan/decorrelacionan según la coherencia obtenida para la señal 304. Ahora, los elementos del codificador 300 (300a, 300b) que se usan para codificar la trama inactiva se analizan en detalle. Como se explicó, se puede usar cualquier otra técnica para codificar las tramas activas 308, y por lo tanto no se analiza aquí.

5

[0059] En términos generales, el codificador 300a, 300b (300) puede incluir una calculadora de parámetros de ruido 3040 para calcular datos de ruido paramétricos 401, 403 para el primer y segundo canales 301, 303. La calculadora de parámetros de ruido 3040 puede calcular datos de ruido paramétricos 401, 403 (por ejemplo, índices y/o ganancias) para el primer canal 301 y el segundo canal 303. Por lo tanto, la calculadora de parámetros de ruido 3040 puede proporcionar datos de audio codificados 232 en una secuencia de tramas que pueden comprender tramas activas 306 y tramas inactivas 308 (que pueden seguir a las tramas activas 306). En particular, en el caso de tramas inactivas 308, los datos de audio codificados 232 se pueden codificar como una o dos tramas de descripción de inserción de silencio (SID) 241, 243. En algunos ejemplos (por ejemplo, en la Fig. 2), solo hay una única trama SID, en alguna otra, hay dos tramas SID (por ejemplo, en la Fig. 4).

10

[0060] Una trama inactiva 308 puede incluir, en particular, al menos uno de:

- información adicional de generación de ruido de confort (por ejemplo, 402, p_{frame});
- datos de parámetros de ruido de confort 401 para el primer canal 301 o una primera combinación lineal de datos de parámetros de ruido de confort para el primer canal 301 y datos de parámetros de ruido de confort para el segundo canal ($V_{l,\text{ind}}$, $V_{m,\text{ind}}$, p_{noise} , ganancia $g_{l,q}$);
- datos de parámetros de ruido de confort 403 para el segundo canal 303 o una segunda combinación lineal de datos de parámetros de ruido de confort para el primer canal 301 y datos de parámetros de ruido de confort para el segundo canal ($V_{r,\text{ind}}$, $V_{s,\text{ind}}$, p_{noise} , ganancia $g_{r,q}$);
- información sobre coherencia (datos de coherencia) (c, 404).

15

[0061] En algunos ejemplos, una primera trama de descriptor de inserción de silencio 241 puede incluir los dos primeros elementos de la lista anterior, y una segunda trama de descriptor de inserción de silencio 243 puede incluir las dos últimas características en los campos de datos específicos. No obstante, diferentes protocolos pueden proporcionar diferentes campos de datos u organización diferente del flujo de bits. Sin embargo, en algunos casos (por ejemplo, en la Fig. 2), puede haber solo una única trama inactiva para los parámetros de ruido para ambos canales.

20

[0062] Se mostrará que la información de coherencia (por ejemplo, parte del "descriptor de inserción de silencio") puede incluir un único valor (por ejemplo, codificado en pocos bits, como cuatro bits) que indica información de coherencia (por ejemplo, datos de correlación), por ejemplo, la coherencia entre el primer canal 301 y el segundo canal 303 de la misma trama inactiva 308. Por otro lado, los datos de parámetros de ruido de confort 401, 403 pueden indicar, para cada canal 301, 303, energía de señal para la trama inactiva 308 (por ejemplo, puede proporcionar sustancialmente una envolvente), o de todos modos puede proporcionar información de forma de ruido. La envolvente o la información de forma de ruido puede estar en forma de múltiples coeficientes para intervalos de frecuencia y una ganancia para cada canal. La información de forma de ruido se puede obtener en la etapa 312 (ver más adelante) usando los canales de entrada originales (301, 303) y, a continuación, se realiza la codificación media/lateral en los vectores de parámetros de forma de ruido. Se mostrará que en el decodificador puede ser posible generar algunos canales de ruido (por ejemplo, 201, 203 como en la Fig. 3) que pueden estar influenciados por la información de coherencia 404. Por lo tanto, los canales de ruido 201, 203 generados por el CNG 220 (220a-220) se pueden modificar por un modificador de señal 250 controlado por los datos de ruido de control (datos de parámetros de ruido de confort 401, 403, 2312) que indican energías de señal para el primer canal de audio L_{out} y el segundo canal de audio R_{out} .

35

40

45

[0063] El codificador de audio 300 (300a, 300b) puede incluir una calculadora de coherencia 320, que puede obtener la información de coherencia (404) que se va a codificar en el flujo de bits (por ejemplo, la señal 232, la trama 241 o 243). La información de coherencia (c, 404) puede indicar una situación de coherencia entre el primer canal 301 (por ejemplo, el canal izquierdo) y el segundo canal 303 (por ejemplo, el canal derecho) en la trama inactiva 308. Ejemplos de los mismos se analizarán más adelante.

50

[0064] El codificador 300 (300a, 300b) puede incluir una interfaz de salida 310 configurada para generar la señal de audio multicanal 232 (flujo de bits) con los datos de audio codificados para la trama activa 306 y, para la trama inactiva 308, los primeros datos paramétricos (datos paramétricos de ruido de confort) 401 (p_{noise} , izquierda) los segundos datos de ruido paramétricos (p_{noise} , derecha 403) y los datos de coherencia c (404). Los primeros datos 401 paramétricos pueden ser datos paramétricos del primer canal (por ejemplo, canal izquierdo) o una primera combinación lineal del primer y segundo canales (por ejemplo, canal medio). Los segundos datos paramétricos 403 pueden ser datos paramétricos del segundo canal (por ejemplo, canal derecho) o una segunda combinación lineal del primer y segundo canales (por ejemplo, canal lateral) diferentes de la primera combinación lineal.

55

60

[0065] En el flujo de bits 232, también puede haber información adicional 402, que incluye una indicación de si la trama actual es una trama activa 306 o una trama inactiva 308, por ejemplo, para informar al decodificador de las

técnicas de decodificación que se van a utilizar.

[0066] En particular, la Fig. 4 muestra la calculadora de parámetros de ruido (etapa de cálculo de parámetros de ruido) 3040 que incluye una primera etapa de calculadora de parámetros de ruido 304-1 en la que se pueden calcular los datos de parámetros de ruido de confort 401 para el primer canal 301, y una segunda etapa de calculadora de parámetros de ruido 304-3, en la que se puede calcular el segundo parámetro de ruido de confort 403 para el segundo canal 303. La Fig. 2 muestra un ejemplo donde los parámetros de ruido se procesan y cuantifican conjuntamente. Las partes internas (por ejemplo, conversión de los vectores de forma de ruido en representación M/S) se muestran en la Fig. 5. Básicamente, podemos tener una forma de ruido del primer canal M y una forma de ruido del segundo canal S que se pueden codificar como índices medios e índices laterales, mientras que también se puede codificar una ganancia para la forma de ruido del canal izquierdo 301 y ganancias para la forma de ruido del canal derecho 303.

[0067] Una calculadora de coherencia 320 puede calcular los datos de coherencia (información de coherencia) c (404) que indican la situación de coherencia entre el primer canal L y el segundo canal R. En este caso, la calculadora de coherencia 320 puede operar en el dominio de frecuencia.

[0068] Como se puede ver, la calculadora de coherencia 320 puede incluir una etapa de coherencia de canal de cálculo 320' en la que se obtiene el valor de coherencia c (404). Corriente abajo de la misma, se puede usar una etapa de cuantificación uniforme 320". Por lo tanto, se puede obtener una versión cuantificada c_{ind} del valor de coherencia c.

[0069] Más adelante aquí, hay algunas explicaciones sobre cómo obtener la coherencia y cómo cuantificarla.

[0070] La calculadora de coherencia 320 puede, en algunos ejemplos:

- calcular un valor intermedio real y un valor intermedio imaginario a partir de valores espectrales complejos para el primer canal y el segundo canal (303) en la trama inactiva;
- calcular un primer valor de energía para el primer canal y un segundo valor de energía para el segundo canal (303) en la trama inactiva; y
- calcular los datos de coherencia (404, c) usando el valor intermedio real, el valor intermedio imaginario, el primer valor de energía y el segundo valor de energía, y/o suavizar al menos uno del valor intermedio real, el valor intermedio imaginario, el primer valor de energía y el segundo valor de energía, y calcular los datos de coherencia usando al menos un valor suavizado.

[0071] La calculadora de coherencia 320 puede elevar al cuadrado un valor intermedio real suavizado y para elevar al cuadrado un valor intermedio imaginario suavizado y para sumar los valores al cuadrado para obtener un primer número de componente. La calculadora de coherencia 320 puede multiplicar los valores de energía primero y segundo suavizados para obtener un segundo número de componente, y combinar los números de componente primero y segundo para obtener un número de componente resultante para el valor de coherencia, en el que se basan los datos de coherencia. La calculadora de coherencia 320 puede calcular una raíz cuadrada del número de resultado para obtener un valor de coherencia en el que se basan los datos de coherencia. Ejemplos de fórmulas se proporcionan más adelante.

[0072] Se explica ahora cómo se obtiene la forma de la forma de ruido (u otra energía de señal) que se va a renderizar en el decodificador. Lo que se codificará es básicamente la forma (u otra información relacionada con la energía) del ruido de la señal de entrada original 302, que en el decodificador se aplicará al ruido generado 203 y le dará forma, para renderizar un ruido 252 (señal de audio de salida) que se asemeja al ruido original de la señal 304.

[0073] En primer lugar, se observa que la señal 304 como tal no se codifica en el flujo de bits 232 por el codificador. Sin embargo, la información de ruido (por ejemplo, información de energía, información de envolvente) se puede codificar en el flujo de bits 232, para generar posteriormente una señal de ruido que tiene la forma de ruido codificada por el codificador.

[0074] Se puede aplicar un bloque de obtener forma de ruido 312 a la señal de entrada 304 del codificador. El bloque "obtener forma de ruido" 312 puede calcular una representación paramétrica de baja resolución 1312 de la envolvente espectral del ruido en la señal de entrada 304. Esto se puede hacer, por ejemplo, al calcular valores de energía en bandas de frecuencia de la representación de dominio de frecuencia de la señal de entrada 304. Los valores de energía se pueden convertir en una representación logarítmica (si es necesario) y se pueden condensar en un número menor (N) de parámetros que se usan posteriormente en el decodificador para generar el ruido de confort. Estas representaciones de baja resolución del ruido se denominan aquí "formas de ruido" 1312. Por lo tanto, lo que está corriente abajo del bloque "obtener forma de ruido" 312 no se debe entender como que representa la señal de entrada 304, sino como que representa su forma de ruido (representaciones paramétricas de las envolventes espec-

trales del ruido en los canales respectivos). Esto es importante, ya que el codificador solo puede transmitir esta representación de menor resolución de la envolvente espectral del ruido en la trama SID. Por lo tanto, en la Fig. 2, se puede entender que toda la parte de "calculadora de parámetros de ruido" (3040) opera solo en estos vectores de parámetros relacionados con el ruido (por ejemplo, identificados como v_l , v_r , $v_{m,ind}$ y $v_{s,ind}$) y no en representaciones de señal de la señal 304.

[0075] La Fig. 5 muestra un ejemplo de la parte de "calculadora de parámetros de ruido" 3040 (cuantificación de forma de ruido conjunta). Se puede aplicar una etapa de convertidor de L/R a M/S 314 para obtener la representación de canal medio v_m de la forma de ruido 1312 (primera combinación lineal de las formas de ruido de los canales L y R) y la representación de canal lateral v_r de la forma de ruido 1312 (segunda combinación lineal de las formas de ruido de las formas de ruido de los canales L y R). Más adelante, se mostrará una forma de cómo obtenerlo. Por consiguiente, la forma de ruido 304 puede resultar dividida en dos canales v_m y v_r .

[0076] Posteriormente, en la etapa de normalización 316, al menos una de la representación de canal medio v_m de la forma de ruido 1312 y la representación de canal lateral v_r de la forma de ruido 1312 se pueden normalizar, para obtener una versión normalizada $v_{m,n}$ de la representación de canal medio v_m de la forma de ruido 1312 y/o una versión normalizada $v_{r,n}$ de la representación de canal lateral v_r de la forma de ruido 1312.

[0077] Posteriormente, se puede aplicar una etapa de cuantificación (por ejemplo, cuantificación vectorial, VQ) 318 a la versión normalizada de la señal 1304, por ejemplo, en forma de una versión cuantificada $v_{m,ind}$ de la representación de canal medio normalizada $v_{m,n}$ de la forma de ruido 1312 y una versión cuantificada $v_{s,ind}$ de la representación de canal lateral normalizada $v_{s,n}$ de la forma de ruido 1312. Se puede usar una cuantificación de vector (por ejemplo, a través de un cuantificador de vector de múltiples etapas). Por lo tanto, los índices $v_{m,ind}[k]$ (k que es el índice del intervalo de frecuencia particular) pueden describir la representación media de la forma de ruido y los índices $v_{s,ind}[k]$ pueden describir la representación lateral de la forma de ruido. Por lo tanto, los índices $v_{m,ind}[k]$ y $v_{s,ind}[k]$ se pueden codificar en el flujo de bits 232 como una primera combinación lineal de datos de parámetros de ruido de confort para el primer canal y datos de parámetros de ruido de confort para el segundo canal y una segunda combinación lineal de datos de parámetros de ruido de confort para el primer canal y datos de parámetros de ruido de confort para el segundo canal.

[0078] En la etapa de descuantificación 322, se puede realizar una descuantificación en la versión cuantificada $v_{m,ind}$ de la representación de canal medio normalizada $v_{m,n}$ de la forma de ruido 1312 y la versión cuantificada $v_{s,ind}$ de la representación de canal lateral normalizada $v_{s,n}$ de la forma de ruido 1312.

[0079] Se puede aplicar un convertidor de M/S a L/R 324 a las versiones descuantificadas de las representaciones medias y laterales descuantificadas $v_{m,q}$ y $v_{s,q}$ de la forma de ruido 1312, para obtener una versión de la forma de ruido 1312 en los canales originales (izquierdo y derecho) v'_l y v'_r .

[0080] Posteriormente, en la etapa 326, se pueden calcular las ganancias g_l y g_r . En particular, las ganancias son válidas para todas las muestras de la forma de ruido del mismo canal (v'_l y v'_r) de la misma trama inactiva 306. Las ganancias g_l y g_r se pueden obtener al tener en cuenta la totalidad (o casi la totalidad) de los intervalos de frecuencia en las representaciones de forma de ruido v'_l y v'_r .

[0081] La ganancia g_l se puede obtener al comparar:

- los valores de los intervalos de frecuencia de la forma de ruido del primer canal 301 en el dominio L/R (corriente arriba del convertidor L/R-a-M/S 314); con
- los valores de los intervalos de frecuencia de la forma de ruido 1312, una vez reconvertidos en el dominio L/R, del primer canal 301 (corriente abajo del convertidor de M/S a L/R 324).

[0082] De manera análoga, la ganancia g_r se puede obtener al comparar:

- los valores de los coeficientes de la forma de ruido del segundo canal 303 en el dominio L/R (corriente arriba del convertidor L/R-a-M/S 314); con
- los valores de los coeficientes de la forma de ruido 1312, reconvertidos en el dominio L/R, del segundo canal 303 (corriente abajo del convertidor de M/S a L/R 324).

[0083] Más adelante se propone un ejemplo de cómo obtener las ganancias. Sin embargo, la ganancia puede ser, en el dominio lineal, por ejemplo, proporcional a un promedio geométrico de una multiplicidad de fracciones, cada fracción que es una fracción entre los coeficientes de forma de ruido de un canal particular en el dominio L/R (corriente arriba del convertidor L/R a M/S 314) y los coeficientes del mismo canal una vez reconvertidos en el dominio L/R corriente abajo al convertidor de M/S a L/R 324. En el dominio logarítmico, para cada canal, la ganancia se puede obtener como que es proporcional a un promedio algebraico entre las diferencias entre los coeficientes, los coeficientes

de la versión FD de la forma de ruido en el dominio L/R (corriente arriba del convertidor L/R a M/S 314) y los coeficientes de la forma de ruido una vez reconvertidos en el dominio L/R corriente abajo del convertidor de M/S a L/R 324. En general, en el dominio logarítmico o escalar, la ganancia puede proporcionar una relación entre una versión de la forma de ruido del canal izquierdo o derecho antes de la conversión de L/R a M/S y la cuantificación con una versión de la forma de ruido del canal izquierdo o derecho después de la descuantificación y la reconversión de M/S a L/R.

5 **[0084]** Se puede aplicar una etapa de cuantificación 328 a la ganancia g_l para obtener una versión cuantificada de la misma indicada con $g_{l,q}$, a la ganancia g_r para obtener una versión cuantificada de la misma indicada con $g_{r,q}$ que se puede obtener a partir de la ganancia no cuantificada g_r . Las ganancias $g_{l,q}$ y $g_{r,q}$ se pueden codificar en el flujo de bits 232 (por ejemplo, como datos de parámetros de ruido de confort 401 y/o 403) para que el decodificador las lea.

[0085] En algunos ejemplos, también es posible comparar la energía del vector de forma de ruido de canal lateral (por ejemplo, antes de normalizarse, por ejemplo, entre las etapas 314 y 316) con un umbral de energía predeterminado α (que puede ser un valor real positivo) (que en este caso es 0,1, pero también podría ser un valor diferente, tal como un valor entre 0,05 y 0,15). En un bloque de comparación 435 es posible determinar si la representación lateral v_s de la forma de ruido de la trama inactiva 308 tiene suficiente energía. Si la energía de la representación lateral v_s de la forma de ruido es menor que el umbral de energía α , entonces un resultado binario ("bandera no lateral"), como información adicional 402 se señala en el flujo de bits 232. Aquí se imagina que la bandera no lateral = 1 si la energía de la representación lateral v_s de la forma de ruido es menor que el umbral de energía α , y la bandera no lateral = 0 si la energía de la representación lateral v_s de la forma de ruido es mayor que el umbral de energía α . En algunos casos, la bandera puede ser 1 o 0 según la aplicación particular en caso de que la energía sea exactamente igual al umbral de energía. El bloque 436 niega el valor binario de la bandera no lateral 436 (si la entrada del bloque 436 es 1, entonces la salida 436' es 0; si la entrada del bloque 436 es 0, entonces la salida 436' es 1). El bloque 436 se muestra como que proporciona como salida 436' el valor opuesto de la bandera. Por consiguiente, si la energía de la representación lateral v_s de la forma de ruido es mayor que el umbral de energía, entonces el valor 436' puede ser 1, y si la energía de la representación lateral v_s de la forma de ruido es menor que el umbral predeterminado, entonces el valor 436' es 0. Se observa que el valor descuantificado $v_{s,q}$ se puede multiplicar por el valor binario 436'. Esta es simplemente una forma posible de obtener que, si la energía de la representación lateral v_s de la forma de ruido es menor que el umbral de energía predeterminado α , entonces los intervalos de la representación lateral descuantificada $v_{s,q}$ de la forma de ruido se ponen a cero artificialmente (la salida 437' del bloque 437 sería 0). Por otro lado, si la energía de la representación lateral v_s de la forma de ruido es suficientemente grande ($> \alpha$), entonces la salida 437' del bloque 437 (multiplicador) puede ser exactamente la misma que $v_{s,q}$. Por consiguiente, si la energía de la representación lateral v_s de la forma de ruido es menor que el umbral de energía predeterminado α , la representación lateral v_s de la forma de ruido (y en particular su versión descuantificada $v_{s,q}$) no se toma en consideración obteniendo las representaciones izquierda/derecha de la forma de ruido. (Se mostrará que además o como alternativa también el decodificador puede tener un mecanismo similar que pone a cero los coeficientes de la representación lateral de la forma de ruido). Se observa que la bandera no lateral también se puede codificar en el flujo de bits 232 como parte de la información adicional 402.

40 **[0086]** Se debe señalar que la energía de la representación lateral de la forma de ruido se muestra como que se mide (por el bloque 435) antes de la normalización de la forma de ruido (en el bloque 316), y la energía no se normaliza antes de compararla con el umbral. En principio, también se puede medir mediante el bloque 435 después de normalizar la forma de ruido (por ejemplo, el bloque 435 se podría introducir mediante $v_{s,n}$ en lugar de v_s).

45 **[0087]** Con referencia al umbral α usado para comparar la energía de la representación lateral de la forma de ruido, el valor 0,1 se puede elegir, en algunos ejemplos, arbitrariamente. En ejemplos, el umbral α se puede elegir después de la experimentación y el ajuste (por ejemplo, a través de la calibración). En algunos ejemplos, en principio se podría utilizar cualquier número que funcione para el formato de número (punto flotante o punto fijo) o la precisión de una implementación individual. Por lo tanto, el umbral α puede ser un parámetro específico de implementación que se puede introducir después de una calibración.

[0088] Se observa que la interfaz de salida (310) se puede configurar:

55 para generar la señal de audio multicanal codificada (232) que tiene datos de audio codificados para la trama activa (306) usando una primera pluralidad de coeficientes para un primer número de intervalos de frecuencia; y para generar los primeros datos de ruido paramétricos, los segundos datos de ruido paramétricos, o la primera combinación lineal de los primeros datos de ruido paramétricos y los segundos datos de ruido paramétricos y la segunda combinación lineal de los primeros datos de ruido paramétricos y los segundos datos de ruido paramétricos usando una segunda pluralidad de coeficientes que describen un segundo número de intervalos de frecuencia, donde el primer número de intervalos de frecuencia es mayor que el segundo número de intervalos de frecuencia.

[0089] De hecho, se puede utilizar una resolución reducida para las tramas inactivas, lo que reduce aún más la cantidad de bits utilizados para codificar el flujo de bits. Lo mismo se aplica al decodificador.

[0090] Cualquiera de los ejemplos del codificador se puede controlar mediante un controlador adecuado.

Decodificador

- 5 **[0091]** Ahora, se analizan decodificadores según ejemplos. Un decodificador puede incluir, por ejemplo, un generador de ruido de confort 220 (220a-220e) analizado anteriormente, por ejemplo, mostrado en Las Figs. 3a-3f. El ruido de confort 204 (señal de audio multicanal) se puede conformar en un modificador de señal 250, para obtener la señal de salida 252. Aquí estamos interesados en mostrar las operaciones para generar el ruido en las tramas inactivas 308, y no las de las tramas activas 206.
- 10 **[0092]** La Fig. 4 muestra un primer ejemplo de decodificador 200', indicado aquí con 200' (200b). Se observa que el decodificador 200' incluye un generador de ruido de confort 220 que puede incluir un generador 220 (220a-220e) según cualquiera de Las Figs. 3a-3f. Corriente abajo del generador 220 (220a-220e), puede estar presente un modificador de señal 250 (no mostrado, pero mostrado en la Fig. 4), para conformar el ruido multicanal generado 204 según los parámetros de energía codificados en los datos de parámetros de ruido de confort (401, 403). A través de la interfaz de entrada de decodificador 210, el decodificador 200' puede obtener del flujo de bits 232 los datos de parámetros de ruido de confort (401, 403), que pueden incluir datos de parámetros de ruido de confort que describen la energía de la señal (por ejemplo, para un primer canal y un segundo canal, o para una primera combinación lineal y segunda combinación lineal del primer y segundo canales, la primera y segunda combinaciones lineales que son linealmente independientes entre sí). A través de la interfaz de entrada del decodificador 210, el decodificador 200' puede obtener datos de coherencia 404, que indican la coherencia entre diferentes canales. La Fig. 4 se muestra que en el flujo de bits 232, para la codificación de las tramas inactivas, se proporcionan dos tramas descriptoras de silencio diferentes 241 y 243, respectivamente, pero existe la posibilidad de usar más de dos tramas descriptoras, o solo una trama de descriptor individual. La salida del decodificador 200b es una salida multicanal.
- 25 **[0093]** Con referencia a la Fig. 2, ahora se analiza un decodificador 200' (aquí indicado con 200a) que es un ejemplo del decodificador 200, que se puede usar para generar la señal de salida 252, por ejemplo, en forma de ruido.
- [0094]** En primer lugar, el decodificador 200a (200') puede incluir una interfaz de entrada 210 para recibir los datos de audio codificados 232 (flujo de bits) en la secuencia de tramas 306, 308, como se codifica por el codificador 300a o 300b, por ejemplo. El decodificador 200a (200') puede ser, o más en general ser parte de, un generador de señales multicanal 200 que puede ser o incluir el generador de ruido de confort 220 (220a-220e) de cualquiera de Las Figs. 3a-3f, a modo de ejemplo.
- 35 **[0095]** En primer lugar, la Fig. 2 muestra un generador de ruido de confort (CNG) estéreo 220 (220a-220e). En particular, el generador de ruido de confort 220 (220a-220e) puede ser como el de Las Figs. 3a-3f o una de sus variantes. Aquí, una información de coherencia 404 (por ejemplo, c , o más precisamente c_q también indicada con "coh" o c_{ind}), como se obtiene del codificador 300a o 300b se puede usar para generar la señal multicanal 204 (en los canales 201, 203) que se han analizado anteriormente. La señal multicanal 204 generada por el CNG 220 (220a-220e) se puede en realidad modificar adicionalmente, por ejemplo, teniendo en cuenta los datos de parámetros de ruido de confort 401 y 403, por ejemplo, información de forma de ruido para un primer canal (izquierdo) y un segundo canal (derecho) de la señal multicanal que se va a conformar. En particular, se mostrará que existe la posibilidad de obtener los índices medios $v_{m, ind}$ (401) y los índices laterales $v_{s, ind}$ (403) generados por el codificador 300a (y en particular por la calculadora de parámetros de ruido 3040) en la etapa 316 y/o 318, y las ganancias $g_{l,q}$ y $g_{r,q}$ obtenidas en la etapa 45 326 y/o 328.
- [0096]** Como se muestra en la Fig. 2, la información adicional 402 puede permitir determinar si la trama actual es una trama activa 306 o una trama inactiva 308. Los elementos de la Fig. 2 se refieren al procesamiento de las tramas inactivas 308, y se pretende que se pueda usar cualquier técnica para la generación de la señal de salida en las tramas activas 306, que por lo tanto no son un objeto del presente documento.
- 50 **[0097]** Como se muestra en la Fig. 2, se obtienen varios ejemplos de datos de ruido de confort a partir del flujo de bits 232. Los datos de ruido de confort pueden incluir, como se explicó anteriormente, información de coherencia (datos) 404, parámetros 401 y 403 ($v_{m, ind}$ y $v_{s, ind}$) que indican la forma del ruido y/o ganancias ($g_{l,q}$ y $g_{r,q}$).
- 55 **[0098]** La etapa 212-C puede descuantificar la versión cuantificada c_{ind} de la información de coherencia 404, para obtener la información de coherencia descuantificada c_q .
- [0099]** La etapa 2120 (descuantificación de forma de ruido conjunta) puede permitir descuantificar los otros datos de ruido de confort obtenidos del flujo de bits 232. Se puede hacer referencia a la Fig. 6. Se forma una etapa de descuantificación 212 mediante otras etapas de descuantificación indicadas aquí con 212-M, 212-S, 212-R, 212-L. La etapa 212-M puede descuantificar los parámetros de forma de ruido de canal medio 401 y 403, para obtener los parámetros de forma de ruido descuantificados $v_{m,q}$ y $v_{s,q}$. La etapa 212-S puede proporcionar la versión descuantificada $v_{s,q}$ de los parámetros de forma de ruido de canal lateral 403 ($v_{s, ind}$). En algunos ejemplos, es posible hacer uso

de la bandera no lateral, para poner a cero la salida de la etapa 212-S en caso de que la energía del vector de forma de ruido v_s se reconozca, por el bloque 435 en el codificador 300a, como menor que el umbral predeterminado α . En caso de que la energía sea menor que el umbral predeterminado α y la bandera no lateral lo señalice, la versión descuantificada $v_{s,q}$ del vector de forma de ruido v_s se puede poner a cero (que conceptualmente se muestra como una multiplicación por una bandera 536' obtenida de un bloque 536 que tiene la misma función que el bloque del codificador 436, aunque el bloque 536 lee realmente una bandera no lateral codificada en la información adicional del flujo de bits 232, sin realizar ninguna comparación con el umbral α). Por lo tanto, si se ha determinado que la energía del canal lateral en el codificador es menor que el umbral predeterminado α , la versión descuantificada $v_{s,q}$ del vector de forma de ruido v_s se pone a cero artificialmente y el valor en la salida 537' del bloque modificador de escala 537 es cero. De lo contrario, si la energía es mayor que el umbral predeterminado, entonces la salida 537' es la misma de la versión cuantificada $v_{s,q}$ de los índices laterales 403 ($v_{s,ind}$) de la forma de ruido del canal lateral. En otros términos, los valores del vector de forma de ruido $v_{s,ind}$ se desprecian en caso de que la energía del canal lateral esté por debajo del umbral de energía predeterminado α .

15 **[0100]** En la etapa de M/S a L/R 516, se realiza una conversión de M/S a L/R, para obtener una versión L/R v'_l , v'_r de los datos paramétricos (forma de ruido). Posteriormente, se puede usar una etapa de ganancia 518 (formada por las etapas 518-L y 518-R), de modo que en la etapa 518-L el canal v'_l se modifica en escala por la ganancia $g_{l,q}$, mientras que en la etapa 518-R, el canal v'_r se modifica en escala por la ganancia $g_{r,q}$. Por lo tanto, los canales de energía $v_{l,q}$ y $v_{r,q}$ se pueden obtener como salida de la etapa de ganancia 518. Los bloques de etapas 518-L y 518-R se muestran con el "+" porque se imagina que la transmisión de los valores está en el dominio logarítmico, y por lo tanto se indica además la modificación de escala de valores. Sin embargo, la etapa de ganancia 518 indica que los vectores de forma de ruido reconstruidos $v_{l,q}$ y $v_{r,q}$ se modifican en escala. Los vectores de forma de ruido reconstruidos $v_{l,q}$ y $v_{r,q}$ se indican aquí de manera compleja con 2312 y son la versión reconstruida de la forma de ruido 1312 obtenida originalmente por el bloque "obtener forma de ruido" 312 en el codificador. En términos generales, cada ganancia es constante para todos los índices (coeficientes) del mismo canal de la misma trama inactiva.

30 **[0101]** Se observa que los índices $v_{m,ind}$, $v_{s,ind}$ y las ganancias $g_{l,q}$, $g_{r,q}$ son coeficientes de forma de ruido y proporcionan información sobre la energía de la trama. Básicamente se refieren a datos paramétricos asociados a la señal de entrada 304 que se usan para generar la señal 252, pero no representan la señal 304 o la señal 252 que se va a generar. Dicho de otra manera, los canales de ruido $v_{r,q}$ y $v_{l,q}$ describen una envolvente que se va a aplicar a la señal multicanal 204 generada por el CNG 220.

35 **[0102]** Volviendo a la Fig. 2, los vectores de forma de ruido reconstruidos $v_{l,q}$ y $v_{r,q}$ (2312) se usan en el modificador de señal 250, para obtener una señal modificada 252 conformando el ruido 204. En particular, el primer canal 201 del ruido generado 204 puede estar conformado por el canal $v_{l,q}$ en la etapa 250-L, y el canal 203 del ruido generado 204 en la etapa 250-R para obtener la señal de audio multicanal de salida 252 (L_{out} y R_{out}).

40 **[0103]** En ejemplos, la propia señal de ruido de confort 204 no se genera en el dominio logarítmico: solo las formas de ruido pueden usar una representación logarítmica. Se puede realizar una conversión del dominio logarítmico al dominio lineal (aunque no se muestra).

[0104] También se puede realizar una conversión del dominio de frecuencia al dominio de tiempo (aunque no se muestra).

45 **[0105]** El decodificador 200' (200a, 200b) también puede comprender un convertidor de espectro-tiempo (por ejemplo, el modificador de señal 250) para convertir el primer canal resultante 201 y el segundo canal resultante 203 que se ajustan espectralmente y se ajustan en coherencia, en representaciones de dominio de tiempo correspondientes para combinarse con o concatenarse con representaciones de dominio de tiempo de canales correspondientes de la señal multicanal decodificada para la trama activa. Esta conversión del ruido de confort generado en una señal en el dominio de tiempo ocurre después del bloque modificador de señal 250 en la Fig. 2. La parte de "combinación con o concatenación a" básicamente significa que antes o después de una trama inactiva que emplea una de estas técnicas CNG, también puede haber tramas activas (otra ruta de procesamiento en la Fig. 1) y para generar una salida continua sin huecos o clics audibles, etc., las tramas se deben concatenar correctamente.

55 **[0106]** En algunos ejemplos:

la señal de audio codificada (232) para la trama activa (306) tiene una primera pluralidad de coeficientes que describen un primer número de intervalos de frecuencia; y
la señal de audio codificada (232) para la trama inactiva (308) tiene una segunda pluralidad de coeficientes que describen un segundo número de intervalos de frecuencia.

[0107] El primer número de intervalos de frecuencia puede ser mayor que el segundo número de intervalos de frecuencia.

[0108] Cualquiera de los ejemplos del decodificador se puede controlar mediante un controlador adecuado.

Etapas de procesamiento: una primera versión

5 [0109] Los parámetros de ruido codificados en las dos tramas SID para los dos canales se calculan como en EVS [6] tal como LP-CNG o FD-CNG o ambos. La conformación de la energía de ruido en el decodificador también es igual que en el EVS, tal como LP-CNG o FD-CNG o ambos.

10 [0110] En el codificador, adicionalmente se calcula la coherencia de los dos canales, se cuantifica uniformemente usando cuatro bits y se envía en el flujo de bits 232. En el decodificador, la operación CNG se puede controlar entonces mediante el valor de coherencia transmitido 404. Se pueden usar tres fuentes de ruido gaussiano N_1, N_2, N_3 (211a, 212a, 213a; 211b, 212b, 213b; 211c, 212c, 213c; 211d, 212d, 213d; 211e, 212e, 213e) como se muestra en Las Figs. 3a-3f. Cuando la coherencia de canal es alta, se puede añadir principalmente ruido correlacionado a ambos canales 221' y 223', en tanto que se añade más ruido no correlacionado si la coherencia 404 es baja.

15 [0111] Para todas las tramas inactivas 306, los parámetros para generación de ruido de confort (parámetros de ruido) se pueden estimar constantemente en el codificador (por ejemplo, 300, 300a, 300b). Esto se puede hacer, por ejemplo, aplicando el algoritmo de estimación de ruido en el dominio de la frecuencia (por ejemplo, [8]), por ejemplo, como se describe en [6] por separado en ambos canales de entrada (por ejemplo, 301, 303) para calcular dos conjuntos de parámetros de ruido (por ejemplo, 401, 403), que también se explican como datos de ruido paramétricos. Además, la coherencia (c, 404) de los dos canales se puede calcular (por ejemplo, en la calculadora de coherencia 320) de la siguiente manera: Dados los espectros DFT de M puntos de los dos canales de entrada $L, R \in C^M(L, R$ puede ser 301, 303) se pueden calcular cuatro valores intermedios, por ejemplo

25

$$c_{real} = \sum_{i=0}^{M-1} R\{L_i \times R_i^*\}$$

$$c_{imag} = \sum_{i=1}^{M-1} I\{L_i \times R_i^*\}$$

y las energías de los dos canales

30

$$e_L = \langle L, L \rangle = \sum_{i=0}^{M-1} L_i \times L_i^*$$

$$e_R = \langle R, R \rangle = \sum_{i=0}^{M-1} R_i \times R_i^*$$

35 [0112] Aquí, puede ser $M = 256$, $R\{\cdot\}$ denota la parte real de un número complejo, $I\{\cdot\}$ denota la parte imaginaria de un número complejo y $\{\cdot\}^*$ denota conjugación compleja. Estos valores intermedios se pueden suavizar a continuación, por ejemplo, utilizando los valores correspondientes de la trama anterior:

40

$$\overline{c_{real}} = 0,95 \times \overline{c_{real_previous}} + 0,05 \times c_{real}$$

$$\overline{c_{imag}} = 0,95 \times \overline{c_{imag_previous}} + 0,05 \times c_{imag}$$

$$\overline{e_L} = 0,95 \times \overline{e_{L_previous}} + 0,05 \times e_L$$

45

$$\overline{e_R} = 0,95 \times \overline{e_{R_previous}} + 0,05 \times e_R$$

[0113] Este pasaje puede ser parte del bloque 320' "Calcular Coherencia de Canal" en el codificador. Este es un suavizado temporal de parámetros internos, para evitar grandes saltos repentinos en los parámetros entre tramas. En otros términos, se aplica aquí un filtro de paso bajo a los parámetros.

50 [0114] En lugar de las constantes 0,95 y 0,05, se pueden usar otras constantes dentro del intervalo $0,95 \pm 0,03$ y $0,05 \mp 0,03$.

[0115] Como alternativa, es posible definir:

55

$$\overline{c_{real}} = \beta \times \overline{c_{real_{previous}}} + \gamma \times c_{real}$$

$$\overline{c_{imag}} = \beta \times \overline{c_{imag_{previous}}} + \gamma \times c_{imag}$$

$$5 \quad \overline{e_L} = \beta \times \overline{e_{L_{previous}}} + \gamma \times e_L$$

$$\overline{e_R} = \beta \times \overline{e_{R_{previous}}} + \gamma \times e_R$$

10 **[0116]** Donde $\beta, \gamma \in [0,1]$ y $\beta + \gamma = 1$, por ejemplo, $\beta = 0,95$ y $\gamma = 0,05$.

[0117] La coherencia (c, 404) ((que puede estar entre 0 y 1) se puede calcular entonces (por ejemplo, en la calculadora de coherencia (320) como

$$c = \sqrt{\frac{\overline{c_{real}^2} + \overline{c_{imag}^2}}{\overline{e_L} \times \overline{e_R}}} \approx \sqrt{\frac{|L^* \times R|}{\langle L, L \rangle \times \langle R, R \rangle}}$$

15 y cuantificada uniformemente (por ejemplo, en el cuantificador 320") usando, por ejemplo, cuatro bits como

$$c_{ind} = 0, \min(15, \text{floor}(15 \times c + 0,5))$$

20 **[0118]** La codificación de los parámetros de ruido estimados 1312, 2312 para ambos canales se puede realizar por separado, por ejemplo, como se especifica en [6]. Dos tramas SID 241, 243 se pueden entonces codificar y enviar al decodificador. La primera trama SID 241 puede contener los parámetros de ruido estimados 401 del canal L y (por ejemplo, cuatro) bits de información adicional 402, por ejemplo, como se describe en [6]. En la segunda trama SID 243, los parámetros de ruido 403 del canal R se pueden enviar junto con el valor de coherencia cuantificado de cuatro bits c, 404 (se pueden elegir diferentes cantidades de bits en diferentes ejemplos).

[0119] En el decodificador (por ejemplo, 200', 200a, 200b), tanto los parámetros de ruido de la trama SID (401, 403) como la información adicional de la primera trama 402 se pueden decodificar, por ejemplo, como se describe en [6]. El valor de coherencia 404 en la segunda trama se puede descuantificar en la etapa 212-C como

$$30 \quad \hat{c} = 15 \times c_{ind}$$

(en la Fig. 2, \hat{c} se sustituye por c_q).

35 **[0120]** Para la generación de ruido de confort (por ejemplo, en el generador 220 o cualquiera de los generadores 220a-220e, que puede incluir una de cualquiera de Las Figs. 3a-3e), según un ejemplo, se pueden usar tres fuentes de ruido gaussiano 211, 212, 213 como se muestra en la Fig. 3. Las fuentes de ruido 211, 212, 213 se pueden sumar de manera adaptativa (por ejemplo, en las etapas de sumador 206-1 y 206-3), por ejemplo, con base en el valor de coherencia (c, 404). Los espectros DFT de las señales de ruido de canal izquierdo y derecho $N_l[k], N_r[k]$ se pueden calcular como

$$N_l[k] = \sqrt{1 - \hat{c}} \times (N_1[k] + j \times N_1[k + M]) + \sqrt{\hat{c}} \times (N_2[k] + j \times N_2[k + M])$$

$$N_r[k] = \sqrt{1 - \hat{c}} \times (N_3[k] + j \times N_3[k + M]) + \sqrt{\hat{c}} \times (N_2[k] + j \times N_2[k + M])$$

45 con $k \in \{0, 1, \dots, M - 1\}$ (que es el índice del intervalo de frecuencia particular, mientras que cada canal tiene M intervalos de frecuencia) y $j^2 = -1$ (es decir, j es la unidad imaginaria), y " \times " es la multiplicación normal. En la presente, "intervalo de frecuencia" se refiere al número de valores complejos en los espectros N_l y N_r , respectivamente. M es la longitud de transformada de la FFT o DFT que se usa, por lo que la longitud de los espectros es M . Se observa que el ruido insertado en la parte real y el ruido insertado en la parte imaginaria pueden ser diferentes. Entonces, para una longitud de espectro de M , se necesitan $2 \times M$ valores (uno real y uno imaginario) generados a partir de cada fuente de ruido. O en otras palabras: N_l y N_r son vectores de valor complejo de longitud M , mientras que N_1, N_2 y N_3 son vectores de valor real de longitud $2 \times M$.

55 **[0121]** Posteriormente, la señal de ruido 204 en los dos canales se forma espectralmente (por ejemplo, dentro de las etapas 250-L, 250-R en la Fig. 2) usando sus parámetros de ruido correspondientes (2312) decodificados a partir de la trama SID respectiva y posteriormente transformados de nuevo al dominio de tiempo (por ejemplo, como se describe en [6]) para la generación de ruido de confort de dominio de frecuencia.

60 **[0122]** Se puede realizar cualquiera de los ejemplos del procesamiento mediante un controlador adecuado.

Etapas de procesamiento: una segunda versión

[0123] Los aspectos de las etapas de procesamiento como se analizó anteriormente se pueden integrar con al menos uno de los aspectos más adelante. Aquí se hace referencia principalmente a Las Figs. 2 y 5, pero también se podría hacer referencia a la Fig. 4.

[0124] En la Fig. 1 se ilustra un diagrama de bloques del marco genérico del codificador. Para cada trama en el codificador, la señal actual se puede clasificar como activa o inactiva al ejecutar un VAD en cada canal por separado como se describe en [6]. La decisión VAD luego se puede sincronizar entre los dos canales. En ejemplos, una trama se clasifica como una trama inactiva 308 solo si ambos canales se clasifican como inactivos. De lo contrario, se clasifica como activo y ambos canales se codifican conjuntamente en un sistema basado en MDCT utilizando M/S por banda como se describe en [10]. Cuando se conmuta de una trama activa a una trama inactiva, las señales pueden entrar en la ruta de codificación SID como se muestra en la Fig. 1.

[0125] Los parámetros (por ejemplo, 1312, 401, 403, $q_{l,q}$, $g_{r,q}$) para generación de ruido de confort (por ejemplo, parámetros de ruido) se pueden estimar constantemente en el codificador (por ejemplo, 300, 300a, 300b) tanto para tramas activas como inactivas (306, 308). Esto se puede hacer, por ejemplo, al aplicar un proceso de estimación de ruido en el dominio de la frecuencia como el analizado en [8] y/o como se describe en [6], por ejemplo, por separado en ambos canales de entrada 301, 303 para calcular dos conjuntos de parámetros de ruido, incluidas las formas de ruido espectral (M ; 401 y/o l_s o 403), por ejemplo, en el dominio logarítmico para cada canal.

[0126] Además, la coherencia (404, c) de los dos canales se puede calcular (por ejemplo, en la calculadora de coherencia 320) de la siguiente manera: Dados los espectros DFT de M puntos de los dos canales de entrada $L, R \in C^M$, se pueden calcular cuatro valores intermedios, que son

$$c_{real} = \sum_{i=1}^M R\{L_i \times R_i^*\}$$

$$c_{imag} = \sum_{i=1}^M I\{L_i \times R_i^*\}$$

30 y las energías de los dos canales

$$e_L = \langle L, L \rangle = \sum_{i=1}^M L_i \times L_i^*$$

$$e_R = \langle R, R \rangle = \sum_{i=1}^M R_i \times R_i^*$$

[0127] Aquí, puede ser $M = 256$ (se pueden usar otros valores para M), $R\{\cdot\}$ denota la parte real de un número complejo, $I\{\cdot\}$ denota la parte imaginaria de un número complejo y $\{\cdot\}^*$ denota conjugación compleja. Estos valores intermedios se suavizan luego sobre una base de subtrama de 10 ms. Con $\{\cdot\}_{previous}$ que denota el valor correspondiente de la subtrama anterior, los valores suavizados se pueden calcular como:

$$\overline{c_{real}} = 0,95 \times \overline{c_{real}_{previous}} + 0,05 \times c_{real}$$

$$\overline{c_{imag}} = 0,95 \times \overline{c_{imag}_{previous}} + 0,05 \times c_{imag}$$

$$\overline{e_L} = 0,95 \times \overline{e_{L_{previous}}} + 0,05 \times e_L$$

$$\overline{e_R} = 0,95 \times \overline{e_{R_{previous}}} + 0,05 \times e_R$$

[0128] En lugar de las constantes 0,95 y 0,05, se pueden usar otras constantes dentro del intervalo $0,95 \pm 0,03$ y $0,05 \mp 0,03$.

[0129] Como alternativa, es posible definir:

$$\overline{c_{real}} = \beta \times \overline{c_{real}_{previous}} + \gamma \times c_{real}$$

$$\overline{c_{imag}} = \beta \times \overline{c_{imag,previous}} + \gamma \times c_{imag}$$

$$\overline{e_L} = \beta \times \overline{e_{L,previous}} + \gamma \times e_L$$

$$\overline{e_R} = \beta \times \overline{e_{R,previous}} + \gamma \times e_R$$

5

[0130] Donde $\beta, \gamma \in [0,1]$ y $\beta + \gamma = 1$, por ejemplo, $\beta = 0,95$ y $\gamma = 0,05$ ($\beta > \gamma$, por ejemplo $\beta > 3 \times \gamma$, o $\beta > 6 \times \gamma$).

10

[0131] A continuación, la coherencia $c \in [0,1]$ se puede calcular (por ejemplo, en 320') como

$$c = \sqrt{\frac{\overline{c_{real}^2} + \overline{c_{imag}^2}}{\overline{e_L} \times \overline{e_R}}} \approx \sqrt{\frac{|L^* \times R|}{\langle L, L \rangle \times \langle R, R \rangle}}$$

15 y cuantificada uniformemente (por ejemplo, a 320") usando cuatro bits (pero son posibles diferentes cantidades de bits) como

$$c_{ind} = \min(15, [15 \times c + 0,5]) \in [0,15],$$

20 donde $\lfloor \cdot \rfloor$ denota el redondeo hacia abajo al número entero más cercano (función floor).

[0132] La codificación de las formas de ruido estimadas de ambos canales se puede realizar conjuntamente. A partir de las formas de ruido de canal izquierdo (v_l) y derecho (v_r), se pueden obtener diferentes canales (por ejemplo, a través de una combinación lineal), tal como una forma de ruido de canal medio (v_m) y se puede calcular una forma de ruido de canal lateral (v_s), (por ejemplo, en el bloque 314) como

25

$$v_m = \left\{ \frac{v_{l,1} + v_{r,1}}{2}, \dots, \frac{v_{l,N} + v_{r,N}}{2} \right\},$$

$$v_s = \left\{ \frac{v_{l,1} - v_{r,1}}{2}, \dots, \frac{v_{l,N} - v_{r,N}}{2} \right\},$$

30

donde N denota la longitud de los vectores de forma de ruido (por ejemplo, para cada trama inactiva 308), por ejemplo, en el dominio de la frecuencia. N denota la longitud del vector de forma de ruido, por ejemplo, como se estima en EVS [6], que puede estar entre 17 y 24. Los vectores de forma de ruido se pueden ver como una representación más compacta de la envolvente espectral del ruido en una trama de entrada. O, de manera más abstracta, una descripción espectral paramétrica de la señal de ruido usando N parámetros. N no está relacionado con la longitud de transformada de una FFT o una DFT.

35

[0133] Entonces, estas formas de ruido se pueden normalizar (por ejemplo, en la etapa 316) y/o cuantificar. Por ejemplo, se pueden cuantificar vectorialmente (por ejemplo, en la etapa 318), por ejemplo, usando cuantificadores vectoriales de múltiples etapas (MSVQ) (se describe un ejemplo en [6, pág. 442]).

40

[0134] El MSVQ usada en la etapa 318 para cuantificar la forma de v_m (para obtener $v_{m,ind}$ 401) puede tener 6 etapas (pero es posible otro número de etapas) y/o usar 37 bits (pero es posible otra cantidad de bits), por ejemplo, como se implementa para canales mono en [6], en tanto que el MSVQ usada, en la etapa 318, para cuantificar la forma de v_s (para obtener $v_{s,ind}$ 403) se puede haber reducido a 4 etapas (o en cualquier caso un número de etapas menor que el número de etapas usadas en la etapa 318) y/o puede usar en total 25 bits (o en cualquier caso una cantidad de bits menor que la cantidad de bits usada en la etapa 318 para codificar la forma v_m).

45

[0135] Los índices de libro de códigos de los MSVQ se pueden transmitir en el flujo de bits (por ejemplo, en los datos 232, y más en particular en los datos de parámetros de ruido de confort 401, 403). Los índices entonces se descuantifican dando como resultado las formas de ruido descuantificadas $v_{m,q}$ y $v_{s,q}$.

50

[0136] En el caso de que el ruido de fondo sea una única fuente de ruido en el centro de la imagen estéreo, se espera que las formas de ruido estimadas de ambos canales v_m, v_s sean muy similares o incluso iguales. La forma de ruido del canal S resultante contendrá entonces solo ceros. Sin embargo, el cuantificador vectorial (etapa 322) utilizado para cuantificar la implementación actual de v_s puede ser tal que no pueda modelar un vector de todo cero y después de la descuantificación, la forma de ruido de v_s descuantificada ($v_{s,q}$) podría resultar que ya no sea todo cero. Esto puede conducir a problemas de percepción con la representación de estos ruidos de fondo centrados. Para eludir este defecto del VQ 322, se puede calcular un valor no_side (bandera no_side) (y también se puede señalar en el flujo

55

de bits) dependiendo de la energía del vector de forma v_s no cuantificado (por ejemplo, la energía del vector de forma de ruido v_s después de la etapa 314 y/o antes de la etapa 316). La bandera no_side puede ser:

$$no_{side} = \begin{cases} 1, si \sum_{j=1}^N v_{s,j}^2 \leq \alpha \\ 0, de otraforma \end{cases}$$

5

[0137] El umbral de energía α podría ser, solo para dar un ejemplo, 0,1 u otro valor en el intervalo [0,05, 0,15]. Sin embargo, el umbral α puede ser arbitrario y, en una implementación, puede depender del formato numérico utilizado (por ejemplo, punto fijo o punto flotante) y/o de las posibles normalizaciones de señal utilizadas. En ejemplos, se podría usar un valor real positivo, dependiendo de lo dura que sea la definición empleada de un canal S "silencioso".

10 Por lo tanto, el intervalo puede ser (0, 1). El valor no_{side} se puede usar para indicar si se debe usar una forma de ruido v_s para reconstruir las formas de ruido de canal v_l y v_r (por ejemplo, en el decodificador). Si no_side es 1, la forma v_s descuantificada se establece a cero (por ejemplo, al modificar la escala del canal $v_{s,q}$ por el valor de 436' en la Fig. 2, que es un valor lógico NOT(NO_side)). no_side se transmite (señaliza) en el flujo de bits 232, por ejemplo, como información adicional 402. Posteriormente, la transformada M/S inversa (por ejemplo, la etapa 324) se puede aplicar a los vectores de forma de ruido descuantificados $v_{m,q}$ y $v_{s,q}$ (este último que se sustituye, por ejemplo, por 0 en caso de que la energía sea baja, por lo tanto, se indica con 437' en la Fig. 2), para obtener los vectores intermedios v'_l y v'_r , como:

$$v'_l = \left\{ \frac{v_{m,q,1} + v_{m,q,1}}{2}, \dots, \frac{v_{m,q,N} + v_{s,q,N}}{2} \right\}$$

20

$$v'_r = \left\{ \frac{v_{m,q,1} - v_{s,q,1}}{2}, \dots, \frac{v_{m,q,N} - v_{s,q,N}}{2} \right\}$$

[0138] Usando estos vectores intermedios v'_l y v'_r y los vectores de forma de ruido no cuantificados v_l y v_r , se calculan dos valores de ganancia como

25

$$g_{l,q} = \frac{\sum_{i=1}^N v_{l,i} - \sum_{i=1}^N v'_{l,i}}{N}$$

$$g_{r,q} = \frac{\sum_{i=1}^N v_{r,i} - \sum_{i=1}^N v'_{r,i}}{N}$$

30 **[0139]**
como

Entonces, los dos valores de ganancia se pueden cuantificar linealmente (por ejemplo, en la etapa 328)

$$g_{X,q} = \min(\max(\lfloor g_X \times 1,5 + 45 \rfloor, 0), 127) \in [0,127].$$

35 son posibles otras cuantificaciones).

[0140] Las ganancias cuantificadas se pueden codificar en el flujo de bits SID (por ejemplo, como parte de los datos de parámetros de ruido de confort 401 o 403, y más en particular $g_{l,q}$ pueden ser parte de los primeros datos de ruido paramétricos, y $g_{r,q}$ puede ser parte de los segundos datos de ruido paramétricos), por ejemplo, usando siete bits para el valor de ganancia $g_{l,q}$ y/o siete bits para el valor de ganancia $g_{r,q}$ (también son posibles diferentes cantidades para cada valor de ganancia).

40 **[0141]** En el decodificador (por ejemplo, 200', 200a, 200b), los vectores de forma de ruido cuantificados (por ejemplo, parte de los datos de parámetros de ruido de confort 401 o 403, y más en particular de los primeros datos de ruido paramétricos y los segundos datos de ruido paramétricos) se pueden descuantificar, por ejemplo, en la etapa 212 (en particular, en cualquiera de las subetapas 212-M, 212-S).

[0142] Los valores de ganancia se pueden descuantificar, por ejemplo, en la etapa 212 (en particular, en cualquiera de las subetapas 212-L, 212-R) como

50

$$g_{l,deq} = \frac{(g_{l,q} - 45)}{1,5}$$

$$g_{r,deq} = \frac{(g_{r,q} - 45)}{1,5}$$

(el valor 45 depende de la cuantificación y puede ser diferente con diferentes cuantificaciones). (En la Fig. 2, se usan $g_{l,d}$ y $g_{r,d}$ en lugar de $g_{l,deq}$ y $g_{r,deq}$).

[0143] El valor de coherencia 404 se puede descuantificar (por ejemplo, en la etapa 212-C) como

$$c_q = 15 \times c_{ind}.$$

[0144] Si la bandera *no_side* (en la información adicional 402) es 1, la forma v_s descuantificada $v_{s,q}$ se establece a cero (valor 537') antes de calcular los vectores intermedios v'_l y v'_r (por ejemplo, en la etapa 516). El valor de ganancia correspondiente entonces se suma a todos los elementos del vector intermedio correspondiente para generar las formas de ruido descuantificadas $v_{l,q}$ y $v_{r,q}$ indicadas de manera compleja con 522) como

$$v_{l,q} = \{v'_{l,1} + g_{L,deq}, \dots, v'_{l,N} + g_{L,deq}\}$$

$$v_{r,q} = \{v'_{r,1} + g_{R,deq}, \dots, v'_{r,N} + g_{R,deq}\}.$$

(La suma se debe a que estamos en el dominio logarítmico y corresponde a una multiplicación con un factor en el dominio lineal).

[0145] Para la generación de ruido de confort, se pueden usar tres fuentes de ruido gaussiano N_1, N_2, N_3 (por ejemplo, 211a, 212a, 213a en la Fig. 3a, 211b, 212b, 212c en la Fig. 3b, etc.) como se muestra en cualquiera de Las Figs. 3a-3f (o se puede usar cualquiera de las otras técnicas). Cuando la coherencia de canal es alta, se suma principalmente ruido correlacionado a ambos canales, en tanto que se suma más ruido no correlacionado si la coherencia es baja.

[0146] Usando las tres fuentes de ruido, los espectros DFT de las señales de ruido de canal izquierdo y derecho N_l (201) y N_r (203) se pueden calcular como

$$N_l[k] = \sqrt{1 - c_q} \times (N_1[k] + j \times N_1[k + M]) + \sqrt{c_q} \times (N_2[k] + j \times N_2[k + M])$$

$$N_r[k] = \sqrt{1 - c_q} \times (N_3[k] + j \times N_3[k + M]) + \sqrt{c_q} \times (N_2[k] + j \times N_2[k + M])$$

con $k \in \{0, 1, \dots, M - 1\}$ y $j^2 = -1$. Aquí, M denota la longitud de bloque de la DFT. Para generar ruido independiente tanto en la parte real como en la imaginaria del espectro complejo, cada fuente de ruido tiene que generar $2 \times M$ valores (dos para un intervalo de frecuencia) por trama. Por lo tanto, N_1, N_2 y N_3 (en respectivamente 211, 212, 213 en la Fig. 3f) se pueden ver como vectores de ruido de valor real que tienen una longitud de $2 \times M$, mientras que N_r y N_k (respectivamente en 201, 203) son vectores de valor complejo de longitud M.

[0147] Posteriormente, las señales de ruido en los dos canales pueden tener forma espectral (por ejemplo, en el modificador de señal 252) usando su forma de ruido correspondiente ($v_{l,q}$ o $v_{r,q}$) decodificada a partir del flujo de bits 232 y posteriormente transformada de nuevo desde el dominio logarítmico al dominio escalar, y desde el dominio de frecuencia al dominio de tiempo, por ejemplo, como se describe en [6] para generar una señal de ruido de confort estereofónico.

[0148] Se puede realizar cualquiera de los ejemplos del procesamiento mediante un controlador adecuado.

Algunas ventajas

[0149] La presente invención proporciona una técnica para generación de ruido de confort estéreo especialmente adecuada para esquemas de codificación estéreo discretos. Al codificar y transmitir conjuntamente parámetros de forma de ruido para ambos canales, el CNG estéreo se puede aplicar sin la necesidad de una mezcla descendente mono.

[0150] Junto con los dos conjuntos individuales de parámetros de ruido, la mezcla de una fuente de ruido común y dos fuentes de ruido individuales controladas por un único valor de coherencia permite la reconstrucción fiel de la imagen estéreo del ruido de fondo sin necesidad de transmitir parámetros estéreo refinados que normalmente solo están presentes en codificadores de audio paramétricos. Dado que solo se emplea este parámetro, la codificación del SID es sencilla sin la necesidad de procedimientos de compresión sofisticados en tanto que se mantiene bajo el tamaño de trama SID.

Algunos aspectos importantes:

[0151] En algunos ejemplos, se obtiene al menos uno de los siguientes aspectos:

1. Se genera ruido de confort para la señal estereofónica al mezclar tres fuentes de ruido gaussiano, una para cada canal y la tercera fuente de ruido común para crear ruido de fondo correlacionado.
2. Controlar la mezcla de las fuentes de ruido con el valor de coherencia que se transmite con la trama SID.
3. Transmitir parámetros de forma de ruido individuales para ambos canales estéreo al codificar conjuntamente las formas de ruido de una manera M/S. Velocidad de bits de trama SID inferior al codificar la forma S con menos bits que M.

Ventajas de las realizaciones

[0152] La inserción de una fuente de ruido común para los dos canales para imitar el ruido correlacionado para generar el ruido de confort final juega un papel importante en la imitación de grabación de ruido de fondo estereofónico.

[0153] Las realizaciones de la invención también se pueden considerar como un procedimiento para generar ruido de confort para la señal estereofónica al mezclar tres fuentes de ruido gaussiano, una para cada canal y la tercera fuente de ruido común para crear ruido de fondo correlacionado, o adicional o separadamente, para controlar la mezcla de las fuentes de ruido con el valor de coherencia que se transmite con la trama SID, o además o de forma separada, como sigue: En un sistema estéreo, generar el ruido de fondo por separado conduce a un ruido completamente no correlacionado que suena desagradable y es muy diferente del ruido de fondo real que causa transiciones audibles abruptas cuando cambiamos a/desde el fondo de modo activo a los fondos de modo DTX. En una realización, en el lado de codificador, además de los parámetros de ruido se calcula la coherencia de los dos canales, se cuantifica uniformemente y se añade a la trama SID. En el decodificador, la operación CNG se controla entonces mediante el valor de coherencia transmitido. Se utilizan tres fuentes de ruido gaussiano N_1, N_2, N_3; cuando la coherencia de canal es alta, se añade principalmente ruido correlacionado a ambos canales, mientras se añade más ruido no correlacionado si la coherencia es baja.

[0154] Una señal codificada se puede almacenar en un medio de almacenamiento digital o un medio de almacenamiento no transitorio o se puede transmitir en un medio de transmisión tal como un medio de transmisión inalámbrico o un medio de transmisión alámbrico tal como Internet.

[0155] Aunque algunos aspectos se han descrito en el contexto de un aparato, es evidente que estos aspectos también representan una descripción del procedimiento correspondiente, donde un bloque o aparato corresponde a una etapa de procedimiento o una característica de una etapa de procedimiento. Análogamente, los aspectos descritos en el contexto de una etapa de procedimiento también representan una descripción de un bloque o elemento o característica correspondiente de un aparato correspondiente.

[0156] Dependiendo de ciertos requisitos de implementación, las realizaciones de la invención se pueden implementar en hardware o en software. La implementación se puede realizar utilizando un medio de almacenamiento digital, por ejemplo, un disquete, un DVD, un CD, una ROM, una PROM, una EPROM, una EEPROM o una memoria FLASH, que tenga señales de control electrónicamente legibles almacenadas en el mismo, que cooperan (o sean capaces de cooperar) con un sistema informático programable de manera que se realiza el procedimiento respectivo.

[0157] Algunas realizaciones según la invención comprenden un soporte de datos que tiene señales de control electrónicamente legibles, que son capaces de cooperar con un sistema informático programable, de modo que se lleva a cabo uno de los procedimientos descritos en esta invención.

[0158] En general, las realizaciones de la presente invención se pueden implementar como un producto de programa informático con un código de programa, el código de programa que es operativo para realizar uno de los procedimientos, cuando el producto de programa informático se ejecuta en un ordenador. El código de programa se puede almacenar, por ejemplo, en un soporte legible por máquina.

[0159] Otras realizaciones comprenden el programa informático para realizar uno de los procedimientos descritos en esta invención, almacenado en un soporte legible por máquina o un medio de almacenamiento no transitorio.

[0160] En otras palabras, una realización del procedimiento inventivo es, por lo tanto, un programa informático que tiene un código de programa para realizar uno de los procedimientos descritos en esta invención, cuando el programa informático se ejecuta en un ordenador.

[0161] Una realización adicional de los procedimientos inventivos es, por lo tanto, un soporte de datos (o un medio de almacenamiento digital, o un medio legible por ordenador) comprendiendo, registrado en el mismo, el programa informático para realizar uno de los procedimientos descritos en esta invención.

[0162] Una realización adicional comprende un medio de procesamiento, por ejemplo, un ordenador o un dispositivo lógico programable, configurado para, o adaptado para, realizar uno de los procedimientos descritos en esta invención.

5 **[0163]** Una realización comprende además un ordenador que tiene instalado en sí el programa informático para realizar uno de los procedimientos descritos en esta invención.

[0164] En algunas realizaciones, un dispositivo lógico programable (por ejemplo, una matriz de compuertas programable en el campo) se puede usar para realizar algunas o todas las funcionalidades de los procedimientos descritos en esta invención. En algunas realizaciones, una matriz de compuertas programable en el campo puede cooperar con un microprocesador con el fin de realizar uno de los procedimientos descritos en esta invención. En general, los procedimientos se llevan a cabo preferentemente mediante cualquier aparato de hardware.

15 **[0165]** Las realizaciones anteriormente descritas son sólo ilustrativas de los principios de la presente invención. Se entiende que las modificaciones y variaciones de las disposiciones y los detalles descritos en esta invención serán evidentes para otros expertos en la técnica. Por lo tanto, se pretende que se limite solo por el alcance de las reivindicaciones de patente inminentes y no por los detalles específicos presentados a modo de descripción y explicación de las realizaciones en esta invención.

20 **Bibliografía o referencias**

[0166]

- [1] UIT-T G.729 Anexo B Un esquema de compresión de silencio para G.729 optimizado para terminales conforme a la Recomendación UIT-T v.70. Unión Internacional de Telecomunicaciones (UIT) Serie G, 2007.
- 25 [2] UIT-T G.729.1 Anexo C esquema DTX/GNC. Unión Internacional de Telecomunicaciones (UIT) Serie G, 2008.
- [3] ITU-T G.718 Codificación integrada de velocidad de bits variable de voz y audio de 8 a 32 kbit/s de banda estrecha y banda ancha robusta contra errores de trama. Unión Internacional de Telecomunicaciones (UIT) Serie G, 2008.
- [4] Funciones de procesamiento de voz del códec de voz obligatorio; códec de voz adaptativo de múltiples velocidades (AMR); Funciones de transcodificación, Especificación técnica TS 26.090 de 3GPP, 2014.
- 30 [5] Códec de voz adaptativo de velocidad múltiple - banda ancha (AMR-WB); Funciones de transcodificación, 3GPP, 2014.
- [6] 3GPP TS 26.445, Códec para Servicios de Voz Mejorados (EVS); Descripción algorítmica detallada.
- [7] Z. Wang y col., "Linear prediction based comfort noise generation in the EVS codec", en la Conferencia Internacional de IEEE sobre Acústica, Procesamiento de Señales y Habla (ICASSP), Brisbane, QLD, 2015.
- 35 [8] A. Lombard, S. Wilde, E. Ravelli, S. Döhla, G. Fuchs y M. Dietz, "Frequency-domain Comfort Noise Generation for Discontinuous Transmission in EVS", en la Conferencia Internacional de IEEE sobre Acústica, Procesamiento de Señales y Habla (ICASSP), Brisbane, QLD, 2015.
- [9] A. Lombard, M. Dietz, S. Wilde, E. Ravelli, P. Setiawan y M. Multrus, "Generation of a comfort noise with high spectro-temporal resolution in discontinuous transmission of audio signals". Patente de Estados Unidos 9583114B2, 19 de junio de 2015.
- 40 [10] E. NORVELL y F. JANSSON, "SUPPORT FOR GENERATION OF COMFORT NOISE. AND GENERATION OF COMFORT NOISE". WO Patente WO 2019/193149 A1, 5 de abril de 2019.

45

REIVINDICACIONES

1. Generador de señales multicanal (200) para generar una señal multicanal (204) que tiene un primer canal (201) y un segundo canal (203), comprendiendo:

5 una primera fuente de audio (211) para generar una primera señal de audio (221);
 una segunda fuente de audio (213) para generar una segunda señal de audio (223);
 una fuente de ruido de mezcla (212) para generar una señal de ruido de mezcla (222); y
 10 un mezclador (206) para mezclar la señal de ruido de mezcla (222) y la primera señal de audio (221) para obtener
 el primer canal (201) y para mezclar la señal de ruido de mezcla (222) y la segunda señal de audio (222) para
 obtener el segundo canal (203), donde el mezclador (206) comprende:

un primer elemento de amplitud (208-1) para influir en una amplitud de la primera señal de audio (221);
 un primer sumador (206-1) para sumar una señal de salida (221) del primer elemento de amplitud y al menos
 15 una porción de la señal de ruido de mezcla (222);
 un segundo elemento de amplitud (208-3) para influir en una amplitud de la segunda señal de audio (223);
 un segundo sumador (206-3) para sumar una salida (223) del segundo elemento de amplitud (208-3) y al menos
 una porción de la señal de ruido de mezcla (222),
 20 donde una cantidad de influencia realizada por el primer elemento de amplitud (208-1) y una cantidad de in-
 fluencia realizada por el segundo elemento de amplitud (208-3) son iguales entre sí o la cantidad de influencia
 realizada por el segundo elemento de amplitud (208-3) es diferente en menos de 20 por ciento de la cantidad
 realizada por el primer elemento de amplitud (208-1),
caracterizado porque el mezclador (206) comprende un tercer elemento de amplitud (208-2) para influir en
 una amplitud de la señal de ruido de mezcla (222),
 25 donde una cantidad de influencia realizada por el tercer elemento de amplitud (208-2) depende de la cantidad
 de influencia realizada por el primer elemento de amplitud (208-1) o el segundo elemento de amplitud (208-3),
 de modo que la cantidad de influencia realizada por el tercer elemento de amplitud (208-2) se vuelve mayor
 cuando la cantidad de influencia realizada por el primer elemento de amplitud o la cantidad de influencia reali-
 zada por el segundo elemento de amplitud (208-3) se vuelve menor.

2. El generador de señales de canal según la reivindicación 1, donde la primera fuente de audio (211) es
 una primera fuente de ruido y la primera señal de audio (221) es una primera señal de ruido, y/o la segunda fuente de
 audio (213) es una segunda fuente de ruido y la segunda señal de audio (223) es una segunda señal de ruido,
 35 donde la primera fuente de ruido (211) y/o la segunda fuente de ruido (213) se configura para generar la primera señal
 de ruido (221) y/o la segunda señal de ruido (223) de modo que la primera señal de ruido (221) y/o la segunda señal
 de ruido (223) se decorrelaciona de la señal de ruido de mezcla (222).

3. Generador de señales multicanal según la reivindicación 1 o 2, donde el mezclador (206) se configura
 para generar el primer canal (201) y el segundo canal (203) de modo que una cantidad de la señal de ruido de mezcla
 40 (222) en el primer canal (201) es igual a una cantidad de la señal de ruido de mezcla (222) en el segundo canal (203)
 o está dentro de un intervalo de 80 por ciento a 120 por ciento de la cantidad de la señal de ruido de mezcla (222) en
 el segundo canal (203).

4. Generador de señales multicanal según una de las reivindicaciones anteriores, donde el mezclador
 45 (206) comprende una entrada de control para recibir un parámetro de control (404, c), y donde el mezclador (206) se
 configura para controlar una cantidad de la señal de ruido de mezcla (222) en el primer canal (201) y el segundo canal
 (203) en respuesta al parámetro de control (404, c).

5. Generador de señales multicanal según una de las reivindicaciones anteriores, donde cada una de la
 50 primera fuente de audio (211), la segunda fuente de audio (213) y la fuente de ruido de mezcla (212) es una fuente de
 ruido gaussiano.

6. Generador de señales multicanal según una de las reivindicaciones anteriores,
 55 donde la primera fuente de audio (211) comprende un primer generador de ruido para generar la primera señal de
 audio (221) como una primera señal de ruido, donde la segunda fuente de audio (213) comprende un decorrela-
 cionador para decorrelacionar la primera señal de ruido (221) para generar la segunda señal de audio (213) como
 una segunda señal de ruido, y donde la fuente de ruido de mezcla (212) comprende un segundo generador de
 ruido, o
 60 donde la primera fuente de audio (211) comprende un primer generador de ruido (211) para generar la primera
 señal de audio (221) como una primera señal de ruido, donde la segunda fuente de audio (213) comprende un
 segundo generador de ruido (213) para generar la segunda señal de audio (223) como una segunda señal de
 ruido, y donde la fuente de ruido de mezcla (212) comprende un decorrelacionador para la decorrelación de la
 primera señal de ruido (221) o la segunda señal de ruido (223) para generar la señal de ruido de mezcla (222), o

donde una de las primeras fuentes de audio (211), la segunda fuente de audio (213) y la fuente de ruido de mezcla (212) comprende un generador de ruido para generar una señal de ruido, y donde otra de las primeras fuentes de audio (211), la segunda fuente de audio (213) y la fuente de ruido de mezcla (212) comprende un primer decorrelacionador para la decorrelación de la señal de ruido, y donde además, una de las primeras fuentes de audio (211), la segunda fuente de audio (213) y la fuente de ruido de mezcla (212) comprende un segundo decorrelacionador para la decorrelación de la señal de ruido, donde el primer y el segundo decorrelacionador son diferentes, de modo que las señales de salida del primer y del segundo decorrelacionador están decorrelacionadas entre sí, o donde la primera fuente de audio (211) comprende un primer generador de ruido, donde la segunda fuente de audio (213) comprende un segundo generador de ruido, y donde la fuente de ruido de mezcla (212) comprende un tercer generador de ruido, donde el primer, el segundo y el tercer generador de ruido están configurados para generar señales de ruido mutuamente decorrelacionadas.

7. Generador de señales multicanal según una de las reivindicaciones anteriores, donde una de la primera fuente de audio (211), la segunda fuente de audio (213) y la fuente de ruido de mezcla (212) comprende un generador de secuencia de número pseudoaleatorio configurado para generar una secuencia de número pseudoaleatorio en respuesta a una semilla, y donde al menos dos de la primera fuente de audio (211), la segunda fuente de audio (213) y la fuente de ruido de mezcla (212) están configuradas para inicializar el generador de secuencia de números pseudoaleatorios utilizando diferentes semillas.

8. Generador de señales multicanal según una de las reivindicaciones 1 a 6, donde al menos una de la primera fuente de audio (211), la segunda fuente de audio (213) y la fuente de ruido de mezcla (212) se configura para operar usando una tabla de ruido prealmacenada, o donde al menos una de la primera fuente de audio (211), la segunda fuente de audio (213) y la fuente de ruido de mezcla (212) está configurada para generar un espectro complejo para una trama utilizando un primer valor de ruido para una parte real y un segundo valor de ruido para una parte imaginaria, donde, opcionalmente, al menos un generador de ruido está configurado para generar un valor espectral de ruido complejo para un intervalo de frecuencia k utilizando para una de las partes real e imaginaria un primer valor aleatorio a un índice k , y utilizando, para la otra de la parte real y la parte imaginaria, un segundo valor aleatorio con un índice $(k+M)$, donde el primer y el segundo valor de ruido están incluidos en una matriz de ruido, por ejemplo, derivada de un generador de secuencias de números aleatorios, una tabla de ruido o un proceso de ruido, que abarca desde un índice inicial hasta un índice final, siendo el índice inicial menor que M y el índice final igual o menor que $2M$, donde M y k son números enteros.

9. Generador de señales multicanal según una de las reivindicaciones anteriores, donde la cantidad de influencia realizada por el tercer elemento de amplitud (208-2) es la raíz cuadrada de un valor predeterminado (c_q) y una cantidad de influencia realizada por el primer elemento de amplitud (208-1) y una cantidad de influencia realizada por el segundo elemento de amplitud (208-3) es la raíz cuadrada de la diferencia entre 1 y el valor predeterminado (c_q).

10. Generador de señales multicanal según una de las reivindicaciones anteriores, comprendiendo además: una interfaz de entrada (210) para recibir datos de audio codificados (232) en una secuencia de tramas (306, 308) comprendiendo una trama activa (306) y una trama inactiva (308) después de la trama activa (306); y un decodificador de audio (200', 200a, 200b) para decodificar datos de audio codificados para la trama activa (306) para generar una señal multicanal decodificada para la trama activa, donde la primera fuente de audio (211), la segunda fuente de audio (213), la fuente de ruido de mezcla (212) y el mezclador (206) están activos en la trama inactiva (308) para generar la señal multicanal (204) para la trama inactiva.

11. Generador de señales multicanal según la reivindicación 10, donde: la señal de audio codificada (232) para la trama activa (306) tiene una primera pluralidad de coeficientes que describen un primer número de intervalos de frecuencia; y la señal de audio codificada (232) para la trama inactiva (308) tiene una segunda pluralidad de coeficientes que describen un segundo número de intervalos de frecuencia, donde el primer número de intervalos de frecuencia es mayor que el segundo número de intervalos de frecuencia.

12. Generador de señales multicanal según la reivindicación 10 u 11, donde los datos de audio codificados (232) para la trama inactiva (308) comprenden datos descriptores de inserción de silencio (p_noise , c) comprendiendo datos de ruido de confort (c , p_noise) que indican una energía de señal (1312) para cada canal de los dos canales (301, 303), o para cada una de las primeras combinaciones lineales del primer y segundo canal y una segunda combinación lineal del primer y segundo canal, para la trama inactiva e

indican una coherencia (404, c) entre el primer canal (301) y el segundo canal (303) en la trama inactiva, y donde el mezclador (206, 220) está configurado para mezclar (206-1, 206-3) la señal de ruido de mezcla (222) y la primera señal de audio (221) o la segunda señal de audio (223) basándose en los datos de ruido de confort que indican la coherencia (404, c), y

5 donde el generador de señales multicanal (200, 220, 220a-220e) comprende además un modificador de señal (250) para modificar el primer canal (201) y el segundo canal (203) o la primera señal de audio (221) o la segunda señal de audio (223) o la señal de ruido de mezcla (222),

10 donde el modificador de señal (250) está configurado para ser controlado por los datos de ruido de confort (p_noise), que indican las energías de señal para el primer canal de audio (301) y el segundo canal de audio (303), o que indican las energías de la señal para una primera combinación lineal de los canales primero y segundo, y una segunda combinación lineal de los canales primero y segundo.

13. Generador de señales multicanal según cualquiera de las reivindicaciones 10-12, donde los datos de audio (232) para la trama inactiva comprenden:

15 una primera trama de descriptor de inserción de silencio (241) para el primer canal (201) y una segunda trama de descriptor de inserción de silencio (243) para el segundo canal (203), donde la primera trama de descriptor de inserción de silencio (241) comprende datos de parámetros de ruido de confort (p_noise) para el primer canal (201), y/o para una primera combinación lineal del primer y segundo canales, y

20 información adicional de generación de ruido de confort (p_frame) para el primer canal y el segundo canal (203), y donde la segunda trama de descriptor de inserción de silencio (243) comprende

datos de parámetros de ruido de confort (p_noise) para el segundo canal (203), y/o para una segunda combinación lineal del primer y segundo canales e

25 información de coherencia (404, c) que indica una coherencia entre el primer canal (201) y el segundo canal (203) en la trama inactiva, y

30 donde el generador de señales multicanal comprende un controlador para controlar la generación de la señal multicanal (204) en la trama inactiva usando la información lateral de generación de ruido de confort (p_frame) para la primera trama de descriptor de inserción de silencio (241) para determinar un modo de generación de ruido de confort para el primer canal (201) y el segundo canal (203), y/o para una primera combinación lineal del primer y segundo canales y una segunda combinación lineal del primer y segundo canales, usando la información de coherencia (404, c) en la segunda trama de descriptor de inserción de silencio (243) para establecer una coherencia (404, c) entre el primer canal (201) y el segundo canal (203) en la trama inactiva, y usando los datos de parámetros de ruido de confort (p_noise) de la primera trama de descriptor de inserción de silencio (241) y usar los datos de parámetros de ruido de confort (p_noise) de la segunda trama de descriptor de inserción de silencio (243) para establecer una situación de energía ($v_{i, q}$) del primer canal (301) y una situación de energía ($v_{r, q}$) del segundo canal (303).

40 14. Generador de señales multicanal según cualquiera de las reivindicaciones 10-13, donde los datos de audio (232) para la trama inactiva comprenden:

45 al menos una trama de descriptor de inserción de silencio (241) para una primera combinación lineal del primer y segundo canales y una segunda combinación lineal del primer y segundo canales, donde la al menos una trama de descriptor de inserción de silencio (241) comprende

datos de parámetros de ruido de confort (p_noise) para la primera combinación lineal del primer y segundo canales, e

50 información adicional de generación de ruido de confort (p_frame) para la segunda combinación lineal del primer y segundo canales,

55 donde el generador de señales multicanal comprende un controlador para controlar la generación de la señal multicanal (204) en la trama inactiva usando la información adicional de generación de ruido de confort (p_frame) para la primera combinación lineal del primer y segundo canales y la segunda combinación lineal del primer y segundo canales, usando la información de coherencia (404, c) en la segunda trama de descriptor de inserción de silencio (243) para establecer una coherencia (404, c) entre el primer canal (201) y el segundo canal (203) en la trama inactiva, y usando los datos de parámetros de ruido de confort (p_noise) de la al menos una trama de descriptor de inserción de silencio (241) y usando los datos de parámetros de ruido de confort (p_noise) de la al menos una trama de descriptor de inserción de silencio (243) para establecer una situación de energía ($v_{i, q}$) del primer canal (301) y una situación de energía ($v_{r, q}$) del segundo canal (303).

15. Generador de señales multicanal según cualquiera de las reivindicaciones 12-14, comprendiendo además un convertidor de espectro-tiempo para convertir un primer canal resultante y un segundo canal resultante que se ajustan espectralmente y se ajustan en coherencia, en representaciones de dominio de tiempo correspondientes

para combinarse con o concatenarse con representaciones de dominio de tiempo de canales correspondientes de la señal multicanal decodificada para la trama activa.

5 **16.** Generador de señales multicanal según cualquiera de las reivindicaciones 10-15, donde los datos de audio para la trama inactiva comprenden:

una trama de descriptor de inserción de silencio (241, 243), donde la trama de descriptor de inserción de silencio (241, 243) comprende datos de parámetros de ruido de confort (p_{noise}) para el primer y segundo canales (201, 203) e información adicional de generación de ruido de confort (p_{frame}) para el primer canal (203) y el segundo canal (203) y/o para una primera combinación lineal del primer y segundo canales y una segunda combinación lineal del primer y segundo canales, e información de coherencia (404, c) que indica una coherencia entre el primer canal (201) y el segundo canal (203) en la trama inactiva, y
 10 donde el generador de señales multicanal (200) comprende un controlador para controlar la generación de la señal multicanal (202) en la trama inactiva usando la información adicional de generación de ruido de confort (p_{frame}) para la trama de descriptor de inserción de silencio (241, 243) para determinar un modo de generación de ruido de confort para el primer canal (201) y el segundo canal (203), usando la información de coherencia (404, c) en la trama de descriptor de inserción de silencio (241) para establecer una coherencia (404, c) entre el primer canal (201) y el segundo canal (203) en la trama inactiva, y usando los datos de parámetros de ruido de confort (p_{noise}) de la trama de descriptor de inserción de silencio (241, 243) para establecer una situación de energía ($v_{l,q}$) del primer canal (301) y una situación de energía ($v_{r,q}$) del segundo canal (303).
 15
 20

17. Generador de señales multicanal según cualquiera de las reivindicaciones 10-16,

25 donde los datos de audio codificados (232) para la trama inactiva comprenden datos de descriptor de inserción de silencio (p_{noise} , c) comprendiendo datos de ruido de confort (c, p_{noise}) que indican una energía de señal para cada canal en una representación media/lateral y datos de coherencia (404, c) que indican la coherencia entre el primer canal y el segundo canal en la representación izquierda/derecha, donde el generador de señales multicanal se configura para convertir la representación media/lateral de la energía de señal en una representación izquierda/derecha de la energía de señal en el primer canal (301) y el segundo canal (303),
 30 donde el mezclador (206, 220) se configura para mezclar (206-1, 206-3) la señal de ruido de mezcla (222) con la primera señal de audio (221) y la segunda señal de audio (223) con base en los datos de coherencia (404, c) para obtener el primer canal (201) y el segundo canal (203), y
 donde el generador de señales multicanal comprende además un modificador de señal (250) configurado para modificar el primer y segundo canales (201, 203) al conformar el primer y segundo canales (201, 203) con base en
 35 la energía de señal en el dominio izquierdo/derecho.

18. Generador de señales multicanal según la reivindicación 17, configurado, en caso de que los datos de audio contengan señalización que indica que la energía en el canal lateral es menor que un umbral predeterminado, para poner a cero (337) los coeficientes del canal lateral ($v_{s,q}$).

40 **19.** Generador de señales multicanal según la reivindicación 17 o 18, donde los datos de audio para la trama inactiva comprenden:

al menos una trama de descriptor de inserción de silencio (241, 243), donde la al menos una trama de descriptor de inserción de silencio (241, 243) comprende datos de parámetros de ruido de confort (p_{noise} , $v_{m,ind}$, $q_{l,q}$, $q_{r,q}$, $v_{s,ind}$) para el canal medio y el canal lateral ($v_{m,q}$, $v_{s,q}$) e información adicional de generación de ruido de confort (p_{frame}) para el canal medio y el canal lateral ($v_{m,q}$, $v_{s,q}$), e información de coherencia (404, c) que indica una coherencia entre el primer canal (201) y el segundo canal (203) en la trama inactiva, y
 45 donde el generador de señales multicanal (200) comprende un controlador para controlar la generación de la señal multicanal (202) en la trama inactiva usando la información adicional de generación de ruido de confort (p_{frame}) para la trama de descriptor de inserción de silencio (241, 243) para determinar un modo de generación de ruido de confort para el primer canal (201) y el segundo canal (203), usando la información de coherencia (404, c) en la trama de descriptor de inserción de silencio (241) para establecer una coherencia (404, c) entre el primer canal (201) y el segundo canal (203) en la trama inactiva, y usando los datos de parámetros de ruido de confort (p_{noise}),
 50 o una versión procesada de los mismos, de la trama de descriptor de inserción de silencio (241, 243) para establecer una situación de energía ($v_{l,q}$) del primer canal (301) y una situación de energía ($v_{r,q}$) del segundo canal (303).
 55

20. Generador de señales multicanal según cualquiera de las reivindicaciones 10-19, configurado además para modificar la escala de los coeficientes de energía de señal (1312, v'_l , v'_r) para el primer y segundo canales mediante información de ganancia ($g_{l,q}$, $q_{r,q}$), codificada con los datos de parámetros de ruido de confort (401, 403) para el primer y segundo canales.

21. Generador de señales multicanal según cualquiera de las reivindicaciones anteriores, configurado para

convertir la señal multicanal generada (252) de una versión de dominio de frecuencia a una versión de dominio de tiempo.

22. Generador de señales multicanal según cualquiera de las reivindicaciones anteriores, donde la primera fuente de audio (211) es una primera fuente de ruido y la primera señal de audio (221) es una primera señal de ruido, o la segunda fuente de audio (213) es una segunda fuente de ruido y la segunda señal de audio (223) es una segunda señal de ruido,

10 donde la primera fuente de ruido o la segunda fuente de ruido se configura para generar la primera señal de ruido (201) o la segunda señal de ruido (203) de modo que la primera señal de ruido (201) o la segunda señal de ruido (203) se correlacionan al menos parcialmente, y
 donde la fuente de ruido de mezcla (212) se configura para generar la señal de ruido de mezcla (222) con una primera porción de ruido de mezcla (221a) y una segunda porción de ruido de mezcla (221b), la segunda porción de ruido de mezcla (221b) que se decorrelaciona al menos parcialmente de la primera porción de ruido de mezcla (221a); y
 15 donde el mezclador (206) se configura para mezclar la primera porción de ruido de mezcla (221a) de la señal de ruido de mezcla (222) y la primera señal de audio (221) para obtener el primer canal (201) y para mezclar la segunda porción de ruido de mezcla (221b) de la señal de ruido de mezcla (222) y la segunda señal de audio (223) para obtener el segundo canal (203).

23. Procedimiento para generar una señal multicanal que tiene un primer canal y un segundo canal (203), comprendiendo:

25 generar una primera señal de audio (221) usando una primera fuente de audio (211);
 generar una segunda señal de audio (223) usando una segunda fuente de audio (213);
 generar una señal de ruido de mezcla (222) usando una fuente de ruido de mezcla (212); y
 mezclar (206) la señal de ruido de mezcla (222) y la primera señal de audio (221) para obtener el primer canal (201) y
 30 mezclar la señal de ruido de mezcla (222) y la segunda señal de audio (223) para obtener el segundo canal (202), comprendiendo el procedimiento:

usar un primer elemento de amplitud (208-1) que influye en una amplitud de la primera señal de audio (221);
 usar un primer sumador (206-1) que suma una señal de salida (221) del primer elemento de amplitud y al menos una porción de la señal de ruido de mezcla (222);
 35 usar un segundo elemento de amplitud (208-3) que influye en una amplitud de la segunda señal de audio (223);
 usar un segundo sumador (206-3) que suma una salida (223) del segundo elemento de amplitud (208-3) y al menos una porción de la señal de ruido de mezcla (222),
 donde una cantidad de influencia realizada por el primer elemento de amplitud (208-1) y una cantidad de influencia realizada por el segundo elemento de amplitud (208-3) son iguales entre sí o la cantidad de influencia realizada por el segundo elemento de amplitud (208-3) es diferente en menos de 20 por ciento de la cantidad realizada por el primer elemento de amplitud (208-1),
 40 **caracterizado porque** la mezcla (206) usa un tercer elemento de amplitud (208-2) que influye en una amplitud de la señal de ruido de mezcla (222),
 donde una cantidad de influencia realizada por el tercer elemento de amplitud (208-2) depende de la cantidad de influencia realizada por el primer elemento de amplitud (208-1) o el segundo elemento de amplitud (208-3),
 45 de modo que la cantidad de influencia realizada por el tercer elemento de amplitud (208-2) se vuelve mayor cuando la cantidad de influencia realizada por el primer elemento de amplitud o la cantidad de influencia realizada por el segundo elemento de amplitud (208-3) se vuelve menor.

50 24. Programa informático para realizar, cuando se ejecuta en un ordenador o un procesador, el procedimiento de la reivindicación 23.

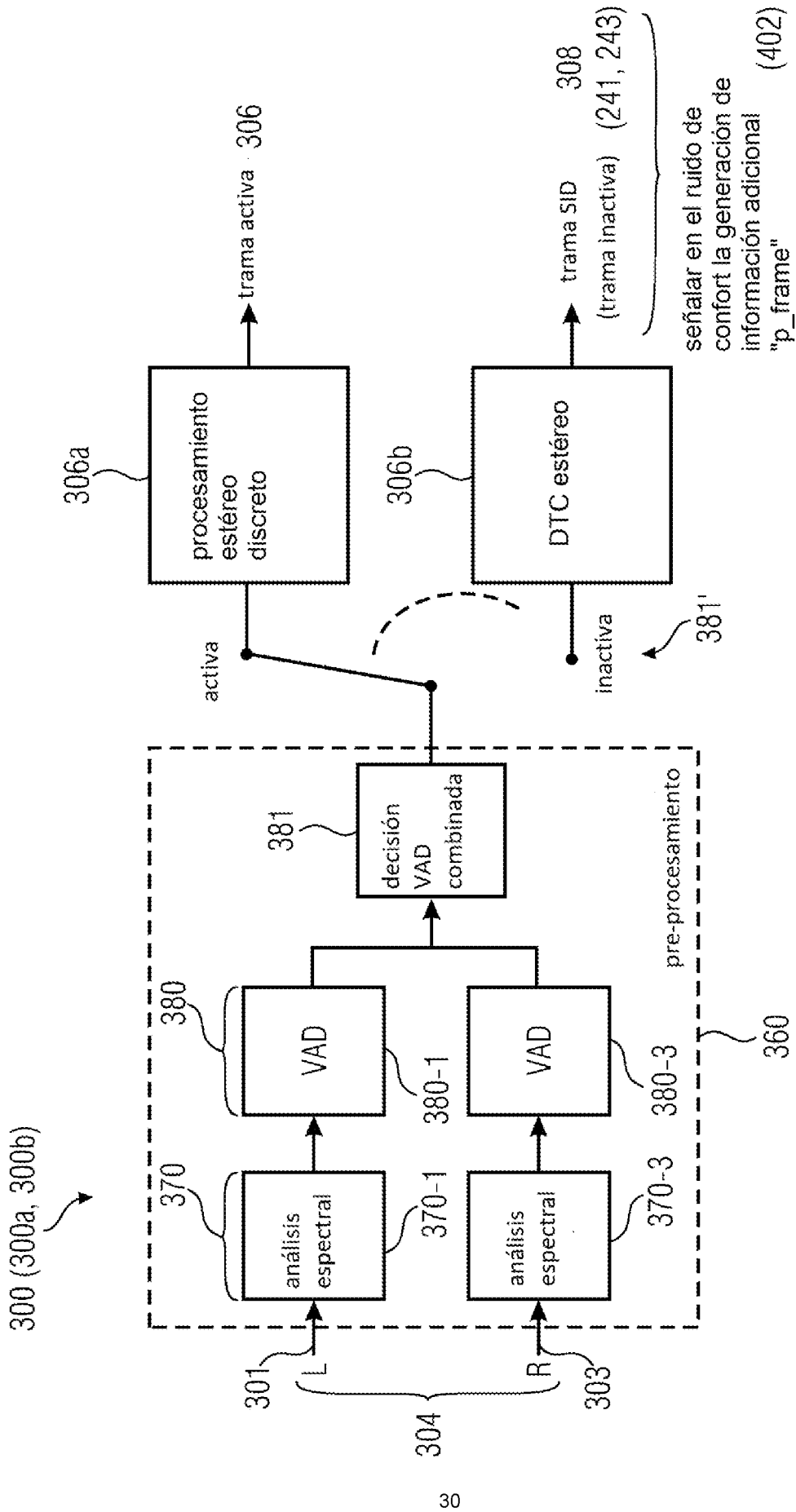


Fig. 1

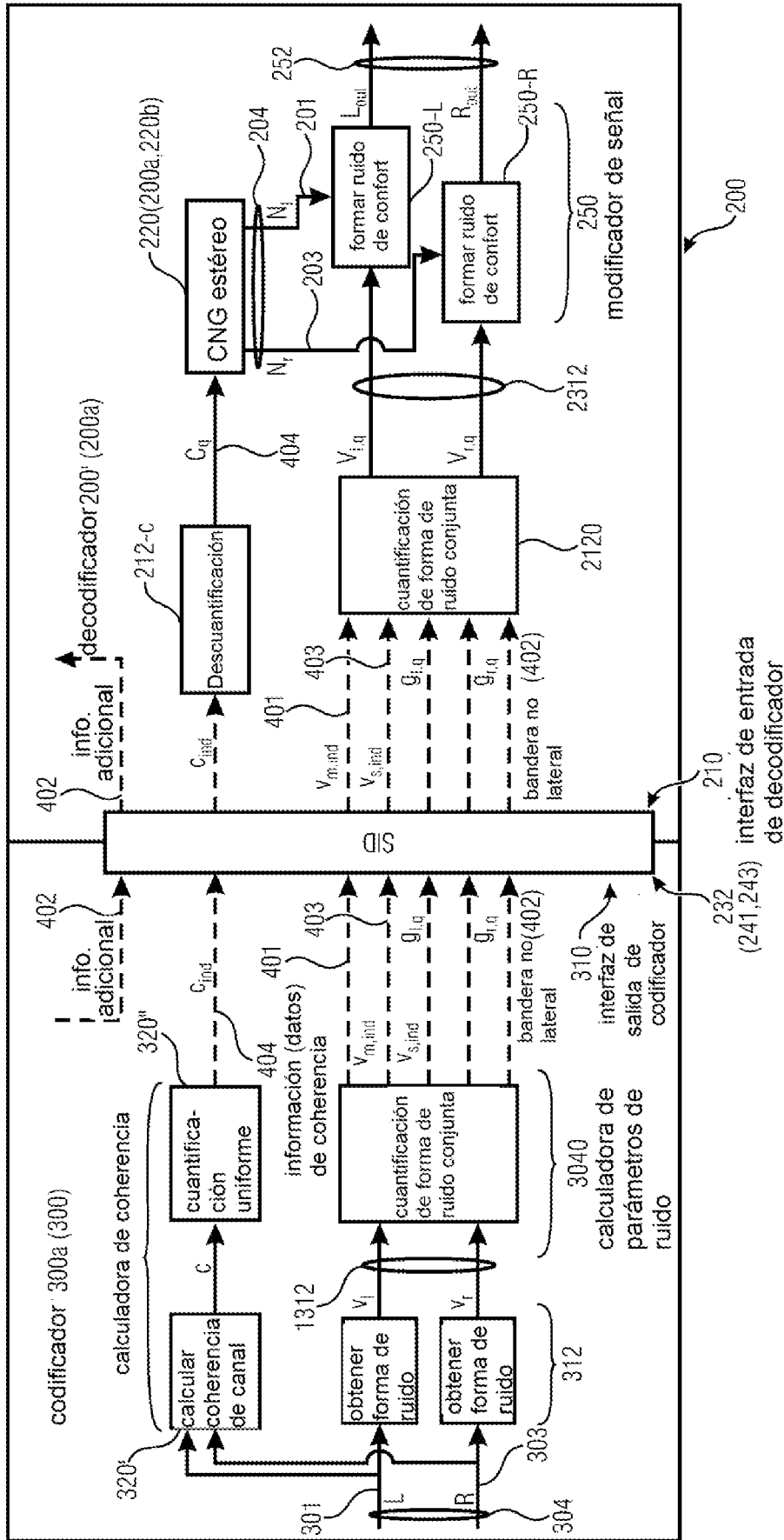


Fig. 2

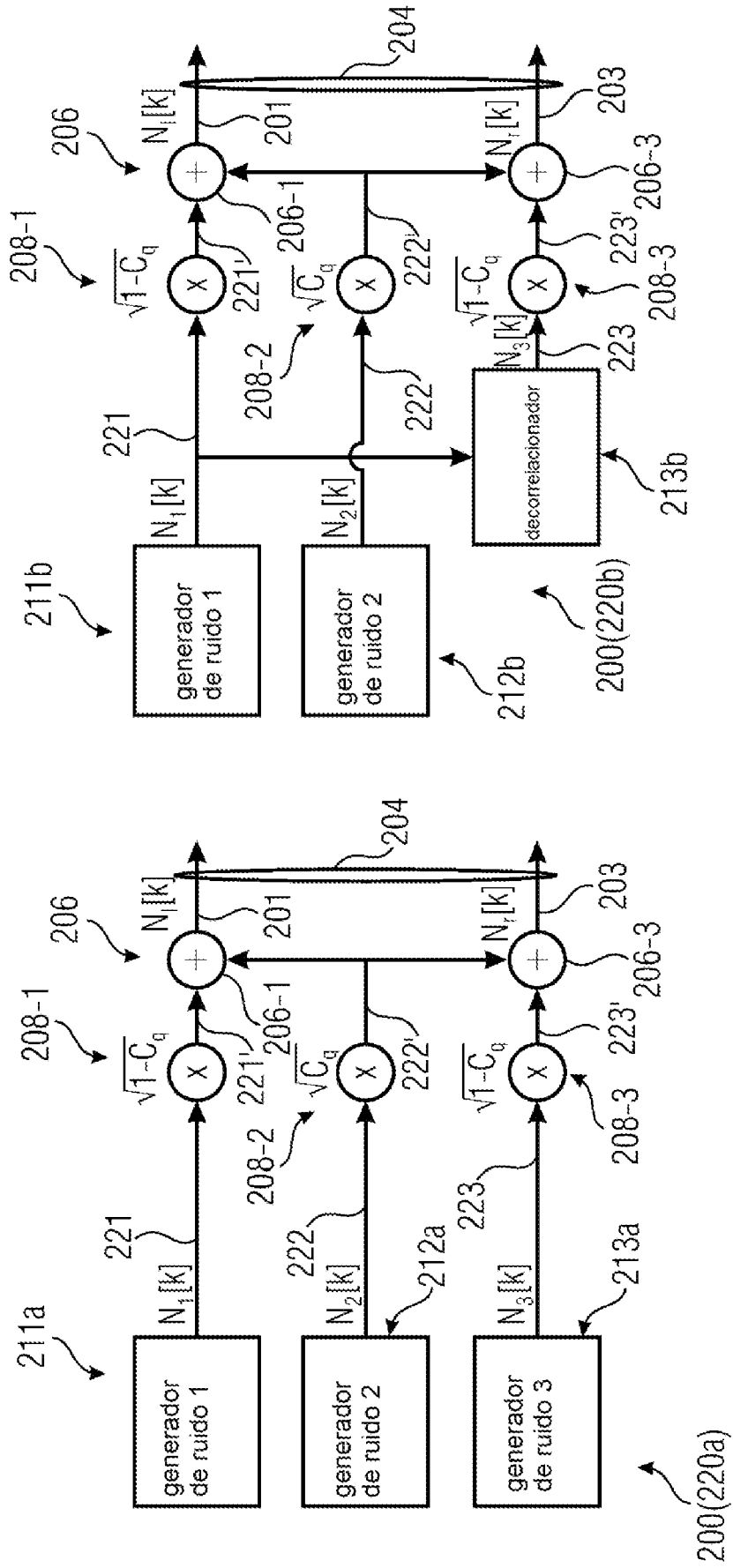


Fig. 3b

Fig. 3a

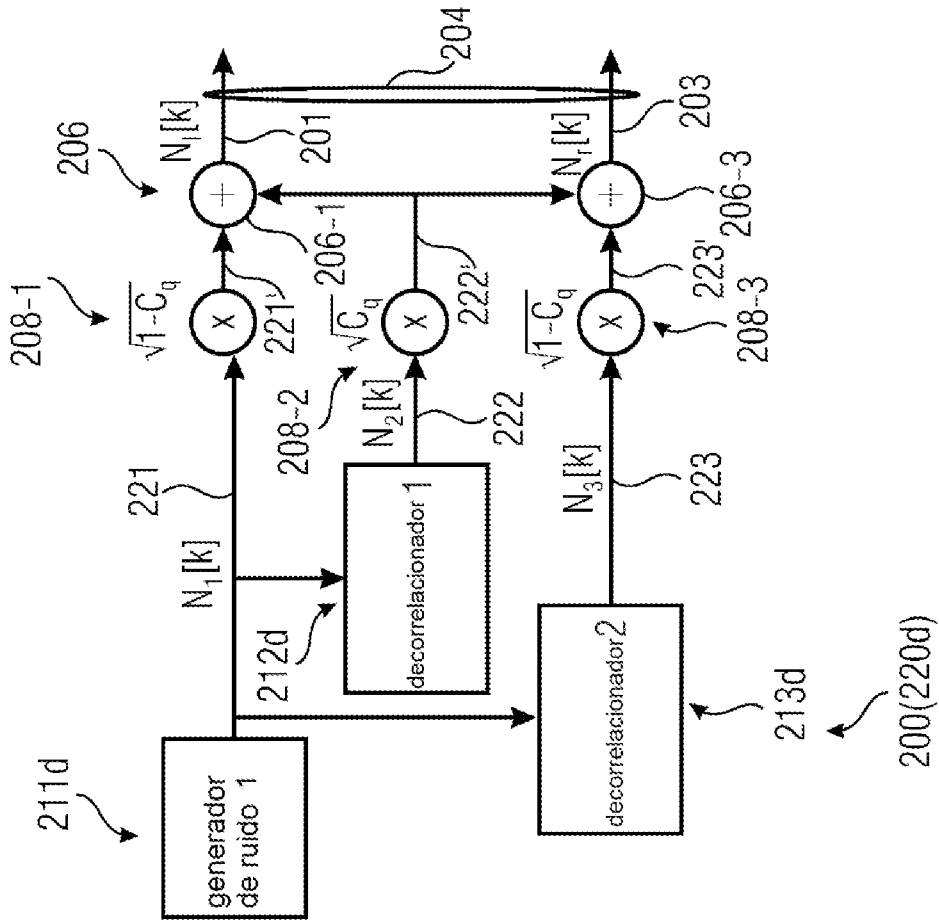


Fig. 3d

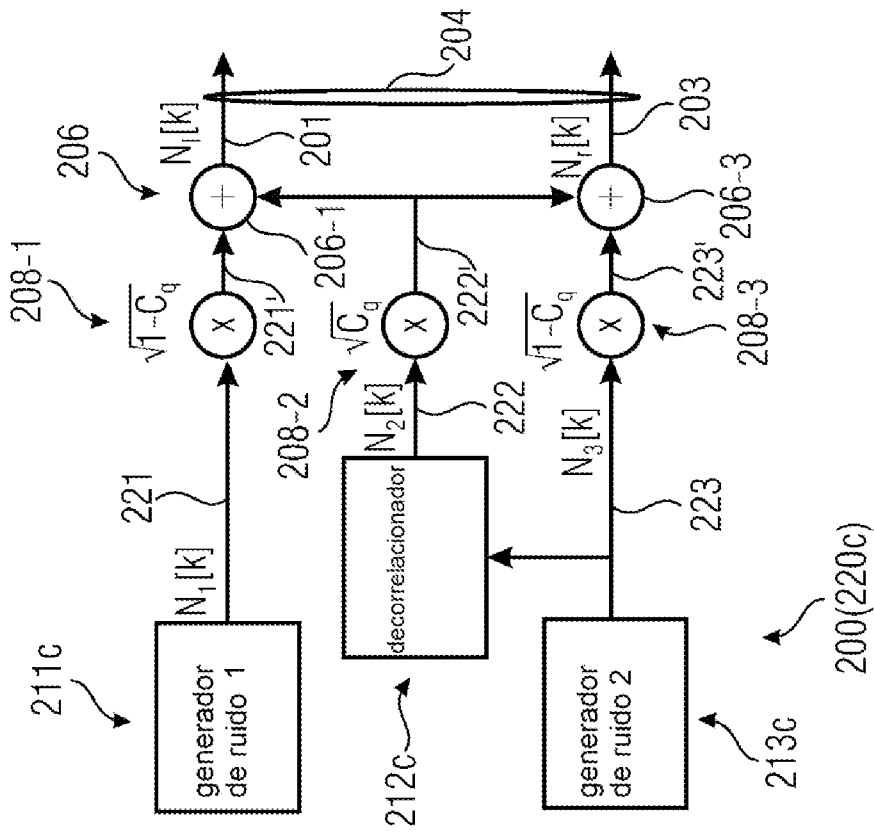


Fig. 3c

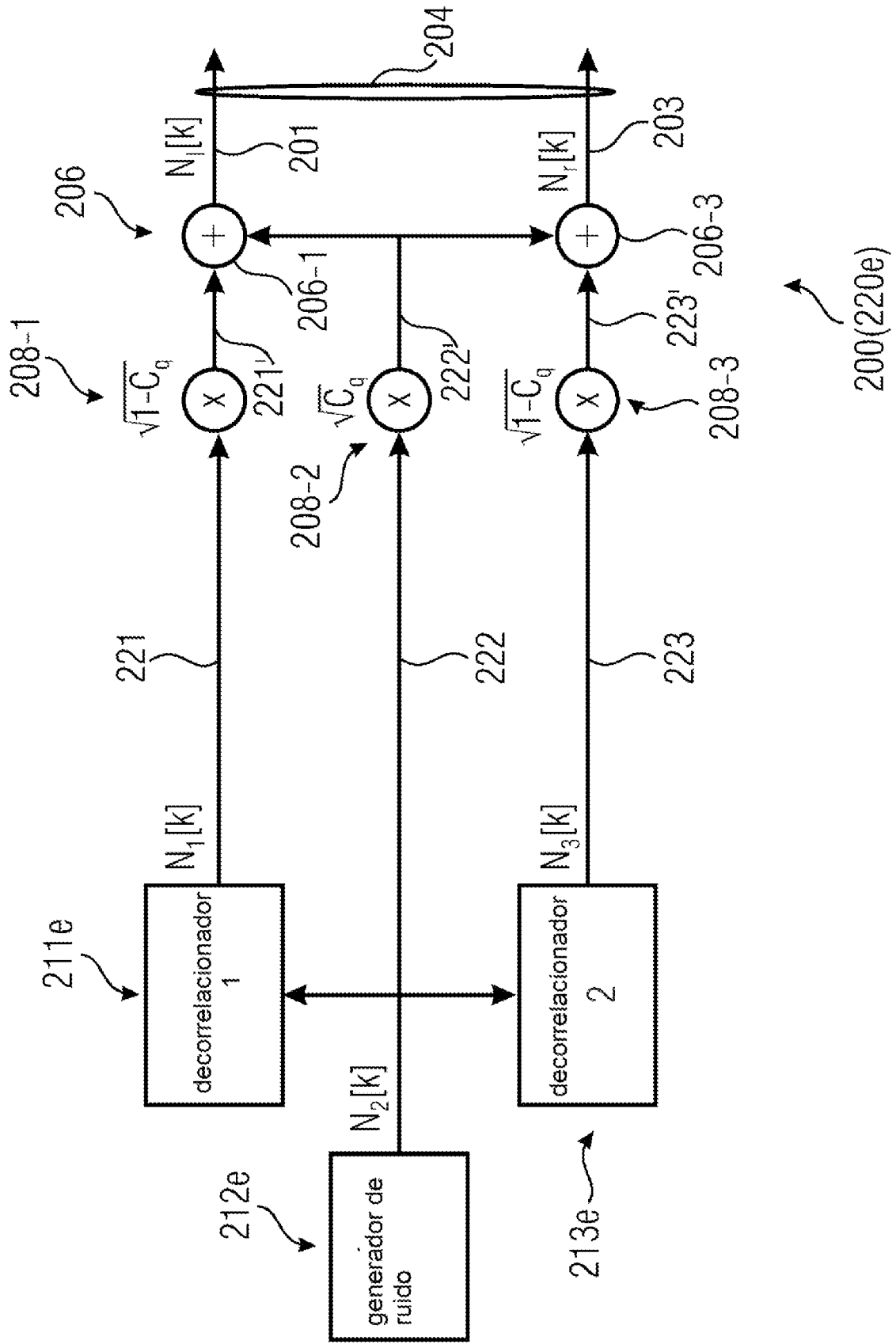


Fig. 3e

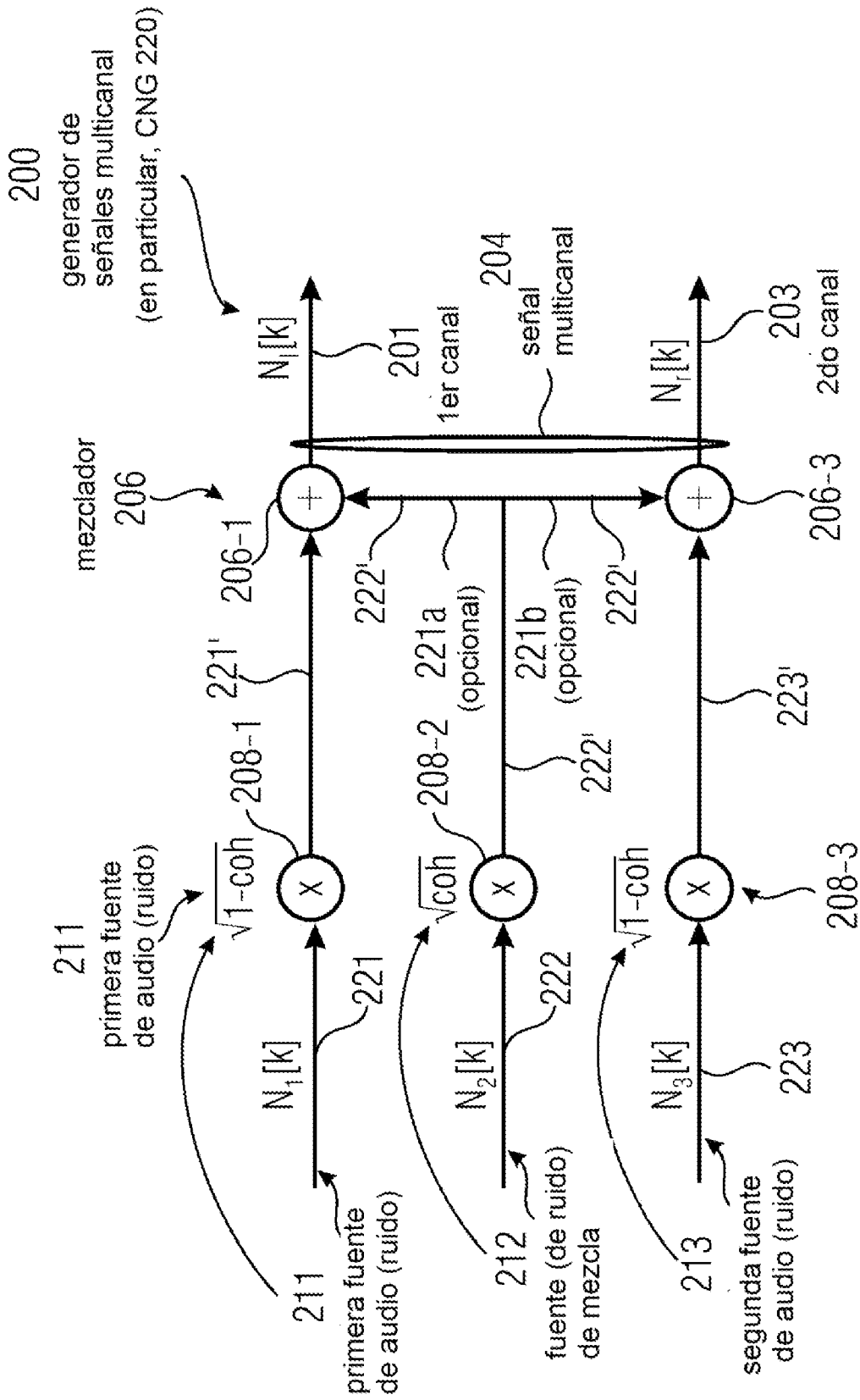


Fig. 3f

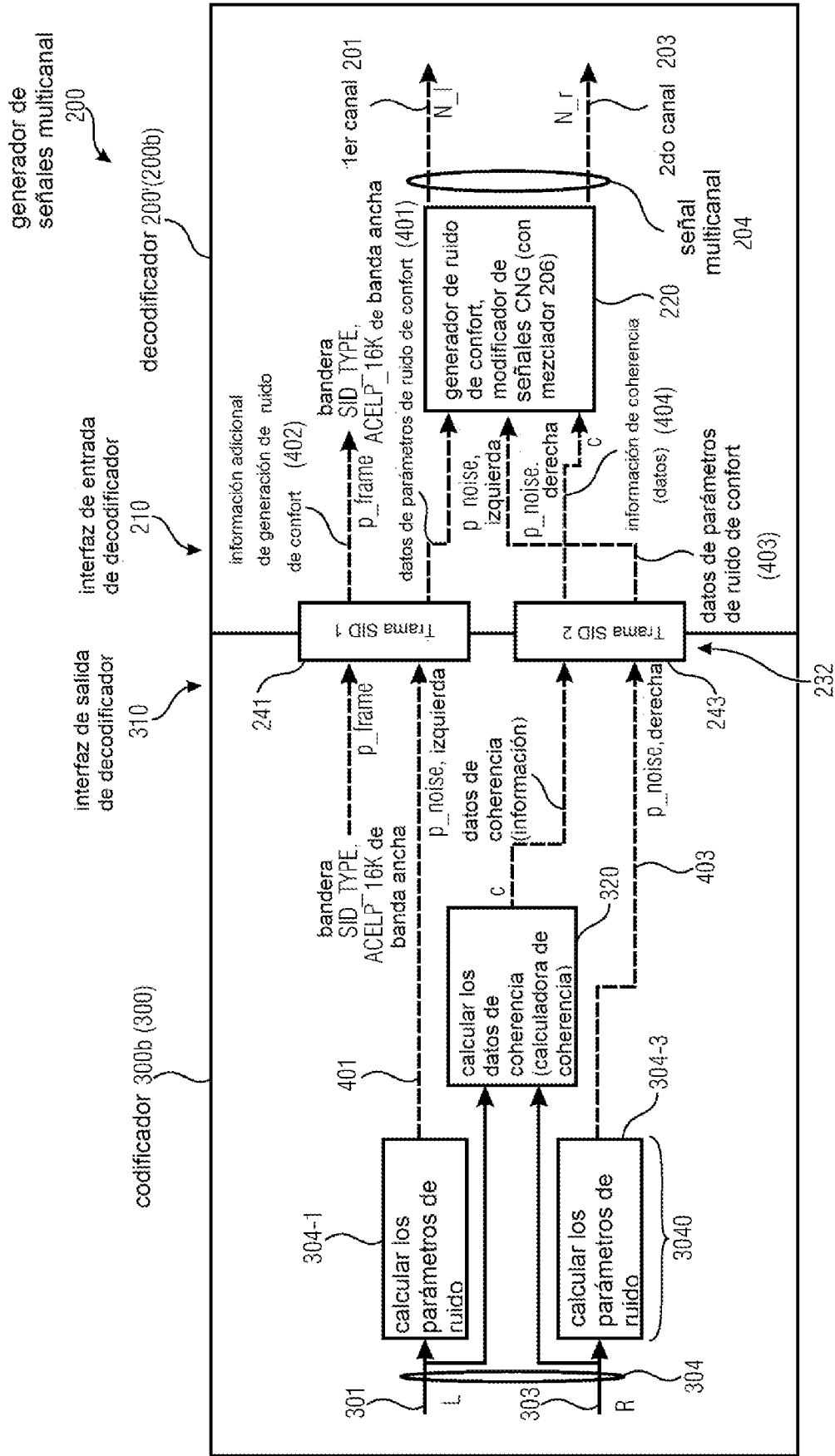
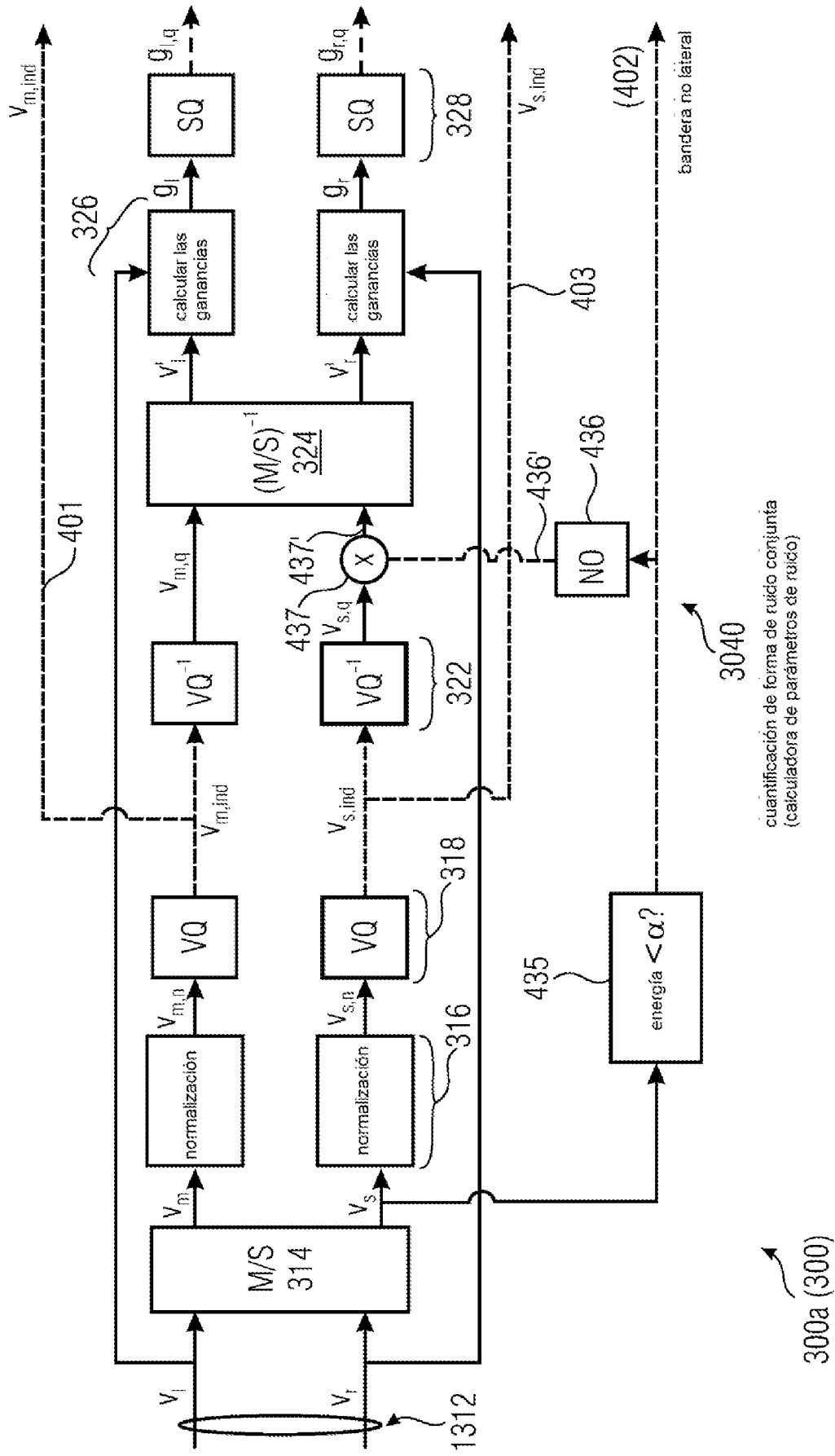


Fig. 4



300a (300)

3040
cuantificación de forma de ruido conjunta
(calculadora de parámetros de ruido)

Fig. 5

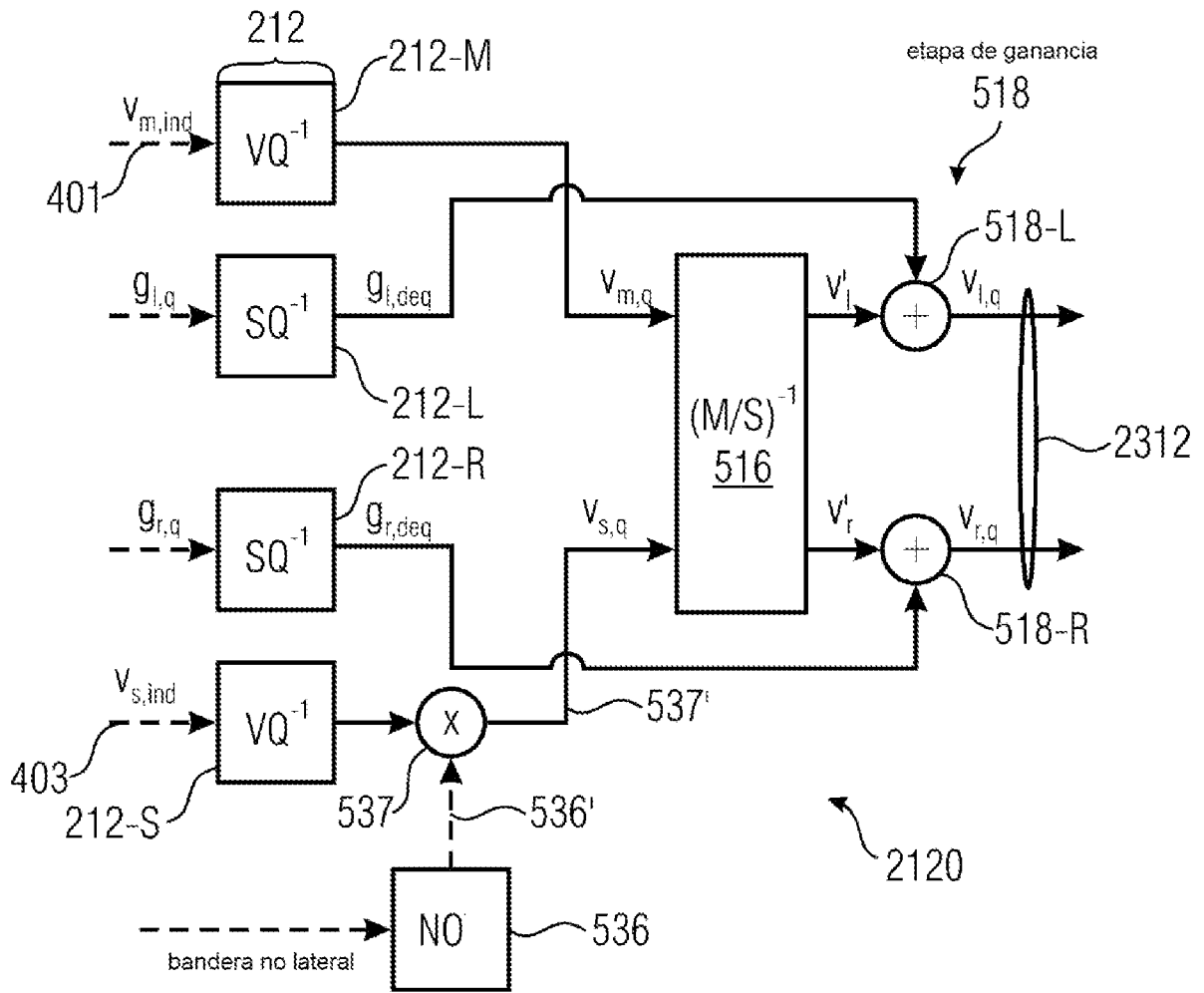


Fig. 6