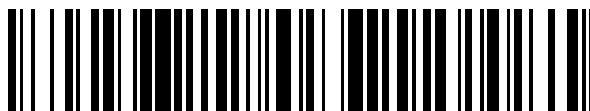


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 780 696**

51 Int. Cl.:

G10L 19/005 (2013.01)

G10L 19/09 (2013.01)

G10L 19/02 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **23.06.2014 PCT/EP2014/063176**

87 Fecha y número de publicación internacional: **24.12.2014 WO14202789**

96 Fecha de presentación y número de la solicitud europea: **23.06.2014 E 14732196 (2)**

97 Fecha y número de publicación de la concesión europea: **25.12.2019 EP 3011563**

54 Título: **Decodificación de audio con reconstrucción de tramas no recibidas o alteradas mediante el uso de TCX LTP**

30 Prioridad:

21.06.2013 EP 13173154
05.05.2014 EP 14166998

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
26.08.2020

73 Titular/es:

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

72 Inventor/es:

**SCHNABEL, MICHAEL;
MARKOVIC, GORAN;
SPERSCHNEIDER, RALPH;
LECOMTE, JÉRÉMIE y
HELMRICH, CHRISTIAN**

74 Agente/Representante:

SALVÀ FERRER, Joan

ES 2 780 696 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

Decodificación de audio con reconstrucción de tramas no recibidas o alteradas mediante el uso de TCX LTP

5 **[001]** La presente invención se refiere a la codificación, el procesamiento y la decodificación de señales de audio y, en particular, a un aparato y un procedimiento para el desvanecimiento de la señal mejorado para sistemas de codificación de audio cambiados durante el ocultamiento de errores.

[002] En lo que sigue, se describe el estado de la técnica con respecto al desvanecimiento de la voz y códecs de audio durante el ocultamiento de pérdida de paquetes (PLC), conforme a sus siglas en inglés). Las explicaciones con respecto al estado de la técnica con los códecs ITU-T (Normalización de la Unión Internacional de Telecomunicaciones) de la serie G (G.718, G.719, G.722, G.722.1, G.729, G.729.1) son seguidas de los códecs 3GPP (AMR, AMR-WB, AMR-WB+) y un códec IETF (OPUS), y concluyen con dos códecs MPEG (HE-AAC, HILN) (ITU = sigla en inglés de "Unión Internacional de Telecomunicaciones"; 3GPP = sigla en inglés de "Proyecto Asociación de Tercera Generación"; AMR = sigla en inglés de "Multitasa Adaptativa"; WB = sigla en inglés de "banda ancha"; IETF = sigla en inglés de: Grupo de Trabajo de Ingeniería de Internet"). A continuación, se analiza el estado de la técnica con respecto al rastreo del nivel de ruido de fondo, seguido de una síntesis que proporciona una reseña.

[003] En primer lugar, se considera G.718. G.718 es un códec de voz de banda estrecha y de banda ancha, que soporta DTX/CNG (DTX = Digital Theater Systems (Sistemas de Teatro Digital); CNG = Generación de Ruido de Confort). Como ejemplos útiles para la comprensión de la invención se hace referencia en particular al código de retardo bajo, el modo de versión de retardo bajo se describirá aquí en más detalle.

[004] Considerando ACELP (Capa 1) (ACELP = Predicción Lineal Excitada por Código Algebraico), la ITU-T recomienda para G.718 [ITU08a, sección 7.11] un desvanecimiento adaptativo en el dominio predictivo lineal para controlar la velocidad de desvanecimiento. En general, el ocultamiento sigue este principio: Según G.718, en el caso de eliminaciones o borrados de tramas, la estrategia de ocultamiento puede sintetizarse como una convergencia de la energía de señal y la envoltura espectral para los parámetros estimados del ruido de fondo. La periodicidad de la señal converge a cero. La velocidad de la convergencia depende de los parámetros de la última trama correctamente recibida y el número de tramas borradas consecutivas, y es controlada por un factor de atenuación, α . El factor de atenuación α depende además de la estabilidad, θ , del filtro LP (LP = sigla en inglés de Predicción Lineal) para tramas SIN VOZ. En general, la convergencia es lenta si la última trama recibida buena se encuentra en un segmento estable y es rápida si la trama se encuentra en un segmento de transición.

35 **[005]** El factor de atenuación α depende de la clase de señal de la voz, que deriva por la clasificación de señal descrita en [ITU08a, sección 6.8.1.3.1 y 7.11.1.1]. El factor de estabilidad θ se computa sobre la base de una medida de distancia entre los filtros adyacentes ISF (sigla en inglés de Frecuencia Espectral de Impedancia y Admitancia) [ITU08a, sección 7.1.2.4.2].

40 **[006]** La Tabla 1 muestra el esquema de cálculo de α :

Tabla 1: valores del factor de atenuación α , el valor θ es un factor de estabilidad computado a partir de una medida de distancia entre los filtros adyacentes LP. [ITU08a, sección 7.1.2.4.2].

Última trama recibida buena	N. ° de tramas borradas sucesivas	α
INICIO ARTIFICIAL		0,6
INICIO, CON VOZ	≤ 3 > 3	1,0 0,4
TRANSICIÓN CON VOZ		0,4
TRANSICIÓN SIN VOZ		0,8
SIN VOZ	= 1 = 2 > 2	$0,2 \theta + 0,8$ 0,6 0,4

45 **[007]** Además, G.718 proporciona un procedimiento de desvanecimiento a fin de modificar la envoltura espectral. La idea general es hacer converger los últimos parámetros de Frecuencia Espectral de Impedancia y Admitancia (ISF) hacia un vector de la media de ISF adaptativa. En primer lugar, se calcula un vector de frecuencia espectral de impedancia y admitancia (ISF) promedio a partir de los últimos 3 vectores de ISF conocidos. A continuación, se promedia de nuevo el vector de ISF promedio con un vector de ISF a largo plazo capacitado desconectado (que es un vector constante) [ITU08a, sección 7.11.1.2].

[008] Además, G.718 proporciona un procedimiento de desvanecimiento para el control del comportamiento

a largo plazo, y, en consecuencia, la interacción con el ruido de fondo, donde la energía de excitación de altura (y, en consecuencia, la periodicidad de excitación) converge en 0, mientras que la energía de excitación aleatoria converge en la energía de excitación de generación de ruido de confort (CNG) [ITU08a, sección 7.11.1.6]. La atenuación de ganancia de innovación se calcula como:

5

$$g_s^{[1]} = \alpha g_s^{[0]} + (1 - \alpha)g_n \quad (1)$$

donde $g_s^{[1]}$ es la ganancia innovadora al comienzo de la siguiente trama, $g_s^{[0]}$ es la ganancia innovadora al comienzo de la trama actual, g_n es la ganancia de la excitación usada durante la generación de ruido de confort y el factor de atenuación α .

10

[009] De manera similar a la atenuación de excitación periódica, la ganancia es atenuada de forma lineal a lo largo de la trama sobre una base de muestra por muestra, iniciando con $g_s^{[0]}$, y alcanza $g_s^{[1]}$ al comienzo de la siguiente trama.

15

[0010] La figura 2 reseña la estructura del decodificador de G.718. En particular, la figura 2 ilustra una estructura de decodificador de G.718 de alto nivel para el ocultamiento de pérdida de paquetes (PLC), con rasgo de un filtro de paso alto.

20 **[0011]**

Por medio de la estrategia anteriormente descrita de G.718, la ganancia innovadora g_s converge con la ganancia utilizada durante la generación de ruido de confort g_n para largas ráfagas de pérdidas de paquetes.

Como se describe en [ITU08a, sección 6.12.3], la ganancia de ruido de confort g_n se proporciona como la raíz cuadrada de la energía \tilde{E} . Las condiciones de la actualización de \tilde{E} no se describen en detalle. Después de la implementación de referencia (punto de flotación código C, estat_ruido_uv_mod.c), \tilde{E} se deriva de la siguiente manera:

25

```

    si (sin voz_vad == 0) {
        si (sinv_cnt > 20) {
            ftmp = lp_gananciadc * lp_gananciadc;
            lp_ener = 0.7f * lp_ener + 0.3f * ftmp;
        }
        de otro modo {
            sinv_cnt++;
        }
    }
    de otro modo {
        sinv_cnt = 0;
    }

```

35

donde sin voz_vad mantiene la detección de actividad de voz, donde sinv_cnt mantiene el número de tramas sin voz en una fila, donde la ganancia mantiene las ganancias de paso bajo del libro de código establecido, y donde lp_ener mantiene la estimación de energía de generación de ruido de confort (CNG) de paso bajo \tilde{E} , que se inicializa con 0.

[0012] Asimismo, G.718 proporciona un filtro de paso alto, introducido en la vía de señal de la excitación sin voz, si la señal de la última trama buena fue clasificada de forma diferente de SIN VOZ, véase la figura 2; véase, además, [ITU08a, sección 7.11.1.6]. Este filtro tiene una característica de estante bajo, con una respuesta de frecuencia en DC de alrededor de 5 dB menos que en la frecuencia Nyquist.

[0013] Además, G.718 propone un circuito de retroalimentación de LTP desacoplado (LTP = Predicción a largo plazo). Aunque durante la operación normal el circuito de retroalimentación para el libro de código adaptativo es actualizado de manera subtrama ([ITU08a, sección 7.1.2.1.4]) sobre la base de la excitación completa, durante el ocultamiento, este circuito de retroalimentación es actualizado de manera de trama (véase [ITU08a, secciones 7.11.1.4, 7.11.2.4, 7.11.1.6, 7.11.2.6; dec_GV_exc@dec_gen_voic.c y syn_bfi_post@syn_bfi_pre_post.c]) sobre la base solamente de la excitación con voz. Con esta estrategia, el libro de código adaptativo no es "contaminado" con ruido que tiene su origen en la excitación de innovación aleatoriamente seleccionada.

55

[0014] Con respecto a las capas de mejoramiento codificado de transformada (3–5) de G.718, durante el

ocultamiento, el decodificador se comporta con respecto a la decodificación de capa alta de manera similar a la operación normal, solo que el espectro de MDCT (transformada de coseno discreta modificada, conforme a sus siglas en inglés) se establece en cero. No se aplica comportamiento de desvanecimiento especial durante el ocultamiento.

5 **[0015]** Con respecto a la generación de ruido de confort (CNG), en G.718, la síntesis de generación de ruido de confort (CNG) se realiza en el siguiente orden. En primer lugar, se decodifican los parámetros de una trama de ruido de confort. A continuación, se sintetiza una trama de ruido de confort. A continuación, se restablece el búfer de altura. A continuación, se guarda la síntesis para la clasificación de FER (recuperación de error de trama, conforme a sus siglas en inglés). Más adelante, se conduce el desénfasis de espectro. A continuación, se conduce el post-filtrado de baja frecuencia. A continuación, las variables de generación de ruido de confort (CNG) son actualizadas.

10 **[0016]** En el caso de ocultamiento, se realiza exactamente lo mismo, excepto que los parámetros de generación de ruido de confort (CNG) no son decodificados a partir de la corriente de bits. Esto significa que los parámetros no son actualizados durante la pérdida de trama, pero se usan los parámetros decodificados a partir del último SID bueno (descriptor de inserción de silencio, conforme a sus siglas en inglés).

20 **[0017]** Ahora, se considera G.719. G.719, que se basa en Siren 22, es un códec de audio de banda completa sobre la base de la transformada. La ITU-T recomienda para G.719 un desvanecimiento con repetición de trama en el dominio espectral [ITU08b, sección 8.6]. Según G.719, se incorpora un mecanismo de ocultamiento de borrado de trama en el decodificador. Cuando una trama es correctamente recibida, los coeficientes de transformada reconstruidos son almacenados en un búfer. Si el decodificador es informado de que se ha perdido una trama o que una trama está alterada, los coeficientes de transformada reconstruidos en la trama recibida más recientemente son sometidos a escala decreciente con un factor 0,5 y, a continuación, son usados como los coeficientes de transformada reconstruidos para la trama actual. El decodificador procede mediante su transformación al dominio de tiempo, y la realización de la operación de formación de ventana-superposición-adición.

30 **[0018]** En lo que sigue, se describe G.722. G.722 es un sistema de codificación de 50 a 7000 Hz que utiliza la modulación por impulsos codificados de diferencial adaptativo de subbanda (SB-ADPCM, conforme a sus siglas en inglés) dentro de una tasa de bits de hasta 64 kbit/s. La señal se divide en una subbanda superior y una inferior, usando un análisis de QMF (QMF = sigla en inglés de filtro espejo en cuadratura). Las dos bandas resultantes son codificadas por ADPCM (ADPCM = sigla en inglés de "modulación por impulsos codificados de diferencial adaptativo").

35 **[0019]** Para G.722, se especifica un algoritmo de alta complejidad para el ocultamiento de pérdida de paquetes, en el Apéndice III [ITU06a], y se especifica un algoritmo de baja complejidad para el ocultamiento de pérdida de paquetes, en el Apéndice IV [ITU07]. G.722 – Apéndice III ([ITU06a, sección III.5]) propone un silenciamiento efectuado de forma gradual, iniciando después de 20ms de pérdida de trama, y completado después de 60 ms de pérdida de trama. Además, G.722 – Apéndice IV propone una técnica de desvanecimiento que aplica "a cada muestra, un factor de ganancia que es computado y adaptado muestra por muestra" [ITU07, sección IV.6.1.2.7].

40 **[0020]** En G.722, el proceso de silenciamiento tiene lugar en el dominio de subbanda justo antes de la síntesis de filtro espejo en cuadratura (QMF) y como la última etapa del módulo de ocultamiento de pérdida de paquetes (PLC). El cálculo del factor de silenciamiento se realiza usando información de clase del clasificador de señales que además es parte del módulo de ocultamiento de pérdida de paquetes (PLC). La distinción se hace entre clases TRANSIENTE, UV_TRANSICIÓN y otras. Además, se hace la distinción entre pérdidas únicas de tramas de 10 ms y otros casos (múltiples pérdidas de tramas de 10 ms y pérdidas únicas o múltiples de tramas de 20 ms).

[0021] Esto se ilustra por medio de la figura 3. En particular, la figura 3 representa un escenario donde el factor de desvanecimiento de G.722 depende de la información de clase y donde 80 muestras son equivalentes a 10 ms.

50 **[0022]** Según G.722, el módulo de ocultamiento de pérdida de paquetes (PLC) crea la señal para la trama perdida, y alguna señal adicional (10 ms) que se supone es mezclada gradualmente con la siguiente trama buena. El silenciamiento para esta señal adicional sigue las mismas reglas. En el ocultamiento de banda alta de G.722, la mezcla gradual no tiene lugar.

55 **[0023]** En lo que sigue, se considera G.722.1. G.722.1, que se basa en Siren 7, es un códec de audio de banda ancha sobre la base de la transformada con un modo de extensión de banda súper ancha, referido como G.722.1C. G.722.1C en sí mismo se basa en Siren 14. La ITU-T recomienda para G.722.1 una repetición de tramas, con subsiguiente silenciamiento [ITU05, sección 4.7]. Si el decodificador es informado, por medio de un mecanismo de señalización externo no definido en esta recomendación, que una trama se ha perdido o ha sido alterada, repite los coeficientes de MLT decodificados de la trama previa (MLT: transformada superpuesta modulada, conforme a sus siglas en inglés). Procede mediante su transformación al dominio de tiempo y realizando la operación de superposición y adición con la información decodificada de la trama previa y de la siguiente. Si la trama previa también fue pérdida o alterada, entonces el decodificador establece todos los coeficientes de MLT de las tramas actuales en cero.

65 **[0024]** Ahora se considera G.729. G.729 es un algoritmo de compresión de datos de audio para voz, que

comprime voz digital en paquetes de 10 milisegundos de duración. Oficialmente, se describe como Codificación de la voz a 8 kbit/s usando la codificación de la voz de predicción lineal excitada por código (CS-ACELP, conforme a sus siglas en inglés) [ITU12].

- 5 **[0025]** Como se reseña en [CPK08], G.729 recomienda un desvanecimiento en el dominio de predicción lineal (LP). El algoritmo de ocultamiento de pérdida de paquetes (PLC) empleado en el estándar G.729 reconstruye la señal de la voz para la trama actual sobre la base de la información de la voz previamente recibida. En otras palabras, el algoritmo de ocultamiento de pérdida de paquetes (PLC) reemplaza la excitación perdida con una característica equivalente de una trama previamente recibida, aunque la energía de excitación decae de forma gradual, finalmente, las ganancias de los libros de códigos adaptativos y establecidos son atenuadas por un factor constante.

[0026] La ganancia de libro de código establecido atenuada se proporciona mediante:

$$g_c^{(m)} = 0.98 \cdot g_c^{(m-1)}$$

15 donde m es el índice de subtrama.

[0027] La ganancia de libro de código adaptativa se basa en una versión atenuada de la ganancia de libro de código adaptativa previa:

20
$$g_p^{(m)} = 0.9 \cdot g_p^{(m-1)}, \text{ unido por } g_p^{(m)} < 0.9$$

[0028] Nam en Park y col. sugieren para G.729 un control de amplitud de señal usando la predicción por medio de la regresión lineal [CPK08, PKJ+11]. Se dirige a la pérdida de paquetes de ráfaga, y utiliza la regresión lineal como una técnica de núcleo. La regresión lineal se basa en el modo lineal como:

$$g'_i = a + bi \quad (2)$$

donde g'_i es la amplitud actual recién predicha, a y b son coeficientes para la función lineal de primer orden, e i es el índice de la trama. A fin de hallar los coeficientes optimizados a^* y b^* , la operación de suma del error de predicción cuadrado se minimiza:

$$\epsilon = \sum_{j=i-4}^{i-1} (g_j - g'_j)^2 \quad (3)$$

35 ϵ es el error cuadrado, g_j es la amplitud pasada original n.º j . A fin de minimizar este error, simplemente el derivado respecto de a y b se establece en cero. Usando los parámetros optimizados a^* y b^* , una estimación de cada g_i^* es denotada por:

$$g_i^* = a^* + b^*i \quad (4)$$

40

[0029] La figura 4 muestra la predicción de amplitud, en particular, la predicción de la amplitud g_i^* , usando la regresión lineal.

[0030] A fin de obtener la amplitud A'_i del paquete perdido i , una relación σ_i

45

$$\sigma_i = \frac{g_i^*}{g_{i-1}} \quad (5)$$

se multiplica con un factor de escala S_i :

$$A'_i = S_i * \sigma_i \quad (6)$$

50

en el que el factor de escala S_i depende del número de tramas ocultadas consecutivas $l(i)$:

$$S_i = \begin{cases} 1.0, & \text{si } l(i) = 1, 2 \\ 0.9, & \text{si } l(i) = 3, 4 \\ 0.8, & \text{si } l(i) = 5, 6 \\ 0, & \text{de otro modo} \end{cases} \quad (7)$$

5 En [PKJ+11], se propone una escala levemente diferente.

[0031] Según G.729, a continuación, A'_i será suavizado a fin de prevenir la atenuación discreta en los bordes de la trama. La amplitud suavizada final $A_i(n)$ se multiplica a la excitación, obtenida de los componentes previos de
10 ocultamiento de pérdida de paquetes (PLC).

[0032] En lo que sigue, se considera G.729.1. G.729.1 es un codificador de tasa de bits variable integrado sobre la base de G.729: una corriente de bits de codificador de banda ancha escalable de 8–32 kbit/s interoperable con G.729 [ITU06b].

15 **[0033]** Según G.729.1, como en G.718 (véase más arriba), se propone un desvanecimiento adaptativo, que depende de la estabilidad de las características de la señal ([ITU06b, sección 7.6.1]). Durante el ocultamiento, la señal es habitualmente atenuada sobre la base de un factor de atenuación α que depende de los parámetros de la clase de la última trama recibida buena y del número de tramas borradas consecutivas. El factor de atenuación α depende
20 adicionalmente de la estabilidad del filtro de predicción lineal (LP) para tramas SIN VOZ. En general, la atenuación es lenta si la última trama recibida buena se encuentra en un segmento estable, y es rápida si la trama se encuentra en un segmento de transición.

[0034] Adicionalmente, el factor de atenuación α depende de la ganancia de altura promedio por subtrama \bar{g}_p
25 ([ITU06b, eq. 163, 164]):

$$\bar{g}_p = 0.1g_p^{(0)} + 0.2g_p^{(1)} + 0.3g_p^{(2)} + 0.4g_p^{(3)} \quad (8)$$

donde $g_p^{(i)}$ es la ganancia de altura en subtrama i .

30 **[0035]** La Tabla 2 muestra el esquema de cálculo de α , donde:

$$\beta = \sqrt{\bar{g}_p} \quad \text{con} \quad 0.85 \geq \beta \geq 0.98 \quad (9)$$

35 **[0036]** Durante el proceso de ocultamiento, se usa α en las siguientes herramientas de ocultamiento:

Tabla 2: Valores del factor de atenuación α , el valor θ es un factor de estabilidad computado a partir de una medida de distancia entre los filtros adyacentes de LP. [ITU06b, sección 7.6.1].

Última trama recibida buena	N.º de tramas borradas sucesivas	α
CON VOZ	1	β
	2,3	\bar{g}_p
	> 3	0,4
INICIO	1	0,8 β
	2,3	\bar{g}_p
	> 3	0,4
INICIO ARTIFICIAL	1	0,6 β
	2,3	\bar{g}_p
	> 3	0,4

(continuación)

Última trama recibida buena	N.º de tramas borradas sucesivas	α
TRANSICIÓN CON VOZ	≤ 2	0,8
	> 2	0,2
TRANSICIÓN SIN VOZ		0,88
SIN VOZ	1	0,95
	2,3	$0,6 \theta + 0,4$
	> 3	0,4

[0037] Según G.729.1, con respecto a la resincronización de pulso glotal, como el último pulso de la excitación de la trama previa se usa para la construcción de la parte periódica, su ganancia es aproximadamente correcta al comienzo de la trama ocultada, y puede establecerse en 1. La ganancia es atenuada a continuación linealmente a lo largo de toda la trama sobre una base de muestra por muestra, a fin de lograr el valor de α al final de la trama. La evolución de energía de los segmentos con voz es extrapolada usando los valores de ganancia de excitación de altura de cada subtrama de la última trama buena. En general, si estas ganancias son mayores que 1, la energía de señal es creciente; si son menores que 1, la energía es decreciente. Por lo tanto, α se establece en $\beta = \sqrt{g_p}$ como se describe con anterioridad; véase [ITU06b, eq. 163, 164]. El valor de β es clipeado entre 0,98 y 0,85 a fin de evitar fuertes incrementos y disminuciones de energía; véase [ITU06b, sección 7.6.4].

[0038] Con respecto a la construcción de la parte aleatoria de la excitación, según G.729.1, al comienzo de un bloque borrado, la ganancia de innovación g_s se inicializa mediante el uso de las ganancias de excitación de innovación de cada subtrama de la última trama buena:

$$g_s = 0.1g^{(0)} + 0.2g^{(1)} + 0.3g^{(2)} + 0.4g^{(3)}$$

donde $g^{(0)}$, $g^{(1)}$, $g^{(2)}$ y $g^{(3)}$ son el libro de código establecido, o innovación, ganancias de las cuatro subtramas de la última trama correctamente recibida. La atenuación de ganancia de innovación se realiza como:

$$g_s^{(1)} = \alpha \cdot g_s^{(0)}$$

en la que $g_s^{(1)}$ es la ganancia de innovación al comienzo de la siguiente trama, $g_s^{(0)}$ es la ganancia de innovación al comienzo de la trama actual, y α es como se define en la Tabla 2 anterior. De manera similar a la atenuación de excitación periódica, la ganancia es de ese modo linealmente atenuada a lo largo de toda la trama, sobre una base de muestra por muestra, iniciando con $g_s^{(0)}$ y llegando al valor de $g_s^{(1)}$ que se lograría al comienzo de la siguiente trama.

[0039] Según G.729.1, si la última buena trama es SIN VOZ, solo se usa la excitación de innovación, y es adicionalmente atenuada por un factor de 0,8. En este caso, el búfer de excitación pasado es actualizado con la excitación de innovación, ya que no hay parte periódica de la excitación disponible; véase [ITU06b, sección 7.6.6].

[0040] En lo que sigue, se considera la multitasa adaptativa (AMR). La multitasa adaptativa 3GPP (AMR) [3GP12b] es un códec de voz que utiliza el algoritmo de ACELP. La multitasa adaptativa (AMR) es capaz de codificar la voz con una tasa de muestreo de 8.000 muestras/s y una tasa de bits entre 4,75 y 12,2 kbit/s, y apoya la señalización de tramas de descriptores de silencio (DTX/CNG).

[0041] En la multitasa adaptativa (AMR), durante el ocultamiento de errores (véase [3GP12a]), se distingue entre tramas que son propensas al error (errores de bits) y tramas que están perdidas por completo (sin ningún dato).

[0042] Para el ocultamiento de predicción lineal excitada por código algebraico (ACELP), la multitasa adaptativa (AMR) introduce una máquina de estado que estima la calidad del canal. Cuanto mayor es el valor del contador de estado, peor es la calidad del canal. El sistema inicia en el estado 0. Cada vez que se detecta una mala trama, el contador de estado se incrementa en uno, y se satura cuando llega a 6. Cada vez que se detecta una trama buena de la voz, el contador de estado se restablece a cero, excepto cuando el estado es 6, donde el contador de estado se establece en 5. El flujo de control de la máquina de estado puede ser descrito por el siguiente código C (BFI es un indicador de mala trama (conforme a sus siglas en inglés), Estado es una variable de estado):

```

50 si (BFI != 0) {
    Estado = Estado + 1;
}
de otro modo, si (Estado == 6) {

```

```

    Estado = 5;
}
de otro modo {
    Estado = 0;
5 }
si (Estado > 6) {
    Estado = 6;
}

```

10 **[0043]** Además de esta máquina de estado, en la multitasa adaptativa (AMR), las banderas de mala trama de las tramas actuales y previas son verificadas (prevBFI).

[0044] Son posibles tres combinaciones diferentes:

15 La primera de las tres combinaciones es BFI = 0, prevBFI = 0, Estado = 0: No se detecta error en la trama recibida o en la trama de la voz recibida previa. Los parámetros de voz recibidos se usan de la manera normal en la síntesis de voz. Se guarda la trama actual de parámetros de voz.
 La segunda de las tres combinaciones es BFI = 0, prevBFI = 1, Estado = 0 o 5: No se detecta error en la trama de voz recibida, pero la trama de voz recibida previa fue mala. La ganancia de predicción a largo plazo (LTP) y la ganancia de libro de código establecido están limitadas por debajo de los valores utilizados para la última subtrama buena recibida:

$$g_p = \begin{cases} g_p, & g_p \leq g_p(-1) \\ g_p(-1), & g_p > g_p(-1) \end{cases} \quad (10)$$

25 donde g_p = ganancia de predicción a largo plazo (LTP) decodificada actual, $g_p(-1)$ = ganancia de predicción a largo plazo (LTP) utilizada para la última subtrama buena (BFI = 0), y

$$g_c = \begin{cases} g_c, & g_c \leq g_c(-1) \\ g_c(-1), & g_c > g_c(-1) \end{cases} \quad (11)$$

30 donde g_c = ganancia de libro de código establecido decodificada actual, y $g_c(-1)$ = ganancia de libro de código establecido utilizada para la última subtrama buena (BFI = 0).

35 El resto de los parámetros de voz recibidos se usan normalmente en la síntesis de voz. Se guarda la trama actual de los parámetros de voz.

La tercera de las tres combinaciones es BFI = 1, prevBFI = 0 o 1, Estado = 1...6: se detecta un error en la trama de la voz recibida, y se inicia el procedimiento de sustitución y silenciamiento. La ganancia de predicción a largo plazo (LTP) y la ganancia de libro de código establecido son reemplazadas por valores atenuados de las subtramas previas:

$$g_p = \begin{cases} P(\text{estad.}) \cdot g_p(-1), & g_p(-1) \leq \text{median5}(g_p(-1), \dots, g_p(-5)) \\ P(\text{estad.}) \cdot \text{median5}(g_p(-1), \dots, g_p(-5)) & g_p(-1) > \text{median5}(g_p(-1), \dots, g_p(-5)) \end{cases} \quad (12)$$

45 donde g_p indica la ganancia decodificada actual de predicción a largo plazo (LTP) y $g_p(-1), \dots, g_p(-n)$ indican las ganancias de predicción a largo plazo (LTP) utilizadas para la últimas subtramas n , y $\text{median5}()$ indica una operación de mediana de 5 puntos, y

$P(\text{estado})$ = factor de atenuación, donde ($P(1) = 0,98$, $P(2) = 0,98$, $P(3) = 0,8$, $P(4) = 0,3$, $P(5) = 0,2$, $P(6) = 0,2$) y estado = número de estado, y

$$g_c = \begin{cases} C(\text{estad.}) \cdot g_c(-1), & g_c(-1) \leq \text{median5}(g_c(-1), \dots, g_c(-5)) \\ C(\text{estad.}) \cdot \text{median5}(g_c(-1), \dots, g_c(-5)) & g_c(-1) > \text{median5}(g_c(-1), \dots, g_c(-5)) \end{cases} \quad (13)$$

donde g_c indica la ganancia de libro de código establecido decodificada actual, y $g_c(-1), \dots, g_c(-n)$ indican las ganancias de libro de código establecido utilizadas para las últimas subtramas n , y $\text{median5}()$ indica una operación de mediana de 5 puntos, y $C(\text{estad}) = \text{factor de atenuación}$, donde $C(1) = 0,98, C(2) = 0,98, C(3) = 0,98, C(4) = 0,98, C(5) = 0,98, C(6) = 0,7$ y estado = número de estado.

[0045] En la multitasa adaptativa (AMR), los valores de retraso de predicción a largo plazo (LTP) (LTP = Predicción a largo plazo) son reemplazados por el valor pasado de la subtrama $n - 4$ de la trama previa (modo 12.2), o valores levemente modificados sobre la base del último valor recibido correctamente (todos los demás modos).

[0046] Según la multitasa adaptativa (AMR), los pulsos de innovación de libro de código establecido recibidos de la trama errónea se usan en el estado en el cual fueron recibidos, cuando se reciben los datos alterados. En el caso de que no se reciban datos, deben emplearse los índices de libro de código establecido aleatorios.

[0047] Con respecto a la generación de ruido de confort (CNG) en multitasa adaptativa (AMR), según [3GP12a, sección 6.4], cada primera trama de descriptor de inserción de silencio (SID) perdida es sustituida usando la información de SID de tramas de SID válidas recibidas anteriormente, y se aplica el procedimiento para tramas de SID válidas. Para subsiguientes tramas de SID perdidas, se aplica una técnica de atenuación al ruido de confort, que disminuirá de forma gradual el nivel de salida. Por lo tanto, se verifica si la última actualización de descriptor de inserción de silencio (SID) fue hace más de 50 tramas ($=1$ s); en ese caso, la salida será silenciada (atenuación de nivel por $-6/8$ dB por trama [3GP12d, $\text{dtx_dec}\{\}\text{@sp_dec.c}$], lo que produce 37,5 dB por segundo). Obsérvese que el desvanecimiento aplicado a la generación de ruido de confort (CNG) se realiza en el dominio de predicción lineal (LP).

[0048] En lo que sigue, se considera la multitasa adaptativa (AMR)–WB (banda ancha). La multitasa adaptativa – banda ancha [ITU03, 3GP09c] es un códec de voz, de predicción lineal excitada por código algebraico (ACELP), sobre la base de multitasa adaptativa (AMR) (véase la sección 1.8). Utiliza extensión de ancho de banda paramétrica y, además, soporta sistemas de teatro digital/generación de ruido de confort (DTX/CNG). En la descripción del estándar [3GP12g], hay soluciones de ejemplos de ocultamiento proporcionadas, que son las mismas que para la multitasa adaptativa (AMR) [3GP12a], con desviaciones menores. Por lo tanto, solo las diferencias con respecto a la multitasa adaptativa (AMR) se describen aquí. Para la descripción del estándar, véase la descripción anterior.

[0049] Con respecto a la predicción lineal excitada por código algebraico (ACELP), en la multitasa adaptativa (AMR)–banda ancha (WB), se realiza el desvanecimiento de predicción lineal excitada por código algebraico (ACELP) sobre la base del código de fuente de referencia [3GP12c], mediante la modificación de la ganancia de altura g_p (para multitasa adaptativa (AMR) arriba referida como ganancia de predicción a largo plazo (LTP)) y mediante la modificación de la ganancia de código g_c .

[0050] En el caso de trama perdida, la ganancia de altura g_p para la primera subtrama es la misma que en la última trama buena, excepto que está limitada entre 0,95 y 0,5. Para la segunda, la tercera y las siguientes subtramas, la ganancia de altura g_p disminuye por un factor de 0,95 y es limitada de nuevo.

[0051] La multitasa adaptativa (AMR)–banda ancha (WB) propone que, en una trama ocultada, g_c se basa en la última g_c :

$$g_{c, \text{actual}} = g_{c, \text{pasada}} * (1.4 - g_{p, \text{pasada}}) \quad (14)$$

$$g_c = g_{c, \text{actual}} * g_{c, \text{inov}} \quad (15)$$

$$g_{c, \text{inov}} = \frac{1.0}{\sqrt{\frac{\text{ener}_{\text{inov}}}{\text{subtrama_tamaño}}}} \quad (16)$$

$$ener_{inov} = \sum_{i=0}^{subtrama_tamaño-1} código[i] \quad (17)$$

[0052] Para el ocultamiento de los retrasos de la predicción a largo plazo (LTP), en la multitasa adaptativa (AMR)–banda ancha (WB), el historial de los cinco últimos retrasos de predicción a largo plazo (LTP) buenos y ganancias de predicción a largo plazo (LTP) se usan para hallar el mejor procedimiento para la actualización, en el caso de una pérdida de trama. En el caso de que la trama sea recibida con errores de bits, se realiza una predicción, tanto si el retraso recibido de predicción a largo plazo (LTP) se puede utilizar o no [3GP12g].

[0053] Con respecto a la generación de ruido de confort (CNG), en multitasa adaptativa (AMR)–banda ancha (WB), si la última trama recibida correctamente fue una trama de descriptor de inserción de silencio (SID), y una trama se clasifica como perdida, será sustituida por la última información de trama de SID válida, y debe aplicarse el procedimiento para tramas de SID válidas.

[0054] Para subsiguientes tramas de SID perdidas, la multitasa adaptativa (AMR)–banda ancha (WB) propone la aplicación de una técnica de atenuación al ruido de confort que disminuirá de forma gradual el nivel de salida. Por lo tanto, se verifica si la última actualización de descriptor de inserción de silencio (SID) fue hace más de 50 tramas (=1 s); en ese caso, la salida será silenciada (atenuación de nivel por $-3/8$ dB por trama [3GP12f, dtx_dec{}@dtx.c], lo que produce 18,75 dB por segundo). Cabe observar que el desvanecimiento aplicado a la generación de ruido de confort (CNG) se realiza en el dominio de predicción lineal (LP).

[0055] Ahora se considera la multitasa adaptativa (AMR)–banda ancha (WB)+. La multitasa adaptativa – banda ancha (WB)+ [3GP09a] es un códec cambiado mediante el uso de la predicción lineal excitada por código algebraico (ACELP) y la TCX (TCX = excitación codificada por la transformada), como códec núcleo. Utiliza la extensión de ancho de banda paramétrica y además soporta Sistemas de teatro digital (DTX)/generación de ruido de confort (CNG).

[0056] En la multitasa adaptativa (AMR)–banda ancha (WB)+, se aplica una lógica de extrapolación de modos para extrapolar los modos de las tramas perdidas dentro de una supertrama distorsionada. Esta extrapolación de modos se sustenta en el hecho de que existe redundancia en la definición de indicadores de modos. La lógica de decisión (proporcionada en [3GP09a, figure 18]) propuesta por la multitasa adaptativa (AMR)–banda ancha (WB)+ es la siguiente:

- Se define un modo de vector ($m_{-1}, m_0, m_1, m_2, m_3$), donde m_{-1} indica el modo de la última trama de la supertrama previa, y m_0, m_1, m_2, m_3 indican los modos de las tramas en la supertrama actual (decodificadas a partir de la corriente de bits), donde $m_k = -1, 0, 1, 2$ o 3 (-1 : perdida, 0 : predicción lineal excitada por código algebraico (ACELP), 1 : TCX20, 2 : TCX40, 3 : TCX80), y donde el número de tramas perdidas *npérdida* puede ser entre 0 y 4.

- Si $m_{-1} = 3$ y dos de los indicadores de modos de las tramas 0 – 3 son iguales a tres, todos los indicadores se establecerán en tres, ya que entonces, es seguro que una trama TCX80 fue indicada dentro de la supertrama.

- Si solo un indicador de las tramas 0 – 3 es tres (y el número de tramas perdidas *npérdida* es tres), el modo se establecerá en (1, 1, 1, 1), ya que, entonces, 3/4 del espectro objetivo TCX80 está perdido, y es muy probable que la ganancia global de excitación codificada por la transformada (TCX) esté perdida.

- Si el modo está indicando (x, 2, -1, x, x) o (x, -1, 2, x, x), será extrapolado a (x, 2, 2, x, x), de manera que indique una trama TCX40. Si el modo indica (x, x, x, 2, -1) o (x, x, -1, 2), será extrapolado a (x, x, x, 2, 2), de modo que indique además una trama TCX40. Debe observarse además que (x, [0, 1], 2, 2, [0, 1]) son configuraciones inválidas.

- Después de esto, para cada trama perdida (*modo* = -1), se establece el modo en predicción lineal excitada por código algebraico (ACELP) (*modo* = 0) si la trama precedente fue predicción lineal excitada por código algebraico (ACELP), y el modo se establece en TCX20 (*modo* = 1), para todos los demás casos.

[0057] Con respecto a la predicción lineal excitada por código algebraico (ACELP), según la multitasa adaptativa (AMR)–banda ancha (WB)+, si un modo de tramas perdidas logra $m_k = 0$ después de la extrapolación de modos, se aplica la misma estrategia que en [3GP12g] para esta trama (véase más arriba).

[0058] En multitasa adaptativa (AMR)–banda ancha (WB)+, según el número de tramas perdidas y el modo extrapolado, se distinguen las siguientes estrategias de ocultamiento relacionadas con TCX (TCX = excitación codificada por la transformada):

- Si una trama completa está perdida, entonces se aplica un ocultamiento de tipo predicción lineal excitada por código algebraico (ACELP): se repite la última excitación, y se usan los coeficientes de frecuencia espectral de impedancia y

admitancia (ISF) ocultados (levemente cambiados hacia sus respectivos medios adaptativos) para sintetizar la señal de dominio de tiempo. Además, se multiplica un factor de desvanecimiento de 0,7 por trama (20 ms) [3GP09b, dec_tcx.c] en el dominio predictivo lineal, justo antes de la síntesis de LPC (Codificación Predictiva lineal).

- 5 – Si el último modo fue TCX80, al igual que si el modo extrapolado de la supertrama (parcialmente perdida) es TCX80 (npérdida = [1, 2], modo = (3, 3, 3, 3, 3)), se realiza el ocultamiento en el dominio FFT, utilizando extrapolación de fases y amplitud, teniendo en cuenta la última trama recibida correctamente. La estrategia de extrapolación de la información de fase no es de interés aquí (sin relación con la estrategia de desvanecimiento), y por lo tanto, no se describe. Para más detalles, véase [3GP09a, sección 6.5.1.2.4]. Con respecto a la modificación de amplitud de multitasa adaptativa (AMR)–WB+, la estrategia realizada para el ocultamiento de excitación codificada por la transformada (TCX) consiste en las siguientes etapas [3GP09a, sección 6.5.1.2.3]:

– Se computa el espectro de magnitud de trama previa:

$$15 \quad \text{antiguo } A[k] = \left| \text{antiguo } \hat{X}[k] \right|$$

– Se computa el espectro de magnitud de trama actual:

$$20 \quad A[k] = \left| \hat{X}[k] \right|$$

– Se computa la diferencia de ganancia de energía de coeficientes espectrales no perdidos entre la trama previa y la trama actual:

$$25 \quad \text{ganancia} = \sqrt{\frac{\sum A[k]^2}{\sum \text{antiguo } A[k]^2}}$$

– La amplitud de los coeficientes espectrales perdidos es extrapolada mediante el uso de:

$$\text{si } (k \text{ perdido}) \quad A[k] = \text{ganancia} \cdot \text{antiguo } A[k]$$

- 30 – En cualquier otro caso de una trama perdida con $m_k = [2, 3]$, el objetivo de la excitación codificada por la transformada (TCX) (FFT inversa de espectro decodificado más relleno de ruido (usando un nivel de ruido decodificado a partir de la corriente de bits)) se sintetiza usando toda la información disponible (que incluye ganancia global de excitación codificada por la transformada (TCX)). No se aplica desvanecimiento en este caso.

- 35 **[0059]** Con respecto a la generación de ruido de confort (CNG) en multitasa adaptativa (AMR)–banda ancha (WB)+, se usa la misma estrategia que en la multitasa adaptativa (AMR)–banda ancha (WB) (véase más arriba).

- [0060]** En lo que sigue, se considera OPUS. OPUS [IET12] incorpora tecnología de dos códecs: el SILK orientado a la voz (conocido como el códec Skype) y el CELT de baja latencia (CELT = Transformada Superpuesta de Energía Restrignida). Opus puede ajustarse de una manera sin problemas, entre altas y bajas tasas de bits, e internamente, cambia entre un códec de predicción lineal a menores tasas de bits (SILK) y un códec de transformada a tasas de bits más altas (CELT), al igual que un híbrido para una corta superposición.

- [0061]** Con respecto a la compresión y descompresión de datos de audio SILK, en OPUS, hay varios parámetros que son atenuados durante el ocultamiento en la rutina del decodificador SILK. El parámetro de ganancia de predicción a largo plazo (LTP) es atenuado mediante la multiplicación de todos los coeficientes de codificación predictiva lineal (LPC) o bien con 0,99, 0,95 o 0,90 por trama, dependiendo del número de tramas perdidas consecutivas, donde la excitación se construye usando el último ciclo de altura de la excitación de la trama previa. El parámetro de retraso de altura es aumentado muy lentamente durante las consecutivas pérdidas. Para pérdidas únicas, se mantiene constante, en comparación con la última trama. Además, el parámetro de ganancia de excitación es exponencialmente atenuado con $0,99^{\text{pérdidocont}}$ por trama, de manera que el parámetro de ganancia de excitación es 0,99 para el primer parámetro de ganancia de excitación, de modo que el parámetro de ganancia de excitación es 0,992 para el segundo parámetro de ganancia de excitación, etc. La excitación se genera mediante el uso de un generador de número aleatorio que genera ruido blanco por el sobreflujo variable. Además, los coeficientes de codificación predictiva lineal (LPC) son extrapolados/promediados sobre la base del último grupo de coeficientes recibidos correctamente. Después de la generación del vector de excitación atenuado, los coeficientes de codificación

predictiva lineal (LPC) ocultos se usan en OPUS para la síntesis de la señal de salida de dominio de tiempo.

[0062] Ahora, en el contexto de OPUS, se considera la transformada superpuesta de energía restringida (CELT). La transformada superpuesta de energía restringida (CELT) es un códec sobre la base de la transformada. El ocultamiento de la transformada superpuesta de energía restringida (CELT) caracteriza una estrategia de ocultamiento de pérdida de paquetes (PLC) sobre la base de la altura, que se aplica para hasta cinco tramas perdidas consecutivamente. Iniciando con la trama 6, se aplica una estrategia de ocultamiento de tipo ruido, que genera ruido de fondo, donde dicha característica se supone que suena como ruido de fondo precedente.

10 **[0063]** La figura 5 ilustra el comportamiento de pérdida de ráfaga de la transformada superpuesta de energía restringida (CELT). En particular, la figura 5 representa un espectrograma (eje x: tiempo; eje y: frecuencia) de un segmento de voz oculto de transformada superpuesta de energía restringida (CELT). La caja de color gris claro indica las primeras 5 tramas consecutivamente perdidas, donde se aplica la estrategia de ocultamiento de pérdida de paquetes (PLC) sobre la base de la altura. Más allá de ello, se muestra el ocultamiento de tipo ruido. Debe observarse
15 que el cambio se realiza de manera instantánea, y no transita suavemente.

[0064] Con respecto al ocultamiento sobre la base de la altura, en OPUS, el ocultamiento sobre la base de la altura consiste en hallar la periodicidad en la señal decodificada por autocorrelación y repetición de la forma de onda en forma de ventana (en el dominio de excitación mediante el uso del análisis de codificación predictiva lineal (LPC) y la síntesis) usando la desviación de altura (retraso de altura). La forma de onda en forma de ventana es superpuesta de tal manera que preserve la cancelación del *aliasing* (superposición periódica sucesiva) en el dominio de tiempo con la trama previa y la siguiente trama [IET12]. Adicionalmente, se deriva un factor de desvanecimiento y se aplica mediante el siguiente código:

```
25 opus_val32 E1=1, E2=1;
   periodo int;
   si (altura_índice <= MAX_PERIODO/2) {
   periodo = altura_índice;
   }
30 de otro modo {
   periodo = MAX_PERIODO/2;
   }
   para (i=0;i<periodo; i++)
   {
35 E1 += exc[MAX_PERIODO- periodo+i] * exc[MAX_PERIODO- periodo + i];
   E2 += exc[MAX_PERIODO-2*periodo+i] * exc[MAX_PERIODO-2*periodo+i];
   }
   si (E1 > E2) {
40 E1 = E2;
   }
   decadencia = raíz cuadrada(E1/E2);
   atenuación = decadencia;
```

[0065] En este código, exc contiene la señal de excitación hasta MAX_PERIODO muestras antes de la pérdida.

45 **[0066]** La señal de excitación se multiplica más tarde con atenuación, a continuación, se sintetiza, y sale mediante la síntesis de codificación predictiva lineal (LPC).

[0067] El algoritmo de desvanecimiento para la estrategia de dominio de tiempo puede ser resumido de la
50 siguiente manera:

- Hallar la energía sincrónica de altura del último ciclo de altura antes de la pérdida.
- Hallar la energía sincrónica de altura del segundo último ciclo de altura antes de la pérdida.
- 55 – Si la energía está en aumento, limitarla para que permanezca constante: atenuación = 1.
- Si la energía está en disminución, continuar con la misma atenuación durante el ocultamiento.

60 **[0068]** Con respecto al ocultamiento de tipo ruido, según OPUS, para la sexta y siguientes tramas perdidas consecutivas, se realiza una estrategia de sustitución de ruido en el dominio de transformada de coseno discreta modificada (MDCT) a fin de estimular el ruido de fondo de confort.

[0069] Con respecto al rastreo del nivel de ruido de fondo y la forma, en OPUS, la estimación de ruido de fondo
65 se realiza de la siguiente manera. Después del análisis de transformada de coseno discreta modificada (MDCT), se

calcula la raíz cuadrada de las energías de la banda de transformada de coseno discreta modificada (MDCT) por banda de frecuencia, donde la agrupación de los bins de la transformada de coseno discreta modificada (MDCT) sigue la escala de Bark según [IET12, Tabla 55]. A continuación, la raíz cuadrada de las energías se transforma en el dominio \log_2 por medio de:

$$5 \quad \text{bandLogE}[i] = \log_2(e) \cdot \log_e(\text{bandE}[i] \cdot eMedias[i]) \quad \text{para} \quad i = 0 \dots 21 \quad (18)$$

donde e es el número de Euler, bandE es la raíz cuadrada de la banda de la transformada de coseno discreta modificada (MDCT) y $eMedias$ es un vector de las constantes (necesarias para obtener el resultado medio cero, que produce una ganancia de codificación mejorada).

[0070] En OPUS, el ruido de fondo es representado en forma logarítmica del lado del decodificador como [IET12, amp2Log2 y log2Amp @ cant_bandas.c]:

$$15 \quad \text{fondoLogE}[i] = \min(\text{fondoLogE}[i] + 8 \cdot 0.001, \text{bandLogE}[i])$$

para $i = 0 \dots 21$ (19)

[0071] La energía mínima trazada se determina básicamente por la raíz cuadrada de la energía de la banda de la trama actual, pero el incremento de una trama a la siguiente es limitado por 0,05 dB.

[0072] Con respecto a la aplicación del nivel de ruido de fondo y la forma, según OPUS, si se aplica el ocultamiento de tipo ruido de pérdida de paquetes (PLC), se usa fondoLogE derivado en la última trama buena y se convierte de nuevo en el dominio lineal:

$$25 \quad \text{bandE}[i] = e^{(\log_e(2) \cdot \text{fondoLogE}[i] + eMedias[i])} \quad \text{para} \quad i = 0 \dots 21 \quad (20)$$

donde e es el número de Euler y $eMedias$ es el mismo vector de constantes que para la transformada de "lineal a log".

[0073] El procedimiento de ocultamiento actual es para completar la trama de la transformada de coseno discreta modificada (MDCT) con ruido blanco producido por un generador de número aleatorio, y escalar este ruido blanco de tal modo que coincida a modo de banda, con la energía de bandE . A continuación, se aplica la transformada de coseno discreta modificada (MDCT) inversa, que logra una señal de dominio de tiempo. Después de la adición de superposición y del desénfasis (como en la decodificación regular), es apagada.

[0074] En lo que sigue, se considera MPEG-4 HE-AAC (MPEG = Grupo de Expertos en Imágenes en Movimiento; HE-AAC = Codificación de Audio Avanzada de Alta Eficiencia). La codificación de audio avanzada de alta eficiencia consiste en un códec de audio sobre la base de la transformada (AAC), suplementado por una extensión de ancho de banda paramétrica (SBR).

[0075] Con respecto a AAC (AAC = codificación de audio avanzada, conforme a sus siglas en inglés), el consorcio DAB especifica para AAC en DAB+, un desvanecimiento a cero en el dominio de frecuencia [EBU10, sección A1.2] (DAB = Transmisión de Audio Digital, conforme a sus siglas en inglés). El comportamiento de desvanecimiento, por ejemplo, la rampa de atenuación, podría establecerse o ser ajustable por el usuario. Los coeficientes espectrales de la última AU (AU = unidad de acceso) son atenuados por un factor correspondiente a las características de desvanecimiento y, a continuación, pasados a la cartografía de frecuencia a dominio. Dependiendo de la rampa de atenuación, el ocultamiento cambia a silenciamiento después de un número de AU inválidas consecutivas, lo que significa que el espectro completo se establecerá en 0.

[0076] El consorcio DRM (DRM = Gerenciamiento de Derechos Digitales) especifica para AAC en DRM un desvanecimiento en el dominio de frecuencia [EBU12, sección 5.3.3]. El ocultamiento trabaja en los datos espectrales justo antes de la conversión final de frecuencia a tiempo. Si múltiples tramas son alteradas, el ocultamiento implementa primero un desvanecimiento, sobre la base de valores espectrales levemente modificados, a partir de la última trama válida. Además, de manera similar a DAB+, el comportamiento de desvanecimiento, por ejemplo, la rampa de atenuación, podría fijarse o ser ajustable por el usuario. Los coeficientes espectrales de la última trama son atenuados por un factor correspondiente a las características de desvanecimiento y, a continuación, pasados a la cartografía de frecuencia a tiempo. Dependiendo de la rampa de atenuación, el ocultamiento cambia al silenciamiento después de un número de tramas inválidas consecutivas, lo que significa que el espectro completo se establecerá en 0.

[0077] 3GPP introduce para AAC en aacPlus Mejorado el desvanecimiento en el dominio de frecuencia similar a DRM [3GP12e, sección 5.1]. El ocultamiento trabaja en los datos espectrales justo antes de la conversión final de frecuencia a tiempo. Si múltiples tramas son alteradas, el ocultamiento implementa primero un desvanecimiento, sobre

la base de valores espectrales levemente modificados, a partir de la última trama buena. Un desvanecimiento completo lleva 5 tramas. Los coeficientes espectrales de la última trama buena son copiados y atenuados por un factor de:

$$\text{Fac desvanec.} = 2^{-1 \cdot (n\text{Trama desv./2})}$$

5

con $n\text{Trama desv.}$ como el contador de tramas desde la última trama buena. Después de cinco tramas de desvanecimiento, el ocultamiento cambia a silenciamiento, lo que significa que el espectro completo se establecerá en 0.

10 **[0078]** Lauber y Sperschneider introducen para AAC un desvanecimiento a modo de trama, del espectro de la transformada de coseno discreta modificada (MDCT), sobre la base de la extrapolación de energía [LS01, sección 4.4]. Podrían usarse formas de energía de un espectro precedente, a fin de extrapolar la forma de un espectro estimado. La extrapolación de energía puede realizarse de forma independiente de las técnicas de ocultamiento como una clase de post-ocultamiento.

15

[0079] Con respecto a AAC, el cálculo de energía se realiza sobre una base de banda de factor de escala, a fin de acercarse a las bandas críticas del sistema auditivo humano. Los valores de energía individuales son disminuidos sobre una base de trama por trama a fin de reducir el volumen suavemente, por ejemplo, para desvanecer la señal. Esto se vuelve necesario, ya que la probabilidad de que los valores estimados representen la señal actual disminuye

20 rápidamente en función del tiempo.

[0080] Para la generación del espectro que va a ser desvanecido sugieren la repetición de trama o la sustitución de ruido [LS01, secciones 3.2 y 3.3].

25 **[0081]** Quackenbusch y Driesen sugieren para AAC un desvanecimiento a cero a modo de trama exponencial [QD03]. Se propone una repetición de grupo adyacente de coeficientes de tiempo/frecuencia, en la que cada repetición tiene atenuación exponencialmente en incremento, de manera que se desvanezca hasta su silenciamiento, en el caso de apagones prolongados.

30 **[0082]** Con respecto a SBR (SBR = Replicación de Banda Espectral, conforme a sus siglas en inglés), en MPEG-4 HE-AAC, 3GPP sugiere para SBR en aacPlus Mejorada el almacenamiento temporario de datos en búfer, de los datos de envoltura decodificados, y en el caso de una pérdida de trama, la reutilización de las energías almacenadas en búfer de los datos de envoltura transmitidos y la disminución de estos por una relación constante de 3 dB para cada trama ocultada. El resultado es alimentado en el proceso de decodificación normal donde el ajustador de envoltura lo utiliza para calcular las ganancias, utilizadas para el ajuste de las altas bandas con parches, creadas por el generador de HF. A continuación, tiene lugar la decodificación de SBR de manera habitual. Además, el ruido de fondo codificado delta y los valores de nivel de seno son eliminados. Como no queda disponible ninguna diferencia con la información previa, el ruido de fondo decodificado y los niveles de seno permanecen proporcionales a la energía de la señal generada de HF [3GP12e, sección 5.2].

40

[0083] El consorcio de gerenciamento de derechos digitales (DRM) especificado para la replicación de banda espectral (SBR) en conjunto con la codificación de audio avanzada (AAC) utiliza la misma técnica que 3GPP [EBU12, sección 5.6.3.1]. Además, el consorcio de transmisión de audio digital (DAB) específica para la replicación de banda espectral (SBR) en la transmisión de audio digital (DAB)+ la misma técnica que 3GPP [EBU10, sección A2].

45

[0084] En lo que sigue, se consideran MPEG-4 CELP y MPEG-4 HVXC (HVXC = Codificación de Excitación de vector armónico, conforme a sus siglas en inglés). El consorcio de gerenciamento de derechos digitales (DRM) específica para la replicación de banda espectral (SBR) en conjunto con CELP y HVXC [EBU12, sección 5.6.3.2] que el mínimo ocultamiento de requerimiento para la replicación de banda espectral (SBR) para códecs de voz debe aplicar un conjunto predeterminado de valores de datos, cada vez que una trama alterada de replicación de banda espectral (SBR) haya sido detectada. Dichos valores producen una envoltura espectral de alta banda estática, a un bajo nivel relativo de reproducción, de modo que se exhiba una inclinación hacia las frecuencias más altas. El objetivo es simplemente garantizar que ninguna ráfaga de audio de mal comportamiento, potencialmente alta, llegue al oído del oyente, por medio de la inserción de "ruido de confort" (en contraste al estricto silenciamiento). De hecho, esto no es

50

un desvanecimiento real, sino, en cambio, un salto a un cierto nivel de energía a fin de insertar cierta clase de ruido de confort.

55 **[0085]** A continuación, se menciona una alternativa [EBU12, sección 5.6.3.2] que reutiliza los últimos datos decodificados correctamente y lentamente disminuye los niveles (L) hacia 0, de forma análoga al caso de la codificación de audio avanzada (AAC) + replicación de banda espectral (SBR).

60

[0086] Ahora, se considera MPEG-4 HILN (HILN = Armónicos y líneas individuales más ruido, conforme a sus siglas en inglés). Meine y col. introducen un desvanecimiento para el códec MPEG-4 HILN paramétrico [ISO09] en un dominio paramétrico [MEP01]. Para componentes armónicos continuados, un buen comportamiento por defecto para

el reemplazo de parámetros diferencialmente codificados alterados es mantener la frecuencia constante, a fin de reducir la amplitud por un factor de atenuación (por ejemplo, -6 dB), y dejar que la envoltura espectral converja hacia aquella de la característica de paso bajo promediado. Una alternativa para la envoltura espectral sería mantenerla sin cambios. Con respecto a amplitudes y envolturas espectrales, los componentes de ruido pueden ser tratados de la misma manera que los componentes armónicos.

[0087] En lo que sigue, se considera el rastreo del nivel de ruido de fondo en la técnica anterior. Rangachari y Loizou [RL06] proporcionan una buena reseña de varios procedimientos, y describen algunas de sus limitaciones. Los procedimientos para el rastreo del nivel de ruido de fondo son, por ejemplo, el procedimiento de pista mínima [RL06] [Coh03] [SFB00] [Dob95], sobre la base de VAD (VAD = detección de actividad de voz, conforme a sus siglas en inglés); la filtración de Kalman [Gan05] [BJH06], las descomposiciones de subespacios [BP06] [HJH08]; decisión programada [SS98] [MPC89] [HE95] y estadística mínima.

[0088] Se seleccionó la estrategia de estadística mínima que se va a utilizar dentro del alcance para USAC-2, (USAC = Codificación de voz y Audio Unificados, conforme a sus siglas en inglés) y se señala a continuación en más detalle.

[0089] La estimación de densidad espectral de potencia de ruido sobre la base de un suavizado óptimo y mínima estadística [Mar01] introduce un estimador de ruido que es capaz de funcionar independientemente de que la señal sea la voz activa o ruido de fondo. En contraste a otros procedimientos, el algoritmo de mínima estadística no utiliza ningún umbral explícito para distinguir entre actividad de la voz y pausa de la voz, y por lo tanto, se relaciona más cercanamente con los procedimientos de decisión programada, que con los procedimientos tradicionales de detección de actividad de voz. De manera similar a los procedimientos de decisión programada, puede actualizar también la PSD (Densidad espectral de potencia, conforme a sus siglas en inglés) de ruido estimado durante la actividad de la voz.

[0090] El procedimiento de mínima estadística se sustenta en dos observaciones, a saber, que la voz y el ruido habitualmente son estadísticamente independientes, y que la potencia de una señal de la voz ruidosa con frecuencia decae hasta el nivel de potencia del ruido. Por lo tanto, es posible derivar una estimación de densidad espectral de potencia (PSD) de ruido exacta mediante el rastreo de la mínima de la densidad espectral de potencia (PSD) de la señal ruidosa. Debido a que la mínima es menor (o en otros casos, igual) que el valor promedio, el procedimiento de rastreo mínimo requiere una compensación de inclinación.

[0091] La inclinación es una función de la variancia de la densidad espectral de potencia (PSD) de la señal suavizada, y como tal, depende del parámetro de suavizado del estimador de densidad espectral de potencia (PSD). En contraste al trabajo previo sobre el mínimo rastreo, que utiliza un parámetro de suavizado constante y una corrección de inclinación mínima constante, se usa un suavizado de densidad espectral de potencia (PSD) dependiente del tiempo y la frecuencia, que también requiere una compensación de inclinación dependiente del tiempo y la frecuencia.

[0092] El uso de rastreo mínimo proporciona una estimación aproximada de la potencia de ruido. Sin embargo, existen algunas desventajas. El suavizado con un parámetro de suavizado fijado amplía los picos de actividad de la voz de la estimación de densidad espectral de potencia (PSD) suavizada. Esto conducirá a estimaciones de ruido inexactas, ya que la ventana de deslizamiento para la búsqueda mínima podría deslizarse hacia picos amplios. En consecuencia, los parámetros de suavizado cercanos a uno no pueden ser usados, y en consecuencia, la estimación de ruido tendrá una variancia relativamente grande. Además, la estimación de ruido se inclina hacia valores menores. Adicionalmente, en el caso de incrementar la potencia de ruido el rastreo mínimo queda retrasado. El rastreo de la densidad espectral de potencia (PSD) de ruido sobre la base de MMSE, con baja complejidad [HHJ10] introduce una estrategia de densidad espectral de potencia (PSD) de ruido de fondo utilizando una búsqueda de MMSE empleada en un espectro de DFT (transformada discreta de Fourier, conforme a sus siglas en inglés). El algoritmo consiste en las siguientes etapas de procesamiento:

– Se computa el máximo estimador de probabilidad sobre la base de la densidad espectral de potencia (PSD) de ruido de la trama previa.

– Se computa el mínimo estimador de cuadrados medios.

– Se estima el máximo estimador de probabilidad usando la estrategia dirigida por decisión [EM84].

– Se computa el factor de inclinación inversa asumiendo que los coeficientes de transformada discreta de Fourier (DFT) de la voz y ruido son distribuidos en forma gaussiana.

– Se suaviza la densidad espectral de potencia de ruido estimada.

[0093] Existe además una estrategia de red de seguridad aplicada a fin de evitar un completo estancamiento

del algoritmo.

[0094] El rastreo de ruido no estacionario sobre la base de la estimación de potencia de ruido recursiva conducida por datos [EH08] introduce un procedimiento para la estimación de la varianza espectral de ruido a partir de señales de voz contaminadas por fuentes de ruido altamente no estacionarias. Este procedimiento utiliza también el suavizado en la dirección de tiempo/frecuencia.

[0095] Un algoritmo de estimación de ruido de baja complejidad sobre la base del suavizado de la estimación de potencia de ruido y la corrección de la inclinación de estimación [Yu09] mejora la estrategia introducida en [EH08]. La principal diferencia es que la función de ganancia espectral para la estimación de potencia de ruido es hallada por un procedimiento conducido por datos iterativos.

[0096] Los procedimientos estadísticos para el mejoramiento de la voz ruidosa [Mar03] combinan la estrategia de mínima estadística proporcionado en [Mar01] por la modificación de ganancia de decisión programada [MCA99], por una estimación de la relación de señal a ruido (SNR) a-priori [MCA99], por una limitación de ganancia adaptativa [MC99] y por un estimador de amplitud espectral log MMSE [EM85].

[0097] El desvanecimiento es de particular interés para una pluralidad de códecs de voz y audio, en particular, multitasa adaptativa (AMR) (véase [3GP12b]) (que incluye la predicción lineal excitada por código algebraico (ACELP) y la generación de ruido de confort (CNG)), multitasa adaptativa (AMR)-banda ancha (WB) (véase [3GP09c]) (que incluye la predicción lineal excitada por código algebraico (ACELP) y la generación de ruido de confort (CNG)), multitasa adaptativa (AMR)-banda ancha (WB)+ (véase [3GP09a]) (que incluye la predicción lineal excitada por código algebraico (ACELP), la excitación codificada por la transformada (TCX) y la generación de ruido de confort (CNG)), G.718 (véase [ITU08a]), G.719 (véase [ITU08b]), G.722 (véase [ITU07]), G.722.1 (véase [ITU05]), G.729 (véase [ITU12, CPK08, PKJ+11]), MPEG-4 HE-codificación de audio avanzada (AAC)/Codificación de audio avanzada (AAC) aacPlus mejorada (véanse [EBU10, EBU12, 3GP12e, LS01, QD03]) (que incluye la codificación de audio avanzada (AAC) y la replicación de banda espectral (SBR)), MPEG-4 HILN (véanse [ISO09, MEP01]) y OPUS (véase [IET12]) (que incluye SILK y la transformada superpuesta de energía restringida (CELT)).

[0098] Dependiendo del códec, el desvanecimiento se realiza en diferentes dominios: Para códecs que utilizan la codificación predictiva lineal (LPC), el desvanecimiento se realiza en el dominio predictivo lineal (también conocido como el dominio de excitación). Esto es cierto para códecs que se sustentan en la predicción lineal excitada por código algebraico (ACELP), por ejemplo, la multitasa adaptativa (AMR), multitasa adaptativa (AMR)-banda ancha (WB), núcleo de la predicción lineal excitada por código algebraico (ACELP) de multitasa adaptativa (AMR)-banda ancha (WB)+, G.718, G.729, G.729.1, el núcleo SILK en OPUS; códecs que además procesan la señal de excitación mediante el uso de una transformación de tiempo-frecuencia, por ejemplo, el núcleo de la excitación codificada por la transformada (TCX) de multitasa adaptativa (AMR)-banda ancha (WB)+, el núcleo de la transformada superpuesta de energía restringida (CELT) en OPUS; y para los esquemas de generación de ruido de confort (CNG) que operan en el dominio predictivo lineal, por ejemplo, la generación de ruido de confort (CNG) en multitasa adaptativa (AMR), generación de ruido de confort (CNG) en multitasa adaptativa (AMR)-banda ancha (WB), generación de ruido de confort (CNG) en multitasa adaptativa (AMR)-banda ancha (WB)+.

[0099] Para códecs que directamente transforman la señal de tiempo en el dominio de frecuencia, el desvanecimiento se realiza en el dominio espectral/subbanda. Esto es cierto para códecs que se sustentan en la transformada de coseno discreta modificada (MDCT) o una transformación similar, tal como la codificación de audio avanzada (AAC) en MPEG-4 HE-codificación de audio avanzada (AAC), G.719, G.722 (dominio de subbanda) y G.722.1.

[0100] Para códecs paramétricos, el desvanecimiento se aplica en el dominio paramétrico. Esto es cierto para MPEG-4 HILN.

[0101] Con respecto a la velocidad del desvanecimiento y la curva del desvanecimiento, un desvanecimiento comúnmente es realizado mediante la aplicación de un factor de atenuación, que se aplica a la representación de señal en el dominio apropiado. El tamaño del factor de atenuación controla la velocidad del desvanecimiento y la curva del desvanecimiento. En la mayoría de los casos, el factor de atenuación se aplica a modo de trama, pero también se utiliza una aplicación a modo de muestra, véase, por ejemplo, G.718 y G.722.

[0102] El factor de atenuación para un cierto segmento de señal podría ser proporcionado de dos maneras, absoluta y relativa.

[0103] En el caso de que se proporcione un factor de atenuación de manera absoluta, el nivel de referencia es siempre el de la última trama recibida. Los factores de atenuación absolutos habitualmente se inician con un valor cercano a 1 para el segmento de señal inmediatamente después de la última trama buena y, a continuación, se degradan de forma más rápida o más lenta, hacia 0. La curva de desvanecimiento depende directamente de estos factores. Es decir, por ejemplo, el caso para el ocultamiento descrito en el Apéndice IV de G.722 (véase, en particular,

[ITU07, figura IV.7]), donde las posibles curvas de desvanecimiento son lineales o gradualmente lineales. Considerando un factor de ganancia $g(n)$, mientras que $g(0)$ representa el factor de ganancia de la última trama buena, un factor de atenuación absoluto $\alpha_{abs}(n)$, el factor de ganancia de cualquier trama perdida subsiguiente puede derivarse como:

5

$$g(n) = \alpha_{abs}(n) \cdot g(0) \quad (21)$$

[0104] En el caso donde se proporciona un factor de atenuación de manera relativa, el nivel de referencia es el de la trama previa. Esto tiene ventajas en el caso de un procedimiento de ocultamiento recursivo, por ejemplo, si la señal ya atenuada es adicionalmente procesada y atenuada de nuevo.

[0105] Si un factor de atenuación se aplica de manera recursiva, entonces este podría ser un valor fijado independiente del número de tramas consecutivamente perdidas, por ejemplo, 0,5 para G.719 (véase más arriba); un valor fijado relativo al número de tramas consecutivamente perdidas, por ejemplo, como se propone para G.729 en [CPK08]: 1,0 para las primeras dos tramas, 0,9 para las siguientes dos tramas, 0,8 para las tramas 5 y 6, y 0, para todas las subsiguientes tramas (véase más arriba); o un valor que es relativo al número de tramas consecutivamente perdidas y que depende de las características de señal, por ejemplo, un desvanecimiento más rápido para una señal inestable, y un desvanecimiento más lento, para una señal estable, por ejemplo, G.718 (véase la sección anterior y [ITU08a, tabla 44]);

20

[0106] Asumiendo un factor de desvanecimiento relativo $0 \leq \alpha_{rel}(n) \leq 1$, mientras que n es el número de la trama perdida ($n \geq 1$); el factor de ganancia de cualquier trama subsiguiente puede derivarse como:

$$g(n) = \alpha_{rel}(n) \cdot g(n-1) \quad (22)$$

25

$$g(n) = \left(\prod_{m=1}^n \alpha(m) \right) \cdot g(0) \quad (23)$$

$$g(n) = \alpha_{rel}^n \cdot g(0) \quad (24)$$

30 resultando en un desvanecimiento exponencial.

[0107] Con respecto al procedimiento de desvanecimiento, habitualmente, se especifica el factor de atenuación, pero en algunos estándares de aplicación (gerenciamiento de derechos digitales (DRM), transmisión de audio digital (DAB)+) este último se deja a criterio del fabricante.

35

[0108] Si diferentes partes de señal son desvanecidas de forma separada, podrían aplicarse diferentes factores de atenuación, por ejemplo, a fin de cambiar gradualmente el volumen de los componentes tonales con una cierta velocidad, y los componentes de tipo ruido, con otra velocidad (por ejemplo, multitasa adaptativa (AMR), SILK).

[0109] Habitualmente, se aplica una cierta señal a la trama entera. Cuando el desvanecimiento se realiza en el dominio espectral, esta es la única manera posible. Sin embargo, si el desvanecimiento se realiza en el dominio de tiempo o en el dominio predictivo lineal, es posible un desvanecimiento más granular. Dicho desvanecimiento más granular se aplica en G.718, donde los factores de ganancia individuales son derivados para cada muestra por interpolación lineal entre el factor de ganancia de la última trama y el factor de ganancia de la trama actual.

45

[0110] Para códecs con una duración de trama variable, un factor de atenuación relativo constante conduce a una velocidad de desvanecimiento diferente dependiendo de la duración de la trama. Este es, por ejemplo, el caso para la codificación de audio avanzada (AAC), donde la duración de la trama depende de la tasa de muestreo.

[0111] A fin de adoptar la curva de desvanecimiento aplicada a la forma temporal de la última señal recibida, los factores de desvanecimiento (estáticos) podrían ser adicionalmente ajustados. Dicho ajuste dinámico adicional se aplica, por ejemplo, para la multitasa adaptativa (AMR), donde se tiene en cuenta la mediana de los cinco factores de ganancia previos (véase [3GP12b] y la sección 1.8.1). Antes de realizar cualquier atenuación, la ganancia actual se establece en la mediana, si la mediana es menor que la última ganancia; de lo contrario, se usa la última ganancia.

50

Además, dicho ajuste dinámico adicional, por ejemplo, se aplica para G729, donde se predice la amplitud usando la regresión lineal de los factores de ganancia previos (véase [CPK08, PKJ+11] y la sección 1.6). En este caso, el factor de ganancia resultante para las primeras tramas ocultadas podría exceder al factor de ganancia de la última trama recibida.

55

[0112] Con respecto al nivel objetivo del desvanecimiento, con la excepción de G.718 y la transformada superpuesta de energía restringida (CELT), el nivel objetivo es 0 para todos los códecs analizados, que incluyen aquellos códecs de generación de ruido de confort (CNG).

[0113] En G.718, el desvanecimiento de la excitación de altura (que representa los componentes tonales) y el desvanecimiento de la excitación aleatoria (que representa los componentes de tipo ruido) se realizan de forma separada. Mientras que el factor de ganancia de altura se desvanece a cero, el factor de ganancia de innovación se desvanece a la energía de excitación de la generación de ruido de confort (CNG).

[0114] Asumiendo que se proporcionan los factores de atenuación relativos, esto conduce –sobre la base de la fórmula (23) –al siguiente factor de atenuación absoluto:

$$g(n) = \alpha_{rel}(n) \cdot g(n - 1) + (1 - \alpha_{rel}(n)) \cdot g_n \quad (25)$$

donde g_n es la ganancia de la excitación usada durante la generación de ruido de confort. Esta fórmula corresponde a la fórmula (23), cuando $g_n = 0$.

[0115] G.718 no realiza desvanecimiento en el caso de Sistemas de Teatro Digital (DTX)/generación de ruido de confort (CNG).

[0116] En la transformada superpuesta de energía restringida (CELT) no hay desvanecimiento hacia el nivel objetivo, pero, después de 5 tramas de ocultamiento tonal (que incluye un desvanecimiento), el nivel es cambiado instantáneamente al nivel objetivo en la 6ª trama consecutivamente perdida. El nivel es derivado a modo de banda usando la fórmula (19).

[0117] Con respecto a la forma espectral objetivo del desvanecimiento, todos los códecs sobre la base de la transformada pura analizados (codificación de audio avanzada (AAC), G.719, G.722, G.722.1), al igual que la replicación de banda espectral (SBR), simplemente prolongan la forma espectral de la última trama buena durante el desvanecimiento.

[0118] Diversos códecs de voz desvanecen la forma espectral a una media, usando la síntesis de codificación predictiva lineal (LPC). La media podría ser estática (multitasa adaptativa (AMR)) o adaptativa (multitasa adaptativa (AMR)–banda ancha (WB), multitasa adaptativa (AMR)–banda ancha (WB)+, G.718), mientras que esta última se deriva a partir de una media estática y una media a corto plazo (derivada promediando los últimos grupos de coeficientes n LP) (LP = Predicción lineal).

[0119] Todos los módulos de generación de ruido de confort (CNG) en los códecs descritos de multitasa adaptativa (AMR), multitasa adaptativa (AMR)–banda ancha (WB), multitasa adaptativa (AMR)–banda ancha (WB)+, G.718 prolongan la forma espectral de la última trama buena durante el desvanecimiento.

[0120] Con respecto al rastreo del nivel de ruido de fondo, existen cinco estrategias diferentes conocidas de la bibliografía:

– Sobre la base del detector de actividad de voz: se sustenta en la relación de señal a ruido (SNR)/VAD (detección de actividad de voz), pero muy difícil de sintonizar y de usar para la voz SNR baja.

– Esquema de decisión programada: La estrategia de decisión programada tiene en cuenta la probabilidad de la presencia de la voz [SS98] [MPC89] [HE95].

– Mínima estadística: se rastrea la mínima de la densidad espectral de potencia (PSD) manteniendo una cierta cantidad de valores con el tiempo en un búfer, de modo que se permita hallar el mínimo ruido de las muestras pasadas [Mar01] [HHJ10] [EH08] [Yu09].

– Filtración de Kalman: el algoritmo utiliza una serie de medidas observadas en función del tiempo, que contienen ruido (variaciones aleatorias), y produce estimaciones de la densidad espectral de potencia (PSD) de ruido que tienden a ser más precisas que aquellas sobre la base de una única medición sola. El filtro de Kalman opera de manera recursiva sobre corrientes de datos de entrada ruidosos para producir una estimación estadísticamente óptima del estado del sistema [Gan05] [BJH06].

– Descomposición subespacial: esta estrategia trata de descomponer una señal de tipo ruido en una señal discursiva limpia y una parte de ruido, utilizando, por ejemplo, la KLT (transformada de Karhunen–Loève, también conocida como análisis de componentes principales) y/o la transformada discreta de Fourier (DFT). A continuación, pueden rastrearse

vectores propios/valores propios usando un algoritmo de suavizado arbitrario [BP06] [HJH08].

[0121] "Sistema digital de telecomunicaciones celulares (Fase 2+); Sistema Universal de Telecomunicaciones Móviles (UMTS); LTE; Funciones de procesamiento de códec de audio; Códec Adaptativo Extendido Multitasa - Banda Ancha (AMR-WB +); Funciones de transcodificación (3GPP TS26.290 versión 9.0. 0 Edición 9) ", INSTITUTO EUROPEO DE NORMAS DE TELECOMUNICACIONES DE ESPECIFICACIÓN TÉCNICA (ETSI), 650, ROUTE DES LUCIOLES; F-06921 SOPHIA-ANTIPOLIS; FRANCIA, n.º V9.0.0, 1 de enero de 2010 describe el mapeo detallado de bloques de entrada de muestras de audio monofónicas o estereofónicas a bloques codificados y además a bloques de salida de muestras de audio monofónicas o estereofónicas reconstruidas. Para codificar la señal mono núcleo, se utiliza la codificación ACELP o TCX para cada trama.

[0122] El documento EP 2 026 330 A1 describe un dispositivo y un procedimiento para el ocultamiento de pérdida de trama. El periodo de tono de una trama perdida actual se obtiene sobre la base de un periodo de tono de la última trama buena anterior a la trama perdida actual. Se recupera una señal de excitación de la trama perdida actual en base al periodo de tono de la trama perdida actual y una señal de excitación de la última trama buena antes de la trama perdida.

[0123] El objeto de la presente invención es proporcionar conceptos mejorados para sistemas de codificación de audio. El objeto de la presente invención es resuelto por un aparato según la reivindicación 1, por un procedimiento según la reivindicación 14 y por un programa informático según la reivindicación 15.

[0124] Se proporciona un aparato para decodificar una señal de audio codificada para obtener una señal de audio reconstruida. El aparato comprende una interfaz receptora para recibir una pluralidad de tramas, un búfer de retardo para almacenar muestras de señales de audio de la señal de audio decodificada, un selector de muestras para seleccionar una pluralidad de muestras de señales de audio seleccionadas de las muestras de señales de audio que están almacenadas en el búfer de retardo, y un procesador de muestras para procesar las muestras de señales de audio seleccionadas para obtener muestras de señales de audio reconstruidas de la señal de audio reconstruida. El selector de muestra se configura para seleccionar, si se recibe una trama actual por la interfaz receptora y si la trama actual que se recibe por la interfaz receptora no está alterada, la pluralidad de muestras de señales de audio seleccionadas de las muestras de señales de audio que están almacenadas en el búfer de retardo dependiendo de una información de retraso de altura que está comprendida por la trama actual. Además, el selector de muestras se configura para seleccionar, si la trama actual no se recibe por la interfaz receptora o si la trama actual que se recibe por la interfaz receptora está alterada, la pluralidad de muestras de señales de audio seleccionadas de las muestras de señales de audio que están almacenadas en el búfer de retardo dependiendo de una información de retraso de altura que está comprendida por otra trama que se recibe previamente por la interfaz receptora.

[0125] Según un ejemplo útil para comprender la invención, el procesador de muestras se puede configurar, por ejemplo, para obtener las muestras de señales de audio reconstruidas, si la trama actual se recibe por la interfaz receptora y si la trama actual que se recibe por la interfaz receptora no está alterada, al reescalar las muestras de señal de audio seleccionadas dependiendo de la información de ganancia que esté comprendida por la trama actual. Además, el selector de muestras se puede configurar, por ejemplo, para obtener las muestras de señales de audio reconstruidas, si la trama actual no se recibe por la interfaz receptora o si la trama actual que se recibe por la interfaz receptora está alterada, al reescalar las muestras de señales de audio seleccionadas dependiendo de la información de ganancia que esté comprendida por la otra trama que se recibe previamente por la interfaz receptora.

[0126] En un ejemplo útil para comprender la invención, el procesador de muestras se puede configurar, por ejemplo, para obtener las muestras de señales de audio reconstruidas, si la trama actual se recibe por la interfaz receptora y si la trama actual que se recibe por la interfaz receptora no está alterada, al multiplicar las muestras de señales de audio seleccionadas y un valor dependiendo de la información de ganancia que esté comprendida por la trama actual. Además, el selector de muestras se configura para obtener las muestras de señales de audio reconstruidas, si la trama actual no se recibe por la interfaz receptora o si la trama actual que se recibe por la interfaz receptora está alterada, al multiplicar las muestras de señales de audio seleccionadas y un valor dependiendo de la información de ganancia que esté comprendida por la otra trama que se recibe previamente por la interfaz receptora.

[0127] Según un ejemplo útil para comprender la invención, el procesador de muestras se puede configurar, por ejemplo, para almacenar las muestras de señales de audio reconstruidas en el búfer de retardo.

[0128] En un ejemplo útil para comprender la invención, el procesador de muestras se puede configurar, por ejemplo, para almacenar las muestras de señales de audio reconstruidas en el búfer de retardo antes de que se reciba una trama adicional por la interfaz receptora.

[0129] Según un ejemplo útil para comprender la invención, el procesador de muestras se puede configurar, por ejemplo, para almacenar las muestras de señales de audio reconstruidas en el búfer de retardo después de que se recibe una trama adicional por la interfaz receptora.

65

[0130] En un ejemplo útil para comprender la invención, el procesador de muestras se puede configurar, por ejemplo, para reescalar las muestras de señales de audio seleccionadas dependiendo de la información de ganancia para obtener las muestras de señales de audio reescaladas y al combinar las muestras de señales de audio reescaladas con las muestras de señales de audio de entrada para obtener las muestras de señales de audio procesadas.

[0131] Según un ejemplo útil para comprender la invención, el procesador de muestras se puede configurar, por ejemplo, para almacenar las muestras de señales de audio procesadas, indicando la combinación de las muestras de señales de audio reescaladas y las muestras de señales de audio de entrada, en el búfer de retardo, y para no almacenar las muestras de señales de audio reescaladas en el búfer de retardo, si la trama actual se recibe por la interfaz receptora y si la trama actual que se recibe por la interfaz receptora no está alterada. Además, el procesador de muestras se configura para almacenar las muestras de señales de audio reescaladas en el búfer de retardo y para no almacenar las muestras de señales de audio procesadas en el búfer de retardo. Si la trama actual no se recibe por la interfaz receptora o si la trama actual que se recibe por la interfaz receptora está alterada.

[0132] Según otro ejemplo útil para comprender la invención, el procesador de muestras se puede configurar, por ejemplo, para almacenar las muestras de señales de audio procesadas en el búfer de retardo, si la trama actual no se recibe por la interfaz receptora o si la trama actual que se recibe por la interfaz receptora está alterada.

[0133] En un ejemplo útil para comprender la invención, el selector de muestras se puede configurar, por ejemplo, para obtener las muestras de señales de audio reconstruidas al reescalar las muestras de señales de audio seleccionadas dependiendo de una ganancia modificada, donde la ganancia modificada se define según la fórmula:

$$\text{ganancia} = \text{ganancia_pasada} * \text{amortiguación};$$

donde ganancia es la ganancia modificada, donde el selector de muestras se puede configurar, por ejemplo, para ajustar la ganancia_pasada a la ganancia después de la ganancia y se ha calculado, y donde la amortiguación es un valor real.

[0134] Según un ejemplo útil para comprender la invención, el selector de muestra se puede configurar, por ejemplo, para calcular la ganancia modificada.

[0135] En un ejemplo útil para comprender la invención, la amortiguación se puede definir, por ejemplo, según: $0 \leq \text{amortiguación} \leq 1$.

[0136] Según un ejemplo útil para comprender la invención, la ganancia modificada, ganancia, se puede ajustar, por ejemplo, a cero, si no se ha recibido al menos un número predefinido de tramas por la interfaz receptora antes de que se haya recibido la última trama por la interfaz receptora.

[0137] Además, se proporciona un procedimiento para decodificar una señal de audio codificada para obtener una señal de audio reconstruida. El procedimiento comprende:

- recibir una pluralidad de tramas,
- almacenar las muestras de señales de audio de la señal de audio decodificada,
- seleccionar una pluralidad de muestras de señales de audio seleccionadas de las muestras de señales de audio que están almacenadas en el búfer de retardo, y;
- procesar las muestras de señales de audio seleccionadas para obtener muestras de señales de audio reconstruidas de la señal de audio reconstruida.

[0138] Si se recibe una trama actual y si la trama actual que se recibe no está alterada, la etapa de seleccionar la pluralidad de muestras de señales de audio seleccionadas de las muestras de señales de audio que están almacenadas en el búfer de retardo se lleva a cabo dependiendo de una información de retraso de altura que está comprendida por la trama actual. Además, si la trama actual no se recibe o si la trama actual que se recibe está alterada, la etapa de seleccionar la pluralidad de muestras de señales de audio seleccionadas de las muestras de señales de audio que están almacenadas en el búfer de retardo se lleva a cabo dependiendo de una información de retraso de altura que está comprendida por otra trama que se recibe previamente por la interfaz receptora.

[0139] Además, se proporciona un programa informático para implementar el procedimiento descrito anteriormente cuando se ejecuta en un ordenador o procesador de señal.

[0140] Ejemplos útiles para comprender la invención emplean TCX LTP (TCX LTP= predicción a largo plazo por excitación codificada por transformada). Durante la operación normal, la memoria TCX LTP se actualiza con la señal sintetizada, que contiene componentes tonales reconstruidos y ruido.

[0141] En lugar de deshabilitar la TCX LTP durante la ocultación, su operación normal se puede continuar durante la ocultación con los parámetros recibidos en la última trama buena. Esto preserva la forma espectral de la señal, particularmente aquellos componentes tonales que se modelan por el filtro LTP.

5 **[0142]** Además, los ejemplos útiles para comprender la invención desacoplan el circuito de retroalimentación TCX LTP. Una continuación simple de la operación TCX LTP normal introduce ruido adicional, puesto que, con cada etapa de actualización, se introduce ruido aleatoriamente generado, adicional de la excitación LTP. Los componentes tonales por lo tanto se obtienen distorsionados más y más cada vez por el ruido adicionado.

10 **[0143]** Para superar esto, solo el búfer TCX LTP actualizado se puede retroalimentar (sin adicionar ruido) a fin de no contaminar la información tonal con el ruido aleatorio iniciado.

[0144] Adicionalmente, según ejemplos útiles para comprender la invención, la ganancia TCX LTP se desvanece a cero.

15

[0145] Estos ejemplos útiles para comprender la invención se basan en el hallazgo de continuar la TCX LTP que ayuda a preservar las características de la señal a corto plazo, pero tiene desventajas a largo plazo: la señal reproducida durante la ocultación incluirá información de voz/tonal que estuvo presente precediendo a la pérdida.

20 Especialmente para voz clara o voz sobre ruido de fondo, es extremadamente improbable que un tono o armónica decaiga más lentamente con un tiempo más prolongado. Al continuar la operación de TCX LTP durante el ocultamiento, particularmente si se desacopla la actualización de memoria LTP (solo los componentes tonales se retroalimentan y no la señal mezclada separada), la información de voz/tonal aún estará presente en la señal ocultada para la pérdida completa, que se atenúa solo por el desvanecimiento total al nivel de ruido cómodo. Además, es imposible alcanzar la envolvente de ruido cómodo durante las pérdidas de paquetes por ráfaga, si la TCX LTP se aplica durante la pérdida de ráfaga sin que se atenúe con el paso del tiempo, debido a que la señal incorporará siempre entonces la información de voz de la LTP.

25

[0146] Por lo tanto, la ganancia de TCX LTP se desvanece hacia cero, de tal forma que se desvanecerán a cero los componentes tonales representados por la LTP, al mismo tiempo, la señal se desvanece al nivel y forma de señal de fondo y de tal forma que el desvanecimiento alcanza la envolvente de fondo espectral deseado (ruido cómodo) sin incorporar componentes tonales no deseados.

30

[0147] En ejemplos útiles para comprender la invención, se usa la misma velocidad de desvanecimiento para el desvanecimiento de la ganancia de LTP como para el desvanecimiento de ruido blanco.

35

[0148] En contraste, en la técnica anterior, no hay un códec de transformada conocido que use LTP durante la ocultación. Para la LTP de MPEG-4 [ISO09] no existe planteamiento de cancelación en la técnica anterior. Otro códec basado en MDCT de la técnica anterior que hace uso a una LTP es CELT, pero este códec usa una ocultación de tipo ACELP para las primeras cinco tramas y para todas las subsiguientes tramas se genera ruido de fondo, lo que no hace uso de la LTP. Una desventaja de la técnica anterior de no usar la TCX LTP es, que todos los componentes tonales que se modelan con la LTP desaparecen de forma abrupta. Además, en los códecs basados en ACELP de la técnica anterior, la operación de LTP se prolonga durante el ocultamiento y la ganancia del libro de códigos adaptables se desvanece hacia cero. Con respecto a la operación de circuito de retroalimentación, la técnica anterior emplea dos estrategias, ya sea la excitación completa, por ejemplo, la suma de la excitación innovativa y adaptativa se retroalimenta (AMR-WB); y solo la excitación adaptativa actualizada, por ejemplo, las partes de la señal tonal, se retroalimenta (G.178). Los ejemplos útiles para comprender la invención mencionados anteriormente superan las desventajas de la técnica anterior.

40

45

[0149] Además, se proporciona un aparato para la decodificación de una señal de audio.

50

[0150] El aparato comprende una interfaz receptora. La interfaz receptora está configurada para recibir una pluralidad de tramas, donde la interfaz receptora está configurada para recibir una primera trama de la pluralidad de tramas, donde dicha primera trama comprende una primera porción de señal de audio de la señal de audio, estando dicha primera porción de señal de audio representada en un primer dominio, y donde la interfaz receptora está configurada para recibir una segunda trama de la pluralidad de tramas, comprendiendo dicha segunda trama una segunda porción de señal de audio de la señal de audio.

55

[0151] Además, el aparato comprende una unidad de transformada para la transformación de la segunda porción de señal de audio o un valor o una señal derivados de la segunda porción de señal de audio, de un segundo dominio a un dominio de rastreo de manera que se obtenga una información de segunda porción de señal, donde el segundo dominio es diferente del primer dominio, donde el dominio de rastreo es diferente del segundo dominio, y donde el dominio de rastreo es igual o diferente con respecto al primer dominio.

60

[0152] Adicionalmente, el aparato comprende una unidad de rastreo de nivel de ruido, donde la unidad de rastreo de nivel de ruido está configurada para recibir una información de primera porción de señal que es representada

65

en el dominio de rastreo, donde la información de primera porción de señal depende de la primera porción de señal de audio. La unidad de rastreo de nivel de ruido está configurada para recibir la segunda porción de señal que es representada en el dominio de rastreo, y donde la unidad de rastreo de nivel de ruido está configurada para determinar la información de nivel de ruido según la información de primera porción de señal que es representada en el dominio de rastreo y según la información de segunda porción de señal que es representada en el dominio de rastreo.

[0153] Además, el aparato comprende una unidad de reconstrucción para la reconstrucción de una tercera porción de señal de audio de la señal de audio según la información de nivel de ruido, si una tercera trama de la pluralidad de tramas no es recibida por la interfaz receptora, pero es alterada.

[0154] Una señal de audio, por ejemplo, puede ser una señal de voz, o una señal de música, o señal que comprende voz y música, etc.

[0155] La afirmación de que la información de primera porción de señal depende de la primera porción de señal de audio significa que la información de primera porción de señal es o bien la primera porción de señal de audio, o que la información de primera porción de señal ha sido obtenida/generada según la primera porción de señal de audio o que de alguna otra manera depende de la primera porción de señal de audio. Por ejemplo, la primera porción de señal de audio puede haber sido transformada de un dominio a otro dominio a fin de obtener la información de primera porción de señal.

[0156] De manera similar, una afirmación de que la información de segunda porción de señal depende de una segunda porción de señal de audio significa que la información de segunda porción de señal es o bien la segunda porción de señal de audio, o que la información de segunda porción de señal ha sido obtenida/generada según la segunda porción de señal de audio o de alguna otra manera depende de la segunda porción de señal de audio. Por ejemplo, la segunda porción de señal de audio puede haber sido transformada de un dominio a otro dominio para obtener información de segunda porción de señal.

[0157] En un ejemplo útil para comprender la invención, la primera porción de señal de audio, por ejemplo, puede ser representada en un dominio de tiempo como el primer dominio. Además, la unidad de transformada puede ser configurada, por ejemplo, para transformar la segunda porción de señal de audio o el valor derivado de la segunda porción de señal de audio, de un dominio de excitación que es el segundo dominio, al dominio de tiempo que es el dominio de rastreo. Adicionalmente, la unidad de rastreo de nivel de ruido, por ejemplo, puede ser configurada para recibir la información de primera porción de señal que es representada en el dominio de tiempo como el dominio de rastreo. Además, la unidad de rastreo de nivel de ruido, por ejemplo, puede ser configurada para recibir la segunda porción de señal que es representada en el dominio de tiempo como el dominio de rastreo.

[0158] Según un ejemplo útil para comprender la invención, la primera porción de señal de audio, por ejemplo, puede ser representada en un dominio de excitación como el primer dominio. Además, la unidad de transformada, por ejemplo, puede ser configurada para transformar la segunda porción de señal de audio o el valor derivado de la segunda porción de señal de audio, de un dominio de tiempo que es el segundo dominio, al dominio de excitación que es el dominio de rastreo. Adicionalmente, la unidad de rastreo de nivel de ruido, por ejemplo, puede ser configurada para recibir la información de primera porción de señal que es representada en el dominio de excitación como el dominio de rastreo. Además, la unidad de rastreo de nivel de ruido, por ejemplo, puede ser configurada para recibir la segunda porción de señal que es representada en el dominio de excitación como el dominio de rastreo.

[0159] En un ejemplo útil para comprender la invención, la primera porción de señal de audio, por ejemplo, puede ser representada en un dominio de excitación como el primer dominio, donde la unidad de rastreo de nivel de ruido, por ejemplo, puede ser configurada para recibir la información de primera porción de señal, donde dicha información de primera porción de señal es representada en el dominio FFT, que es el dominio de rastreo, y donde dicha información de primera porción de señal depende de que dicha primera porción de señal de audio está representada en el dominio de excitación, donde la unidad de transformada, por ejemplo, puede ser configurada para transformar la segunda porción de señal de audio o el valor derivado de la segunda porción de señal de audio, de un dominio de tiempo que es el segundo dominio, a un dominio FFT que es el dominio de rastreo, y donde la unidad de rastreo de nivel de ruido, por ejemplo, puede ser configurada para recibir la segunda porción de señal de audio que es representada en el dominio FFT.

[0160] En un ejemplo útil para comprender la invención, el aparato, por ejemplo, puede comprender adicionalmente una primera unidad de agregación para la determinación de un primer valor agregado según la primera porción de señal de audio. Además, el aparato, por ejemplo, puede comprender adicionalmente una segunda unidad de agregación para la determinación, según la segunda porción de señal de audio, de un segundo valor agregado como el valor derivado de la segunda porción de señal de audio. Adicionalmente, la unidad de rastreo de nivel de ruido, por ejemplo, puede ser configurada para recibir el primer valor agregado como la información de primera porción de señal que es representada en el dominio de rastreo, donde la unidad de rastreo de nivel de ruido, por ejemplo, puede ser configurada para recibir el segundo valor agregado como la información de segunda porción de señal que es representado en el dominio de rastreo, y donde la unidad de rastreo de nivel de ruido, por ejemplo, puede estar

configurada para determinar información de nivel de ruido según el primer valor agregado que es representado en el dominio de rastreo y dependiendo del segundo valor agregado que es representado en el dominio de rastreo.

- 5 **[0161]** Según un ejemplo útil para comprender la invención, la primera unidad de agregación, por ejemplo, puede estar configurada para determinar el primer valor agregado de tal manera que el primer valor agregado indica una media cuadrática de la primera porción de señal de audio o de una señal derivada de la primera porción de señal de audio. Además, la segunda unidad de agregación, por ejemplo, puede estar configurada para determinar el segundo valor agregado de tal manera que el segundo valor agregado indica una media cuadrática de la segunda porción de señal de audio o de una señal derivada de la segunda porción de señal de audio.
- 10 **[0162]** En un ejemplo útil para comprender la invención, la unidad de transformada, por ejemplo, puede ser configurada para transformar el valor derivado de la segunda porción de señal de audio del segundo dominio al dominio de rastreo, mediante la aplicación de un valor de ganancia sobre el valor derivado de la segunda porción de señal de audio.
- 15 **[0163]** Según ejemplos útiles para comprender la invención, el valor de ganancia, por ejemplo, puede indicar una ganancia introducida por La síntesis de codificación predictiva lineal, o el valor de ganancia, por ejemplo, puede indicar una ganancia introducida por la síntesis de codificación predictiva lineal y el desénfasis.
- 20 **[0164]** En un ejemplo útil para comprender la invención, la unidad de rastreo de nivel de ruido, por ejemplo, puede estar configurada para determinar información de nivel de ruido mediante la aplicación de una estrategia de estadística mínima.
- 25 **[0165]** Según un ejemplo útil para comprender la invención, la unidad de rastreo de nivel de ruido, por ejemplo, puede estar configurada para determinar un nivel de ruido de confort como la información de nivel de ruido. La unidad de reconstrucción, por ejemplo, puede ser configurada para reconstruir la tercera porción de señal de audio según la información de nivel de ruido, si dicha tercera trama de la pluralidad de tramas no es recibida por la interfaz receptora o si dicha tercera trama es recibida por la interfaz receptora, pero es alterada.
- 30 **[0166]** En un ejemplo útil para comprender la invención, la unidad de rastreo de nivel de ruido, por ejemplo, puede estar configurada para determinar un nivel de ruido de confort como la información de nivel de ruido derivada de un espectro de nivel de ruido, donde dicho espectro de nivel de ruido se obtiene mediante la aplicación de la estrategia de estadística mínima. La unidad de reconstrucción, por ejemplo, puede ser configurada para reconstruir la tercera porción de señal de audio según una pluralidad de coeficientes predictivos lineales, si dicha tercera trama de la pluralidad de tramas no es recibida por la interfaz receptora o si dicha tercera trama es recibida por la interfaz receptora, pero es alterada.
- 35 **[0167]** Según otro ejemplo útil para comprender la invención, la unidad de rastreo de nivel de ruido, por ejemplo, puede estar configurada para determinar una pluralidad de coeficientes predictivos lineales que indican un nivel de ruido de confort como la información de nivel de ruido, y la unidad de reconstrucción, por ejemplo, puede ser configurada para reconstruir la tercera porción de señal de audio según la pluralidad de coeficientes predictivos lineales.
- 40 **[0168]** En un ejemplo útil para comprender la invención, la unidad de rastreo de nivel de ruido está configurada para determinar una pluralidad de coeficientes de FFT que indican un nivel de ruido de confort como la información de nivel de ruido, y la primera unidad de reconstrucción está configurada para reconstruir la tercera porción de señal de audio según un nivel de ruido de confort derivado a partir de dichos coeficientes de FFT, si dicha tercera trama de la pluralidad de tramas no es recibida por la interfaz receptora o si dicha tercera trama es recibida por la interfaz receptora pero es alterada.
- 45 **[0169]** En un ejemplo útil para comprender la invención, la unidad de reconstrucción, por ejemplo, puede ser configurada para reconstruir la tercera porción de señal de audio según la información de nivel de ruido y según la primera porción de señal de audio, si dicha tercera trama de la pluralidad de tramas no es recibida por la interfaz receptora o si dicha tercera trama es recibida por la interfaz receptora, pero es alterada.
- 50 **[0170]** Según un ejemplo útil para comprender la invención, la unidad de reconstrucción, por ejemplo, puede ser configurada para reconstruir la tercera porción de señal de audio mediante la atenuación o la amplificación de una señal derivada de la primera o la segunda porción de señal de audio.
- 55 **[0171]** En un ejemplo útil para comprender la invención, el aparato, por ejemplo, puede comprender adicionalmente una unidad de predicción a largo plazo que comprende un búfer de retardo. Además, la unidad de predicción a largo plazo, por ejemplo, puede ser configurada para generar una señal procesada según la primera o la segunda porción de señal de audio, según una entrada de búfer de retardo que está almacenada en el búfer de retardo y según una ganancia de predicción a largo plazo. Adicionalmente, la unidad de predicción a largo plazo, por ejemplo, puede ser configurada para desvanecer la ganancia de predicción a largo plazo hacia cero, si dicha tercera trama de
- 60 **[0171]** En un ejemplo útil para comprender la invención, el aparato, por ejemplo, puede comprender adicionalmente una unidad de predicción a largo plazo que comprende un búfer de retardo. Además, la unidad de predicción a largo plazo, por ejemplo, puede ser configurada para generar una señal procesada según la primera o la segunda porción de señal de audio, según una entrada de búfer de retardo que está almacenada en el búfer de retardo y según una ganancia de predicción a largo plazo. Adicionalmente, la unidad de predicción a largo plazo, por ejemplo, puede ser configurada para desvanecer la ganancia de predicción a largo plazo hacia cero, si dicha tercera trama de
- 65 **[0171]** En un ejemplo útil para comprender la invención, el aparato, por ejemplo, puede comprender adicionalmente una unidad de predicción a largo plazo que comprende un búfer de retardo. Además, la unidad de predicción a largo plazo, por ejemplo, puede ser configurada para generar una señal procesada según la primera o la segunda porción de señal de audio, según una entrada de búfer de retardo que está almacenada en el búfer de retardo y según una ganancia de predicción a largo plazo. Adicionalmente, la unidad de predicción a largo plazo, por ejemplo, puede ser configurada para desvanecer la ganancia de predicción a largo plazo hacia cero, si dicha tercera trama de

la pluralidad de tramas no es recibida por la interfaz receptora o si dicha tercera trama es recibida por la interfaz receptora, pero es alterada.

5 **[0172]** Según un ejemplo útil para comprender la invención, la unidad de predicción a largo plazo, por ejemplo, puede ser configurada para desvanecer la ganancia de predicción a largo plazo hacia cero, donde una velocidad con la cual la ganancia de predicción a largo plazo es desvanecida a cero depende de un factor de desvanecimiento.

10 **[0173]** En un ejemplo útil para comprender la invención, la unidad de predicción a largo plazo, por ejemplo, puede ser configurada para actualizar la entrada del búfer de retardo mediante el almacenamiento de la señal procesada generada en el búfer de retardo, si dicha tercera trama de la pluralidad de tramas no es recibida por la interfaz receptora o si dicha tercera trama es recibida por la interfaz receptora, pero es alterada.

15 **[0174]** Según un ejemplo útil para comprender la invención, la unidad de transformada, por ejemplo, puede ser una primera unidad de transformada, y la unidad de reconstrucción es una primera unidad de reconstrucción. El aparato comprende adicionalmente una segunda unidad de transformada y una segunda unidad de reconstrucción. La segunda unidad de transformada, por ejemplo, puede ser configurada para transformar la información de nivel de ruido del dominio de rastreo al segundo dominio, si una cuarta trama de la pluralidad de tramas no es recibida por la interfaz receptora o si dicha cuarta trama es recibida por la interfaz receptora, pero es alterada. Además, la segunda
20 unidad de reconstrucción, por ejemplo, puede ser configurada para reconstruir una cuarta porción de señal de audio de la señal de audio según la información de nivel de ruido que es representada en el segundo dominio, si dicha cuarta trama de la pluralidad de tramas no es recibida por la interfaz receptora, o si dicha cuarta trama es recibida por la interfaz receptora, pero es alterada.

25 **[0175]** En un ejemplo útil para comprender la invención, la segunda unidad de reconstrucción puede ser configurada, por ejemplo, para reconstruir la cuarta porción de señal de audio según la información de nivel de ruido y según la segunda porción de señal de audio.

30 **[0176]** Según un ejemplo útil para comprender la invención, la segunda unidad de reconstrucción puede ser configurada, por ejemplo, para reconstruir la cuarta porción de señal de audio mediante la atenuación o la amplificación de una señal derivada de la primera o la segunda porción de señal de audio.

[0177] Además, se proporciona un procedimiento para la decodificación de una señal de audio.

35 **[0178]** El procedimiento comprende:

– La recepción de una primera trama de una pluralidad de tramas, comprendiendo dicha primera trama una primera porción de señal de audio de la señal de audio, estando dicha primera porción de señal de audio representada en un primer dominio.

40 – La recepción de una segunda trama de la pluralidad de tramas, comprendiendo dicha segunda trama una segunda porción de señal de audio de la señal de audio.

45 – La transformación de la segunda porción de señal de audio o un valor o una señal derivados de la segunda porción de señal de audio de un segundo dominio a un dominio de rastreo de manera que se obtenga una información de segunda porción de señal, donde el segundo dominio es diferente del primer dominio, donde el dominio de rastreo es diferente del segundo dominio, y donde el dominio de rastreo es igual o diferente respecto del primer dominio.

50 – La determinación de información de nivel de ruido según información de primera porción de señal, que es representada en el dominio de rastreo, y según la información de segunda porción de señal que es representada en el dominio de rastreo, donde la información de primera porción de señal depende de la primera porción de señal de audio. Y:

55 – La reconstrucción de una tercera porción de señal de audio de la señal de audio dependiendo de la información de nivel de ruido que es representada en el dominio de rastreo, si una tercera trama de la pluralidad de tramas no es recibida, o si dicha tercera trama es recibida, pero es alterada.

[0179] Adicionalmente, se proporciona un programa informático para la implementación del procedimiento anteriormente mencionado cuando se ejecuta en un ordenador o en un procesador de señales.

60 **[0180]** Algunos de los ejemplos útiles para comprender la invención proporcionan un parámetro de suavizado de variación de tiempo, de tal manera que las capacidades de rastreo del periodograma suavizado y su varianza están mejor equilibrados, a fin de desarrollar un algoritmo para la compensación de inclinación, y para acelerar el rastreo de ruido, en general.

65 **[0181]** Los ejemplos útiles para comprender la invención se sustentan en el hallazgo de que, con respecto al

desvanecimiento, son de interés los siguientes parámetros: el dominio de desvanecimiento; la velocidad de desvanecimiento, o, más en general, la curva de desvanecimiento; el nivel objetivo del desvanecimiento; la forma espectral objetivo del desvanecimiento; y/o el rastreo del nivel de ruido de fondo. En este contexto, los ejemplos útiles para comprender la invención se sustentan en el hallazgo de que la técnica anterior posee significativas desventajas.

5 **[0182]** Se proporciona un aparato y un procedimiento para el desvanecimiento mejorado de la señal para sistemas de codificación de audio cambiados durante el ocultamiento de errores.

10 **[0183]** Además, se proporciona un programa informático para la implementación del procedimiento anteriormente mencionado cuando se ejecuta en un ordenador o en un procesador de señales.

15 **[0184]** Los ejemplos útiles para comprender la invención representan un desvanecimiento al nivel de ruido de confort. Según los ejemplos útiles para comprender la invención, se representa un rastreo de nivel de ruido de confort común en el dominio de excitación. El nivel de ruido de confort que es dirigido durante la pérdida de paquetes de ráfaga será el mismo, sin consideración del codificador núcleo (predicción lineal excitada por código algebraico (ACELP)/excitación codificada por la transformada (TCX)) en uso, y siempre estará actualizado. No existe técnica previa conocida, donde es necesario un rastreo de nivel de ruido común. Los ejemplos útiles para comprender la invención proporcionan el desvanecimiento de un códec cambiado a una señal de tipo ruido de confort durante las pérdidas de paquetes de ráfagas.

20 **[0185]** Además, los ejemplos útiles para comprender la invención representan que la complejidad general será menor en comparación con dos módulos de rastreo de nivel de ruido independientes, ya que pueden compartirse las funciones (PROM [memoria de solo lectura programable, conforme a sus siglas en inglés]) y la memoria.

25 **[0186]** En ejemplos útiles para comprender la invención, la derivación de nivel en el dominio de excitación (en comparación con la derivación de nivel en el dominio de tiempo) proporciona más mínimas durante la voz activa, ya que parte de la información de la voz es cubierta por los coeficientes de predicción lineal (LP).

30 **[0187]** En el caso de la predicción lineal excitada por código algebraico (ACELP), según los ejemplos útiles para comprender la invención, la derivación de nivel tiene lugar en el dominio de excitación. En el caso de la excitación codificada por la transformada (TCX), en los ejemplos útiles para comprender la invención, el nivel es derivado en el dominio de tiempo, y la ganancia de la síntesis de la codificación predictiva lineal (LPC) y el desénfasis se aplica como un factor de corrección a fin de modelar el nivel de energía en el dominio de excitación. El rastreo del nivel en el dominio de excitación, por ejemplo, antes de la sustitución de ruido de dominio de frecuencia (FDNS), teóricamente también sería posible, pero la compensación de nivel entre el dominio de excitación de la excitación codificada por la transformada (TCX) y el dominio de excitación de la predicción lineal excitada por código algebraico (ACELP) se considera bastante compleja.

40 **[0188]** Ninguna técnica anterior incorpora dicho rastreo de nivel de fondo común en diferentes dominios. Las técnicas de la técnica anterior no tienen dicho rastreo de nivel de ruido de confort común, por ejemplo, en el dominio de excitación, en un sistema de códec cambiado. Por lo tanto, los ejemplos útiles para comprender la invención son convenientes sobre la técnica anterior, ya que, en cuanto a las técnicas de la técnica anterior, el nivel de ruido de confort que es dirigido durante las pérdidas de paquetes de ráfagas puede ser diferente, según el modo de codificación anterior (predicción lineal excitada por código algebraico (ACELP)/excitación codificada por la transformada (TCX)), donde el nivel fue rastreado; ya que en la técnica anterior, el rastreo que es separado para cada modo de codificación causará complejidad computacional adicional e innecesario exceso; y ya que, en la técnica anterior, podría no haber disponible ningún nivel de ruido de confort actualizado en ninguno de los núcleos, debido a cambio reciente a este núcleo.

50 **[0189]** Según algunos ejemplos útiles para comprender la invención, el rastreo de nivel es conducido en el dominio de excitación, pero el desvanecimiento de la excitación codificada por la transformada (TCX) es conducido en el dominio de tiempo. Mediante el desvanecimiento en el dominio de tiempo, se evitan los errores de la TDAC (cancelación del *aliasing* del dominio de tiempo, conforme a sus siglas en inglés), que causarían el *aliasing*. Esto se vuelve de particular interés cuando se ocultan los componentes de señal tonal. Además, la conversión de nivel entre el dominio de excitación de la predicción lineal excitada por código algebraico (ACELP) y el dominio espectral de la transformada de coseno discreta modificada (MDCT) se evita, y en consecuencia, por ejemplo, se ahorran recursos de computación. Debido al cambio entre el dominio de excitación y el dominio de tiempo, se requiere un ajuste de nivel entre el dominio de excitación y el dominio de tiempo. Esto se resuelve por la derivación de la ganancia que sería introducida por la síntesis de la codificación predictiva lineal (LPC) y el pre-énfasis y el uso de esta ganancia como un factor de corrección para la conversión del nivel entre los dos dominios.

65 **[0190]** En contraste, las técnicas de la técnica anterior no conducen el rastreo de nivel en el dominio de excitación y el desvanecimiento de excitación codificada por la transformada (TCX) en el dominio de tiempo. Con respecto a los códecs sobre la base de la transformada del estado de la técnica, se aplica el factor de atenuación o bien en el dominio de excitación (para estrategias de ocultamiento de tipo dominio de tiempo/predicción lineal excitada

por código algebraico (ACELP), véase [3GP09a]) o en el dominio de frecuencia (para estrategias de dominio de frecuencia como la repetición de trama o la sustitución de ruido, véase [LS01]). Una desventaja de la estrategia de la técnica anterior para la aplicación del factor de atenuación en el dominio de frecuencia es que el *aliasing* será causado en la región de superposición y adición en el dominio de tiempo. Este será el caso para tramas adyacentes a las cuales se aplican diferentes factores de atenuación, debido a que el procedimiento de desvanecimiento causa el error de la TDAC (cancelación del *aliasing* del dominio de tiempo, conforme a sus siglas en inglés). Esto es en particular relevante cuando se ocultan los componentes de señal tonal. Los ejemplos anteriormente mencionados útiles para comprender la invención son ventajosos de este modo en comparación con la técnica anterior.

10 **[0191]** Los ejemplos útiles para comprender la invención compensan la influencia del filtro de paso alto sobre la ganancia de la síntesis de la codificación predictiva lineal (LPC). Según ejemplos útiles para comprender la invención, a fin de compensar el cambio de ganancia no deseado del análisis de la codificación predictiva lineal (LPC) y el énfasis causado por la excitación sin voz filtrada por paso alto, se deriva un factor de corrección. Este factor de corrección considera este cambio de ganancia no deseado y modifica el nivel de ruido de confort objetivo en el dominio de excitación, de tal manera que se alcance el nivel objetivo correcto en el dominio de tiempo.

15 **[0192]** En contraste, la técnica anterior, por ejemplo, G.718 [ITU08a], introduce un filtro de paso alto en la vía de la señal de la excitación sin voz, como se representa en la figura 2, si la señal de la última trama buena no fue clasificada como SIN VOZ. Al realizar esto, las técnicas de la técnica anterior causan efectos secundarios no deseados, ya que la ganancia de la subsiguiente síntesis de codificación predictiva lineal (LPC) depende de las características de la señal, que son alteradas por este filtro de paso alto. Debido a que el nivel de fondo es rastreado y aplicado en el dominio de excitación, el algoritmo se sustenta en la ganancia de la síntesis de la codificación predictiva lineal (LPC), que, a su vez, depende de nuevo de las características de la señal de excitación. En otras palabras: la modificación de las características de la señal de la excitación debido a la filtración de paso alto, como es conducida por la técnica anterior, podría conducir a una ganancia modificada (habitualmente, reducida) de la síntesis de la codificación predictiva lineal (LPC). Esto conduce a un nivel de salida erróneo, incluso cuando el nivel de excitación sea correcto.

[0193] Los ejemplos útiles para comprender la invención superan estas desventajas de la técnica anterior.

30 **[0194]** En particular, los ejemplos útiles para comprender la invención representan una forma espectral adaptativa de ruido de confort. En contraste a G.718, al rastrear la forma espectral del ruido de fondo, y mediante la aplicación (desvanecimiento) de esta forma durante las pérdidas de paquetes de ráfagas, la característica de ruido del ruido de fondo anterior coincidirá, de modo que conduzca a una característica de ruido agradable del ruido de confort. Esto evita desajustes inoportunos de la forma espectral que pueden ser introducidos mediante el uso de una envoltura espectral derivada por capacitación sin conexión y/o la forma espectral de las últimas tramas recibidas.

35 **[0195]** Además, se proporciona un aparato para la decodificación de una señal de audio. El aparato comprende una interfaz receptora, donde la interfaz receptora está configurada para recibir una primera trama que comprende una primera porción de señal de audio de la señal de audio, y donde la interfaz receptora está configurada para recibir una segunda trama que comprende una segunda porción de señal de audio de la señal de audio.

40 **[0196]** Además, el aparato comprende una unidad de rastreo de nivel de ruido, donde la unidad de rastreo de nivel de ruido está configurada para determinar la información de nivel de ruido dependiendo de al menos una de la primera porción de señal de audio y la segunda porción de señal de audio (esto significa: dependiendo de la primera porción de señal de audio y/o la segunda porción de señal de audio), donde la información de nivel de ruido es representada en un dominio de rastreo.

45 **[0197]** Adicionalmente, el aparato comprende una primera unidad de reconstrucción para la reconstrucción de, en un primer dominio de reconstrucción, una tercera porción de señal de audio de la señal de audio dependiendo de la información de nivel de ruido, si una tercera trama de la pluralidad de tramas no es recibida por la interfaz receptora o si dicha tercera trama es recibida por la interfaz receptora pero es alterada, donde el primer dominio de reconstrucción es diferente o igual respecto del dominio de rastreo.

50 **[0198]** Además, el aparato comprende una unidad de transformada para la transformación de la información de nivel de ruido, del dominio de rastreo a un segundo dominio de reconstrucción, si una cuarta trama de la pluralidad de tramas no es recibida por la interfaz receptora o si dicha cuarta trama es recibida por la interfaz receptora pero es alterada, donde el segundo dominio de reconstrucción es diferente del dominio de rastreo, y donde el segundo dominio de reconstrucción es diferente del primer dominio de reconstrucción; y

55 **[0199]** Adicionalmente, el aparato comprende una segunda unidad de reconstrucción para la reconstrucción de, en el segundo dominio de reconstrucción, una cuarta porción de señal de audio de la señal de audio según la información de nivel de ruido que es representada en el segundo dominio de reconstrucción, si dicha cuarta trama de la pluralidad de tramas no es recibida por la interfaz receptora o si dicha cuarta trama es recibida por la interfaz receptora, pero es alterada.

60
65

[0200] Según algunos ejemplos útiles para comprender la invención, el dominio de rastreo puede ser, por ejemplo, aquel en el que el dominio de rastreo es un dominio de tiempo, un dominio espectral, un dominio FFT, un dominio de transformada de coseno discreta modificada (MDCT), o un dominio de excitación. El primer dominio de reconstrucción puede ser, por ejemplo, el dominio de tiempo, el dominio espectral, el dominio FFT, el dominio de la transformada de coseno discreta modificada (MDCT), o el dominio de excitación. El segundo dominio de reconstrucción puede ser, por ejemplo, el dominio de tiempo, el dominio espectral, el dominio FFT, el dominio de la transformada de coseno discreta modificada (MDCT), o el dominio de excitación.

[0201] En un ejemplo útil para comprender la invención, el dominio de rastreo, por ejemplo, puede ser el dominio FFT; el primer dominio de reconstrucción puede ser, por ejemplo, el dominio de tiempo; y el segundo dominio de reconstrucción puede ser, por ejemplo, el dominio de excitación.

[0202] En otro ejemplo útil para comprender la invención, el dominio de rastreo puede ser, por ejemplo, el dominio de tiempo, el primer dominio de reconstrucción puede ser, por ejemplo, el dominio de tiempo, y el segundo dominio de reconstrucción puede ser, por ejemplo, el dominio de excitación.

[0203] Según un ejemplo útil para comprender la invención, dicha primera porción de señal de audio puede ser representada, por ejemplo, en un primer dominio de entrada, y dicha segunda porción de señal de audio puede ser representada, por ejemplo, en un segundo dominio de entrada. La unidad de transformada, por ejemplo, puede ser una segunda unidad de transformada. El aparato, por ejemplo, puede comprender adicionalmente una primera unidad de transformada para la transformación de la segunda porción de señal de audio o un valor o una señal derivados de la segunda porción de señal de audio, del segundo dominio de entrada al dominio de rastreo, de manera que se obtenga una información de segunda porción de señal. La unidad de rastreo de nivel de ruido puede estar configurada, por ejemplo, para recibir una información de primera porción de señal que es representada en el dominio de rastreo, donde la información de primera porción de señal depende de la primera porción de señal de audio, donde la unidad de rastreo de nivel de ruido está configurada para recibir la segunda porción de señal que se representa en el dominio de rastreo, y donde la unidad de rastreo de nivel de ruido está configurada para la determinación de la información de nivel de ruido dependiendo de la información de primera porción de señal que es representada en el dominio de rastreo y dependiendo de la información de segunda porción de señal que es representada en el dominio de rastreo.

[0204] Según un ejemplo útil para comprender la invención, el primer dominio de entrada puede ser, por ejemplo, el dominio de excitación, y el segundo dominio de entrada puede ser, por ejemplo, el dominio de la transformada de coseno discreta modificada (MDCT).

[0205] En otro ejemplo útil para comprender la invención, el primer dominio de entrada puede ser, por ejemplo, el dominio de la transformada de coseno discreta modificada (MDCT), y donde el segundo dominio de entrada puede ser, por ejemplo, el dominio de la transformada de coseno discreta modificada (MDCT).

[0206] Según un ejemplo útil para comprender la invención, la primera unidad de reconstrucción puede estar configurada, por ejemplo, para reconstruir la tercera porción de señal de audio mediante la conducción de un primer desvanecimiento a un espectro de tipo ruido. La segunda unidad de reconstrucción puede estar configurada, por ejemplo, para reconstruir la cuarta porción de señal de audio mediante la conducción de un segundo desvanecimiento a un espectro de tipo ruido y/o un segundo desvanecimiento de una ganancia de predicción a largo plazo (LTP). Además, la primera unidad de reconstrucción y la segunda unidad de reconstrucción pueden estar configuradas, por ejemplo, para conducir el primer desvanecimiento y el segundo desvanecimiento a un espectro de tipo ruido y/o un segundo desvanecimiento de una ganancia de predicción a largo plazo (LTP) con la misma velocidad de desvanecimiento.

[0207] En un ejemplo útil para comprender la invención, el aparato puede comprender, por ejemplo, adicionalmente una primera unidad de agregación para la determinación de un primer valor agregado dependiendo de la primera porción de señal de audio. Además, el aparato adicionalmente puede comprender, por ejemplo, una segunda unidad de agregación para la determinación, dependiendo de la segunda porción de señal de audio, de un segundo valor agregado como el valor derivado de la segunda porción de señal de audio. La unidad de rastreo de nivel de ruido puede estar configurada, por ejemplo, para recibir el primer valor agregado como la información de primera porción de señal que es representada en el dominio de rastreo, donde la unidad de rastreo de nivel de ruido puede estar configurada, por ejemplo, para recibir el segundo valor agregado como la información de segunda porción de señal que es representada en el dominio de rastreo, y donde la unidad de rastreo de nivel de ruido está configurada para determinar la información de nivel de ruido dependiendo del primer valor agregado que es representado en el dominio de rastreo y dependiendo del segundo valor agregado que es representado en el dominio de rastreo.

[0208] Según un ejemplo útil para comprender la invención, la primera unidad de agregación puede estar configurada, por ejemplo, para determinar el primer valor agregado de tal manera que el primer valor agregado indica una media cuadrática de la primera porción de señal de audio o de una señal derivada de la primera porción de señal de audio. La segunda unidad de agregación está configurada para determinar el segundo valor agregado de tal manera que el segundo valor agregado indica una media cuadrática de la segunda porción de señal de audio o de una señal

derivada de la segunda porción de señal de audio.

[0209] En un ejemplo útil para comprender la invención, la primera unidad de transformada puede ser configurada, por ejemplo, para transformar el valor derivado de la segunda porción de señal de audio, del segundo dominio de entrada al dominio de rastreo mediante la aplicación de un valor de ganancia sobre el valor derivado de la segunda porción de señal de audio.

[0210] Según un ejemplo útil para comprender la invención, el valor de ganancia puede indicar, por ejemplo, una ganancia introducida por la síntesis de codificación predictiva lineal, o donde el valor de ganancia indica una ganancia introducida por la síntesis de codificación predictiva lineal y el desénfasis.

[0211] En un ejemplo útil para comprender la invención, la unidad de rastreo de nivel de ruido puede estar configurada, por ejemplo, para determinar la información de nivel de ruido mediante la aplicación de una estrategia de estadística mínima.

[0212] Según un ejemplo útil para comprender la invención, la unidad de rastreo de nivel de ruido puede estar configurada, por ejemplo, para determinar un nivel de ruido de confort como la información de nivel de ruido. La unidad de reconstrucción puede estar configurada, por ejemplo, para reconstruir la tercera porción de señal de audio dependiendo de la información de nivel de ruido, si dicha tercera trama de la pluralidad de tramas no es recibida por la interfaz receptora o si dicha tercera trama es recibida por la interfaz receptora, pero es alterada.

[0213] En un ejemplo útil para comprender la invención, la unidad de rastreo de nivel de ruido puede estar configurada, por ejemplo, para determinar un nivel de ruido de confort como la información de nivel de ruido derivada de un espectro de nivel de ruido, donde dicho espectro de nivel de ruido se obtiene mediante la aplicación de la estrategia de estadística mínima. La unidad de reconstrucción puede estar configurada, por ejemplo, para reconstruir la tercera porción de señal de audio dependiendo de una pluralidad de coeficientes predictivos lineales, si dicha tercera trama de la pluralidad de tramas no es recibida por la interfaz receptora, o si dicha tercera trama es recibida por la interfaz receptora, pero es alterada.

[0214] Según un ejemplo útil para comprender la invención, la primera unidad de reconstrucción puede estar configurada, por ejemplo, para reconstruir la tercera porción de señal de audio dependiendo de la información de nivel de ruido y dependiendo de la primera porción de señal de audio, si dicha tercera trama de la pluralidad de tramas no es recibida por la interfaz receptora o si dicha tercera trama es recibida por la interfaz receptora, pero es alterada.

[0215] En un ejemplo útil para comprender la invención, la primera unidad de reconstrucción puede estar configurada, por ejemplo, para reconstruir la tercera porción de señal de audio mediante la atenuación o amplificación de la primera porción de señal de audio.

[0216] Según un ejemplo útil para comprender la invención, la segunda unidad de reconstrucción puede estar configurada, por ejemplo, para reconstruir la cuarta porción de señal de audio según la información de nivel de ruido y según la segunda porción de señal de audio.

[0217] En un ejemplo útil para comprender la invención, la segunda unidad de reconstrucción puede estar configurada, por ejemplo, para reconstruir la cuarta porción de señal de audio mediante la atenuación o amplificación de la segunda porción de señal de audio.

[0218] Según un ejemplo útil para comprender la invención, el aparato puede comprender, por ejemplo, adicionalmente una unidad de predicción a largo plazo que comprende un búfer de retardo, donde la unidad de predicción a largo plazo puede estar configurada, por ejemplo, para generar una señal procesada dependiendo de la primera o la segunda porción de señal de audio, dependiendo de una entrada de búfer de retardo que está almacenada en el búfer de retardo y dependiendo de una ganancia de predicción a largo plazo, y donde la unidad de predicción a largo plazo está configurada para desvanecer la ganancia de predicción a largo plazo hacia cero, si dicha tercera trama de la pluralidad de tramas no es recibida por la interfaz receptora o si dicha tercera trama es recibida por la interfaz receptora pero es alterada.

[0219] En un ejemplo útil para comprender la invención, la unidad de predicción a largo plazo puede estar configurada, por ejemplo, para desvanecer la ganancia de predicción a largo plazo hacia cero, donde una velocidad con la cual la ganancia de predicción a largo plazo es desvanecida a cero depende de un factor de desvanecimiento.

[0220] En un ejemplo útil para comprender la invención, la unidad de predicción a largo plazo puede estar configurada, por ejemplo, para actualizar la entrada del búfer de retardo mediante el almacenamiento de la señal procesada generada en el búfer de retardo, si dicha tercera trama de la pluralidad de tramas no es recibida por la interfaz receptora o si dicha tercera trama es recibida por la interfaz receptora, pero es alterada.

[0221] Además, se proporciona un procedimiento para la decodificación de una señal de audio. El

procedimiento comprende:

– La recepción de una primera trama que comprende una primera porción de señal de audio de la señal de audio, y la recepción de una segunda trama que comprende una segunda porción de señal de audio de la señal de audio.

5

– La determinación de información de nivel de ruido dependiendo de al menos una de la primera porción de señal de audio y la segunda porción de señal de audio, donde la información de nivel de ruido es representada en un dominio de rastreo.

10 – La reconstrucción de, en un primer dominio de reconstrucción, una tercera porción de señal de audio de la señal de audio dependiendo de la información de nivel de ruido, si una tercera trama de la pluralidad de tramas no es recibida o si dicha tercera trama es recibida, pero es alterada, donde el primer dominio de reconstrucción es diferente o igual respecto del dominio de rastreo.

15 – La transformación de la información de nivel de ruido, del dominio de rastreo a un segundo dominio de reconstrucción, si una cuarta trama de la pluralidad de tramas no es recibida o si dicha cuarta trama es recibida, pero es alterada, donde el segundo dominio de reconstrucción es diferente del dominio de rastreo, y donde el segundo dominio de reconstrucción es diferente del primer dominio de reconstrucción. Y:

20 – La reconstrucción, en el segundo dominio de reconstrucción, de una cuarta porción de señal de audio de la señal de audio dependiendo de la información de nivel de ruido que es representada en el segundo dominio de reconstrucción, si dicha cuarta trama de la pluralidad de tramas no es recibida o si dicha cuarta trama es recibida, pero es alterada.

[0222] Además, se proporciona un programa informático para la implementación del procedimiento anteriormente mencionado, cuando se ejecuta en un ordenador o en un procesador de señales.

[0223] Además, se proporciona un aparato para la decodificación de una señal de audio codificada de manera que se obtenga una señal de audio reconstruida. El aparato comprende una interfaz receptora para la recepción de una o más tramas, un generador de coeficientes y un reconstructor de señal. El generador de coeficientes está configurado para determinar, si una trama actual de una o más tramas es recibida por la interfaz receptora y si la trama actual recibida por la interfaz receptora no está alterada, estando uno o más primeros coeficientes de señal de audio, comprendidos por la trama actual, donde dichos uno o más primeros coeficientes de señal de audio indican una característica de la señal de audio codificada, e indicando uno o más coeficientes de ruido un ruido de fondo de la señal de audio codificada. Además, el generador de coeficientes está configurado para generar uno o más segundos coeficientes de señal de audio, dependiendo de los uno o más primeros coeficientes de señal de audio y dependiendo de los uno o más coeficientes de ruido, si la trama actual no es recibida por la interfaz receptora o si la trama actual recibida por la interfaz receptora es alterada. El reconstructor de señal de audio está configurado para reconstruir una primera porción de la señal de audio reconstruida dependiendo de uno o más primeros coeficientes de señal de audio, si la trama actual es recibida por la interfaz receptora y si la trama actual recibida por la interfaz receptora no está alterada. Además, el reconstructor de señal de audio está configurado para reconstruir una segunda porción de la señal de audio reconstruida dependiendo de uno o más segundos coeficientes de señal de audio, si la trama actual no es recibida por la interfaz receptora, o si la trama actual recibida por la interfaz receptora está alterada.

[0224] En algunos ejemplos útiles para comprender la invención, uno o más primeros coeficientes de señal de audio pueden ser, por ejemplo, uno o más coeficientes de filtro predictivo lineal de la señal de audio codificada. En algunos ejemplos útiles para comprender la invención, uno o más primeros coeficientes de señal de audio pueden ser, por ejemplo, uno o más coeficientes de filtro predictivo lineal de la señal de audio codificada.

[0225] Según un ejemplo útil para comprender la invención, uno o más coeficientes de ruido pueden ser, por ejemplo, uno o más coeficientes de filtro predictivo lineal que indican el ruido de fondo de la señal de audio codificada. En un ejemplo útil para comprender la invención, uno o más coeficientes de filtro predictivo lineal, por ejemplo, pueden representar una forma espectral del ruido de fondo.

[0226] En un ejemplo útil para comprender la invención, el generador de coeficientes, por ejemplo, puede ser configurado para determinar una o más segundas porciones de señal de audio de tal manera que una o más segundas porciones de señal de audio sean uno o más coeficientes de filtro predictivo lineal de la señal de audio reconstruida, o de tal manera que uno o más primeros coeficientes de señal de audio sean uno o más pares espectrales de impedancia y admitancia de la señal de audio reconstruida.

[0227] Según un ejemplo útil para comprender la invención, el generador de coeficientes, por ejemplo, puede ser configurado para generar uno o más segundos coeficientes de señal de audio mediante la aplicación de la fórmula:

$$f_{\text{actual}}[i] = \alpha \cdot f_{\text{último}}[i] + (1 - \alpha) \cdot p_{\text{media}}[i] \quad i = 0 \dots 16$$

donde $f_{actual}[i]$ indica uno de uno o más segundos coeficientes de señal de audio, donde $f_{ultimo}[i]$ indica uno de uno o más primeros coeficientes de señal de audio, donde $p_{tmedia}[i]$ es uno de uno o más coeficientes de ruido, donde α es un número real con $0 \leq \alpha \leq 1$, y donde i es un índice. En un ejemplo útil para comprender la invención, $0 < \alpha < 1$.

5

[0228] Según un ejemplo útil para comprender la invención, $f_{ultimo}[i]$ indica un coeficiente de filtro predictivo lineal de la señal de audio codificada, y donde $f_{actual}[i]$ indica un coeficiente de filtro predictivo lineal de la señal de audio reconstruida.

10 **[0229]** En un ejemplo útil para comprender la invención, $p_{tmedia}[i]$, por ejemplo, puede indicar el ruido de fondo de la señal de audio codificada.

[0230] En un ejemplo útil para comprender la invención, el generador de coeficientes puede estar configurado, por ejemplo, para determinar, si la trama actual de una o más tramas es recibida por la interfaz receptora y si la trama actual recibida por la interfaz receptora no está alterada, uno o más coeficientes de ruido mediante la determinación de un espectro de ruido de la señal de audio codificada.

15 **[0231]** Según un ejemplo útil para comprender la invención, el generador de coeficientes puede estar configurado, por ejemplo, para determinar coeficientes de LPC que representan ruido de fondo mediante el uso de una estrategia de estadística mínima sobre el espectro de señal para determinar un espectro de ruido de fondo y mediante el cálculo de los coeficientes LPC que representan la forma de ruido de fondo a partir del espectro de ruido de fondo.

20 **[0232]** Además, se proporciona un procedimiento para la decodificación de una señal de audio codificada a fin de obtener una señal de audio reconstruida. El procedimiento comprende:

- La recepción de una o más tramas.
- La determinación, si una trama actual de una o más tramas es recibida y si la trama actual recibida no está alterada, estando uno o más primeros coeficientes de señal de audio comprendidos por la trama actual, donde dichos uno o más primeros coeficientes de señal de audio indican una característica de la señal de audio codificada, e indicando uno o más coeficientes de ruido un ruido de fondo de la señal de audio codificada.
- La generación de uno o más segundos coeficientes de señal de audio, dependiendo de uno o más primeros coeficientes de señal de audio y dependiendo de uno o más coeficientes de ruido, si la trama actual no es recibida o si la trama actual recibida está alterada.
- La reconstrucción de una primera porción de la señal de audio reconstruida dependiendo de uno o más primeros coeficientes de señal de audio, si la trama actual es recibida y si la trama actual recibida no está alterada. Y:
- La reconstrucción de una segunda porción de la señal de audio reconstruida dependiendo de uno o más segundos coeficientes de señal de audio, si la trama actual no es recibida o si la trama actual recibida está alterada.

30 **[0233]** Además, se proporciona un programa informático para la implementación del procedimiento anteriormente descrito cuando se ejecuta en un ordenador o en un procesador de señales.

[0234] El tener medios comunes para rastrear y aplicar la forma espectral del ruido de confort durante el desvanecimiento posee varias ventajas. El rastreo y la aplicación de la forma espectral de modo que puedan realizarse de manera similar para ambos códecs núcleo permiten una simple estrategia común. La transformada superpuesta de energía restringida (CELT) enseña solo el rastreo de energías a modo de banda, en el dominio espectral, y la formación a modo de banda, de la forma espectral en el dominio espectral, lo que no es posible para el núcleo CELP.

35 **[0235]** En contraste, en la técnica anterior, la forma espectral del ruido de confort introducido durante las pérdidas de ráfagas es o bien completamente estática, o en parte estática y en parte adaptativa a la media a corto plazo de la forma espectral (como se representa en G.718 [ITU08a]), y habitualmente no coincidirá con el ruido de fondo en la señal antes de la pérdida de paquete. Estas características de no coincidencia del ruido de confort podrían ser molestas. Según la técnica anterior, puede emplearse una forma de ruido de fondo capacitada sin conexión (estática) que pueda ser placentera en términos de sonido para señales particulares, pero menos placentera para otros, por ejemplo, sonidos de ruido de automóviles totalmente diferentes al ruido de oficina.

60

[0236] Además, en la técnica anterior, puede emplearse una adaptación a la media a corto plazo de la forma espectral de las tramas previamente recibidas, que podría llevar las características de la señal más cerca de la señal recibida antes, si bien no necesariamente de las características del ruido de fondo. En la técnica anterior, el rastreo de la forma espectral a modo de banda en el dominio espectral (como se representa en la transformada superpuesta de energía restringida (CELT) [IET12]) no es aplicable para un códec cambiado mediante el uso no solo de un núcleo

65

sobre la base del dominio de la transformada de coseno discreta modificada (MDCT) (excitación codificada por la transformada (TCX)), sino, además, de un núcleo sobre la base de la predicción lineal excitada por código algebraico (ACELP). Los ejemplos útiles para comprender la invención anteriormente mencionados son de este modo ventajosos en comparación con la técnica anterior.

5

[0237] Además, se proporciona un aparato para la decodificación de una señal de audio codificada de manera que se obtenga una señal de audio reconstruida. El aparato comprende una interfaz receptora para recibir una o más tramas que comprenden información en una pluralidad de muestras de señales de audio de un espectro de señal de audio de la señal de audio codificada, y un procesador para generar la señal de audio reconstruida. El procesador se configura para generar la señal de audio reconstruida al desvanecer un espectro modificado hacia un espectro objetivo, si una trama actual no se recibe por la interfaz receptora o si la trama actual se recibe por la interfaz receptora pero está alterada, donde el espectro modificado comprende una pluralidad de muestras de señales modificadas, donde, para cada una de las muestras de señales modificadas del espectro modificado, un valor absoluto de la muestra de señal modificada es igual a un valor absoluto de una de las muestras de señales de audio del espectro de señal de audio. Además, el procesador se configura para no desvanecer el espectro modificado al espectro objetivo, si la trama actual de las una o más tramas se recibe por la interfaz receptora y si la trama actual que se recibe por la interfaz receptora no está alterada.

10

15

[0238] Según un ejemplo útil para comprender la invención, el espectro objetivo puede ser, por ejemplo, un espectro tipo ruido.

20

[0239] En un ejemplo útil para comprender la invención, el espectro tipo ruido puede representar, por ejemplo, ruido blanco.

[0240] Según un ejemplo útil para comprender la invención, el espectro tipo ruido se puede formar, por ejemplo.

25

[0241] En un ejemplo útil para comprender la invención, la forma del espectro tipo ruido puede depender, por ejemplo, de un espectro de señal de audio de una señal previamente recibida.

[0242] Según un ejemplo útil para comprender la invención, el espectro tipo ruido se puede formar, por ejemplo, dependiendo de la forma del espectro de señal de audio.

30

[0243] En un ejemplo útil para comprender la invención, el procesador puede emplear, por ejemplo, un factor de inclinación para formar el espectro tipo ruido.

35

[0244] Según un ejemplo útil para comprender la invención, el procesador puede emplear, por ejemplo, la fórmula

$$\text{ruido formado } [i] = \text{ruido} * \text{energía} (\text{factor de inclinación}, i/N)$$

40

donde N indica el número de muestras, donde i es un índice, donde $0 < i < N$, con $\text{factor_inclinación} > 0$, y donde energía es una función de energía.

energía (x,y) indica x^y

energía (factor_inclinación, i/N) indica $\text{factor_inclinación}^{1/N}$

45

Si el factor de inclinación es más pequeño que 1 esto significa la atenuación con incremento de i. Si el factor de inclinación es mayor que 1 significa amplificación con i creciente.

[0245] Según otro ejemplo útil para comprender la invención, el procesador puede emplear, por ejemplo, la fórmula

50

$$\text{ruido_formado } [i] = \text{ruido} * (1+i / (N-1) * (\text{factor_inclinación}-1))$$

donde N indica el número de muestras, donde i es un índice, donde $0 < i < N$, con $\text{factor_inclinación} > 0$.

55

[0246] Si el $\text{factor_inclinación}$ es más pequeño que 1, esto significa la atenuación con i creciente. Si el $\text{factor_inclinación}$ es mayor que 1 significa amplificación con i creciente.

[0247] Según un ejemplo útil para comprender la invención, el procesador se puede configurar, por ejemplo, para generar el espectro modificado, al cambiar un signo de una o más de las muestras de señales de audio del espectro de señal de audio, si la trama actual no se recibe por la interfaz receptora o si la trama actual que se recibe por la interfaz receptora está alterada.

60

[0248] En un ejemplo útil para comprender la invención, cada una de las muestras de señales de audio del

espectro de señal de audio se puede representar, por ejemplo, por un número real pero no por un número imaginario.

[0249] Según un ejemplo útil para comprender la invención, las muestras de señales de audio del espectro de señal de audio se pueden representar, por ejemplo, en un dominio de transformada por coseno discreto modificado.

[0250] En otro ejemplo útil para comprender la invención, las muestras de señales de audio del espectro de señal de audio se pueden representar por ejemplo en un dominio de transformada por seno discreto modificado.

[0251] Según un ejemplo útil para comprender la invención, el procesador se puede configurar, por ejemplo, para generar el espectro modificado al emplear una función de signo aleatorio que produce de manera aleatoria o pseudo aleatoria ya sea un primero o un segundo valor.

[0252] En un ejemplo útil para comprender la invención, el procesador se puede configurar, por ejemplo, para desvanecer el espectro modificado al espectro objetivo al disminuir subsiguientemente un factor de atenuación.

[0253] Según un ejemplo útil para comprender la invención, el procesador se puede configurar, por ejemplo, para desvanecer el espectro modificado al espectro objetivo al incrementar subsiguientemente un factor de atenuación.

[0254] En un ejemplo útil para comprender la invención, si la trama actual no se recibe por la interfaz receptora o si la trama actual que se recibe por la interfaz receptora está alterada, el procesador se puede configurar, por ejemplo, para generar la señal de audio reconstruida al emplear la fórmula:

$$X [i] = (1 - \text{cum amortiguamiento}) * \text{ruido} [i] + \text{cum amortiguamiento} * \text{signo aleatorio} * x \text{ antiguo} [i]$$

donde i es un índice, donde $x [i]$ indica una muestra de la señal de audio reconstruida, donde cum amortiguamiento es un factor de atenuación, donde $x = \text{antiguo} [i]$ indica una de las muestras de señales de audio del espectro de señal de audio de la señal de audio codificada, donde signo aleatorio () devuelve 1 o -1, donde ruido es un vector aleatorio que indica el espectro objetivo.

[0255] En un ejemplo útil para comprender la invención, dicho ruido de vector aleatorio se puede escalar, por ejemplo, de tal manera que su media cuadrática sea similar a la media cuadrática del espectro de la señal de audio codificada que está comprendida por una de las tramas que se recibe por último por la interfaz receptora.

[0256] Según un ejemplo útil para comprender la invención, el procesador se puede configurar, por ejemplo, para generar la señal de audio reconstruida, empleando un vector aleatorio que se escala de tal manera que su media cuadrática es similar a la media del espectro de la señal de audio codificada que está comprendida por una de las tramas que se recibe por último por la interfaz receptora.

[0257] Además, se proporciona un procedimiento para decodificar una señal de audio codificada para obtener una señal de audio reconstruida. El procedimiento comprende:

- recibir una o más tramas que comprenden información de una pluralidad de muestras de señales de audio de un espectro de señal de audio de la señal de audio codificada, y;

- generar la señal de audio reconstruida.

[0258] La generación de la señal de audio reconstruida se lleva a cabo al desvanecer un espectro modificado a un espectro objetivo, si no recibe una trama actual o si la trama actual se recibe pero está alterada, donde el espectro modificado comprende una pluralidad de muestras de señal modificada, donde, para cada una de las muestras de señal modificadas del espectro modificado, un valor absoluto de la muestra de señal modificada es igual a un valor absoluto de una de las muestras de señal de audio del espectro de señal de audio. El espectro modificado no se desvanece a un espectro de ruido blanco, si la trama actual de la una o más tramas se recibe y si la trama actual que se recibe no está alterada.

[0259] Además, se proporciona un programa informático para implementar el procedimiento descrito anteriormente que se ejecuta en un ordenador o procesador de señales.

[0260] Ejemplos útiles para comprender la invención logran un espectro MDCT desvanecido a ruido blanco antes de la aplicación FDNS (FDNS = sustitución de ruido en dominio de frecuencia).

[0261] Según la técnica anterior, en los códecs basados en ACELP, el libro de códigos innovador se reemplaza con un vector aleatorio (por ejemplo, con ruido). En ejemplos útiles para comprender la invención, el planteamiento ACELP, que consiste en reemplazar el código de libros innovador con un vector aleatorio (por ejemplo, con ruido) se

adopta a la estructura del decodificador TCX. Aquí, el equivalente del libro de códigos innovador es el espectro MDCT usualmente recibido dentro de la corriente de bits y se alimenta al FDNS.

[0262] La estrategia de ocultación MDCT clásica será repetir simplemente este espectro como esta o aplicar un cierto proceso de aleatorización, que prolonga básicamente la forma espectral de la última trama recibida [LS01]. Esto tiene la desventaja de que la forma espectral a corto plazo se prolonga, conduciendo frecuentemente a un sonido metálico repetitivo que no es tipo ruido de fondo, y de esta manera no se puede usar como ruido agradable.

[0263] Usando el procedimiento propuesto, la formación espectral a corto plazo se realiza por el FDNS y la TCX LTP, la formación espectral en ejecución a largo plazo se realiza por el FDNS únicamente. La formación por el FDNS se desvanece de la forma espectral a corto plazo a la forma espectral a largo plazo trazada del ruido de fondo, y la TCX LTP se desvanece a cero.

[0264] El desvanecimiento de los coeficientes de FDNS a los coeficientes de ruido de fondos trazados conduce a tener una transición suave entre la última envolvente espectral buena y la envolvente espectral de fondo que se buscará en la ejecución larga, a fin de lograr un ruido de fondo agradable en el caso de pérdidas prolongadas de tramas por ráfaga.

[0265] En contraste, según el estado de la técnica para los códecs basados en transformadas, el ocultamiento tipo ruido se lleva a cabo por la repetición de tramas o sustitución de ruido en el dominio de frecuencia [LS01]. En la técnica anterior, la sustitución de ruido se realiza usualmente por el mezclado de signos de las bandejas espectrales. Si en la TCX (dominio de frecuencia) de la técnica anterior se usa la mezcla de signos durante la ocultación, los últimos coeficientes MDCT recibidos se reutilizan y cada signo se aleatoriza antes de que se transforme de manera inversa el espectro al dominio de tiempo. La desventaja de este procedimiento de la técnica anterior es que, para las tramas consecutivamente perdidas, se usa el mismo espectro una y otra vez, solo con diferentes aleatorizaciones de signo y atenuación global. Cuando se busca la envolvente espectral con el paso del tiempo en una rejilla de tiempo gruesa, se puede ver que la envolvente es aproximadamente constante durante las pérdidas consecutivas de tramas debido a que las energías de bandas se mantienen constantes con relación una a otra dentro de una trama y solo se atenúan globalmente. En el sistema de codificación usado, según la técnica anterior, los valores espectrales se procesan mediante el uso de FDNS, a fin de restaurar el espectro original. Esto significa, que si uno quiere desvanecer el espectro MDCT a una cierta envolvente espectral (mediante el uso de coeficientes de FDNS, por ejemplo, describiendo el ruido de fondo actual), el resultado no solo es dependiente de los coeficientes de FDNS, sino también dependiente del espectro previamente decodificado que se mezcló con signo. Los ejemplos útiles para comprender la invención mencionados anteriormente superan estas desventajas de la técnica anterior.

[0266] Los ejemplos útiles para comprender la invención se basan en el hallazgo de que es necesario desvanecer el espectro usado para el mezclado de signos a ruido blanco, antes de alimentarlo en el procesamiento de FDNS. De otro modo el espectro producido nunca corresponderá a la envolvente buscada utilizada para el procesamiento FDNS.

[0267] En ejemplos útiles para comprender la invención, se usa la misma velocidad de desvanecimiento para el desvanecimiento de ganancia por LTP como para el desvanecimiento de ruido blanco.

[0268] En lo que sigue, se describen ejemplos útiles para comprender la invención de la presente invención en más detalle con referencia a las figuras, en los cuales:

La figura 1a ilustra un aparato para la decodificación de una señal de audio según un ejemplo útil para comprender la invención,

La figura 1b ilustra un aparato para la decodificación de una señal de audio según otro ejemplo útil para comprender la invención,

La figura 1c ilustra un aparato para la decodificación de una señal de audio según otro ejemplo útil para comprender la invención, donde el aparato comprende adicionalmente una primera y una segunda unidades de agregación,

La figura 1d ilustra un aparato para la decodificación de una señal de audio según un ejemplo adicional útil para comprender la invención, donde el aparato comprende además una unidad de predicción a largo plazo que comprende un búfer de retardo,

La figura 2 ilustra la estructura del decodificador de G.718,

La figura 3 representa un escenario donde el factor de desvanecimiento de G.722 depende de la información de clase,

La figura 4 muestra una estrategia para la predicción de amplitud mediante el uso de la regresión lineal,

La figura 5 ilustra el comportamiento de pérdida de ráfaga de la transformada superpuesta de energía restringida (CELT)),

La figura 6 muestra un rastreo del nivel de ruido de fondo según un ejemplo útil para comprender la invención en el 5 decodificador durante un modo de operación libre de error,

La figura 7 ilustra la derivación de ganancia de la síntesis de codificación predictiva lineal (LPC) y el desénfasis según un ejemplo útil para comprender la invención,

10 La figura 8 representa la aplicación de nivel de ruido de confort durante la pérdida de paquete según un ejemplo útil para comprender la invención,

La figura 9 ilustra la compensación de ganancia de paso alto avanzada durante el ocultamiento de predicción lineal excitada por código algebraico (ACELP) según un ejemplo útil para comprender la invención,

15 La figura 10 representa el desacoplamiento del circuito de retroalimentación de la predicción a largo plazo (LTP) durante el ocultamiento según un ejemplo útil para comprender la invención,

La figura 11 ilustra un aparato para la decodificación de una señal de audio codificada de manera que se obtenga una 20 señal de audio reconstruida según un ejemplo útil para comprender la invención,

La figura 12 muestra un aparato para la decodificación de una señal de audio codificada de manera que se obtenga una señal de audio reconstruida según otro ejemplo útil para comprender la invención, y

25 La figura 13 ilustra un aparato para la decodificación de una señal de audio codificada de manera que se obtenga una señal de audio reconstruida según un ejemplo adicional útil para comprender la invención, y

La figura 14 ilustra un aparato para la decodificación de una señal de audio codificada de manera que se obtenga una 30 señal de audio reconstruida según otro ejemplo útil para comprender la invención.

[0269] La figura 1a ilustra un aparato para la decodificación de una señal de audio según un ejemplo útil para comprender la invención.

[0270] El aparato comprende una interfaz receptora 110. La interfaz receptora está configurada para recibir 35 una pluralidad de tramas, donde la interfaz receptora 110 está configurada para recibir una primera trama de la pluralidad de tramas, comprendiendo dicha primera trama una primera porción de señal de audio de la señal de audio, donde dicha primera porción de señal de audio está representada en un primer dominio.

[0271] Además, la interfaz receptora 110 está configurada para recibir una segunda trama de la pluralidad de 40 tramas, comprendiendo dicha segunda trama una segunda porción de señal de audio de la señal de audio.

[0272] Además, el aparato comprende una unidad de transformada 120 para la transformación de la segunda 45 porción de señal de audio o un valor o una señal derivados de la segunda porción de señal de audio de un segundo dominio a un dominio de rastreo de manera que se obtenga una información de segunda porción de señal, donde el segundo dominio es diferente del primer dominio, donde el dominio de rastreo es diferente del segundo dominio, y donde el dominio de rastreo es igual o diferente respecto del primer dominio.

[0273] Adicionalmente, el aparato comprende una unidad de rastreo de nivel de ruido 130, donde la unidad de 50 rastreo de nivel de ruido está configurada para recibir una información de primera porción de señal que está representada en el dominio de rastreo, donde la información de primera porción de señal depende de la primera porción de señal de audio, donde la unidad de rastreo de nivel de ruido está configurada para recibir la segunda porción de señal que es representada en el dominio de rastreo, y donde la unidad de rastreo de nivel de ruido está configurada para determinar la información de nivel de ruido dependiendo de la información de primera porción de 55 señal que es representada en el dominio de rastreo y dependiendo de la información de segunda porción de señal que es representada en el dominio de rastreo.

[0274] Además, el aparato comprende una unidad de reconstrucción para la reconstrucción de una tercera 60 porción de señal de audio de la señal de audio según la información de nivel de ruido, si una tercera trama de la pluralidad de tramas no es recibida por la interfaz receptora, pero es alterada.

[0275] Con respecto a la primera y/o la segunda porción de señal de audio, por ejemplo, la primera y/o la 65 segunda porción de señal de audio puede alimentarse, por ejemplo, en una o más unidades de procesamiento (no expuestas) para la generación de una o más señales de altavoz para uno o más altavoces, de manera que la información de sonido recibida compuesta por la primera y/o la segunda porción de señal de audio pueda ser reproducida de nuevo.

[0276] Además, sin embargo, la primera y la segunda porciones de señal de audio se usan también para el ocultamiento, por ejemplo, en el caso de que las subsiguientes tramas no lleguen al receptor, o en el caso de que las subsiguientes tramas sean erróneas.

5 **[0277]** Entra otras cosas, la presente invención se sustenta en el hallazgo de que el rastreo de nivel de ruido debe conducirse en un dominio común, referido en esta solicitud como “dominio de rastreo”. El dominio de rastreo, por ejemplo, puede ser un dominio de excitación, por ejemplo, el dominio en el cual la señal es representada por las codificaciones predictivas lineales (LPC) o por los pares espectrales de impedancia y admitancia (ISP, conforme a sus siglas en inglés) como se describe en la multitasa adaptativa (AMR)–banda ancha (WB) y en la multitasa adaptativa
10 (AMR)–banda ancha (WB)+ (véase [3GP12a], [3GP12b], [3GP09a], [3GP09b], [3GP09c]). El rastreo del nivel de ruido en un solo dominio tiene, entre otras cosas, la ventaja de que se evitan los efectos del *aliasing* cuando la señal cambia entre una primera representación en un primer dominio y una segunda representación en un segundo dominio (por ejemplo, cuando la representación de señal cambia de la predicción lineal excitada por código algebraico (ACELP) a la excitación codificada por la transformada (TCX) o viceversa).

15 **[0278]** Con respecto a la unidad de transformada 120, lo que se transforma es o bien la segunda porción de señal de audio en sí misma, o una señal derivada de la segunda porción de señal de audio (por ejemplo, la segunda porción de señal de audio ha sido procesada a fin de obtener la señal derivada), o un valor derivado de la segunda porción de señal de audio (por ejemplo, la segunda porción de señal de audio ha sido procesada a fin de obtener el
20 valor derivado).

[0279] Con respecto a la primera porción de señal de audio, en algunos ejemplos útiles para comprender la invención, la primera porción de señal de audio puede ser procesada y/o transformada al dominio de rastreo.

25 **[0280]** En otros ejemplos útiles para comprender la invención, sin embargo, la primera porción de señal de audio puede estar ya representada en el dominio de rastreo.

[0281] En algunos ejemplos útiles para comprender la invención, la información de primera porción de señal es idéntica a la primera porción de señal de audio. En otros ejemplos útiles para comprender la invención, la información
30 de primera porción de señal es, por ejemplo, un valor agregado dependiendo de la primera porción de señal de audio.

[0282] Ahora, en primer lugar, se considera en más detalle el desvanecimiento a un nivel de ruido de confort.

[0283] La estrategia de desvanecimiento descrita puede ser implementado, por ejemplo, en una versión de
35 bajo retardo de xHE–codificación de audio avanzada (AAC) [NMR+12] (xHE–AAC = codificación de audio avanzada de alta eficiencia extendida, conforme a sus siglas en inglés), que es capaz de cambiar sin problemas entre la codificación de predicción lineal excitada por código algebraico (ACELP) (la voz) y de transformada de coseno discreta modificada (MDCT) (música/ruido) sobre una base por trama.

40 **[0284]** Con respecto al rastreo de nivel común en un dominio de rastreo, por ejemplo, un dominio de excitación, para aplicar un desvanecimiento suave a un nivel de ruido de confort apropiado durante la pérdida de paquetes, es necesario identificar dicho nivel de ruido de confort durante el proceso de decodificación normal. Por ejemplo, puede asumirse que un nivel de ruido similar al ruido de fondo es más confortable. Por lo tanto, el nivel de ruido de fondo puede ser derivado y actualizado de forma constante durante la codificación normal.

45 **[0285]** La presente invención se sustenta en el hallazgo de que cuando se tiene un códec de núcleo cambiado (por ejemplo, la predicción lineal excitada por código algebraico (ACELP) y la excitación codificada por la transformada (TCX)), es en particular adecuada la consideración de un nivel de ruido de fondo común, independientemente del codificador de núcleo seleccionado.

50 **[0286]** La figura 6 representa un rastreo del nivel de ruido de fondo según un ejemplo útil para comprender la invención preferido en el decodificador durante el modo de operación libre de error, por ejemplo, durante la decodificación normal.

55 **[0287]** El rastreo en sí mismo, por ejemplo, puede realizarse usando la estrategia de estadística mínima (véase [Mar01]).

[0288] Este nivel rastreado de ruido de fondo puede considerarse, por ejemplo, la información de nivel de ruido mencionada anteriormente.

60 **[0289]** Por ejemplo, la estimación de ruido de estadística mínima presentada en el documento: “Rainer Martin, *Noise power spectral density estimation based on optimal smoothing and minimum statistics*, IEEE Transactions on Speech and Audio Processing 9 (2001), n. ° 5, 504 –512” [Mar01] puede emplearse para el rastreo del nivel de ruido de fondo.

65

[0290] De manera correspondiente, en algunos ejemplos útiles para comprender la invención, la unidad de rastreo de nivel de ruido 130 está configurada para determinar la información de nivel de ruido mediante la aplicación de una estrategia de estadística mínima, por ejemplo, mediante el empleo de la estimación de ruido de estadística mínima de [Mar01].

5

[0291] A continuación, se describen algunas consideraciones y detalles de esta estrategia de rastreo.

[0292] Con respecto al rastreo de nivel, se supone que el fondo es de tipo ruido. En consecuencia, es preferible realizar el rastreo de nivel en el dominio de excitación, a fin de evitar el rastreo de componentes tonales delanteros que son extraídos por la codificación predictiva lineal (LPC). Por ejemplo, el relleno de ruido de predicción lineal excitada por código algebraico (ACELP) puede emplear además el nivel de ruido de fondo en el dominio de excitación. Con el rastreo en el dominio de excitación, solo un único rastreo del nivel de ruido de fondo puede cumplir dos propósitos, lo que ahorra complejidad computacional. En un ejemplo útil para comprender la invención preferido, el rastreo se realiza en el dominio de excitación de la predicción lineal excitada por código algebraico (ACELP).

10

15

[0293] La figura 7 ilustra la derivación de ganancia de la síntesis de codificación predictiva lineal (LPC) y el desénfasis según un ejemplo útil para comprender la invención.

[0294] Con respecto a la derivación de nivel, por ejemplo, la derivación de nivel puede ser conducida o bien en el dominio de tiempo o en el dominio de excitación, o en cualquier otro dominio adecuado. Si los dominios para la derivación de nivel y el rastreo de nivel difieren, puede ser necesaria, por ejemplo, una compensación de ganancia.

20

[0295] En el ejemplo útil para comprender la invención preferido, la derivación de nivel para la predicción lineal excitada por código algebraico (ACELP) se realiza en el dominio de excitación. En consecuencia, no se requiere compensación de ganancia.

25

[0296] Para la excitación codificada por la transformada (TCX), puede ser necesaria, por ejemplo, una compensación de ganancia a fin de ajustar el nivel derivado al dominio de excitación de la predicción lineal excitada por código algebraico (ACELP).

30

[0297] En el ejemplo útil para comprender la invención preferido, la derivación de nivel para la excitación codificada por la transformada (TCX) tiene lugar en el dominio de tiempo. Se halló una compensación de ganancia manipulable para esta estrategia: la ganancia introducida por la síntesis de codificación predictiva lineal (LPC) y el desénfasis es derivada como se muestra en la figura 7, y el nivel derivado es dividido por esta ganancia.

35

[0298] Alternativamente, la derivación de nivel para la excitación codificada por la transformada (TCX) podría realizarse en el dominio de excitación de la excitación codificada por la transformada (TCX). Sin embargo, la compensación de ganancia entre el dominio de excitación de la excitación codificada por la transformada (TCX) y el dominio de excitación de la predicción lineal excitada por código algebraico (ACELP) se consideró demasiado complicado.

40

[0299] Por lo tanto, volviendo a la figura 1a, en algunos ejemplos útiles para comprender la invención, la primera porción de señal de audio es representada en un dominio de tiempo como el primer dominio. La unidad de transformada 120 está configurada para transformar la segunda porción de señal de audio o el valor derivado de la segunda porción de señal de audio de un dominio de excitación que es el segundo dominio, al dominio de tiempo que es el dominio de rastreo. En tales ejemplos útiles para comprender la invención, la unidad de rastreo de nivel de ruido 130 está configurada para recibir la información de primera porción de señal que es representada en el dominio de tiempo como el dominio de rastreo. Además, la unidad de rastreo de nivel de ruido 130 está configurada para recibir la segunda porción de señal que es representada en el dominio de tiempo como el dominio de rastreo.

50

[0300] En otros ejemplos útiles para comprender la invención, la primera porción de señal de audio está representada en un dominio de excitación como el primer dominio. La unidad de transformada 120 está configurada para transformar la segunda porción de señal de audio o el valor derivado de la segunda porción de señal de audio, de un dominio de tiempo que es el segundo dominio, al dominio de excitación que es el dominio de rastreo. En tales ejemplos útiles para comprender la invención, la unidad de rastreo de nivel de ruido 130 está configurada para recibir la información de primera porción de señal que es representada en el dominio de excitación como el dominio de rastreo. Además, la unidad de rastreo de nivel de ruido 130 está configurada para recibir la segunda porción de señal que es representada en el dominio de excitación como el dominio de rastreo.

55

[0301] En un ejemplo útil para comprender la invención, la primera porción de señal de audio puede estar representada, por ejemplo, en un dominio de excitación como el primer dominio, donde la unidad de rastreo de nivel de ruido 130 puede estar configurada, por ejemplo, para recibir la información de primera porción de señal, donde dicha información de primera porción de señal está representada en el dominio FFT, que es el dominio de rastreo, y donde dicha información de primera porción de señal depende de dicha primera porción de señal de audio que está representada en el dominio de excitación, donde la unidad de transformada 120 puede estar configurada, por ejemplo,

60

65

para transformar la segunda porción de señal de audio o el valor derivado de la segunda porción de señal de audio, de un dominio de tiempo que es el segundo dominio, a un dominio FFT que es el dominio de rastreo, y donde la unidad de rastreo de nivel de ruido 130 puede estar configurada, por ejemplo, para recibir la segunda porción de señal de audio que es representada en el dominio FFT.

5

[0302] La figura 1b ilustra un aparato según otro ejemplo útil para comprender la invención. En la figura 1b, la unidad de transformada 120 de la figura 1a es una primera unidad de transformada 120, y la unidad de reconstrucción 140 de la figura 1a es una primera unidad de reconstrucción 140. El aparato comprende adicionalmente una segunda unidad de transformada 121 y una segunda unidad de reconstrucción 141.

10

[0303] La segunda unidad de transformada 121 está configurada para transformar la información de nivel de ruido del dominio de rastreo al segundo dominio, si una cuarta trama de la pluralidad de tramas no es recibida por la interfaz receptora o si dicha cuarta trama es recibida por la interfaz receptora, pero es alterada.

15 **[0304]**

Además, la segunda unidad de reconstrucción 141 está configurada para reconstruir una cuarta porción de señal de audio de la señal de audio según la información de nivel de ruido que es representada en el segundo dominio si dicha cuarta trama de la pluralidad de tramas no es recibida por la interfaz receptora o si dicha cuarta trama es recibida por la interfaz receptora, pero es alterada.

20 **[0305]**

La figura 1c ilustra un aparato para la decodificación de una señal de audio según otro ejemplo útil para comprender la invención. El aparato comprende adicionalmente una primera unidad de agregación 150 para la determinación de un primer valor agregado dependiendo de la primera porción de señal de audio. Además, el aparato de la figura 1c comprende adicionalmente una segunda unidad de agregación 160 para la determinación de un segundo valor agregado como el valor derivado de la segunda porción de señal de audio dependiendo de la segunda porción de señal de audio. En el ejemplo útil para comprender la invención de la figura 1c, la unidad de rastreo de nivel de ruido 130 está configurada para recibir el primer valor agregado como la información de primera porción de señal que es representada en el dominio de rastreo, donde la unidad de rastreo de nivel de ruido 130 está configurada para recibir el segundo valor agregado como la información de segunda porción de señal que es representada en el dominio de rastreo. La unidad de rastreo de nivel de ruido 130 está configurada para determinar la información de nivel de ruido dependiendo del primer valor agregado que es representado en el dominio de rastreo y según el segundo valor agregado que es representado en el dominio de rastreo.

25

30

[0306] En un ejemplo útil para comprender la invención, la primera unidad de agregación 150 está configurada para determinar el primer valor agregado, de tal manera que el primer valor agregado indica una media cuadrática de la primera porción de señal de audio o de una señal derivada de la primera porción de señal de audio. Además, la segunda unidad de agregación 160 está configurada para determinar el segundo valor agregado, de tal manera que el segundo valor agregado indica una media cuadrática de la segunda porción de señal de audio o de una señal derivada de la segunda porción de señal de audio.

35

40 **[0307]**

La figura 6 ilustra un aparato para la decodificación de una señal de audio según un ejemplo útil para comprender la invención adicional.

[0308] En la figura 6, la unidad de rastreo de nivel de fondo 630 implementa una unidad de rastreo de nivel de ruido 130 según la figura 1a.

45

[0309] Además, en la figura 6, la unidad de media cuadrática (RMS) 650 (RMS = media cuadrática, conforme a sus siglas en inglés) es una primera unidad de agregación, y la unidad de media cuadrática (RMS) 660 es una segunda unidad de agregación.

50 **[0310]**

Según algunos ejemplos útiles para comprender la invención, la (primera) unidad de transformada 120 de la figura 1a, figura 1b y figura 1c está configurada para transformar el valor derivado de la segunda porción de señal de audio, del segundo dominio al dominio de rastreo, mediante la aplicación de un valor de ganancia (x) sobre el valor derivado de la segunda porción de señal de audio, por ejemplo, mediante la división del valor derivado de la segunda porción de señal de audio por un valor de ganancia (x). En otros ejemplos útiles para comprender la invención, un valor de ganancia puede ser multiplicado, por ejemplo.

55

[0311] En algunos ejemplos útiles para comprender la invención, el valor de ganancia (x puede indicar), por ejemplo, una ganancia introducida por la síntesis de codificación predictiva lineal, o el valor de ganancia (x) puede indicar, por ejemplo, una ganancia introducida por la síntesis de codificación predictiva lineal y el desénfasis.

60

[0312] En la figura 6, la unidad 622 proporciona el valor (x) que indica la ganancia introducida por la síntesis de codificación predictiva lineal y el desénfasis. La unidad 622 divide a continuación el valor proporcionado por la segunda unidad de agregación 660, que es un valor derivado de la segunda porción de señal de audio, por el valor de ganancia proporcionado (x) (por ejemplo, ya sea mediante la división por x, ya sea mediante la multiplicación del valor 1/x). Por lo tanto, la unidad 620 de la figura 6 que comprende las unidades 621 y 622 implementa la primera unidad

65

de transformada de la figura 1a, figura 1b o figura 1c.

[0313] El aparato de la figura 6 recibe una primera trama con una primera porción de señal de audio que es una excitación con voz y/o una excitación sin voz y que es representada en el dominio de rastreo, en la figura 6, un dominio de codificación predictiva lineal (LPC) (predicción lineal excitada por código algebraico (ACELP)). La primera porción de señal de audio es alimentada en una unidad de síntesis de codificación predictiva lineal (LPC) y desénfasis 671 para el procesamiento de manera que se obtenga una salida de primera porción de señal de audio en el dominio de tiempo. Además, la primera porción de señal de audio es alimentada en el módulo de media cuadrática (RMS) 650 de manera que se obtenga un primer valor que indica una media cuadrática de la primera porción de señal de audio. Este primer valor (primer valor de media cuadrática (RMS)) está representado en el dominio de rastreo. El primer valor de media cuadrática (RMS), que es representado en el dominio de rastreo, es alimentado entonces en la unidad de rastreo de nivel de ruido 630.

[0314] Además, el aparato de la figura 6 recibe una segunda trama con una segunda porción de señal de audio que comprende un espectro de transformada de coseno discreta modificada (MDCT) y que es representada en un dominio de transformada de coseno discreta modificada (MDCT). El relleno de ruido se conduce por medio de un módulo de relleno de ruido 681, la formación de ruido en el dominio de frecuencia se conduce por medio de un módulo de formación de ruido en el dominio de frecuencia 682; transformación al dominio de tiempo se conduce por medio de un módulo de transformada de coseno discreta modificada (iMDCT)/OLA 683 (OLA = superposición y adición, conforme a sus siglas en inglés), y la predicción a largo plazo se conduce por medio de una unidad de predicción a largo plazo 684. La unidad de predicción a largo plazo puede comprender, por ejemplo, búfer de retardo (no expuesto en la figura 6).

[0315] La señal derivada de la segunda porción de señal de audio es alimentada a continuación en el Módulo de media cuadrática (RMS) 660 de manera que se obtenga un segundo valor que indica la obtención de una media cuadrática de dicha señal derivada de la segunda porción de señal de audio. Este segundo valor (segundo valor de media cuadrática (RMS)) está aún representado en el dominio de tiempo. La unidad 620 transforma a continuación el segundo valor de media cuadrática (RMS) del dominio de tiempo al dominio de rastreo, aquí, el dominio de codificación predictiva lineal (LPC). El segundo valor de media cuadrática (RMS), que es representado en el dominio de rastreo, es alimentado a continuación en la unidad de rastreo de nivel de ruido 630.

[0316] En ejemplos útiles para comprender la invención, el rastreo de nivel se conduce en el dominio de excitación, si bien el desvanecimiento de excitación codificada por la transformada (TCX) se conduce en el dominio de tiempo.

[0317] Mientras que durante la decodificación normal se rastrea el nivel de ruido de fondo, puede usarse, por ejemplo, durante la pérdida de paquetes, como un indicador de un nivel apropiado de ruido de confort, al cual la última señal recibida es suavemente desvanecida a modo de nivel.

[0318] La derivación del nivel para el rastreo y la aplicación del desvanecimiento de nivel, en general, son independientes entre sí, y podrían realizarse en diferentes dominios. En el ejemplo útil para comprender la invención preferido, la aplicación de nivel se realiza en los mismos dominios que la derivación de nivel, de modo que conduzca a los mismos beneficios que para la predicción lineal excitada por código algebraico (ACELP), sin compensación de ganancia necesaria, y que para la excitación codificada por la transformada (TCX), la compensación de ganancia inversa como para la derivación de nivel (véase la figura 6) es necesaria y, en consecuencia, puede usarse la misma derivación de ganancia, como se ilustra en la figura 7.

[0319] En lo que sigue, se describe la compensación de una influencia del filtro de paso alto sobre la ganancia de síntesis de la codificación predictiva lineal (LPC) según ejemplos útiles para comprender la invención.

[0320] La figura 8 destaca esta estrategia. En particular, la figura 8 ilustra la aplicación de nivel de ruido de confort durante la pérdida de paquetes.

[0321] En la figura 8, la unidad de filtro de ganancia de paso alto 643, la unidad de multiplicación 644, la unidad de desvanecimiento 645, la unidad de filtro de paso alto 646, la unidad de desvanecimiento 647 y la unidad de combinación 648 forman en conjunto una primera unidad de reconstrucción.

[0322] Además, en la figura 8, la unidad de provisión de nivel de fondo 631 proporciona la información de nivel de ruido. Por ejemplo, la unidad de provisión de nivel de fondo 631 puede ser implementada de igual manera que la unidad de rastreo de nivel de fondo 630 de la figura 6.

[0323] Adicionalmente, en la figura 8, la unidad de ganancia de síntesis de codificación predictiva lineal (LPC) y desénfasis 649, y la unidad de multiplicación 641 forman juntas una segunda unidad de transformada 640.

[0324] Además, en la figura 8, la unidad de desvanecimiento 642 representa una segunda unidad de

reconstrucción.

[0325] En el ejemplo útil para comprender la invención de la figura 8, la excitación con voz y la excitación sin voz son desvanecidas de forma separada. La excitación con voz es desvanecida a cero, si bien la excitación sin voz es desvanecida hacia el nivel de ruido de confort. La figura 8 representa adicionalmente un filtro de paso alto, que es introducido en la cadena de señal de la excitación sin voz a fin de suprimir los componentes de baja frecuencia para todos los casos, excepto cuando la señal fue clasificada como sin voz.

[0326] A fin de modelar la influencia del filtro de paso alto, el nivel después de la síntesis de codificación predictiva lineal (LPC) y el desénfasis es computado una vez con el filtro de paso alto, y una vez sin el filtro de paso alto. A continuación, la relación de dichos dos niveles es derivada y utilizada para alterar el nivel de fondo aplicado.

[0327] Esto se ilustra por medio de la figura 9. En particular, la figura 9 representa compensación de ganancia de paso alto avanzada durante el ocultamiento de predicción lineal excitada por código algebraico (ACELP) según un ejemplo útil para comprender la invención.

[0328] En lugar de la señal de excitación actual, se usa solo un simple impulso como entrada para este cómputo. Esto permite una reducida complejidad, ya que la respuesta de impulso decae rápidamente y, de ese modo, puede realizarse la derivación de media cuadrática (RMS) en una trama de tiempo más corto. En la práctica, solo se usa una subtrama en lugar de la trama entera.

[0329] Según un ejemplo útil para comprender la invención, la unidad de rastreo de nivel de ruido 130 está configurada para determinar un nivel de ruido de confort como la información de nivel de ruido. La unidad de reconstrucción 140 está configurada para reconstruir la tercera porción de señal de audio dependiendo de la información de nivel de ruido, si dicha tercera trama de la pluralidad de tramas no es recibida por la interfaz receptora 110 o si dicha tercera trama es recibida por la interfaz receptora 110 pero es alterada.

[0330] Según un ejemplo útil para comprender la invención, la unidad de rastreo de nivel de ruido 130 está configurada para determinar un nivel de ruido de confort como la información de nivel de ruido. La unidad de reconstrucción 140 está configurada para reconstruir la tercera porción de señal de audio según la información de nivel de ruido, si dicha tercera trama de la pluralidad de tramas no es recibida por la interfaz receptora 110 o si dicha tercera trama es recibida por la interfaz receptora 110 pero es alterada.

[0331] En un ejemplo útil para comprender la invención, la unidad de rastreo de nivel de ruido 130 está configurada para determinar un nivel de ruido de confort como la información de nivel de ruido derivada de un espectro de nivel de ruido, donde dicho espectro de nivel de ruido se obtiene mediante la aplicación de la estrategia de estadística mínima. La unidad de reconstrucción 140 está configurada para reconstruir la tercera porción de señal de audio según una pluralidad de coeficientes predictivos lineales, si dicha tercera trama de la pluralidad de tramas no es recibida por la interfaz receptora 110 o si dicha tercera trama es recibida por la interfaz receptora 110 pero es alterada.

[0332] En un ejemplo útil para comprender la invención, la (primera y/o segunda) unidad de reconstrucción 140, 141 puede estar configurada, por ejemplo, para reconstruir la tercera porción de señal de audio dependiendo de la información de nivel de ruido y dependiendo de la primera porción de señal de audio, si dicha tercera (cuarta) trama de la pluralidad de tramas no es recibida por la interfaz receptora 110 o si dicha tercera (cuarta) trama es recibida por la interfaz receptora 110 pero es alterada.

[0333] Según un ejemplo útil para comprender la invención, la (primera y/o segunda) unidad de reconstrucción 140, 141, por ejemplo, puede ser configurada para reconstruir la tercera (o cuarta) porción de señal de audio mediante la atenuación o amplificación de la primera porción de señal de audio.

[0334] La figura 14 ilustra un aparato para la decodificación de una señal de audio. El aparato comprende una interfaz receptora 110, donde la interfaz receptora 110 está configurada para recibir una primera trama que comprende una primera porción de señal de audio de la señal de audio, y donde la interfaz receptora 110 está configurada para recibir una segunda trama que comprende una segunda porción de señal de audio de la señal de audio.

[0335] Además, el aparato comprende una unidad de rastreo de nivel de ruido 130, donde la unidad de rastreo de nivel de ruido 130 está configurada para determinar la información de nivel de ruido dependiendo de al menos una de la primera porción de señal de audio y la segunda porción de señal de audio (esto significa: según la primera porción de señal de audio y/o la segunda porción de señal de audio), donde la información de nivel de ruido es representada en un dominio de rastreo.

[0336] Adicionalmente, el aparato comprende una primera unidad de reconstrucción 140 para la reconstrucción de, en un primer dominio de reconstrucción, una tercera porción de señal de audio de la señal de audio según la información de nivel de ruido, si una tercera trama de la pluralidad de tramas no es recibida por la interfaz receptora 110 o si dicha tercera trama es recibida por la interfaz receptora 110 pero es alterada, donde el primer dominio de

reconstrucción es diferente o igual respecto del dominio de rastreo.

5 **[0337]** Además, el aparato comprende una unidad de transformada 121 para la transformación de la información de nivel de ruido, del dominio de rastreo a un segundo dominio de reconstrucción, si una cuarta trama de la pluralidad de tramas no es recibida por la interfaz receptora 110 o si dicha cuarta trama es recibida por la interfaz receptora 110 pero es alterada, donde el segundo dominio de reconstrucción es diferente del dominio de rastreo, y donde el segundo dominio de reconstrucción es diferente del primer dominio de reconstrucción; y

10 **[0338]** Adicionalmente, el aparato comprende una segunda unidad de reconstrucción 141 para la reconstrucción de, en el segundo dominio de reconstrucción, una cuarta porción de señal de audio de la señal de audio dependiendo de la información de nivel de ruido que es representada en el segundo dominio de reconstrucción, si dicha cuarta trama de la pluralidad de tramas no es recibida por la interfaz receptora 110 o si dicha cuarta trama es recibida por la interfaz receptora 110 pero es alterada.

15 **[0339]** Según algunos ejemplos útiles para comprender la invención, el dominio de rastreo puede ser, por ejemplo, donde el dominio de rastreo es un dominio de tiempo, un dominio espectral, un dominio FFT, un dominio de transformada de coseno discreta modificada (MDCT), o un dominio de excitación. El primer dominio de reconstrucción puede ser, por ejemplo, el dominio de tiempo, el dominio espectral, el dominio FFT, el dominio de transformada de coseno discreta modificada (MDCT), o el dominio de excitación. El segundo dominio de reconstrucción puede ser, por
20 ejemplo, el dominio de tiempo, el dominio espectral, el dominio FFT, el dominio de transformada de coseno discreta modificada (MDCT), o el dominio de excitación.

[0340] En un ejemplo útil para comprender la invención, el dominio de rastreo puede ser, por ejemplo, el dominio FFT, el primer dominio de reconstrucción puede ser, por ejemplo, el dominio de tiempo, y el segundo dominio de reconstrucción puede ser, por ejemplo, el dominio de excitación.
25

[0341] En otro ejemplo útil para comprender la invención, el dominio de rastreo puede ser, por ejemplo, el dominio de tiempo, el primer dominio de reconstrucción puede ser, por ejemplo, el dominio de tiempo, y el segundo dominio de reconstrucción puede ser, por ejemplo, el dominio de excitación.
30

[0342] Según un ejemplo útil para comprender la invención, dicha primera porción de señal de audio puede ser representada, por ejemplo, en un primer dominio de entrada, y dicha segunda porción de señal de audio puede ser representada, por ejemplo, en un segundo dominio de entrada. La unidad de transformada puede ser, por ejemplo, una segunda unidad de transformada. El aparato puede comprender, por ejemplo, adicionalmente una primera unidad de transformada para la transformación de la segunda porción de señal de audio o un valor o una señal derivados de la segunda porción de señal de audio, del segundo dominio de entrada al dominio de rastreo, de manera que se obtenga una información de segunda porción de señal. La unidad de rastreo de nivel de ruido puede ser configurada, por ejemplo, para recibir una información de primera porción de señal que es representada en el dominio de rastreo, donde la información de primera porción de señal depende de la primera porción de señal de audio, donde la unidad de rastreo de nivel de ruido está configurada para recibir la segunda porción de señal que es representada en el dominio de rastreo, y donde la unidad de rastreo de nivel de ruido está configurada para la determinación de la información de nivel de ruido dependiendo de la información de primera porción de señal que es representada en el dominio de rastreo y dependiendo de la información de segunda porción de señal que es representada en el dominio de rastreo.
35
40
45

[0343] Según un ejemplo útil para comprender la invención, el primer dominio de entrada puede ser, por ejemplo, el dominio de excitación, y el segundo dominio de entrada puede ser, por ejemplo, el dominio de transformada de coseno discreta modificada (MDCT).

50 **[0344]** En otro ejemplo útil para comprender la invención, el primer dominio de entrada puede ser, por ejemplo, el dominio de transformada de coseno discreta modificada (MDCT), y donde el segundo dominio de entrada puede ser, por ejemplo, el dominio de transformada de coseno discreta modificada (MDCT).

[0345] Si, por ejemplo, una señal está representada en un dominio de tiempo, puede ser representada, por ejemplo, por muestras de dominio de tiempo de la señal. O, por ejemplo, si una señal está representada en un dominio espectral puede ser representada, por ejemplo, por muestras espectrales de un espectro de la señal.
55

[0346] En un ejemplo útil para comprender la invención, el dominio de rastreo puede ser, por ejemplo, el dominio FFT, el primer dominio de reconstrucción puede ser, por ejemplo, el dominio de tiempo, y el segundo dominio de reconstrucción puede ser, por ejemplo, el dominio de excitación.
60

[0347] En otro ejemplo útil para comprender la invención, el dominio de rastreo puede ser, por ejemplo, el dominio de tiempo, el primer dominio de reconstrucción puede ser, por ejemplo, el dominio de tiempo, y el segundo dominio de reconstrucción puede ser, por ejemplo, el dominio de excitación.
65

[0348] En algunos ejemplos útiles para comprender la invención, las unidades ilustradas en la figura 14 pueden estar configuradas, por ejemplo, como se describe para las figuras 1a, 1b, 1c y 1d.

[0349] Con respecto a ejemplos útiles para comprender la invención particular, en, por ejemplo, un modo de tasa baja, un aparato según un ejemplo útil para comprender la invención puede recibir, por ejemplo, tramas de predicción lineal excitada por código algebraico (ACELP) como una entrada, que son representadas en un dominio de excitación, y que son transformadas a continuación en un dominio de tiempo por medio de la síntesis de codificación predictiva lineal (LPC). Además, en el modo de tasa baja, el aparato según un ejemplo útil para comprender la invención puede recibir, por ejemplo, tramas de excitación codificada por la transformada (TCX) como una entrada, que son representadas en un dominio de transformada de coseno discreta modificada (MDCT), y que son transformadas a continuación en un dominio de tiempo por medio de una transformada de coseno discreta modificada (MDCT) inversa.

[0350] A continuación, se conduce el rastreo en un dominio de FFT, donde la señal de FFT es derivada de la señal de dominio de tiempo mediante la conducción de una FFT (Transformada rápida de Fourier). El rastreo, por ejemplo, puede ser conducido por la conducción de una estrategia de estadística mínima, separada para todas las líneas espectrales de manera que se obtenga un espectro de ruido de confort.

[0351] A continuación, se realiza el ocultamiento mediante la conducción de la derivación de nivel sobre la base del espectro de ruido de confort. La derivación de nivel se conduce sobre la base del espectro de ruido de confort. La conversión de nivel en el dominio de tiempo se conduce para el ocultamiento de pérdida de paquetes (PLC) de excitación codificada por la transformada (TCX) de dominio de frecuencia (FD). Se conduce un desvanecimiento en el dominio de tiempo. Se conduce una derivación de nivel en el dominio de excitación para el ocultamiento de pérdida de paquetes (PLC) de predicción lineal excitada por código algebraico (ACELP) y para el ocultamiento de pérdida de paquetes (PLC) de excitación codificada por la transformada (TCX) de dominio de tiempo (TD) (de tipo ACELP). A continuación, se lleva a cabo un desvanecimiento en el dominio de excitación.

[0352] El siguiente listado resume esto:
tasa baja:

- entrada:
 - predicción Lineal Excitada por Código Algebraico (ACELP) (dominio de excitación → dominio de tiempo, por medio de la síntesis de codificación predictiva lineal (LPC)
 - excitación codificada por la transformada (TCX) (dominio de transformada de coseno discreta modificada (MDCT) → dominio de tiempo, mediante la transformada de coseno discreta modificada (MDCT) inversa)
- rastreo:
 - dominio FFT, derivado de dominio de tiempo por medio de FFT
 - estadística mínima, separada para todas las líneas espectrales → espectro de ruido de confort
- ocultamiento:
 - derivación de nivel sobre la base del espectro de ruido de confort
 - conversión de nivel en el dominio de tiempo para ocultamiento de pérdida de paquetes (PLC), de excitación codificada por la transformada (TCX) de dominio de frecuencia (FD) → desvanecimiento en el dominio de tiempo
 - conversión de nivel en el dominio de excitación para ocultamiento de pérdida de paquetes (PLC) de predicción lineal excitada por código algebraico (ACELP) y para el ocultamiento de pérdida de paquetes (PLC) de excitación codificada por la transformada (TCX) de dominio de tiempo (TD) (de tipo ACELP) → desvanecimiento en el dominio de excitación

[0353] Por ejemplo, en un modo de tasa alta, pueden recibirse tramas de excitación codificada por la transformada (TCX) como una entrada, que son representadas en el dominio de la transformada de coseno discreta modificada (MDCT), y que son transformadas a continuación al dominio de tiempo mediante una transformada de coseno discreta modificada (MDCT) inversa.

[0354] El rastreo puede conducirse a continuación en el dominio de tiempo. Por ejemplo, el rastreo puede conducirse mediante la conducción de una estrategia de estadística mínima sobre la base del nivel de energía, de

manera que se obtenga un nivel de ruido de confort.

[0355] Para el ocultamiento, para el ocultamiento de pérdida de paquetes (PLC) de excitación codificada por la transformada (TCX) de dominio de frecuencia (FD), el nivel puede usarse como tal, y puede conducirse solamente un desvanecimiento en el dominio de tiempo. Para el ocultamiento de pérdida de paquetes (PLC) de excitación codificada por la transformada (TCX) de dominio de tiempo (TD), (de tipo ACELP)), se conduce la conversión de nivel en el dominio de excitación, y el desvanecimiento se conduce en el dominio de excitación.

[0356] El siguiente listado resume esto:
10 tasa alta:

- entrada:
 - excitación codificada por la transformada (TCX) (dominio de transformada de coseno discreta modificada (MDCT) → dominio de tiempo, mediante transformada de coseno discreta modificada (MDCT) inversa)

15

- rastreo:

- dominio de tiempo

20

- estadística mínima sobre el nivel de energía → nivel de ruido de confort

- ocultamiento:

- uso de nivel "como tal"

25

- ocultamiento de pérdida de paquetes (PLC) de excitación codificada por la transformada (TCX) de dominio de frecuencia (FD) → desvanecimiento en el dominio de tiempo

- conversión de nivel en el dominio de excitación para

30

- ocultamiento de pérdida de paquetes (PLC) de excitación codificada por la transformada (TCX) de dominio de tiempo (TD) (de tipo ACELP) → desvanecimiento en el dominio de excitación

[0357] El dominio FFT y el dominio de la transformada de coseno discreta modificada (MDCT) son ambos dominios espectrales, mientras que el dominio de excitación es cierta clase de dominio de tiempo.

[0358] Según un ejemplo útil para comprender la invención, la primera unidad de reconstrucción 140 puede estar configurada, por ejemplo, para reconstruir la tercera porción de señal de audio mediante la conducción de un primer desvanecimiento a un espectro de tipo ruido. La segunda unidad de reconstrucción 141 puede estar configurada, por ejemplo, para reconstruir la cuarta porción de señal de audio mediante la conducción de un segundo desvanecimiento a un espectro de tipo ruido y/o un segundo desvanecimiento de una ganancia de predicción a largo plazo (LTP). Además, la primera unidad de reconstrucción 140 y la segunda unidad de reconstrucción 141 pueden estar configuradas, por ejemplo, para conducir el primer desvanecimiento y el segundo desvanecimiento a un espectro de tipo ruido y/o un segundo desvanecimiento de una ganancia de predicción a largo plazo (LTP) con la misma velocidad de desvanecimiento.

[0359] Ahora, se considera la forma espectral adaptativa de ruido de confort.

[0360] A fin de lograr la formación adaptativa al ruido de confort durante la pérdida de paquetes de ráfaga, como una primera etapa, puede conducirse el hallazgo de coeficientes apropiados de codificación predictiva lineal (LPC) que representan el ruido de fondo. Estos coeficientes de codificación predictiva lineal (LPC) pueden derivarse durante la voz activa mediante el uso de una estrategia de estadística mínima para hallar el espectro de ruido de fondo y, a continuación, el cálculo de los coeficientes de codificación predictiva lineal (LPC) a partir de este, mediante el uso de un algoritmo arbitrario para la derivación de codificación predictiva lineal (LPC) conocido de la bibliografía. Algunos ejemplos útiles para comprender la invención, por ejemplo, pueden convertir directamente el espectro de ruido de fondo en una representación que puede usarse directamente para la sustitución de ruido de dominio de frecuencia (FDNS) en el dominio de la transformada de coseno discreta modificada (MDCT).

[0361] El desvanecimiento a ruido de confort puede realizarse en el dominio de frecuencia espectral de impedancia y admitancia (ISF) (también aplicable en el dominio de LSF; LSF: frecuencia espectral lineal):

$$f_{\text{actual}}[i] = \alpha \cdot f_{\text{último}}[i] + (1 - \alpha) \cdot pt_{\text{media}}[i] \quad i = 0 \dots 16 \quad (26)$$

estableciendo pt_{media} en coeficientes de predicción lineal (LP) apropiados que describen el ruido de confort.

[0362] Con respecto a la forma espectral adaptativa anteriormente descrita del ruido de confort, se ilustra un ejemplo útil para comprender la invención más general en la figura 11.

[0363] La figura 11 ilustra un aparato para la decodificación de una señal de audio codificada de manera que se obtenga una señal de audio reconstruida según un ejemplo útil para comprender la invención.

[0364] El aparato comprende una interfaz receptora 1110 para la recepción de una o más tramas, un generador de coeficientes 1120, y un reconstructor de señal 1130.

[0365] El generador de coeficientes 1120 está configurado para determinar, si una trama actual de una o más tramas es recibida por la interfaz receptora 1110, y si la trama actual que es recibida por la interfaz receptora 1110 no es alterada/errónea, siendo uno o más primeros coeficientes de señal de audio, comprendidos por la trama actual, donde dichos uno o más primeros coeficientes de señal de audio indican una característica de la señal de audio codificada, e indicando uno o más coeficientes de ruido un ruido de fondo de la señal de audio codificada. Además, el generador de coeficientes 1120 está configurado para generar uno o más segundos coeficientes de señal de audio, dependiendo de los uno o más primeros coeficientes de señal de audio y dependiendo de los uno o más coeficientes de ruido, si la trama actual no es recibida por la interfaz receptora 1110 o si la trama actual recibida por la interfaz receptora 1110 es alterada/errónea.

[0366] El reconstructor de señal de audio 1130 está configurado para reconstruir una primera porción de la señal de audio reconstruida dependiendo de los uno o más primeros coeficientes de señal de audio, si la trama actual es recibida por la interfaz receptora 1110 y si la trama actual recibida por la interfaz receptora 1110 no está alterada. Además, el reconstructor de señal de audio 1130 está configurado para reconstruir una segunda porción de la señal de audio reconstruida dependiendo de uno o más segundos coeficientes de señal de audio, si la trama actual no es recibida por la interfaz receptora 1110 o si la trama actual recibida por la interfaz receptora 1110 está alterada.

[0367] La determinación de un ruido de fondo es bien conocida en la técnica (véase, por ejemplo, [Mar01]: Rainer Martin, *Noise power spectral density estimation based on optimal smoothing and minimum statistics*, IEEE Transactions on Speech and Audio Processing 9 (2001), n.º 5, 504–512), y en un ejemplo útil para comprender la invención, el aparato procede de forma consecutiva.

[0368] En algunos ejemplos útiles para comprender la invención, uno o más primeros coeficientes de señal de audio pueden ser, por ejemplo, uno o más coeficientes de filtro predictivo lineal de la señal de audio codificada. En algunos ejemplos útiles para comprender la invención, uno o más primeros coeficientes de señal de audio pueden ser, por ejemplo, uno o más coeficientes de filtro predictivo lineal de la señal de audio codificada.

[0369] Se conoce bien en la técnica la manera de reconstruir una señal de audio, por ejemplo, una señal de voz, a partir de coeficientes de filtro predictivo lineal o de pares espectrales de impedancia y admitancia (véase, por ejemplo, [3GP09c]: *Speech codec speech processing functions; adaptive multi-rate – wideband (AMRWB) speech codec; transcoding functions*, 3GPP TS 26.190, 3rd Generation Partnership Project, 2009), y en un ejemplo útil para comprender la invención, el reconstructor de señal procede consecutivamente.

[0370] Según un ejemplo útil para comprender la invención, uno o más coeficientes de ruido, por ejemplo, pueden ser uno o más coeficientes de filtro predictivo lineal que indican el ruido de fondo de la señal de audio codificada. En un ejemplo útil para comprender la invención, uno o más coeficientes de filtro predictivo lineal pueden representar, por ejemplo, una forma espectral del ruido de fondo.

[0371] En un ejemplo útil para comprender la invención, el generador de coeficientes 1120 puede estar configurado, por ejemplo, para determinar una o más segundas porciones de señal de audio de tal manera que una o más segundas porciones de señal de audio son uno o más coeficientes de filtro predictivo lineal de la señal de audio reconstruida, o de tal manera que uno o más primeros coeficientes de señal de audio son uno o más pares espectrales de impedancia y admitancia de la señal de audio reconstruida.

[0372] Según un ejemplo útil para comprender la invención, el generador de coeficientes 1120, por ejemplo, puede estar configurado para generar uno o más segundos coeficientes de señal de audio mediante la aplicación de la fórmula:

$$f_{actual}[i] = \alpha \cdot f_{último}[i] + (1 - \alpha) \cdot pt_{media}[i]$$

donde $f_{actual}[i]$ indica uno de los uno o más segundos coeficientes de señal de audio, donde $f_{último}[i]$ indica uno de los uno o más primeros coeficientes de señal de audio, donde $pt_{media}[i]$ es uno de uno o más coeficientes de ruido, donde

α es un número real con $0 \leq \alpha \leq 1$, y donde i es un índice.

[0373] Según un ejemplo útil para comprender la invención, $f_{\text{ultimo}}[i]$ indica un coeficiente de filtro predictivo lineal de la señal de audio codificada, y donde $f_{\text{actua}}[i]$ indica un coeficiente de filtro predictivo lineal de la señal de audio
5 reconstruida.

[0374] En un ejemplo útil para comprender la invención, $pt_{\text{media}}[i]$, por ejemplo, puede ser un coeficiente de filtro predictivo lineal que indica el ruido de fondo de la señal de audio codificada.

10 **[0375]** Según un ejemplo útil para comprender la invención, el generador de coeficientes 1120 puede estar configurado, por ejemplo, para generar coeficientes de señal de audio de al menos 10 segundos como uno o más segundos coeficientes de señal de audio.

15 **[0376]** En un ejemplo útil para comprender la invención, el generador de coeficientes 1120 puede estar configurado, por ejemplo, para determinar, si la trama actual de una o más tramas es recibida por la interfaz receptora 1110 y si la trama actual recibida por la interfaz receptora 1110 no está alterada, uno o más coeficientes de ruido mediante la determinación de un espectro de ruido de la señal de audio codificada.

20 **[0377]** En lo que sigue, se considera el desvanecimiento del espectro de transformada de coseno discreta modificada (MDCT) a ruido blanco antes de la aplicación de la sustitución de ruido de dominio de frecuencia (FDNS).

25 **[0378]** En lugar de modificar aleatoriamente el signo de un bin de transformada de coseno discreta modificada (MDCT) (mezcla de signos), se llena el espectro completo con ruido blanco, al que se le da forma usando la sustitución de ruido de dominio de frecuencia (FDNS). A fin de evitar un cambio instantáneo en las características del espectro, se aplica una mezcla gradual entre la mezcla de signos y el relleno de ruido. La mezcla gradual puede realizarse de la siguiente manera:

```

para (i=0; i<L_trama; i++) {
    si (antiguo_x[i] != 0) {
30     x[i] = (1 - cum_amortiguamiento)*ruido[i] + cum_amortiguamiento *
        signo_aleatorio () * x_antiguo[i];
    }
}

```

35 donde:

40 cum_amortiguamiento es el factor de atenuación (absoluto) – disminuye de trama a trama, recomenzando en 1 y disminuyendo hacia 0;
x_antiguo es el espectro de la última trama recibida
signo_aleatorio devuelve 1 o -1
ruido contiene un vector aleatorio (ruido blanco) que se somete a escala de tal manera que su media cuadrática (RMS) es similar al último espectro bueno.

45 **[0379]** El término signo_aleatorio () *antiguo_x[i] caracteriza el proceso de mezcla de signos a fin de aleatorizar las fases y de ese modo evitar repeticiones de armónicos.

[0380] A continuación, podría efectuarse otra normalización del nivel de energía después de la mezcla gradual, a fin de garantizar que la energía de la operación de suma no se desvíe debido a la correlación de los dos vectores.

50 **[0381]** Según ejemplos útiles para comprender la invención, la primera unidad de reconstrucción 140 puede estar configurada, por ejemplo, para reconstruir la tercera porción de señal de audio dependiendo de la información de nivel de ruido y dependiendo de la primera porción de señal de audio. En un ejemplo útil para comprender la invención particular, la primera unidad de reconstrucción 140 puede estar configurada, por ejemplo, para reconstruir la tercera porción de señal de audio mediante la atenuación o amplificación de la primera porción de señal de audio.
55

60 **[0382]** En algunos ejemplos útiles para comprender la invención, la segunda unidad de reconstrucción 141 puede estar configurada, por ejemplo, para reconstruir la cuarta porción de señal de audio dependiendo de la información de nivel de ruido y dependiendo de la segunda porción de señal de audio. En un ejemplo útil para comprender la invención particular, la segunda unidad de reconstrucción 141 puede estar configurada, por ejemplo, para reconstruir la cuarta porción de señal de audio mediante la atenuación o amplificación de la segunda porción de señal de audio.

65 **[0383]** Con respecto al desvanecimiento anteriormente descrito del espectro de transformada de coseno discreta modificada (MDCT) a ruido blanco antes de la aplicación de la sustitución de ruido de dominio de frecuencia (FDNS), un ejemplo útil para comprender la invención más general es ilustrado mediante la figura 12.

[0384] La figura 12 ilustra un aparato para la decodificación de una señal de audio codificada de manera que se obtenga una señal de audio reconstruida según un ejemplo útil para comprender la invención.

5 **[0385]** El aparato comprende una interfaz receptora 1210 para la recepción de una o más tramas que comprenden información sobre una pluralidad de muestras de señal de audio de un espectro de señales de audio de la señal de audio codificada, y un procesador 1220 para la generación de la señal de audio reconstruida.

[0386] El procesador 1220 está configurado para generar la señal de audio reconstruida mediante el desvanecimiento de un espectro modificado a un espectro objetivo, si una trama actual no es recibida por la interfaz receptora 1210 o si la trama actual es recibida por la interfaz receptora 1210 pero es alterada, donde el espectro modificado comprende una pluralidad de muestras de señales modificadas, donde, para cada una de las muestras de señales modificadas del espectro modificado, un valor absoluto de dicha muestra de señal modificada es igual a un valor absoluto de una de las muestras de señal de audio del espectro de señales de audio.

15 **[0387]** Además, el procesador 1220 está configurado para no realizar el desvanecimiento del espectro modificado al espectro objetivo, si la trama actual de una o más tramas es recibida por la interfaz receptora 1210 y si la trama actual recibida por la interfaz receptora 1210 no está alterada.

20 **[0388]** Según un ejemplo útil para comprender la invención, el espectro objetivo es un espectro de tipo ruido.

[0389] En un ejemplo útil para comprender la invención, el espectro de tipo ruido representa ruido blanco.

[0390] Según un ejemplo útil para comprender la invención, se forma el espectro de tipo ruido.

25 **[0391]** En un ejemplo útil para comprender la invención, la forma del espectro de tipo ruido depende de un espectro de señales de audio de una señal previamente recibida.

[0392] Según un ejemplo útil para comprender la invención, el espectro de tipo ruido se forma dependiendo de la forma del espectro de señales de audio.

[0393] En un ejemplo útil para comprender la invención, el procesador 1220 emplea un factor de inclinación para dar forma al espectro de tipo ruido.

35 **[0394]** Según un ejemplo útil para comprender la invención, el procesador 1220 emplea la fórmula

$$\text{ruido_formado}[i] = \text{ruido} * \text{energía}(\text{factor_inclinación}, i/N)$$

donde N indica el número de muestras,

40 donde i es un índice,

donde $0 \leq i < N$, con $\text{factor_inclinación} > 0$,

45 donde energía es una función de potencia.

[0395] Si el $\text{factor_inclinación}$ es menor que 1, esto significa la atenuación con i creciente. Si el $\text{factor_inclinación}$ es mayor que 1, esto significa la amplificación con i creciente.

50 **[0396]** Según otro ejemplo útil para comprender la invención, el procesador 1220 puede emplear la fórmula

$$\text{ruido_formado}[i] = \text{ruido} * (1 + i / (N-1) * (\text{factor_inclinación}-1))$$

donde N indica el número de muestras,

55 donde i es un índice, donde $0 \leq i < N$,
con $\text{factor_inclinación} > 0$.

[0397] Según un ejemplo útil para comprender la invención, el procesador 1220 está configurado para generar el espectro modificado, mediante el cambio de un signo de una o más de las muestras de señal de audio del espectro de señales de audio, si la trama actual no es recibida por la interfaz receptora 1210 o si la trama actual recibida por la interfaz receptora 1210 está alterada.

[0398] En un ejemplo útil para comprender la invención, cada una de las muestras de señal de audio del espectro de señales de audio está representada por un número real, pero no por un número imaginario.

[0399] Según un ejemplo útil para comprender la invención, las muestras de señal de audio del espectro de señales de audio están representadas en un dominio de transformada de coseno discreta modificada.

5 **[0400]** En otro ejemplo útil para comprender la invención, las muestras de señal de audio del espectro de señales de audio están representadas en un dominio de transformada de seno discreta modificada. Según un ejemplo útil para comprender la invención, el procesador 1220 está configurado para generar el espectro modificado mediante el empleo de una función de signos aleatorios que, de forma aleatoria, o pseudo-aleatoria, emite un primer o un segundo valor.

10 **[0401]** En un ejemplo útil para comprender la invención, el procesador 1220 está configurado para desvanecer el espectro modificado al espectro objetivo mediante la disminución subsiguiente de un factor de atenuación.

15 **[0402]** Según un ejemplo útil para comprender la invención, el procesador 1220 está configurado para desvanecer el espectro modificado al espectro objetivo mediante el incremento subsiguiente de un factor de atenuación.

20 **[0403]** En un ejemplo útil para comprender la invención, si la trama actual no es recibida por la interfaz receptora 1210 o si la trama actual recibida por la interfaz receptora 1210 está alterada, el procesador 1220 está configurado para generar la señal de audio reconstruida mediante el empleo de la fórmula:

$$x[i] = (1 - \text{cum_amortiguamiento}) * \text{ruido}[i] + \text{cum_amortiguamiento} * \text{signo_aleatorio}() * x_antiguo[i]$$

25 donde i es un índice, donde $x[i]$ indica una muestra de la señal de audio reconstruida, donde $\text{cum_amortiguamiento}$ es un factor de atenuación, donde $x_antiguo[i]$ indica una de las muestras de señal de audio del espectro de señales de audio de la señal de audio codificada, donde $\text{signo_aleatorio}()$ retorna 1 o -1 , y donde ruido es un vector aleatorio que indica el espectro objetivo.

30 **[0404]** Algunos ejemplos útiles para comprender la invención continúan una operación de predicción a largo plazo (LTP) de la excitación codificada por la transformada (TCX). En tales ejemplos útiles para comprender la invención, la operación de predicción a largo plazo (LTP) de la excitación codificada por la transformada (TCX) es continuada durante el ocultamiento con los parámetros de la predicción a largo plazo (LTP) (retraso de predicción a largo plazo (LTP) y ganancia de predicción a largo plazo (LTP)) derivados de la última trama buena.

35 **[0405]** Las operaciones de predicción a largo plazo (LTP) pueden resumirse como:

- Alimentación de búfer de retardo de predicción a largo plazo (LTP) sobre la base de la salida previamente derivada.
- Sobre la base del retraso de predicción a largo plazo (LTP): selección de la porción de señal apropiada, del búfer de
- 40 retardo de predicción a largo plazo (LTP) que se usa como contribución de predicción a largo plazo (LTP) para dar forma a la señal actual.
- Reescalar esta contribución de predicción a largo plazo (LTP) mediante el uso de la ganancia de predicción a largo plazo (LTP).
- Sumar esta contribución reescalada de predicción a largo plazo (LTP) a la entrada de señal de predicción a largo
- 45 plazo (LTP) a fin de generar la señal de salida de predicción a largo plazo (LTP).

[0406] Podrían considerarse diferentes estrategias con respecto al tiempo, cuando se realiza la actualización del búfer de retardo de predicción a largo plazo (LTP):

50 Como la primera operación de predicción a largo plazo (LTP) en trama n usando la salida de la última trama $n-1$. Esto actualiza el búfer de retardo de predicción a largo plazo (LTP) en trama n que se va a usar durante el procesamiento de predicción a largo plazo (LTP) en trama n .

55 **[0407]** Como la última operación de predicción a largo plazo (LTP) en trama n usando la salida de la trama actual n . Esto actualiza el búfer de retardo de predicción a largo plazo (LTP) en trama n que se va a usar durante el procesamiento de predicción a largo plazo (LTP) en trama $n + 1$.

[0408] En lo que sigue, se considera el desacoplado del circuito de retroalimentación de la predicción a largo plazo (LTP) de la excitación codificada por la transformada (TCX).

60 **[0409]** El desacoplamiento del circuito de retroalimentación de la predicción a largo plazo (LTP) de la excitación codificada por la transformada (TCX) evita la introducción de ruido adicional (resultante de la sustitución de ruido aplicado a la señal de entrada de LPT) durante cada circuito de retroalimentación del decodificador de predicción a largo plazo (LTP) cuando se encuentra en modo de ocultamiento.

[0410] La figura 10 ilustra este desacoplamiento. En particular, la figura 10 representa el desacoplamiento del circuito de retroalimentación de la predicción a largo plazo (LTP) durante el ocultamiento ($bfi=1$).

[0411] La figura 10 ilustra un búfer de retardo 1020, un selector de muestras 1030, y un procesador de muestras 1040 (el procesador de muestras 1040 es indicado por la línea de puntos).

[0412] Hacia el tiempo, cuando se realiza la actualización del búfer de retardo de la predicción a largo plazo (LTP) 1020, algunos ejemplos útiles para comprender la invención proceden de la siguiente manera:

10 – Para la operación normal: a fin de actualizar el búfer de retardo de predicción a largo plazo (LTP) 1020, podría preferirse como la primera operación de predicción a largo plazo (LTP, conforme a sus siglas en inglés), ya que la señal de salida sumada habitualmente es almacenada de forma persistente. Con esta estrategia, puede omitirse un búfer dedicado.

15 – Para la operación desacoplada: a fin de actualizar el búfer de retardo de predicción a largo plazo (LTP), podría preferirse como la última operación de predicción a largo plazo (LTP), ya que la contribución de predicción a largo plazo (LTP) a la señal habitualmente solo es almacenada de forma temporal. Con esta estrategia, se preserva la señal de contribución de predicción a largo plazo (LTP) transitoriamente. La implementación de esta manera de este búfer de contribución de predicción a largo plazo (LTP) podría hacerse persistente.

20 **[0413]** Asumiendo que se usa esta última estrategia en cualquier caso (operación normal y ocultamiento), los ejemplos útiles para comprender la invención pueden implementar, por ejemplo, lo siguiente:

– Durante la operación normal: se usa la salida de señal de dominio de tiempo del decodificador de predicción a largo plazo (LTP) después de su adición a la señal de entrada de la predicción a largo plazo (LTP) para alimentar el búfer de retardo de predicción a largo plazo (LTP).

25 – Durante el ocultamiento: se usa la salida de señal de dominio de tiempo del decodificador de predicción a largo plazo (LTP) antes de su adición a la señal de entrada de predicción a largo plazo (LTP) para alimentar el búfer de retardo de predicción a largo plazo (LTP).

30 **[0414]** Algunos ejemplos útiles para comprender la invención desvanecen la ganancia de la predicción a largo plazo (LTP) de la excitación codificada por la transformada (TCX) hacia cero. En tal ejemplo útil para comprender la invención, la ganancia de la predicción a largo plazo (LTP) de la excitación codificada por la transformada (TCX) puede ser desvanecida, por ejemplo, hacia cero con un cierto factor de desvanecimiento adaptativo de señal. Esto, por ejemplo, puede ser realizado iterativamente, por ejemplo, según el siguiente pseudo-código:

```
35 ganancia = ganancia_pasada * amortiguamiento;
   [...]
   ganancia_pasada = ganancia;
```

40 donde:

ganancia es la ganancia del decodificador de predicción a largo plazo (LTP) de la excitación codificada por la transformada (TCX) aplicada en la trama actual;

45 ganancia_pasada es la ganancia del decodificador de predicción a largo plazo (LTP) de la excitación codificada por la transformada (TCX) aplicada en la trama previa;

amortiguamiento es el factor de desvanecimiento (relativo).

50 **[0415]** La figura 1d ilustra un aparato según un ejemplo útil para comprender la invención adicional, donde el aparato comprende adicionalmente una unidad de predicción a largo plazo 170 que comprende un búfer de retardo 180. La unidad de predicción a largo plazo 170 está configurada para generar una señal procesada dependiendo de la segunda porción de señal de audio, dependiendo de una entrada de búfer de retardo que está almacenada en el búfer de retardo 180 y dependiendo de una ganancia de predicción a largo plazo. Además, la unidad de predicción a largo plazo está configurada para desvanecer la ganancia de predicción a largo plazo hacia cero, si dicha tercera trama de la pluralidad de tramas no es recibida por la interfaz receptora 110 o si dicha tercera trama es recibida por la interfaz receptora 110 pero es alterada.

60 **[0416]** En otros ejemplos útiles para comprender la invención (no expuestos), la unidad de predicción a largo plazo puede estar configurada, por ejemplo, para generar una señal procesada dependiendo de la primera porción de señal de audio, dependiendo de una entrada de búfer de retardo que está almacenada en el búfer de retardo y dependiendo de una ganancia de predicción a largo plazo.

65 **[0417]** En la figura 1d, la primera unidad de reconstrucción 140 puede generar, por ejemplo, la tercera porción de señal de audio adicionalmente dependiendo de la señal procesada.

[0418] En un ejemplo útil para comprender la invención, la unidad de predicción a largo plazo 170 puede estar configurada, por ejemplo, para desvanecer la ganancia de predicción a largo plazo hacia cero, donde una velocidad con la cual la ganancia de predicción a largo plazo es desvanecida a cero depende de un factor de desvanecimiento.

5

[0419] De manera alternativa o adicional, la unidad de predicción a largo plazo 170 puede estar configurada, por ejemplo, para actualizar la entrada del búfer de retardo 180 mediante el almacenamiento de la señal procesada generada en el búfer de retardo 180 si dicha tercera trama de la pluralidad de tramas no es recibida por la interfaz receptora 110 o si dicha tercera trama es recibida por la interfaz receptora 110 pero es alterada.

10 **[0420]** Con respecto al uso anteriormente descrito de la predicción a largo plazo (LTP) de la excitación codificada por la transformada (TCX), un ejemplo útil más general para comprender la invención se ilustra en la figura 13.

15 **[0421]** La figura 13 ilustra un aparato para la decodificación de una señal de audio codificada de manera que se obtenga una señal de audio reconstruida.

20 **[0422]** El aparato comprende una interfaz receptora 1310 para la recepción de una pluralidad de tramas, un búfer de retardo 1320 para el almacenamiento de muestras de señal de audio de la señal de audio codificada, un selector de muestras 1330 para la selección de una pluralidad de muestras de señal de audio seleccionadas de las muestras de señal de audio almacenadas en el búfer de retardo 1320, y un procesador de muestras 1340 para el procesamiento de las muestras de señal de audio seleccionadas a fin de obtener muestras de la señal de audio reconstruida de la señal de audio reconstruida.

25 **[0423]** El selector de muestras 1330 está configurado para seleccionar, si una trama actual es recibida por la interfaz receptora 1310 y si la trama actual recibida por la interfaz receptora 1310 no está alterada, estando la pluralidad de muestras de señal de audio seleccionadas de las muestras de señal de audio almacenadas en el búfer de retardo 1320 dependiendo de una información de retraso de altura comprendida por la trama actual. Además, el selector de muestras 1330 está configurado para seleccionar, si la trama actual no es recibida por la interfaz receptora 1310 o si la trama actual recibida por la interfaz receptora 1310 está alterada, la pluralidad de muestras de señal de audio seleccionadas de las muestras de señal de audio almacenadas en el búfer de retardo 1320 dependiendo de una información de retraso de altura compuesta por otra trama recibida previamente por la interfaz receptora 1310.

35 **[0424]** Según un ejemplo útil para comprender la invención, el procesador de muestras 1340 puede estar configurado, por ejemplo, para obtener las muestras de señal de audio reconstruida, si la trama actual es recibida por la interfaz receptora 1310 y si la trama actual recibida por la interfaz receptora 1310 no está alterada, mediante el reescalado de las muestras de señal de audio seleccionadas dependiendo de la información de ganancia comprendida por la trama actual. Además, el selector de muestras 1330 puede estar configurado, por ejemplo, para obtener las muestras de señal de audio reconstruida, si la trama actual no es recibida por la interfaz receptora 1310 o si la trama actual recibida por la interfaz receptora 1310 está alterada, mediante el reescalado de las muestras de señal de audio seleccionadas según la información de ganancia comprendida por dicha otra trama recibida previamente por la interfaz receptora 1310.

45 **[0425]** En un ejemplo útil para comprender la invención, el procesador de muestras 1340 puede estar configurado, por ejemplo, para obtener las muestras de señal de audio reconstruida, si la trama actual es recibida por la interfaz receptora 1310 y si la trama actual recibida por la interfaz receptora 1310 no está alterada, multiplicando las muestras de señal de audio seleccionadas y un valor dependiendo de la información de ganancia comprendida por la trama actual. Además, el selector de muestras 1330 está configurado para obtener las muestras de señal de audio reconstruida, si la trama actual no es recibida por la interfaz receptora 1310 o si la trama actual recibida por la interfaz receptora 1310 está alterada, multiplicando las muestras de señal de audio seleccionadas y un valor dependiendo de la información de ganancia comprendida por dicha otra trama recibida previamente por la interfaz receptora 1310.

50 **[0426]** Según un ejemplo útil para comprender la invención, el procesador de muestras 1340 puede estar configurado, por ejemplo, para almacenar las muestras de señal de audio reconstruida en el búfer de retardo 1320.

55 **[0427]** En un ejemplo útil para comprender la invención, el procesador de muestras 1340 puede estar configurado, por ejemplo, para almacenar las muestras de señal de audio reconstruida en el búfer de retardo 1320 antes de que otra trama sea recibida por la interfaz receptora 1310.

60 **[0428]** Según un ejemplo útil para comprender la invención, el procesador de muestras 1340 puede estar configurado, por ejemplo, para almacenar las muestras de señal de audio reconstruida en el búfer de retardo 1320 después de que otra trama sea recibida por la interfaz receptora 1310.

65 **[0429]** En un ejemplo útil para comprender la invención, el procesador de muestras 1340 puede estar configurado, por ejemplo, para reescalar las muestras de señal de audio seleccionadas dependiendo de la información de ganancia a fin de obtener muestras de señal de audio reescaladas y mediante la combinación de las muestras de

señal de audio reescaladas con muestras de señal de audio de entrada para obtener las muestras de señal de audio procesadas.

5 **[0430]** Según un ejemplo útil para comprender la invención, el procesador de muestras 1340 puede estar configurado, por ejemplo, para almacenar las muestras de señal de audio procesadas, que indican la combinación de las muestras de señal de audio reescaladas y las muestras de señal de audio de entrada, en el búfer de retardo 1320, y para no almacenar las muestras de señal de audio reescaladas en el búfer de retardo 1320, si la trama actual es recibida por la interfaz receptora 1310 y si la trama actual recibida por la interfaz receptora 1310 no está alterada. Además, el procesador de muestras 1340 está configurado para almacenar las muestras de señal de audio
10 reescaladas en el búfer de retardo 1320 y para no almacenar las muestras de señal de audio procesadas en el búfer de retardo 1320, si la trama actual no es recibida por la interfaz receptora 1310 o si la trama actual recibida por la interfaz receptora 1310 está alterada.

15 **[0431]** Según otro ejemplo útil para comprender la invención, el procesador de muestras 1340 puede estar configurado, por ejemplo, para almacenar las muestras de señal de audio procesadas en el búfer de retardo 1320, si la trama actual no es recibida por la interfaz receptora 1310 o si la trama actual recibida por la interfaz receptora 1310 está alterada.

20 **[0432]** En un ejemplo útil para comprender la invención, el selector de muestras 1330 puede estar configurado, por ejemplo, para obtener las muestras de señal de audio reconstruida mediante el reescalado de las muestras de señal de audio seleccionadas dependiendo de una ganancia modificada, donde la ganancia modificada se define según la fórmula:

$$\text{ganancia} = \text{ganancia_pasada} * \text{amortiguamiento};$$

25 donde ganancia es la ganancia modificada, donde el selector de muestras 1330 puede estar configurado, por ejemplo, para establecer ganancia_pasada en la ganancia después de la ganancia, y se ha calculado, y donde el amortiguamiento es un número real.

30 **[0433]** Según un ejemplo útil para comprender la invención, el selector de muestras 1330 puede estar configurado, por ejemplo, para calcular la ganancia modificada.

[0434] En un ejemplo útil para comprender la invención, el amortiguamiento puede ser definido, por ejemplo, según: $0 < \text{amortiguamiento} < 1$.

35 **[0435]** Según un ejemplo útil para comprender la invención, la ganancia ganancia modificada puede ser establecida, por ejemplo, en cero, si al menos un número predefinido de tramas no ha sido recibido por la interfaz receptora 1310, desde que una última trama ha sido recibida por la interfaz receptora 1310.

40 **[0436]** En lo que sigue, se considera la velocidad de desvanecimiento. Existen varios módulos de ocultamiento que se aplican a una cierta clase de desvanecimiento. Aunque la velocidad de este desvanecimiento podría ser seleccionada de manera diferente a través de estos módulos, es beneficioso usar la misma velocidad de desvanecimiento para todos los módulos de ocultamiento para un núcleo (predicción lineal excitada por código algebraico (ACELP) o excitación codificada por la transformada (TCX)). Por ejemplo:

45 Para la predicción lineal excitada por código algebraico (ACELP), debe usarse la misma velocidad de desvanecimiento, en particular, para el libro de código adaptativo (alterando la ganancia), y/o para la señal de libro de código innovadora (alterando la ganancia).

50 **[0437]** Además, para la excitación codificada por la transformada (TCX), debe usarse la misma velocidad de desvanecimiento, en particular, para la señal de dominio de tiempo, y/o para la ganancia de la predicción a largo plazo (LTP) (disminuida a cero), y/o para el peso de la codificación predictiva lineal (LPC) (disminuida a uno), y/o para los coeficientes de LP (disminuidos a forma espectral de fondo), y/o para la mezcla gradual a ruido blanco.

55 **[0438]** Adicionalmente, podría ser preferible además usar la misma velocidad de desvanecimiento para la predicción lineal excitada por código algebraico (ACELP) y la excitación codificada por la transformada (TCX), si bien, debido a la diferente naturaleza de los núcleos, podría seleccionarse también el uso de diferentes velocidades de desvanecimiento.

60 **[0439]** Esta velocidad de desvanecimiento podría ser estática, si bien, preferentemente, es adaptativa a las características de la señal. Por ejemplo, la velocidad de desvanecimiento puede depender del factor de estabilidad de la codificación predictiva lineal (LPC) (excitación codificada por la transformada (TCX)) y/o de una clasificación, y/o de un número de tramas consecutivamente perdidas.

65 **[0440]** La velocidad de desvanecimiento, por ejemplo, puede ser determinada según el factor de atenuación, que podría ser proporcionado de forma absoluta o relativa, y que podría además cambiar en función del tiempo durante

un cierto desvanecimiento.

[0441] En ejemplos útiles para comprender la invención, se usa la misma velocidad de desvanecimiento para el desvanecimiento de la ganancia de predicción a largo plazo (LTP) que para el desvanecimiento de ruido blanco.

5

[0442] Se han proporcionado un aparato, un procedimiento y un programa informático para la generación de una señal de ruido de confort como se describe anteriormente.

[0443] Aunque algunos aspectos se han descrito en el contexto de un aparato, es claro que estos aspectos representan además una descripción del correspondiente procedimiento, donde un bloque o dispositivo corresponde a una etapa de procedimiento o un rasgo de una etapa de procedimiento. De forma análoga, los aspectos descritos en el contexto de una etapa de procedimiento representan además una descripción de un correspondiente bloque o elemento o rasgo de un aparato correspondiente.

[0444] La señal descompuesta de la invención puede ser almacenada en un medio de almacenamiento digital, o puede ser transmitida en un medio de transmisión tal como un medio de transmisión inalámbrico o un medio de transmisión por cable tal como Internet.

[0445] Dependiendo de ciertos requisitos de implementación, las realizaciones de la invención pueden ser implementadas en hardware o en software. La implementación puede realizarse usando un medio de almacenamiento digital, por ejemplo, un disquete, un DVD (disco versátil digital), un CD (disco compacto, conforme a sus siglas en inglés), una ROM (memoria de solo lectura, conforme a sus siglas en inglés), una PROM (memoria de solo lectura programable, conforme a sus siglas en inglés), una EPROM (memoria de solo lectura programable de borrado, conforme a sus siglas en inglés), una EEPROM (memoria de solo lectura programable de borrado electrónico, conforme a sus siglas en inglés) o una memoria FLASH, que tiene señales de control electrónicamente legibles allí almacenadas, que cooperan (o son capaces de cooperar) con un sistema informático programable de tal manera que se lleva a cabo el respectivo procedimiento.

[0446] Algunas realizaciones según la invención comprenden un soporte de datos no transitorio que tiene señales de control electrónicamente legibles, que son capaces de cooperar con un sistema informático programable, de manera que se lleve a cabo uno de los procedimientos que se describen en esta invención.

[0447] En general, las realizaciones de la presente invención pueden ser implementadas como un producto de programa informático con un código de programa, siendo el código de programa operativo para llevar a cabo uno de los procedimientos cuando se ejecuta el producto de programa informático en un ordenador. El código de programa puede ser almacenado, por ejemplo, en un soporte legible en una máquina.

[0448] Otras realizaciones comprenden el programa informático para llevar a cabo uno de los procedimientos que se describen en esta invención, almacenado en un soporte legible en una máquina.

40

[0449] En otras palabras, una realización del procedimiento de la invención es, por lo tanto, un programa informático que tiene un código de programa para la realización de uno de los procedimientos que se describen en esta invención, cuando se ejecuta el programa informático en un ordenador.

[0450] Una realización adicional de los procedimientos de la invención es, por lo tanto, un soporte de datos (o un medio de almacenamiento digital, o un medio legible por ordenador) que comprende, allí grabado, el programa informático para llevar a cabo uno de los procedimientos que se describen en esta invención.

[0451] Una realización adicional del procedimiento de la invención es, por lo tanto, una corriente de datos o una secuencia de señales que representan el programa informático para llevar a cabo uno de los procedimientos que se describen en esta invención. La corriente de datos o la secuencia de señales puede estar configurada, por ejemplo, para ser transferida por medio de una conexión de comunicación de datos, por ejemplo, por medio de Internet.

[0452] Una realización adicional comprende un medio de procesamiento, por ejemplo, un ordenador, o un dispositivo lógico programable, configurado o adaptado para llevar a cabo uno de los procedimientos que se describen en esta invención.

[0453] Una realización adicional comprende un ordenador que tiene instalado en sí el programa informático para llevar a cabo uno de los procedimientos que se describen en esta invención.

60

[0454] En algunas realizaciones, puede usarse un dispositivo lógico programable (por ejemplo, una matriz de puertas programables de campo) para realizar algunas o la totalidad de las funcionalidades de los procedimientos que se describen en esta invención. En algunas realizaciones, una matriz de puertas programables de campo puede cooperar con un microprocesador a fin de llevar a cabo uno de los procedimientos que se describen en esta invención.

[0454] En general, los procedimientos son llevados a cabo, preferentemente, por cualquier aparato de soporte físico.

65

[0455] Las realizaciones descritas con anterioridad son meramente ilustrativas de los principios de la presente invención. Se entiende que las modificaciones y variaciones de las disposiciones y los detalles que se describen en esta invención serán evidentes para los expertos en la materia. Por lo tanto, se tiene la intención de limitación 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 de esta invención.

Referencias

10 [0456]

[3GP09a] 3GPP; Technical Specification Group Services and System Aspects, *Extended adaptive multi-rate - wideband (AMR-WB+) codec*, 3GPP TS 26.290, 3rd Generation Partnership Project, 2009.

15 [3GP09b] *Extended adaptive multi-rate - wideband (AMR-WB+) codec; floating-point ANSI-C code*, 3GPP TS 26.304, 3rd Generation Partnership Project, 2009.

[3GP09c] *Speech codec speech processing functions; adaptive multi-rate - wideband (AMRWB) speech codec; transcoding functions*, 3GPP TS 26.190, 3rd Generation Partnership Project, 2009.

20

[3GP12a] *Adaptive multi-rate (AMR) speech codec; error concealment of lost frames (release 11)*, 3GPP TS 26.091, 3rd Generation Partnership Project, sept. de 2012.

25 [3GP12b] *Adaptive multi-rate (AMR) speech codec; transcoding functions (release 11)*, 3GPP TS 26.090, 3rd Generation Partnership Project, sept. de 2012. [3GP12c] , *ANSI-C code for the adaptive multi-rate - wideband (AMR-WB) speech codec*, 3GPP TS 26.173, 3rd Generation Partnership Project, sept. de 2012.

[3GP12d] *ANSI-C code for the floating-point adaptive multi-rate (AMR) speech codec (release 11)*, 3GPP TS 26.104, 3rd Generation Partnership Project, sept. de 2012.

30

[3GP12e] *General audio codec audio processing functions; Enhanced aacPlus general audio codec; additional decoder tools (release 11)*, 3GPP TS 26.402, 3rd Generation Partnership Project, sept. de 2012.

35 [3GP12f] *Speech codec speech processing functions; adaptive multi-rate - wideband (amr-wb) speech codec; ansi-c code*, 3GPP TS 26.204, 3rd Generation Partnership Project, 2012.

[3GP12g] *Speech codec speech processing functions; adaptive multi-rate - wideband (AMR-WB) speech codec; error concealment of erroneous or lost frames*, 3GPP TS 26.191, 3rd Generation Partnership Project, sept. de 2012.

40 [BJH06] I. Batina, J. Jensen y R. Heusdens, *Noise power spectrum estimation for speech enhancement using an autoregressive model for speech power spectrum dynamics*, en Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. 3 (2006), 1064–1067.

45 [BP06] A. Borowicz y A. Petrovsky, *Minima controlled noise estimation for klt-based speech enhancement*, CD-ROM, 2006, Italia, Florencia.

[Coh03] I. Cohen, *Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging*, IEEE Trans. Speech Audio Process. 11 (2003), n. ° 5, 466–475.

50 [CPK08] Choong Sang Cho, Nam In Park y Hong Kook Kim, *A packet loss concealment algorithm robust to burst packet loss for celp-type speech coders*, Tech. report, Korea Electronics Technology Institute, Gwang Institute of Science and Technology, 2008, The 23rd International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC 2008). [Informe técnico, Instituto de Tecnología Electrónica de Korea, Instituto Gwang de Ciencia y Tecnología, 2008, 23 Conferencia Técnica Internacional sobre Circuitos/Sistemas, Ordenadores y Comunicaciones].

55

[Dob95] G. Doblinger, *Computationally efficient speech enhancement by spectral minima tracking in subbands*, en Proc. Eurospeech (1995), 1513–1516.

60 [EBU10] EBU/ETSI JTC Broadcast, *Digital audio broadcasting (DAB); transport of advanced audio coding (AAC) audio*, ETSI TS 102 563, European Broadcasting Union, mayo de 2010.

[EBU12] *Digital radio mondiale (DRM); system specification*, ETSI ES 201 980, ETSI, junio de 2012.

65 [EH08] Jan S. Erkelens y Richards Heusdens, *Tracking of Nonstationary Noise Based on Data-Driven Recursive*

- Noise Power Estimation*, Audio, Speech, and Language Processing, IEEE Transactions on 16 (2008), n. ° 6, 1112–1123.
- 5 [EM84] Y. Ephraim y D. Malah, *Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator*, IEEE Trans. Acoustics, Speech and Signal Processing 32 (1984), n. ° 6, 1109–1121.
- [EM85] *Speech enhancement using a minimum mean-square error log-spectral amplitude estimator*, IEEE Trans. Acoustics, Speech and Signal Processing 33 (1985), 443–445.
- 10 [Gan05] S. Gannot, *Speech enhancement: Application of the kalman filter in the estimate-maximize (em framework)*, Springer, 2005.
- [HE95] H. G. Hirsch y C. Ehrlicher, *Noise estimation techniques for robust speech recognition*, Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing, n. ° págs. 153-156, IEEE, 1995.
- 15 [HHJ10] Richard C. Hendriks, Richard Heusdens y Jesper Jensen, *MMSE based noise PSD tracking with low complexity*, Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference (Conferencia Internacional IEEE), marzo de 2010, págs. 4266–4269.
- 20 [HJH08] Richard C. Hendriks, Jesper Jensen y Richard Heusdens, *Noise tracking using dft domain subspace decompositions*, IEEE Trans. Audio, Speech, Lang. Process. 16 (2008), n. ° 3, 541–553.
- [IET12] IETF, *Definition of the Opus Audio Codec*, Tech. Report RFC 6716, Internet Engineering Task Force, sept. de 2012.
- 25 [ISO09] ISO/IEC JTC1/SC29/WG11, *Information technology – coding of audio-visual objects – part 3: Audio*, ISO/IEC IS 14496-3, International Organization for Standardization [Organización Internacional de Normalización], 2009.
- 30 [ITU03] ITU-T, *Wideband coding of speech at around 16 kbit/s using adaptive multi-rate wideband (amr-wb)*, Recommendation ITU-T G.722.2, Telecommunication Standardization Sector of ITU [Recomendación ITU–T, G.722.2, Sector de Normalización de Telecomunicaciones de ITU (Unión Internacional de Telecomunicaciones)], julio de 2003.
- 35 [ITU05] *Low-complexity coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss*, Recommendation ITU-T G.722.1, Telecommunication Standardization Sector of ITU [Recomendación ITU-T G.722.1, Sector de Normalización de Telecomunicaciones de la ITU], mayo de 2005.
- 40 [ITU06a] *G.722 Appendix III: A high-complexity algorithm for packet loss concealment for G.722*, ITU-T Recommendation, Recomendación ITU–T, ITU-T, nov. de 2006.
- [ITU06b] *G.729.1: G.729-based embedded variable bit-rate coder: An 8-32 kbit/s scalable wideband coder bitstream interoperable with g.729*, Recommendation ITU-T G.729.1, Telecommunication Standardization Sector of ITU [Recomendación ITU-T G.729.1, Sector de Normalización de Telecomunicaciones de la ITU], mayo de 2006.
- 45 [ITU07] *G.722 Appendix IV: A low-complexity algorithm for packet loss concealment with G.722*, Recomendación ITU–T, ITU-T, agosto de 2007.
- [ITU08a] *G.718: Frame error robust narrow-band and wideband embedded variable bit-rate coding of speech and audio from 8-32 kbit/s*, Recommendation ITU-T G.718, Telecommunication Standardization Sector of ITU [Recomendación ITU-T G.718, Sector de Normalización de Telecomunicaciones de la ITU], junio de 2008.
- 50 [ITU08b] *G.719: Low-complexity, full-band audio coding for high-quality, conversational applications*, Recommendation ITU-T G.719, Telecommunication Standardization Sector of ITU [Recomendación ITU-T G.719, Sector de Normalización de Telecomunicaciones de la ITU], junio de 2008.
- 55 [ITU12] *G.729: Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear prediction (cs-acelp)* Recommendation ITU-T G.729, Telecommunication Standardization Sector of ITU [Recomendación ITU-T G.729, Sector de Normalización de Telecomunicaciones de la ITU], junio de 2012.
- 60 [LS01] Pierre Lauber y Ralph Sperschneider, *Error concealment for compressed digital audio*, Audio Engineering Society Convention 111, n. ° 5460, sept. de 2001.
- 65 [Mar01] Rainer Martin, *Noise power spectral density estimation based on optimal smoothing and minimum statistics*, IEEE Transactions on Speech and Audio Processing 9 (2001), n. ° 5, 504–512.

- [Mar03] *Statistical methods for the enhancement of noisy speech*, International Workshop on Acoustic Echo and Noise Control (IWAENC2003), Technical University of Braunschweig, sept. de 2003.
- 5 [MC99] R. Martin y R. Cox, *New speech enhancement techniques for low bit rate speech coding*, in Proc. IEEE Workshop on Speech Coding (1999), 165–167.
- [MCA99] D. Malah, R. V. Cox y A. J. Accardi, *Tracking speech-presence uncertainty to improve speech enhancement in nonstationary noise environments*, Proc. IEEE Int. Conf. on Acoustics Speech and Signal Processing (1999), 789–792.
- 10 [MEP01] Nikolaus Meine, Bernd Edler y Heiko Purnhagen, *Error protection and concealment for HILN MPEG-4 parametric audio coding*, Audio Engineering Society Convention 110, n. ° 5300, mayo de 2001.
- 15 [MPC89] Y. Mahieux, J.-P. Petit y A. Charbonnier, *Transform coding of audio signals using correlation between successive transform blocks*, Acoustics, Speech, and Signal Processing, 1989. ICASSP-89., 1989 International Conference on, 1989, págs. 2021–2024 vol.3.
- [NMR+12] Max Neuendorf, Markus Multrus, Nikolaus Rettelbach, Guillaume Fuchs, Julien Robilliard, Jérémie Lecomte, Stephan Wilde, Stefan Bayer, Sascha Disch, Christian Helmrich, Roch Lefebvre, Philippe Gournay, Bruno Bessette, Jimmy Lapierre, Kristopher Kjörling, Heiko Purnhagen, Lars Villemoes, Werner Oomen, Erik Schuijers, Kei Kikuri, Toru Chinen, Takeshi Norimatsu, Chong Kok Seng, Eunmi Oh, Miyoung Kim, Schuyler Quackenbush y Berndhard Grill, *MPEG Unified Speech and Audio Coding - The ISO / MPEG Standard for High-Efficiency Audio Coding of all Content Types*, Convention Paper 8654, AES, abril de 2012, presentado en la 132 Convención de Budapest, Hungría.
- 20 [PKJ+11] Nam In Park, Hong Kook Kim, Min A Jung, Seong Ro Lee, y Seung Ho Choi, *Burst packet loss concealment using multiple codebooks and comfort noise for celp-type speech coders in wireless sensor networks*, Sensors 11 (2011), 5323–5336.
- 25 [QD03] Schuyler Quackenbush and Peter F. Driessen, *Error mitigation in MPEG-4 audio packet communication systems*, Audio Engineering Society Convention 115, n. ° 5981, oct. de 2003.
- 30 [RL06] S. Rangachari y P. C. Loizou, *A noise-estimation algorithm for highly non-stationary environments*, Speech Commun. 48 (2006), 220–231.
- 35 [SFB00] V. Stahl, A. Fischer y R. Bippus, *Quantile based noise estimation for spectral subtraction and wiener filtering*, in Proc. IEEE Int. Conf. Acoust., Speech and Signal Process. (2000), 1875–1878.
- 40 [SS98] J. Sohn y W. Sung, *A voice activity detector employing soft decision based noise spectrum adaptation*, Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing, n. ° págs. 365-368, IEEE, 1998.
- [Yu09] Rongshan Yu, *A low-complexity noise estimation algorithm based on smoothing of noise power estimation and estimation bias correction*, Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference(Conferencia Internacional IEE) en, abril de 2009, págs. 4421–4424.
- 45

REIVINDICACIONES

1. Un aparato para transformar la decodificación de excitación codificada de una señal de audio codificada a fin de obtener una señal de audio reconstruida, donde el aparato está configurado para recibir una pluralidad de 5 tramas, y donde el aparato comprende:

una unidad de predicción a largo plazo para llevar a cabo una predicción a largo plazo, que comprende:

10 un búfer de retardo (1020) para almacenar las muestras de señales de audio de la señal de audio decodificada, un selector de muestras (1030) para la selección de una pluralidad de muestras de señales de audio seleccionadas de las muestras de señales de audio que se almacenan en el búfer de retardo (1020), y un procesador de muestras (1040) para el procesamiento de las muestras de señales de audio seleccionadas para obtener muestras de señales de audio reconstruidas de la señal de audio reconstruida,

15 donde el selector de muestras (1030) está configurado para seleccionar, si una trama actual es recibida por el aparato y si la trama actual recibida por el aparato no está alterada, la pluralidad de muestras de señales de audio seleccionadas de las muestras de señales de audio que se almacenan en el búfer de retardo (1020) dependiendo de una información de retardo de tono que está comprendida por la trama actual, y
 20 donde el selector de muestras (1030) está configurado para seleccionar, si la trama actual no es recibida por el aparato o si la trama actual recibida por el aparato está alterada, la pluralidad de muestras de señales de audio seleccionadas de las muestras de señales de audio que se almacenan en el búfer de retardo (1020) dependiendo de una información de retardo de tono que está comprendida por otra trama recibida previamente por el aparato, donde el selector de muestras (1030) está configurado para obtener las muestras de señales de audio reconstruidas mediante el reescalado de las muestras de señales de audio seleccionadas dependiendo de una
 25 ganancia de decodificador de predicción a largo plazo de excitación codificada transformada modificada, donde la ganancia modificada se define según la fórmula:

$$\text{ganancia} = \text{ganancia_pasada} * \text{amortiguamiento};$$

30 donde la ganancia es la ganancia de decodificador de predicción a largo plazo de excitación codificada transformada modificada, donde el selector de muestras (1030) está configurado para establecer ganancia_pasada en la ganancia después de que se haya calculado la ganancia, donde ganancia_pasada es la ganancia de decodificador de predicción a largo plazo de excitación codificada transformada aplicada en la trama anterior, y
 35 donde el amortiguamiento es un valor real con $0 \leq \text{amortiguamiento} \leq 1$.

2. Un aparato según la reivindicación 1, donde el selector de muestras (1030) está configurado para calcular la ganancia de decodificador de predicción a largo plazo de excitación codificada transformada modificada.

3. Un aparato según la reivindicación 1 o 2, donde la ganancia de ganancia de decodificador de predicción a largo plazo de excitación codificada transformada modificada se establece en cero, si al menos un número predefinido de tramas no ha sido recibido por el aparato desde que una trama ha sido recibida por último por el aparato.

4. Un procedimiento para la decodificación de excitación codificada transformada de una señal de audio codificada a fin de obtener una señal de audio reconstruida, donde el procedimiento comprende:

la recepción de una pluralidad de tramas,
 la realización de una predicción a largo plazo mediante

50 el almacenamiento de las muestras de señales de audio de la señal de audio decodificada, la selección de una pluralidad de muestras de señales de audio seleccionadas de las muestras de señales de audio que se almacenan en un búfer de retardo (1020), y el procesamiento de las muestras de señales de audio seleccionadas para obtener muestras de señales de audio reconstruidas de la señal de audio reconstruida,

60 donde, si se recibe una trama actual y si la trama actual que se recibe no está alterada, se lleva a cabo la etapa de la selección de la pluralidad de muestras de señales de audio seleccionadas de las muestras de señales de audio que se almacenan en el búfer de retardo (1020) dependiendo de una información de retardo de tono que está comprendida por la trama actual, y
 donde, si no se recibe la trama actual o si la trama actual que se recibe está alterada, se lleva a cabo la etapa de la selección de la pluralidad de muestras de señales de audio seleccionadas de las muestras de señales de audio que se almacenan en el búfer de retardo (1020) dependiendo de una información de retardo de tono que está comprendida por otra trama que se recibe previamente,
 65 donde el procedimiento comprende además la etapa de reescalado de las muestras de señales de audio

seleccionadas dependiendo de una ganancia de decodificador de predicción a largo plazo de excitación codificada transformada modificada, donde la ganancia de decodificador de predicción a largo plazo de excitación codificada transformada modificada está definida según la fórmula:

5 $ganancia = ganancia_pasada * amortiguamiento;$

donde ganancia es la ganancia de decodificador de predicción a largo plazo de excitación codificada transformada modificada,

10 donde ganancia_pasada se establece a ganancia después de que se haya calculado la ganancia, donde ganancia_pasada es la ganancia de decodificador de predicción a largo plazo de excitación codificada transformada aplicada en la trama anterior, y

donde el amortiguamiento es un valor real con $0 \leq amortiguamiento \leq 1$.

5. Un programa informático para la implementación del procedimiento de la reivindicación 4 cuando se
15 ejecuta en un ordenador o procesador de señal.

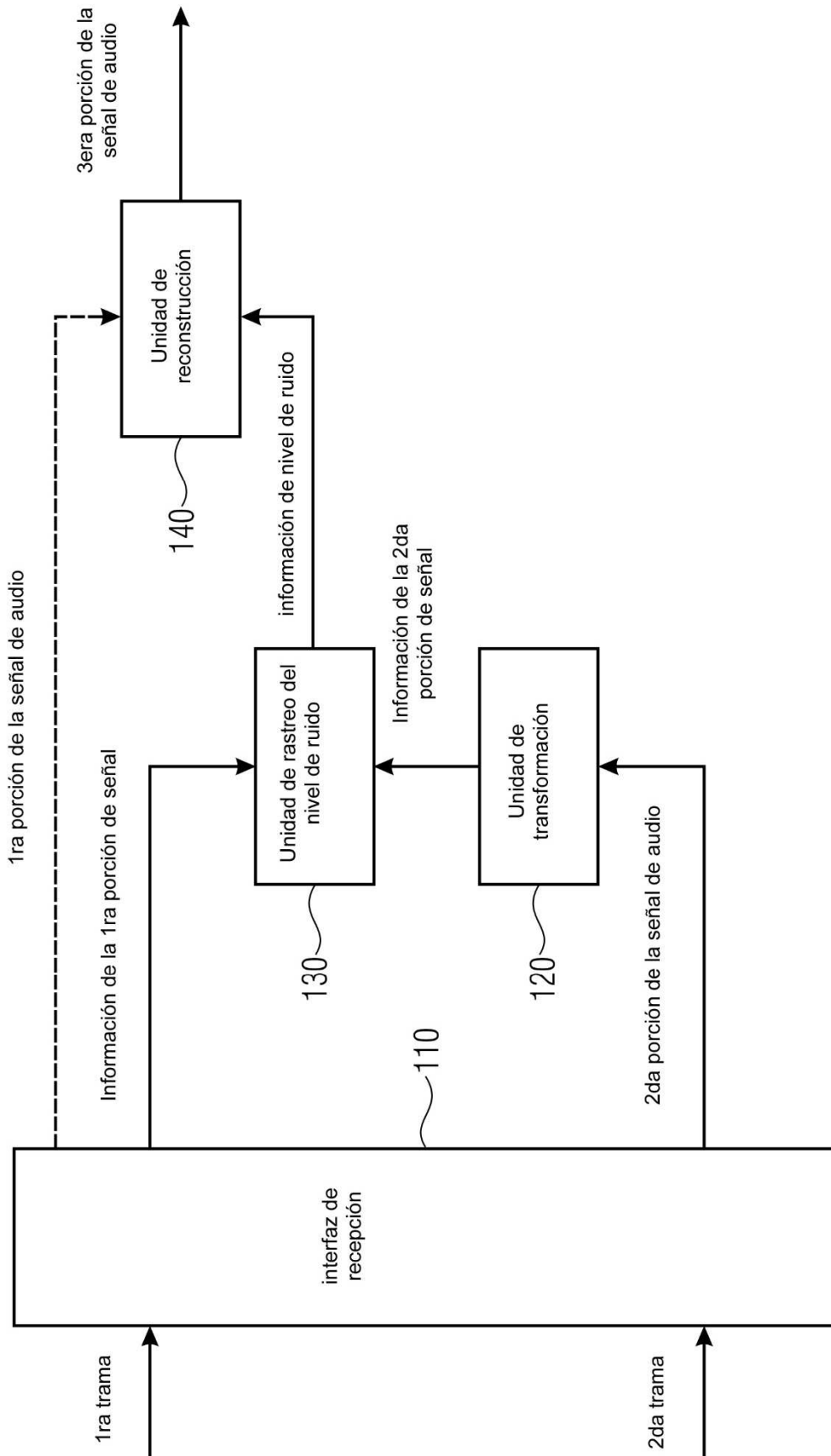


FIG 1A

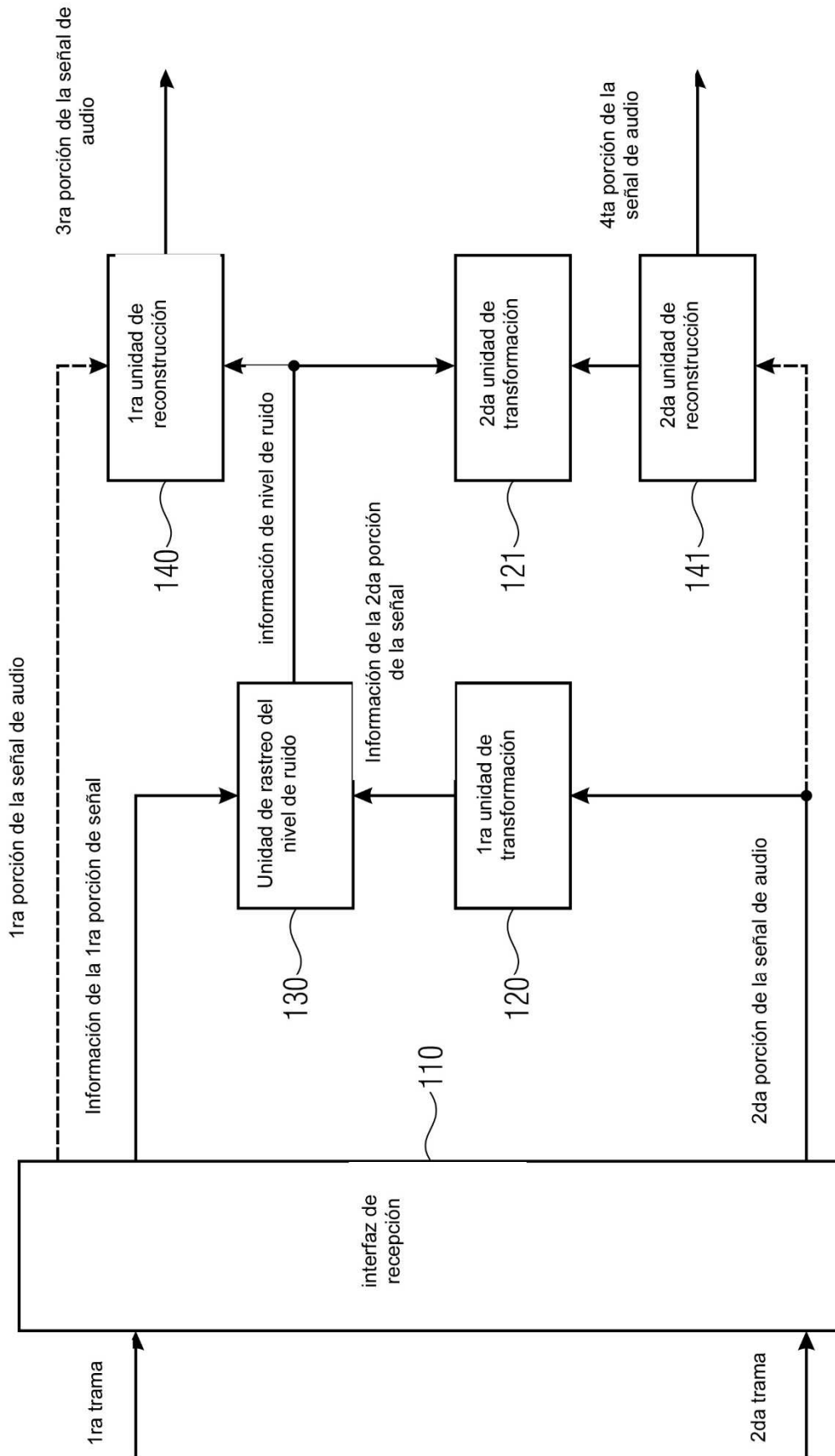


FIG 1B

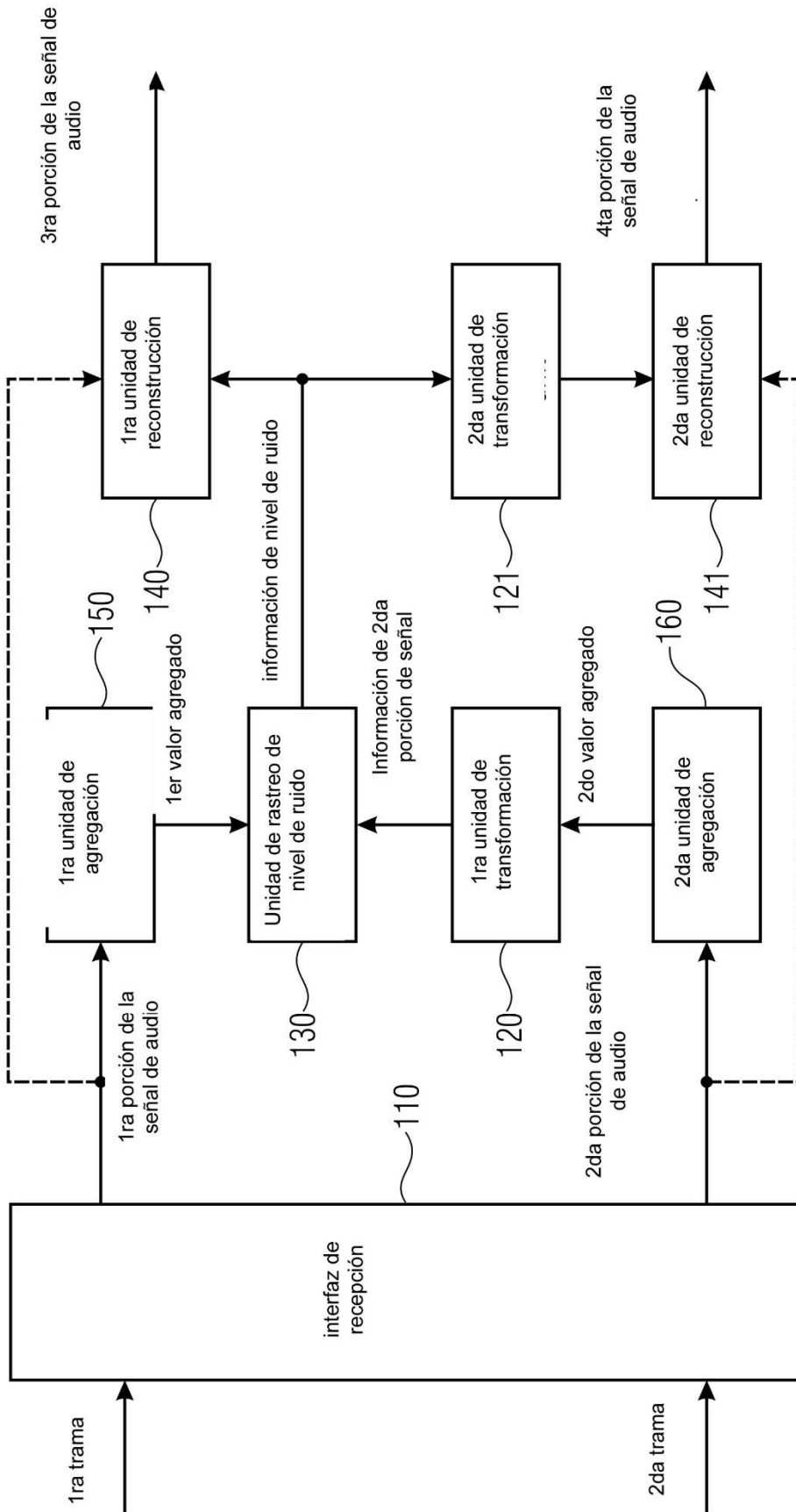


FIG 1C

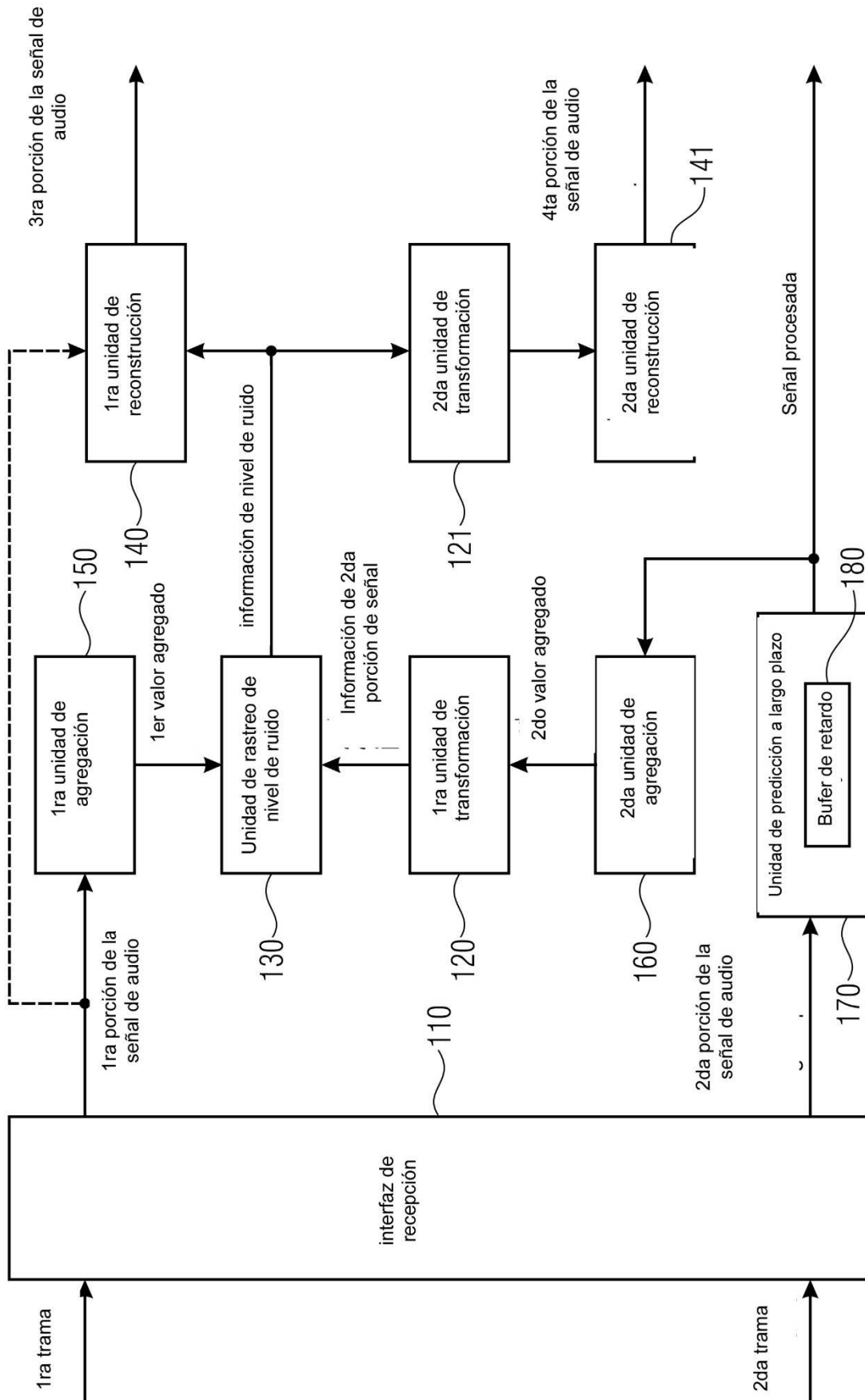


FIG 1D

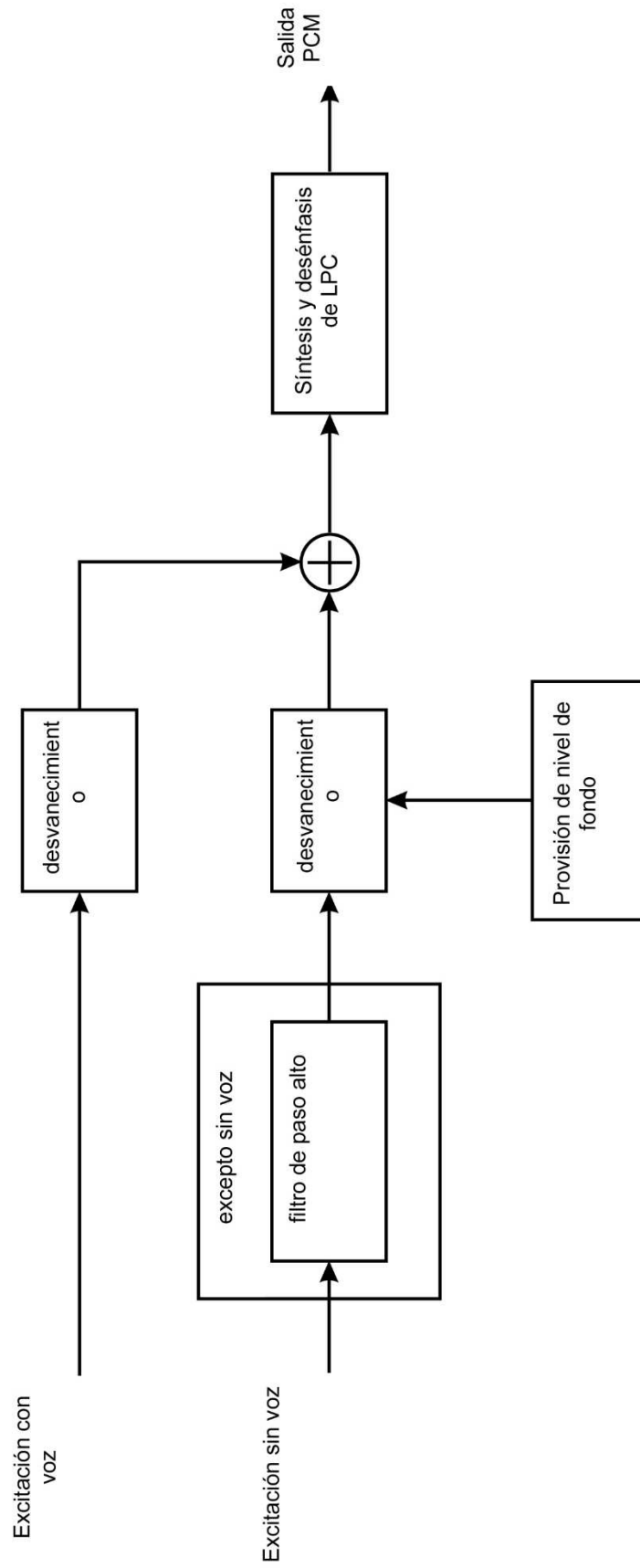
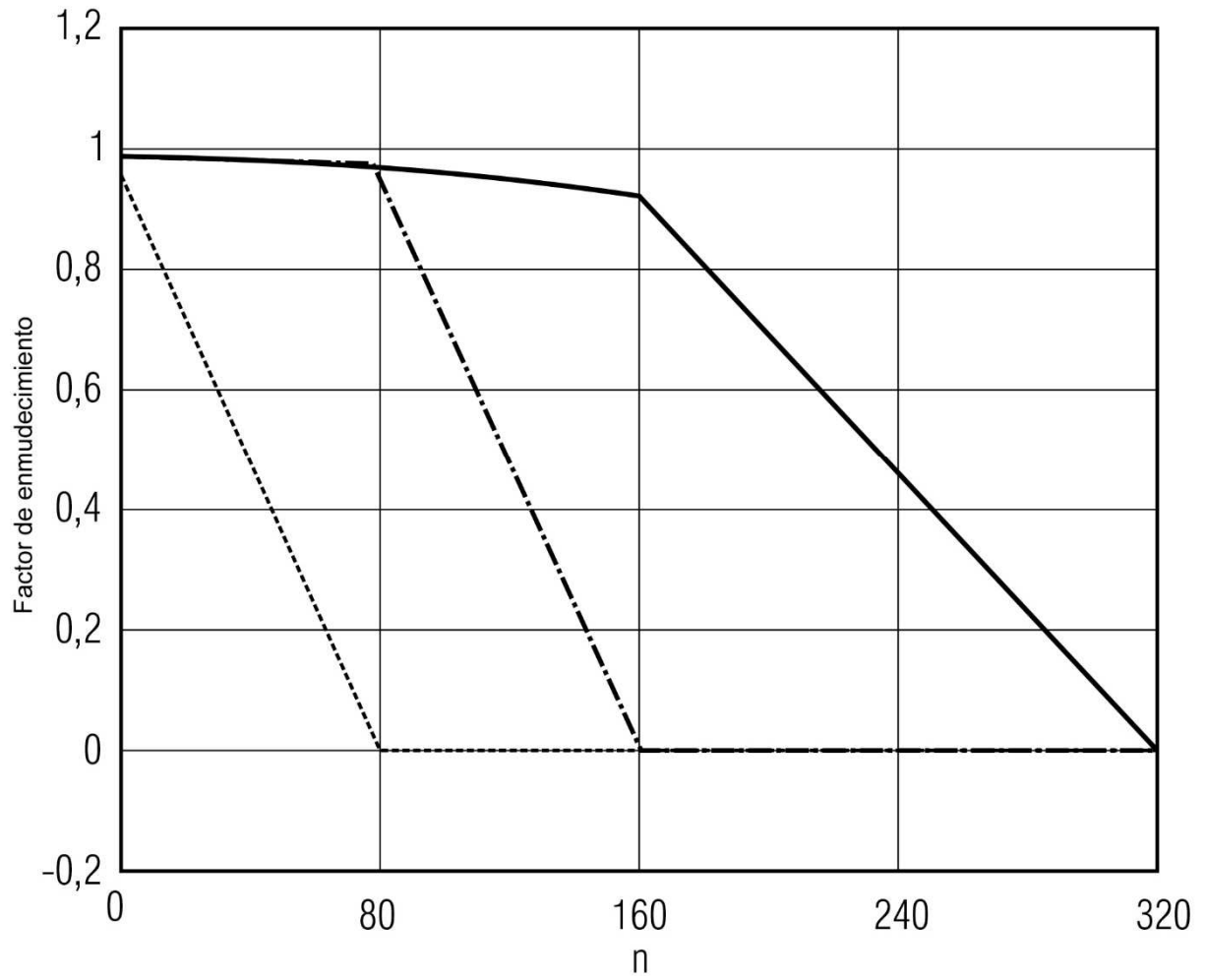


FIG 2



G.722APP1V_F07

- transitorio
- . - . VUV_TRANSITION
- Otras clases

FIG 3

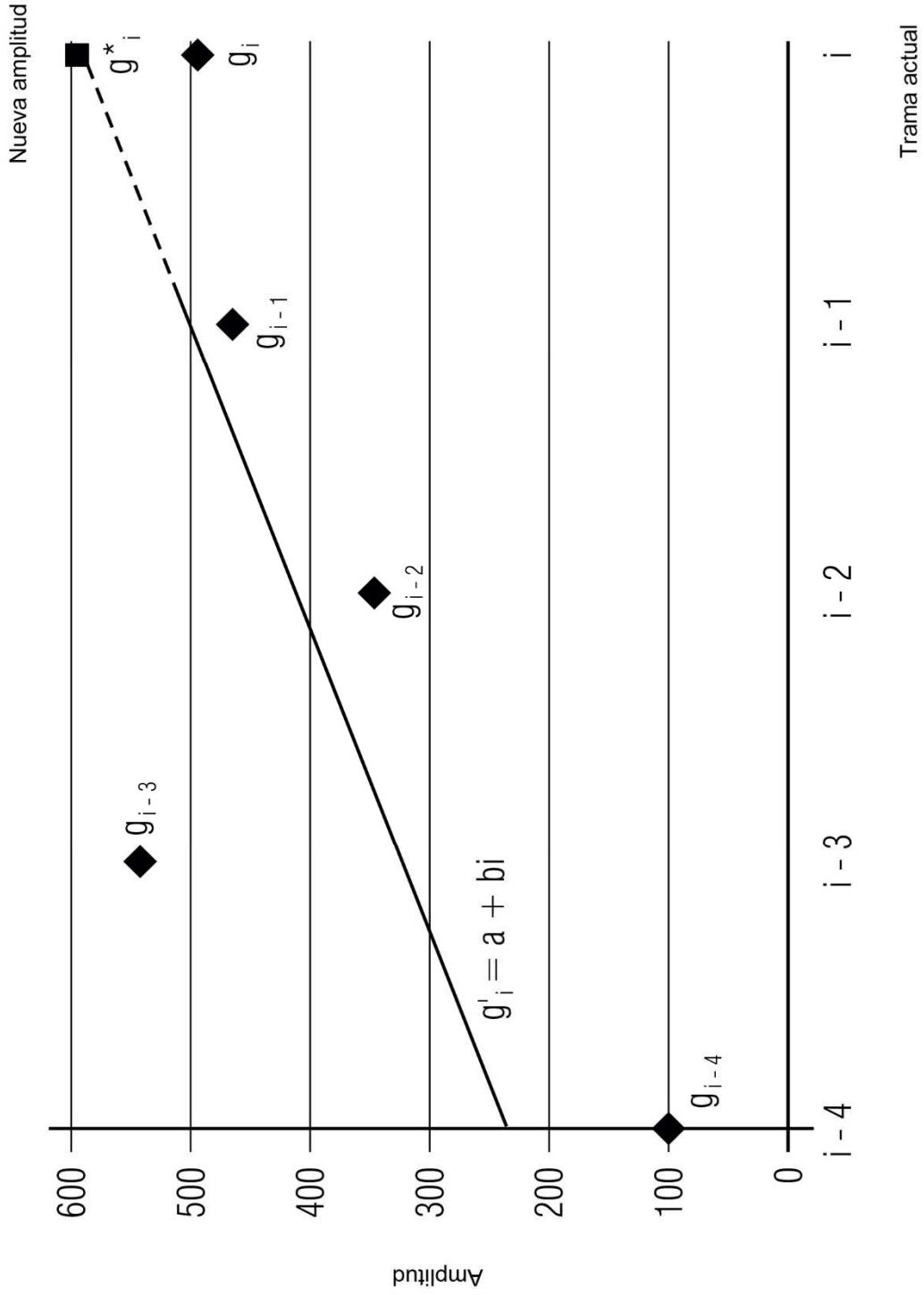


FIG 4

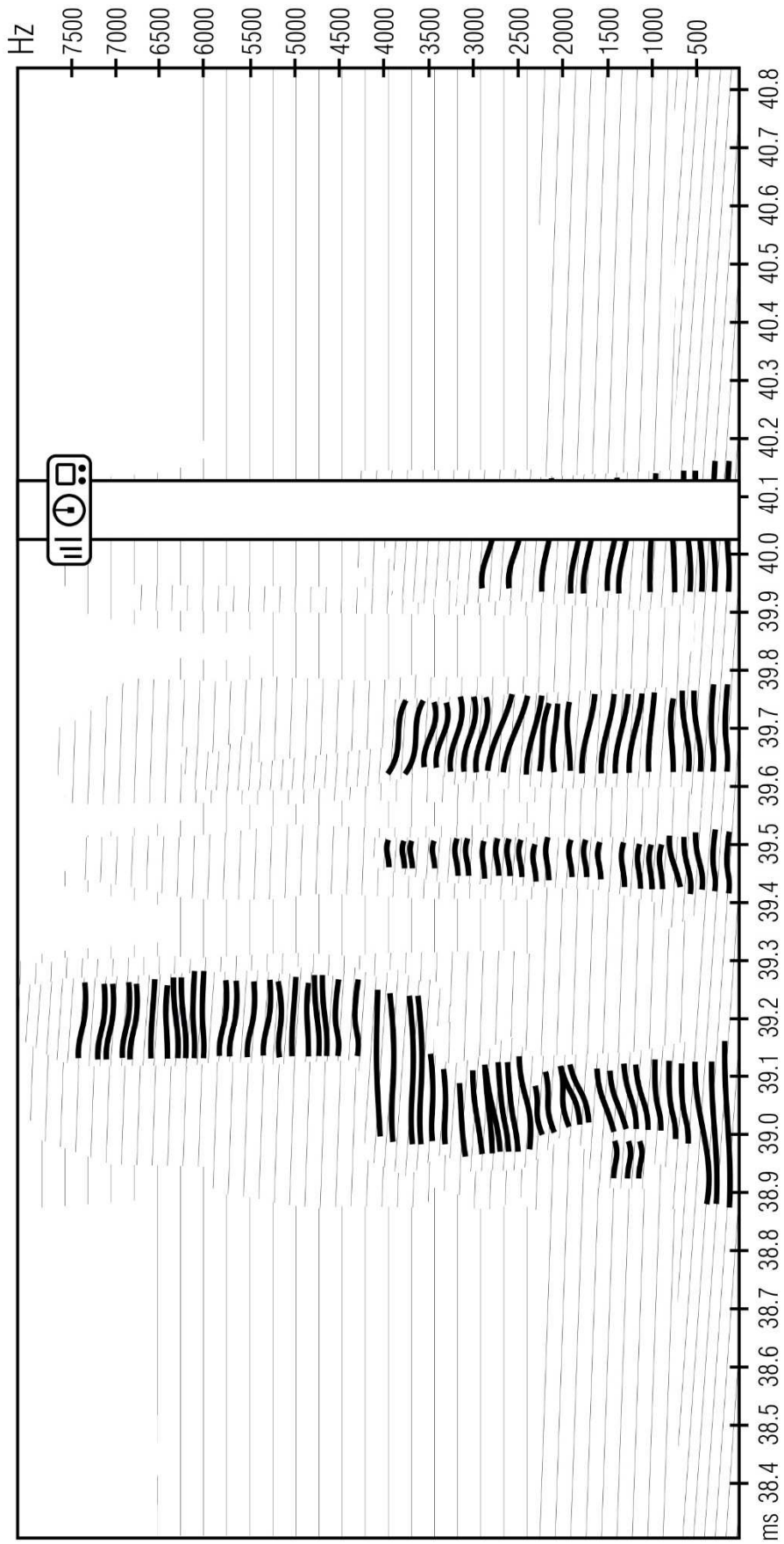


FIG 5

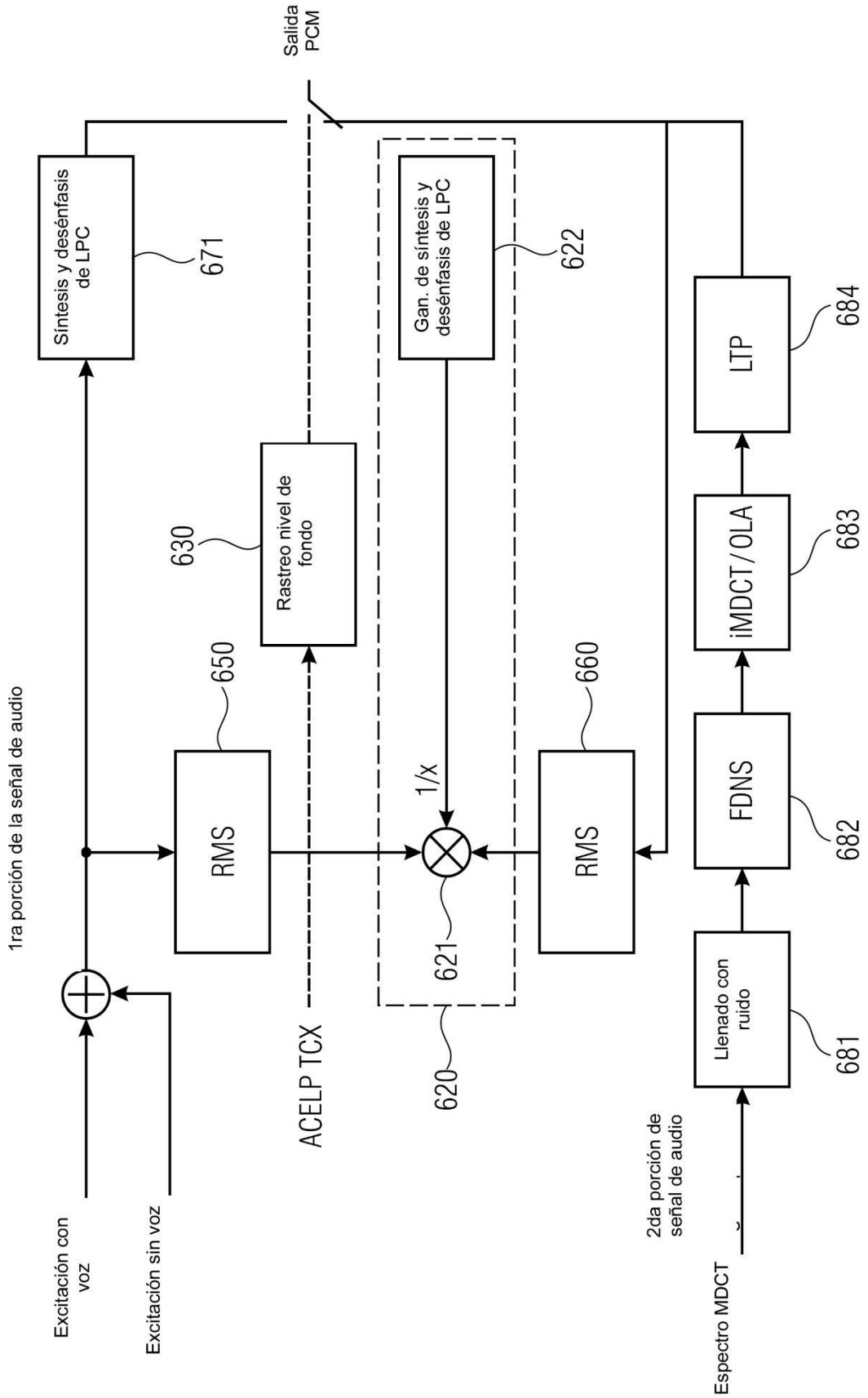


FIG 6

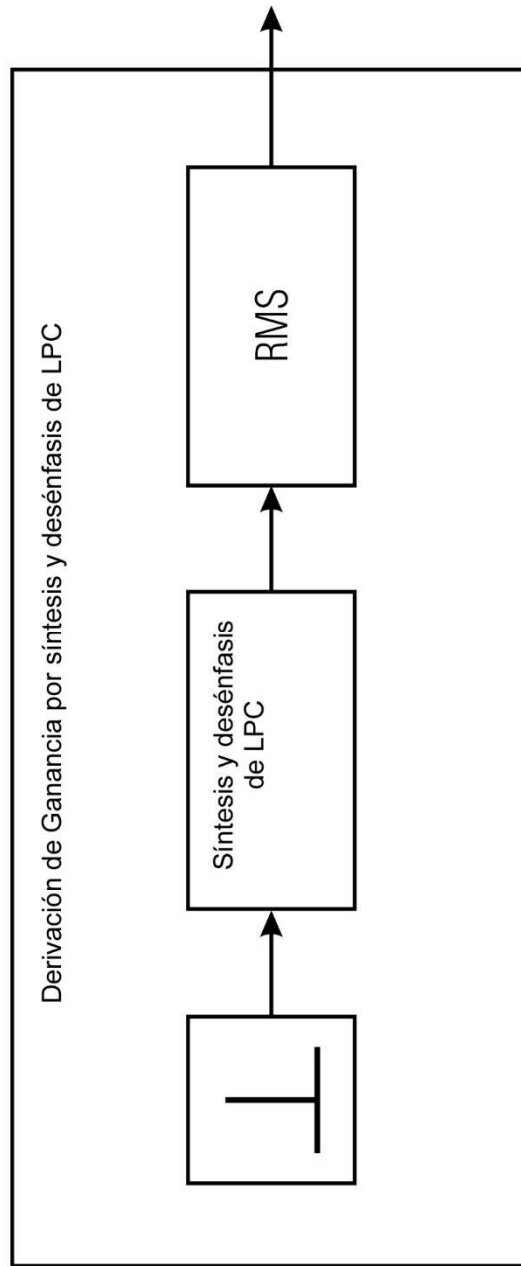


FIG 7

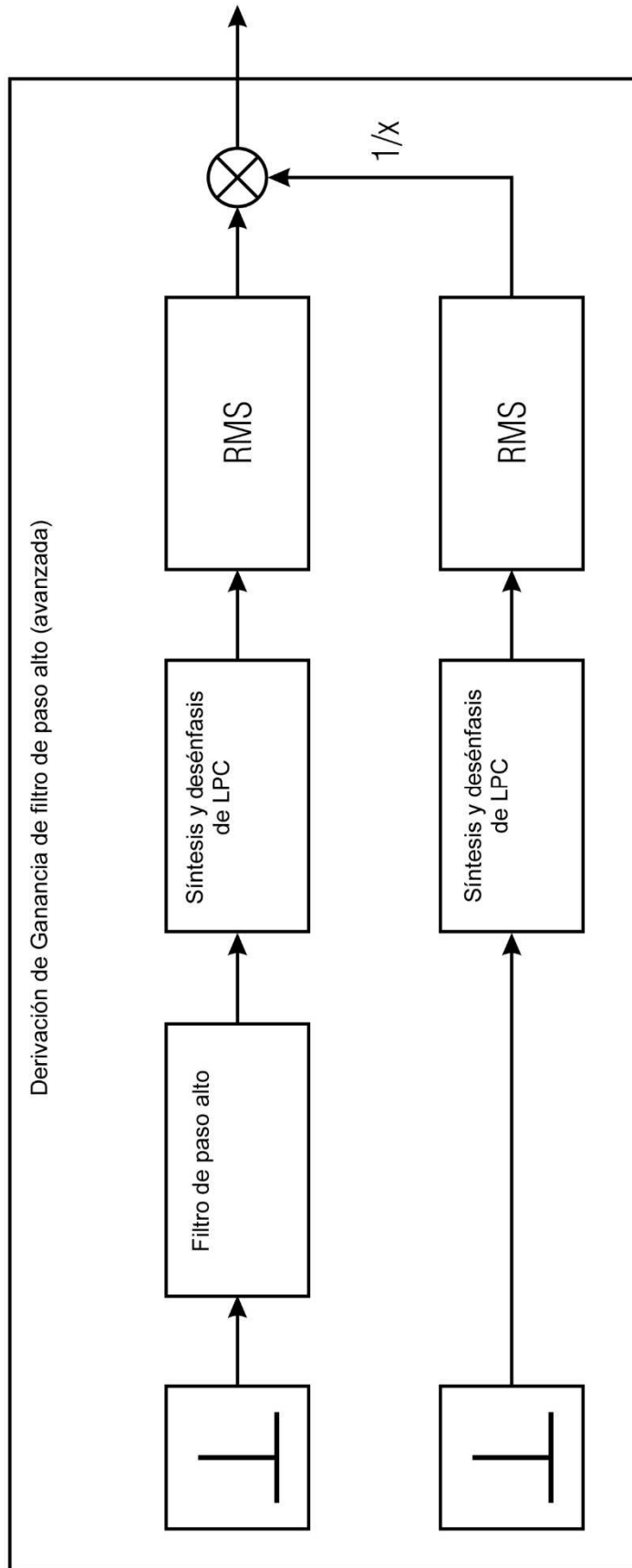


FIG 9

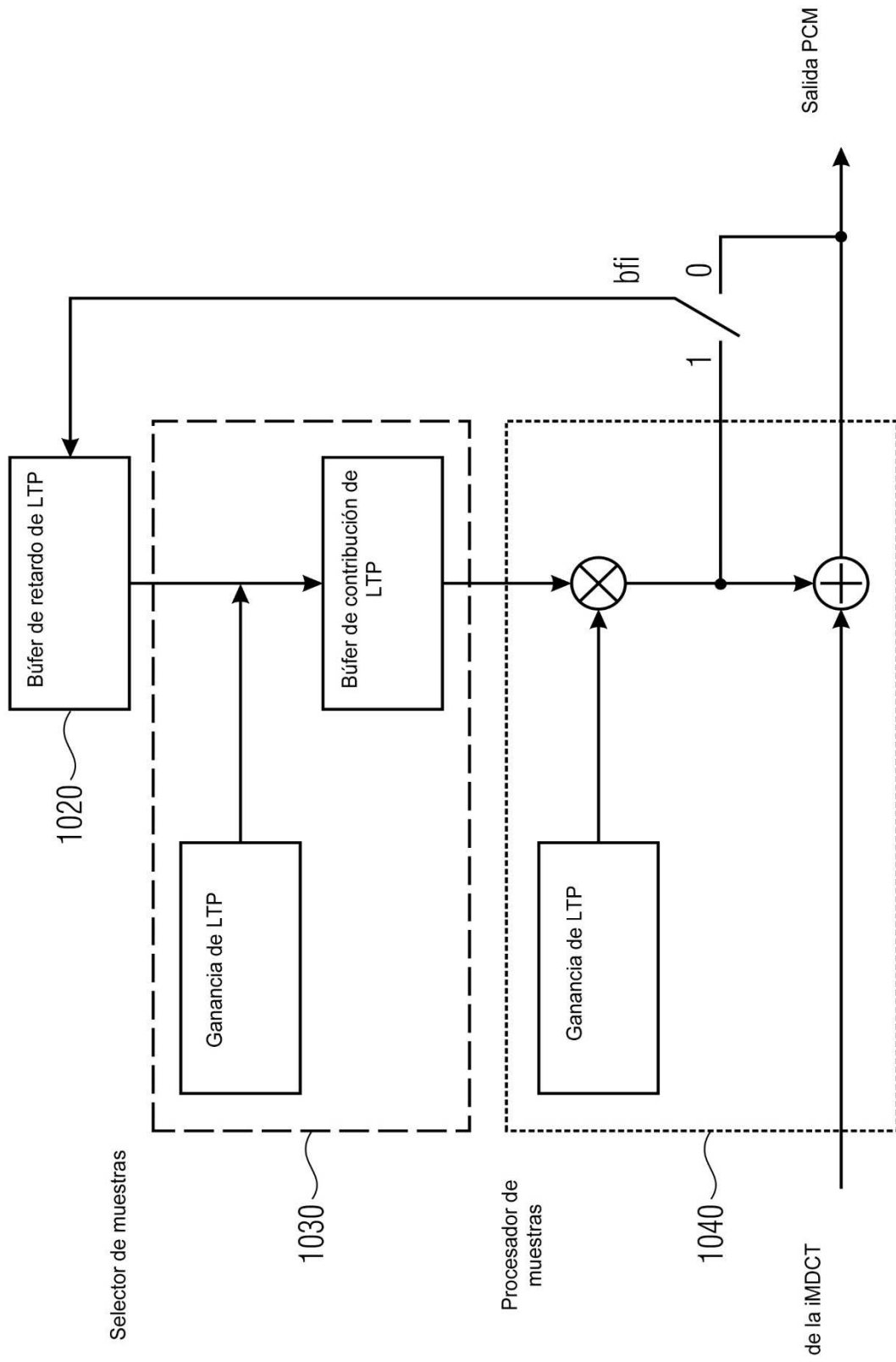


FIG 10

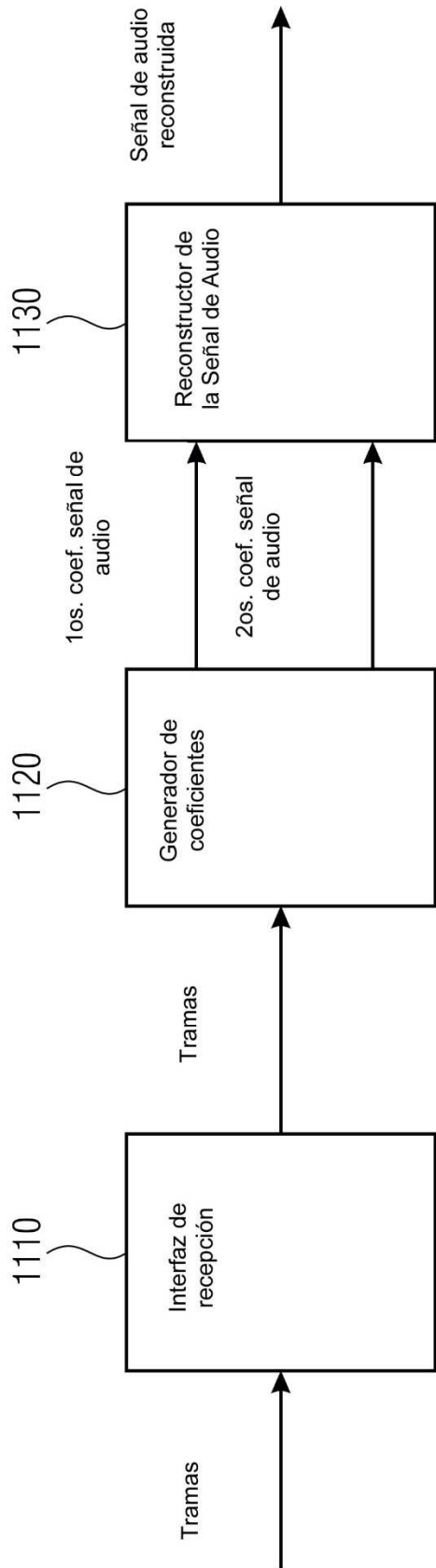


FIG 11

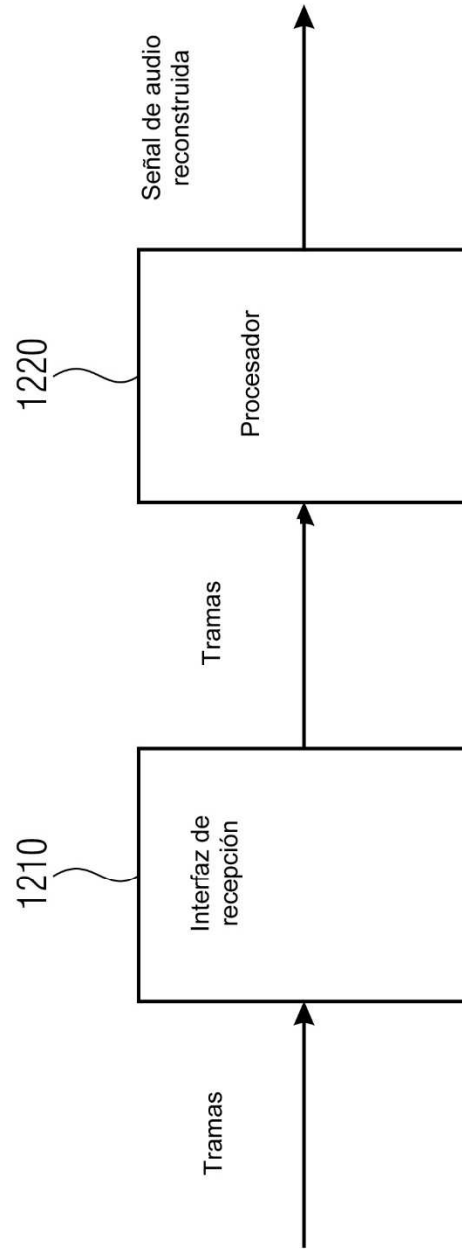


FIG 12

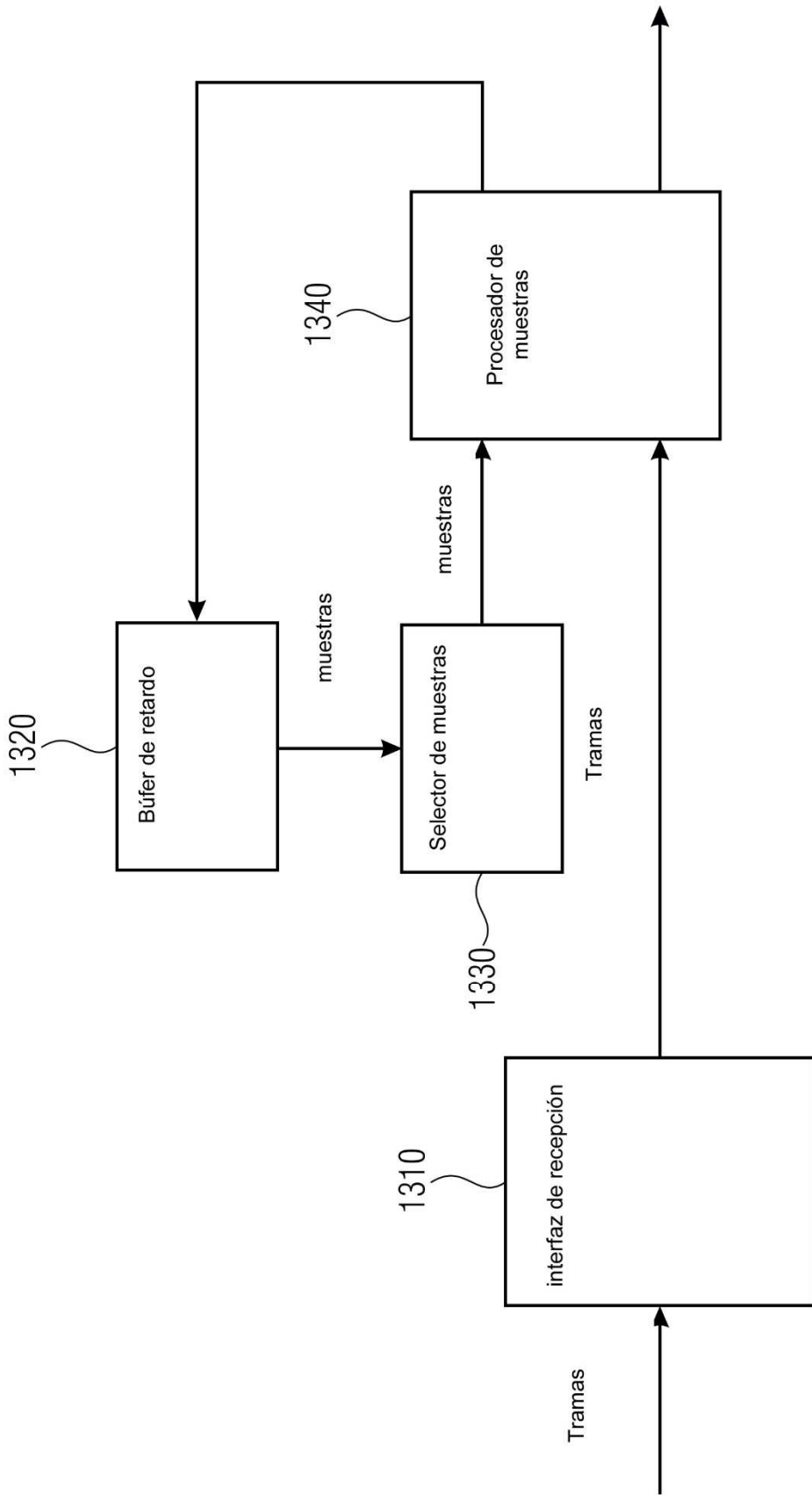


FIG 13

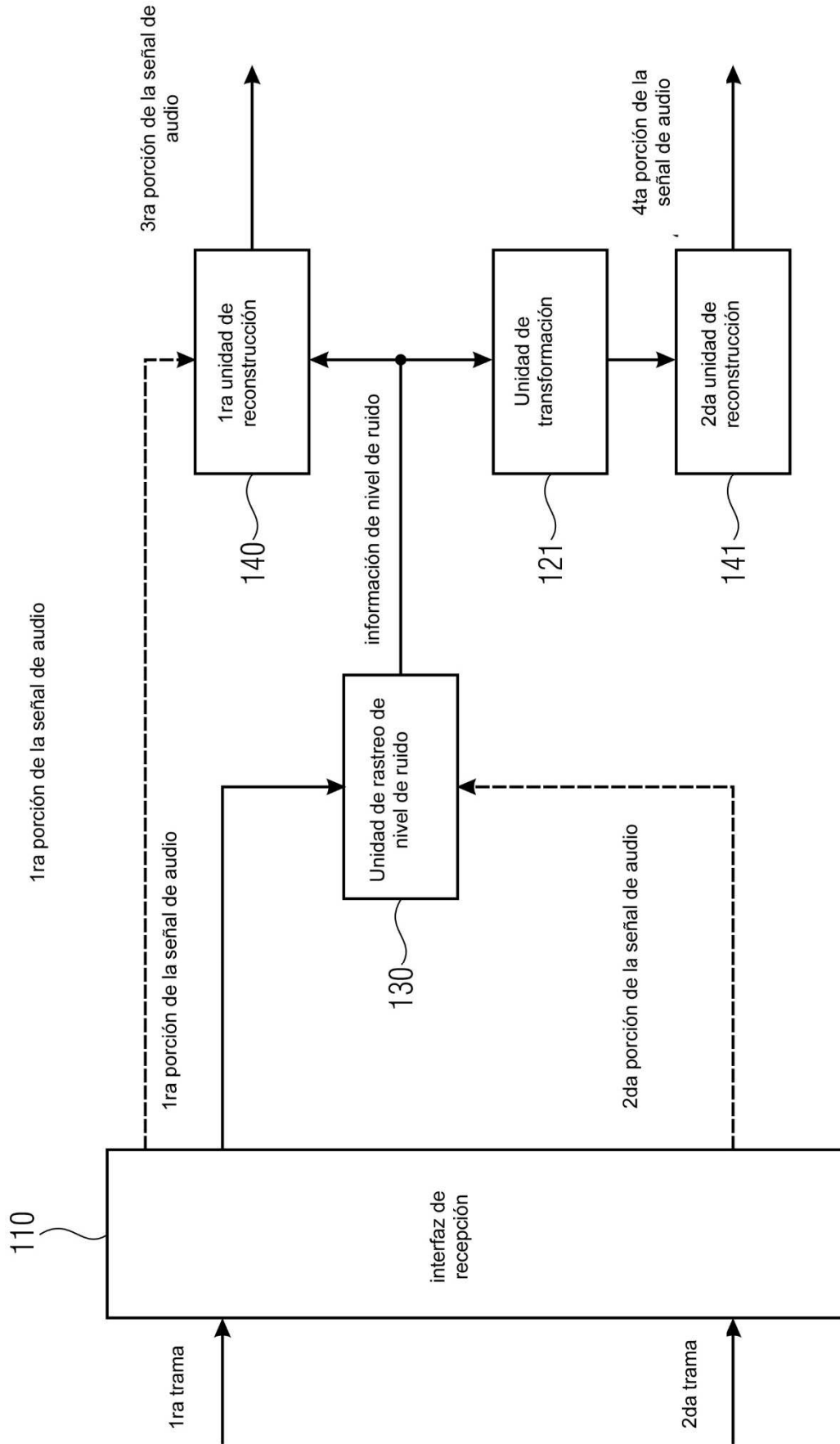


FIG 14