

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 644 693**

51 Int. Cl.:

G10L 19/005 (2013.01)

G10L 25/90 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **23.06.2014 PCT/EP2014/063173**

87 Fecha y número de publicación internacional: **24.12.2014 WO14202786**

96 Fecha de presentación y número de la solicitud europea: **23.06.2014 E 14732194 (7)**

97 Fecha y número de publicación de la concesión europea: **26.07.2017 EP 3011558**

54 Título: **Aparato y método para generar una forma espectral adaptativa de ruido de confort**

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:

30.11.2017

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:

SALVA FERRER, Joan

ES 2 644 693 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Aparato y método para generar una forma espectral adaptativa de ruido de confort

5 **[0001]** 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 método para el desvanecimiento mejorado de la señal para sistemas de codificación de audio cambiados durante el ocultamiento de errores.

[0002] A continuación, se describe el estado de la técnica respecto al desvanecimiento de códec vocales y de audio durante el ocultamiento de pérdida de paquetes (PLC). Las explicaciones con respecto al estado de la técnica con los códecos ITU-T de la serie G (G.718, G.719, G.722, G.722.1, G.729, G.729.1), son seguidas de los códecos 3GPP (AMR, AMR-WB, AMR-WB+) y un códec IETF (OPUS), y concluyen con dos códecos MPEG (HE-AAC, HILN) (ITU = Unión Internacional de Telecomunicaciones; 3GPP = Proyecto Asociación de Tercera Generación; AMR = Multitasa Adaptativa; WB = banda ancha; IETF = Grupo de Trabajo de Ingeniería de Internet). A continuación, se analiza el estado de la técnica respecto del rastreo del nivel de ruido de fondo, seguido de un resumen que proporciona una perspectiva general.

[0003] En primer lugar, se considera G.718. G.718 es un códec vocal de banda estrecha y de banda ancha, que soporta DTX/CNG (DTX = Digital Theater Systems (Sistemas de Cine Digital); CNG = Generación de Ruido de Confort). Como realizaciones se refieren en particular al código de retardo bajo, se describirá aquí en más detalle el modo de versión de retardo bajo.

[0004] 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: De acuerdo con G.718, en el caso de 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 se controla por un factor de atenuación, α . El factor de atenuación α , depende además de la estabilidad, θ , del filtro LP (LP = Predicción Lineal) para tramas SIN VOZ. En general, la convergencia es lenta si la última buena trama recibida se encuentra en un segmento estable y es rápida si la trama se encuentra en un segmento de transición.

[0005] 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 (Frecuencia Espectral de Impedancia y Admitancia) [ITU08a, sección 7.1.2.4.2].

[0006] La Tabla 1 muestra el esquema de cálculo de α :
 40 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 buena trama recibida	Número de tramas borradas sucesivas	α
INICIO ARTIFICIAL		0,6
INICIO, CON VOZ	≥ 3	1,0
	> 3	0,4
TRANSICIÓN CON VOZ		0,4
TRANSICIÓN SIN VOZ		0,8
SIN VOZ	= 1	$0,2 \theta + 0,8$
	= 2	0,6
	> 2	0,4

[0007] Además, G.718 proporciona un método de desvanecimiento a fin de modificar la envoltura espectral. La idea general es hacer converger los últimos parámetros de ISF hacia un vector de la media de ISF adaptativa. En primer lugar, se calcula un vector 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 de largo plazo en tren fuera de línea (que es un vector constante) [ITU08a, sección 7.11.1.2].

50 **[0008]** Además, G.718 proporciona un método de desvanecimiento para el control del comportamiento de

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 CNG [ITU08a, sección 7.11.1.6]. La atenuación de ganancia de innovación se calcula como:

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

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 .

[0009] De manera similar a la atenuación de excitación periódica, la ganancia es atenuada en forma lineal a lo largo de la trama sobre una base de muestra por muestra, partiendo de

$$g_s^{[0]},$$

y alcanza

$g_s^{[1]}$

al comienzo de la siguiente trama.

[0010] La Fig. 2 describe la estructura del decodificador de G.718. En particular, la Fig. 2 ilustra una estructura de decodificador de G.718 de alto nivel para PLC, con rasgo de un filtro de paso alto.

[0011] Por medio del enfoque descrito anteriormente 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 (código C de punto flotante, stat_noise_uv_mod.c), \tilde{E} se obtiene como se indica a continuación:

```

if (unvoiced_vad == 0) {
40     if ( unv_cnt > 20 ) {
        ftmp = lp_gainc *
            lp_gainc; lp_ener = 0.7f * lp_ener + 0.3f * ftmp;
        }
        else {
45             unv_cnt++;
        }
    }
    else {
        unv_cnt = 0;
50 }

```

en la que unvoiced_vad sostiene la detección de actividad vocal, en la que unv_cnt sostiene el número de tramas sin voz en una fila, en la que lp_gainc sostiene las ganancias de paso bajo del libro de código fijo, y en la que lp_ener sostiene la estimación de energía CNG de paso bajo \tilde{E} , se inicializa con 0.

[0012] Además, G.718 proporciona un filtro de paso alto, introducido en la trayectoria de señal de la

excitación sin voz, si la señal de la última buena trama fue clasificada de forma diferente de SIN VOZ, véase la Fig. 2; véase además [ITU08a, sección 7.11.1.6]. Este filtro tiene una característica de meseta baja, con una respuesta de frecuencia en DC de aproximadamente 5 dB menor que en la frecuencia Nyquist.

5 **[0013]** Además, G.718 propone un circuito de retroalimentación de LTP desacoplado (LTP = Predicción de 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]) en base a la completa excitación. Durante el ocultamiento, este circuito de retroalimentación es actualizado a modo 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]) basándose en la
10 excitación con voz únicamente. Con este enfoque, el libro de código adaptativo no es "contaminado" con ruido que tiene su origen en la excitación de innovación aleatoriamente seleccionada.

[0014] Con respecto a las capas de mejora codificadas de transformada (3-5) de G.718, durante el ocultamiento, el decodificador se comporta respecto de la capa alta, de modo de decodificar de manera similar a la
15 operación normal, solo que el espectro de MDCT se ajusta a cero. No se aplica comportamiento de desvanecimiento especial durante el ocultamiento.

[0015] Con respecto a la CNG, en G.718, la síntesis de 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
20 ruido de confort. Después, se restablece el búfer de altura. A continuación, se guarda la síntesis para la clasificación de FER (recuperación de error de trama). Posteriormente, se realiza el desénfasis de espectro. A continuación, se realiza el posfiltrado de baja frecuencia. A continuación, las variables de CNG son actualizadas.

[0016] En el caso de ocultamiento, se realiza exactamente lo mismo, excepto que los parámetros de CNG no
25 son decodificados a partir del flujo de bits. Esto significa que los parámetros no están actualizados durante la pérdida de trama, pero se usan los parámetros decodificados a partir del último buen SID (descriptor de inserción de silencio).

[0017] Ahora, se considera G.719. G.719, que se basa en Siren 22, es un códec de audio de banda completa
30 basado en la transformada. La ITU-T recomienda para G.719 un desvanecimiento con repetición de trama en el dominio espectral [ITU08b, sección 8.6]. De acuerdo con 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 se almacenan 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
35 sometidos a escala decreciente con un factor 0,5, y después, 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 ventaneado-superposición-adición.

[0018] A continuación, se describe G.722. G.722 es un sistema de codificación de 50 a 7000 Hz que utiliza la
40 modulación por impulsos codificados de diferencial adaptativo de subbanda (SB-ADPCM) 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 = filtro espejo en cuadratura). Las dos bandas resultantes son codificadas por ADPCM (ADPCM = modulación por impulsos codificados de diferencial adaptativo).

[0019] Para G.722, se especifica un algoritmo de alta complejidad para el ocultamiento de pérdida de
45 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 111.5]) propone un silenciamiento realizado de forma gradual, comenzando después de 20 ms de pérdida de trama, y completándose después de 60 ms de pérdida de trama. Además, G.722 -Apéndice IV propone una técnica de desvanecimiento que
50 se aplica "a cada muestra, un factor de ganancia que es computado y adaptado muestra por muestra " [ITU07, sección IV.6.1.2.7].

[0020] En G.722, el proceso de silenciamiento tiene lugar en el dominio de subbanda justo antes de la
55 síntesis de QMF y como la última etapa del módulo PLC. El cálculo del factor de silenciamiento se realiza usando información de clase del clasificador de señales que además es parte del módulo de PLC. La distinción se hace entre clases TRANSIENT, 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/múltiples de tramas de 20 ms).

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

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

5

[0023] A continuación, se considera G.722.1. G.722.1, que se basa en Siren 7, es un códec de audio de banda ancha basado en la transformada con un modo de extensión de banda súper ancha, denominado como G.722.1C. El propio G. 722.1C se basa en Siren 14. La ITU-T recomienda para G.722.1 una repetición de tramas, con el posterior silenciamiento [ITU05, sección 4.7]. Si el decodificador es informado, por medio de un mecanismo de señalización externo no definido en esta recomendación, de que una trama se ha perdido o ha sido alterada, repite los coeficientes de MLT (transformada superpuesta modulada) decodificados de la trama previa. 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 se perdió o se alteró, entonces el decodificador establece todos los coeficientes de MLT de las tramas actuales en cero.

10

15

[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 vocal de predicción lineal excitada por código (CS-ACELP) [ITU12].

[0025] Como se describe en [CPK08], G.729 recomienda un desvanecimiento en el dominio LP. El algoritmo PLC empleado en el estándar G.729 reconstruye la señal vocal para la trama actual en base a la información vocal previamente recibida. En otras palabras, el algoritmo 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 se atenúan por un factor constante.

20

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

$$g_c^{(m)} = 0,98 \cdot g_c^{(m-1)}$$

30 con 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 anterior:

$$g_p^{(m)} = 0,9 \cdot g_p^{(m-1)} \quad , \text{ unido por } g_p^{(m)} < 0,9$$

35

[0028] Nam en Park et al. 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:

40

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

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

45 cuadrado se minimiza:

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

[0029] ϵ es el error cuadrado, g_j es la amplitud pasada original j ésima. A fin de minimizar este error, simplemente el derivado con respecto a a y b se ajusta a cero. Usando los parámetros optimizados a^* y b^* , una estimación de cada g_i^* se representa por

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

[0030] La Fig. 4 muestra la predicción de amplitud, en particular, la predicción de la amplitud g_i^* usando la regresión lineal.

5 **[0031]** Para obtener la amplitud A'_i del paquete perdido i , una relación σ_i

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

se multiplica con un factor de escala S_i :

10

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

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

$$S_i = \begin{cases} 1,0, & \text{si } l(i) = 1, 2 \\ 0,9, & \text{si } l(i) = 3, 4 \\ 0,8, & \text{si } l(i) = 5, 6 \\ 0, & \text{de otro modo} \end{cases} \quad (7)$$

15

[0032] En [PKJ+11], se propone una escala levemente diferente. De acuerdo con G.729, posteriormente, 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 PLC anteriores.

20

[0033] A continuación, se considera G.729.1. G.729.1 es un codificador de tasa de bits variable embutido basado en G.729: Un flujo de bits de codificador de banda ancha escalable de 8-32 kbit/s interoperable con G.729 [ITU06b].

25 **[0034]**

De acuerdo con G.729.1, como en G.718 (véase anteriormente), 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 en base a un factor de atenuación α que depende de los parámetros de la clase de la última buena trama recibida y el número de tramas borradas consecutivas. El factor de atenuación α depende adicionalmente de la estabilidad del filtro LP para tramas SIN VOZ. En general, la atenuación

30

es lenta si la última buena trama recibida se encuentra en un segmento estable y es rápida si la trama se encuentra en un segmento de transición.

[0035] Además, el factor de atenuación α depende de la ganancia de paso promedio por subtrama g_p ([ITU06b, ec. 163, 164]):

35

$$\bar{g}_p = 0,1g_p^{(0)} + 0,2g_p^{(1)} + 0,3g_p^{(2)} + 0,4g_p^{(3)} \quad (8)$$

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

[0036] La Tabla 2 muestra el esquema de cálculo de α , donde

40

$$\beta = \sqrt{\bar{g}_p} \quad \text{con} \quad 0,85 \geq \beta \geq 0,98 \quad (9)$$

[0037] 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 LP. [ITU06b, sección 7.6.1].

45

última buena trama recibida	Número de tramas borradas sucesivas	α
CON VOZ	1	β
	2,3	g_p
	>3	0,4
INICIO	1	$0,8 \beta$
	2,3	g_p
	>3	0,4
INICIO ARTIFICIAL	1	$0,6 \beta$
	2,3	g_p
	>3	0,4
TRANSICIÓN CON VOZ	≥ 2	0,8
	>2	0,2
TRANSICIÓN SIN VOZ		0,88
	1	0,95
	2,3	$0,6 \theta + 0,4$
	>3	0,4

[0038] De acuerdo con G.729.1, con respecto a la resincronización de pulso glotal, como el último pulso de la excitación de la trama anterior se usa para la construcción de la parte periódica, su ganancia es aproximadamente correcta al comienzo de la trama oculta, y puede ajustarse en 1. Después, la ganancia se atenúa linealmente a lo largo de toda la trama en 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 se extrapola usando los valores de ganancia de excitación de altura de cada subtrama de la última buena trama. En general, si estas ganancias son mayores de 1, la energía de señal

es creciente; si son menores de 1, la energía es decreciente. Por lo tanto, α se ajusta a $\beta = \sqrt{g_p}$ como se ha descrito anteriormente, véase [ITU06b, ec. 163, 164]. El valor de β se acorta entre 0,98 y 0,85 para evitar fuertes aumentos y descensos de energía, véase [ITU06b, sección 7.6.4].

[0039] Con respecto a la construcción de la parte aleatoria de la excitación, de acuerdo con 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 buena trama:

$$g_s = 0,1g^{(0)} + 0,2g^{(1)} + 0,3g^{(2)} + 0,4g^{(3)}$$

en la que $g^{(0)}$, $g^{(1)}$, $g^{(2)}$ y $g^{(3)}$ son las ganancias del libro de código establecido, o innovación, 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, partiendo de $g_s^{(0)}$ y llegando al valor de $g_s^{(1)}$ que se lograría al comienzo de la siguiente trama.

[0040] De acuerdo con 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].

[0041] A continuación, se considera AMR. 3GPP AMR [3GP12b] es un códec vocal que utiliza en algoritmo ACELP. AMR es capaz de codificar la voz con una tasa de muestreo de 8000 muestras/s y una tasa de bits entre 4,75 y 12,2 kbit/s, y sostiene la señalización de tramas de descriptores de silencio (DTX/CNG).

[0042] En la AMR, durante el ocultamiento de errores (véase [3GP12a]), se distingue entre tramas que son

propensas a error (errores de bits) y tramas que están perdidas por completo (sin ningún dato).

[0043] Para el ACELP, la 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 aumenta en uno, y se satura cuando llega a 6. Cada vez que se detecta una buena trama de la voz, el contador de estado se restablece a cero, excepto cuando el estado es 6, donde el contador de estado se ajusta a 5. El flujo de control de la máquina de estado puede describirse por el siguiente código C (BFI es un indicador de mala trama, Estado es una variable de estado):

```

10 if (BFI != 0) {
    State = State + 1;
}
else if (State == 6) {
15 State = 5;
}
else {
    State = 0;
}
if (State > 6) {
20 State = 6;
}

```

[0044] Además de esta máquina de estado, en la AMR, los indicadores de mala trama de las tramas actuales y previas son verificadas (prevBFI).

[0045] Son posibles tres diferentes combinaciones:

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 anterior. Los parámetros vocales recibidos se usan de la manera normal en la síntesis vocal. Se guarda la trama actual de parámetros vocales.

[0046] La segunda de las tres combinaciones es BFI = 0, prevBFI = 1, Estado = 0 o 5: No se detecta error en la trama de la voz recibida, pero la trama de la voz recibida anterior fue mala. La ganancia LTP y la ganancia de libro de código establecido están limitadas debajo de los valores utilizados para la última buena subtrama recibida:

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

donde g_p = ganancia LTP decodificada actual, $g_p(-1)$ = ganancia LTP usada para la última buena subtrama (BFI = 0), y

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

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

[0047] El resto de los parámetros vocales recibidos se usan normalmente en la síntesis vocal. Se guarda la trama actual de parámetros vocales.

[0048] La tercera de las tres combinaciones es BFI = 1, prevBFI = 0 o 1, Estado = 1..6: Se detecta un error en la trama vocal recibida, y se inicia el procedimiento de sustitución y silenciamiento. La ganancia LTP y la ganancia de libro de código fijo se reemplazan por valores atenuados de las subtramas anteriores:

$$g_p = \begin{cases} P(\text{estado}) \cdot g_p(-1), & g_p(-1) \leq \text{mediana5}(g_p(-1), \dots, g_p(-5)) \\ P(\text{estado}) \cdot \text{mediana5}(g_p(-1), \dots, g_p(-5)) & g_p(-1) > \text{mediana5}(g_p(-1), \dots, g_p(-5)) \end{cases} \quad (12)$$

5 y donde g_p indica la ganancia LTP decodificada actual de predicción de largo plazo y $g_p(-1), \dots, g_p(-n)$ indican las ganancias (LTP) utilizadas para las últimas n subtramas y $\text{mediana5}()$ indica una operación de mediana de 5 puntos

P(estado) = factor de atenuación,

10 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

$$g_c = \begin{cases} C(\text{estado}) \cdot g_c(-1), & g_c(-1) \leq \text{mediana5}(g_c(-1), \dots, g_c(-5)) \\ C(\text{estado}) \cdot \text{mediana5}(g_c(-1), \dots, g_c(-5)) & g_c(-1) > \text{mediana5}(g_c(-1), \dots, g_c(-5)) \end{cases} \quad (13)$$

15 donde g_c indica la ganancia de libro de código fijo decodificada actual y $g_c(-1), \dots, g_c(-n)$ indican las ganancias de libro de código fijo utilizadas para las últimas n subtramas y $\text{mediana5}()$ indica una operación mediana de 5 puntos y $C(\text{estado}) = \text{factor de atenuación}$, 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.

20 **[0049]** En la AMR, los valores de retraso de LTP (LTP = Predicción de largo plazo) se reemplazan por el valor pasado de la 4ª subtrama de la trama anterior (modo 12.2), o valores ligeramente modificados en base al último valor recibido correctamente (todos los demás modos).

25 **[0050]** De acuerdo con la AMR, los pulsos de innovación de libro de código fijo recibidos de la trama errónea se usan en el estado en el que se recibieron, cuando se reciben los datos alterados. En el caso de que no se reciban datos, deben emplearse los índices de libro de código fijo aleatorios.

30 **[0051]** Con respecto a la CNG en AMR, de acuerdo con [3GP12a, sección 6.4], cada primera trama 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 posteriores 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 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, dtx_dec { } @sp_dec. c], que produce 37,5 dB por segundo). Obsérvese que el desvanecimiento aplicado a la CNG se realiza en el dominio LP.

35 **[0052]** A continuación, se considera la AMR-WB. La multitasa adaptativa - WB [ITU03, 3GP09c] es un códec vocal, ACELP, en base a AMR (véase la sección 1.8). Utiliza una extensión de ancho de banda paramétrica y, además, sostiene DTX/CNG. En la descripción del estándar [3GP12g], hay soluciones de ejemplos de ocultamiento proporcionadas, que son las mismas que para la AMR [3GP12a], con desviaciones menores. Por lo tanto, sólo las diferencias con respecto a la AMR se describen aquí. Para la descripción del estándar, véase la descripción anterior.

40 **[0053]** Con respecto a ACELP, en AMR-WB, el desvanecimiento ACELP se realiza basándose en el código fuente de referencia [3GP12c] modificando la ganancia de altura g_p (para AMR anterior denominada ganancia LTP) y modificando la ganancia de código g_c .

45 **[0054]** En el caso de trama perdida, la ganancia de altura g_p para la primera subtrama es la misma que en la última buena trama, 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 en un factor de 0,95 y se limita de nuevo.

[0055] La AMR-WB propone que en una trama oculta, g_c se basa en la última g_c :

$$g_{c,actual} = g_{c,pasado} * (1,4 - g_{p,pasado}) \quad (14)$$

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

$$g_{c_{inov}} = \frac{1.0}{\sqrt{\frac{ener_{inov}}{\text{tamaño_subtrama}}}} \quad (16)$$

$$ener_{inov} = \sum_{i=0}^{\text{tamaño_subtrama}-1} code[i] \quad (17)$$

5 **[0056]** Para el ocultamiento de los retrasos de LTP, en la AMR-WB, el historial de los cinco últimos buenos retrasos de LTP y ganancias de LTP se usan para hallar el mejor método 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, ya sea si el retraso recibido de predicción de LTP es utilizable o no [3GP12g].

10 **[0057]** Con respecto a la CNG, en AMR-WB, si la última trama recibida correctamente fue una trama de 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.

[0058] Para posteriores tramas de SID perdidas, la AMR-WB propone la aplicación de una técnica de atenuación al ruido de confort que disminuirá en forma gradual el nivel de salida. Por lo tanto, se verifica si la última actualización de 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] que produce 18,75 dB por segundo). Obsérvese que el desvanecimiento aplicado a la CNG se realiza en el dominio LP.

15 **[0059]** Ahora, se considera AMR-WB+. La multitasa adaptativa - WB+ [3GP09a] es un códec cambiado usando ACELP y TCX (TCX = excitación codificada por la transformada), como códecs de núcleo. Utiliza una extensión de ancho de banda paramétrica y, además, sostiene DTX/CNG.

[0060] En AMR-WB+, se aplica una lógica de extrapolación de modos para extrapolar los modos de las 25 tramas perdidas dentro de una supertrama distorsionada. Esta extrapolación de modos se basa en el hecho de que existe redundancia en la definición de indicadores de modos. La lógica de decisión (dada en [3GP09a, figura 18]) propuesta por AMR-WB+ es como se indica a continuación:

- Se define un modo de vector, (m₋₁, m₀, m₁, m₂, m₃), donde m₋₁ indica el modo de la última trama de la supertrama anterior y m₀, m₁, m₂, m₃ indican los modos de las tramas en la supertrama actual (decodificada a partir del flujo de bits), donde m_k = -1, 0, 1, 2 o 3 (-1: perdida, 0: ACELP, 1: TCX20, 2: TCX40, 3: TCX80), y donde el número de tramas perdidas *n*loss puede estar entre 0 y 4.

30 - Si m₋₁ = 3 y dos de los indicadores de modo de las tramas 0 - 3 son iguales a tres, todos los indicadores se ajustarán a tres ya que entonces, es seguro que una trama TCX80 fue indicada dentro de la supertrama.

35 - Si solo un indicador de las tramas 0 - 3 es tres (y el número de tramas perdidas *n*loss es tres), el modo se ajustará a (1, 1, 1, 1), ya que, entonces, 3/4 del espectro objetivo TCX80 está perdido, y es muy probable que la ganancia global TCX esté perdida.

- si el modo indica (x, 2,-1, x, x) o (x,-1, 2, x, x), se extrapolará a (x, 2, 2, x, x), lo que indica una trama TCX40. Si el modo indica (x, x, x, 2,-1) o (x, x,-1, 2) se extrapolará a (x, x, x, 2, 2), lo que también indica una trama TCX40. Ha de 40 apreciarse que (x, [0, 1], 2, 2, [0, 1]) son configuraciones inválidas.

- Después de esto, para cada trama que está perdida (*modo* = -1), el modo se ajusta a ACELP (*modo* = 0) si la trama anterior fue ACELP y el modo se ajusta a TCX20 (*modo* = 1) para todos los demás casos.

[0061] Con respecto a ACELP, de acuerdo con AMR-WB+, si un modo de tramas perdidas da como resultado $m_k = 0$ después de la extrapolación de modo, se aplica el mismo enfoque que en [3GP12g] para esta trama (véase anteriormente).

5

[0062] En AMR-WB+, dependiendo del número de tramas perdidas y el modo extrapolado, se distinguen los siguientes enfoques de ocultamiento relacionados con TCX (TCX = Excitación Codificada de la Transformada):

- Si se pierde una trama completa, entonces se aplica una ACELP como ocultamiento: La última excitación se repite y se usan coeficientes ISF ocultos (ligeramente cambiados hacia su media adaptativa) para sintetizar la señal de dominio temporal. Adicionalmente, un factor de desvanecimiento de 0,7 por trama (20 ms) [3GP09b, dec_tcxc] se multiplica en el dominio predictivo lineal, justo antes de la síntesis LPC (codificación predictiva lineal).

- Si el último modo fue TCX80, así como el modo extrapolado de la supertrama (parcialmente perdida) es TCX80 (nloss = [1, 2], modo = (3, 3, 3, 3, 3)), se realiza el ocultamiento en el dominio FFT, utilizando extrapolación de fase y amplitud, teniendo en cuenta la última trama correctamente recibida. El enfoque 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 AMR-WB+, el enfoque realizado para el ocultamiento de TCX consiste en las siguientes etapas [3GP09a, sección 6.5.1.2.3]:

20 - Se computa el espectro de magnitud de trama anterior:

$$oldA[k] = |old\hat{X}[k]|$$

- Se computa el espectro de magnitud de trama actual:

25

$$A[k] = |\hat{X}[k]|$$

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

30

$$ganancia = \sqrt{\frac{\sum A[k]^2}{\sum oldA[k]^2}}$$

- La amplitud de los coeficientes espectrales perdidos se extrapola usando:

35

$$if(lost[k]) A[k] = ganancia \cdot oldA[k]$$

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

40

[0063] Con respecto a la CNG en AMR-WB+, se usa el mismo enfoque que en la AMR-WB (véase anteriormente).

[0064] A continuación, se considera OPUS. OPUS [IET12] incorpora tecnología de dos códecs: SILK orientado a la voz (conocido como el códec Skype) y la CELT de baja latencia (CELT = Transformada Superpuesta de Energía Restringida). Opus puede ajustarse de una manera sin costuras, 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), así como un híbrido para una corta superposición.

50

[0065] 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 LTP está atenuado mediante la multiplicación de todos los coeficientes 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 anterior. 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 se atenúa exponencialmente con $0,99^{lostcnt}$ por trama, de manera que el parámetro de ganancia de excitación sea de 0,99 para el primer parámetro de ganancia de excitación, de manera que el parámetro de ganancia de excitación sea 0,992 para el segundo parámetro de ganancia de excitación, etc. La excitación se genera usando un generador de número aleatorio que genera ruido blanco por el sobreflujo variable. Además, los coeficientes de LPC son extrapolados/promediados en base al último grupo de coeficientes recibidos correctamente. Después de la generación del vector de excitación atenuado, los coeficientes de LPC ocultos se usan en OPUS para la síntesis de la señal de salida de dominio de tiempo.

[0066] Ahora, en el contexto de OPUS, se considera la CELT. La CELT es un códec basado en la transformada. El ocultamiento de la CELT caracteriza un enfoque PLC basado en la altura, que se aplica para hasta cinco tramas perdidas consecutivamente. Partiendo de la trama 6, se aplica un enfoque de ocultamiento de tipo ruido, que genera ruido de fondo, cuya característica se supone que suena como ruido de fondo precedente.

[0067] La Fig. 5 ilustra el comportamiento de pérdida de ráfaga de CELT. En particular, la Fig. 5 representa un espectrograma (eje x: tiempo; eje y: frecuencia) de un segmento vocal oculto de CELT. La caja de color gris claro indica las primeras 5 tramas consecutivamente perdidas, donde se aplica el enfoque de PLC basado en la altura. Más allá de ello, se muestra el ocultamiento de tipo ruido. Debe observarse que el cambio se realiza de manera instantánea, y no transita suavemente.

[0068] Con respecto al ocultamiento basado en el altura, en OPUS, el ocultamiento basado en el altura consiste en hallar la periodicidad en la señal decodificada por autocorrelación y la repetición de la forma de onda ventaneada (en el dominio de excitación usando el análisis y síntesis de LPC) usando la desviación de altura (retraso de altura). La forma de onda ventaneada se superpone de tal forma como para preservar la cancelación del solapamiento en el dominio de tiempo con la trama anterior y la siguiente trama [JET12]. Adicionalmente, se deriva un factor de desvanecimiento y se aplica mediante el siguiente código:

```
opus_val32 E1=1, E2=1;
int period;
if (pitch_index <= MAX_PERIOD/2) {
35     period = pitch_index;
}
else {
    period = MAX_PERIOD/2;
}
40 for (i=0;i<period;i++)
{
    E1 += exc[MAX_PERIOD- period+i] * exc[MAX_PERIOD- period+i];
    E2 += exc[MAX_PERIOD-2*period+i] * exc[MAX_PERIOD-2*period+i];
}
45 if (E1 > E2) {
    E1 = E2;
}
decay = sqrt(E1/E2);
attenuation = decay;
```

[0069] En este código, exc contiene la señal de excitación hasta MAX_PERIOD muestras antes de la pérdida.

[0070] La señal de excitación se multiplica más tarde con atenuación, después, se sintetiza y sale a través de la síntesis de LPC.

[0071] El algoritmo de desvanecimiento para el enfoque de dominio de tiempo puede ser resumido de la siguiente manera:

- Hallar la energía sincrónica de altura del último ciclo de altura antes de la pérdida.
- 60 - Hallar la energía sincrónica de altura del segundo ciclo de altura antes de la pérdida.

- 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.

[0072] Con respecto al ocultamiento de tipo ruido, de acuerdo con OPUS, para la 6ª y siguientes tramas perdidas consecutivas, se realiza un enfoque de sustitución de ruido en el dominio de MDCT a fin de simular el ruido de fondo de confort.

[0073] Con respecto al rastreo del nivel de ruido de fondo y la forma, en OPUS, la estimación de ruido de fondo se realiza de la siguiente manera. Después del análisis de MDCT, se calcula la raíz cuadrada de las energías de la banda de MDCT por banda de frecuencia, donde la agrupación de los intervalos de MDCT sigue la escala de Bark de acuerdo con [IET12, Tabla 55]. Después, la raíz cuadrada de las energías se transforma en el dominio \log_2 por medio de:

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

15

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

20 **[0074]** En OPUS, el ruido de fondo es representado en el lado del decodificador como [IET12, amp2Log2 y log2Amp @ quant_bands.c]:

$$backgroundLogE[i] = \min(backgroundLogE[i] + 8 \cdot 0,001, bandLogE[i]) \\ \text{para } i = 0 \dots 21 \quad (19)$$

25 **[0075]** La energía mínima trazada se determina básicamente por la raíz cuadrada de la energía de la banda de la trama actual, pero el aumento de una trama a la siguiente se limita en 0,05 dB.

[0076] Con respecto a la aplicación del nivel de ruido de fondo y la forma, de acuerdo con OPUS, si se aplica el ruido como PLC, se usa $backgroundLogE$ como se obtiene en la última buena trama y se convierte de nuevo en el dominio lineal:

30

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

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

35

[0077] El procedimiento de ocultamiento actual es para llenar la trama de la MDCT con ruido blanco producido por un generador de número aleatorio, y escalar este ruido blanco de modo tal que coincida a modo de banda, con la energía de $bandE$. Posteriormente, se aplica la MDCT inversa, que logra una señal de dominio de tiempo. Después de la superposición y adición y del desénfasis (como en la decodificación regular), es apagada.

40

[0078] A continuación, 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 de avanzada de alta eficiencia consiste en un códec de audio basado en la transformada (AAC), complementado por una extensión de ancho de banda paramétrica (SBR).

45

[0079] Con respecto a AAC (AAC = codificación de audio avanzada), el consorcio DAB especifica AAC en DAB+, un desvanecimiento a cero en el dominio de frecuencia [EBU10, sección A1.2] (DAB = Transmisión de Audio Digital). El comportamiento de desvanecimiento, por ejemplo, la rampa de atenuación, podría fijarse o ser ajustable por el usuario. Los coeficientes espectrales de la última AU (AU = unidad de acceso) se atenúan por un factor correspondiente a las características de desvanecimiento y después se pasan a la asignación de frecuencia-tiempo.

50

De acuerdo con 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 ajustará a 0.

[0080] El consorcio DRM (DRM = gestión de derechos digitales) especifica 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-tiempo. Si se alteran múltiples tramas, el ocultamiento implementa primero un desvanecimiento, en base a valores espectrales ligeramente 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 se atenúan por un factor correspondiente a las características de desvanecimiento y después se pasan a la asignación de frecuencia-tiempo. De acuerdo con la rampa de atenuación, el ocultamiento cambia a silenciamiento después de un número de tramas inválidas consecutivas, lo que significa que el espectro completo se ajustará a 0.

[0081] 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-tiempo. Si se alteran múltiples tramas, el ocultamiento implementa primero un desvanecimiento, en base a valores espectrales ligeramente modificados, a partir de la última trama buena. Un desvanecimiento completo tarda 5 tramas. Los coeficientes espectrales de la última buena trama se copian y se atenúan por un factor de:

$$fadeOutFac = 2^{-(nFadeOutFrame/2)}$$

con $nFadeOutFrame$ como el contador de tramas desde la última buena trama. Después de cinco tramas de desvanecimiento, el ocultamiento cambia a silenciamiento, lo que significa que el espectro completo se ajustará a 0.

[0082] Lauber y Sperschneider introducen para AAC un desvanecimiento a modo de trama, del espectro de la MDCT, en base a 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 posocultamiento.

[0083] Con respecto a AAC, el cálculo de energía se realiza en 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 en 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 rápidamente en función del tiempo.

[0084] Para la generación del espectro a desvanecer, sugieren la repetición de trama o la sustitución de ruido [LS01, secciones 3.2 y 3.3].

[0085] Quackenbusch y Driesen sugieren para AAC un desvanecimiento a cero a modo de trama exponencial [QD03]. Se propone una repetición del conjunto adyacente de coeficientes de tiempo/frecuencia, en la que cada repetición tiene atenuación exponencialmente en aumento, por lo tanto, el desvanecimiento hasta el silenciamiento, en el caso de prolongados apagones.

[0086] Con respecto a SBR (SBR = Replicación de Banda Espectral), en MPEG-4 HE-AAC, 3GPP sugiere SBR en aacPlus mejorada para almacenar en memoria intermedia los datos de envoltura decodificados, y en el caso de una pérdida de trama, para reutilizar las energías almacenadas en memoria intermedia de los datos de envoltura transmitidos y disminuirlos por una relación constante de 3 dB para cada trama oculta. El resultado se suministra 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. Después tiene lugar la decodificación de SBR de manera habitual. Además, el suelo de ruido codificado delta y los valores de nivel de seno son eliminados. Como no queda disponible ninguna diferencia con la información anterior, el suelo de ruido decodificado y los niveles de seno permanecen proporcionales a la energía de la señal generada de HF [3GP12e, sección 5.2].

[0087] El consorcio de DRM especificado para la SBR junto con la AAC utiliza la misma técnica que 3GPP [EBU12, sección 5.6.3.1]. Además, el consorcio de DAB especifica para SBR en DAB+ la misma técnica que 3GPP [EBU10, sección A2].

- [0088]** A continuación, se consideran MPEG-4 CELP y MPEG-4 HVXC (HVXC = codificación de excitación de vector armónico). El consorcio de DRM especifica para la SBR junto con CELP y HVXC [EBU12, sección 5.6.3.2] que el mínimo ocultamiento de requerimiento para la SBR para códecs vocales sirve para aplicar un conjunto predeterminado de valores de datos, cada vez que se ha detectado una trama alterada de SBR. Estos valores producen una envoltura espectral de alta banda estática, a un bajo nivel relativo de reproducción, que muestra una reducción gradual de respuesta 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 un desvanecimiento real, sino, en cambio, un salto a un cierto nivel de energía a fin de insertar cierta clase de ruido de confort.
- [0089]** Posteriormente, 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 AAC + SBR.
- [0090]** Ahora, se considera MPEG-4 HILN (HILN = armónicos y líneas individuales más ruido). Meine et al. 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 la 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.
- [0091]** A continuación, se considera el rastreo del nivel de ruido de fondo en la técnica anterior. Rangachari y Loizou [RL06] proporcionan una buena reseña de varios métodos, y analizan algunas de sus limitaciones. Los métodos para el rastreo del nivel de ruido de fondo son, por ejemplo, el procedimiento de seguimiento mínimo [RL06] [Coh03] [SFBOO] [Dob95], basado en VAD (VAD = detección de actividad de voz); la filtración de Kalman [Gan05] [BJH06], las descomposiciones de subespacios [BP06] [HJH08]; decisión blanda [SS98] [MPC89] [HE95] y estadística mínima.
- [0092]** Se seleccionó el enfoque de estadística mínima para usarse dentro del alcance para USAC-2, (USAC = codificación de voz y audio unificados) y se describe a continuación en más detalle.
- [0093]** La estimación de densidad espectral de potencia de ruido basada en un suavizado óptimo y estadística mínima [Mar01] introduce un estimador de ruido, que es capaz de funcionar independientemente de la señal que es la voz activa o ruido de fondo. A diferencia de otros métodos, el algoritmo de mínima estadística no utiliza ningún umbral explícito para distinguir entre actividad vocal y pausa vocal, y por lo tanto, se relaciona más estrechamente con los métodos de decisión blanda, que con los métodos tradicionales de detección de actividad vocal. De manera similar a los métodos de decisión blanda, puede también actualizar la PSD (Densidad espectral de potencia) de ruido estimado durante la actividad vocal.
- [0094]** El método de mínima estadística se basa en dos observaciones, concretamente, 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 PSD (PSD = densidad espectral de potencia) de ruido exacta mediante el rastreo de la mínima de la PSD de la señal ruidosa. Debido a que la mínima es menor (o en otros casos, igual) que el valor promedio, el método de rastreo mínimo requiere una compensación de inclinación.
- [0095]** La inclinación es una función de la varianza de la PSD de la señal suavizada, y como tal, depende del parámetro de suavizado del estimador de PSD. A diferencia del 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 PSD dependiente del tiempo y la frecuencia, que también requiere una compensación de inclinación dependiente del tiempo y la frecuencia.
- [0096]** 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 PSD suavizada. Esto conducirá a estimaciones de ruido inexactas, ya que la ventana de deslizamiento para la búsqueda mínima podría deslizarse hacia picos amplios. Por lo tanto, los parámetros de suavizado cercanos a uno no puede ser usados, y en consecuencia, la estimación de ruido tendrá una varianza relativamente grande. Además, la estimación de ruido se inclina hacia valores menores.

Adicionalmente, en el caso de aumentar la potencia de ruido el rastreo mínimo queda retrasado.

[0097] El rastreo de PSD de ruido basado en MMSE, con baja complejidad [HHJ10] introduce un enfoque de PSD de ruido de fondo utilizando una búsqueda de MMSE usada en un espectro de DFT (Transformada Discreta de Fourier). El algoritmo consiste en estas etapas de procesamiento:

- Se calcula el máximo estimador de probabilidad basándose en la PSD de ruido de la trama previa.
- Se computa el mínimo estimador de cuadrados medios.
- Se estima el máximo estimador de probabilidad usando el enfoque dirigido por decisión [EM84].
- 10 - Se computa el factor de inclinación inversa asumiendo que los coeficientes de DFT de la voz y ruido son distribuidos en forma gaussiana.
- Se suaviza la densidad espectral de potencia de ruido estimada.

[0098] Existe además un enfoque de red de seguridad aplicado con el fin de evitar un completo estancamiento del algoritmo.

[0099] El rastreo de ruido no estacionario basado en la estimación de potencia de ruido recursiva conducida por datos [EH08] introduce un método para la estimación de la varianza espectral de ruido a partir de señales de la voz contaminadas por fuentes de ruido altamente no estacionarias. Este método también está usando el suavizado en la dirección de tiempo/frecuencia.

[0100] Un algoritmo de estimación de ruido de baja complejidad basado en el suavizado de la estimación de potencia de ruido y la corrección de la inclinación de estimación [Yu09] mejora el enfoque introducido en [EH08]. La principal diferencia es que la función de ganancia espectral para la estimación de potencia de ruido es hallada por un método conducido por datos iterativos.

[0101] Los métodos estadísticos para la mejora de la voz ruidosa [Mar03] combinan el enfoque de mínima estadística proporcionado en [Mar01] por la modificación de ganancia de decisión blanda [MCA99], por una estimación de la SNR *a priori* [MCA99], por una limitación de ganancia adaptativa [MC99] y por un estimador de amplitud espectral log MMSE [EM85].

[0102] El desvanecimiento es de particular interés para una pluralidad de códecs de la voz y audio, en particular AMR (véase [3GP12b]) (incluyendo ACELP y CNG), AMR-WB (véase [3GP09c]) (incluyendo ACELP y CNG), AMR-WB+ (véase [3GP09a]) (incluyendo ACELP, TCX y 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-AAC/aacPlus mejorado (véase [EBU10, EBU12, 3GP12e, LS01, QD03]) (incluyendo AAC y SBR), MPEG-4 HILN (véase [ISO09, MEP01]) y OPUS (véase [IET12]) (incluyendo SILK y CELT).

[0103] Dependiendo del códec, el desvanecimiento se realiza en diferentes dominios:

Para códecs que utilizan LPC, el desvanecimiento se realiza en el dominio predictivo lineal (también conocido como el dominio de excitación). Esto es cierto para códecs que se basan en ACELP, por ejemplo, AMR, AMR-WB, el núcleo de ACELP de AMR-WB+, G.718, G.729, G.729.1, el núcleo SILK en OPUS; códecs que además procesan la señal de excitación usando una transformación de tiempo-frecuencia, por ejemplo, el núcleo de TCX de AMR-WB+, el núcleo de CELT en OPUS; y para los esquemas de generación de ruido de confort (CNG) que operan en el dominio predictivo lineal, por ejemplo, CNG en AMR, CNG en AMR-WB, CNG en AMR-WB+.

[0104] 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 basan en la MDCT o una transformación similar, tal como AAC en MPEG-4 HE-AAC, G.719, G.722 (dominio de subbanda) y G.722.1.

[0105] Para códecs paramétricos, el desvanecimiento se aplica en el dominio paramétrico. Esto es cierto para MPEG-4 HILN.

[0106] 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, si bien se utiliza además una aplicación a modo de muestra, véase, por ejemplo, G.718 y G.722.

[0107] El factor de atenuación para un cierto segmento de señal podría ser proporcionado de dos maneras, absoluta y relativa.

[0108] En el caso de que se provea un factor de atenuación de manera absoluta, el nivel de referencia es siempre aquel de la última trama recibida. Los factores de atenuación absolutos habitualmente inician con un valor cercano a 1 para el segmento de señal inmediatamente después de la última buena trama, y luego, se degradan en forma más veloz 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 buena trama, un factor de atenuación absoluto $\alpha_{abs}(n)$, el factor de ganancia de cualquier trama perdida posterior puede derivarse como

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

[0109] En el caso en el que se proporciona un factor de atenuación de manera relativa, el nivel de referencia es aquel 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 nuevamente.

[0110] 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 anteriormente); 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 posteriores tramas (véase anteriormente); 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 veloz 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]);

[0111] Asumiendo un factor de desvanecimiento relativo $0 \geq \alpha_{rel}(n) \geq 1$, mientras que n es el número de la trama perdida ($n \leq 1$); el factor de ganancia de cualquier trama posterior puede derivarse como

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

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

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

dando como resultado un desvanecimiento exponencial.

[0112] Con respecto al procedimiento de desvanecimiento, habitualmente, se especifica el factor de atenuación, si bien en algunos estándares de aplicación (DRM, DAB+) este último se deja al fabricante.

[0113] Si diferentes partes de señal son desvanecidas en forma separada, podrían aplicarse diferentes factores de atenuación, por ejemplo, con el 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, AMR, SILK).

[0114] Habitualmente, se aplica una cierta ganancia 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 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.

[0115] 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. Es decir, por ejemplo, el caso para AAC, donde la duración de la trama depende de la tasa de muestreo.

[0116] 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 AMR, donde se tiene en cuenta la mediana de los cinco factores de ganancia previos (véase [3GP12b] y 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. Además, dicho ajuste dinámico adicional se aplica, por ejemplo, para G729, donde se predice la amplitud usando la regresión lineal de los factores de ganancia previos (véase [CPK08, PKJ+11] y 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.

10 **[0117]** Con respecto al nivel objetivo del desvanecimiento, excepto G.718 y CELT, el nivel objetivo es 0 para todos los códecs analizados, que incluyen aquellos códecs de generación de ruido de confort (CNG).

[0118] 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 en forma separada. Si bien el factor de ganancia de altura se desvanece a cero, el factor de ganancia de innovación se desvanece a la energía de excitación de CNG.

[0119] Asumiendo que se proporcionan los factores de atenuación relativos, esto conduce - basándose en la fórmula (23) - al siguiente factor de atenuación absoluto:

$$20 \quad g(n) = \alpha_{rel}(n) \cdot g(n-1) + (1 - \alpha_{rel}(n)) \cdot g_n \quad (25)$$

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

25 **[0120]** G.718 no realiza desvanecimiento en el caso de DTX/CNG.

[0121] En CELT no hay desvanecimiento hacia el nivel objetivo, si bien, 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).

[0122] Con respecto a la forma espectral objetivo del desvanecimiento, todos los códecs basados en la transformada pura analizados (AAC, G.719, G.722, G.722.1), al igual que la SBR, simplemente prolongan la forma espectral de la última buena trama durante el desvanecimiento.

35 **[0123]** Diversos códecs de la voz desvanecen la forma espectral a una media, usando la síntesis de LPC. La media podría ser estática (AMR) o adaptativa (AMR-WB, AMR-WB+, G.718), mientras que esta última se deriva a partir de una media estática y una media de corto plazo (derivada promediando los últimos grupos de coeficientes n LP) (LP = Predicción Lineal).

40 **[0124]** Todos los módulos de CNG en los códecs descritos de AMR, AMR-WB, AMR-WB+, G.718 prolongan la forma espectral de la última buena trama durante el desvanecimiento.

[0125] Con respecto al rastreo del nivel de ruido de fondo, existen cinco enfoques diferentes conocidos de la bibliografía:

- Basado del detector de actividad de voz: se basa en SNR/VAD, pero muy difícil de sintonizar y de usar para la voz SNR baja.
- Esquema de decisión blanda: El enfoque de decisión blanda tiene en cuenta la probabilidad de la presencia de la voz [SS98] [MPC89] [HE95],
- Mínima estadística: Se rastrea la mínima de la PSD manteniendo una cierta cantidad de valores con el tiempo en un búfer, permitiendo así encontrar el mínimo ruido de las muestras pasadas [Mar01] [HHJ10] [EH08] [Yu09].
- Filtración de Kalman: El algoritmo usa una serie de mediciones observadas en función del tiempo, que contienen ruido (variaciones aleatorias), y produce estimaciones de la PSD de ruido que tienden a ser más precisas que aquellas basándose en 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: Este enfoque 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 DFT (Transformada Discreta de Fourier). A continuación, pueden rastrearse vectores propios/valores propios usando un algoritmo de suavizado arbitrario [BP06] [HJH08].

[0126] El documento EP 2 026 330 A1 describe un dispositivo y un método para el ocultamiento perdido de la trama. Se obtiene un periodo de tono de una trama perdida actual sobre la base de un periodo de altura de la última trama buena antes de la trama perdida actual. Una señal de excitación de la trama perdida actual se recupera sobre la base del periodo de altura de la trama perdida actual y una señal de excitación de la última trama buena antes de la trama perdida. De esta manera, se reduce el contraste auditivo de un receptor, y se mejora la calidad vocal. Además, en el documento EP 2 026 330 A1, se ajusta un periodo de altura de tramas perdidas continuas en base a la tendencia de cambio del periodo de altura de la última trama bueno antes de la trama perdida.

[0127] El documento EP 1 775 717 A1 describe un dispositivo de decodificación de audio capaz de mejorar la calidad de audio de una señal decodificada considerando el cambio de energía de una señal pasada en el procesamiento de ocultación de borrado. En este dispositivo, una unidad de cálculo de cambio de energía calcula una energía media de una señal de fuente de audio de un ciclo de paso desde el final del vector ACB emitido a partir de un libro de códigos adaptativo. Además, la unidad de cálculo de cambio de energía calcula una relación de la energía media de la subtrama actual y la subtrama inmediatamente antes y emite la relación a una unidad de generación de ganancia ACB. La unidad de generación de ganancia de ACB emite una ganancia ACB de procesamiento de ocultación definida por la ganancia de ACB descodificada en el pasado o información sobre la relación de cambio de energía emitida desde la unidad de cálculo de cambio de energía, a un multiplicador.

[0128] 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 método según la reivindicación 12 y por un programa informático según la reivindicación 13. Se proporciona un aparato para la decodificación de una señal de audio codificada para obtener 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 pero está alterada, uno o más primeros coeficientes de señal de audio, que están comprendidos por la trama actual, en el que dichos uno o más primeros coeficientes de señal de audio indican una característica de la señal de audio codificada, y uno o más coeficientes de ruido que indican una forma espectral de 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, que dependen de los uno o más primeros coeficientes de señal de audio y que dependen 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 pero está alterada. El reconstructor de señal de audio está configurado para la reconstrucción de una primera porción de la señal de audio reconstruida que depende de los 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 la reconstrucción de una segunda porción de la señal de audio reconstruida que depende 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.

[0129] En algunas realizaciones, el uno o más primeros coeficientes de señal de audio, por ejemplo, pueden ser uno o más coeficientes de filtro predictivo lineal de la señal de audio codificada. En algunas realizaciones, el uno o más primeros coeficientes de señal de audio, por ejemplo, pueden ser uno o más coeficientes de filtro predictivo lineal de la señal de audio codificada.

[0130] Según una realización, el 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 una realización, el uno o más coeficientes de filtro predictivo lineal, por ejemplo, pueden representar una forma espectral del ruido de fondo.

[0131] En una realización, el generador de coeficientes, por ejemplo, puede estar configurado para determinar una o más segundas porciones de señal de audio de 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 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.

[0132] Según una realización, el generador de coeficientes, por ejemplo, puede estar configurado para generar el uno o más segundos coeficientes de señal de audio mediante la aplicación de la fórmula:

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

en la que $f_{\text{actual}}[i]$ indica uno de los uno o más segundos coeficientes de señal de audio,

- 5 en la que $f_{\text{as}}[i]$ indica uno de los uno o más primeros coeficientes de señal de audio, en la que $pt_{\text{media}}[i]$ es uno de los uno o más coeficientes de ruido, en la que α es un número real con $0 \leq \alpha \leq 1$, y en la que i es un índice. En una realización, $0 < \alpha < 1$.

- 10 **[0133]** Según una realización, $f_{\text{último}}[i]$ indica un coeficiente de filtro predictivo lineal de la señal de audio codificada, y en la que $f_{\text{actual}}[i]$ indica un coeficiente de filtro predictivo lineal de la señal de audio reconstruida.

[0134] En una realización, $pt_{\text{media}}[i]$ puede indicar, por ejemplo, el ruido de fondo de la señal de audio codificada.

15

[0135] En una realización, el generador de coeficientes, por ejemplo, puede estar configurado 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, el uno o más coeficientes de ruido mediante la determinación de un espectro de ruido de la señal de audio codificada.

20

[0136] Según una realización, el generador de coeficientes, por ejemplo, puede estar configurado para determinar coeficientes de LPC que representan ruido de fondo usando un enfoque 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.

25

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

- Recibir una o más tramas.

- 30 - Determinar, si una trama actual de una o más tramas es recibida y si la trama actual recibida no está alterada, de uno o más primeros coeficientes de señal de audio, estando comprendidos por la trama actual, en el que 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.

- Generar uno o más segundos coeficientes de señal de audio, que dependen de los uno o más primeros coeficientes de señal de audio y que dependen de los uno o más coeficientes de ruido, si la trama actual no es recibida o si la trama actual recibida está alterada.

- 35 - Reconstruir una primera porción de la señal de audio reconstruida que depende de los uno o más primeros coeficientes de señal de audio, si la trama actual es recibida y si la trama de actual que se recibe no está alterada. Y:

- Reconstruir una segunda porción de la señal de audio reconstruida que depende de los uno o más segundos coeficientes de señal de audio, si la trama actual no es recibida o si la trama de actual que se recibe está alterada.

40

[0138] Además, se proporciona un programa informático para la implementación del método anteriormente descrito cuando se ejecuta en un ordenador o procesador de señales.

- 45 **[0139]** 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 un simple enfoque común. 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, que no es posible para el núcleo CELP.

50

[0140] Por el contrario, en el estado de la técnica, 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 de 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 el estado de la técnica, puede emplearse una forma de ruido de fondo en tren fuera de línea (estática) que pueda ser placentera en términos de sonido para señales particulares, si bien menos placentera para otras, por ejemplo, sonidos de ruido de coches totalmente diferentes al ruido de oficina.

55

[0141] Además, en el estado de la técnica, puede emplearse una adaptación a la media de 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 el estado de la técnica, el rastreo de la forma espectral a modo de banda en el dominio espectral (como se representa en CELT [IET12]) no es aplicable para un códec cambiado usando no solo un núcleo basado en el dominio de la MDCT (TCX), sino, además, un núcleo basado en ACELP. Las realizaciones mencionadas anteriormente, en consecuencia, son ventajosas en comparación con la técnica anterior.

[0142] Además, se proporciona un aparato para la decodificación de una señal de audio.

[0143] El aparato comprende una interfaz receptora. La interfaz receptora está configurada para recibir una pluralidad de tramas, en el que la interfaz receptora 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, estando dicha primera porción de señal de audio representada en un primer dominio, y en el que 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.

[0144] 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 derivado de la segunda porción de señal de audio de un segundo dominio a un dominio de rastreo para obtener una información de segunda porción de señal, en el que el segundo dominio es diferente del primer dominio, en el que el dominio de rastreo es diferente del segundo dominio, y en el que el dominio de rastreo es igual o diferente del primer dominio.

[0145] Adicionalmente, el aparato comprende una unidad de rastreo de nivel de ruido, en el que la unidad de rastreo de nivel de ruido está configurada para recibir una información de primera porción de señal que es representada en el dominio de rastreo, en el que 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 en el que 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 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.

[0146] Además, el aparato comprende una unidad de reconstrucción para la reconstrucción de 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 pero está alterada.

[0147] Una señal de audio, por ejemplo, puede ser una señal de la voz, o una señal de música, o señal que comprende la voz y música, etc.

[0148] 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 dependiendo de 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 para obtener la información de primera porción de señal.

[0149] 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 dependiendo de 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.

[0150] En una realización, 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, 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 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.

[0151] Según una realización, 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.

[0152] En una realización, la primera porción de señal de audio, por ejemplo, puede ser representada en un dominio de excitación como el primer dominio, en la que la unidad de rastreo de nivel de ruido, por ejemplo, puede ser configurada para recibir la información de primera porción de señal, en la que dicha información de primera porción de señal está representada en el dominio FFT, que es el dominio de rastreo, y en la que dicha información de primera porción de señal depende de dicha primera porción de señal de audio que está representada en el dominio de excitación, en la que 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 en la que 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.

[0153] En una realización, el aparato, por ejemplo, puede comprender adicionalmente una primera unidad de agregación para determinar un primer valor agregado dependiendo de la primera porción de señal de audio. Además, el aparato, por ejemplo, puede comprender adicionalmente una segunda unidad de agregación para determinar, dependiendo de la segunda porción de señal de audio, 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, en la que 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 representada en el dominio de rastreo, y en la que la unidad de rastreo de nivel de ruido, por ejemplo, puede estar configurada para determinar 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.

[0154] Según una realización, la primera unidad de agregación por ejemplo, puede estar configurada para determinar el primer valor agregado de manera que el primer valor agregado indique 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 manera que el segundo valor agregado indique una media cuadrática de la segunda porción de señal de audio o de una señal derivada de la segunda porción de señal de audio.

[0155] En una realización, 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.

[0156] Según realizaciones, 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.

[0157] En una realización, la unidad de rastreo de nivel de ruido, por ejemplo, puede estar configurada para determinar información de nivel de ruido mediante la aplicación de un enfoque de estadística mínima.

[0158] Según una realización, 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 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 está alterada.

[0159] En una realización, 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, en la que dicho espectro de nivel de ruido se obtiene mediante la aplicación del enfoque de estadística mínima. La unidad de reconstrucción, por ejemplo, puede ser configurada 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 está alterada.

[0160] Según otra realización, 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 dependiendo de la pluralidad de coeficientes predictivos lineales.

[0161] En una realización, 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 dependiendo de 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 está alterada.

[0162] En una realización, la unidad de reconstrucción, por ejemplo, puede ser configurada 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 está alterada.

[0163] Según una realización, 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.

[0164] En una realización, el aparato, por ejemplo, puede comprender adicionalmente una unidad de predicción de largo plazo que comprende un búfer de retardo. Además, la unidad de predicción de largo plazo, por ejemplo, puede ser configurada 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 de largo plazo. Adicionalmente, la unidad de predicción de largo plazo, por ejemplo, puede ser configurada para desvanecer la ganancia de predicción de 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 está alterada.

[0165] Según una realización, la unidad de predicción de largo plazo, por ejemplo, puede ser configurada para desvanecer la ganancia de predicción de largo plazo hacia cero, en la que una velocidad con la cual la ganancia de predicción de largo plazo es desvanecida a cero depende de un factor de desvanecimiento.

[0166] En una realización, la unidad de predicción de 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 está alterada.

[0167] Según una realización, 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 adicionalmente comprende 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 está alterada. Además, la segunda unidad de reconstrucción, por ejemplo, puede ser configurada para reconstruir 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, 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 está alterada.

[0168] En una realización, la segunda unidad de reconstrucción, por ejemplo, puede ser configurada 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.

[0169] Según una realización, la segunda unidad de reconstrucción, por ejemplo, puede ser configurada 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.

5

[0170] Además, se proporciona un método para la decodificación de una señal de audio.

[0171] El método comprende:

10 - Recibir 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.

- 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.

15 - Transformación de la segunda porción de señal de audio o un valor o una señal derivado de la segunda porción de señal de audio de un segundo dominio a un dominio de rastreo para obtener una información de segunda porción de señal, en el que el segundo dominio es diferente del primer dominio, en el que el dominio de rastreo es diferente del segundo dominio, y en el que el dominio de rastreo es igual o diferente del primer dominio.

- Determinar 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, en el que la información de primera porción de señal depende de la primera porción de señal de audio. Y:

20 - Reconstruir 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
25 si dicha tercera trama es recibida pero está alterada.

[0172] Adicionalmente, se proporciona un programa informático para la implementación del método anteriormente descrito cuando se ejecuta en un ordenador o procesador de señales.

30 **[0173]** Algunas de las realizaciones de la presente invención proporcionan un parámetro de suavizado de variación de tiempo, de manera que las capacidades de rastreo del periodograma suavizado y su varianza sean mejor equilibrados, para desarrollar un algoritmo para la compensación de inclinación, y para acelerar el rastreo de ruido, en general.

35 **[0174]** Las realizaciones de la presente invención se basan 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, las realizaciones se basan en el hallazgo de que la técnica anterior tiene desventajas significativas.

40

[0175] Se proporciona un aparato y un método para la mejora del desvanecimiento de la señal para sistemas de codificación de audio cambiados durante el ocultamiento de errores.

45 **[0176]** Además, se proporciona un programa informático para la implementación del método anteriormente descrito cuando se ejecuta en un ordenador o procesador de señales.

[0177] Las realizaciones representan un desvanecimiento al nivel de ruido de confort. Según las realizaciones, 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, independientemente del codificador núcleo (ACELP/TCX) en uso, y siempre estará actualizado. No existe estado de la técnica conocido, donde sea necesario un rastreo de nivel de ruido común. Las realizaciones proporcionan el desvanecimiento de un códec cambiado a una señal de tipo ruido de confort durante las pérdidas de paquetes de ráfagas.

50

[0178] Además, las realizaciones 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) y la memoria.

[0179] En las realizaciones, la derivación de nivel en el dominio de excitación (en comparación con la derivación de nivel en el dominio de tiempo) proporciona más mínimos durante la voz activa, ya que parte de la información de la voz es cubierta por los coeficientes de LP.

60

[0180] En el caso de ACELP, según las realizaciones, la derivación de nivel tiene lugar en el dominio de excitación. En el caso de TCX, en las realizaciones, el nivel es derivado en el dominio de tiempo, y la ganancia de la síntesis de LPC y el desénfasis se aplica como un factor de corrección con el 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 FDNS, teóricamente también sería posible, si bien la compensación de nivel entre el dominio de excitación de TCX y el dominio de excitación de ACELP se considera bastante compleja.

[0181] 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, las realizaciones son convenientes sobre la técnica anterior, ya que, en cuanto a las técnicas del estado de la técnica, el nivel de ruido de confort que es dirigido durante las pérdidas de paquetes de ráfagas puede ser diferente, dependiendo del modo de codificación precedente (ACELP/TCX), donde el nivel fue rastreado; ya que en el estado de la técnica, el rastreo que es separado para cada modo de codificación causará complejidad computacional adicional e innecesario en exceso; y ya que, en el estado de la técnica, podría no haber disponible ningún nivel de ruido de confort actualizado en ninguno de los núcleos, debido a cambio reciente a este núcleo.

[0182] Según algunas realizaciones, el rastreo de nivel es conducido en el dominio de excitación, si bien el desvanecimiento de TCX se realiza en el dominio de tiempo. Mediante el desvanecimiento en el dominio de tiempo, se evitan fallos de la TDAC, que causarían la superposición periódica sucesiva. Esto llega a ser de particular interés cuando se ocultan componentes de señal tonal. Además, la conversión de nivel entre el dominio de excitación de ACELP y el dominio espectral de la MDCT se evita, y en consecuencia, por ejemplo, se ahorran recursos de cálculo. 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 es resuelto por la derivación de la ganancia que sería introducida por la síntesis de LPC y el preénfasis y el uso de esta ganancia como un factor de corrección para la conversión del nivel entre los dos dominios.

[0183] Por el contrario, las técnicas del estado de la técnica no realizan el rastreo de nivel en el dominio de excitación y el desvanecimiento de TCX en el dominio de tiempo. Con respecto a los códecs basados en la transformada del estado de la técnica, se aplica el factor de atenuación o bien en el dominio de excitación (para enfoques de ocultamiento de tipo dominio de tiempo/ACELP), véase [3GP09a] o en el dominio de frecuencia (para enfoques de dominio de frecuencia como la repetición de trama o la sustitución de ruido, véase [LS01]). Una desventaja del enfoque del estado de la técnica para la aplicación del factor de atenuación en el dominio de frecuencia es que la superposición periódica sucesiva 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 fallo de la TDAC (cancelación de la superposición periódica sucesiva del dominio de tiempo). Esto es en particular relevante cuando se ocultan los componentes de señal tonal. Las realizaciones mencionadas anteriormente, en consecuencia, son convenientes en comparación con la técnica anterior.

[0184] Las realizaciones compensan la influencia del filtro paso alto sobre la ganancia de la síntesis de LPC. Según realizaciones, para compensar el cambio de ganancia no deseado del análisis de 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 manera que se alcance el nivel objetivo correcto en el dominio de tiempo.

[0185] Por el contrario, el estado de la técnica, por ejemplo, G.718 [ITU08a], introduce un filtro paso alto en la vía de la señal de la excitación sin voz, como se representa en la Fig. 2, si la señal de la última buena trama no fue clasificada como SIN VOZ. Al realizar esto, las técnicas del estado de la técnica causan efectos secundarios no deseados, ya que la ganancia de la posterior síntesis de LPC depende de las características de la señal, que son alteradas por este filtro paso alto. Debido a que el nivel de fondo es rastreado y aplicado en el dominio de excitación, el algoritmo se basa en la ganancia de la síntesis de 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 paso alto, como es conducida por el estado de la técnica, podría conducir a una ganancia modificada (habitualmente reducida) de la síntesis de LPC. Esto conduce a un nivel de salida erróneo, aún cuando el nivel de excitación sea correcto.

[0186] Las realizaciones superan estas desventajas de la técnica anterior.

60

[0187] En particular, las realizaciones representan una forma espectral adaptativa de ruido de confort. A diferencia de 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 precedente coincidirá, conduciendo a una característica de ruido agradable del ruido de confort. Esto evita no coincidencias
5 demasiado visibles de la forma espectral que pueden ser introducidas mediante el uso de una envoltura espectral derivada en tren fuera de línea y/o la forma espectral de las últimas tramas recibidas.

[0188] Además, se proporciona un aparato para la decodificación de una señal de audio. El aparato comprende una interfaz receptora, en la que la interfaz receptora está configurada para recibir una primera trama
10 que comprende una primera porción de señal de audio de la señal de audio, y en la que la interfaz receptora está configurada para recibir una segunda trama que comprende una segunda porción de señal de audio de la señal de audio.

[0189] Además, el aparato comprende una unidad de rastreo de nivel de ruido, en el que la unidad de rastreo de nivel de ruido está configurada para determinar la información de nivel de ruido que depende al menos de una de
15 la primera porción de señal de audio y la segunda porción de señal de audio (esto significa: que depende de la primera porción de señal de audio y/o la segunda porción de señal de audio), en la que la información de nivel de ruido se representa en un dominio de rastreo.

[0190] 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 está alterada, en el que el primer dominio de reconstrucción es diferente o igual del dominio de rastreo.
20

[0191] 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 está alterada, en el que el segundo dominio de reconstrucción es diferente del dominio de rastreo, y
25 en el que el segundo dominio de reconstrucción es diferente del primer dominio de reconstrucción; y

[0192] 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 dependiendo de la información de nivel de ruido que es representada en el segundo dominio de reconstrucción, si
35 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 está alterada.

[0193] Según algunas realizaciones, el dominio de rastreo, por ejemplo, puede ser en el que el dominio de rastreo es un dominio de tiempo, un dominio espectral, un dominio FFT, un dominio de MDCT, o un dominio de excitación. El primer dominio de reconstrucción, por ejemplo, puede ser el dominio de tiempo, el dominio espectral, el dominio FFT, el dominio de MDCT, o el dominio de excitación. El segundo dominio de reconstrucción, por ejemplo, puede ser el dominio de tiempo, el dominio espectral, el dominio FFT, el dominio de MDCT, o el dominio de excitación.
40

[0194] En una realización, el dominio de rastreo, por ejemplo, puede ser el dominio FFT, el primer dominio de reconstrucción, por ejemplo, puede ser el dominio de tiempo, y el segundo dominio de reconstrucción, por ejemplo, puede ser el dominio de excitación.
45

[0195] En otra realización, el dominio de rastreo, por ejemplo, puede ser el dominio de tiempo, el primer dominio de reconstrucción, por ejemplo, puede ser el dominio de tiempo, y el segundo dominio de reconstrucción, por ejemplo, puede ser el dominio de excitación.
50

[0196] Según una realización, dicha primera porción de señal de audio, por ejemplo, puede ser representada en un primer dominio de entrada, y dicha segunda porción de señal de audio, por ejemplo, puede ser representada 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 derivado de la segunda porción de señal de audio, del segundo dominio de entrada al dominio de rastreo, para obtener una información de segunda porción de señal. La unidad de rastreo de nivel de ruido, por ejemplo, puede ser configurada para recibir una
55 información de primera porción de señal que es representada en el dominio de rastreo, en la que la información de
60

primera porción de señal depende de la primera porción de señal de audio, en la que 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 en la que 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 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.

[0197] Según una realización, el primer dominio de entrada, por ejemplo, puede ser el dominio de excitación, y el segundo dominio de entrada, por ejemplo, puede ser el dominio de MDCT.

10 **[0198]** En otra realización, el primer dominio de entrada, por ejemplo, puede ser el dominio de MDCT, y donde el segundo dominio de entrada, por ejemplo, puede ser el dominio de MDCT.

[0199] Según una realización, la primera unidad de reconstrucción, por ejemplo, puede ser configurada 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, por ejemplo, puede ser configurada 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 LTP. Además, la primera unidad de reconstrucción y la segunda unidad de reconstrucción, por ejemplo, pueden ser configuradas para conducir el primer desvanecimiento y el segundo desvanecimiento a un espectro de tipo ruido y/o un segundo desvanecimiento de una ganancia de LTP con la misma velocidad de desvanecimiento.

[0200] En una realización, el aparato, por ejemplo, puede comprender adicionalmente una primera unidad de agregación para determinar un primer valor agregado dependiendo de la primera porción de señal de audio. Además, el aparato, por ejemplo, puede comprender además una segunda unidad de agregación para determinar, dependiendo de la segunda porción de señal de audio, un segundo valor agregado como el valor derivado de la segunda porción de señal de audio. 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, en la que 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 representada en el dominio de rastreo, y en la que 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.

[0201] Según una realización, la primera unidad de agregación por ejemplo, puede estar configurada para determinar el primer valor agregado de manera que el primer valor agregado indique 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 manera que el segundo valor agregado indique una media cuadrática de la segunda porción de señal de audio o de una señal derivada de la segunda porción de señal de audio.

[0202] En una realización, la primera unidad de transformada, por ejemplo, puede ser configurada 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.

[0203] Según una realización, el valor de ganancia, por ejemplo, puede indicar una ganancia introducida por la síntesis de codificación predictiva lineal, o en el que el valor de ganancia indica una ganancia introducida por la síntesis de codificación predictiva lineal y el desénfasis.

50 **[0204]** En una realización, la unidad de rastreo de nivel de ruido, por ejemplo, puede estar configurada para determinar la información de nivel de ruido mediante la aplicación de un enfoque de estadística mínima.

[0205] Según una realización, 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 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 está alterada.

[0206] En una realización, 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, en la que dicho espectro de nivel de ruido se obtiene mediante la aplicación del enfoque de estadística mínima. La unidad de reconstrucción, por ejemplo, puede ser configurada 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 está alterada.

[0207] Según una realización, la primera unidad de reconstrucción, por ejemplo, puede ser configurada 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 está alterada.

[0208] En una realización, la primera unidad de reconstrucción, por ejemplo, puede ser configurada para reconstruir la tercera porción de señal de audio mediante la atenuación o amplificación de la primera porción de señal de audio.

[0209] Según una realización, la segunda unidad de reconstrucción, por ejemplo, puede ser configurada 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.

[0210] En una realización, la segunda unidad de reconstrucción, por ejemplo, puede ser configurada para reconstruir la cuarta porción de señal de audio mediante la atenuación o amplificación de la segunda porción de señal de audio.

[0211] Según una realización, el aparato, por ejemplo, puede comprender adicionalmente una unidad de predicción de largo plazo que comprende un búfer de retardo, en el que la unidad de predicción de largo plazo, por ejemplo, puede ser configurada 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 de largo plazo, y en el que la unidad de predicción de largo plazo está configurada para desvanecer la ganancia de predicción de 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 está alterada.

[0212] En una realización, la unidad de predicción de largo plazo, por ejemplo, puede ser configurada para desvanecer la ganancia de predicción de largo plazo hacia cero, en la que una velocidad con la que la ganancia de predicción de largo plazo es desvanecida a cero depende de un factor de desvanecimiento.

[0213] En una realización, la unidad de predicción de 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 está alterada.

[0214] Además, se proporciona un método para la decodificación de una señal de audio. El método comprende:

- 45 - Recibir 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.
- Determinar 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, en el que la información de nivel de ruido es representada en un dominio de rastreo.
- 50 - Reconstruir, 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 está alterada, donde el primer dominio de reconstrucción es diferente o igual del dominio de rastreo.
- Transformar 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 está alterada, en el que el segundo dominio de reconstrucción es diferente del dominio de rastreo, y en el que el segundo dominio de reconstrucción es diferente del primer dominio de reconstrucción. Y:
- Reconstruir, 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 o si dicha cuarta trama es recibida pero está alterada.

[0215] Además, se proporciona un programa informático para la implementación del método anteriormente descrito cuando se ejecuta en un ordenador o procesador de señales.

5 **[0216]** Además, se proporciona un aparato para la decodificación de una señal de audio codificada para obtener una señal de audio reconstruida. El aparato comprende una interfaz receptora 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 para la generación de la señal de audio reconstruida. El procesador está configurado para generar la señal de audio reconstruida mediante el
 10 desvanecimiento de un espectro modificado a un espectro objetivo, si una trama actual no es recibida por la interfaz receptora o si la trama actual es recibida por la interfaz receptora pero está alterada, en el que el espectro modificado comprende una pluralidad de muestras de señales modificadas, en el que, 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. Además, el
 15 procesador 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 y si la trama actual recibida por la interfaz receptora no está alterada

20 **[0217]** Según una realización, el espectro objetivo puede ser, por ejemplo, un espectro de tipo ruido.

[0218] En una realización, el espectro de tipo ruido puede representar, por ejemplo, ruido blanco.

[0219] Según una realización, el espectro de tipo ruido puede ser, por ejemplo, con forma.

25 **[0220]** En una realización, la forma del espectro de tipo ruido puede depender, por ejemplo, de un espectro de señales de audio de una señal previamente recibida.

[0221] Según una realización, el espectro de tipo ruido puede ser, por ejemplo, con forma dependiendo de la forma del espectro de señales de audio.

30 **[0222]** En una realización, el procesador puede emplear, por ejemplo, un factor de inclinación para dar forma al espectro de tipo ruido.

35 **[0223]** Según una realización, el procesador puede emplear, por ejemplo, la fórmula

$$\text{shaped_noise}[i] = \text{ruido} \times \text{potencia}(\text{tilt_factor}, i/N)$$

en la que N indica el número de muestras, en la que i es un índice, en la que $0 \leq i < N$, con $\text{tilt_factor} > 0$, y en la que potencia es una función de potencia.

40 potencia (x, y) indica x^y

potencia (tilt_factor, i/N) indica $\text{tilt_factor}^{\frac{i}{N}}$

45 **[0224]** Si el tilt_factor es menor de 1, esto significa la atenuación con i creciente. Si el tilt_factor es mayor que 1, esto significa la amplificación con i creciente.

[0225] Según otra realización, el procesador puede emplear, por ejemplo, la fórmula

50
$$\text{shaped_noise}[i] = \text{ruido} \times (1 + i/(N-1) * (\text{tilt_factor} - 1))$$

en la que N indica el número de muestras, en la que i es un índice, en la que $0 \leq i < N$, con $\text{tilt_factor} > 0$.

55 **[0226]** Si el factor de inclinación es menor de 1, esto significa la atenuación con i creciente. Si el tilt_factor es mayor que 1, esto significa la amplificación con i creciente.

[0227] Según una realización, el procesador puede estar configurado, por ejemplo, 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 o si la trama actual recibida por la interfaz receptora está alterada.

[0228] En una realización, cada una de las muestras de señal de audio del espectro de señales de audio puede estar representada, por ejemplo, por un número real, pero no por un número imaginario.

[0229] Según una realización, las muestras de señal de audio del espectro de señales de audio pueden estar representadas, por ejemplo, en un dominio de transformada de coseno discreta modificada.

[0230] En otra realización, las muestras de señal de audio del espectro de señales de audio pueden estar representadas, por ejemplo, en un dominio de transformada de seno discreta modificada.

[0231] Según una realización, el procesador puede estar configurado, por ejemplo, para generar el espectro modificado mediante el empleo de una función de signos aleatorios que, en forma aleatoria, o pseudoaleatoria, da salida o bien a un primer o un segundo valor.

[0232] En una realización, el procesador puede estar configurado, por ejemplo, para desvanecer el espectro modificado al espectro objetivo mediante la disminución posterior de un factor de atenuación.

[0233] Según una realización, el procesador puede estar configurado, por ejemplo, para desvanecer el espectro modificado al espectro objetivo mediante el aumento posterior de un factor de atenuación.

[0234] En una realización, si la trama actual no es recibida por la interfaz receptora o si la trama actual recibida por la interfaz receptora está alterada, el procesador puede estar configurado, por ejemplo, para generar la señal de audio reconstruida mediante el empleo de la fórmula:

$$x[i] = (1 - \text{cum_damping}) * \text{noise}[i] + \text{cum_damping} * \text{random_sign}() * x_old[i]$$

en la que i es un índice, en la que $x[i]$ indica una muestra de la señal de audio reconstruida, en la que cum_damping es un factor de atenuación, en la que $x_old[i]$ indica una de las muestras de señal de audio del espectro de señales de audio de la señal de audio codificada, en la que $\text{random_sign}()$ devuelve a 1 o -1, y en la que el ruido es un vector aleatorio que indica el espectro objetivo.

[0235] En una realización, dicho ruido de vector aleatorio, por ejemplo, puede ser escalado de 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 recibió por último lugar por la interfaz receptora.

[0236] Según una general realización, el procesador puede, por ejemplo, estar configurado para generar la señal de audio reconstruida, empleando un vector aleatorio que es escalado de modo 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 han sido recibidas por último lugar por la interfaz receptora.

[0237] Además, se proporciona un método para la decodificación de una señal de audio codificada para obtener una señal de audio reconstruida. El método comprende:

- Recibir una o más tramas que comprenden información sobre 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.

[0238] La generación de la señal de audio reconstruida se realiza mediante el desvanecimiento de un espectro modificado a un espectro objetivo, si una trama actual no es recibida o si la trama actual es recibida pero está alterada, en el que el espectro modificado comprende una pluralidad de muestras de señales modificadas, en el que, 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 señales de audio del espectro de señales de audio. El espectro modificado no se desvanece a un espectro de ruido blanco, si la trama actual de la una o más tramas es recibida y si la trama actual que es recibida no está alterada.

[0239] Además, se proporciona un programa informático para la implementación del método anteriormente

descrito cuando se ejecuta en un ordenador o procesador de señales.

[0240] Las realizaciones realizan un espectro de MDCT de desvanecimiento a ruido blanco antes de la aplicación de FDNS (FDNS = sustitución de ruido de dominio de frecuencia).

5

[0241] Según el estado de la técnica, en códecs basados en ACELP, el innovador libro de código se sustituye por un vector aleatorio (por ejemplo, con ruido). En realizaciones, el enfoque de ACELP, que consiste en sustituir el innovador libro de código con un vector aleatorio (por ejemplo, con ruido) es adoptado para la estructura de decodificador de TCX. Aquí, el equivalente del innovador libro de código es el espectro de MDCT normalmente recibido dentro de la corriente de bits y alimentado en el FDNS.

10

[0242] El enfoque de ocultamiento de MDCT clásico sería repetir simplemente este espectro como tal o aplicarlo a un cierto proceso de aleatorización, que prolonga básicamente la forma espectral de la última trama recibida [LS01]. Esto tiene el inconveniente que la forma espectral de corto plazo se prolonga, conduciendo frecuentemente a un sonido metálico repetitivo que no es de tipo ruido de fondo, y así no puede usarse como ruido de confort.

15

[0243] Usando el método propuesto, la forma espectral de corto plazo se realiza por la FDNS y la LTP de TCX, la forma espectral en la ejecución larga se realiza por la FDNS solo. La forma por la FDNS se desvanece a partir de la forma espectral a corto plazo a la forma espectral de largo plazo rastreada del ruido de fondo, y la LTP de TCX se desvanece a cero.

20

[0244] El desvanecimiento de los coeficientes de FDNS a coeficientes de ruido de fondo rastreados conduce a que tenga una transición suave entre la última buena envoltura espectral y la envoltura de fondo espectral que debe ser el objetivo de la ejecución larga, con el fin de lograr un ruido de fondo agradable en caso de largas pérdidas de trama de ráfaga.

25

[0245] Por el contrario, según el estado de la técnica, para códecs basados en la transformada, se realiza ocultamiento de tipo ruido por repetición de tramas o sustitución de ruido en el dominio de frecuencias [LS01]. En el estado de la técnica, la sustitución de ruido se realiza normalmente por mezcla de signos de los intervalos espectrales. Si en el estado de la técnica la mezcla de signos de TCX (dominio de frecuencia) se usa durante el ocultamiento, los últimos coeficientes de MDCT recibidos se reutilizan y cada signo se aleatoriza antes de que el espectro sea inversamente transformado al dominio de tiempo. El inconveniente de este procedimiento del estado de la técnica es, que para tramas consecutivamente perdidas, el mismo espectro se usa una y otra vez, solo con aleatorizaciones de signo diferentes y atenuación global. Cuando se mira a la envoltura espectral con el tiempo en una rejilla de tiempo gruesa, puede observarse que la envoltura es aproximadamente constante durante la pérdida de tramas consecutiva, debido a que las energías de banda se mantienen constantes relativamente entre sí dentro de una trama y solo son globalmente atenuadas. En el sistema de codificación usado, según el estado de la técnica, los valores espectrales se procesan usando FDNS, con el fin de restaurar el espectro original. Esto significa que si se quiere desvanecer el espectro de MDCT a una cierta envoltura espectral (usando coeficientes de FDNS, por ejemplo, que describen el ruido de fondo actual), el resultado no solo es dependiente de los coeficientes de FDNS, sino también dependiente del espectro previamente decodificado que era de mezcla de signos. Las realizaciones anteriormente mencionadas superan las desventajas del estado de la técnica.

30

35

40

[0246] Las realizaciones se basan en el hallazgo de que es necesario desvanecer el espectro usado para la mezcla de signos a ruido blanco, antes de alimentarlo en el procesamiento de FDNS. De otro modo, el espectro de salida nunca coincidirá con la envoltura objetivo usada para el procedimiento de FDNS.

45

[0247] En realizaciones, se usa la misma velocidad de desvanecimiento para el desvanecimiento de la ganancia de LTP que para el desvanecimiento de ruido blanco.

50

[0248] Además, se proporciona un aparato para la decodificación de una señal de audio codificada para obtener una señal de audio reconstruida. El aparato comprende una interfaz receptora para la recepción de una pluralidad de tramas, un búfer de retardo para el almacenamiento de muestras de señal de audio de la señal de audio codificada, un selector de muestras 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, y un procesador de muestras para el procesamiento de las muestras de señal de audio seleccionadas para obtener muestras de la señal de audio reconstruida de la señal de audio reconstruida. El selector de muestras está configurado para seleccionar, si una trama actual es recibida por la interfaz receptora y si la trama actual recibida por la interfaz receptora no está alterada, la pluralidad de muestras de señal de audio seleccionadas de las muestras de señal de audio almacenadas

55

60

en el búfer de retardo que depende de una información de retraso de altura comprendida por la trama actual. Además, el selector de muestras está configurado para seleccionar, si la trama actual no es recibida por la interfaz receptora o si la trama actual recibida por la interfaz receptora está alterada, la pluralidad de muestras de señal de audio seleccionadas de las muestras de señal de audio almacenadas en el búfer de retardo que depende de una información de retraso de altura compuesta por otra trama recibida previamente por la interfaz receptora.

5 **[0249]** Según una realización, el procesador de muestras, por ejemplo, puede ser configurado para obtener las muestras de señal de audio reconstruida, si la trama actual es recibida por la interfaz receptora y si la trama actual recibida por la interfaz receptora no está alterada, mediante la reescala de las muestras de señal de audio seleccionadas dependiendo de la información de ganancia comprendida por la trama actual. Además, el selector de muestras, por ejemplo, puede ser configurado para obtener las muestras de señal de audio reconstruida, si la trama actual no es recibida por la interfaz receptora o si la trama actual recibida por la interfaz receptora está alterada, mediante la reescala de las muestras de señal de audio seleccionadas dependiendo de la información de ganancia comprendida por dicha otra trama recibida previamente por la interfaz receptora.

10 **[0250]** En una realización, el procesador de muestras, por ejemplo, puede ser configurado para obtener las muestras de señal de audio reconstruida, si la trama actual es recibida por la interfaz receptora y si la trama actual recibida por la interfaz receptora 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 está configurado para obtener las muestras de señal de audio reconstruida, si la trama actual no es recibida por la interfaz receptora o si la trama actual recibida por la interfaz receptora 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.

15 **[0251]** Según una realización, el procesador de muestras, por ejemplo, puede ser configurado para almacenar las muestras de señal de audio reconstruida en el búfer de retardo.

20 **[0252]** En una realización, el procesador de muestras, por ejemplo, puede ser configurado para almacenar las muestras de señal de audio reconstruida en el búfer de retardo antes de que otra trama sea recibida por la interfaz receptora.

25 **[0253]** Según una realización, el procesador de muestras, por ejemplo, puede ser configurado para almacenar las muestras de señal de audio reconstruida en el búfer de retardo después de que otra trama sea recibida por la interfaz receptora.

30 **[0254]** En una realización, el procesador de muestras, por ejemplo, puede ser configurado para reescalar las muestras de señal de audio seleccionadas dependiendo de la información de ganancia para 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.

35 **[0255]** Según una realización, el procesador de muestras, por ejemplo, puede ser configurado 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, y para no almacenar las muestras de señal de audio reescaladas en el búfer de retardo, 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 procesador de muestras está configurado para almacenar las muestras de señal de audio reescaladas en el búfer de retardo y para no almacenar las muestras de señal de audio procesadas en el búfer de retardo, si la trama actual no es recibida por la interfaz receptora o si la trama actual recibida por la interfaz receptora está alterada.

40 **[0256]** Según otra realización, el procesador de muestras, por ejemplo, puede ser configurado para almacenar las muestras de señal de audio procesadas en el búfer de retardo, si la trama actual no es recibida por la interfaz receptora o si la trama actual recibida por la interfaz receptora está alterada.

45 **[0257]** En una realización, el selector de muestras, por ejemplo, puede ser configurado para obtener las muestras de señal de audio reconstruida mediante la reescala de las muestras de señal de audio seleccionadas dependiendo de una ganancia modificada, en la que la ganancia modificada se define según la fórmula:

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

50 en la que ganancia es la ganancia modificada, en la que el selector de muestras, por ejemplo, puede ser configurado

para establecer `past_gain` en la ganancia después de la ganancia, y se ha calculado, y donde el amortiguamiento es un valor real.

[0258] Según una realización, el selector de muestras, por ejemplo, puede ser configurado para calcular la ganancia modificada.

[0259] En una realización, el amortiguamiento, por ejemplo, puede ser definido según: $0 \geq \text{amortiguamiento} \geq 1$.

10 **[0260]** Según una realización, la ganancia de ganancia modificada, por ejemplo, puede ser ajustada a cero, si al menos un número predefinido de tramas no ha sido recibido por la interfaz receptora, desde que una última trama ha sido recibida por la interfaz receptora.

15 **[0261]** Además, se proporciona un método para la decodificación de una señal de audio codificada para obtener una señal de audio reconstruida. El método comprende:

- Recibir una pluralidad de tramas.
- Almacenar muestras de señal de audio de la señal de audio codificada.
- Seleccionar una pluralidad de muestras de señal de audio seleccionadas de las muestras de señal de audio que están almacenadas en el búfer de retardo. Y:
- Procesar las muestras de señal de audio seleccionadas para obtener muestras de la señal de audio reconstruidas de la señal de audio reconstruida.

25 **[0262]** Si una trama actual es recibida, y si la trama actual recibida no está alterada, la etapa de la selección de la pluralidad de muestras de señal de audio seleccionadas de las muestras de señal de audio que están almacenadas en el búfer de retardo se realiza dependiendo de una información de retraso de altura comprendida por la trama actual. Además, si la trama actual no es recibida, o si la trama actual que es recibida está alterada, la etapa de la selección de la pluralidad de muestras de señal de audio seleccionadas de las muestras de señal de audio que están almacenadas en el búfer de retardo se realiza dependiendo de una información de retraso de altura
30 comprendida por otra trama que ha sido recibida previamente por la interfaz receptora.

[0263] Además, se proporciona un programa informático para la implementación del método anteriormente descrito cuando se ejecuta en un ordenador o procesador de señales.

35 **[0264]** Las realizaciones emplean LTP de TCX (LTP de TCX = Predicción de largo plazo de la Excitación codificada por la transformada). Durante la operación normal, la memoria de la LTP de TCX se actualiza con la señal sintetizada, que contiene ruido y componentes tonales reconstruidos.

40 **[0265]** En lugar de deshabilitar la LTP de TCX durante el ocultamiento, su operación normal puede ser continuada durante el ocultamiento con los parámetros recibidos en la última buena trama. Esto conserva la forma espectral de la señal, en particular, aquellos componentes tonales que son modelados por el filtro de LTP.

45 **[0266]** Además, las realizaciones desacoplan el circuito de retroalimentación de LTP de TCX. Una simple continuación de la operación de predicción de LTP de TCX normal introduce ruido adicional, ya que con cada etapa de actualización se introduce más ruido aleatoriamente generado de la excitación de LTP. Los componentes tonales, en consecuencia, se distorsionan cada vez más en función del tiempo, por el ruido agregado.

50 **[0267]** Para superar esto, solo puede ser retroalimentado el búfer de LTP de TCX actualizado (sin la adición de ruido), con el fin de no contaminar la información tonal con ruido aleatorio no deseado.

[0268] Adicionalmente, según las realizaciones, la ganancia de la predicción de LTP de TCX se desvanece a cero.

55 **[0269]** Estas realizaciones se basan en el hallazgo de que la continuación de la LTP de TCX ayuda a conservar las características de señal a corto plazo, pero posee desventajas a largo plazo. La señal transmitida durante el ocultamiento incluirá la información de voz/tonal que estaba presente antes de la pérdida. Especialmente para la voz limpia o la voz sobre ruido de fondo, es extremadamente improbable que un tono o armónico decaiga en forma muy lenta durante un tiempo muy largo. Al continuar la operación de LTP de TCX durante el ocultamiento, en particular, si la actualización de la memoria de LTP es desacoplada (solo se retroalimentan los componentes tonales
60 y no la parte con mezcla de signos), la información de voz/tonal permanecerá presente en la señal ocultada para la

pérdida entera, siendo atenuada solo por el desvanecimiento general al nivel de ruido de confort. Además, es imposible alcanzar la envoltura de ruido de confort durante las pérdidas de paquetes de ráfagas, si la LTP de TCX se aplica durante la pérdida de ráfaga sin ser atenuada en función del tiempo, ya que la señal entonces siempre incorporará la información de voz de la LTP.

5

[0270] Por lo tanto, la ganancia de la LTP de TCX se desvanece hacia cero, de modo que los componentes tonales representados por la LTP se desvanecerán a cero, al mismo tiempo que la señal se desvanece al nivel y forma de señal de fondo, y de manera que el desvanecimiento alcance la envoltura de fondo espectral deseada (ruido de confort) sin incorporar componentes tonales no deseados.

10

[0271] En realizaciones, se usa la misma velocidad de desvanecimiento para el desvanecimiento de la ganancia de LTP que para el desvanecimiento de ruido blanco.

[0272] Por el contrario, en el estado de la técnica, no hay códec de transformación conocido que use la LTP durante el ocultamiento. Para la MPEG-4 LTP [IS009], no existen enfoques de ocultamiento en el estado de la técnica. Otro códec basado en la MDCT del estado de la técnica, que hace uso de una LTP, es CELT, si bien este códec usa un ocultamiento de tipo ACELP para las primeras cinco tramas, y para todas las posteriores tramas, se genera ruido de fondo, que no hace uso de la LTP. Una desventaja del estado de la técnica del no uso de la LTP de TCX es que todos los componentes tonales que están modelados con la LTP desaparecen abruptamente. Además, en los códecs basados en ACELP del estado de la técnica, la operación de LTP es prolongada durante el ocultamiento, y la ganancia del libro de código adaptativo es desvanecida a cero. Con respecto a la operación de circuito de retroalimentación, el estado de la técnica emplea dos enfoques, o bien la excitación entera, por ejemplo, la suma de la excitación innovadora y la adaptativa, es retroalimentada (AMR-WB); o solo es retroalimentada la excitación adaptativa actualizada, por ejemplo, las partes de señal tonal (G.718). Las realizaciones mencionadas anteriormente superan las desventajas del estado de la técnica.

[0273] A continuación, se describen las realizaciones de la presente invención en más detalle con referencia a las figuras, en las que:

- 30 La Fig. 1a ilustra un aparato para la decodificación de una señal de audio según una realización,
 La Fig. 1b ilustra un aparato para la decodificación de una señal de audio según otra realización,
 La Fig. 1c ilustra un aparato para la decodificación de una señal de audio según otra realización, en la que el aparato adicionalmente comprende una primera y una segunda unidades de agregación,
 La Fig. 1d ilustra un aparato para la decodificación de una señal de audio según una realización adicional, en la que
 35 el aparato además comprende una unidad de predicción de largo plazo que comprende un búfer de retardo,
 La Fig. 2 ilustra la estructura del decodificador de G.718,
 La Fig. 3 representa un escenario, en la que el factor de desvanecimiento de G.722 depende de la información de clase,
 La Fig. 4 muestra un enfoque para la predicción de amplitud usando la regresión lineal,
 40 La Fig. 5 ilustra el comportamiento de pérdida de ráfaga de transformada superpuesta de energía restringida (CELT),
 La Fig. 6 muestra un rastreo del nivel de ruido de fondo según una realización en el decodificador durante un modo de operación libre de error,
 La Fig. 7 ilustra la derivación de ganancia de la síntesis de LPC y el desénfasis según una realización,
 45 La Fig. 8 representa la aplicación de nivel de ruido de confort durante la pérdida de paquete según una realización,
 La Fig. 9 ilustra la compensación de ganancia paso alto avanzada durante el ACELP según una realización,
 La Fig. 10 representa el desacoplamiento del circuito de retroalimentación de la LTP durante el ocultamiento según una realización,
 La Fig. 11 ilustra un aparato para la decodificación de una señal de audio codificada para obtener una señal de
 50 audio reconstruida según una realización,
 La Fig. 12 muestra un aparato para la decodificación de una señal de audio codificada para obtener una señal de audio reconstruida según otra realización, y
 La Fig. 13 ilustra un aparato para la decodificación de una señal de audio codificada para obtener una señal de audio reconstruida según una realización adicional, y
 55 La Fig. 14 ilustra un aparato para la decodificación de una señal de audio codificada para obtener una señal de audio reconstruida según otra realización.

[0274] La Fig. 1 a ilustra un aparato para la decodificación de una señal de audio según una realización.

60 **[0275]** El aparato comprende una interfaz receptora 110. La interfaz receptora está configurada para recibir

una pluralidad de tramas, en el que 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, estando dicha primera porción de señal de audio representada en un primer dominio. Además, la interfaz receptora 110 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.

[0276] Además, el aparato comprende una unidad de transformada 120 para la transformación de la segunda porción de señal de audio o un valor o una señal derivado de la segunda porción de señal de audio de un segundo dominio a un dominio de rastreo para obtener una información de segunda porción de señal, en el que el segundo dominio es diferente del primer dominio, en el que el dominio de rastreo es diferente del segundo dominio, y en el que el dominio de rastreo es igual o diferente del primer dominio.

[0277] Adicionalmente, el aparato comprende una unidad de rastreo de nivel de ruido 130, en el que la unidad de rastreo de nivel de ruido está configurada para recibir una información de primera porción de señal que es representada en el dominio de rastreo, en el que la información de primera porción de señal depende de la primera porción de señal de audio, en el que 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 en el que 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 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.

[0278] Además, el aparato comprende una unidad de reconstrucción para la reconstrucción de 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 pero está alterada.

[0279] Con respecto a la primera y/o la segunda porción de señal de audio, por ejemplo, la primera y/o la segunda porción de señal de audio, por ejemplo, pueden alimentarse 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 comprendida por la primera y/o la segunda porción de señal de audio pueda ser reproducida nuevamente.

[0280] 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 posteriores tramas no lleguen al receptor, o en el caso de que las posteriores tramas sean erróneas.

[0281] Entre otras cosas, la presente invención se basa en el hallazgo de que el rastreo de nivel de ruido debe realizarse en un dominio común, referido en el presente documento como "dominio de rastreo". El dominio de rastreo, por ejemplo, puede ser un dominio de excitación, por ejemplo, el dominio en el que la señal es representada por LPC (LPC = coeficiente predictivo lineal) o por ISP (par espectral de impedancia y admitancia) como se describe en AMR-WB y en AMR-WB+ (véase [3GP12a], [3GP12b], [3GP09a], [3GP09b], [3GP09c]). El rastreo del nivel de ruido en el único dominio tiene, entre otras cosas, la ventaja de que se evitan los efectos de la superposición periódica sucesiva 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 ACELP a TCX o viceversa).

[0282] 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 para 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 para obtener el valor derivado).

[0283] Con respecto a la primera porción de señal de audio, en algunas realizaciones, la primera porción de señal de audio puede ser procesada y/o transformada al dominio de rastreo.

[0284] En otras realizaciones, sin embargo, la primera porción de señal de audio puede estar ya representada en el dominio de rastreo.

[0285] En algunas realizaciones, la información de primera porción de señal es idéntica a la primera porción de señal de audio. En otras realizaciones, la información de primera porción de señal es, por ejemplo, un valor agregado que depende de la primera porción de señal de audio.

[0286] Ahora, en primer lugar, se considera en más detalle el desvanecimiento a un nivel de ruido de confort.

[0287] El enfoque de desvanecimiento descrito, por ejemplo, puede ser implementado en una versión de bajo retardo de xHE-AAC [NMR+12] (xHE-AAC = AAC de Alta Eficiencia Extendida), que es capaz de cambiar sin costuras entre la codificación de ACELP (la voz) y de MDCT (música/ruido) sobre una base por trama.

[0288] Con respecto al rastreo de nivel común en un dominio de rastreo, por ejemplo, como 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 cómodo. Por lo tanto, el nivel de ruido de fondo puede ser derivado y actualizado en forma constante durante la codificación normal.

[0289] La presente invención se basa en el hallazgo de que cuando se tiene un códec de núcleo cambiado (por ejemplo, ACELP y TCX), es en particular adecuada la consideración de un nivel de ruido de fondo común, independientemente del codificador de núcleo seleccionado.

[0290] La Fig. 6 representa un rastreo del nivel de ruido de fondo según una realización preferida en el decodificador durante el modo de operación libre de error, por ejemplo, durante la decodificación normal.

[0291] El rastreo en sí mismo, por ejemplo, puede realizarse usando el enfoque de estadística mínima (véase [Mar01]).

[0292] Este nivel rastreado de ruido de fondo, por ejemplo, puede considerarse como la información de nivel de ruido mencionada anteriormente.

[0293] 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.º 5, 504 -512", [Mar01] puede emplearse para el rastreo del nivel de ruido de fondo.

[0294] De manera correspondiente, en algunas realizaciones, la unidad de rastreo de nivel de ruido 130 está configurada para determinar la información de nivel de ruido mediante la aplicación de un enfoque de estadística mínima, por ejemplo, mediante el empleo de la estimación de ruido de estadística mínima de [Mar01].

[0295] A continuación, se describen algunas consideraciones y detalles de este enfoque de rastreo.

[0296] Con respecto al rastreo de nivel, se supone que el fondo es de tipo ruido. Por lo tanto, es preferible realizar el rastreo de nivel en el dominio de excitación para evitar el rastreo de componentes tonales delanteros que son extraídos por LPC. Por ejemplo, el relleno de ruido de 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 una realización preferida, el rastreo se realiza en el dominio de excitación de ACELP.

[0297] La Fig. 7 ilustra la derivación de ganancia de la síntesis de LPC y el desénfasis según una realización.

[0298] Con respecto a la derivación de nivel, por ejemplo, la derivación de nivel puede realizarse 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.

[0299] En la realización preferida, la derivación de nivel para ACELP se realiza en el dominio de excitación. Por lo tanto, no se requiere compensación de ganancia.

[0300] Para TCX, puede ser necesaria, por ejemplo, una compensación de ganancia para ajustar el nivel derivado al dominio de excitación de ACELP.

[0301] En la realización preferida, la derivación de nivel para TCX tiene lugar en el dominio de tiempo. Se halló una compensación de ganancia manipulable para este enfoque: La ganancia introducida por la síntesis de LPC y el desénfasis se deriva como se muestra en la Fig. 7, y el nivel derivado se divide entre esta ganancia.

[0302] Como alternativa, la derivación de nivel para TCX podría realizarse en el dominio de excitación de TCX. Sin embargo, la compensación de ganancia entre el dominio de excitación de TCX y el dominio de excitación de ACELP se consideró demasiado complicada.

5 **[0303]** Por lo tanto, volviendo a la Fig. 1a, en algunas realizaciones, 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 dichas realizaciones, la unidad de rastreo de nivel de ruido 130 está configurada para recibir la información de
10 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.

[0304] En otras realizaciones, la primera porción de señal de audio está representada en un dominio de
15 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 dichas realizaciones, 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
20 130 está configurada para recibir la segunda porción de señal que es representada en el dominio de excitación como el dominio de rastreo.

[0305] En una realización, la primera porción de señal de audio, por ejemplo, puede ser representada en un dominio de excitación como el primer dominio, en la que la unidad de rastreo de nivel de ruido 130, por ejemplo,
25 puede ser configurada para recibir la información de primera porción de señal, en la que dicha información de primera porción de señal está representada en el dominio FFT, que es el dominio de rastreo, y en la que dicha información de primera porción de señal depende de dicha primera porción de señal de audio que está representada en el dominio de excitación, en la que la unidad de transformada 120, 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
30 dominio de tiempo que es el segundo dominio, a un dominio FFT que es el dominio de rastreo, y en la que la unidad de rastreo de nivel de ruido 130, por ejemplo, puede ser configurada para recibir la segunda porción de señal de audio que es representada en el dominio FFT.

[0306] La Fig. 1b ilustra un aparato según otra realización. En la Fig. 1b, la unidad de transformada 120 de la
35 Fig. 1a es una primera unidad de transformada 120, y la unidad de reconstrucción 140 de la Fig. 1a es una primera unidad de reconstrucción 140. El aparato adicionalmente comprende una segunda unidad de transformada 121 y una segunda unidad de reconstrucción 141.

[0307] La segunda unidad de transformada 121 está configurada para transformar la información de nivel de
40 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 está alterada.

[0308] Además, la segunda unidad de reconstrucción 141 está configurada para reconstruir una cuarta
45 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 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 está alterada.

[0309] La Fig. 1c ilustra un aparato para la decodificación de una señal de audio según otra realización. El
50 aparato adicionalmente comprende una primera unidad de agregación 150 para determinar un primer valor agregado que depende la primera porción de señal de audio. Además, el aparato de la Fig. 1c adicionalmente comprende una segunda unidad de agregación 160 para determinar un segundo valor agregado como el valor derivado de la segunda porción de señal de audio que depende de la segunda porción de señal de audio. En la realización de la Fig. 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, en la que la unidad de
55 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 dependiendo del segundo valor agregado que es representado en el dominio de rastreo.

60

[0310] En una realización, la primera unidad de agregación 150 está configurada para determinar el primer valor agregado, de manera que el primer valor agregado indique 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 manera que el segundo valor agregado indique una media cuadrática de la segunda porción de señal de audio o de una señal derivada de la segunda porción de señal de audio.

[0311] La Fig. 6 ilustra un aparato para la decodificación de una señal de audio según una realización adicional.

[0312] En la Fig. 6, la unidad de rastreo de nivel de fondo 630 implementa una unidad de rastreo de nivel de ruido 130 según la Fig. 1 a.

[0313] Además, en la Fig. 6, la unidad de RMS 650 (RMS = media cuadrática) es una primera unidad de agregación, y la unidad de RMS 660 es una segunda unidad de agregación.

[0314] Según algunas realizaciones, la (primera) unidad de transformada 120 de la Fig. 1a, Fig. 1b y Fig. 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 entre un valor de ganancia (x). En otras realizaciones, un valor de ganancia, por ejemplo, puede ser multiplicado.

[0315] En algunas realizaciones, el valor de ganancia (x), por ejemplo, puede indicar una ganancia introducida por la síntesis de codificación predictiva lineal, o el valor de ganancia (x), por ejemplo, puede indicar una ganancia introducida por la síntesis de codificación predictiva lineal y el desénfasis.

[0316] En la Fig. 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 entonces divide el valor proporcionado por la segunda unidad de agregación 660, que es un valor derivado de la segunda porción de señal de audio, entre el valor de ganancia proporcionado (x) (por ejemplo, ya sea mediante la división entre x, ya sea mediante la multiplicación del valor $1/x$). Por lo tanto, la unidad 620 de la Fig. 6 que comprende las unidades 621 y 622 implementa la primera unidad de transformada de la Fig. 1a, Fig. 1b o Fig. 1c.

[0317] El aparato de la Fig. 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 Fig. 6, un dominio de LPC (ACELP). La primera porción de señal de audio es alimentada en una unidad de síntesis de LPC y desénfasis 671 para el procesamiento para obtener 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 RMS 650 para obtener un primer valor que indica una media cuadrática de la primera porción de señal de audio. Este primer valor (primer valor de RMS) está representado en el dominio de rastreo. El primer valor de RMS, que es representado en el dominio de rastreo, es entonces alimentado en la unidad de rastreo de nivel de ruido 630.

[0318] Además, el aparato de la Fig. 6 recibe una segunda trama con una segunda porción de señal de audio que comprende un espectro de MDCT y que es representada en un dominio de MDCT. El relleno de ruido se realiza por medio de un módulo de relleno de ruido 681, la formación de ruido en el dominio de frecuencia se realiza por medio de un módulo de formación de ruido en el dominio de frecuencia 682; la transformación al dominio de tiempo se realiza por un módulo de iMDCT/OLA 683 (OLA = superposición y adición), y la predicción de largo plazo se realiza por medio de una unidad de predicción de largo plazo 684. La unidad de predicción de largo plazo, por ejemplo, puede comprender un búfer de retardo (no mostrado en la Fig. 6).

[0319] La señal derivada de la segunda porción de señal de audio es luego alimentada en el módulo de RMS 660 para obtener un segundo valor que indica una media cuadrática de dicha señal derivada de la segunda porción de señal de audio. Este segundo valor (segundo valor de RMS) es aún representado en el dominio de tiempo. La unidad 620 entonces transforma el segundo valor de media cuadrática (RMS) del dominio de tiempo al dominio de rastreo, aquí, el dominio de LPC (ACELP). El segundo valor de RMS, que es representado en el dominio de rastreo, es entonces alimentado en la unidad de rastreo de nivel de ruido 630.

[0320] En realizaciones, el rastreo de nivel se realiza en el dominio de excitación, si bien el desvanecimiento de TCX se realiza en el dominio de tiempo.

[0321] 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 que la última señal recibida es suavemente desvanecida a modo de nivel.

5 **[0322]** 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 la realización preferida, la aplicación de nivel se realiza en los mismos dominios que la derivación de nivel, conduciendo a los mismos beneficios que para ACELP, sin compensación de ganancia necesaria, y que para TCX, es necesaria la compensación de ganancia inversa como para la derivación de nivel (véase la Fig. 6), y en consecuencia, puede usarse la misma derivación de ganancia,
10 como se ilustra por la Fig. 7.

[0323] A continuación, se describe la compensación de una influencia del filtro paso alto sobre la ganancia de síntesis de LPC según realizaciones.

15 **[0324]** La Fig. 8 explica resumidamente este enfoque. En particular, la Fig. 8 ilustra la aplicación de nivel de ruido de confort durante la pérdida de paquetes.

[0325] En la Fig. 8, la unidad de filtro de ganancia paso alto 643, la unidad de multiplicación 644, la unidad de desvanecimiento 645, la unidad de filtro paso alto 646, la unidad de desvanecimiento 647 y la unidad de combinación 648 forman en conjunto una primera unidad de reconstrucción.
20

[0326] Además, en la Fig. 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 Fig. 6.
25

[0327] Adicionalmente, en la Fig. 8, la unidad de ganancia de síntesis de LPC y desénfasis 649, y la unidad de multiplicación 641 forman juntas una segunda unidad de transformada 640.

[0328] Además, en la Fig. 8, la unidad de desvanecimiento 642 representa una segunda unidad de reconstrucción.
30

[0329] En la realización de la Fig. 8, la excitación con voz y la excitación sin voz son desvanecidas en 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 Fig. 8 adicionalmente representa un filtro paso alto, que es introducido en la cadena de señal de la excitación sin voz con el fin de suprimir los componentes de baja frecuencia para todos los casos, excepto cuando la señal fue clasificada como sin voz.
35

[0330] Para modelar la influencia del filtro paso alto, el nivel después de la síntesis de LPC y el desénfasis se calcula una vez con y una vez sin el filtro paso alto. A continuación, la relación de dichos dos niveles es derivada y usada para alterar el nivel de fondo aplicado.
40

[0331] Esto se ilustra por la Fig. 9. En particular, la Fig. 9 representa la compensación de ganancia paso alto avanzada durante el ocultamiento de ACELP según una realización.

45 **[0332]** 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 RMS en una trama de tiempo más corto. En la práctica, solo se usa una subtrama en lugar de la trama entera.

50 **[0333]** Según una realización, 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 la reconstrucción de 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 está alterada.
55

[0334] Según una realización, 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 la reconstrucción de 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 está alterada.
60

- [0335]** En una realización, 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, en la que dicho espectro de nivel de ruido se obtiene mediante la aplicación del enfoque de estadística mínima. La unidad de reconstrucción 140 está configurada para la reconstrucción de 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 110 o si dicha tercera trama es recibida por la interfaz receptora 110 pero está alterada.
- [0336]** En una realización, la (primera y/o segunda) unidad de reconstrucción 140, 141, por ejemplo, puede ser configurada 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 está alterada.
- [0337]** Según una realización, 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.
- [0338]** La Fig. 14 ilustra un aparato para la decodificación de una señal de audio. El aparato comprende una interfaz receptora 110, en la que 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 en la que 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.
- [0339]** Además, el aparato comprende una unidad de rastreo de nivel de ruido 130, en el que la unidad de rastreo de nivel de ruido 130 está configurada para determinar la información de nivel de ruido que depende al menos de una de la primera porción de señal de audio y la segunda porción de señal de audio (esto significa: que depende de la primera porción de señal de audio y/o la segunda porción de señal de audio), en la que la información de nivel de ruido se representa en un dominio de rastreo.
- [0340]** 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 dependiendo de 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 está alterada, en el que el primer dominio de reconstrucción es diferente o igual del dominio de rastreo.
- [0341]** 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 está alterada, en el que el segundo dominio de reconstrucción es diferente del dominio de rastreo, y en el que el segundo dominio de reconstrucción es diferente del primer dominio de reconstrucción; y
- [0342]** 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 está alterada.
- [0343]** Según algunas realizaciones, el dominio de rastreo, por ejemplo, puede ser en el que el dominio de rastreo es un dominio de tiempo, un dominio espectral, un dominio FFT, un dominio de MDCT, o un dominio de excitación. El primer dominio de reconstrucción, por ejemplo, puede ser el dominio de tiempo, el dominio espectral, el dominio FFT, el dominio de MDCT, o el dominio de excitación. El segundo dominio de reconstrucción, por ejemplo, puede ser el dominio de tiempo, el dominio espectral, el dominio FFT, el dominio de MDCT, o el dominio de excitación.
- [0344]** En una realización, el dominio de rastreo, por ejemplo, puede ser el dominio FFT, el primer dominio de reconstrucción, por ejemplo, puede ser el dominio de tiempo, y el segundo dominio de reconstrucción, por ejemplo, puede ser el dominio de excitación.
- [0345]** En otra realización, el dominio de rastreo, por ejemplo, puede ser el dominio de tiempo, el primer

dominio de reconstrucción, por ejemplo, puede ser el dominio de tiempo, y el segundo dominio de reconstrucción, por ejemplo, puede ser el dominio de excitación.

[0346] Según una realización, dicha primera porción de señal de audio, por ejemplo, puede ser representada en un primer dominio de entrada, y dicha segunda porción de señal de audio, por ejemplo, puede ser representada 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 derivado de la segunda porción de señal de audio, del segundo dominio de entrada al dominio de rastreo, para obtener una información de segunda porción de señal. La unidad de rastreo de nivel de ruido, por ejemplo, puede ser configurada para recibir una información de primera porción de señal que es representada en el dominio de rastreo, en la que la información de primera porción de señal depende de la primera porción de señal de audio, en la que 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 en la que 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 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.

[0347] Según una realización, el primer dominio de entrada, por ejemplo, puede ser el dominio de excitación, y el segundo dominio de entrada, por ejemplo, puede ser el dominio de MDCT.

[0348] En otra realización, el primer dominio de entrada, por ejemplo, puede ser el dominio de MDCT, y donde el segundo dominio de entrada, por ejemplo, puede ser el dominio de MDCT.

[0349] Si, por ejemplo, una señal está representada en un dominio de tiempo, por ejemplo, puede ser representada por muestras de dominio de tiempo de la señal. O, por ejemplo, si una señal está representada en un dominio espectral, por ejemplo, puede ser representada por muestras espectrales de un espectro de la señal.

[0350] En una realización, el dominio de rastreo, por ejemplo, puede ser el dominio FFT, el primer dominio de reconstrucción, por ejemplo, puede ser el dominio de tiempo, y el segundo dominio de reconstrucción, por ejemplo, puede ser el dominio de excitación.

[0351] En otra realización, el dominio de rastreo, por ejemplo, puede ser el dominio de tiempo, el primer dominio de reconstrucción, por ejemplo, puede ser el dominio de tiempo, y el segundo dominio de reconstrucción, por ejemplo, puede ser el dominio de excitación.

[0352] En algunas realizaciones, las unidades ilustradas en la Fig. 14, por ejemplo, pueden estar configuradas como se describe para las Figs. 1a, 1b, 1c y 1d.

[0353] Con respecto a realizaciones particulares, en, por ejemplo, un modo de tasa baja, un aparato según una realización, por ejemplo, puede recibir tramas de ACELP como una entrada, que son representadas en un dominio de excitación, y que son luego transformadas en un dominio de tiempo por medio de la síntesis de LPC. Además, en el modo de tasa baja, el aparato según una realización, por ejemplo, puede recibir tramas de TCX como una entrada, que son representadas en un dominio de MDCT, y que son luego transformadas en un dominio de tiempo por medio de una MDCT inversa.

[0354] Luego se realiza el rastreo en un dominio de FFT, en el que la señal de FFT es derivada de la señal de dominio de tiempo mediante la conducción de una FFT (transformada de Fourier rápida). El rastreo, por ejemplo, puede ser realizado por la conducción de un enfoque de estadística mínima, separado para todas las líneas espectrales para obtener un espectro de ruido de confort.

[0355] A continuación se realiza el ocultamiento mediante la conducción de la derivación de nivel basada en el espectro de ruido de confort. La derivación de nivel se realiza basándose en el espectro de ruido de confort. La conversión de nivel en el dominio de tiempo se realiza para el PLC de TCX de FD. Se realiza un desvanecimiento en el dominio de tiempo. Se realiza una derivación de nivel en el dominio de excitación para el PLC de ACELP y para el PLC de TCX de TD (de tipo ACELP). Después, se realiza un desvanecimiento en el dominio de excitación.

[0356] El siguiente listado resume esto: tasa baja:

low rate:
• input:

- + acelp (excitation domain -> time domain, via lpc synthesis)
- + tcx (mdct domain -> time domain, via inverse MDCT)

- tracing:

- 5 + fft-domain, derived from time domain via FFT
- + minimum statistics, separate for all spectral lines
- > comfort noise spectrum

- concealment:

- 10 + level derivation based on the comfort noise spectrum
- + level conversion into time domain for
 - FD TCX PLC
 -> fading in the time domain
- + level conversion into excitation domain for
 - 15 ■ ACELP PLC
 - TD TCX PLC (ACELP like)
 -> fading in the excitation domain

[0357] En, por ejemplo, un modo de tasa alta, pueden recibirse tramas de TCX como una entrada, que son representadas en el dominio de MDCT, y que son luego transformadas al dominio de tiempo mediante una MDCT inversa.

[0358] El rastreo entonces puede realizarse en el dominio de tiempo. Por ejemplo, el rastreo puede realizarse mediante la conducción de un enfoque de estadística mínima basado en el nivel de energía para obtener un nivel de ruido de confort.

[0359] Para el ocultamiento, para el PLC de TCX de FD, el nivel puede usarse como tal, y puede realizarse solamente un desvanecimiento en el dominio de tiempo. Para el PLC de TCX de TD, (de tipo ACELP), se realiza la conversión de nivel en el dominio de excitación y el desvanecimiento en el dominio de excitación.

[0360] El siguiente listado resume esto:

tasa alta:

- entrada:

- 35 ○ tcx (dominio de MDCT -> dominio de tiempo, mediante MDCT inversa)

- rastreo:

- dominio de tiempo
- estadística mínima sobre el nivel de energía -> nivel de ruido de confort

- ocultamiento:

- 40 ○ uso de nivel "como tal"
 - PLC de TCX de FD
 - > desvanecimiento en el dominio de tiempo
- conversión de nivel en el dominio de excitación para
 - PLC de TCX de TD (de tipo ACELP)
 - 45 -> desvanecimiento en el dominio de excitación

[0361] El dominio FFT y el dominio de la MDCT son ambos dominios espectrales, mientras que el dominio de excitación es cierta clase de dominio de tiempo.

[0362] Según una realización, la primera unidad de reconstrucción 140, por ejemplo, puede ser configurada 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, por ejemplo, puede ser configurada 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 LTP. Además, la primera unidad de reconstrucción 140 y la segunda unidad de reconstrucción 141, por ejemplo, pueden ser configuradas para conducir el primer desvanecimiento y el segundo desvanecimiento a un espectro de tipo ruido y/o un segundo desvanecimiento de una ganancia de LTP con la misma velocidad de desvanecimiento.

[0363] Ahora, se considera la forma espectral adaptativa de ruido de confort.

[0364] A fin de lograr la formación adaptativa al ruido de confort durante la pérdida de paquetes de ráfaga, como una primera etapa, puede realizarse el hallazgo de coeficientes apropiados de LPC que representan el ruido de fondo. Estos coeficientes de LPC pueden derivarse durante la voz activa usando un enfoque de estadística mínima para hallar el espectro de ruido de fondo, y luego, el cálculo de los coeficientes de LPC a partir de este, usando un algoritmo arbitrario para la derivación de LPC conocido de la bibliografía. Algunas realizaciones, por ejemplo, pueden convertir directamente el espectro de ruido de fondo en una representación que puede usarse directamente para FDNS en el dominio de la MDCT.

[0365] El desvanecimiento a ruido de confort puede realizarse en el dominio de ISF (también aplicable en el dominio de LSF; LSF: Frecuencia Espectral Lineal):

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

ajustando pt_{media} en coeficientes de LP apropiados que describen el ruido de confort.

15

[0366] Con respecto a la forma espectral adaptativa anteriormente descrita del ruido de confort, se ilustra una realización más general por la Fig. 11.

[0367] La Fig. 11 ilustra un aparato para la decodificación de una señal de audio codificada para obtener una señal de audio reconstruida según una realización.

[0368] 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.

[0369] 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 recibida por la interfaz receptora 1110 no pero está alterada/errónea, uno o más primeros coeficientes de señal de audio, que están comprendidos por la trama actual, en el que dichos uno o más primeros coeficientes de señal de audio indican una característica de la señal de audio codificada, y uno o más coeficientes de ruido que indican 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, que dependen de los uno o más primeros coeficientes de señal de audio y que dependen 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 pero está alterada/errónea.

[0370] El reconstructor de señal de audio 1130 está configurado para la reconstrucción de una primera porción de la señal de audio reconstruida que depende 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 la reconstrucción de una segunda porción de la señal de audio reconstruida que depende 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.

[0371] La determinación de un ruido de fondo es muy conocida en la técnica (ver, 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), no. 5, 504 -512), y en una realización, el aparato procede consecuentemente.

[0372] En algunas realizaciones, el uno o más primeros coeficientes de señal de audio, por ejemplo, pueden ser uno o más coeficientes de filtro predictivo lineal de la señal de audio codificada. En algunas realizaciones, el uno o más primeros coeficientes de señal de audio, por ejemplo, pueden ser uno o más coeficientes de filtro predictivo lineal de la señal de audio codificada.

[0373] Se conoce bien en la técnica la manera de reconstruir una señal de audio, por ejemplo, una señal de la 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 (AMR-WB) speech codec; transcoding functions*, 3GPP TS 26.190, 3rd Generation Partnership Project, 2009), y en una realización, el reconstructor de señal procede consecuentemente.

[0374] Según una realización, el 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 una

60

realización, el uno o más coeficientes de filtro predictivo lineal, por ejemplo, pueden representar una forma espectral del ruido de fondo.

[0375] En una realización, el generador de coeficientes 1120, por ejemplo, puede estar configurado para determinar una o más segundas porciones de señal de audio de 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 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.

10 **[0376]** Según una realización, el generador de coeficientes 1120, por ejemplo, puede estar configurado para generar el uno o más segundos coeficientes de señal de audio mediante la aplicación de la fórmula:

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

15 en la que $f_{actual}[i]$ indica uno de los uno o más segundos coeficientes de señal de audio, en la que $f_{último}[i]$ indica uno de los uno o más primeros coeficientes de señal de audio, en la que $pt_{media}[i]$ es uno de uno o más coeficientes de ruido, en la que α es un número real con $0 \geq \alpha \geq 1$, y en la que i es un índice.

[0377] Según una realización, $f_{último}[i]$ indica un coeficiente de filtro predictivo lineal de la señal de audio codificada, y en la que $f_{actual}[i]$ indica un coeficiente de filtro predictivo lineal de la señal de audio reconstruida.

[0378] En una realización, $pt_{media}[i]$ por ejemplo, puede ser un coeficiente de filtro predictivo lineal que indica el ruido de fondo de la señal de audio codificada.

25 **[0379]** Según una realización, el generador de coeficientes 1120, por ejemplo, puede ser configurado para generar coeficientes de señal de audio de al menos 10 segundos como el uno o más segundos coeficientes de señal de audio.

[0380] En una realización, el generador de coeficientes 1120, por ejemplo, puede estar configurado 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, el uno o más coeficientes de ruido mediante la determinación de un espectro de ruido de la señal de audio codificada.

[0381] A continuación, se considera el desvanecimiento del espectro de MDCT a ruido blanco antes de la aplicación de FDNS.

[0382] En lugar de modificar aleatoriamente el signo de un intervalo de MDCT (mezcla de signos), se llena el espectro completo con ruido blanco, que se le da forma usando FDNS. Para 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:

```

for (i=0; i<L_frame; i++) {
    if (old_x[i] != 0)
        x[i] = (1 - cum_damping)*noise[i] + cum_damping *
45 random_sign() * x_old[i];
    }
}

```

donde:

50 cum_damping es el factor de atenuación (absoluto) - disminuye de trama a trama, empezando de 1 y disminuyendo hacia 0

x_old es el espectro de la última trama recibida

random_sign devuelve 1 o -1

55 ruido contiene un vector aleatorio (ruido blanco) que se somete a escala de manera tal que su media cuadrática (RMS) sea similar al último espectro bueno.

[0383] El término random_sign () *old_x[i] caracteriza el proceso de mezcla de signos con el fin de aleatorizar las fases y de ese modo evitar repeticiones de armónicos.

60

[0384] Posteriormente, podría efectuarse otra normalización del nivel de energía después de la mezcla gradual, con el fin de garantizar que la energía de la operación de suma no se desvíe debido a la correlación de los dos vectores.

5 **[0385]** Según realizaciones, la primera unidad de reconstrucción 140, por ejemplo, puede ser configurada para reconstruir la tercera porción de señal de audio dependiendo de la información de nivel de ruido y que depende de la primera porción de señal de audio. En una realización particular, la primera unidad de reconstrucción 140, por ejemplo, puede ser configurada para reconstruir la tercera porción de señal de audio mediante la atenuación o amplificación de la primera porción de señal de audio.

10 **[0386]** En algunas realizaciones, la segunda unidad de reconstrucción 141, por ejemplo, puede ser configurada para reconstruir la cuarta porción de señal de audio dependiendo de la información de nivel de ruido y que depende de la segunda porción de señal de audio. En una realización particular, la segunda unidad de reconstrucción 141, por ejemplo, puede ser configurada para reconstruir la cuarta porción de señal de audio
15 mediante la atenuación o amplificación de la segunda porción de señal de audio.

[0387] Con respecto al desvanecimiento anteriormente descrito del espectro de MDCT a ruido blanco antes de la aplicación de FDNS, una realización más general es ilustrada por la Fig. 12.

20 **[0388]** La Fig. 12 ilustra un aparato para la decodificación de una señal de audio codificada para obtener una señal de audio reconstruida según una realización.

[0389] 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
25 la señal de audio codificada, y un procesador 1220 para la generación de la señal de audio reconstruida.

[0390] 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 está alterada, en el que el
30 espectro modificado comprende una pluralidad de muestras de señales modificadas, en el que, 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.

[0391] 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
35

[0392] Según una realización, el espectro objetivo es un espectro de tipo ruido.

40 **[0393]** En una realización, el espectro de tipo ruido representa ruido blanco.

[0394] Según una realización, el espectro de tipo ruido es con forma.

[0395] En una realización, la forma del espectro de tipo ruido depende de un espectro de señales de audio de
45 una señal previamente recibida.

[0396] Según una realización, el espectro de tipo ruido es con forma dependiendo de la forma del espectro de señales de audio.

50 **[0397]** En una realización, el procesador 1220 emplea un factor de inclinación para dar forma al espectro de tipo ruido.

[0398] Según una realización, el procesador 1220 emplea la fórmula

55
$$\text{shaped_noise}[i] = \text{ruido} \times \text{potencia}(\text{tilt_factor}, i/N)$$

en la que N indica el número de muestras,

en la que i es un índice,

en la que $0 \leq i < N$, con $\text{tilt_factor} > 0$,

60 en la que potencia es una función de potencia.

[0399] Si el tilt_factor es menor de 1, esto significa la atenuación con i creciente. Si el tilt_factor es mayor que 1, esto significa la amplificación con i creciente.

5 **[0400]** Según otra realización, el procesador 1220 puede emplear la fórmula

$$\text{shaped_noise}[i] = \text{ruido} \times (1 + i/(N-1) * (\text{tilt_factor} - 1))$$

en la que N indica el número de muestras,

10 en la que i es un índice, en la que $0 \leq i < N$, con tilt_factor > 0.

[0401] Según una realización, 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, 15 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.

[0402] En una realización, 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. 20

[0403] Según una realización, las muestras de señal de audio del espectro de señales de audio están representadas en un dominio de transformada de coseno discreta modificada.

[0404] En otra realización, las muestras de señal de audio del espectro de señales de audio están representadas en un dominio de transformada de seno discreta modificada. 25

[0405] Según una realización, el procesador 1220 está configurado para generar el espectro modificado mediante el empleo de una función de signos aleatorios que, en forma aleatoria, o pseudoaleatoria, da salida o bien a un primer o un segundo valor. 30

[0406] En una realización, el procesador 1220 está configurado para desvanecer el espectro modificado al espectro objetivo mediante la disminución posterior de un factor de atenuación.

[0407] Según una realización, el procesador 1220 está configurado para desvanecer el espectro modificado al espectro objetivo mediante el aumento posterior de un factor de atenuación. 35

[0408] En una realización, 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: 40

$$x[i] = (1 - \text{cum_damping}) * \text{noise}[i] + \text{cum_damping} * \text{random_sign}() * x_old[i]$$

en la que i es un índice, en la que x[i] indica una muestra de la señal de audio reconstruida, en la que cum_damping es un factor de atenuación, en la que x_old[i] indica una de las muestras de señal de audio del espectro de señales 45 de audio de la señal de audio codificada, en la que random_sign() devuelve a 1 o -1, y en la que el ruido es un vector aleatorio que indica el espectro objetivo.

[0409] Algunas realizaciones continúan una operación de LTP de TCX. En estas realizaciones, la operación de LTP de TCX es continuada durante el ocultamiento con los parámetros de la LTP (retraso de LTP y ganancia de 50 LTP) derivados de la última buena trama.

[0410] Las operaciones de LTP pueden resumirse como:

- Alimentar el búfer de retardo de LTP basándose en la salida previamente derivada.
- 55 - Basándose en el retraso de LTP: seleccionar la porción de señal apropiada, del búfer de retardo de LTP que se usa como contribución de LTP para dar forma a la señal actual.
- Reescalar esta contribución de LTP usando la ganancia de LTP.
- Sumar esta contribución reescalada de LTP a la entrada de señal de LTP con el fin de generar la señal de salida de

LTP.

[0411] Podrían considerarse diferentes enfoques con respecto al tiempo, cuando se realiza la actualización del búfer de retardo de LTP:

5

Como la primera operación de LTP en trama n usando la salida de la última trama n-1. Esto actualiza el búfer de retardo de LTP en trama n que va a usarse durante el procesamiento de LTP en trama n.

Como la última operación de LTP en trama n usando la salida de la trama actual n. Esto actualiza el búfer de retardo de LTP en trama n que va a usarse durante el procesamiento de LTP en trama n+1.

10

[0412] A continuación, se considera el desacoplado del circuito de retroalimentación de la LTP de TCX.

[0413] El desacoplado del circuito de retroalimentación de la LTP de 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 LTP cuando se encuentra en modo de ocultamiento.

15

[0414] La Fig. 10 ilustra este desacoplado. En particular, la Fig. 10 representa el desacoplado del circuito de retroalimentación de la LTP durante el ocultamiento ($bfi=1$).

20

[0415] La Fig. 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).

[0416] Hacia el tiempo, cuando se realiza la actualización del búfer de retardo de la LTP 1020, algunas realizaciones proceden de la siguiente manera:

25

- Para la operación normal: Para actualizar el búfer de retardo de LTP 1020, podría preferirse como la primera operación de LTP, ya que la señal de salida sumada habitualmente es almacenada en forma persistente. Con este enfoque, puede omitirse un búfer dedicado.

- Para la operación desacoplada: Para actualizar el búfer de retardo de LTP, podría preferirse como la última operación de LTP, ya que la contribución de LTP a la señal habitualmente solo es almacenada en forma temporaria. Con este enfoque, se preserva la señal de contribución de LTP transitoriamente. La implementación de esta manera de este búfer de contribución de LTP podría hacerse persistente.

30

[0417] Asumiendo que se usa este último enfoque en cualquier caso (operación normal y ocultamiento), las realizaciones, por ejemplo, pueden implementar lo siguiente:

35

- Durante la operación normal: Se usa la salida de señal de dominio de tiempo del decodificador de LTP después de su adición a la señal de entrada de la LTP para alimentar el búfer de retardo de LTP.

- Durante el ocultamiento: Se usa la salida de señal de dominio de tiempo del decodificador de LTP antes de su adición a la señal de entrada de LTP para alimentar el búfer de retardo de LTP.

40

[0418] Algunas realizaciones desvanecen la ganancia de la LTP de TCX hacia cero. En dicha realización, la ganancia de la LTP de TCX, por ejemplo, puede ser desvanecida hacia cero con un cierto factor de desvanecimiento adaptativo de señal. Esto, por ejemplo, puede ser realizado iterativamente, por ejemplo, según el siguiente pseudocódigo:

45

ganancia = gain_past * damping;

[...]

gain_past = ganancia;

50

donde:

ganancia es la ganancia del decodificador de LTP de TCX aplicada en la trama actual;

gain_past es la ganancia del decodificador de LTP de TCX aplicada en la trama previa;

55

amortiguamiento es el factor de desvanecimiento (relativo).

[0419] La Fig. 1d ilustra un aparato según una realización adicional, en el que el aparato adicionalmente comprende una unidad de predicción de largo plazo 170 que comprende un búfer de retardo 180. La unidad de predicción de largo plazo 170 está configurada para generar una señal procesada que depende de la segunda porción de señal de audio, que depende de una entrada de búfer de retardo que está almacenada en el búfer de

60

retardo 180 y que depende de una ganancia de predicción de largo plazo. Además, la unidad de predicción de largo plazo está configurada para desvanecer la ganancia de predicción de 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 está alterada.

5

[0420] En otras realizaciones (no mostradas), la unidad de predicción de largo plazo, por ejemplo, puede ser configurada para generar una señal procesada que depende de la primera porción de señal de audio, que depende de una entrada de búfer de retardo que está almacenada en el búfer de retardo y que depende de una ganancia de predicción de largo plazo.

10

[0421] En la Fig. 1d, la primera unidad de reconstrucción 140, por ejemplo, puede generar la tercera porción de señal de audio adicionalmente que depende de la señal procesada.

[0422] En una realización, la unidad de predicción de largo plazo 170, por ejemplo, puede ser configurada para desvanecer la ganancia de predicción de largo plazo hacia cero, en la que una velocidad con la que la ganancia de predicción de largo plazo es desvanecida a cero depende de un factor de desvanecimiento.

[0423] Como alternativa o adicional, la unidad de predicción de largo plazo 170, por ejemplo, puede ser configurada 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 está alterada.

[0424] Con respecto al uso anteriormente descrito de la LTP de TCX, una realización más general se ilustra por la Fig. 13.

25

[0425] La Fig. 13 ilustra un aparato para la decodificación de una señal de audio codificada para obtener una señal de audio reconstruida.

[0426] 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 para obtener muestras de la señal de audio reconstruida de la señal de audio reconstruida.

35

[0427] 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, la pluralidad de muestras de señal de audio seleccionadas de las muestras de señal de audio almacenadas en el búfer de retardo 1320 que depende 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 que depende de una información de retraso de altura compuesta por otra trama recibida previamente por la interfaz receptora 1310.

[0428] Según una realización, el procesador de muestras 1340, por ejemplo, puede ser configurado 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 la reescala 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, por ejemplo, puede ser 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, mediante la reescala de las muestras de señal de audio seleccionadas dependiendo de la información de ganancia comprendida por dicha otra trama recibida previamente por la interfaz receptora 1310.

[0429] En una realización, el procesador de muestras 1340, por ejemplo, puede ser configurado 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

60

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.

5 **[0430]** Según una realización, el procesador de muestras 1340, por ejemplo, puede ser configurado para almacenar las muestras de señal de audio reconstruida en el búfer de retardo 1320.

10 **[0431]** En una realización, el procesador de muestras 1340, por ejemplo, puede ser configurado 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.

10 **[0432]** Según una realización, el procesador de muestras 1340, por ejemplo, puede ser configurado 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.

15 **[0433]** En una realización, el procesador de muestras 1340, por ejemplo, puede ser configurado para reescalar las muestras de señal de audio seleccionadas dependiendo de la información de ganancia para 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.

20 **[0434]** Según una realización, el procesador de muestras 1340, por ejemplo, puede ser configurado 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
25 de muestras 1340 está configurado para almacenar las muestras de señal de audio 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.

30 **[0435]** Según otra realización, el procesador de muestras 1340, por ejemplo, puede ser configurado 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.

35 **[0436]** En una realización, el selector de muestras 1330, por ejemplo, puede ser configurado para obtener las muestras de señal de audio reconstruida mediante la reescala de las muestras de señal de audio seleccionadas dependiendo de una ganancia modificada, en la que la ganancia modificada se define según la fórmula:

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

40 en la que ganancia es la ganancia modificada, en la que el selector de muestras 1330, por ejemplo, puede ser configurado para establecer `past_gain` en la ganancia después de la ganancia, y se ha calculado, y donde el amortiguamiento es un número real.

45 **[0437]** Según una realización, el selector de muestras 1330, por ejemplo, puede ser configurado para calcular la ganancia modificada.

[0438] En una realización, el amortiguamiento, por ejemplo, puede ser definido según: $0 < \text{amortiguamiento} < 1$.

50 **[0439]** Según una realización, la ganancia de ganancia modificada, por ejemplo, puede ser ajustada a 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.

[0440] A continuación, se considera la velocidad de desvanecimiento. Existen varios módulos de
55 ocultamiento que se aplican a una cierta clase de desvanecimiento. Si bien 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 (ACELP o TCX). Por ejemplo:

60 Para 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).

- 5 **[0441]** Además, para TCX, debe usarse la misma velocidad de desvanecimiento, en particular, para la señal de dominio de tiempo, y/o para la ganancia de la LTP (disminuida a cero), y/o para la ponderación de 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.
- 10 **[0442]** Adicionalmente, podría ser preferible además usar la misma velocidad de desvanecimiento para ACELP y TCX, si bien, debido a la diferente naturaleza de los núcleos, podría también seleccionarse el uso de diferentes velocidades de desvanecimiento.
- 15 **[0443]** Esta velocidad de desvanecimiento podría ser estática, pero preferentemente es adaptativa a las características de la señal. Por ejemplo, la velocidad de desvanecimiento, por ejemplo, puede depender del factor de estabilidad de LPC (TCX) y/o de una clasificación, y/o de un número de tramas consecutivamente perdidas.
- 20 **[0444]** La velocidad de desvanecimiento, por ejemplo, puede ser determinada dependiendo del factor de atenuación, que podría ser proporcionado en forma absoluta o relativa, y que podría además cambiar en función del tiempo durante un cierto desvanecimiento.
- [0445]** En realizaciones, se usa la misma velocidad de desvanecimiento para el desvanecimiento de la ganancia de LTP que para el desvanecimiento de ruido blanco.
- [0446]** Se han proporcionado un aparato, método y programa informático para la generación de una señal de ruido de confort como se describe anteriormente.
- 25 **[0447]** Si bien algunos aspectos se han descrito en el contexto de un aparato, es claro que estos aspectos además representan una descripción del método correspondiente, donde un bloque o dispositivo se corresponde con una etapa de método o una característica de una etapa de método. Análogamente, los aspectos descritos en el contexto de una etapa de método además representan una descripción de un bloque o artículo correspondiente o característica de un aparato correspondiente.
- 30 **[0448]** La señal descompuesta inventiva 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 alámbrico tal como Internet.
- 35 **[0449]** Dependiendo de ciertos requisitos de implementación, las realizaciones de la invención pueden ser implementadas en hardware o software. La implementación puede realizarse usando un medio de almacenamiento digital, por ejemplo, un disquete, un DVD, un CD, una ROM, una PROM, una EPROM, una EEPROM o una memoria FLASH, que tiene señales de control electrónicamente legibles allí almacenadas, que cooperan (o son capaces de cooperar) con un sistema de ordenador programable, de manera que se lleve a cabo el método respectivo.
- 40 **[0450]** Algunas realizaciones según la invención comprenden un portador de datos no transitorio que tiene señales de control electrónicamente legibles, que son capaces de cooperar con un sistema de ordenador programable, de manera que se lleve a cabo uno de los métodos que se describen en el presente documento.
- 45 **[0451]** 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 realizar uno de los métodos cuando el producto de programa informático se ejecuta en un ordenador. El código de programa puede ser almacenado, por ejemplo, en un portador legible en una máquina.
- 50 **[0452]** Otras realizaciones comprenden el programa informático para realizar uno de los métodos descritos en el presente documento, almacenado en un portador legible en una máquina.
- [0453]** En otras palabras, una realización del método inventivo, por lo tanto, es un programa informático que tiene un código de programa para la realización de uno de los métodos descritos en el presente documento, cuando el programa informático se ejecuta en un ordenador.
- 55 **[0454]** Una realización adicional de los métodos inventivos es, por lo tanto, un portador de datos (o un medio de almacenamiento digital, o un medio legible por computadora) que comprende, grabado en el mismo, el programa informático para realizar uno de los métodos descritos en el presente documento.
- 60

[0455] Una realización adicional del método inventivo es, por lo tanto, una corriente de datos o una secuencia de señales que representa el programa informático para realizar uno de los métodos descritos en el presente documento. El flujo de datos o la secuencia de señales, por ejemplo, puede estar configurado para transferirse por medio de una conexión de comunicación de datos, por ejemplo, por medio de Internet.

5

[0456] Una realización adicional comprende un medio de procesamiento, por ejemplo, un ordenador, o un dispositivo lógico programable, configurado o adaptado para realizar uno de los métodos descritos en el presente documento.

10 **[0457]** Una realización adicional comprende un ordenador que tiene en el mismo instalado el programa informático para realizar uno de los métodos descritos en el presente documento.

[0458] 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 métodos descritos en el presente documento. En algunas realizaciones, la matriz de puertas programables de campo puede cooperar con un microprocesador con el fin de realizar uno de los métodos descritos en el presente documento. En general, los métodos se realizan, preferentemente, por cualquier aparato de hardware.

15

[0459] Las realizaciones anteriormente descritas son simplemente ilustrativas de los principios de la presente invención. Se entiende que las modificaciones y variaciones de las disposiciones y los detalles descritos en el presente documento serán evidentes para otros expertos en la técnica. Por lo tanto, se tiene la intención de limitarse solo por el alcance de las reivindicaciones de patente inminentes, y no por los detalles específicos presentados a modo de descripción y explicación de las realizaciones en el presente documento.

20

25 **Bibliografía**

[0460]

[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.

30

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

35 [3GP09c] Speech codec speech processing functions; adaptive multi-rate - wideband (AMRWB) speech codec; transcoding functions, 3GPP TS 26.190, 3rd Generation Partnership Project, 2009.

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

40

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

45 [3GP12d] ANSI-C code for the floating-point adaptive multi-rate (AMR) speech codec (release11), 3GPP TS 26.104, 3rd Generation Partnership Project, Sep 2012.

[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, Sep 2012.

50

[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, Sep 2012.

55

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

60 [BP06] A. Borowicz and A. Petrovsky, Minima controlled noise estimation for klt-based speech enhancement, CD-

ROM, 2006, Italy, Florence.

[Coh03] I. Cohen, Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging, IEEE Trans. Speech Audio Process. 11 (2003), no. 5, 466-475.

5

[CPK08] Choong Sang Cho, Nam In Park, and 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).

10

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

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

15

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

[EH08] Jan S. Erkelens and Richards Heusdens, Tracking of Nonstationary Noise Based on Data-Driven Recursive Noise Power Estimation, Audio, Speech, and Language Processing, IEEE Transactions on 16 (2008), no. 6, 1112-1123.

20

[EM84] Y. Ephraim and D. Malah, Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator, IEEE Trans. Acoustics, Speech and Signal Processing 32 (1984), no. 6, 1109-1121.

25

[EM85] Speech enhancement using a minimum mean-square error log-spectral amplitude estimator, IEEE Trans. Acoustics, Speech and Signal Processing 33 (1985), 443-445.

[Gan05] S. Gannot, Speech enhancement: Application of the kalman filter in the estimate-maximize (em framework), Springer, 2005.

30

[HE95] H. G. Hirsch and C. Ehrlicher, Noise estimation techniques for robust speech recognition, Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing, no. pp. 153-156, IEEE, 1995.

[HHJ10] Richard C. Hendriks, Richard Heusdens, and Jesper Jensen, MMSE based noise PSD tracking with low complexity, Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on, Mar 2010, pp. 4266 -4269.

35

[HJH08] Richard C. Hendriks, Jesper Jensen, and Richard Heusdens, Noise tracking using dft domain subspace decompositions, IEEE Trans. Audio, Speech, Lang. Process. 16 (2008), no. 3, 541-553.

40

[IET12] IETF, Definition of the Opus Audio Codec, Tech. Report RFC 6716, Internet Engineering Task Force, Sep 2012.

[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, 2009.

45

[ITU03] ITU-T, Wideband coding of speech at around 16 kbitts using adaptive multi-rate wide band (amr-wb), Recommendation ITU-T G.722.2, Telecommunication Standardization Sector of ITU, Jul 2003.

50

[ITU05] Low-complexity coding at 24 and 32 kbitts for hands-free operation in systems with low frame loss, Recommendation ITU-T G.722.1, Telecommunication Standardization Sector of ITU, May 2005.

[ITU06a] G.722 Appendix III: A high-complexity algorithm for packet loss concealment for G.722, ITU-T Recommendation, ITU-T, Nov 2006.

55

[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, May 2006.

60

- [ITU07] G.722 Appendix IV: A low-complexity algorithm for packet loss concealment with G. 722, ITU-T Recommendation, ITU-T, Aug 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, Jun 2008.
- [ITU08b] G.719: Low-complexity, full-band audio coding for high-quality, conversational applications, Recommendation ITU-T G.719, Telecommunication Standardization Sector of ITU, Jun 2008.
- 10 [ITU 12] 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, June 2012.
- [LS01] Pierre Lauber and Ralph Sperschneider, Error concealment for compressed digital audio, Audio Engineering Society Convention 111, no. 5460, Sep 2001.
- 15 [Mar01] Rainer Martin, Noise power spectral density estimation based on optimal smoothing and minimum statistics, IEEE Transactions on Speech and Audio Processing 9 (2001), no. 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, Sep 2003.
- 20 [MC99] R. Martin and R. Cox, New speech enhancement techniques for low bit rate speech coding, in Proc. IEEE Workshop on Speech Coding (1999), 165-167.
- 25 [MCA99] D. Malah, R. V. Cox, and 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.
- [MEP01] Nikolaus Meine, Bernd Edler, and Heiko Purnhagen, Error protection and concealment for HILN MPEG-4 parametric audio coding, Audio Engineering Society Convention 110, no. 5300, May 2001.
- 30 [MPC89] Y. Mahieux, J.-P. Petit, and 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, pp. 2021-2024 vol.3.
- 35 [NMR+12] Max Neuendorf, Markus Multus, Nikolaus Rettelbach, Guillaume Fuchs, Julien Robilliard, Jérémie Lecomte, Stephan Wilde, Stefan Bayer, Sascha Disch, Christian Helmrich, Roch Lefebvre, Philippe Gournay, Bruno Bessette, Jimmy Lapiere, 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, and 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, April 2012, Presented at the 132nd Convention Budapest, Hungary.
- 40 [PKJ+11] Nam In Park, Hong Kook Kim, Min A Jung, Seong Ro Lee, and 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.
- 45 [QD03] Schuyler Quackenbush and Peter F. Driessen, Error mitigation in MPEG-4 audio packet communication systems, Audio Engineering Society Convention 115, no. 5981, Oct 2003.
- 50 [RL06] S. Rangachari and P. C. Loizou, A noise-estimation algorithm for highly non-stationary environments, Speech Commun. 48 (2006), 220-231.
- [SFB00] V. Stahl, A. Fischer, and 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.
- [SS98] J. Sohn and W. Sung, A voice activity detector employing soft decision based noise spectrum adaptation, Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing, no. pp. 365-368, IEEE, 1998.
- 60 [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 on, Apr 2009, pp. 4421-4424.

REIVINDICACIONES

1. Un aparato para la decodificación de una señal de audio codificada para obtener una señal de audio reconstruida, en el que el aparato comprende:

5 una interfaz receptora (1110) para recibir una o más tramas, un generador de coeficientes (1120), y un reconstructor de señal (1130), en el que el generador de coeficientes (1120) está configurado para determinar, si una trama actual de una o más
10 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 primeros coeficientes de señal de audio, que están comprendidos por la trama actual, en el que dichos uno o más primeros coeficientes de señal de audio indican una característica de la señal de audio codificada, y uno o más coeficientes de ruido que indican una forma espectral de un ruido de fondo de la señal de audio codificada,
15 en el que el generador de coeficientes (1120) está configurado para generar uno o más segundos coeficientes de señal de audio, que dependen de los uno o más primeros coeficientes de señal de audio y que dependen 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) está alterada, en el que el reconstructor de señal de audio (1130) está configurado para la reconstrucción de una primera porción de la señal de audio reconstruida que depende de los uno o más primeros coeficientes de señal de audio, si la trama
20 actual es recibida por la interfaz receptora (1110) y si la trama actual recibida por la interfaz receptora (1110) no está alterada, y en el que el reconstructor de señal de audio (1130) está configurado para la reconstrucción de una segunda porción de la señal de audio reconstruida que depende de uno o más segundos coeficientes de señal de audio, si la trama
25 actual no es recibida por la interfaz receptora (1110) o si la trama actual recibida por la interfaz receptora (1110) está alterada.

2. Un aparato según la reivindicación 1, en el que uno o más primeros coeficientes de señal de audio son uno o más coeficientes de filtro predictivo lineal de la señal de audio codificada.

30 3. Un aparato según la reivindicación 2, en el que el uno o más coeficientes de filtro predictivos lineales se representan por uno o más pares espectrales de impedancia y admitancia o por uno o más pares espectrales lineales, o por uno o más frecuencias espectrales de impedancia y admitancia, o por una o más frecuencias espectrales lineales de la señal de audio codificada.

35 4. Un aparato según una de las reivindicaciones anteriores, en el que el uno o más coeficientes de ruido son uno o más coeficientes de filtro predictivo lineal que indican el ruido de fondo de la señal de audio codificada.

5. Un aparato según una de las reivindicaciones 2 a 4, en el que el uno o más coeficientes de filtro
40 predictivo lineal representan una forma espectral del ruido de fondo.

6. Un aparato según una de las reivindicaciones anteriores, en el que el generador de coeficientes (1120) está configurado para determinar la una o más segundas porciones de señales de audio de tal forma que la una o más segundas porciones de señal de audio sean uno o más coeficientes de filtro predictivo lineal de la señal de
45 audio reconstruida.

7. Un aparato según la reivindicación 1, en el que el generador de coeficientes (1120) está configurado para generar el uno o más segundos coeficientes de señal de audio aplicando la fórmula:

50
$$f_{actual}[i] = \alpha \cdot f_{último}[i] + (1 - \alpha) \cdot pt_{media}[i]$$

en la que $f_{actual}[i]$ indica uno del uno o más segundos coeficientes de señal de audio, en la que $f_{último}[i]$ indica uno del uno o más primeros coeficientes de señal de audio, en la que $pt_{media}[i]$ es uno del uno o más coeficientes de ruido,
55 en la que α es un número real con $0 \geq \alpha \geq 1$, y en la que i es un índice.

8. Un aparato según la reivindicación 7,

60 en el que $f_{último}[i]$ indica un coeficiente de filtro predictivo lineal de la señal de audio codificada, y

en la que $f_{actual}[i]$ indica un coeficiente de filtro predictivo lineal de la señal de audio reconstruida.

9. Un aparato según la reivindicación 8, en el que $pt_{media}[i]$ indica el ruido de fondo de la señal de audio codificada.
- 5 10. Un aparato de acuerdo con una de las reivindicaciones, en el que el generador de coeficientes (1120) está configurado 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, el uno o más coeficientes de ruido mediante la determinación de un espectro de ruido de la señal de audio codificada.
- 10 11. Un aparato según una de las reivindicaciones anteriores, en el que el generador de coeficientes (1120) está configurado para determinar coeficientes de LPC que representan ruido de fondo usando un enfoque 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 una forma de ruido de fondo a partir del espectro de ruido de fondo.
- 15 12. Un método para la decodificación de una señal de audio codificada para obtener una señal de audio reconstruida, en el que el método comprende:
- recibir una o más tramas.
- 20 determinar, si una trama actual de una o más tramas es recibida y si la trama actual recibida no está alterada, de uno o más primeros coeficientes de señal de audio, estando comprendidos por la trama actual, en el que 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,
- 25 generar uno o más segundos coeficientes de señal de audio, que dependen de los uno o más primeros coeficientes de señal de audio y que dependen de los uno o más coeficientes de ruido, si la trama actual no es recibida o si la trama actual recibida está alterada,
- reconstruir una primera porción de la señal de audio reconstruida que depende de los uno o más primeros coeficientes de señal de audio, si la trama actual es recibida y si la trama de actual que se recibe no está alterada, y reconstruir una segunda porción de la señal de audio reconstruida que depende de los uno o más segundos
- 30 coeficientes de señal de audio, si la trama actual no es recibida o si la trama de actual que se recibe está alterada.
13. Un programa informático adaptado para implementar el método de la reivindicación 12 cuando se ejecuta en un ordenador o procesador de señal.

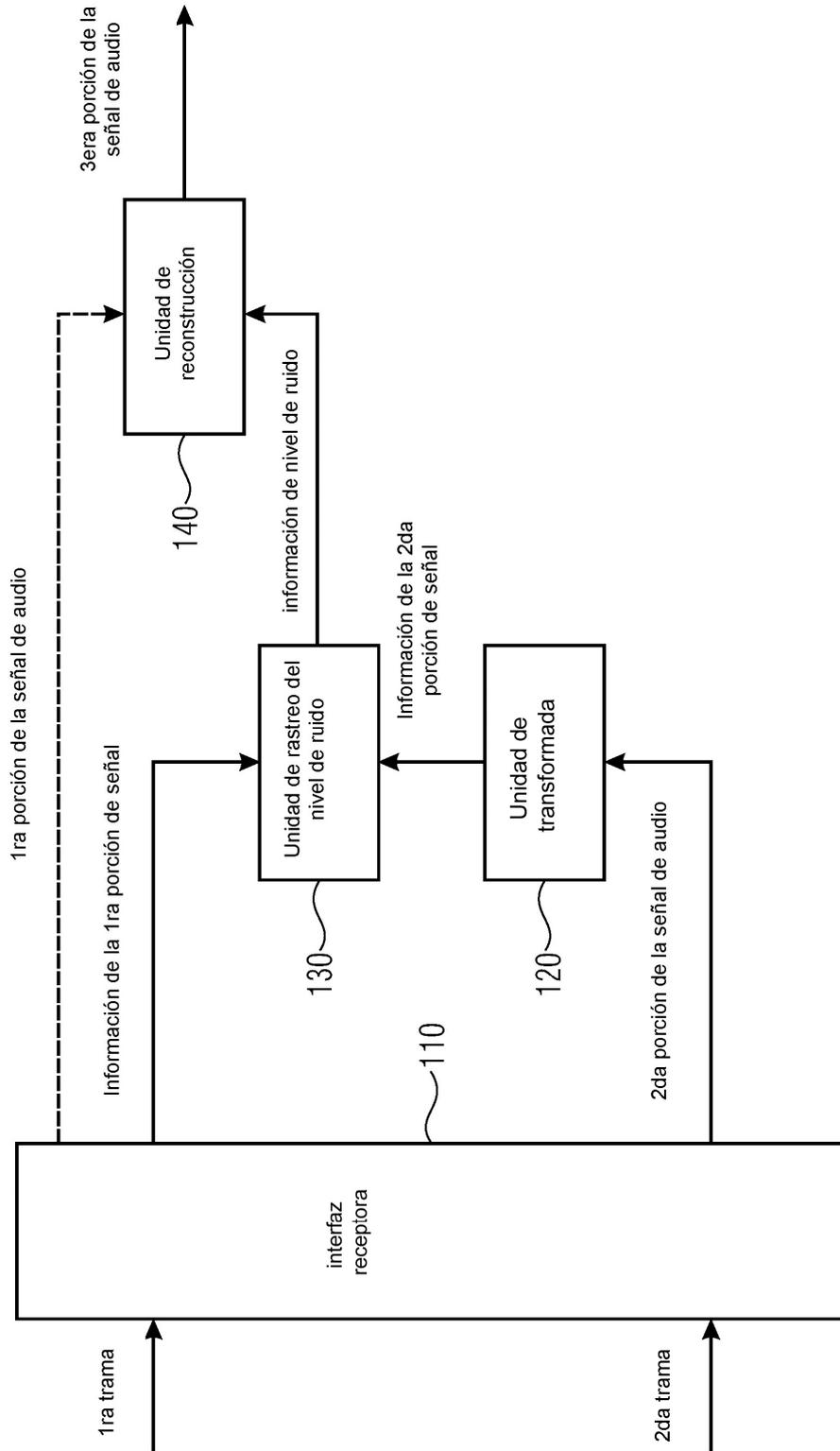


FIG 1A

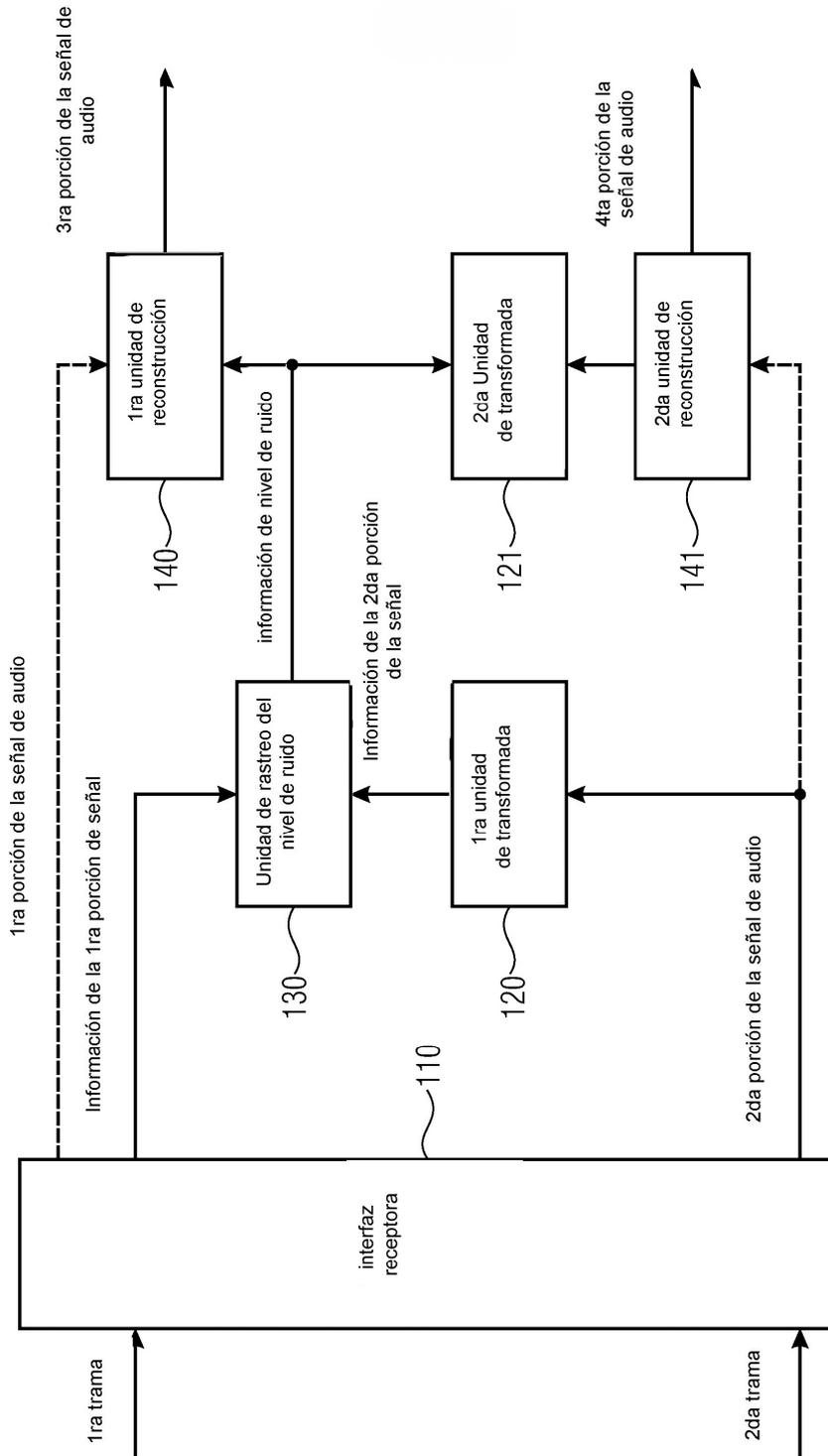


FIG 1B

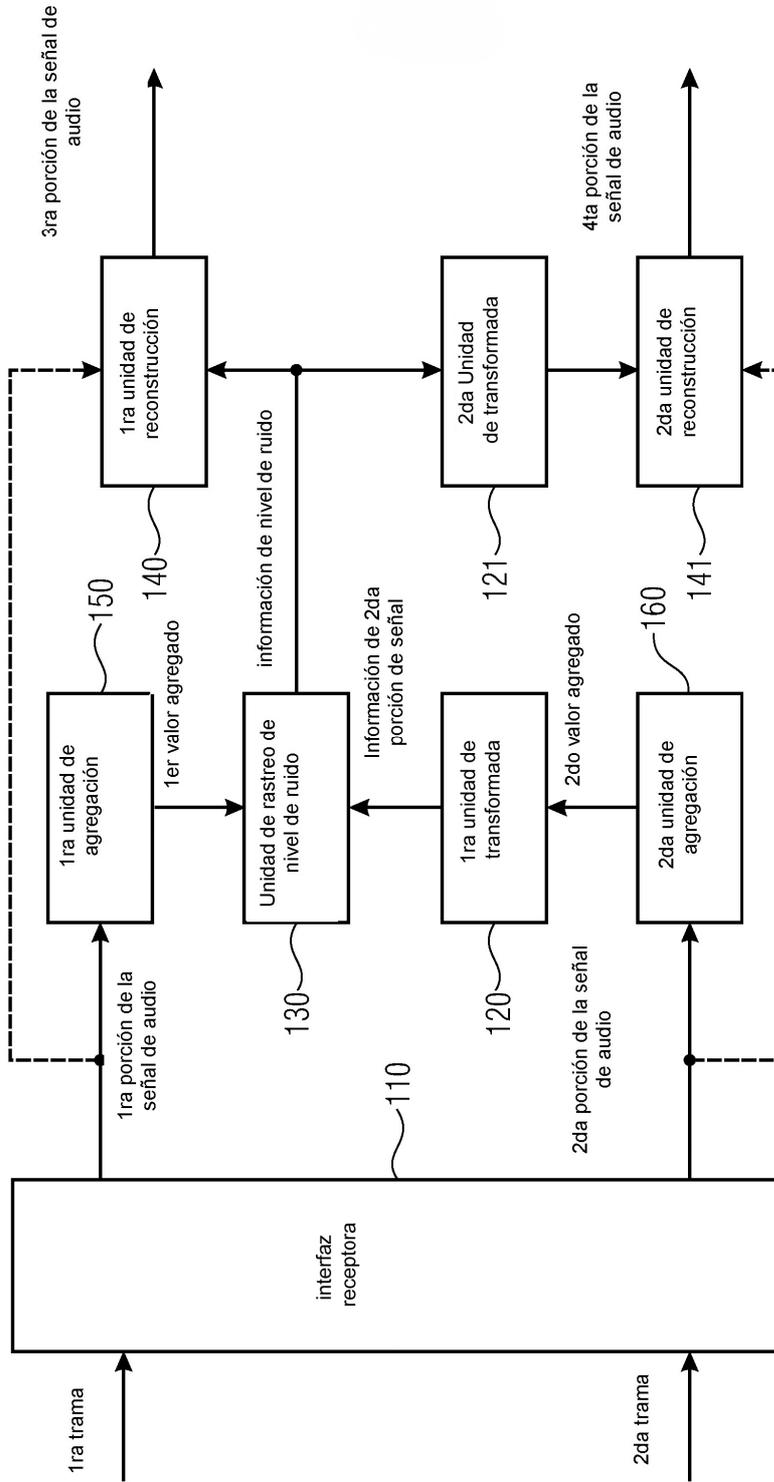


FIG 1C

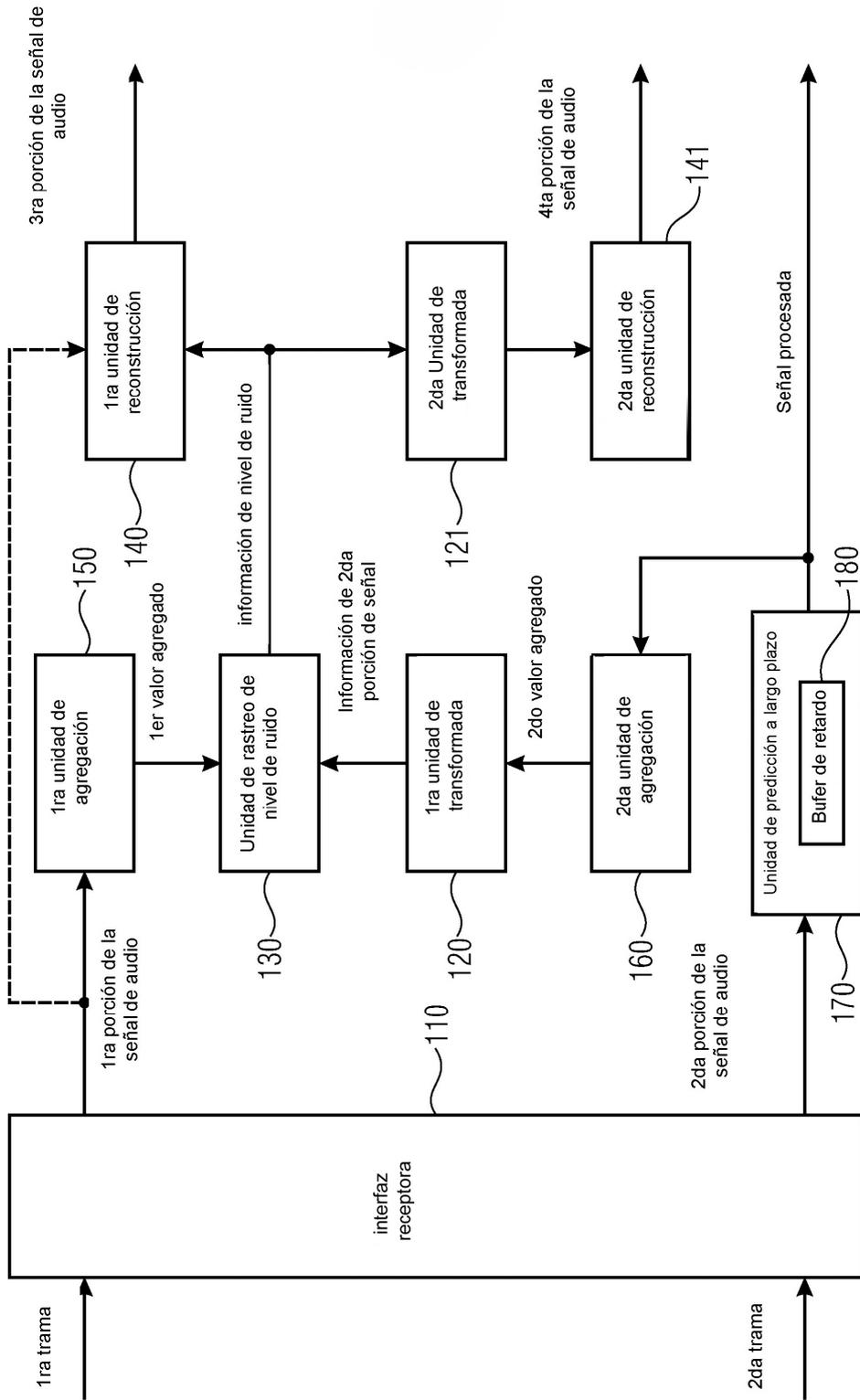


FIG 1D

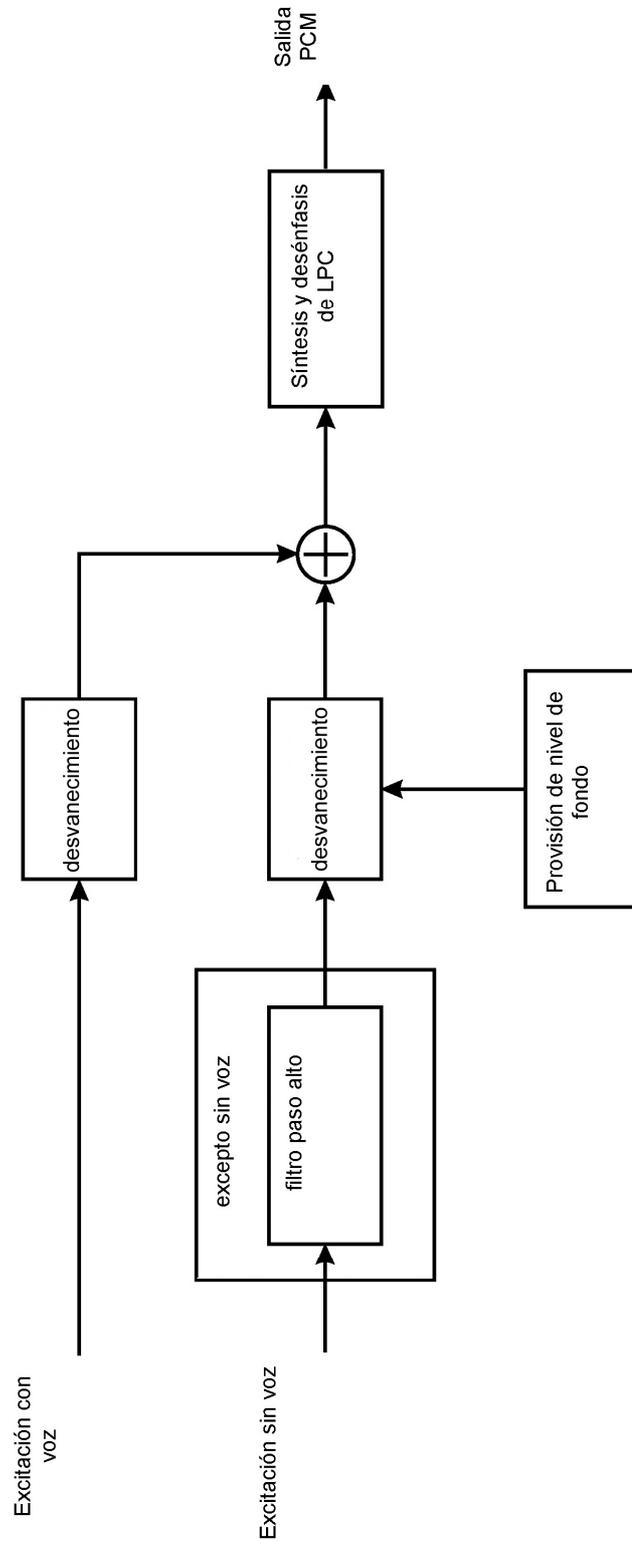
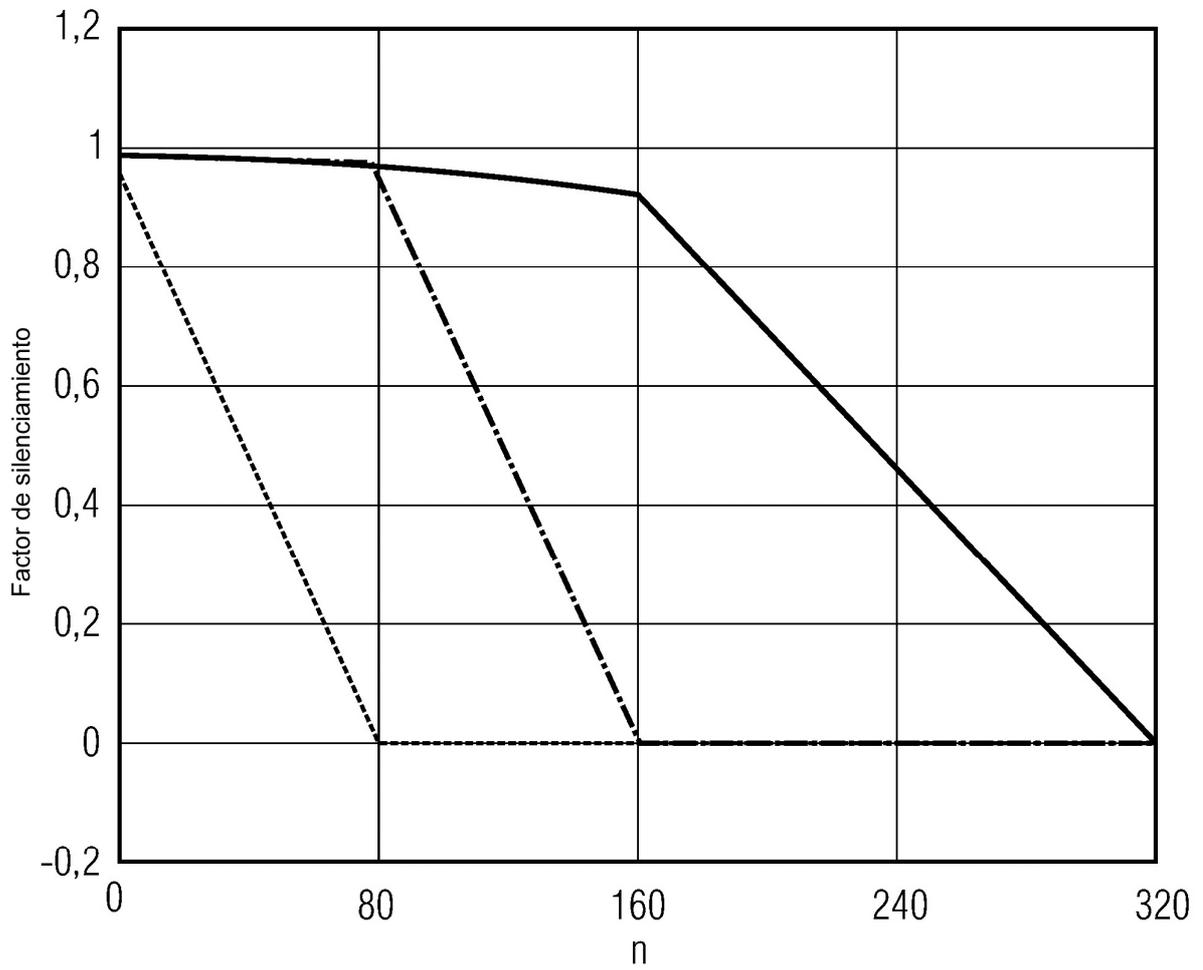


FIG 2



G.722APP:IV_F07

- transitorio
- . - . VUV_TRANSITION
- Otras clases

FIG 3

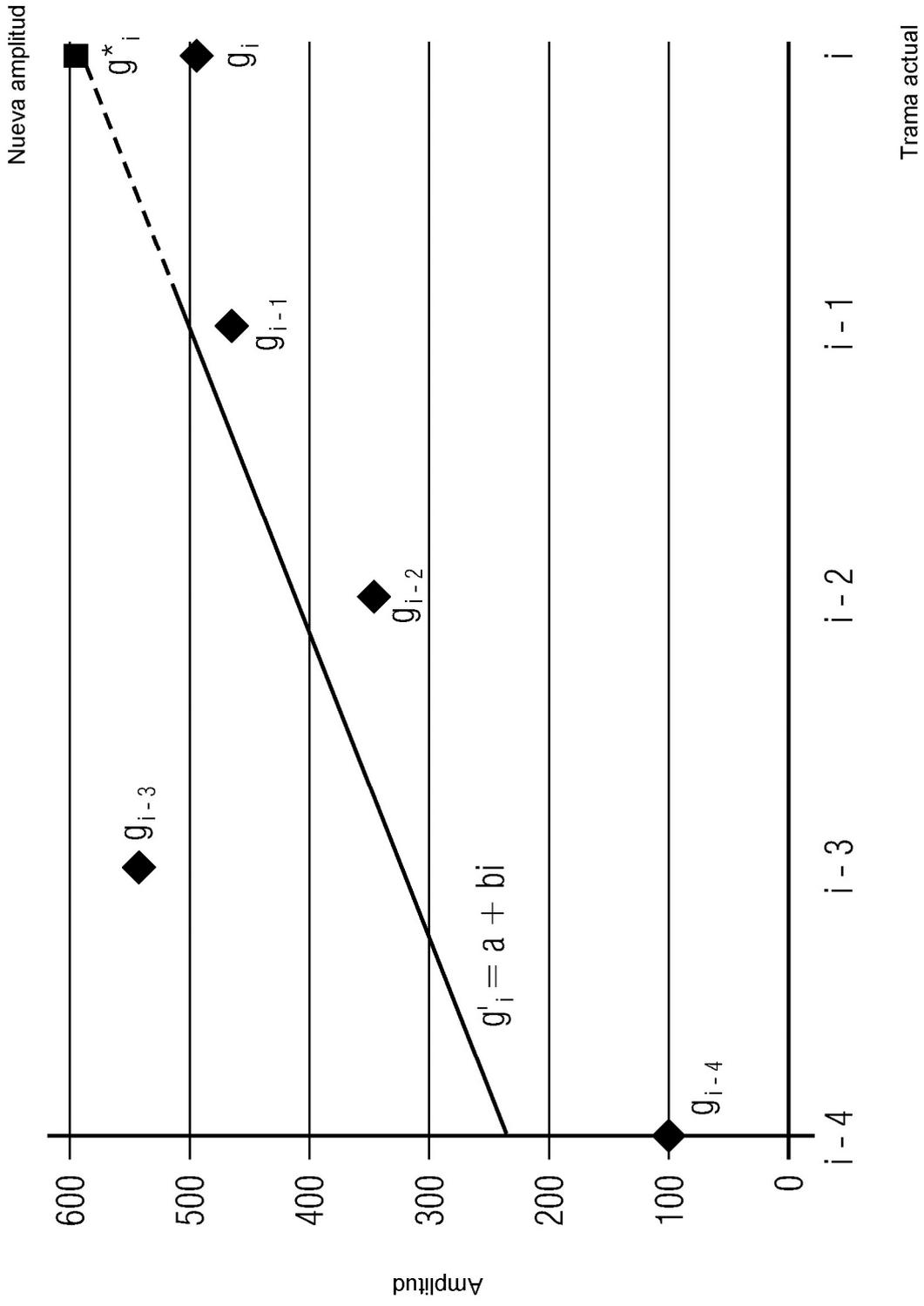


FIG 4

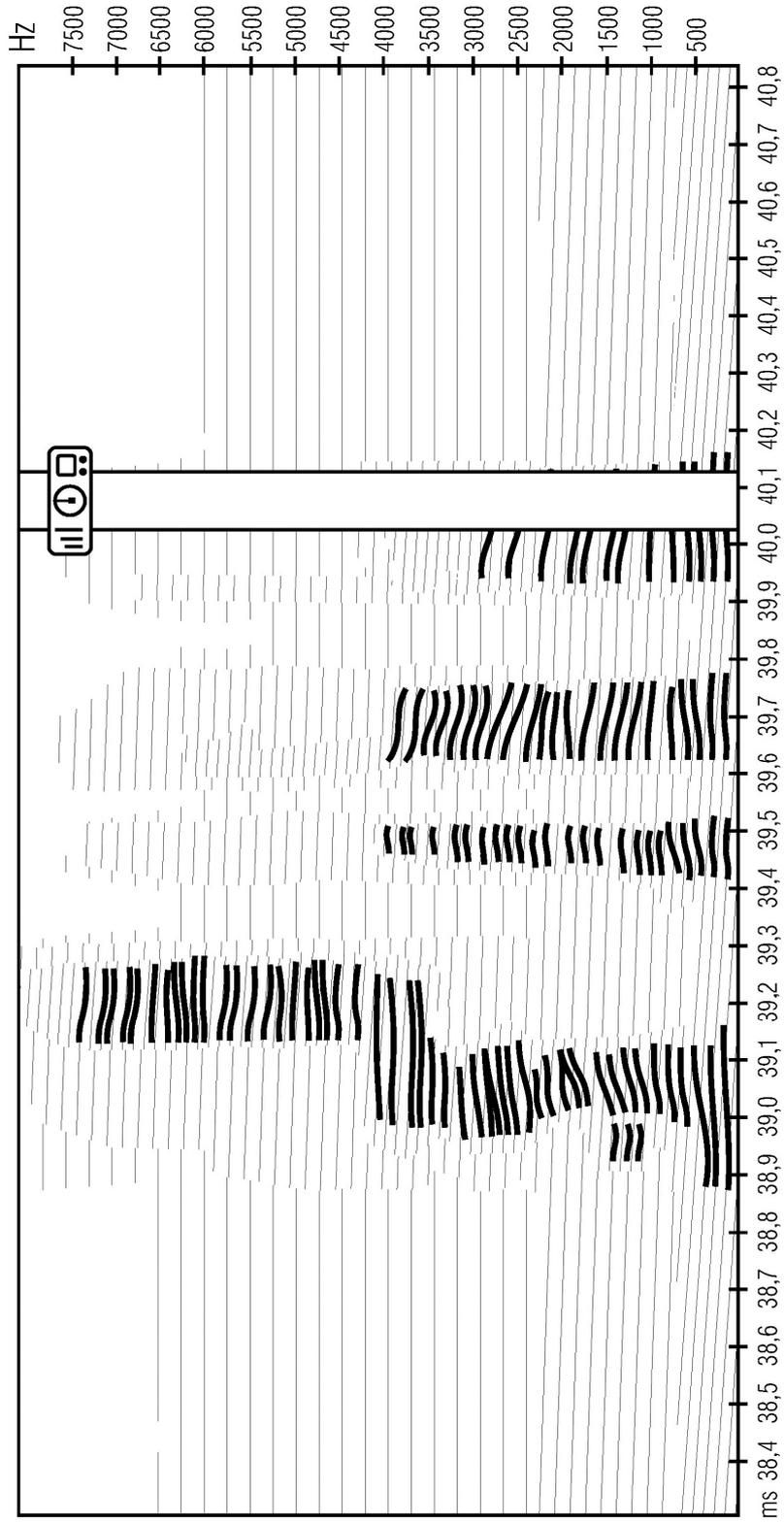


FIG 5

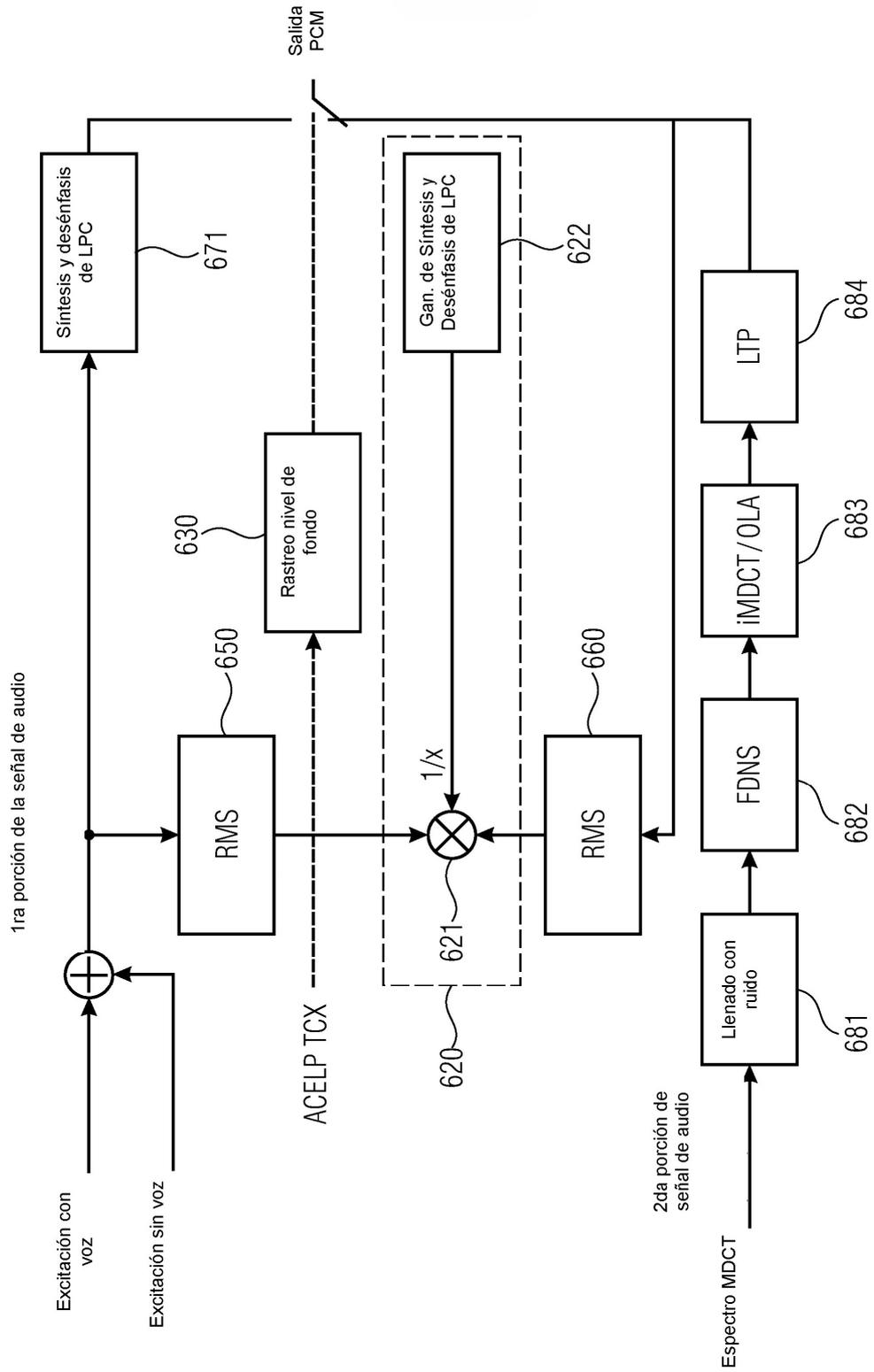


FIG 6

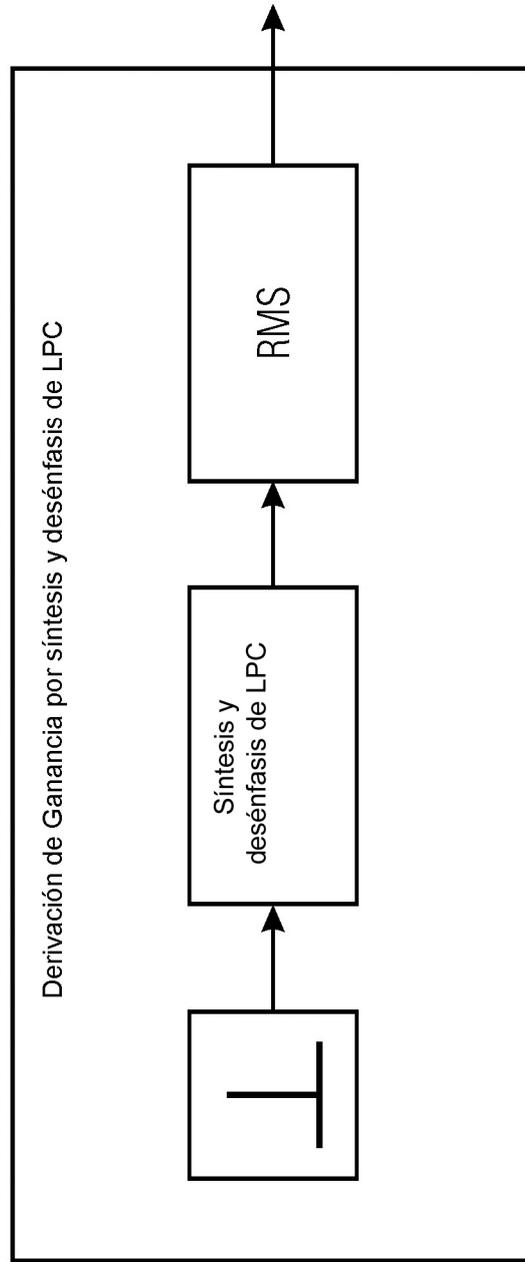


FIG 7

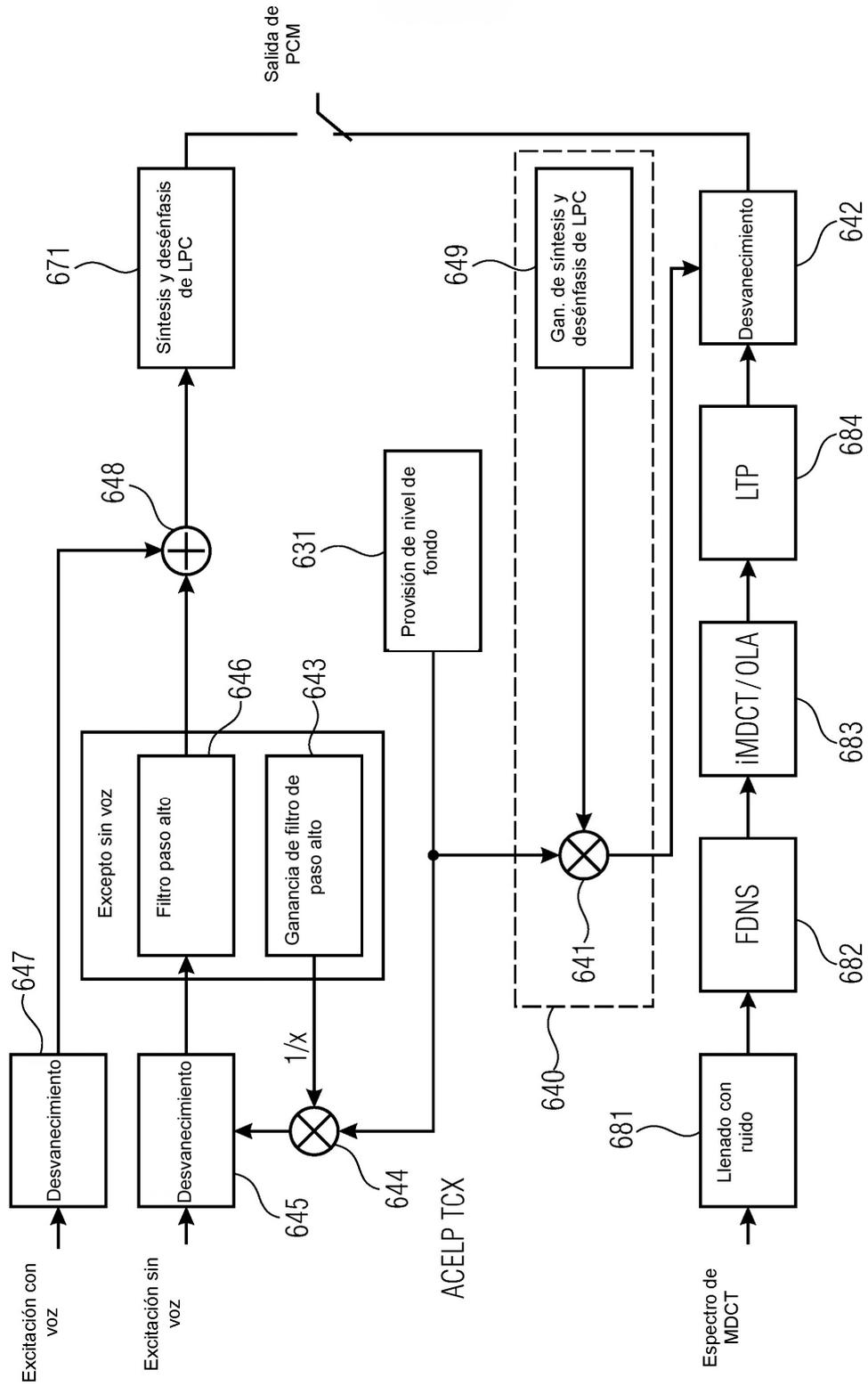


FIG 8

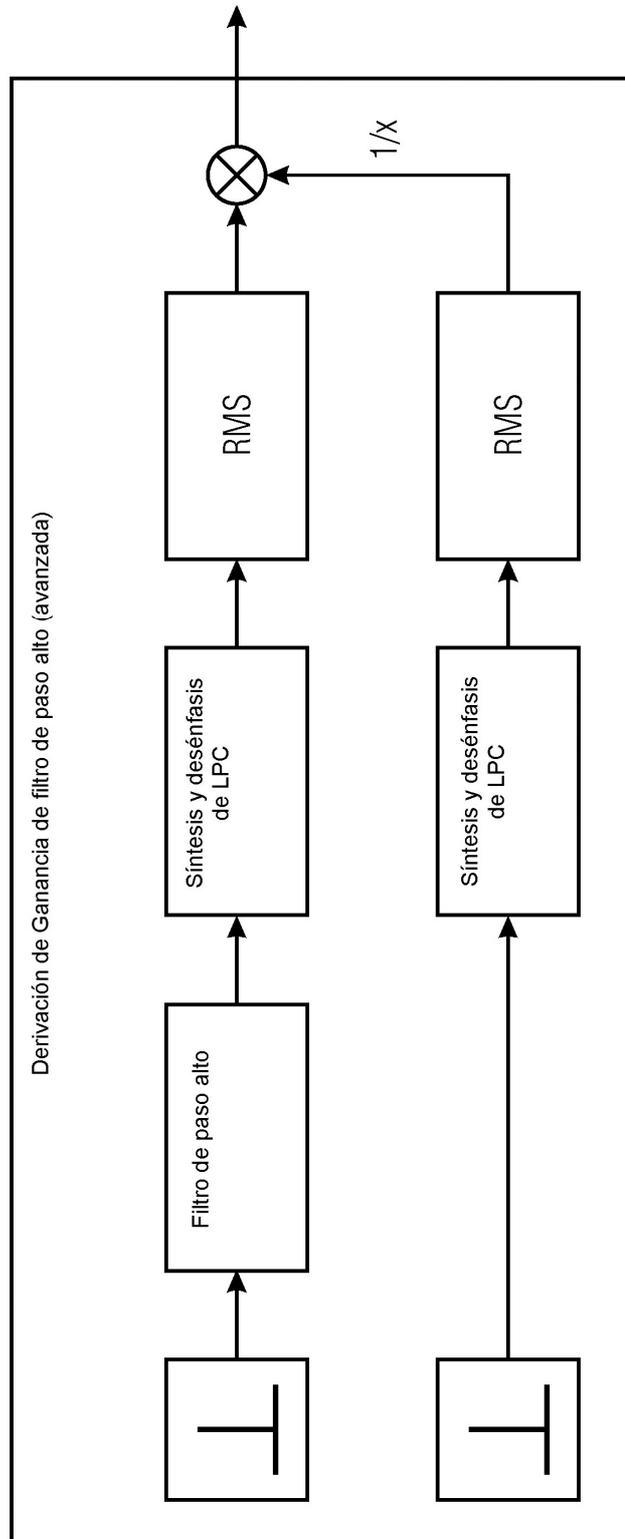


FIG 9

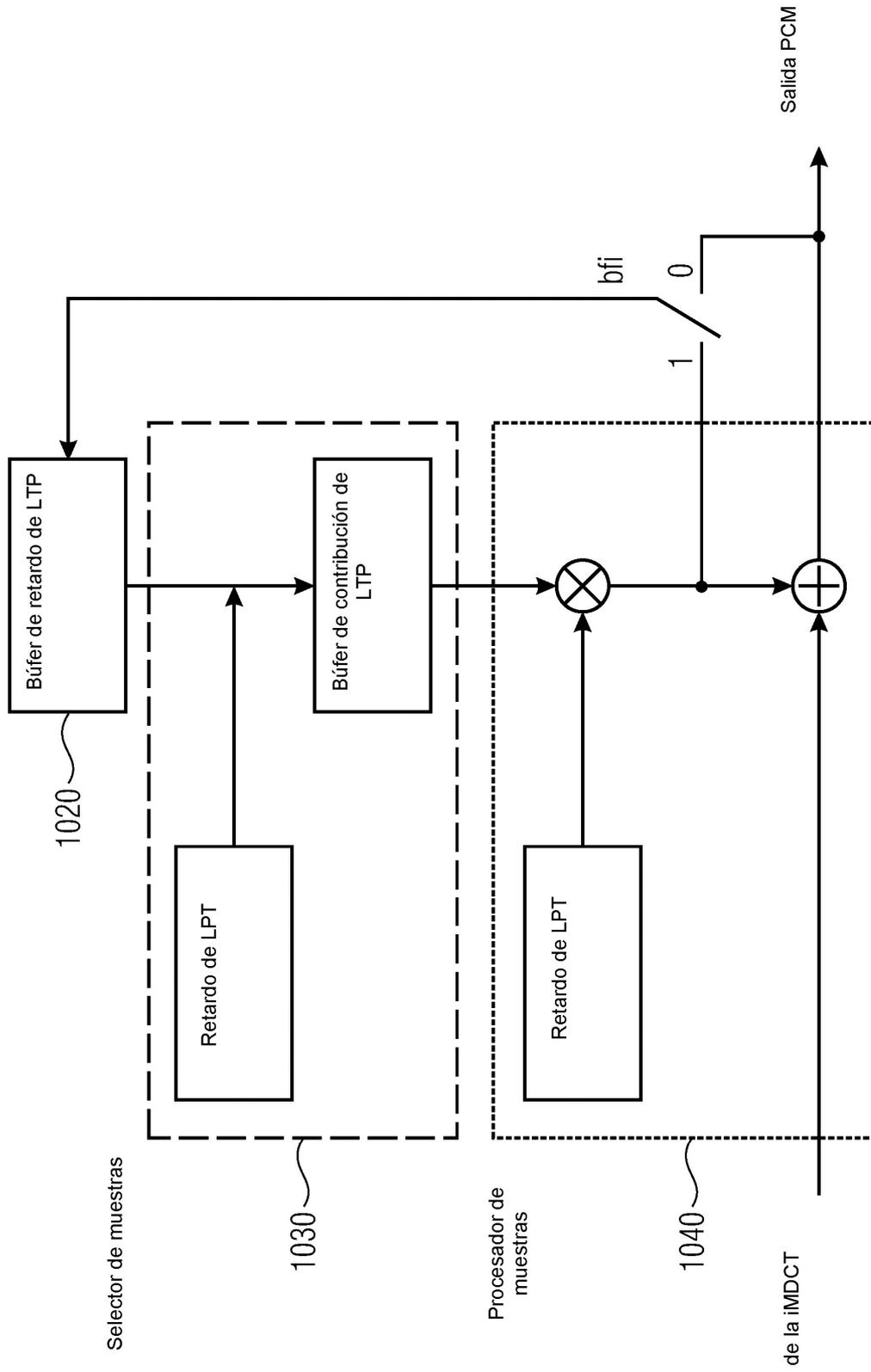


FIG 10

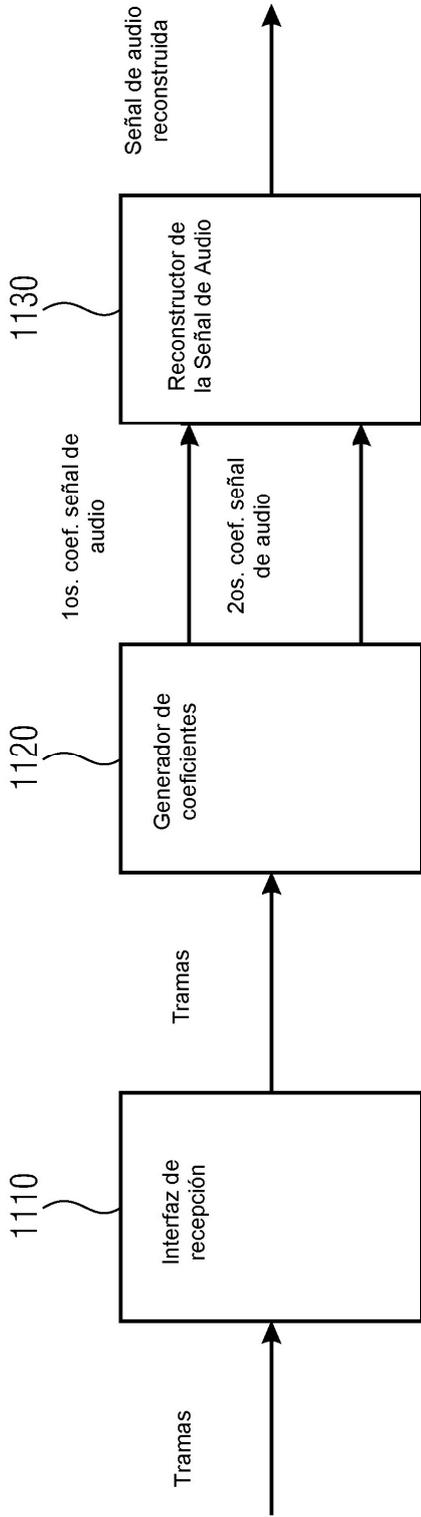


FIG 11

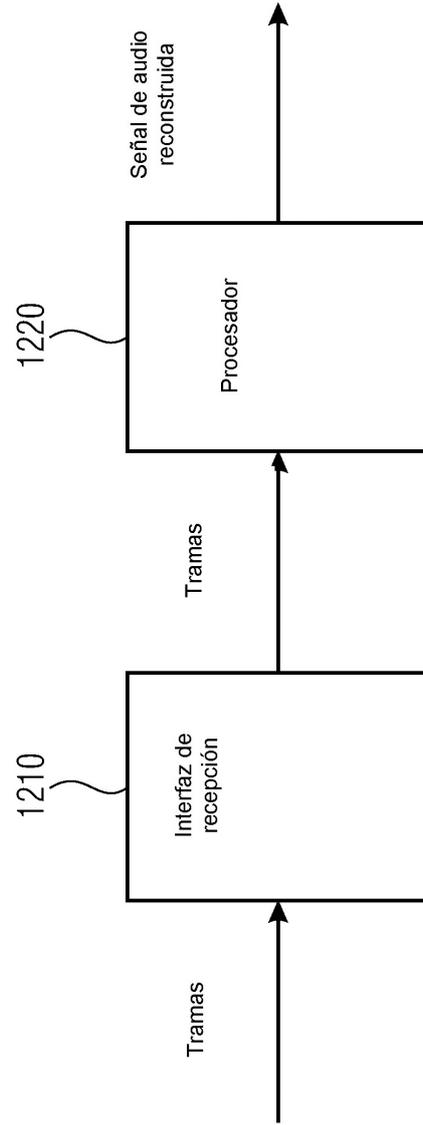


FIG 12

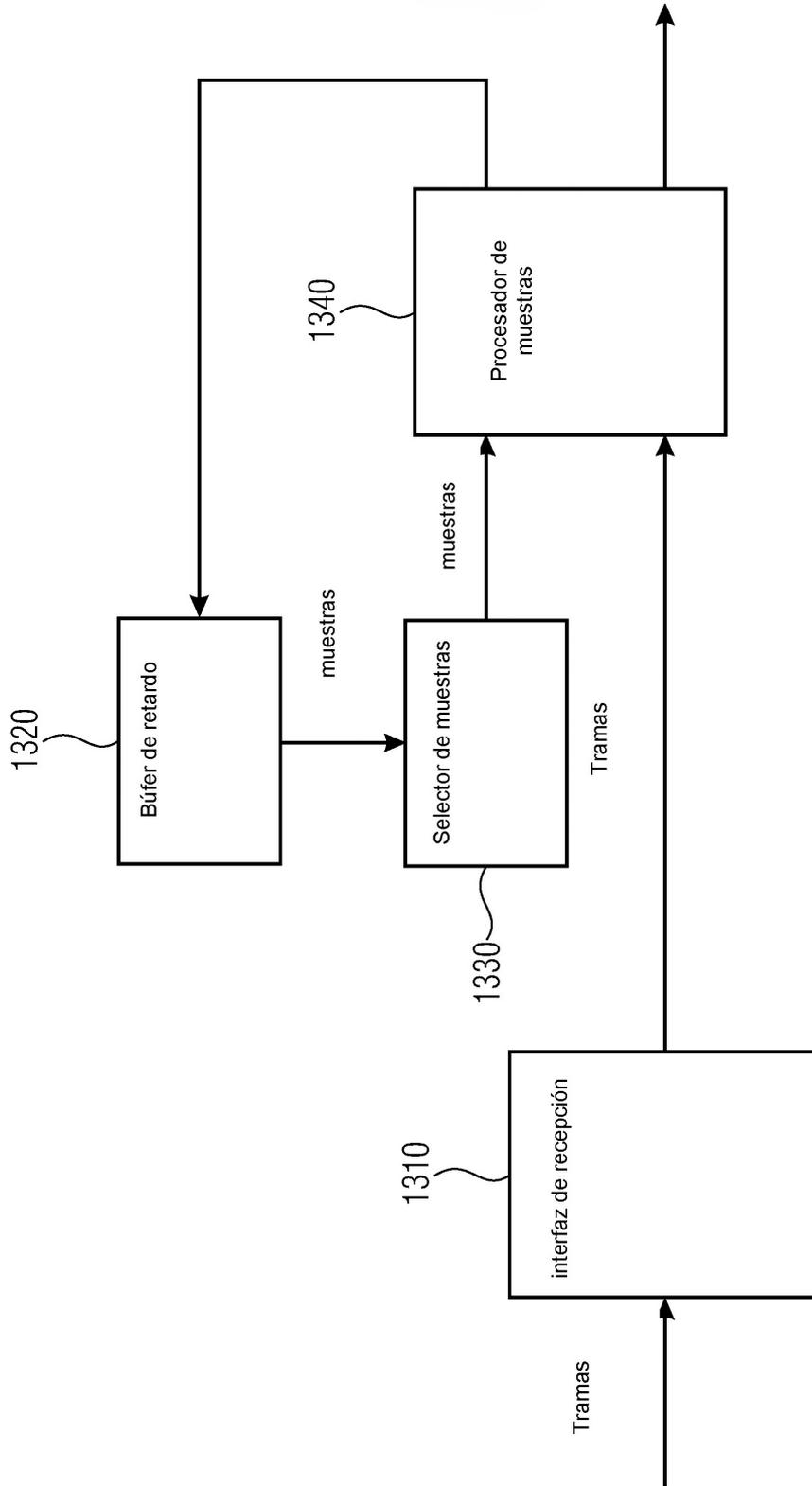


FIG 13

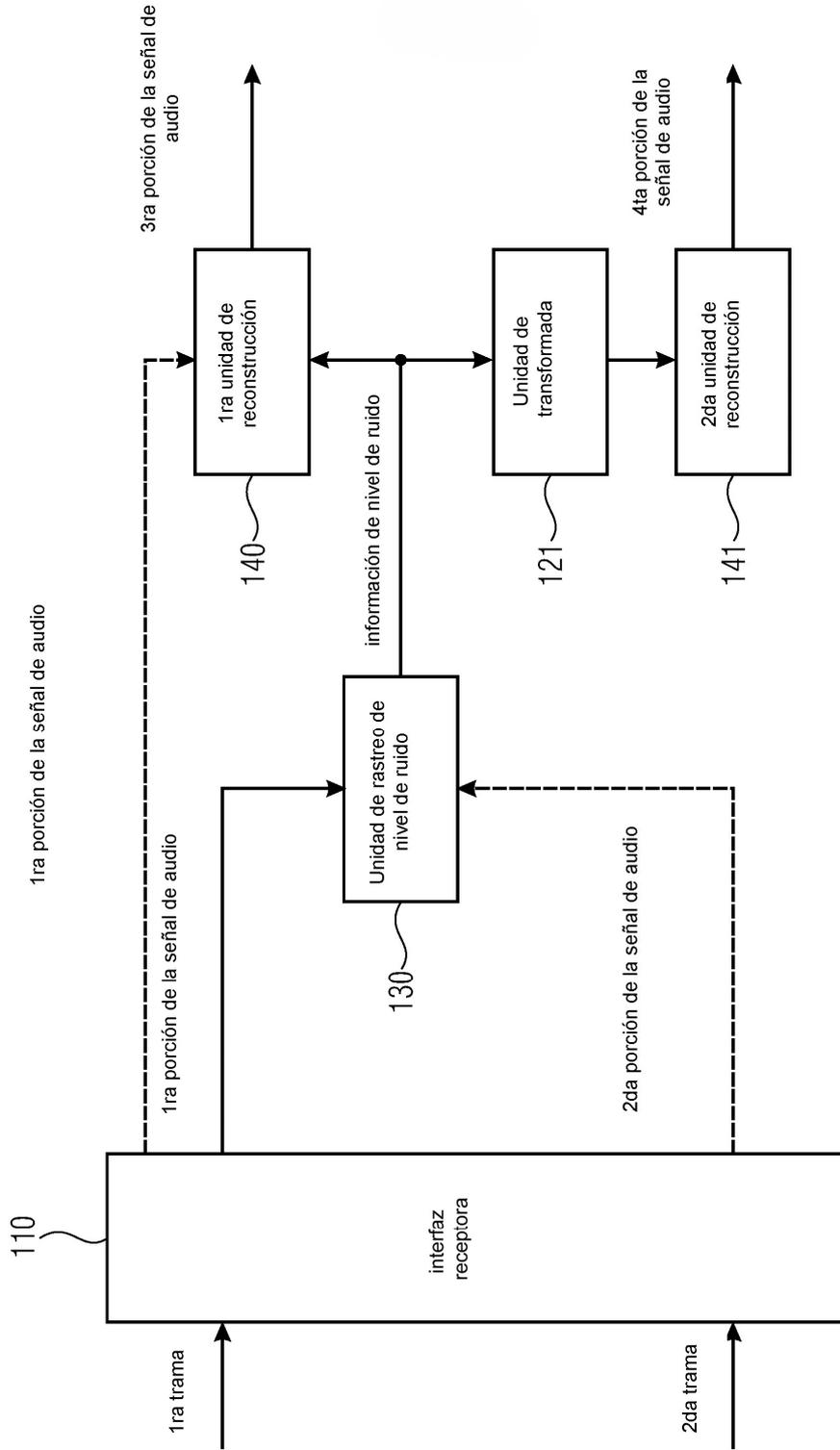


FIG 14