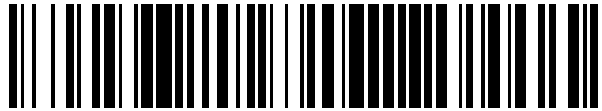


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 639 127**

51 Int. Cl.:

**G10L 19/005** (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **23.06.2014 PCT/EP2014/063175**

87 Fecha y número de publicación internacional: **24.12.2014 WO14202788**

96 Fecha de presentación y número de la solicitud europea: **23.06.2014 E 14732195 (4)**

97 Fecha y número de publicación de la concesión europea: **26.07.2017 EP 3011559**

54 Título: **Aparato y procedimiento que realiza un desvanecimiento de un espectro MDCT a ruido blanco antes de la aplicación de FDNS**

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:

**25.10.2017**

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 2 639 127 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 procedimiento que realiza un desvanecimiento de un espectro MDCT a ruido blanco antes de la aplicación de FDNS

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 procedimiento para el desvanecimiento señales mejorado para sistemas de codificación de audio conmutados durante el ocultamiento de errores.

10 **[0002]** En lo sucesivo, se describe el estado de la técnica respecto al desvanecimiento de códecs [codificadores-decodificadores] de voz y de audio durante el ocultamiento de pérdida de paquetes (PLC). Las explicaciones con respecto al estado de la técnica con los códecs ITU-T de la serie G (G.718, G.719, G.722, G.722.1, G.729, G.729.1) son seguidas por los códecs 3GPP (AMR, AMR-WB, AMR-WB+) y un códec IETF (OPUS), y concluyen con dos códecs MPEG (HE-AAC, HILN) (ITU = Unión Internacional de Telecomunicaciones; 3GPP = Proyecto Asociación de 3ª 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 visión de conjunto

20 **[0003]** En primer lugar, se considera G.718. G.718 es un códec de voz de banda estrecha y de banda ancha, que soporta DTX/CNG (DTX = Digital Theater Systems (marca registrada); CNG = Generación de Ruido de Confort). Dado que realizaciones se refieren en particular al código de retardo bajo, en este contexto se describirá con más detalle el modo de versión de retardo bajo.

25 **[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 resumirse 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 recibida correctamente y el número de tramas borradas consecutivas, y es controlada por un factor de atenuación,  $\alpha$ . El factor de atenuación  $\alpha$  depende además de la estabilidad,  $\theta$ , del filtro LP (LP = Predicción Lineal) para tramas SIN VOZ. En general, la convergencia es lenta si la última buena trama recibida está en un segmento estable y es rápida si la trama está en un segmento de transición.

35

**[0005]** El factor de atenuación  $\alpha$  depende de la clase de señal de voz, que se deriva mediante clasificación de señales descrita en [ITU08a, sección 6.8.1.3.1 y 7.11.1.1]. El factor de estabilidad  $\theta$  se calcula sobre la base de una medida de distancia entre los filtros ISF (Frecuencia Espectral de Inmitancia) adyacentes [ITU08a, sección 7.1.2.4.2].

40 **[0006]** La tabla 1 muestra el esquema de cálculo de  $\alpha$ :

Tabla 1: Valores del factor de atenuación  $\alpha$ , el valor  $\theta$  es un factor de estabilidad calculado a partir de una medida de distancia entre los filtros de LP adyacentes. [ITU08a, sección 7.1.2.4.2].

última trama buena recibida	Número de tramas borradas sucesivas	$\alpha$
INICIO ARTIFICIAL		0,6
INICIO, CON VOZ	$\leq 3$	1,0
	$> 3$	0,4
TRANSICIÓN CON VOZ		0,4
TRANSICIÓN SIN VOZ		0,8
SIN VOZ	= 1	$0,2 \cdot \theta + 0,8$
	= 2	0,6
	$> 2$	0,4

45 **[0007]** Además, G.718 proporciona un procedimiento de desvanecimiento a fin de modificar la envoltura espectral. La idea general es hacer converger los últimos parámetros de ISF hacia un vector de la media de ISF adaptativo. En primer lugar, se calcula un vector de ISF promedio a partir de los últimos 3 vectores de ISF conocidos. A continuación, se promedia nuevamente el vector de ISF promedio con un vector de ISF a largo plazo capacitado fuera de línea (que es un vector constante) [ITU08a, sección 7.11.1.2].

**[0008]** Además, G.718 proporciona un procedimiento de desvanecimiento para controlar el comportamiento a largo plazo, y, en consecuencia, la interacción con el ruido de fondo, donde la energía de excitación de altura tonal (y, por lo tanto, la periodicidad de excitación) converge en 0, mientras que la energía de excitación aleatoria converge en la energía de excitación de 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  $\alpha$ .

**[0009]** De forma similar a la atenuación de excitación periódica, la ganancia se atenúa de forma lineal a lo largo de la trama muestra por muestra comenzando con,  $g_s^{[0]}$  y alcanza  $g_s^{[1]}$  al comienzo de la siguiente trama.

**[0010]** La figura 2 perfila la estructura del decodificador de G.718. En particular, la figura 2 ilustra una estructura de decodificador de G.718 de alto nivel para PLC, que presenta un filtro de paso alto.

**[0011]** Mediante el enfoque descrito anteriormente de G.718, la ganancia innovadora  $g_s$  converge con la ganancia usada durante la generación de ruido de confort  $g_n$  para ráfagas largas de pérdidas de paquetes. Tal como se describe en [ITU08a, sección 6.12.3], la ganancia de ruido de confort  $g_n$  viene dada como la raíz cuadrada de la energía  $\tilde{E}$ . Las condiciones de la actualización de  $\tilde{E}$  no se describen con detalle. Siguiendo la implementación de referencia (código C de punto flotante, *stat\_noise\_uv\_mod.c*),  $\tilde{E}$  se deriva de la siguiente manera:

```

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

```

en la que *unvoiced\_vad* contiene la detección de actividad de voz, en la que *unv\_cnt* contiene el número de tramas sin voz en una fila, en la que *lp\_gainc* contiene las ganancias de paso bajo del libro de códigos fijo, y en la que *lp\_ener* contiene la estimación de energía de CNG de paso bajo  $E$ , 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 trama buena se clasificó diferente de SIN VOZ, véase la figura 2, véase también [ITU08a, sección 7.11.1.6]. Este filtro tiene una característica paramétrica baja con una respuesta de frecuencia en CC que es aproximadamente 5 dB más baja que en la frecuencia Nyquist.

**[0013]** Además, G.718 propone un bucle de retroalimentación de LTP desacoplado (LTP = predicción a largo plazo): aunque durante el funcionamiento normal el bucle de retroalimentación para el libro de códigos adaptativo se actualiza por subtramas ([ITU08a, sección 7.1.2.1.4]) basándose en la excitación completa. Durante el ocultamiento, este bucle de retroalimentación se actualiza por tramas (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 excitación con voz solamente. Con este enfoque, el libro de códigos adaptativo no está "contaminado" con ruido que tiene su origen en la excitación de innovación seleccionada aleatoriamente.

**[0014]** Respecto a las capas de mejora codificadas por transformada (3-5) de G.718, durante el ocultamiento,

el decodificador se comporta respecto a la decodificación de capa alta, similar al funcionamiento normal, sólo que el espectro MDCT se ajusta a cero. No se aplica comportamiento de desvanecimiento especial durante el ocultamiento.

- [0015]** Con respecto a CNG, en G.718, la síntesis de CNG se realiza en el siguiente orden. En primer lugar, se decodifican parámetros de una trama de ruido de confort. A continuación, se sintetiza una trama de ruido de confort. Seguidamente, el búfer de altura tonal se restablece. A continuación, se guarda la síntesis para la FER (recuperación de error de trama). Seguidamente, se lleva a cabo el desénfasis del espectro. A continuación, se lleva a cabo postfiltrado de baja frecuencia. A continuación, se actualizan las variables de CNG.
- 10 **[0016]** En el caso de ocultamiento, se realiza exactamente lo mismo, excepto que los parámetros de CNG no se decodifican a partir del flujo de bits. Esto significa que los parámetros no se actualizan durante la pérdida de trama, sino que se usan los parámetros decodificados a partir de la última trama de SID (descriptor de inserción de silencio) buena.
- 15 **[0017]** Ahora, se considera G.719. G.719, que se basa en Siren 22, es un códec de audio de banda completa basado en 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 recibida correctamente, los coeficientes de transformada reconstruidos son almacenados en un búfer. Si el decodificador es informado de que se ha perdido una trama o que una trama está alterada, los coeficientes de transformada reconstruidos en la trama recibida más recientemente son convertidos a escala decreciente con un factor 0,5, y a continuación, son usados como los coeficientes de transformada reconstruidos para la trama actual. El decodificador continúa transformándolos en el dominio temporal, y la realización de la operación de generación de ventanas-superposición-adición.
- 20 **[0018]** En lo sucesivo, se describe G.722. G.722 es un sistema de codificación de 50 a 7000 Hz que usa modulación por código de pulsos diferenciales adaptativos de subbanda (SB-ADPCM) dentro de una tasa de bits 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 código de pulsos diferenciales adaptativos).
- 25 **[0019]** Para G.722, se especifica un algoritmo de alta complejidad para ocultamiento de pérdida de paquetes en el Apéndice III [ITU06a] y se especifica un algoritmo de baja complejidad para ocultamiento de pérdida de paquetes en el Apéndice IV [ITU07]. G.722 - Apéndice III ([ITU06a, sección 111.5]) propone un silenciamiento realizado gradualmente, que comienza después de 20 ms de pérdida de tramas, que está completo después de 60 ms de pérdida de tramas. Además, G.722 - Apéndice IV propone una técnica de desvanecimiento que aplica "a cada muestra un factor de ganancia que se calcula y se adapta muestra por muestra" [ITU07, sección IV.6.1.2.7].
- 30 **[0020]** En G.722, el proceso de silenciamiento tiene lugar en el dominio de subbanda justo antes de la síntesis de QMF y como la última etapa del módulo de PLC. El cálculo del factor de silenciamiento se realiza usando información de clase del clasificador de señales que también es parte del módulo de PLC. Se realiza la distinción entre clases *TRANSIENT* (transitoria), *UV\_TRANSITION* (uv\_transición) y otras. Además, se realiza la distinción entre pérdidas individuales de tramas de 10 ms y otros casos (múltiples pérdidas de tramas de 10 ms y pérdidas únicas/múltiples de tramas de 20 ms).
- 35 **[0021]** Esto se ilustra mediante la figura 3. En particular, la figura 3 representa un escenario, donde el factor de desvanecimiento de G.722, depende de información de clase y en el que 80 muestras son equivalentes a 10 ms.
- 40 **[0022]** De acuerdo con G.722, el módulo de PLC crea la señal para la trama que falta y alguna señal adicional (10 ms) que se supone que se somete a desvanecimiento gradual con la siguiente trama buena. El enmudecimiento para esta señal adicional sigue las mismas normas. En ocultamiento de banda alta de G.722, el desvanecimiento gradual no tiene lugar.
- 45 **[0023]** En lo sucesivo, se considera G.722.1. G.722.1, que se basa en Siren 7, es un códec de audio de banda ancha basado en transformada con un modo de extensión de banda superancho, denominado G.722.1C. G.722.1C, a su vez, se basa en Siren 14. La ITU-T recomienda para G.722.1 una repetición de tramas con 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 anterior. Continúa transformándolos al dominio temporal, y realizando la operación de superposición y adición con la información
- 50

decodificada de la trama anterior y siguiente. Si la trama anterior también se perdió o se alteró, entonces el decodificador ajusta todos los coeficientes de MLT de las tramas actuales a cero.

**[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. Se describe oficialmente como codificación de voz a 8 kbit/s usando codificación de voz de predicción lineal con excitación por código (CS-ACELP) [ITU12].

**[0025]** Tal como se perfila en [CPK08], G.729 recomienda un desvanecimiento en el dominio de LP. El algoritmo de PLC empleado en la norma G.729 reconstruye la señal de voz para la trama actual basándose en información de voz recibida previamente. En otras palabras, el algoritmo de PLC sustituye la excitación que falta con una característica equivalente de una trama recibida previamente, aunque la energía de excitación declina gradualmente, finalmente, las ganancias de los libros de códigos adaptativo y fijo están atenuadas por un factor constante.

15 **[0026]** La ganancia del libro de códigos fijo atenuada viene dada por:

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

con m es el índice de subtrama.

20 **[0027]** La ganancia del libro de códigos adaptativo se basa en una versión atenuada de la ganancia del libro de códigos adaptativo anterior:

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

**[0028]** Nam In Park y col. sugieren para G.729, un control de amplitud de señal usando predicción por medio de regresión lineal [CPK08, PKJ+11]. Se dirige a la pérdida de paquetes en ráfagas y usa regresión lineal como una técnica fundamental. La regresión lineal se basa en el modelo lineal como

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

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

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

$\epsilon$  es el error al cuadrado,  $g_j$  es la amplitud j-ésima pasada original. Para minimizar este error, simplemente la derivada respecto a y b se ajusta a cero. Usando los parámetros optimizados  $a^*$  y  $b^*$ , una estimación de cada  $g_i^*$  se indica mediante

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

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

**[0030]** Para obtener la amplitud  $A'_i$  del paquete perdido i, una relación

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

se multiplica con un factor de escala

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

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

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

**[0031]** En [PKJ+11], se propone una conversión a escala ligeramente diferente.

10 **[0032]** De acuerdo con G.729, seguidamente,  $A'_i$  se suavizará para prevenir atenuación discreta en los bordes de trama. La amplitud suavizada final  $A_i(n)$  se multiplica hasta la excitación, obtenida de los componentes de PLC previos.

15 **[0033]** En lo sucesivo, se considera G.729.1. G.729.1 es un codificador de tasa de bits variable insertado basado en G.729: un flujo de bits de codificador de banda ancha convertible a escala de 8-32 kbit/s inter-operable con G.729 [ITU06b].

20 **[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 señal ([ITU06b, sección 7.6.1]). Durante el ocultamiento, la señal se atenúa habitualmente basándose en un factor de atenuación  $\alpha$  que depende de los parámetros de la clase de la última trama buena recibida y el número de tramas borradas consecutivas. El factor de atenuación  $\alpha$  depende además de la estabilidad del filtro de LP para tramas SIN VOZ. En general, la atenuación es lenta si la última trama buena recibida está en un segmento estable y es rápida si la trama está en un segmento de transición.

25

**[0035]** Además, el factor de atenuación  $\alpha$  depende de la ganancia de altura tonal promedio por subtrama  $\bar{g}_p$  ([ITU06b, ec. 163, 164]):

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

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

**[0036]** La tabla 2 muestra el esquema de cálculo de  $\alpha$ , donde

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

35 **[0037]** Durante el proceso de ocultamiento,  $\alpha$  se usa en las siguientes herramientas de ocultamiento:

Tabla 2: valores del factor de atenuación  $\alpha$ , el valor  $\theta$  es un factor de estabilidad calculado a partir de una medida de distancia entre los filtros de LP adyacentes. [ITU06b, sección 7.6.1].

última trama buena recibida	Número de tramas borradas sucesivas	$\alpha$
CON VOZ	1	$\beta$
	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	$\leq 2$	0,8
	$> 2$	0,2
TRANSICIÓN SIN VOZ		0,88
SIN VOZ	1	0,95
	2,3	$0,6 \theta + 0,4$
	> 3	0,4

- [0038]** De acuerdo con G.729.1, respecto a resincronización de pulsos glotales, dado que 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 a 1. La ganancia se atenúa a continuación linealmente a lo largo de la trama muestra por muestra hasta conseguir el valor de  $\alpha$  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 tonal de cada subtrama de la última trama buena. En general, si estas ganancias son mayores de 1, la energía de la señal está aumentando, si son menores que 1, la energía está disminuyendo.  $\alpha$  se ajusta, por lo tanto, a  $\beta = \sqrt{\bar{g}_p}$  tal como se ha descrito anteriormente, véase [ITU06b, ec. 163, 164]. El valor de  $\beta$  se recorta entre 0,98 y 0,85 para evitar aumentos y disminuciones fuertes de energía, véase [ITU06b, sección 7.6.4].

- [0039]** 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 usando las ganancias de excitación de innovación de cada subtrama de la última trama buena:

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

- en la que  $g^{(0)}$ ,  $g^{(1)}$ ,  $g^{(2)}$  y  $g^{(3)}$  son las ganancias del libro de códigos fijo, o de innovación, de las cuatro subtramas de la última trama recibida correctamente. 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  $\alpha$  es tal como se define en la tabla 2 anterior. De manera similar a la atenuación de excitación periódica, la ganancia es atenuada, de este modo, linealmente a lo largo de la trama muestra por muestra comenzando con  $g_s^{(0)}$  y yendo hasta el valor de  $g_s^{(1)}$  que se conseguiría al comienzo de la siguiente trama.

- [0040]** De acuerdo con G.729.1, si la última trama buena es SIN VOZ, solamente se usa la excitación de innovación y está atenuada además mediante un factor de 0,8. En este caso, el búfer de excitación pasado se actualiza con la excitación de innovación dado que no hay ninguna parte periódica de la excitación disponible, véase [ITU06b, sección 7.6.6].

- [0041]** En lo sucesivo, se considera AMR. 3GPP AMR [3GP12b] es un códec de voz que utiliza el algoritmo ACELP. AMR es capaz de codificar voz con una tasa de muestreo de 8000 muestras/s y una tasa de bits entre 4,75 y 12,2 kbit/s y soporta señalización de tramas descriptoras de silencio (DTX/CNG).

**[0042]** En AMR, durante el ocultamiento de errores (véase [3GP12a]), se distingue entre tramas que son propensas al error (errores de bits) y tramas, que están completamente perdidas (no hay datos en absoluto).

**[0043]** Para ocultamiento de ACELP, AMR introduce una máquina de estados que estima la calidad del canal: cuanto mayor sea el valor del contador de estados, peor es la calidad del canal. El sistema comienza en estado 0. Cada vez que se detecta una trama mala, el contador de estados aumenta en uno y se satura cuando llega a 6. Cada vez que se detecta una trama de voz buena, el contador de estados se reajusta a cero, excepto cuando el estado es 6, donde el contador de estados se ajusta a 5. El flujo de control de la máquina de estados puede describirse mediante el siguiente código C (BFI es un indicador de trama mala, *State* es una variable de estado):

10

```

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

**[0044]** Además de esta máquina de estados, en AMR, se verifican las marcas de trama mala de las tramas actual y anterior (*prevBFI*).

15

**[0045]** Son posibles tres combinaciones diferentes:

20 La primera de las tres combinaciones es  $BFI = 0$ ,  $prevBFI = 0$ ,  $State = 0$ : No se detecta ningún error en la trama de voz recibida o en la recibida anterior. Los parámetros de voz recibidos se usan de la manera normal en la síntesis de voz. La trama actual de los parámetros de voz se guarda.

**[0046]** La segunda de las tres combinaciones es  $BFI = 0$ ,  $prevBFI = 1$ ,  $State = 0$  o 5: No se detecta ningún error en la trama de voz recibida, pero la trama de voz recibida anterior era mala. La ganancia de LTP y la ganancia del libro de códigos fijo están limitadas por debajo de los valores usados para la última subtrama buena recibida:

25

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

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

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

35 donde  $g_c$  = ganancia del libro de códigos fijo decodificada actual, y  $g_c(-1)$  = ganancia del libro de códigos fijo usada para la última subtrama buena ( $BFI = 0$ ).



**[0047]** El resto de los parámetros de voz recibidos se usan normalmente en la síntesis de voz. La trama actual de parámetros de voz se guarda.

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

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

donde  $g_p$  indica la ganancia de LTP decodificada actual y  $g_p(-1), \dots, g_p(-n)$  indican las ganancias de LTP usadas para las últimas  $n$  subtramas y  $median5()$  indica una operación de mediana de 5 puntos y  $P(state)$  = factor de atenuación, donde  $(P(1) = 0,98, P(2) = 0,98, P(3) = 0,8, P(4) = 0,3, P(5) = 0,2, P(6) = 0,2)$  y  $state$  = número de estado, y

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

donde  $g_c$  indica la ganancia del libro de códigos fijo decodificada actual y  $g_c(-1), \dots, g_c(-n)$  indican las ganancias de libro de códigos fijo usadas para las últimas  $n$  subtramas y  $median5()$  indica una operación de mediana de 5 puntos y  $C(state)$  = 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  $state$  = número de estado.

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

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

**[0051]** Respecto a CNG en AMR, de acuerdo con [3GP12a, sección 6.4], cada primera trama de SID perdida se sustituye usando la información de SID a partir de tramas de SID válidas recibidas previamente y se aplica el procedimiento para tramas de SID válidas. Para tramas de SID perdidas posteriores, se aplica una técnica de atenuación al ruido de confort que reducirá gradualmente el nivel de salida. Por lo tanto, se verifica si la última actualización de SID fue de hace más de 50 tramas (=1 s), en caso afirmativo, la salida se silenciará (atenuación de nivel en -6/8 dB por trama [3GP12d,  $dtx\_dec\{\}@sp\_dec.c$ ] lo que produce 37,5 dB por segundo). Nótese que el desvanecimiento aplicado a CNG se realiza en el dominio de LP.

**[0052]** En lo sucesivo, se considera AMR-WB. Multitasa adaptativa - WB [ITU03, 3GP09c] es un códec de voz, ACELP, basado en AMR (véase la sección 1.8). Usa extensión de ancho de banda paramétrica y también soporta DTX/CNG. En la descripción de la norma [3GP12g] hay soluciones ejemplares de ocultamiento dadas que son iguales que para AMR [3GP12a] con desviaciones menores. Por lo tanto, en este contexto solamente se describen las diferencias respecto a AMR. Para la descripción de la norma, véase la descripción anteriormente.

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

**[0054]** En el caso de trama perdida, la ganancia de altura tonal  $g_p$  para la primera subtrama es la misma que en la última trama buena, excepto que está limitada entre 0,95 y 0,5. Para la segunda, la tercera y las siguientes subtramas, la ganancia de altura tonal  $g_p$  disminuye en un factor de 0,95 y está, nuevamente, limitada.

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

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

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

$$g_{c_{inov}} = \frac{1,0}{\sqrt{\frac{ener_{inov}}{\text{subframe\_size}}}} \quad (16)$$

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

**[0056]** Para ocultar los retrasos de LTP, en AMR-WB, se usa el historial de los cinco últimos buenos retrasos de LTP y ganancias de LTP para descubrir el mejor procedimiento para actualizar, en el caso de una pérdida de trama. En caso de que la trama se reciba con errores de bits se realiza una predicción, de si el retraso de LTP recibido es utilizable o no [3GP12g].

**[0057]** Respecto a 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, AMR-WB propone aplicar una técnica de atenuación al ruido de confort que disminuirá gradualmente el nivel de salida. Por lo tanto, se verifica si la última actualización de SID fue hace más de 50 tramas (=1 s), en caso afirmativo, la salida se silenciará (atenuación de nivel en -3/8 dB por trama [3GP12f, dtx\_dec{}@dtx.c] que produce 18,75 dB por segundo). Nótese que el desvanecimiento aplicado a CNG se realiza en el dominio de LP.

**[0059]** Ahora, se considera AMR-WB+. Multitasa adaptativa - WB+ [3GP09a] es un códec conmutado usando ACELP y TCX (TCX = Excitación codificada por transformada) como códecs de núcleo. Usa extensión de ancho de banda paramétrica y también soporta DTX/CNG.

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

- Se define un modo de vector,  $(m_{-1}, m_0, m_1, m_2, m_3)$ , donde  $m_{-1}$  indica el modo de la última trama de la supertrama anterior y  $m_0, m_1, m_2, m_3$  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 tramas perdidas  $n_{loss}$  puede estar entre 0 y 4.

- 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 dado que entonces es seguro que una trama TCX80 estaba indicada dentro de la supertrama.

- Si solamente un indicador de las tramas 0 - 3 es tres (y el número de tramas perdidas  $n_{loss}$  es tres), el modo se ajustará a  $(1, 1, 1, 1)$ , dado que entonces 3/4 del espectro diana TCX80 se ha perdido y es muy probable que la

ganancia de TCX global se haya perdido.

5 - Si el modo está indicando (x, 2,-1, x, x) o (x,-1, 2, x, x), se extrapolará a (x, 2, 2, x, x), indicando 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), indicando también una trama TCX40. Debe observarse que (x, [0, 1], 2, 2, [0, 1]) son configuraciones inválidas.

- Después de eso, 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.

10 **[0061]** 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, el mismo enfoque que en [3GP12g] se aplica para esta trama (véase anteriormente).

15 **[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 por transformada):

20 - Si una trama completa está perdida, entonces se aplica un ocultamiento de tipo ACELP: la última excitación se repite y se usan coeficientes de ISF ocultos (ligeramente desplazados 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 de LPC (Codificación predictiva lineal).

25 - 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 ocultamiento en el dominio de FFT, utilizando extrapolación de fase y amplitud, teniendo en cuenta la última trama recibida correctamente. El enfoque de extrapolación de la información de fase no es de ningún interés en este contexto (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 ocultamiento de TCX consiste en las siguientes etapas [3GP09a, sección 6.5.1.2.3]:

30 - Se calcula el espectro de magnitud de la trama anterior:

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

35 - Se calcula el espectro de magnitud de la trama actual:

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

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

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

45 - La amplitud de los coeficientes espectrales que faltan se extrapola usando:

$$si \ (lost[k]) \quad A[k] = gain \cdot oldA[k]$$

50 - En cualquier otro caso de una trama perdida con  $m_k = [2, 3]$ , la diana 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 ganancia de TCX global). No se aplica desvanecimiento en este caso.

**[0063]** Respecto a CNG en AMR-WB+, se usa el mismo enfoque que e AMR-WB (véase anteriormente).

5

**[0064]** En lo sucesivo, se considera OPUS. OPUS [IET12] incorpora tecnología de dos códecs: el SILK orientado a la voz (conocido como códec Skype) y el CELT de baja latencia (CELT = transformada superpuesta de energía restringida). Opus puede ajustarse de manera continua entre tasas de bits altas y bajas, e internamente, conmuta entre un códec de predicción lineal a tasas de bits bajas (SILK) y un códec de transformada a tasas de bits más altas (CELT) así como un híbrido para una superposición corta.

10

**[0065]** Respecto a la compresión y descompresión de datos de audio de SILK, en OPUS, hay varios parámetros que se atenúan durante el ocultamiento en la rutina del decodificador SILK. El parámetro de ganancia de LTP se atenúa multiplicando todos los coeficientes de LPC por 0,99, 0,95 o 0,90 por trama, dependiendo del número de tramas perdidas consecutivas, donde la excitación se acumula usando el último ciclo de altura tonal desde la excitación de la trama anterior. El parámetro de retraso de altura tonal aumenta muy lentamente durante pérdidas consecutivas. Para pérdidas individuales, se mantiene constante en comparación con la última trama. Además, el parámetro de ganancia de excitación se atenúa exponencialmente con  $0,99^{\text{lostcnt}}$  por trama, de este modo el parámetro de ganancia de excitación es 0,99 para el primer parámetro de ganancia de excitación, de modo que el parámetro de ganancia de excitación es 0,992 para el segundo parámetro de ganancia de excitación, y así sucesivamente. La excitación se genera usando un generador de números aleatorios que está generando ruido blanco mediante desbordamiento variable. Además, los coeficientes de LPC se extrapolan/promedian basándose en el último conjunto de coeficientes recibido correctamente. Después de generar el vector de excitación atenuado, los coeficientes de LPC ocultos se usan en OPUS para sintetizar la señal de salida de dominio temporal.

15

20

25

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

30

**[0067]** La figura 5 ilustra el comportamiento de pérdida de ráfaga de CELT. En particular, la figura 5 representa un espectrograma (eje x: tiempo; eje y: frecuencia) de un segmento de voz oculto de CELT. El recuadro gris claro indica las 5 primeras tramas perdidas consecutivamente, donde se aplica el enfoque de PLC basado en altura tonal. Más allá de esto, se muestra el ocultamiento de tipo ruido. Debe observarse que la conmutación se realiza instantáneamente, y no transita suavemente.

35

**[0068]** Respecto al ocultamiento basado en altura tonal, en OPUS, el ocultamiento basado en altura tonal consiste en descubrir la periodicidad en la señal decodificada mediante autocorrelación repitiendo la forma de onda en ventanas (en el dominio de excitación usando análisis y síntesis de LPC) usando la desviación de altura tonal (retraso de altura tonal). La forma de onda en ventanas se superpone de tal manera que preserve la cancelación del solapamiento sucesivo "aliasing" del dominio temporal con la trama anterior y la siguiente trama [IET12]. Adicionalmente, un factor de desvanecimiento se deriva y se aplica mediante el siguiente código:

40

```
opus_val32 E1=1, E2=1;
int period;
if (pitch_index <= MAX_PERIOD/2) {
    period = pitch_index;
}
else {
    period = MAX_PERIOD/2;
}
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];
}
if (E1 > E2) {
    E1 = E2;
}
decay = sqrt(E1/E2);
attenuation = decay;
```

**[0069]** En este código, *exc* contiene la señal de excitación hasta muestras de *MAX\_PERIOD* antes de la pérdida.

5 **[0070]** La señal de excitación se multiplica más adelante con atenuación, a continuación, se sintetiza y es emitida mediante síntesis de LPC.

**[0071]** El algoritmo de desvanecimiento para el enfoque de dominio temporal puede resumirse de la siguiente manera:

10

- Descubrir la energía síncrona de altura tonal del último ciclo de altura tonal antes de la pérdida.

- Descubrir la energía síncrona de altura tonal del segundo último ciclo de altura tonal antes de la pérdida.

15

- Si la energía está aumentando, limitarla para que permanezca constante: atenuación = 1

- Si la energía está disminuyendo, continuar con la misma atenuación durante el ocultamiento.

**[0072]** 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, con el fin de simular ruido de fondo de confort.

**[0073]** Respecto al rastreo del nivel y la forma del ruido de fondo, en OPUS, la estimación del ruido de fondo se realiza de la siguiente manera: después del análisis de MDCT, se calcula la raíz cuadrada de las energías de banda de MDCT por banda de frecuencia, donde el agrupamiento de los "bins" de MDCT sigue la escala de Bark de acuerdo con [IET12, Tabla 55]. Entonces, la raíz cuadrada de las energías se transforma en el dominio  $\log_2$  mediante:

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

30

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

35 **[0074]** En OPUS, el ruido de fondo se registra en el lado del decodificador de la siguiente manera [IET12, `amp2Log2` y `log2Amp@quant_bands.c`]:

$$backgroundLogE[i] = \min(backgroundLogE[i] + 8 \cdot 0,001, bandLogE[i])$$

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

40 **[0075]** La energía mínima rastreada se determina básicamente mediante la raíz cuadrada de la energía de la banda de la trama actual, pero el aumento de una trama a la siguiente está limitado por 0,05 dB.

**[0076]** Respecto a la aplicación del nivel y la forma del ruido de fondo, de acuerdo con OPUS, si se aplica PLC de tipo ruido, se usa  $backgroundLogE$  tal como se deriva en la última trama buena y se convierte de vuelta al dominio lineal:

45

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

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

50

**[0077]** El procedimiento de ocultamiento actual es para llenar la trama de MDCT con ruido blanco producido

por un generador de números aleatorios, y convertir a escala este ruido blanco de manera que coincida en banda con la energía de *bandE*. Posteriormente, se aplica la MDCT inversa, lo que da como resultado una señal de dominio temporal. Después de la superposición, adición y desénfasis (como en decodificación regular) ésta se apaga.

5

**[0078]** En lo sucesivo, se considera MPEG-4 HE-AAC (MPEG = grupo de expertos en imágenes en movimiento; HE-AAC = codificación de audio avanzada de alta eficiencia). La codificación de audio avanzada de alta eficiencia consiste en un códec de audio basado en transformada (AAC), suplementado mediante una extensión de ancho de banda paramétrica (SBR).

10

**[0079]** Respecto a AAC (AAC = codificación de audio avanzada), el consorcio de DAB especifica para AAC en DAB+, un desvanecimiento a cero en el dominio de frecuencia [EBU10, sección A1.2] (DAB = difusión 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) están atenuados en un factor correspondiente a las características de desvanecimiento y a continuación se pasan al mapeo de frecuencia a tiempo. Dependiendo de la rampa de atenuación, el ocultamiento conmuta a silenciamiento después de un número de AU inválidas consecutivas, lo que significa que el espectro completo se ajustará a 0.

15

**[0080]** El consorcio de DRM (DRM = gestión de derechos digitales) especifica para AAC en DRM un desvanecimiento en el dominio de frecuencia [EBU12, sección 5.3.3]. El ocultamiento trabaja sobre los datos espectrales justo antes de la conversión final de frecuencia a tiempo. Si múltiples tramas están alteradas, el ocultamiento implementa en primer lugar un desvanecimiento basado en valores espectrales ligeramente modificados desde la última trama válida. Además, similar a DAB+, el comportamiento de desvanecimiento, por ejemplo, la rampa de atenuación, podría ser fija o ajustable por el usuario. Los coeficientes espectrales de la última trama se atenúan en un factor correspondiente a las características de desvanecimiento y a continuación se pasa al mapeo de frecuencia a tiempo. Dependiendo de la rampa de atenuación, el ocultamiento conmuta a silenciamiento después de un número de tramas inválidas consecutivas, lo que significa que el espectro completo se ajustará a 0.

25

**[0081]** 3GPP presentan para AAC en aacPlus mejorado el desvanecimiento en el dominio de frecuencia similar a DRM [3GP12e, sección 5.1]. El ocultamiento trabaja sobre los datos espectrales justo antes de la conversión final de frecuencia a tiempo. Si múltiples tramas están alteradas, el ocultamiento implementa en primer lugar un desvanecimiento basado en valores espectrales ligeramente modificados desde la última trama buena. Un desvanecimiento completo requiere 5 tramas. Los coeficientes espectrales desde la última trama buena se copian y se atenúan en un factor de:

35

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

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

40

**[0082]** Lauber y Sperschneider presentan para AAC un desvanecimiento por tramas del espectro MDCT, basado en extrapolación de energía [LS01, sección 4.4]. Podrían usarse formas de energía de un espectro anterior para extrapolar la forma de un espectro estimado. La extrapolación de energía puede realizarse independiente de las técnicas de ocultamiento como una clase de postocultamiento.

45

**[0083]** Respecto a AAC, el cálculo de energía se realiza sobre la base de una banda de factor de escala con el fin de acercarse a las bandas críticas del sistema auditivo humano. Los valores de energía individuales disminuyen trama a trama con el fin de reducir el volumen suavemente, por ejemplo, para desvanecer la señal. Esto se vuelve necesario dado que la probabilidad de que los valores estimados representen la señal actual, disminuye rápidamente con el tiempo.

50

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

55

**[0085]** Quackenbusch y Driesen sugieren para AAC un desvanecimiento por tramas a cero exponencial [QD03]. Se propone una repetición de un conjunto adyacente de coeficientes de tiempo/frecuencia, en la que cada repetición ha aumentado exponencialmente la atenuación, desvaneciéndose de este modo gradualmente a silencio

en el caso de apagones prolongados.

**[0086]** Respecto a SBR (SBR = replicación de banda espectral) in MPEG-4 HE-AAC, 3GPP sugiere SBR en aacPlus mejorada para almacenar temporalmente los datos de envoltura decodificados y, en el caso de una pérdida de trama, para reutilizar las energías almacenadas temporalmente de los datos de envoltura transmitidos y para disminuirlas en una relación constante de 3 dB para cada trama oculta. El resultado se introduce en el proceso de decodificación normal donde el ajustador de envoltura lo usa para calcular las ganancias, usadas para ajustar las bandas altas parcheadas creadas por el generador de HF. La decodificación de SBR tiene lugar a continuación de la manera habitual. Además, los valores de umbral mínimo de ruido y nivel del seno codificados en delta están siendo eliminados. Dado que queda disponible ninguna diferencia con la información previa, el umbral mínimo de ruido y los niveles del seno decodificados siguen siendo proporcionales a la energía de la señal generada de HF [3GP12e, sección 5.2].

**[0087]** El consorcio de DRM especificó para SBR junto con AAC 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]** En lo sucesivo, se consideran MPEG-4 CELP y MPEG-4 HVXC (HVXC = codificación de excitación de vector armónico). El consorcio de DRM especifica para SBR junto con CELP y HVXC [EBU12, sección 5.6.3.2] que el ocultamiento de requisito mínimo para SBR para los códecs de voz es aplicar un conjunto predeterminado de valores de datos, cada vez que se ha detectado una trama de SBR alterada. Esos valores dan una envoltura espectral de banda alta estática a un nivel de reproducción relativo bajo, mostrando una caída hacia las frecuencias más altas. El objetivo es simplemente garantizar que ninguna ráfaga de audio de mal comportamiento, potencialmente alta, alcance los oídos del oyente, por medio de la inserción de "ruido de confort" (en oposición al silenciamiento estricto). Esto no es, de hecho, desvanecimiento real sino en su lugar un salto a cierto nivel de energía con el fin de insertar alguna 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 desvaneciendo lentamente los niveles (L) hacia 0, análogamente al caso de AAC + SBR.

**[0090]** Ahora, se considera MPEG-4 HILN (HILN = armónicos y líneas individuales más ruido). Meine y col., presentan 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 sustituir parámetros codificados de forma diferencial alterados es mantener la frecuencia constante, reducir la amplitud en un factor de atenuación (por ejemplo, -6 dB), y dejar que la envoltura espectral converja hacia la de la característica de paso bajo promediada. Una alternativa para la envoltura espectral sería mantenerla inalterada. Con respecto a amplitudes y envolturas espectrales, componentes de ruido pueden tratarse de la misma manera que componentes armónicos.

**[0091]** En lo sucesivo, se considera el rastreo del nivel de ruido de fondo en la técnica anterior. Rangachari y Loizou [RL06] proporcionan una buena visión de conjunto de varios procedimientos y describen algunas de sus limitaciones. Procedimientos para rastrear el nivel de ruido de fondo son, por ejemplo, procedimiento de seguimiento mínimo [RL06] [Coh03] [SFB00] [Dob95], basado en VAD (VAD = detección de actividad de voz); filtración de Kalman [Gan05] [BJH06], descomposición subespacial [BP06] [HJH08]; Decisión programada [SS98] [MPC89] [HE95] y estadística mínima.

**[0092]** El enfoque de estadística mínima se seleccionó para usarlo dentro del alcance para USAC-2, (USAC = codificación de voz y audio unificada) y se perfila posteriormente con más detalle.

**[0093]** La estimación de densidad espectral de potencia del ruido basada en suavizado óptimo y estadística mínima [Mar01] presenta un estimador de ruido, que es capaz de funcionar independientemente de que la señal sea voz activa o ruido de fondo. En contraste con otros procedimientos, el algoritmo de estadística mínima no usa ningún umbral explícito para distinguir entre actividad de voz y pausa de voz y está, por lo tanto, más estrechamente relacionada con procedimientos de decisión programada que con los procedimientos de detección de actividad de voz tradicionales. Similar a los procedimientos de decisión programada, también puede actualizar la PSD (densidad espectral de potencia) del ruido estimada durante la actividad de voz.

**[0094]** El procedimiento de estadística mínima se basa en dos observaciones, concretamente que la voz y el ruido son habitualmente estadísticamente independientes y que la potencia de una señal de voz ruidosa

frecuentemente decae al nivel de potencia del ruido. Es posible, por lo tanto, derivar una estimación de PSD (PSD = densidad espectral de potencia) de ruido precisa rastreando el mínimo de la PSD de señal ruidosa. Dado que el mínimo es más pequeño que (o en otros casos igual a) el valor promedio, el procedimiento de rastreo mínimo requiere una compensación del sesgo.

5

**[0095]** El sesgo está en función de la varianza de la PSD de señal suavizada y, por lo tanto, depende del parámetro de suavizado del estimador de PSD. En contraste con un trabajo anterior sobre rastreo mínimo, que utiliza un parámetro de suavizado constante y una corrección del sesgo mínima constante, se usa un suavizado de PSD dependiente del tiempo y la frecuencia, lo que también requiere una compensación del sesgo dependiente del tiempo y la frecuencia.

10

**[0096]** El uso de rastreo mínimo proporciona una estimación aproximada de la potencia del ruido. Sin embargo, existen algunas desventajas. El suavizado con un parámetro de suavizado fijo amplía los picos de actividad de voz de la estimación de PSD suavizada. Esto causará estimaciones de ruido inexactas, dado que la ventana deslizante para la búsqueda mínima podría deslizarse hacia picos amplios. Por lo tanto, no pueden usarse parámetros de suavizado cercanos a uno, y, como consecuencia, la estimación de ruido tendrá una varianza relativamente grande. Además, la estimación de ruido está sesgada hacia valores más bajos. Además, en el caso de aumento de la potencia del ruido, el rastreo mínimo se retrasa.

15

**[0097]** El rastreo de la PSD de ruido basado en MMSE con baja complejidad [HHJ10] presenta un enfoque de PSD de ruido de fondo que utiliza una búsqueda de MMSE usada en un espectro de DFT (transformada de Fourier discreta). El algoritmo consiste en estas etapas de procesamiento:

20

- Se calcula el estimador de máxima probabilidad basándose en la PSD de ruido de la trama anterior.

25

- Se calcula el estimador de media cuadrática mínima.

- Se estima el estimador de máxima probabilidad usando el enfoque dirigido por decisión [EM84].

30

- Se calcula el factor de sesgo inverso suponiendo que los coeficientes de DFT de voz y ruido tienen una distribución Gaussiana.

- Se suaviza la densidad espectral de potencia del ruido estimada.

35

**[0098]** También hay un enfoque de red de seguridad aplicado con el fin de evitar un bloqueo completo del algoritmo.

**[0099]** El rastreo de ruido no estacionario basándose en estimación de ruido recursiva impulsada por datos [EH08] presenta un procedimiento para la estimación de la varianza espectral de ruido a partir de señales de voz contaminadas por fuentes de ruido altamente no estacionario. Este procedimiento también está usando suavizado en dirección de tiempo/frecuencia.

40

**[0100]** Un algoritmo de estimación de ruido de baja complejidad basado en el suavizado de estimación de potencia de ruido y corrección del sesgo de estimación [Yu09] mejora el enfoque presentado en [EH08]. La principal diferencia es, que la función de ganancia espectral para estimación de potencia de ruido se descubre mediante un procedimiento impulsado por datos iterativo.

45

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

50

**[0102]** El desvanecimiento es de particular interés para una pluralidad de códecs de voz y de 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 mejorada (véase [EBU10, EBU12, 3GP12e, LS01, QD03]) (incluyendo AAC y SBR), MPEG-4 HILN (véase [ISO09, MEP01]) y OPUS (véase [IET12]) (incluyendo SILK y CELT).

55



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

5 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 de SILK en OPUS; códecs que procesan además 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 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+.

10 **[0104]** Para códecs que transforman directamente la señal temporal en el dominio de frecuencia, el desvanecimiento se realiza en el dominio espectral/de subbanda. Esto es cierto para códecs que se basan en MDCT o una transformación similar, tales como AAC en MPEG-4 HE-AAC, G.719, G.722 (dominio de subbanda) y G.722.1.

15 **[0105]** Para códecs paramétricos, se aplica desvanecimiento en el dominio paramétrico. Esto es cierto para MPEG-4 HILN.

20 **[0106]** Respecto a la velocidad de desvanecimiento y la curva de desvanecimiento, un desvanecimiento se realiza comúnmente 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 de desvanecimiento y la curva de desvanecimiento. En la mayoría de los casos, el factor de atenuación se aplica por tramas, pero también se utiliza una aplicación por muestras véase, por ejemplo, G.718 y G.722.

25 **[0107]** El factor de atenuación para cierto segmento de señal podría proporcionarse de dos maneras, absoluta y relativa.

30 **[0108]** En el caso en que un factor de atenuación se proporciona de forma absoluta, el nivel de referencia es siempre el de la última trama recibida. Los factores de atenuación absolutos habitualmente comienzan con un valor cercano a 1 para el segmento de señal inmediatamente después de la última trama buena y a continuación se degradan más rápido o más lento hacia 0. La curva de desvanecimiento depende directamente de estos factores. Éste es, 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 lineal o gradualmente lineales. Considerando un factor de ganancia  $g(n)$ , mientras que  $g(0)$  representa el factor de ganancia de la última trama buena, un factor de atenuación absoluto  $\alpha_{abs}(n)$ , el factor de ganancia de cualquier trama perdida posterior puede derivarse como

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

40 **[0109]** En el caso en que un factor de atenuación se proporciona de forma relativa, el nivel de referencia es el de la trama anterior. Esto tiene ventajas en el caso de un procedimiento de ocultamiento recursivo, por ejemplo, si la señal ya atenuada es procesada adicionalmente y atenuada de nuevo.

45 **[0110]** Si un factor de atenuación se aplica de forma recursiva, entonces éste podría ser un valor fijo independiente del número de tramas perdidas consecutivamente, por ejemplo, 0,5 para G.719 (véase anteriormente); un valor fijo con respecto al número de tramas perdidas consecutivamente, por ejemplo, según lo propuesto para G.729 en [CPK08]: 1,0 para las dos primeras tramas, 0,9 para las dos tramas siguientes, 0,8 para las tramas 5 y 6, y 0 para todas las tramas posteriores (véase anteriormente); o un valor que es relativo al número de tramas perdidas consecutivamente y que depende de características de señal, por ejemplo, un desvanecimiento más rápido para una señal inestable y un desvanecimiento más lento para una señal estable, por ejemplo, G.718 (véase la sección anterior y [ITU08a, tabla 44]);

50 **[0111]** Suponiendo un factor de desvanecimiento relativo  $0 \leq \alpha_{rel}(n) \leq 1$ , mientras que  $n$  es el número de la trama perdida ( $n \geq 1$ ); el factor de ganancia de cualquier trama 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.

5 **[0112]** Respecto al procedimiento de desvanecimiento, habitualmente, se especifica el factor de atenuación, pero en algunas normas de aplicación (DRM, DAB+) este último se deja al fabricante.

10 **[0113]** Si diferentes partes de señal se desvanecen por separado, podrían aplicarse diferentes factores de atenuación, por ejemplo, para desvanecer componentes tonales con cierta velocidad y componentes de tipo ruido con otra velocidad (por ejemplo, AMR, SILK).

15 **[0114]** Habitualmente, se aplica cierta ganancia a toda la trama. Cuando se realiza el desvanecimiento en el dominio espectral, ésta es la única manera posible. Sin embargo, si el desvanecimiento se realiza en el dominio temporal o el dominio predictivo lineal, es posible un desvanecimiento más granular. Dicho desvanecimiento más granular se aplica en G.718, donde factores de ganancia individuales se derivan para cada muestra mediante interpolación lineal entre el factor de ganancia de la última trama y el factor de ganancia de la trama actual.

20 **[0115]** Para códecs con una duración de trama variable, un factor de atenuación relativo constante causa una velocidad de desvanecimiento diferente dependiendo de la duración de la trama. Éste es, por ejemplo, el caso para AAC, donde la duración de la trama depende de la tasa de muestreo.

25 **[0116]** Para adoptar la curva de desvanecimiento aplicada a la forma temporal de la última señal recibida, los factores de desvanecimiento (estáticos) podrían ajustarse adicionalmente. Dicho ajuste dinámico adicional se aplica, por ejemplo, para AMR donde la mediana de los cinco factores de ganancia previos se tiene en cuenta (véase [3GP12b] y sección 1.8.1). Antes de que se realice cualquier atenuación, la ganancia actual se ajusta a la mediana, si la mediana es más pequeña que la última ganancia, de lo contrario se usa la última ganancia. Además, dicho ajuste dinámico adicional se aplica, por ejemplo, para G729, donde la amplitud se predice usando 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 ocultas podría superar el factor de ganancia de la última trama recibida.

30 **[0117]** Respecto al nivel diana del desvanecimiento, con la excepción de G.718 y CELT, el nivel diana es 0 para todos los códecs analizados, incluyendo aquellos códecs de generación de ruido de confort (CNG).

35 **[0118]** En G.718, el desvanecimiento de la excitación de altura tonal (que representa componentes tonales) y el desvanecimiento de la excitación aleatoria (que representa componentes de tipo ruido) se realiza por separado. Aunque la ganancia de altura tonal factor se desvanece a cero, la ganancia de innovación factor se desvanece a la energía de excitación de CNG.

40 **[0119]** Suponiendo que se dan factores de atenuación relativos, esto conduce - basándose en la fórmula (23) - al siguiente factor de atenuación absoluto:

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

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

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

**[0121]** En CELT no hay ningún desvanecimiento hacia el nivel diana, pero después de 5 tramas de ocultamiento tonal (incluyendo un desvanecimiento) el nivel se conmuta instantáneamente al nivel diana en la 6ª trama perdida consecutivamente. El nivel se deriva por bandas usando la fórmula (19).

5

**[0122]** Respecto a la forma espectral diana del desvanecimiento, todos los códecs basados en transformada pura analizados (AAC, G.719, G.722, G.722.1) así como SBR simplemente prolongan la forma espectral de la última trama buena durante el desvanecimiento.

10 **[0123]** Diversos códecs de 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 de una media estática y una media a corto plazo (derivada promediando los últimos  $n$  conjuntos de coeficientes de LP) (LP = predicción lineal).

15 **[0124]** Todos los módulos de CNG en los códecs descritos AMR, AMR-WB, AMR-WB+, G.718 prolongan la forma espectral de la última trama buena durante el desvanecimiento.

**[0125]** Respecto al rastreo del nivel del ruido de fondo, hay cinco enfoques diferentes conocidos a partir de la bibliografía:

20

- Basado en detector de actividad de voz: basado en SNR/VAD, pero muy difícil de ajustar finamente y difícil de usar para voz de SNR baja.

25

- Esquema de decisión programada: el enfoque de decisión programada tiene en cuenta la probabilidad de presencia de voz [SS98] [MPC89] [HE95].

30

- Estadística mínima: el mínimo de la PSD se rastrea reteniendo cierta cantidad de valores a lo largo del tiempo en un búfer, permitiendo de este modo descubrir el ruido mínimo a partir de las muestras pasadas [Mar01] [HHJ10] [EH08] [Yu09].

35

- Filtración de Kalman: el algoritmo usa una serie de mediciones observadas a lo largo del tiempo, que contienen ruido (variaciones aleatorias), y produce estimaciones de la PSD del ruido que tiende a ser más precisa que las basadas en una única medición solamente. El filtro de Kalman opera de forma recursiva sobre flujos de datos de entrada ruidosos para producir una estimación estadísticamente óptima del estado del sistema [Gan05] [BJH06].

40

- Descomposición subespacial: este enfoque intenta descomponer una señal de tipo ruido en una señal de voz 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 de Fourier temporal discreta). A continuación, los vectores propios/valores propios pueden rastrearse usando un algoritmo de suavizado arbitrario [BP06] [HJH08].

Una disposición de ocultamiento de pérdida de paquetes de audio basada en interpolación se propone además en el documento EP 2 360 682 A1. El objetivo de la presente invención es proporcionar conceptos mejorados para sistemas de codificación de audio. El objetivo de la presente invención se resuelve mediante un aparato de acuerdo con la reivindicación 1, mediante un procedimiento de acuerdo con la reivindicación 19 y mediante un programa informático de acuerdo con la reivindicación 20.

45

**[0126]** Se proporciona un aparato para decodificar una señal de audio codificada para obtener una señal de audio reconstruida. El aparato comprende una interfaz receptora para recibir una o más tramas que comprenden información sobre una pluralidad de muestras de señal de audio de un espectro de señal de audio de la señal de audio codificada, y un procesador para generar la señal de audio reconstruida. El procesador está configurado para generar la señal de audio reconstruida desvaneciendo un espectro modificado a un espectro diana, 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 señal modificadas, en la que, para cada una de las muestras de señal 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 señal de audio. Además, el procesador está configurado para no desvanecer el espectro modificado al espectro diana, si la trama actual de las una o más tramas es recibida por la interfaz receptora y si la trama actual que es recibida por la interfaz receptora no está alterada.

50

55

[0127] De acuerdo con una realización, el espectro diana puede ser, por ejemplo, un espectro de tipo ruido.

[0128] En una realización, el espectro de tipo ruido puede representar, por ejemplo, ruido blanco.

[0129] De acuerdo con una realización, el espectro de tipo ruido puede estar, por ejemplo, conformado.

[0130] En una realización, la forma del espectro de tipo ruido puede depender, por ejemplo, de un espectro de señal de audio de una señal recibida previamente.

[0131] De acuerdo con una realización, el espectro de tipo ruido puede estar, por ejemplo, conformado dependiendo de la forma del espectro de señal de audio.

[0132] En una realización, el procesador puede emplear, por ejemplo, un factor de inclinación para conformar el espectro de tipo ruido.

[0133] De acuerdo con una realización, el procesador puede emplear, por ejemplo, la fórmula

$$\text{shaped\_noise}[i] = \text{noise} * \text{power}(\text{tilt\_factor}, i/N)$$

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 *power* es una función de potencia.

*power*(x, y) indica  $x^y$

*power*(*tilt\_factor*, *i/N*) indica  $\text{tilt\_factor}^{\frac{i}{N}}$

[0134] Si *tilt\_factor* es menor que 1 esto significa atenuación con i creciente. Si *tilt\_factor* es mayor que 1 significa amplificación con i creciente.

[0135] De acuerdo con otra realización, el procesador puede emplear, por ejemplo, la fórmula

$$\text{shaped\_noise}[i] = \text{noise} * (1 + i / (N-1) * (\text{tilt\_factor}-1))$$

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

[0136] Si el *tilt\_factor* es menor que 1 esto significa atenuación con i creciente. Si *tilt\_factor* es mayor que 1 significa amplificación con i creciente.

[0137] De acuerdo con una realización, el procesador puede estar configurado, por ejemplo, para generar el espectro modificado, cambiando un signo de una o más de las muestras de señal de audio del espectro de señal de audio, si la trama actual no es recibida por la interfaz receptora o si la trama actual que es recibida por la interfaz receptora está alterada.

[0138] En una realización, cada una de las muestras de señal de audio del espectro de señal de audio puede estar, por ejemplo, representada por un número real pero no por un número imaginario.

[0139] De acuerdo con una realización, las muestras de señal de audio del espectro de señal de audio pueden estar, por ejemplo, representadas en un dominio de transformada de coseno discreta modificada.

[0140] En otra realización, las muestras de señal de audio del espectro de señal de audio pueden estar, por ejemplo, representadas en un dominio de transformada de seno discreta modificada.

[0141] De acuerdo con una realización, el procesador puede estar configurado, por ejemplo, para generar el espectro modificado empleando una función de signo aleatorio que emite de forma aleatoria o pseudoaleatoria bien un primer o bien un segundo valor.

[0142] En una realización, el procesador puede estar, por ejemplo, configurado para desvanecer el espectro modificado al espectro diana disminuyendo posteriormente un factor de atenuación.

**[0143]** De acuerdo con una realización, el procesador puede estar, por ejemplo, configurado para desvanecer el espectro modificado al espectro diana aumentando posteriormente un factor de atenuación.

5 **[0144]** En una realización, si la trama actual no es recibida por la interfaz receptora o si la trama actual que es recibida por la interfaz receptora está alterada, el procesador puede estar configurado, por ejemplo, para generar la señal de audio reconstruida empleando la fórmula:

$$x[i] = (1 - \text{cum\_damping}) * \text{noise}[i] + \text{cum\_damping} * \text{random\_sign}() * x\_old[i]$$

10

en la que  $i$  es un índice, en la que  $x[i]$  indica una muestra de la señal de audio reconstruida, en la que  $\text{cum\_damping}$  es un factor de atenuación, en la que  $x\_old[i]$  indica una de las muestras de señal de audio del espectro de señal de audio de la señal de audio codificada, en la que  $\text{random\_sign}()$  retorna 1 o -1, y en la que  $\text{noise}$  es un vector aleatorio que indica el espectro diana.

15

**[0145]** En una realización, dicho vector aleatorio  $\text{noise}$  puede, por ejemplo, convertirse a escala de modo que su media cuadrática es similar a la media cuadrática del espectro de la señal de audio codificada que está compuesta por una de las tramas que son recibidas las últimas por la interfaz receptora.

20 **[0146]** De acuerdo con una realización general, el procesador puede estar configurado, por ejemplo, para generar la señal de audio reconstruida, empleando un vector aleatorio que es convertido a escala, de modo que su media cuadrática sea similar a la media cuadrática del espectro de la señal de audio codificada que está compuesta por una de las tramas que son recibidas las últimas por la interfaz receptora.

25 **[0147]** Además, se proporciona un procedimiento para decodificar una señal de audio codificada para obtener una señal de audio reconstruida. El procedimiento comprende:

- Recibir una o más tramas que comprenden información sobre una pluralidad de muestras de señal de audio de un espectro de señal de audio de la señal de audio codificada. Y:

30

- Generar la señal de audio reconstruida.

**[0148]** Generar la señal de audio reconstruida se lleva a cabo desvaneciendo un espectro modificado a un espectro diana, si una trama actual no es recibida o si la trama actual es recibida pero está alterada, en la que el espectro modificado comprende una pluralidad de muestras de señal modificadas, en la que, para cada una de las muestras de señal 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 señal de audio. El espectro modificado no se desvanece a un espectro de ruido blanco, si la trama actual de las una o más tramas es recibida y si la trama actual que es recibida no está alterada.

40

**[0149]** Además, se proporciona un programa informático para implementar el procedimiento descrito anteriormente cuando es ejecutado en un ordenador o un procesador de señales.

**[0150]** Ejemplos realizan un espectro de MDCT de desvanecimiento a ruido blanco antes de la aplicación de FDNS (FDNS = Sustitución de ruido en el dominio de frecuencia).

45

**[0151]** De acuerdo con la técnica anterior, en códecs basados en ACELP, el libro de códigos innovador se sustituye con un vector aleatorio (por ejemplo, con ruido). En ejemplos, el enfoque de ACELP, que consiste en sustituir el libro de códigos innovador con un vector aleatorio (por ejemplo, con ruido) se adopta para la estructura del decodificador de TCX. En este contexto, el equivalente del libro de códigos innovador es el espectro de MDCT habitualmente recibido dentro del flujo de bits e introducido en la FDNS.

50

**[0152]** El enfoque de ocultación de MDCT clásico sería simplemente repetir este espectro como es o aplicar cierto proceso de aleatorización, que básicamente prolonga la forma espectral de la última trama recibida [LS01]. Esto tiene la desventaja de que la forma espectral a corto plazo se prolonga, conduciendo frecuentemente a un sonido metálico repetitivo, que no es de tipo ruido de fondo, y, por lo tanto, no puede usarse como ruido de confort.

55

**[0153]** Usando el procedimiento propuesto la conformación espectral a corto plazo se realiza mediante la FDNS y la LTP de TCX, la conformación espectral a la larga se realiza mediante la FDNS solamente. La conformación por la FDNS se desvanece desde forma espectral a corto plazo a la forma espectral a largo plazo rastreada del ruido de fondo, y la LTP de TCX se desvanece a cero.

5

**[0154]** El desvanecimiento de los coeficientes de FDNS a coeficientes de ruido de fondo rastreados conduce a tener una transición suave entre la última envoltura espectral buena y la envoltura de fondo espectral que debe ser la diana a la larga, con el fin de conseguir un ruido de fondo agradable en el caso de pérdidas de trama de ráfaga larga.

10

**[0155]** En contraste, de acuerdo con el estado de la técnica, para códecs basados en transformada, se lleva a cabo ocultamiento de tipo ruido mediante repetición de tramas o sustitución de ruido en el dominio de frecuencia [LS01]. En la técnica anterior, la sustitución de ruido se realiza habitualmente mediante mezcla de signos de los "bins" espectrales. Si en la técnica anterior se usa mezcla de signos de TCX (dominio de frecuencia) durante el ocultamiento, los últimos coeficientes de MDCT recibidos se reutilizan y cada signo se aleatoriza antes de que el espectro se transforme de forma inversa en el dominio temporal. La desventaja de este procedimiento de la técnica anterior es que, para tramas perdidas consecutivamente, se usa el mismo espectro una y otra vez, justo con diferentes aleatorizaciones de signo y atenuación global. Cuando se observa la envoltura espectral a lo largo del tiempo en una cuadrícula temporal grosera, puede verse que la envoltura es aproximadamente constante durante pérdida de tramas consecutivas, dado que las energías de banda se mantienen constante unas con respecto a otras dentro de una trama y solamente son atenuadas globalmente. En el sistema de codificación usado, de acuerdo con la técnica anterior, los valores espectrales se procesan usando FDNS, con el fin de restaurar el espectro original. Esto significa que, si se desea desvanecer el espectro MDCT a cierta envoltura espectral (usando coeficientes de FDNS, por ejemplo, que describen el ruido de fondo actual), el resultado es no solamente dependiente de los coeficientes de FDNS, sino también dependiente del espectro decodificado previamente que se sometió a mezcla de signos. Los ejemplos mencionados anteriormente superan estas desventajas de la técnica anterior.

**[0156]** Los ejemplos se basan en el descubrimiento de que es necesario desvanecer el espectro usado para la mezcla de signos a ruido blanco, antes de introducirlo en el procesamiento de FDNS. En caso contrario, el espectro emitido no coincidirá nunca con la envoltura diana usada para procesamiento por FDNS.

**[0157]** En ejemplos, se usa la misma velocidad de desvanecimiento para desvanecimiento de la ganancia de LTP que para el desvanecimiento de ruido blanco.

35 **[0158]** Además, se proporciona un aparato para decodificar una señal de audio.

**[0159]** El aparato comprende una interfaz receptora. La interfaz receptora está configurada para recibir una pluralidad de tramas, en la 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 la 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.

**[0160]** Además, el aparato comprende una unidad de transformada para transformar la segunda porción de señal de audio o un valor o señal derivada de la segunda porción de señal de audio a partir 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 a o diferente del primer dominio.

50 **[0161]** Además, el aparato comprende una unidad de rastreo del nivel de ruido, en la que la unidad de rastreo del nivel de ruido está configurada para recibir una información de primera porción de señal que está 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. La unidad de rastreo del nivel de ruido está configurada para recibir la segunda porción de señal que está representada en el dominio de rastreo, y en la que la unidad de rastreo del nivel de ruido está configurada para determinar información sobre el nivel de ruido dependiendo de la información de primera porción de señal que está representada en el dominio de rastreo y dependiendo de la información de segunda porción de señal que está representada en el dominio de rastreo.

**[0162]** Además, el aparato comprende una unidad de reconstrucción para reconstruir una tercera porción de

señal de audio de la señal de audio dependiendo de la información sobre el nivel de ruido, si una tercera trama de la pluralidad de tramas no es recibida por la interfaz receptora, pero está alterada.

5 **[0163]** Una señal de audio puede ser, por ejemplo, una señal de voz, o una señal de música, o señal que comprende voz y música, etc.

10 **[0164]** 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 la primera porción de señal de audio, o que la información de primera porción de señal se ha obtenido/generado dependiendo de la primera porción de señal de audio o, 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.

15 **[0165]** Del mismo modo, 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 la segunda porción de señal de audio, o que la información de segunda porción de señal se ha obtenido/generado 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.

20 **[0166]** En un ejemplo, la primera porción de señal de audio puede estar, por ejemplo, representada en un dominio temporal como el primer dominio. Además, la unidad de transformada puede estar, por ejemplo, configurada para transformar la segunda porción de señal de audio o el valor derivado de la segunda porción de señal de audio desde un dominio de excitación que es el segundo dominio hasta el dominio temporal que es el dominio de rastreo. Además, la unidad de rastreo del nivel de ruido puede estar, por ejemplo, configurada para recibir la información de primera porción de señal que está representada en el dominio temporal como el dominio de rastreo. Además, la unidad de rastreo del nivel de ruido puede estar, por ejemplo, configurada para recibir la segunda porción de señal que está representada en el dominio temporal como el dominio de rastreo.

30 **[0167]** De acuerdo con un ejemplo, la primera porción de señal de audio puede estar, por ejemplo, representada en un dominio de excitación como el primer dominio. Además, la unidad de transformada puede estar, por ejemplo, configurada para transformar la segunda porción de señal de audio o el valor derivado de la segunda porción de señal de audio desde un dominio temporal que es el segundo dominio hasta el dominio de excitación que es el dominio de rastreo. Además, la unidad de rastreo del nivel de ruido puede estar, por ejemplo, configurada para recibir la información de primera porción de señal que está representada en el dominio de excitación como el dominio de rastreo. Además, la unidad de rastreo del nivel de ruido puede estar, por ejemplo, configurada para recibir la segunda porción de señal que está representada en el dominio de excitación como el dominio de rastreo.

40 **[0168]** En un ejemplo, la primera porción de señal de audio puede estar, por ejemplo, representada en un dominio de excitación como el primer dominio, en la que la unidad de rastreo del nivel de ruido puede estar, por ejemplo, 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 de 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 puede estar, por ejemplo, configurada para transformar la segunda porción de señal de audio o el valor derivado de la segunda porción de señal de audio desde un dominio temporal que es el segundo dominio hasta un dominio de FFT que es el dominio de rastreo, y en la que la unidad de rastreo del nivel de ruido puede estar, por ejemplo, configurada para recibir la segunda porción de señal de audio que está representada en el dominio de FFT.

50 **[0169]** En un ejemplo, el aparato puede comprender, por ejemplo, además una primera unidad de agregación para determinar un primer valor agregado dependiendo de la primera porción de señal de audio. Además, el aparato puede comprender, por ejemplo, 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. Además, la unidad de rastreo del nivel de ruido puede estar, por ejemplo, configurada para recibir el primer valor agregado como la información de primera porción de señal que está representada en el dominio de rastreo, en la que la unidad de rastreo del nivel de ruido puede estar, por ejemplo, configurada para recibir el segundo valor agregado como la información de segunda porción de señal que está representada en el dominio de rastreo, y en la que la unidad de rastreo del nivel de ruido puede estar, por ejemplo, configurada para determinar información sobre el nivel de ruido dependiendo del primer valor agregado que está representado en el dominio de

rastreo y dependiendo del segundo valor agregado que está representado en el dominio de rastreo.

- [0170]** De acuerdo con un ejemplo, la primera unidad de agregación puede estar, por ejemplo, configurada para determinar el primer valor agregado de modo que el primer valor agregado indica una media cuadrática de la primera porción de señal de audio o de una señal derivada de la primera porción de señal de audio. Además, la segunda unidad de agregación puede estar, por ejemplo, configurada para determinar el segundo valor agregado de modo que el segundo valor agregado indica a media cuadrática de la segunda porción de señal de audio o de una señal derivada de la segunda porción de señal de audio.
- 10 **[0171]** En un ejemplo, la unidad de transformada puede estar, por ejemplo, configurada para transformar el valor derivado de la segunda porción de señal de audio desde el segundo dominio hasta el dominio de rastreo aplicando un valor de ganancia sobre el valor derivado de la segunda porción de señal de audio. De acuerdo con ejemplos, el valor de ganancia puede indicar, por ejemplo, una ganancia introducida por síntesis de codificación predictiva lineal, o el valor de ganancia puede indicar, por ejemplo, una ganancia introducida por síntesis y desénfasis de codificación predictiva lineal.
- 15 **[0172]** En un ejemplo, la unidad de rastreo del nivel de ruido puede estar, por ejemplo, configurada para determinar información sobre el nivel de ruido aplicando un enfoque de estadística mínima.
- 20 **[0173]** De acuerdo con un ejemplo, la unidad de rastreo del nivel de ruido puede estar, por ejemplo, configurada para determinar un nivel de ruido de confort como la información sobre el nivel de ruido. La unidad de reconstrucción puede estar, por ejemplo, configurada para reconstruir la tercera porción de señal de audio dependiendo de la información sobre el 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.
- 25 **[0174]** En un ejemplo, la unidad de rastreo del nivel de ruido puede estar, por ejemplo, configurada para determinar un nivel de ruido de confort como la información sobre el nivel de ruido derivada de un espectro del nivel de ruido, en la que dicho espectro del nivel de ruido se obtiene aplicando el enfoque de estadística mínima. La unidad de reconstrucción puede estar, por ejemplo, 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.
- 30 **[0175]** De acuerdo con otro ejemplo, la unidad de rastreo del nivel de ruido puede estar, por ejemplo, configurada para determinar una pluralidad de coeficientes predictivos lineales que indican un nivel de ruido de confort como la información sobre el nivel de ruido, y la unidad de reconstrucción puede estar, por ejemplo, configurada para reconstruir la tercera porción de señal de audio dependiendo de la pluralidad de coeficientes predictivos lineales.
- 35 **[0176]** En un ejemplo, la unidad de rastreo del nivel de ruido está configurada para determinar una pluralidad de coeficientes de FFT que indican un nivel de ruido de confort como la información sobre el 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 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.
- 40 **[0177]** En un ejemplo, la unidad de reconstrucción puede estar, por ejemplo, configurada para reconstruir la tercera porción de señal de audio dependiendo de la información sobre el 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.
- 45 **[0178]** De acuerdo con un ejemplo, la unidad de reconstrucción puede estar, por ejemplo, configurada para reconstruir la tercera porción de señal de audio atenuando o amplificando una señal derivada de la primera o la segunda porción de señal de audio.
- 50 **[0179]** En un ejemplo, el aparato puede comprender, por ejemplo, además una unidad de predicción a largo plazo que comprende un búfer de retardo. Además, la unidad de predicción a largo plazo puede estar configurada, por ejemplo, para generar una señal procesada dependiendo de la primera o la segunda porción de señal de audio, dependiendo de una entrada del búfer de retardo que se almacena en el búfer de retardo y dependiendo de una
- 55



ganancia de predicción a largo plazo. Además, la unidad de predicción a largo plazo puede estar, por ejemplo, configurada para desvanecer la ganancia de predicción a largo plazo hacia cero, si dicha tercera trama de la pluralidad de tramas no es recibida por la interfaz receptora o si dicha tercera trama es recibida por la interfaz receptora, pero está alterada.

5

**[0180]** De acuerdo con un ejemplo, la unidad de predicción a largo plazo puede estar, por ejemplo, configurada para desvanecer la ganancia de predicción a largo plazo hacia cero, en la que una velocidad con la que la ganancia de predicción a largo plazo se desvanece a cero depende de un factor de desvanecimiento.

10 **[0181]** En un ejemplo, la unidad de predicción a largo plazo puede estar, por ejemplo, configurada para actualizar la entrada del búfer de retardo almacenando 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.

15 **[0182]** De acuerdo con un ejemplo, la unidad de transformada puede ser, por ejemplo, una primera unidad de transformada, y la unidad de reconstrucción es una primera unidad de reconstrucción. El aparato comprende además una segunda unidad de transformada y una segunda unidad de reconstrucción. La segunda unidad de transformada puede estar, por ejemplo, configurada para transformar la información sobre el nivel de ruido desde el dominio de rastreo hasta el 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 puede estar, por ejemplo, configurada para reconstruir una cuarta porción de señal de audio de la señal de audio dependiendo de la información sobre el nivel de ruido que está 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.

20

25 **[0183]** En un ejemplo, la segunda unidad de reconstrucción puede estar, por ejemplo, configurada para reconstruir la cuarta porción de señal de audio dependiendo de la información sobre el nivel de ruido y dependiendo de la segunda porción de señal de audio.

30 **[0184]** De acuerdo con un ejemplo, la segunda unidad de reconstrucción puede estar, por ejemplo, configurada para reconstruir la cuarta porción de señal de audio atenuando o amplificando una señal derivada de la primera o la segunda porción de señal de audio.

**[0185]** Además, se proporciona un procedimiento para decodificar una señal de audio.

35

**[0186]** El procedimiento comprende:

40 - 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.

45 - Transformar la segunda porción de señal de audio o un valor o señal derivada de la segunda porción de señal de audio desde un segundo dominio hasta un dominio de rastreo para obtener una información de segunda porción de señal, en la que el segundo dominio es diferente del primer dominio, en la que el dominio de rastreo es diferente del segundo dominio, y en el que el dominio de rastreo es igual a o diferente del primer dominio.

50 - Determinar información sobre el nivel de ruido dependiendo de información de primera porción de señal, que está representada en el dominio de rastreo, y dependiendo de la información de segunda porción de señal que está 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. Y:

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

**[0187]** Además, se proporciona un programa informático para implementar el procedimiento descrito

anteriormente cuando es ejecutado en un ordenador o un procesador de señales.

**[0188]** Algunos de los ejemplos proporcionan un parámetro de suavizado que varía con el tiempo de modo que las capacidades de rastreo del periodograma suavizado y su varianza estén mejor equilibradas, para desarrollar un algoritmo para compensación de sesgo, y para acelerar el rastreo de ruido en general.

**[0189]** Los ejemplos se basan en el descubrimiento de que, con respecto al desvanecimiento, los siguientes parámetros son de interés: el dominio de desvanecimiento; la velocidad de desvanecimiento, o, más general, curva de desvanecimiento; el nivel diana del desvanecimiento; la forma espectral diana del desvanecimiento; y/o el rastreo del nivel de ruido de fondo. En este contexto, los ejemplos se basan en el descubrimiento de que la técnica anterior presenta desventajas significativas.

**[0190]** Se proporciona un aparato y un procedimiento para desvanecimiento de señales mejorado para sistemas de codificación de audio conmutados durante el ocultamiento de errores.

**[0191]** Además, se proporciona un programa informático para implementar el procedimiento descrito anteriormente cuando es ejecutado en un ordenador o un procesador de señales.

**[0192]** Los ejemplos realizan un desvanecimiento al nivel de ruido de confort. De acuerdo con ejemplos, se realiza un rastreo del nivel de ruido de confort común en el dominio de excitación. El nivel de ruido de confort que es la diana durante la pérdida de paquetes en ráfagas será el mismo, independientemente del codificador de núcleo (ACELP/TCX) en uso, y siempre está actualizado. No existe técnica anterior conocida, donde un rastreo del nivel de ruido común es necesario. Los ejemplos proporcionan el desvanecimiento de un códec conmutado a una señal de tipo ruido de confort durante pérdidas de paquetes en ráfagas.

**[0193]** Además, los ejemplos realizan que la complejidad global será más baja en comparación con tener dos módulos de rastreo del nivel de ruido independientes, dado que las funciones (PROM) y memoria pueden compartirse.

**[0194]** En los ejemplos, la derivación de nivel en el dominio de excitación (en comparación con la derivación de nivel en el dominio temporal) proporciona más mínimos durante la voz activa, dado que parte de la información de voz está cubierta por los coeficientes de LP.

**[0195]** En el caso de ACELP, de acuerdo con ejemplos, la derivación de nivel tiene lugar en el dominio de excitación. En el caso de TCX, en los ejemplos, el nivel se deriva en el dominio temporal, y la ganancia de la síntesis y desénfasis de LPC se aplica como un factor de corrección con el fin de modelar el nivel de energía en el dominio de excitación. Rastrear el nivel en el dominio de excitación, por ejemplo, antes de la FDNS, teóricamente también sería posible, pero la compensación de nivel entre el dominio de excitación de TCX y el dominio de excitación de ACELP se considera bastante compleja.

**[0196]** Ninguna técnica anterior incorpora dicho rastreo de nivel de fondo común en diferentes dominios. Las técnicas de la técnica anterior no tienen dicho rastreo del nivel de ruido de confort común, por ejemplo, en el dominio de excitación, en un sistema de códec conmutado. Por lo tanto, los ejemplos son ventajosos respecto a la técnica anterior, puesto que para las técnicas de la técnica anterior, el nivel de ruido de confort que es diana durante pérdidas de paquetes en ráfagas puede ser diferente, dependiendo del modo de codificación precedente (ACELP/TCX), donde el nivel fue rastreado; como en la técnica anterior, siendo el rastreo independiente para cada modo de codificación causará complejidad computacional adicional y sobrecarga innecesaria; y dado que en la técnica anterior, podría no haber disponible ningún nivel de ruido de confort actualizado en ningún núcleo debido a reciente conmutación a este núcleo.

**[0197]** De acuerdo con algunos ejemplos, el rastreo de nivel se lleva a cabo en el dominio de excitación, pero el desvanecimiento de TCX se lleva a cabo en el dominio temporal. Mediante desvanecimiento en el dominio temporal, se evitan fallos de la TDAC, que causarían solapamiento sucesivo. Esto se vuelve de particular interés cuando se ocultan componentes de señal tonal. Además, se evita la conversión de nivel entre el dominio de excitación de ACELP y el dominio espectral de MDCT y de este modo, por ejemplo, se ahorran recursos de computación. Debido a la conmutación entre el dominio de excitación y el dominio temporal, se requiere un ajuste de nivel entre el dominio de excitación y el dominio temporal. Esto se resuelve mediante la derivación de la ganancia que sería introducida por la síntesis de LPC y el preénfasis y para usar esta ganancia como un factor de corrección para convertir el nivel entre los dos dominios.

**[0198]** En contraste, las técnicas de la técnica anterior no llevan a cabo rastreo de nivel en el dominio de excitación y desvanecimiento de TCX en el dominio temporal. Respecto a códecs basados en transformada del estado de la técnica, el factor de atenuación se aplica bien en el dominio de excitación (para enfoques de ocultamiento de tipo dominio temporal/ACELP, véase [3GP09a]) o bien en el dominio de frecuencia (para enfoques de dominio de frecuencia como repetición de tramas o sustitución de ruido, véase [LS01]). Una desventaja del enfoque de la técnica anterior para aplicar el factor de atenuación en el dominio de frecuencia es que se causará solapamiento sucesivo en la región de superposición y adición en el dominio temporal. Éste será el caso para tramas adyacentes a las que se aplican factores de atenuación diferentes, dado que el proceso de desvanecimiento hace que la TDAC (cancelación del solapamiento sucesivo en el dominio temporal) falle. Esto es particularmente relevante cuando se ocultan componentes de señal tonal. Los ejemplos mencionados anteriormente son, por lo tanto, ventajosos respecto a la técnica anterior.

**[0199]** Los ejemplos compensan la influencia del filtro de paso alto sobre la ganancia de la síntesis de LPC. De acuerdo con ejemplos, para compensar el cambio de ganancia no deseado del análisis y énfasis de LPC causados por la excitación sin voz filtrada por paso alto, se deriva un factor de corrección. Este factor de corrección tiene en cuenta este cambio de ganancia no deseado y modifica el nivel de ruido de confort diana en el dominio de excitación de modo que se alcance el nivel diana correcto en el dominio temporal.

**[0200]** En contraste, la técnica anterior, por ejemplo, G.718 [ITU08a], presenta un filtro de paso alto en la trayectoria de señal de la excitación sin voz, tal como se representa en la figura 2, si la señal de la última trama buena no se clasificó como SIN VOZ. Mediante esto, las técnicas de la técnica anterior causan efectos secundarios no deseados, dado que la ganancia de la síntesis de LPC posterior depende de las características de señal, que son alteradas por este filtro de paso alto. Dado que el nivel de fondo es rastreado y aplicado en el dominio de excitación, el algoritmo depende de la ganancia de la síntesis de LPC, que, a su vez, depende e nuevo de las características de la señal de excitación. En otras palabras: la modificación de las características de señal de la excitación debido a la filtración de paso alto, tal como se lleva a cabo por la técnica anterior, podría conducir a una ganancia modificada (habitualmente reducida) de la síntesis de LPC. Esto causa un nivel de salida erróneo incluso aunque el nivel de excitación sea correcto.

**[0201]** Los ejemplos superan estas desventajas de la técnica anterior.

**[0202]** En particular, ejemplos realizan una forma espectral adaptativa de ruido de confort. En contraste con G.718, rastreando la forma espectral del ruido de fondo, y aplicando (desvaneciendo a) esta forma durante las pérdidas de paquetes en ráfagas, la característica de ruido del ruido de fondo precedente coincidirá, causando una característica de ruido agradable del ruido de confort. Esto evita no coincidencias demasiado evidentes de la forma espectral que pueden introducirse usando una envoltura espectral que se derivó mediante capacitación fuera de línea y/o la forma espectral de las últimas tramas recibidas.

**[0203]** Además, se proporciona un aparato para decodificar una señal de audio. El aparato comprende una interfaz receptora, en la que la interfaz receptora está configurada para recibir una primera trama que comprende una primera porción de señal de audio de la señal de audio, y 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.

**[0204]** Además, el aparato comprende una unidad de rastreo del nivel de ruido, en la que la unidad de rastreo del nivel de ruido está configurada para determinar información sobre el nivel de ruido dependiendo de al menos una de la primera porción de señal de audio y la segunda porción de señal de audio (esto significa: dependiendo de la primera porción de señal de audio y/o la segunda porción de señal de audio), en la que la información sobre el nivel de ruido está representada en un dominio de rastreo.

**[0205]** Además, el aparato comprende una primera unidad de reconstrucción para reconstruir, en un primer dominio de reconstrucción, una tercera porción de señal de audio de la señal de audio dependiendo de la información sobre el 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 la que el primer dominio de reconstrucción es diferente de o igual al dominio de rastreo.

**[0206]** Además, el aparato comprende una unidad de transformada para transformar la información sobre el nivel de ruido desde el dominio de rastreo hasta 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 la que el segundo dominio de reconstrucción es diferente del dominio de rastreo, y en la que el segundo dominio de reconstrucción es diferente del primer dominio de reconstrucción, y

- [0207]** Además, el aparato comprende una segunda unidad de reconstrucción para reconstruir, en el segundo dominio de reconstrucción, una cuarta porción de señal de audio de la señal de audio dependiendo de la información sobre el nivel de ruido que está representada en el segundo dominio de reconstrucción, si dicha cuarta trama de la pluralidad de tramas no es recibida por la interfaz receptora o si dicha cuarta trama es recibida por la interfaz receptora, pero está alterada.
- 10 **[0208]** De acuerdo con algunos ejemplos, el dominio de rastreo puede ser, por ejemplo, aquel en el que el dominio de rastreo es un dominio temporal, un dominio espectral, un dominio de FFT, un dominio de MDCT o un dominio de excitación. El primer dominio de reconstrucción puede ser, por ejemplo, el dominio temporal, el dominio espectral, el dominio de FFT, el dominio de MDCT o el dominio de excitación. El segundo dominio de reconstrucción puede ser, por ejemplo, el dominio temporal, el dominio espectral, el dominio de FFT, el dominio de MDCT o el
- 15 dominio de excitación.
- [0209]** En un ejemplo, el dominio de rastreo puede ser, por ejemplo, el dominio de FFT, el primer dominio de reconstrucción puede ser, por ejemplo, el dominio temporal, y el segundo dominio de reconstrucción puede ser, por ejemplo, el dominio de excitación.
- 20 **[0210]** En otro ejemplo, el dominio de rastreo puede ser, por ejemplo, el dominio temporal, el primer dominio de reconstrucción puede ser, por ejemplo, el dominio temporal, y el segundo dominio de reconstrucción puede ser, por ejemplo, el dominio de excitación.
- 25 **[0211]** De acuerdo con un ejemplo, dicha primera porción de señal de audio puede estar, por ejemplo, representada en un primer dominio de entrada, y dicha segunda porción de señal de audio puede estar, por ejemplo, representada en un segundo dominio de entrada. La unidad de transformada puede ser, por ejemplo, una segunda unidad de transformada. El aparato puede comprender, por ejemplo, además una primera unidad de transformada para transformar la segunda porción de señal de audio o un valor o señal derivada de la segunda porción de señal
- 30 de audio desde el segundo dominio de entrada hasta el dominio de rastreo para obtener una información de segunda porción de señal. La unidad de rastreo del nivel de ruido puede estar, por ejemplo, configurada para recibir una información de primera porción de señal que está 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 del nivel de ruido está configurada para recibir la segunda porción de señal que está representada en el
- 35 dominio de rastreo, y en la que la unidad de rastreo del nivel de ruido está configurada para determinar la información sobre el nivel de ruido dependiendo de la información de primera porción de señal que está representada en el dominio de rastreo y dependiendo de la información de segunda porción de señal que está representada en el dominio de rastreo.
- 40 **[0212]** De acuerdo con un ejemplo, el primer dominio de entrada puede ser, por ejemplo, el dominio de excitación, y el segundo dominio de entrada puede ser, por ejemplo, el dominio de MDCT.
- [0213]** En otro ejemplo, el primer dominio de entrada puede ser, por ejemplo, el dominio de MDCT, y en el que el segundo dominio de entrada puede ser, por ejemplo, el dominio de MDCT.
- 45 **[0214]** De acuerdo con un ejemplo, la primera unidad de reconstrucción puede estar, por ejemplo, configurada para reconstruir la tercera porción de señal de audio llevando a cabo un primer desvanecimiento a un espectro de tipo ruido. La segunda unidad de reconstrucción puede estar, por ejemplo, configurada para reconstruir la cuarta porción de señal de audio llevando a cabo un segundo desvanecimiento a un espectro de tipo ruido y/o un
- 50 segundo desvanecimiento de una ganancia de LTP. Además, la primera unidad de reconstrucción y la segunda unidad de reconstrucción puede estar, por ejemplo, configurada para llevar a cabo 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.
- 55 **[0215]** En un ejemplo, el aparato puede comprender, por ejemplo, además una primera unidad de agregación para determinar un primer valor agregado dependiendo de la primera porción de señal de audio. Además, el aparato puede comprender, por ejemplo, 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 del nivel de ruido puede estar, por ejemplo, configurada para recibir el primer

valor agregado como la información de primera porción de señal que está representada en el dominio de rastreo, en la que la unidad de rastreo del nivel de ruido puede estar, por ejemplo, configurada para recibir el segundo valor agregado como la información de segunda porción de señal que está representada en el dominio de rastreo, y en la que la unidad de rastreo del nivel de ruido está configurada para determinar la información sobre el nivel de ruido dependiendo del primer valor agregado que está representado en el dominio de rastreo y dependiendo del segundo valor agregado que está representado en el dominio de rastreo.

10 **[0216]** De acuerdo con un ejemplo, la primera unidad de agregación puede estar, por ejemplo, configurada para determinar el primer valor agregado de modo que el primer valor agregado indica una media cuadrática de la primera porción de señal de audio o de una señal derivada de la primera porción de señal de audio. La segunda unidad de agregación está configurada para determinar el segundo valor agregado de modo que el segundo valor agregado indica una media cuadrática de la segunda porción de señal de audio o de una señal derivada de la segunda porción de señal de audio.

15 **[0217]** En un ejemplo, la primera unidad de transformada puede estar, por ejemplo, configurada para transformar el valor derivado de la segunda porción de señal de audio desde el segundo dominio de entrada hasta el dominio de rastreo aplicando un valor de ganancia sobre el valor derivado de la segunda porción de señal de audio.

20 **[0218]** De acuerdo con un ejemplo, el valor de ganancia puede indicar, por ejemplo, una ganancia introducida por síntesis de codificación predictiva lineal, o en el que el valor de ganancia indica una ganancia introducida por síntesis y desénfasis de codificación predictiva lineal.

25 **[0219]** En un ejemplo, la unidad de rastreo del nivel de ruido puede estar, por ejemplo, configurada para determinar la información sobre el nivel de ruido aplicando un enfoque de estadística mínima.

30 **[0220]** De acuerdo con un ejemplo, la unidad de rastreo del nivel de ruido puede estar, por ejemplo, configurada para determinar un nivel de ruido de confort como la información sobre el nivel de ruido. La unidad de reconstrucción puede estar, por ejemplo, configurada para reconstruir la tercera porción de señal de audio dependiendo de la información sobre el 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.

35 **[0221]** En un ejemplo, la unidad de rastreo del nivel de ruido puede estar, por ejemplo, configurada para determinar un nivel de ruido de confort como la información sobre el nivel de ruido derivada de un espectro del nivel de ruido, en la que dicho espectro del nivel de ruido se obtiene aplicando el enfoque de estadística mínima. La unidad de reconstrucción puede estar, por ejemplo, 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.

40 **[0222]** De acuerdo con un ejemplo, la primera unidad de reconstrucción puede estar, por ejemplo, configurada para reconstruir la tercera porción de señal de audio dependiendo de la información sobre el 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.

45 **[0223]** En un ejemplo, la primera unidad de reconstrucción puede estar, por ejemplo, configurada para reconstruir la tercera porción de señal de audio atenuando o amplificando la primera porción de señal de audio.

50 **[0224]** De acuerdo con un ejemplo, la segunda unidad de reconstrucción puede estar, por ejemplo, configurada para reconstruir la cuarta porción de señal de audio dependiendo de la información sobre el nivel de ruido y dependiendo de la segunda porción de señal de audio.

**[0225]** En un ejemplo, la segunda unidad de reconstrucción puede estar, por ejemplo, configurada para reconstruir la cuarta porción de señal de audio atenuando o amplificando la segunda porción de señal de audio.

55 **[0226]** De acuerdo con un ejemplo, el aparato puede comprender, por ejemplo, además una unidad de predicción a largo plazo que comprende un búfer de retardo, en la que la unidad de predicción a largo plazo puede estar, por ejemplo, configurada para generar una señal procesada dependiendo de la primera o la segunda porción de señal de audio, dependiendo de una entrada del búfer de retardo que se almacena en el búfer de retardo y dependiendo de una ganancia de predicción a largo plazo, y en la que la unidad de predicción a largo plazo está

configurada para desvanecer la ganancia de predicción a largo plazo hacia cero, si dicha tercera trama de la pluralidad de tramas no es recibida por la interfaz receptora o si dicha tercera trama es recibida por la interfaz receptora pero está alterada.

5 **[0227]** En un ejemplo, la unidad de predicción a largo plazo puede estar, por ejemplo, configurada para desvanecer la ganancia de predicción a largo plazo hacia cero, en la que una velocidad con la que la ganancia de predicción a largo plazo se desvanece a cero depende de un factor de desvanecimiento.

10 **[0228]** En un ejemplo, la unidad de predicción a largo plazo puede estar, por ejemplo, configurada para actualizar la entrada del búfer de retardo almacenando 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.

15 **[0229]** Además, se proporciona un procedimiento para decodificar una señal de audio. El procedimiento comprende:

- Recibir una primera trama que comprende una primera porción de señal de audio de la señal de audio, y recibir una segunda trama que comprende una segunda porción de señal de audio de la señal de audio.

20 - Determinar información sobre el 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 la que la información sobre el nivel de ruido está representada en un dominio de rastreo.

25 - Reconstruir, en un primer dominio de reconstrucción, una tercera porción de señal de audio de la señal de audio dependiendo de la información sobre el 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, en la que el primer dominio de reconstrucción es diferente de o igual al dominio de rastreo.

30 - Transformar la información sobre el nivel de ruido desde el dominio de rastreo hasta 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 la 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:

35 - Reconstruir, en el segundo dominio de reconstrucción, una cuarta porción de señal de audio de la señal de audio dependiendo de la información sobre el nivel de ruido que está 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.

40 **[0230]** Además, se proporciona un programa informático para implementar el procedimiento descrito anteriormente cuando es ejecutado en un ordenador o un procesador de señales.

45 **[0231]** Además, se proporciona un aparato para decodificar una señal de audio codificada para obtener una señal de audio reconstruida. El aparato comprende una interfaz receptora para recibir una o más tramas, un generador de coeficientes, y un reconstructor de señales. El generador de coeficientes está configurado para determinar, si una trama actual de las una o más tramas es recibida por la interfaz receptora y si la trama actual que es recibida por la interfaz receptora no está alterada, uno o más primeros coeficientes de señal de audio, que están compuestos 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 está configurado para generar uno o más segundos coeficientes de señal de audio, dependiendo de los uno o más primeros coeficientes de señal de audio y dependiendo de los uno o más coeficientes de ruido, si la trama actual no es recibida por la interfaz receptora o si la trama actual que es recibida por la interfaz receptora está alterada. El reconstructor de señales de audio está configurado para reconstruir una primera porción de la señal de audio reconstruida dependiendo de los uno o más primeros coeficientes de señal de audio, si la trama actual es recibida por la interfaz receptora y si la trama actual que es recibida por la interfaz receptora no está alterada. Además, el reconstructor de señales de audio está configurado para reconstruir una segunda porción de la señal de audio reconstruida dependiendo de los 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 que es recibida por la interfaz receptora está alterada.

**[0232]** En algunos ejemplos, los uno o más primeros coeficientes de señal de audio pueden ser, por ejemplo, uno o más coeficientes de filtro predictivo lineal de la señal de audio codificada. En algunos ejemplos, los uno o más primeros coeficientes de señal de audio pueden ser, por ejemplo, uno o más coeficientes de filtro predictivo lineal de la señal de audio codificada.

5

**[0233]** De acuerdo con un ejemplo, los uno o más coeficientes de ruido pueden ser, por ejemplo, uno o más coeficientes de filtro predictivo lineal que indican el ruido de fondo de la señal de audio codificada. En un ejemplo, los uno o más coeficientes de filtro predictivo lineal pueden representar, por ejemplo, una forma espectral del ruido de fondo.

10

**[0234]** En un ejemplo, el generador de coeficientes puede estar, por ejemplo, configurado para determinar las una o más segundas porciones de señal de audio de modo que las 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 modo que los uno o más primeros coeficientes de señal de audio sean uno o más pares espectrales de inmitancia de la señal de audio reconstruida.

15

**[0235]** De acuerdo con un ejemplo, el generador de coeficientes puede estar configurado, por ejemplo, para generar los uno o más segundos coeficientes de señal de audio aplicando la fórmula:

$$f_{current}[i] = \alpha \cdot f_{last}[i] + (1 - \alpha) \cdot pt_{mean}[i]$$

20

en la que  $f_{current}[i]$  indica uno de los uno o más segundos coeficientes de señal de audio, en la que  $f_{last}[i]$  indica uno de los uno o más primeros coeficientes de señal de audio, en la que  $pt_{mean}[i]$  es uno de los uno o más coeficientes de ruido, en la que  $\alpha$  es un número real con  $0 \leq \alpha \leq 1$ , y en la que  $i$  es un índice. En un ejemplo,  $0 < \alpha < 1$ .

25

**[0236]** De acuerdo con un ejemplo,  $f_{last}[i]$  indica un coeficiente de filtro predictivo lineal de la señal de audio codificada, y en la que  $f_{current}[i]$  indica un coeficiente de filtro predictivo lineal de la señal de audio reconstruida.

**[0237]** En un ejemplo,  $pt_{mean}[i]$  puede indicar, por ejemplo, el ruido de fondo de la señal de audio codificada.

30

**[0238]** En un ejemplo, el generador de coeficientes puede estar, por ejemplo, configurado para determinar, si la trama actual de las una o más tramas es recibida por la interfaz receptora y si la trama actual que es recibida por la interfaz receptora no está alterada, los uno o más coeficientes de ruido determinando un espectro de ruido de la señal de audio codificada.

35

**[0239]** De acuerdo con un ejemplo, el generador de coeficientes puede estar, por ejemplo, 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 calculando los coeficientes de LPC que representan la forma del ruido de fondo a partir del espectro de ruido de fondo.

40

**[0240]** Además, se proporciona un procedimiento para decodificar una señal de audio codificada para obtener una señal de audio reconstruida. El procedimiento comprende:

- Recibir una o más tramas.

45

- Determinar, si una trama actual de las una o más tramas es recibida y si la trama actual que es recibida no está alterada, uno o más primeros coeficientes de señal de audio, que están compuestos por la trama actual, en los que 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.

50

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

55

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

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

**[0241]** Además, se proporciona un programa informático para implementar el procedimiento descrito anteriormente cuando es ejecutado en un ordenador o un procesador de señales.

**[0242]** El tener medios comunes para rastrear y aplicar la forma espectral de ruido de confort durante el desvanecimiento presenta varias desventajas. Rastrear y aplicar la forma espectral de modo que pueda hacerse de forma similar para ambos códecs permite un enfoque común sencillo. CELT enseña solamente el rastreo por bandas de energías en el dominio espectral y la formación por bandas de la forma espectral en el dominio espectral, lo cual no es posible para el núcleo de CELP.

**[0243]** En contraste, en la técnica anterior, la forma espectral del ruido de confort introducida durante pérdidas de ráfagas es bien completamente estática, o bien parcialmente estática y parcialmente adaptativa a la media a corto plazo de la forma espectral (tal como se constata en G.718 [ITU08a]), y habitualmente no coincidirá con el ruido de fondo en la señal antes de la pérdida de paquetes. Esta no coincidencia de las características del ruido de confort podría ser molesta. De acuerdo con la técnica anterior, puede emplearse una forma de ruido de fondo capacitada (estática) fuera de línea que puede ser de sonido agradable para señales particulares, pero menos agradable para otras, por ejemplo, el ruido de un coche suena totalmente diferente del ruido de oficina.

**[0244]** Además, en la técnica anterior, puede emplearse una adaptación a la media a corto plazo de la forma espectral de las tramas recibidas previamente que podría acercar las características de señal a la señal recibida antes, pero no necesariamente a las características del ruido de fondo. En la técnica anterior, el rastreo de la forma espectral por bandas en el dominio espectral (tal como se constata en CELT [IET12]) no es aplicable para un códec conmutado usando no solamente un núcleo basado en dominio de MDCT (TCX) sino también un núcleo basado en ACELP. Los ejemplos mencionados anteriormente son, por lo tanto, ventajoso respecto a la técnica anterior.

**[0245]** Además, se proporciona un aparato para decodificar una señal de audio codificada para obtener una señal de audio reconstruida. El aparato comprende una interfaz receptora para recibir una pluralidad de tramas, un búfer de retardo para almacenar muestras de señal de audio de la señal de audio decodificada, un selector de muestras para seleccionar una pluralidad de muestras de señal de audio seleccionadas a partir de las muestras de señal de audio que se almacenan en el búfer de retardo, y un procesador de muestras para procesar las muestras de señal de audio seleccionadas para obtener muestras de 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 que es recibida por la interfaz receptora no está alterada, la pluralidad de muestras de señal de audio seleccionadas a partir de las muestras de señal de audio que se almacenan en el búfer de retardo dependiendo de una información de retraso de altura tonal que está compuesta 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 que es recibida por la interfaz receptora está alterada, la pluralidad de muestras de señal de audio seleccionadas a partir de las muestras de señal de audio que se almacenan en el búfer de retardo dependiendo de una información de retraso de altura tonal que está compuesta por otra trama que es recibida previamente por la interfaz receptora.

**[0246]** De acuerdo con un ejemplo, el procesador de muestras puede estar, por ejemplo, configurada para obtener las muestras de señal de audio reconstruida, si la trama actual es recibida por la interfaz receptora y si la trama actual que es recibida por la interfaz receptora no está alterada, sustituyendo la escala de las muestras de señal de audio seleccionadas dependiendo de la información de ganancia que está compuesta por la trama actual. Además, el selector de muestras puede estar, por ejemplo, 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 que es recibida por la interfaz receptora está alterada, sustituyendo la escala de las muestras de señal de audio seleccionadas dependiendo de la información de ganancia que están compuestas por dicha otra trama que es recibida previamente por la interfaz receptora.

**[0247]** En un ejemplo, el procesador de muestras puede estar, por ejemplo, configurada para obtener las muestras de señal de audio reconstruida, si la trama actual es recibida por la interfaz receptora y si la trama actual que es recibida por la interfaz receptora no está alterada, multiplicando las muestras de señal de audio seleccionadas y un valor que depende de la información de ganancia que está compuesto 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 que es recibida por la interfaz receptora está



alterada, multiplicando las muestras de señal de audio seleccionadas y un valor que depende de la información de ganancia que está compuesta por dicha otra trama que es recibida previamente por la interfaz receptora.

5 **[0248]** De acuerdo con un ejemplo, el procesador de muestras puede estar, por ejemplo, configurado para almacenar las muestras de señal de audio reconstruida en el búfer de retardo.

10 **[0249]** En un ejemplo, el procesador de muestras puede estar, por ejemplo, configurado para almacenar las muestras de señal de audio reconstruida en el búfer de retardo antes de que una trama adicional sea recibida por la interfaz receptora.

15 **[0250]** De acuerdo con un ejemplo, el procesador de muestras puede estar, por ejemplo, configurado para almacenar las muestras de señal de audio reconstruida en el búfer de retardo después de que una trama adicional sea recibida por la interfaz receptora.

20 **[0251]** En un ejemplo, el procesador de muestras puede estar, por ejemplo, configurado para sustituir la escala de las muestras de señal de audio seleccionadas dependiendo de la información de ganancia para obtener muestras de señal de audio con la escala sustituida y combinando las muestras de señal de audio con la escala sustituida con muestras de señal de audio de entrada para obtener las muestras de señal de audio procesadas.

25 **[0252]** De acuerdo con un ejemplo, el procesador de muestras puede estar, por ejemplo, configurado para almacenar las muestras de señal de audio procesadas, que indican la combinación de las muestras de señal de audio con la escala sustituida y las muestras de señal de audio de entrada, en el búfer de retardo, y para no almacenar las muestras de señal de audio con la escala sustituida en el búfer de retardo, si la trama actual es recibida por la interfaz receptora y si la trama actual que es recibida por la interfaz receptora no está alterada.

30 Además, el procesador de muestras está configurado para almacenar las muestras de señal de audio con la escala sustituida 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 que es recibida por la interfaz receptora está alterada.

35 **[0253]** De acuerdo con otro ejemplo, el procesador de muestras puede estar, por ejemplo, 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 que es recibida por la interfaz receptora está alterada.

40 **[0254]** En un ejemplo, el selector de muestras puede estar, por ejemplo, configurado para obtener las muestras de señal de audio reconstruida sustituyendo la escala de las muestras de señal de audio seleccionadas dependiendo de una ganancia modificada, en el que la ganancia modificada se define de acuerdo con la fórmula:

$$\text{gain} = \text{gain\_past} * \text{damping};$$

45 en la que *gain* es la ganancia modificada, en la que el selector de muestras puede estar, por ejemplo, configurado para ajustar *gain\_past* a la ganancia después de *gain* y ha sido calculado, y en la que *damping* es un valor real.

50 **[0255]** De acuerdo con un ejemplo, el selector de muestras puede estar, por ejemplo, configurado para calcular la ganancia modificada.

55 **[0256]** En un ejemplo, el amortiguamiento puede definirse, por ejemplo, de acuerdo con:  $0 \leq \text{damping} \leq 1$ .

**[0257]** De acuerdo con un ejemplo, la ganancia modificada puede ajustarse, por ejemplo, a cero, si al menos un número predeterminado de tramas no ha sido recibido por la interfaz receptora desde que una trama ha sido recibida la última por la interfaz receptora.

**[0258]** Además, se proporciona un procedimiento para decodificar una señal de audio codificada para obtener una señal de audio reconstruida. El procedimiento comprende:

- 55 - Recibir una pluralidad de tramas.
- Almacenar muestras de señal de audio de la señal de audio decodificada.
- Seleccionar una pluralidad de muestras de señal de audio seleccionadas a partir de las muestras de señal de

audio que se almacenan en el búfer de retardo. Y:

- Procesar las muestras de señal de audio seleccionadas para obtener muestras de señal de audio reconstruida de la señal de audio reconstruida.

5

**[0259]** Si una trama actual es recibida y si la trama actual que es recibida no está alterada, se lleva a cabo la etapa de seleccionar la pluralidad de muestras de señal de audio seleccionadas a partir de las muestras de señal de audio que se almacenan en el búfer de retardo dependiendo de una información de retraso de altura tonal que está compuesta por la trama actual. Además, si la trama actual no es recibida o si la trama actual que es recibida está alterada, se lleva a cabo la etapa de seleccionar la pluralidad de muestras de señal de audio seleccionadas a partir de las muestras de señal de audio que se almacenan en el búfer de retardo dependiendo de una información de retraso de altura tonal que está compuesta por otra trama que es recibida previamente por la interfaz receptora.

10

**[0260]** Además, se proporciona un programa informático para implementar el procedimiento descrito anteriormente cuando es ejecutado en un ordenador o un procesador de señales.

15

**[0261]** Los ejemplos emplean LTP de TCX (LTP de TCX = predicción a largo plazo de excitación codificada por transformada). Durante el funcionamiento normal, la memoria de LTP de TCX se actualiza con la señal sintetizada, que contiene ruido y componentes tonales reconstruidos.

20

**[0262]** En lugar de inhabilitar la LTP de TCX durante el ocultamiento, su funcionamiento normal puede continuar durante el ocultamiento con los parámetros recibidos en la última trama buena. Esto preserva la forma espectral de la señal, particularmente aquellos componentes tonales que son modelados por el filtro de LTP.

25

**[0263]** Además, los ejemplos desacoplan el bucle de retroalimentación de LTP de TCX. Una simple continuación del funcionamiento de LTP de TCX normal introduce ruido adicional, dado que, con cada etapa de actualización, se introduce ruido generado aleatoriamente adicional a partir de la excitación de LTP. Los componentes tonales se distorsionan, por lo tanto, cada vez más a lo largo del tiempo por el ruido añadido.

30

**[0264]** Para superar esto, solamente se puede retroalimentar el búfer de LTP de TCX (sin añadir ruido), con el fin de no contaminar la información tonal con ruido aleatorio no deseado.

**[0265]** Además, de acuerdo con ejemplos, la ganancia de LTP de TCX se desvanece a cero.

35

**[0266]** Estos ejemplos se basan en el descubrimiento de que continuar la LTP de TCX ayuda a preservar las características de señal a corto plazo, pero presenta desventajas a largo plazo: la señal transmitida durante el ocultamiento incluirá la información de voz/tonal que estaba presente precediendo a la pérdida. Especialmente para voz limpia o voz sobre ruido de fondo, es extremadamente improbable que un tono o armónico decaiga muy lentamente durante un periodo muy prolongado. Continuando el funcionamiento de LTP de TCX durante el ocultamiento, particularmente si la actualización de la memoria de LTP está desacoplada (sólo se retroalimentan componentes tonales y no la parte con mezcla de signos), la información de voz/tonal permanecerá presente en la señal oculta para toda la pérdida, que es atenuada solamente por el desvanecimiento global al nivel de ruido de confort. Además, es imposible alcanzar la envoltura de ruido de confort durante pérdidas de paquetes en ráfagas, si la LTP de TCX se aplica durante la pérdida de ráfagas sin ser atenuada a lo largo del tiempo, dado que la señal incorporará siempre a continuación la información de voz de la LTP.

40

45

**[0267]** Por lo tanto, la ganancia de LTP de TCX se desvanece hacia cero, de modo que los componentes tonales representados por la LTP se desvanecerán a cero, al mismo tiempo la señal se desvanece al nivel y la forma de la señal de fondo, y de modo que el desvanecimiento alcance la envoltura de fondo espectral (ruido de confort) deseada sin incorporar componentes tonales no deseados.

50

**[0268]** En los ejemplos, se usa la misma velocidad de desvanecimiento para el desvanecimiento de la ganancia de LTP que para el desvanecimiento de ruido blanco.

55

**[0269]** En contraste, en la técnica anterior, no hay ningún códec de transformada conocido que usa LTP durante el ocultamiento. Para la LTP de MPEG-4 [ISO09] no existen enfoques de ocultamiento en la técnica anterior. Otro códec basado en MDCT de la técnica anterior que utiliza una LTP es CELT, pero este códec usa un ocultamiento de tipo ACELP para las cinco primeras tramas, y para todas las tramas posteriores se genera ruido de fondo, que no utiliza la LTP. Una desventaja de la técnica anterior de no usar la LTP de TCX es que todos los

componentes tonales que están siendo modelados con la LTP desaparecen abruptamente. Además, en códecs basados en ACELP de la técnica anterior, el funcionamiento de LTP se prolonga durante el ocultamiento, y la ganancia del libro de códigos adaptativo se desvanece hacia cero. Con respecto al funcionamiento del bucle de retroalimentación, la técnica anterior emplea dos enfoques, bien la excitación completa, por ejemplo, la suma de la 5 excitación innovadora y la adaptativa, se retroalimenta (AMR-WB); o solamente la excitación adaptativa actualizada, por ejemplo, las partes de señal tonal, se retroalimenta (G.718). Los ejemplos mencionados anteriormente superan las desventajas de la técnica anterior.

**[0270]** En lo sucesivo, se describen ejemplos de la presente invención con más detalle con referencia a las 10 figuras, en las que:

La figura 1a ilustra un aparato para decodificar una señal de audio de acuerdo con una realización,

15 La figura 1b ilustra un aparato para decodificar una señal de audio de acuerdo con otra realización,

La figura 1c ilustra un aparato para decodificar una señal de audio de acuerdo con otra realización, en el que el aparato comprende además una primera y una segunda unidad de agregación,

20 La figura 1d ilustra un aparato para decodificar una señal de audio de acuerdo con una realización adicional, en la que el aparato comprende además una unidad de predicción a largo plazo que comprende un búfer de retardo,

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

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

La figura 4 muestra un enfoque para predicción de amplitud usando regresión lineal,

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

La figura 6 muestra un rastreo del nivel del ruido de fondo de acuerdo con una realización en el decodificador durante un modo de funcionamiento libre de errores,

35 La figura 7 ilustra la derivación de ganancia de la síntesis y el desénfasis de LPC de acuerdo con una realización,

La figura 8 representa la aplicación de nivel de ruido de confort durante la pérdida de paquetes de acuerdo con una realización,

40 La figura 9 ilustra la compensación de ganancia de paso alto avanzada durante ocultamiento de ACELP de acuerdo con una realización,

La figura 10 representa el desacoplamiento del bucle de retroalimentación de LTP durante el ocultamiento de acuerdo con una realización,

45 La figura 11 ilustra un aparato para decodificar una señal de audio codificada para obtener una señal de audio reconstruida de acuerdo con una realización,

La figura 12 muestra un aparato para decodificar una señal de audio codificada para obtener una señal de audio reconstruida de acuerdo con otra realización, y

50 La figura 13 ilustra un aparato para decodificar una señal de audio codificada para obtener una señal de audio reconstruida una realización adicional, y

55 La figura 14 ilustra un aparato para decodificar una señal de audio codificada para obtener una señal de audio reconstruida de acuerdo con otra realización.

**[0271]** La figura 1a ilustra un aparato para decodificar una señal de audio de acuerdo con una realización.

**[0272]** El aparato comprende una interfaz receptora 110. La interfaz receptora está configurada para recibir

una pluralidad de tramas, en la 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.

**[0273]** Además, el aparato comprende una unidad de transformada 120 para transformar la segunda porción de señal de audio o un valor o señal derivada de la segunda porción de señal de audio desde un segundo dominio hasta un dominio de rastreo para obtener una información de segunda porción de señal, en la que el segundo dominio es diferente del primer dominio, en la que el dominio de rastreo es diferente del segundo dominio, y en la que el dominio de rastreo es igual a o diferente del primer dominio.

**[0274]** Además, el aparato comprende una unidad de rastreo del nivel de ruido 130, en la que la unidad de rastreo del nivel de ruido está configurada para recibir una información de primera porción de señal que está 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 del nivel de ruido está configurada para recibir la segunda porción de señal que está representada en el dominio de rastreo, y en la que la unidad de rastreo del nivel de ruido está configurada para determinar información sobre el nivel de ruido dependiendo de la información de primera porción de señal que está representada en el dominio de rastreo y dependiendo de la información de segunda porción de señal que está representada en el dominio de rastreo.

**[0275]** Además, el aparato comprende una unidad de reconstrucción para reconstruir una tercera porción de señal de audio de la señal de audio dependiendo de la información sobre el nivel de ruido, si una tercera trama de la pluralidad de tramas no es recibida por la interfaz receptora, pero está alterada.

**[0276]** 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 puede, por ejemplo, alimentarse a una o más unidades de procesamiento (no mostradas) para generar una o más señales de altavoz para uno o más altavoces, de modo que la información de sonido recibida compuesta por la primera y/o la segunda porción de señal de audio pueda ser reproducida.

**[0277]** Además, sin embargo, la primera y segunda porción de señal de audio también se usan para ocultamiento, por ejemplo, en caso de que tramas posteriores no lleguen al receptor o en caso de que tramas posteriores sean erróneas.

**[0278]** Entre otras cosas, la presente invención se basa en el descubrimiento de que el rastreo del nivel de ruido debe llevarse a cabo en un dominio común, denominado en el presente documento "dominio de rastreo". El dominio de rastreo, puede ser, por ejemplo, un dominio de excitación, por ejemplo, el dominio en el que la señal está representada por los LPC (LPC = coeficiente predictivo lineal) o por los ISP (ISP = par espectral de inmitancia) tal como se describe en AMR-WB y AMR-WB+ (véase [3GP12a], [3GP12b], [3GP09a], [3GP09b], [3GP09c]). El rastreo del nivel de ruido en un único dominio presenta, entre otras, la ventaja de que se evitan efectos de solapamiento sucesivo cuando la señal conmuta entre una primera representación en un primer dominio y una segunda representación en un segundo dominio (por ejemplo, cuando la representación de señal conmuta desde ACELP a TCX o viceversa).

**[0279]** Respecto a la unidad de transformada 120, que se transforma en bien la propia segunda porción de señal de audio o bien 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).

**[0280]** Respecto a la primera porción de señal de audio, en algunas realizaciones, la primera porción de señal de audio puede procesarse y/o transformarse en el dominio de rastreo.

**[0281]** En otras realizaciones, sin embargo, la primera porción de señal de audio puede estar ya representada en el dominio de rastreo.

**[0282]** 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.

- [0283]** Ahora, en primer lugar, se considera un desvanecimiento a un nivel de ruido de confort con más detalle.
- 5 **[0284]** El enfoque de desvanecimiento descrito puede implementarse, por ejemplo, en una versión de bajo retardo de xHE-AAC [NMR+12] (xHE-AAC = AAC de alta eficiencia extendida), que es capaz de conmutar de forma ininterrumpida entre codificación de ACELP (voz) y MDCT (música/ruido) trata a trama.
- 10 **[0285]** Respecto al rastreo de nivel común en un dominio de rastreo, por ejemplo, un dominio de excitación, para aplicar un desvanecimiento suave a un nivel de ruido de confort apropiado durante la pérdida de paquetes, es necesario identificar dicho nivel de ruido de confort durante el proceso de decodificación normal. Puede suponerse, por ejemplo, que un nivel de ruido similar al ruido de fondo es el más confortable. Por lo tanto, el nivel de ruido de fondo puede derivarse y actualizarse constantemente durante la decodificación normal.
- 15 **[0286]** La presente invención se basa en el descubrimiento de que cuando se tiene un códec de núcleo conmutado (por ejemplo, ACELP y TCX), considerar un nivel del ruido de fondo común independiente del codificador de núcleo seleccionado es particularmente adecuado.
- 20 **[0287]** La figura 6 representa un rastreo del nivel del ruido de fondo de acuerdo con una realización preferida en el decodificador durante el modo de funcionamiento libre de errores, por ejemplo, durante la decodificación normal.
- 25 **[0288]** El propio rastreo puede realizarse, por ejemplo, usando el enfoque de estadística mínima (véase [Mar01]).
- 30 **[0289]** Este nivel del ruido de fondo rastreado puede considerarse, por ejemplo, como la información sobre el nivel de ruido mencionada anteriormente.
- 35 **[0290]** Por ejemplo, la estimación de ruido por 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), no. 5, 504-512" [Mar01] puede emplearse para rastreo del nivel del ruido de fondo.
- 40 **[0291]** De manera correspondiente, en algunas realizaciones, la unidad de rastreo del nivel de ruido 130 está configurada para determinar información sobre el nivel de ruido aplicando un enfoque de estadística mínima, por ejemplo, empleando la estimación de ruido por estadística mínima de [Mar01].
- 45 **[0292]** Posteriormente, se describen algunas consideraciones y detalles de este enfoque de rastreo.
- 50 **[0293]** 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 rastrear componentes tonales de primer plano que son extraídos por LPC. Por ejemplo, el relleno de ruido de ACELP también puede emplear el nivel de ruido de fondo en el dominio de excitación. Con rastreo en el dominio de excitación, solamente un único rastreo del nivel de ruido de fondo puede servir para dos fines, lo que ahorra complejidad computacional. En una realización preferida, el rastreo se realiza en el dominio de excitación de ACELP.
- 55 **[0294]** La figura 7 ilustra derivación de ganancia de síntesis y desénfasis de LPC de acuerdo con una realización.
- 60 **[0295]** Respecto a la derivación de nivel, la derivación de nivel puede llevarse a cabo, por ejemplo, en el dominio temporal 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.
- 65 **[0296]** 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.
- 70 **[0297]** Para TCX, puede ser necesaria, por ejemplo, una compensación de ganancia para ajustar el nivel derivado al dominio de excitación de ACELP.

**[0298]** En la realización preferida, la derivación de nivel para TCX tiene lugar en el dominio temporal. Para este enfoque se descubrió una compensación de ganancia manejable: la ganancia introducida por síntesis y desénfasis de LPC se deriva tal como se muestra en la figura 7 y el nivel derivado se divide por esta ganancia.

5 **[0299]** 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.

**[0300]** Por lo tanto, volviendo a la figura 1a, en algunas realizaciones, la primera porción de señal de audio se representa en un dominio temporal 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 desde un dominio de excitación que es el segundo dominio hasta el dominio temporal que es el dominio de rastreo. En dichas realizaciones, la unidad de rastreo del nivel de ruido 130 está configurada para recibir la información de primera porción de señal que está representada en el dominio temporal como el dominio de rastreo. Además, la  
10  
15 unidad de rastreo del nivel de ruido 130 está configurada para recibir la segunda porción de señal que está representada en el dominio temporal como el dominio de rastreo.

**[0301]** En otras realizaciones, la primera porción de señal de audio se representa en un dominio de excitación como el primer dominio. La unidad de transformada 120 está configurada para transformar la segunda porción de señal de audio o el valor derivado de la segunda porción de señal de audio desde un dominio temporal que es el segundo dominio hasta el dominio de excitación que es el dominio de rastreo. En dichas realizaciones, la unidad de rastreo del nivel de ruido 130 está configurada para recibir la información de primera porción de señal que está representada en el dominio de excitación como el dominio de rastreo. Además, la unidad de rastreo del nivel de ruido 130 está configurada para recibir la segunda porción de señal que está representada en el dominio de  
20  
25 excitación como el dominio de rastreo.

**[0302]** En una realización, la primera porción de señal de audio puede estar, por ejemplo, representada en un dominio de excitación como el primer dominio, en la que la unidad de rastreo del nivel de ruido 130 puede estar, por ejemplo, configurada para recibir la información de primera porción de señal, en la que dicha información de primera  
30 porción de señal se representa en el dominio de 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 puede estar, por ejemplo, configurada para transformar la segunda porción de señal de audio o el valor derivado de la segunda porción de señal de audio desde un dominio temporal que es el segundo dominio hasta un dominio de FFT que es el dominio de rastreo, y en la que la unidad de  
35 rastreo del nivel de ruido 130 puede estar, por ejemplo, configurada para recibir la segunda porción de señal de audio que está representada en el dominio de FFT.

**[0303]** La figura 1b ilustra un aparato de acuerdo con otra realización. En la figura 1b, la unidad de transformada 120 de la figura 1a es una primera unidad de transformada 120, y la unidad de reconstrucción 140 de la figura 1a es una primera unidad de reconstrucción 140. El aparato comprende además una segunda unidad de transformada 121 y una segunda unidad de reconstrucción 141.

**[0304]** La segunda unidad de transformada 121 está configurada para transformar la información sobre el nivel de ruido desde el dominio de rastreo hasta el 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.  
45

**[0305]** Además, la segunda unidad de reconstrucción 141 está configurada para reconstruir una cuarta porción de señal de audio de la señal de audio dependiendo de la información sobre el nivel de ruido que está 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.  
50

**[0306]** La figura 1c ilustra un aparato para decodificar una señal de audio de acuerdo con otra realización. El aparato comprende además una primera unidad de agregación 150 para determinar un primer valor agregado dependiendo de la primera porción de señal de audio. Además, el aparato de la figura 1c comprende además una segunda unidad de agregación 160 para determinar un segundo valor agregado como el valor derivado de la segunda porción de señal de audio dependiendo de la segunda porción de señal de audio. En la realización de la figura 1c, la unidad de rastreo del nivel de ruido 130 está configurada para recibir un primer valor agregado como la información de primera porción de señal que está representada en el dominio de rastreo, en la que la unidad de  
55

rastreo del nivel de ruido 130 está configurada para recibir el segundo valor agregado como la información de segunda porción de señal que está representada en el dominio de rastreo. La unidad de rastreo del nivel de ruido 130 está configurada para determinar información sobre el nivel de ruido dependiendo del primer valor agregado que está representado en el dominio de rastreo y dependiendo del segundo valor agregado que está representado en el dominio de rastreo.

**[0307]** En una realización, la primera unidad de agregación 150 está configurada para determinar el primer valor agregado de modo que el primer valor agregado indica una media cuadrática de la primera porción de señal de audio o de una señal derivada de la primera porción de señal de audio. Además, la segunda unidad de agregación 160 está configurada para determinar el segundo valor agregado de modo que el segundo valor agregado indica una media cuadrática de la segunda porción de señal de audio o de una señal derivada de la segunda porción de señal de audio.

**[0308]** La figura 6 ilustra un aparato para decodificar una señal de audio de acuerdo con una realización adicional.

**[0309]** En la figura 6, la unidad de rastreo del nivel de fondo 630 implementa una unidad de rastreo del nivel de ruido 130 de acuerdo con la figura 1a.

**[0310]** Además, en la figura 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.

**[0311]** De acuerdo con algunas realizaciones, la (primera) unidad transformada 120 de la figura 1a, la figura 1b y la figura 1c está configurada para transformar el valor derivado de la segunda porción de señal de audio desde el segundo dominio hasta el dominio de rastreo aplicando un valor de ganancia (x) sobre el valor derivado de la segunda porción de señal de audio, por ejemplo, dividiendo el valor derivado de la segunda porción de señal de audio por un valor de ganancia (x). En otras realizaciones, un valor de ganancia puede, por ejemplo, multiplicarse.

**[0312]** En algunas realizaciones, el valor de ganancia (x) puede indicar, por ejemplo, una ganancia introducida por síntesis de codificación predictiva lineal, o el valor de ganancia (x) puede indicar, por ejemplo, una ganancia introducida por síntesis y desénfasis de codificación predictiva lineal.

**[0313]** En la figura 6, la unidad 622 proporciona el valor (x) que indica la ganancia introducida mediante síntesis y desénfasis de codificación predictiva lineal. La unidad 622 divide a continuación el valor, proporcionado por la segunda unidad de agregación 660, que es un valor derivado de la segunda porción de señal de audio, por el valor de ganancia proporcionado (x) (por ejemplo, bien dividiendo por x, o bien multiplicando el valor 1/x). De este modo, la unidad 620 de la figura 6 que comprende las unidades 621 y 622 implementa la primera unidad de transformada de la figura 1a, la figura 1b o la figura 1c.

**[0314]** El aparato de la figura 6 recibe una primera trama con una primera porción de señal de audio que es una excitación con voz y/o una excitación sin voz y que está representada en el dominio de rastreo, en la figura 6 un dominio de LPC (ACELP). La primera porción de señal de audio es alimentada en una unidad de síntesis y desénfasis de LPC 671 para procesamiento para obtener una salida de primera porción de señal de audio en el dominio temporal. 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 está representado en el dominio de rastreo, es alimentado a continuación a la unidad de rastreo del nivel de ruido 630.

**[0315]** Además, el aparato de la figura 6 recibe una segunda trama con una segunda porción de señal de audio que comprende un espectro de MDCT y que está representado en un dominio de MDCT. El relleno de ruido es llevado a cabo por un módulo de relleno de ruido 681, la conformación de ruido en el dominio de frecuencia es llevada a cabo por un módulo de conformación de ruido en el dominio de frecuencia 682, la transformación al dominio temporal es llevada a cabo por un módulo de iMDCT/OLA 683 (OLA = superposición y adición) y la predicción a largo plazo es llevada a cabo por una unidad de predicción a largo plazo 684. La unidad de predicción a largo plazo puede comprender, por ejemplo, un búfer de retardo (no mostrado en la figura 6).

**[0316]** La señal derivada de la segunda porción de señal de audio es alimentada a continuación al módulo de RMS 660 para obtener un segundo valor que indica que se obtiene una media cuadrática de esa señal derivada de la segunda porción de señal de audio. Este segundo valor (segundo valor de RMS) aún está representado en el

dominio temporal. La unidad 620 transforma a continuación el segundo valor de RMS desde el dominio temporal hasta el dominio de rastreo, en este contexto, el dominio de LPC (ACELP). El segundo valor de RMS, que está representado en el dominio de rastreo, es alimentado a continuación a la unidad de rastreo del nivel de ruido 630.

5 **[0317]** En realizaciones, el rastreo de nivel se lleva a cabo en el dominio de excitación, pero el desvanecimiento de TCX se lleva a cabo en el dominio temporal.

**[0318]** Aunque durante la decodificación normal el nivel de ruido de fondo es rastreado, éste puede usarse, por ejemplo, durante la pérdida de paquetes como un indicador de un nivel de ruido de confort apropiado, al que la  
10 última señal recibida se desvanece suavemente por niveles.

**[0319]** Derivar el nivel para rastreo y aplicar el desvanecimiento de nivel son, en general, 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, no se  
15 necesita compensación de ganancia, y que para TCX, la compensación de ganancia inversa que para la derivación de nivel (véase la figura 6) es necesaria y, por lo tanto, puede usarse la misma derivación de ganancia, tal como se ilustra mediante la figura 7.

**[0320]** En lo sucesivo, se describe la compensación de una influencia del filtro de paso alto sobre la ganancia de la síntesis de LPC de acuerdo con realizaciones.

**[0321]** La figura 8 perfila este enfoque. En particular, la figura 8 ilustra la aplicación del nivel de ruido de confort durante la pérdida de paquetes.

25 **[0322]** En la figura 8, la unidad de filtro de ganancia de paso alto 643, la unidad de multiplicación 644, la unidad de desvanecimiento 645, la unidad de filtro de paso alto 646, la unidad de desvanecimiento 647 y la unidad de combinación 648 juntas forman una primera unidad de reconstrucción.

**[0323]** Además, en la figura 8, la unidad de provisión de nivel de fondo 631 proporciona la información sobre  
30 el nivel de ruido. Por ejemplo, la unidad de provisión de nivel de fondo 631 puede implementarse igualmente como unidad de rastreo del nivel de fondo 630 de la figura 6.

**[0324]** Además, en la figura 8, la unidad de ganancia de síntesis y desénfasis de LPC 649 y la unidad de multiplicación 641 juntas forman una segunda unidad de transformada 640.

35 **[0325]** Además, en la figura 8, la unidad de desvanecimiento 642 representa una segunda unidad de reconstrucción.

**[0326]** En la realización de la figura 8, la excitación con voz y sin voz se desvanecen por separado: la  
40 excitación con voz se desvanece a cero, pero la excitación sin voz se desvanece hacia el nivel de ruido de confort. La figura 8 representa además un filtro de paso alto, que se introduce en la cadena de señales de la excitación sin voz para suprimir componentes de baja frecuencia para todos los casos excepto cuando la señal se clasificó como sin voz.

45 **[0327]** En relación con modelar la influencia del filtro de paso alto, el nivel después de síntesis y desénfasis de LPC se calcula una vez con y una vez sin el filtro de paso alto. Posteriormente, la relación de esos dos niveles se deriva y se usa para alterar el nivel de fondo aplicado.

**[0328]** Esto se ilustra mediante la figura 9. En particular, la figura 9 representa compensación de ganancia de  
50 paso alto avanzada durante ocultamiento de ACELP de acuerdo con una realización.

**[0329]** En lugar de la señal de excitación actual solamente se usa un simple impulso como entrada para este cálculo. Esto permite una complejidad reducida, dado que la respuesta al impulso decae rápidamente y, de este modo, la derivación de RMS puede realizarse en un marco temporal más corto. En la práctica, solamente se usa una  
55 subtrama en lugar de toda la trama.

**[0330]** De acuerdo con una realización, la unidad de rastreo del nivel de ruido 130 está configurada para determinar un nivel de ruido de confort como la información sobre el nivel de ruido. La unidad de reconstrucción 140 está configurada para reconstruir la tercera porción de señal de audio dependiendo de la información sobre el 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.

5 **[0331]** De acuerdo con una realización, la unidad de rastreo del nivel de ruido 130 está configurada para determinar un nivel de ruido de confort como la información sobre el nivel de ruido. La unidad de reconstrucción 140 está configurada para reconstruir la tercera porción de señal de audio dependiendo de la información sobre el 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.

10 **[0332]** En una realización, la unidad de rastreo del nivel de ruido 130 está configurada para determinar un nivel de ruido de confort como la información sobre el nivel de ruido derivada de un espectro del nivel de ruido, en la que dicho espectro del nivel de ruido se obtiene aplicando el enfoque de estadística mínima. La unidad de reconstrucción 140 está 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  
15 interfaz receptora 110 o si dicha tercera trama es recibida por la interfaz receptora 110 pero está alterada.

**[0333]** En una realización, la (primera y/o segunda) unidad de reconstrucción 140, 141 puede estar, por ejemplo, configurada para reconstruir la tercera porción de señal de audio dependiendo de la información sobre el nivel de ruido y dependiendo de la primera porción de señal de audio, si dicha tercera (cuarta) trama de la pluralidad  
20 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.

**[0334]** De acuerdo con una realización, la (primera y/o segunda) unidad de reconstrucción 140, 141 puede estar, por ejemplo, configurada para reconstruir la tercera (o cuarta) porción de señal de audio atenuando o  
25 amplificando la primera porción de señal de audio.

**[0335]** La figura 14 ilustra un aparato para decodificar 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  
30 para recibir una segunda trama que comprende una segunda porción de señal de audio de la señal de audio.

**[0336]** Además, el aparato comprende una unidad de rastreo del nivel de ruido 130, en la que la unidad de rastreo del nivel de ruido 130 está configurada para determinar información sobre el nivel de ruido dependiendo de al menos una de la primera porción de señal de audio y la segunda porción de señal de audio (esto significa:  
35 dependiendo de la primera porción de señal de audio y/o la segunda porción de señal de audio), en la que la información sobre el nivel de ruido está representada en un dominio de rastreo.

**[0337]** Además, el aparato comprende una primera unidad de reconstrucción 140 para reconstruir, en un primer dominio de reconstrucción, una tercera porción de señal de audio de la señal de audio dependiendo de la  
40 información sobre el 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 la que el primer dominio de reconstrucción es diferente de o igual al dominio de rastreo.

**[0338]** Además, el aparato comprende una unidad de transformada 121 para transformar la información sobre el nivel de ruido desde el dominio de rastreo hasta 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 la que el segundo dominio de reconstrucción es diferente del dominio de rastreo, y en la que el segundo dominio de reconstrucción es diferente del primer dominio de reconstrucción, y

50 **[0339]** Además, el aparato comprende una segunda unidad de reconstrucción 141 para reconstruir, en el segundo dominio de reconstrucción, una cuarta porción de señal de audio de la señal de audio dependiendo de la información sobre el nivel de ruido que está 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.

55 **[0340]** De acuerdo con algunas realizaciones, el dominio de rastreo puede ser, por ejemplo, aquel en el que el dominio de rastreo es un dominio temporal, un dominio espectral, un dominio de FFT, un dominio de MDCT o un dominio de excitación. El primer dominio de reconstrucción puede ser, por ejemplo, el dominio temporal, el dominio espectral, el dominio de FFT, el dominio de MDCT o el dominio de excitación. El segundo dominio de reconstrucción

puede ser, por ejemplo, el dominio temporal, el dominio espectral, el dominio de FFT, el dominio de MDCT o el dominio de excitación.

**[0341]** En una realización, el dominio de rastreo puede ser, por ejemplo, el dominio de FFT, el primer dominio de reconstrucción puede ser, por ejemplo, el dominio temporal, y el segundo dominio de reconstrucción puede ser, por ejemplo, el dominio de excitación.

**[0342]** En otra realización, el dominio de rastreo puede ser, por ejemplo, el dominio temporal, el primer dominio de reconstrucción puede ser, por ejemplo, el dominio temporal, y el segundo dominio de reconstrucción puede ser, por ejemplo, el dominio de excitación.

**[0343]** De acuerdo con una realización, dicha primera porción de señal de audio puede estar, por ejemplo, representada en un primer dominio de entrada, y dicha segunda porción de señal de audio puede estar, por ejemplo, representada en un segundo dominio de entrada. La unidad de transformada puede ser, por ejemplo, una segunda unidad de transformada. El aparato puede comprender, por ejemplo, además una primera unidad de transformada para transformar la segunda porción de señal de audio o un valor o señal derivada de la segunda porción de señal de audio desde el segundo dominio de entrada hasta el dominio de rastreo para obtener una información de segunda porción de señal. La unidad de rastreo del nivel de ruido puede estar, por ejemplo, configurada para recibir una información de primera porción de señal que está 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 del nivel de ruido está configurada para recibir la segunda porción de señal que está representada en el dominio de rastreo, y en la que la unidad de rastreo del nivel de ruido está configurada para determinar la información sobre el nivel de ruido dependiendo de la información de primera porción de señal que está representada en el dominio de rastreo y dependiendo de la información de segunda porción de señal que está representada en el dominio de rastreo.

**[0344]** De acuerdo con una realización, el primer dominio de entrada puede ser, por ejemplo, el dominio de excitación, y el segundo dominio de entrada puede ser, por ejemplo, el dominio de MDCT.

**[0345]** En otra realización, el primer dominio de entrada puede ser, por ejemplo, el dominio de MDCT, y en la que el segundo dominio de entrada puede ser, por ejemplo, el dominio de MDCT.

**[0346]** Si, por ejemplo, una señal está representada en un dominio temporal, puede estar, por ejemplo, representada mediante muestras en el dominio temporal de la señal. O, por ejemplo, si una señal está representada en un dominio espectral, puede estar, por ejemplo, representada mediante muestras espectrales de un espectro de la señal.

**[0347]** En una realización, el dominio de rastreo puede ser, por ejemplo, el dominio de FFT, el primer dominio de reconstrucción puede ser, por ejemplo, el dominio temporal, y el segundo dominio de reconstrucción puede ser, por ejemplo, el dominio de excitación.

**[0348]** En otra realización, el dominio de rastreo puede ser, por ejemplo, el dominio temporal, el primer dominio de reconstrucción puede ser, por ejemplo, el dominio temporal, y el segundo dominio de reconstrucción puede ser, por ejemplo, el dominio de excitación.

**[0349]** En algunas realizaciones, las unidades ilustradas en la figura 14, pueden estar, por ejemplo, configuradas tal como se describe para las figuras 1a, 1b, 1c y 1d.

**[0350]** Respecto a realizaciones particulares, en, por ejemplo, un modo de tasa baja, un aparato de acuerdo con una realización puede recibir, por ejemplo, tramas de ACELP como entrada, que están representadas en un dominio de excitación, y que son transformadas a continuación a un dominio temporal mediante síntesis de LPC. Además, en el modo de tasa baja, el aparato de acuerdo con una realización puede recibir, por ejemplo, tramas de TCX como entrada, que están representadas en un dominio de MDCT, y que son transformadas a continuación a un dominio temporal mediante una MDCT inversa.

**[0351]** A continuación se lleva a cabo el rastreo en un dominio de FFT, en el que la señal de FFT se deriva a partir de la señal de dominio temporal llevando a cabo una FFT (transformada rápida de Fourier). El rastreo puede, por ejemplo, llevarse a cabo llevando a cabo un enfoque de estadística mínima, independiente para todas las líneas espectrales para obtener un espectro de ruido de confort.

**[0352]** A continuación se lleva a cabo ocultamiento llevando a cabo derivación de nivel basándose en el espectro de ruido de confort. La derivación de nivel se lleva a cabo basándose en el espectro de ruido de confort. La conversión de nivel en el dominio temporal se lleva a cabo para PLC de TCX de FD. Se lleva a cabo un desvanecimiento en el dominio temporal. Una derivación de nivel en el dominio de excitación se lleva a cabo para PLC de ACELP y para PLC de TCX de TD (de tipo ACELP). A continuación, se lleva a cabo un desvanecimiento en el dominio de excitación.

**[0353]** La siguiente lista resume esto:

10

tasa baja:

- entrada:

15

- o acelp (dominio de excitación -> dominio temporal, mediante síntesis de lpc)

- o tcx (dominio de mdct -> dominio temporal, mediante MDCT inversa)

20

- rastreo:

- o dominio de FFT, derivado del dominio temporal mediante FFT

- o estadística mínima, independiente para todas las líneas espectrales -> espectro de ruido de confort

25

- ocultamiento:

- o derivación de nivel basada en el espectro de ruido de confort

- o conversión de nivel en el dominio temporal para

30

- PLC de TCX de FD

- > desvanecimiento en el dominio temporal

- o conversión de nivel en el dominio de excitación para

35

- PLC de ACELP

- PLC de TCX de TD (de tipo ACELP)

- > desvanecimiento en el dominio de excitación

40

**[0354]** En, por ejemplo, un modo de tasa alta, puede recibir, por ejemplo, tramas de TCX como entrada, que están representadas en el dominio de MDCT, y que son transformadas a continuación al dominio temporal mediante una MDCT inversa.

45

**[0355]** A continuación se lleva a cabo rastreo en el dominio temporal. El rastreo puede, por ejemplo, llevarse a cabo llevando a cabo un enfoque de estadística mínima basado en el nivel de energía para obtener un nivel de ruido de confort.

50

**[0356]** Para ocultamiento, para PLC de TCX de FD, el nivel puede usarse como tal y solamente puede llevarse a cabo un desvanecimiento en el dominio temporal. Para PLC de TCX de TD (de tipo ACELP), se lleva a cabo conversión en el dominio de excitación y desvanecimiento en el dominio de excitación.

**[0357]** La siguiente lista resume esto:

55

tasa alta:

- entrada:

- o tcx (dominio de mdct -> dominio temporal, mediante MDCT inversa)

- rastreo:

- dominio temporal

5

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

- ocultamiento:

10

- uso del nivel "como tal"

- PLC de TCX de FD

- > desvanecimiento en el dominio temporal

15

- conversión de nivel en el dominio de excitación para

- PLC de TCX de TD (de tipo ACELP)

- > desvanecimiento en el dominio de excitación

20 **[0358]** El dominio de FFT y el dominio de MDCT son ambos dominios espectrales, mientras que el dominio de excitación es alguna clase de dominio temporal.

**[0359]** De acuerdo con una realización, la primera unidad de reconstrucción 140 puede estar, por ejemplo, configurada para reconstruir la tercera porción de señal de audio llevando a cabo un primer desvanecimiento a un espectro de tipo ruido. La segunda unidad de reconstrucción 141 puede estar, por ejemplo, configurada para reconstruir la cuarta porción de señal de audio llevando a cabo 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 pueden estar, por ejemplo, configuradas para llevar a cabo 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.

30

**[0360]** Ahora se considera conformación espectral adaptativa de ruido de confort.

**[0361]** Para conseguir conformación adaptativa para ruido de confort durante la pérdida de paquetes en ráfagas, como primera etapa, puede llevarse a cabo el descubrimiento de coeficientes de LPC apropiados que representan el ruido de fondo. Estos coeficientes de LPC pueden derivarse durante la voz activa usando un enfoque de estadística mínima para descubrir el espectro de ruido de fondo y a continuación calcular coeficientes de LPC a partir de éste usando un algoritmo arbitrario para derivación de LPC conocida a partir 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 MDCT.

40

**[0362]** El desvanecimiento a ruido de confort puede realizarse en el dominio de ISF (también aplicable en el dominio de LSF; LSF frecuencia espectral lineal):

45

$$f_{current}[i] = \alpha \cdot f_{last}[i] + (1 - \alpha) \cdot p_{tmean}[i] \quad i = 0 \dots 16 \quad (26)$$

ajustando  $p_{tmean}$  a coeficientes de LP apropiados que describen el ruido de confort.

**[0363]** Respecto a la conformación espectral adaptativa descrita anteriormente del ruido de confort, una realización más general se ilustra mediante la figura 11.

50

**[0364]** La figura 11 ilustra un aparato para decodificar una señal de audio codificada para obtener una señal de audio reconstruida de acuerdo con una realización.

**[0365]** El aparato comprende una interfaz receptora 1110 para recibir una o más tramas, un generador de coeficientes 1120, y un reconstructor de señales 1130.

55

**[0366]** El generador de coeficientes 1120 está configurado para determinar, si una trama actual de las una o

más tramas es recibida por la interfaz receptora 1110 y si la trama actual que es recibida por la interfaz receptora 1110 no está alterada/es errónea, uno o más primeros coeficientes de señal de audio, que están compuestos por la trama actual, en la 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  
 5 codificada. Además, el generador de coeficientes 1120 está configurado para generar uno o más segundos coeficientes de señal de audio, dependiendo de los uno o más primeros coeficientes de señal de audio y dependiendo de los uno o más coeficientes de ruido, si la trama actual no es recibida por la interfaz receptora 1110 o si la trama actual que es recibida por la interfaz receptora 1110 está alterada/es errónea.

10 **[0367]** El reconstructor de señales de audio 1130 está configurado para reconstruir una primera porción de la señal de audio reconstruida dependiendo de los uno o más primeros coeficientes de señal de audio, si la trama actual es recibida por la interfaz receptora 1110 y si la trama actual que es recibida por la interfaz receptora 1110 no está alterada. Además, el reconstructor de señales de audio 1130 está configurada para reconstruir una segunda  
 15 porción de la señal de audio reconstruida dependiendo de los 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 que es recibida por la interfaz receptora 1110 está alterada.

**[0368]** La determinación de un ruido de fondo es bien conocida en la técnica (véase, por ejemplo, [Mar01]: Rainer Martin, Noise power spectral density estimation based on optimal smoothing and minimum statistics, IEEE  
 20 Transactions on Speech and Audio Processing 9 (2001), no. 5, 504 -512), y en una realización, el aparato procede en consecuencia.

**[0369]** En algunas realizaciones, los uno o más primeros coeficientes de señal de audio pueden ser, por ejemplo, uno o más coeficientes de filtro predictivo lineal de la señal de audio codificada. En algunas realizaciones,  
 25 los uno o más primeros coeficientes de señal de audio puede ser, por ejemplo, uno o más coeficientes de filtro predictivo lineal de la señal de audio codificada.

**[0370]** Es bien conocido en la técnica cómo reconstruir una señal de audio, por ejemplo, una señal de voz, a partir de coeficientes de filtro predictivo lineal o a partir de pares espectrales de inmitancia (véase, por ejemplo,  
 30 [3GP09c]: Speech codec speech processing functions; adaptive multi-rate - wideband (AMRWB) speech codec; transcoding functions, 3GPP TS 26.190, 3rd Generation Partnership Project, 2009), y en una realización, el reconstructor de señales procede en consecuencia.

**[0371]** De acuerdo con una realización, los uno o más coeficientes de ruido pueden ser, por ejemplo, uno o  
 35 más coeficientes de filtro predictivo lineal que indican el ruido de fondo de la señal de audio codificada. En una realización, los uno o más coeficientes de filtro predictivo lineal puede representar, por ejemplo, una forma espectral del ruido de fondo.

**[0372]** En una realización, el generador de coeficientes 1120 puede estar, por ejemplo, configurado para  
 40 determinar las una o más segundas porciones de señal de audio de modo que las 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 modo que los uno o más primeros coeficientes de señal de audio sean uno o más pares espectrales de inmitancia de la señal de audio reconstruida.

45 **[0373]** De acuerdo con una realización, el generador de coeficientes 1120 puede estar configurado, por ejemplo, para generar los uno o más segundos coeficientes de señal de audio aplicando la fórmula:

$$f_{current}[i] = \alpha \cdot f_{last}[i] + (1 - \alpha) \cdot pt_{mean}[i]$$

50 en la que  $f_{current}[i]$  indica uno de los uno o más segundos coeficientes de señal de audio, en la que  $f_{last}[i]$  indica uno de los uno o más primeros coeficientes de señal de audio, en la que  $pt_{mean}[i]$  es uno de los uno o más coeficientes de ruido, en la que  $\alpha$  es un número real con  $0 \leq \alpha \leq 1$ , y en la que  $i$  es un índice.

**[0374]** De acuerdo con una realización,  $f_{last}[i]$  indica un coeficiente de filtro predictivo lineal de la señal de  
 55 audio codificada, y en la que  $f_{current}[i]$  indica un coeficiente de filtro predictivo lineal de la señal de audio reconstruida.

**[0375]** En una realización,  $pt_{mean}[i]$  puede ser, por ejemplo, un coeficiente de filtro predictivo lineal que indica el ruido de fondo de la señal de audio codificada.

**[0376]** De acuerdo con una realización, el generador de coeficientes 1120 puede estar configurado, por ejemplo, para generar al menos 10 segundos coeficientes de señal de audio como los uno o más segundos coeficientes de señal de audio.

5

**[0377]** En una realización, el generador de coeficientes 1120 puede estar, por ejemplo, configurado para determinar, si la trama actual de las una o más tramas es recibida por la interfaz receptora 1110 y si la trama actual que es recibida por la interfaz receptora 1110 no está alterada, los uno o más coeficientes de ruido determinando un espectro de ruido de la señal de audio codificada.

10

**[0378]** En lo sucesivo, se considera el desvanecimiento del espectro MDCT a ruido blanco antes de la aplicación de FDNS.

**[0379]** En lugar de modificar aleatoriamente el signo de un "bin" de MDCT (mezcla de signos), el espectro completo se llena con ruido blanco, que es conformado usando la FDNS. Para evitar un cambio instantáneo en las características del espectro, se aplica un desvanecimiento gradual entre mezcla de signos y relleno de ruido. El desvanecimiento gradual puede realizarse de la siguiente manera:

15

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

20 donde:

*cum\_damping* es el factor de atenuación (absoluto) - disminuye de trama a trama, comenzando a partir de 1 y disminuyendo hacia 0

25

*x\_old* es el espectro de la última trama recibida

*random\_sign* retorna 1 o -1

*noise* contiene un vector aleatorio (ruido blanco) que es convertido de escala de modo que su media cuadrática (RMS) es similar al último espectro bueno.

**[0380]** El término *random\_sign()\*old\_x[i]* caracteriza el proceso de mezcla de signos para aleatorizar las fases y, de este modo, evitar repeticiones de armónicos.

**[0381]** Posteriormente, podría realizarse otra normalización del nivel de energía después del desvanecimiento gradual para asegurarse de que la energía de suma no se desvía debido a la correlación de los dos vectores.

35

**[0382]** De acuerdo con realizaciones, la primera unidad de reconstrucción 140 puede estar, por ejemplo, configurada para reconstruir la tercera porción de señal de audio dependiendo de la información sobre el nivel de ruido y dependiendo de la primera porción de señal de audio. En una realización particular, la primera unidad de reconstrucción 140 puede estar, por ejemplo, configurada para reconstruir la tercera porción de señal de audio atenuando o amplificando la primera porción de señal de audio.

40

**[0383]** En algunas realizaciones, la segunda unidad de reconstrucción 141 puede estar, por ejemplo, configurada para reconstruir la cuarta porción de señal de audio dependiendo de la información sobre el nivel de ruido y dependiendo de la segunda porción de señal de audio. En una realización particular, la segunda unidad de reconstrucción 141 puede estar, por ejemplo, configurada para reconstruir la cuarta porción de señal de audio atenuando o amplificando la segunda porción de señal de audio.

45

**[0384]** Respecto al desvanecimiento descrito anteriormente del espectro MDCT a ruido blanco antes de la aplicación de FDNS, una realización más general se ilustra mediante la figura 12.

50

**[0385]** La figura 12 ilustra un aparato para decodificar una señal de audio codificada para obtener una señal de audio reconstruida de acuerdo con una realización.

5 **[0386]** El aparato comprende una interfaz receptora 1210 para recibir una o más tramas que comprenden información sobre una pluralidad de muestras de señal de audio de un espectro de señal de audio de la señal de audio codificada, y un procesador 1220 para generar la señal de audio reconstruida.

10 **[0387]** El procesador 1220 está configurado para generar la señal de audio reconstruida desvaneciendo un espectro modificado a un espectro diana, 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 la que el espectro modificado comprende una pluralidad de muestras de señal modificadas, en la que, para cada una de las muestras de señal 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 señal de audio.

15 **[0388]** Además, el procesador 1220 está configurado para no desvanecer el espectro modificado al espectro diana, si la trama actual de las una o más tramas es recibida por la interfaz receptora 1210 y si la trama actual que es recibida por la interfaz receptora 1210 no está alterada.

20 **[0389]** De acuerdo con una realización, el espectro diana es un espectro de tipo ruido.

**[0390]** En una realización, el espectro de tipo ruido representa ruido blanco.

25 **[0391]** De acuerdo con una realización, el espectro de tipo ruido está conformado.

**[0392]** En una realización, la forma del espectro de tipo ruido depende de un espectro de señal de audio de una señal recibida previamente.

30 **[0393]** De acuerdo con una realización, el espectro de tipo ruido está conformado dependiendo de la forma del espectro de señal de audio.

**[0394]** En una realización, el procesador 1220 emplea un factor de inclinación para conformar el espectro de tipo ruido.

35 **[0395]** De acuerdo con una realización, el procesador 1220 emplea la fórmula

$$\text{shaped\_noise}[i] = \text{noise} * \text{power}(\text{tilt\_factor}, i/N)$$

40 en la que N indica el número de muestras,

en la que i es un índice,

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

45 en la que *power* es una función de potencia.

**[0396]** Si el *tilt\_factor* es menor que 1 esto significa atenuación con i creciente. Si el *tilt\_factor* es mayor que 1 significa amplificación con i creciente.

50 **[0397]** De acuerdo con otra realización, el procesador 1220 puede emplear la fórmula

$$\text{shaped\_noise}[i] = \text{noise} * (1 + i / (N-1) * (\text{tilt\_factor}-1))$$

55 en la que N indica el número de muestras,

en la que i es un índice, en la que  $0 \leq i < N$ ,

con *tilt\_factor* > 0.

**[0398]** De acuerdo con una realización, el procesador 1220 está configurado para generar el espectro modificado, cambiando un signo de una o más de las muestras de señal de audio del espectro de señal de audio, si la trama actual no es recibida por la interfaz receptora 1210 o si la trama actual que es recibida por la interfaz receptora 1210 está alterada.

**[0399]** En una realización, cada una de las muestras de señal de audio del espectro de señal de audio está representada por un número real pero no por un número imaginario.

**[0400]** De acuerdo con una realización, las muestras de señal de audio del espectro de señal de audio están representadas en un dominio de transformada de coseno discreta modificada.

**[0401]** En otra realización, las muestras de señal de audio del espectro de señal de audio están representadas en un dominio de transformada de seno discreta modificada.

**[0402]** De acuerdo con una realización, el procesador 1220 está configurado para generar el espectro modificado empleando una función de signo aleatorio que emite de forma aleatoria o pseudoaleatoria bien un primer o un segundo valor.

**[0403]** En una realización, el procesador 1220 está configurado para desvanecer el espectro modificado al espectro diana disminuyendo posteriormente un factor de atenuación.

**[0404]** De acuerdo con una realización, el procesador 1220 está configurado para desvanecer el espectro modificado al espectro diana aumentando posteriormente un factor de atenuación.

**[0405]** En una realización, si la trama actual no es recibida por la interfaz receptora 1210 o si la trama actual que es recibida por la interfaz receptora 1210 está alterada, el procesador 1220 está configurado para generar la señal de audio reconstruida empleando la fórmula:

```

30      x[i]      =      (1-cum_damping)      *      noise[i]      +      cum_damping      *
      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 señal de audio de la señal de audio codificada, en la que *random\_sign()* retorna 1 o -1, y en la que *noise* es un vector aleatorio que indica el espectro diana.

**[0406]** Algunas realizaciones continúan una operación de LTP de TCX. En esas realizaciones, la operación de LTP de TCX es continuada durante el ocultamiento con los parámetros de LTP (retraso de LTP y ganancia de LTP) derivados de la última trama buena.

**[0407]** Las operaciones de LTP pueden resumirse como:

- Alimentar el búfer de retardo de LTP basándose en la salida derivada previamente.
- Basándose en el retraso de LTP: seleccionar la porción de señal apropiada del búfer de retardo de LTP que se usa como una contribución de LTP para conformar la señal actual.
- Sustituir la escala de esta contribución de LTP usando la ganancia de LTP.
- Añadir esta contribución de LTP con sustitución de escala a la señal de entrada de LTP para generar la señal de salida de LTP.

**[0408]** Podrían considerarse diferentes enfoques con respecto al tiempo, donde se realiza la actualización del búfer de retardo de LTP:

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 se usará durante el procesamiento de LTP en trama *n*.

**[0409]** 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  $n$  to que se usará durante el procesamiento de LTP en trama  $n+1$ .

**[0410]** En lo sucesivo, se considera el desacoplamiento del bucle de retroalimentación de LTP de TCX.

5 **[0411]** Desacoplar el bucle de retroalimentación de LTP de TCX evita la introducción de ruido adicional (que resulta de la sustitución de ruido aplicada a la señal de entrada de LPT) durante cada bucle de retroalimentación del decodificador de LTP cuando está en modo de ocultamiento.

**[0412]** La figura 10 ilustra este desacoplamiento. En particular, la figura 10 representa el desacoplamiento del bucle de retroalimentación de LTP durante el ocultamiento ( $bfi=1$ ).

**[0413]** La figura 10 ilustra un búfer de retardo 1020, un selector de muestras 1030 y un procesador de muestras 1040 (el procesador de muestras 1040 se indica mediante la línea discontinua).

15 **[0414]** Hacia el tiempo, cuando se realiza la actualización del búfer de retardo de LTP 1020, algunas realizaciones proceden de la siguiente manera:

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

- Para el funcionamiento desacoplado: podría ser preferible actualizar el búfer de retardo de LTP 1020 como la última operación de LTP, dado que la contribución de LTP a la señal habitualmente sólo se almacena temporalmente. Con este enfoque, la señal de contribución de LTP transitoriamente se preserva. Como implementación, este búfer de contribución de LTP podría simplemente hacerse persistente.

**[0415]** Suponiendo que este último enfoque se usa en cualquier caso (funcionamiento normal y ocultamiento), realizaciones, pueden implementar, por ejemplo, lo siguiente:

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

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

**[0416]** Algunas realizaciones desvanecen la ganancia de LTP de TCX hacia cero. En dicha realización, la ganancia de LTP de TCX puede, por ejemplo, desvanecerse hacia cero con cierto factor de desvanecimiento adaptativo de señal. Esto puede realizarse, por ejemplo, de forma iterativa, por ejemplo, de acuerdo con el siguiente pseudocódigo:

```
gain = gain_past * damping;
[...];
gain_past = gain;
```

donde:

*gain* es la ganancia del decodificado de LTP de TCX aplicada en la trama actual;

*gain\_past* es la ganancia del decodificador de LTP de TCX aplicada en la trama anterior;

*damping* es el factor de desvanecimiento (relativo).

**[0417]** La figura 1d ilustra un aparato de acuerdo con una realización adicional, en el que el aparato comprende además una unidad de predicción a largo plazo 170 que comprende un búfer de retardo 180. La unidad de predicción a largo plazo 170 está configurada para generar una señal procesada dependiendo de la segunda porción de señal de audio, dependiendo de una entrada del búfer de retardo que se almacena en el búfer de retardo 180 y dependiendo de una ganancia de predicción a largo plazo. Además, la unidad de predicción a largo plazo está

configurada para desvanecer la ganancia de predicción a largo plazo hacia cero, si dicha tercera trama de la pluralidad de tramas no es recibida por la interfaz receptora 110 o si dicha tercera trama es recibida por la interfaz receptora 110 pero está alterada.

- 5 **[0418]** En otras realizaciones (no mostradas), la unidad de predicción a largo plazo puede estar configurada, por ejemplo, para generar una señal procesada dependiendo de la primera porción de señal de audio, dependiendo de una entrada del búfer de retardo que se almacena en el búfer de retardo y dependiendo de una ganancia de predicción a largo plazo.
- 10 **[0419]** En la figura 1d, la primera unidad de reconstrucción 140 puede generar, por ejemplo, la tercera porción de señal de audio dependiendo además de la señal procesada.
- [0420]** En una realización, la unidad de predicción a largo plazo 170 puede estar, por ejemplo, configurada para desvanecer la ganancia de predicción a largo plazo hacia cero, en la que una velocidad con la que la ganancia de predicción a largo plazo se desvanece a cero depende de un factor de desvanecimiento.
- 15 **[0421]** Como alternativa o adicionalmente, la unidad de predicción a largo plazo 170 puede estar, por ejemplo, configurada para actualizar la entrada del búfer de retardo 180 almacenando 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 20 110 o si dicha tercera trama es recibida por la interfaz receptora 110 pero está alterada.
- [0422]** Respecto al uso descrito anteriormente de LTP de TCX, una realización más general se ilustra mediante la figura 13.
- 25 **[0423]** La figura 13 ilustra un aparato para decodificar una señal de audio codificada para obtener una señal de audio reconstruida.
- [0424]** El aparato comprende una interfaz receptora 1310 para recibir una pluralidad de tramas, un búfer de retardo 1320 para almacenar muestras de señal de audio de la señal de audio decodificada, un selector de muestras 30 1330 para seleccionar una pluralidad de muestras de señal de audio seleccionadas a partir de las muestras de señal de audio que se almacenan en el búfer de retardo 1320, y un procesador de muestras 1340 para procesar las muestras de señal de audio seleccionadas para obtener muestras de señal de audio reconstruida de la señal de audio reconstruida.
- 35 **[0425]** 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 que es recibida por la interfaz receptora 1310 no está alterada, la pluralidad de muestras de señal de audio seleccionadas a partir de las muestras de señal de audio que se almacenan en el búfer de retardo 1320 dependiendo de una información de retraso de altura tonal que está compuesta por la trama actual. Además, el selector de muestras 1330 está configurado para seleccionar, si la trama 40 actual no es recibida por la interfaz receptora 1310 o si la trama actual que es recibida por la interfaz receptora 1310 está alterada, la pluralidad de muestras de señal de audio seleccionadas a partir de las muestras de señal de audio que se almacenan en el búfer de retardo 1320 dependiendo de una información de retraso de altura tonal que está compuesta por otra trama que es recibida previamente por la interfaz receptora 1310.
- 45 **[0426]** De acuerdo con una realización, el procesador de muestras 1340 puede estar, por ejemplo, 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 que es recibida por la interfaz receptora 1310 no está alterada, sustituyendo la escala de las muestras de señal de audio seleccionadas dependiendo de la información de ganancia que está compuesta por la trama actual. Además, el selector de muestras 1330 puede estar, por ejemplo, configurado para 50 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 que es recibida por la interfaz receptora 1310 está alterada, sustituyendo la escala de las muestras de señal de audio seleccionadas dependiendo de la información de ganancia que está compuesta por dicha otra trama que es recibida previamente por la interfaz receptora 1310.
- 55 **[0427]** En una realización, el procesador de muestras 1340 puede estar, por ejemplo, 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 que es recibida por la interfaz receptora 1310 no está alterada, multiplicando las muestras de señal de audio seleccionadas y un valor que depende de la información de ganancia que está compuesta 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 que es recibida por la interfaz receptora 1310 está alterada, multiplicando las muestras de señal de audio seleccionadas y un valor que depende de la información de ganancia que está compuesta por dicha otra trama que es recibida previamente por la interfaz receptora 1310.

5

**[0428]** De acuerdo con una realización, el procesador de muestras 1340 puede estar, por ejemplo, configurado para almacenar las muestras de señal de audio reconstruida en el búfer de retardo 1320.

**[0429]** En una realización, el procesador de muestras 1340 puede estar, por ejemplo, configurado para almacenar las muestras de señal de audio reconstruida en el búfer de retardo 1320 antes de que una trama adicional sea recibida por la interfaz receptora 1310.

10

**[0430]** De acuerdo con una realización, el procesador de muestras 1340 puede estar, por ejemplo, configurado para almacenar las muestras de señal de audio reconstruida en el búfer de retardo 1320 después de que una trama adicional sea recibida por la interfaz receptora 1310.

15

**[0431]** En una realización, el procesador de muestras 1340 puede estar, por ejemplo, configurado para sustituir la escala de las muestras de señal de audio seleccionadas dependiendo de la información de ganancia para obtener muestras de señal de audio con la escala sustituida y, combinando las muestras de señal de audio con la escala sustituida con muestras de señal de audio de entrada, para obtener las muestras de señal de audio procesadas.

20

**[0432]** De acuerdo con una realización, el procesador de muestras 1340 puede estar, por ejemplo, configurado para almacenar las muestras de señal de audio procesadas, que indican la combinación de las muestras de señal de audio con la escala sustituida 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 con la escala sustituida en el búfer de retardo 1320, si la trama actual es recibida por la interfaz receptora 1310 y si la trama actual que es recibida por la interfaz receptora 1310 no está alterada. Además, el procesador de muestras 1340 está configurado para almacenar las muestras de señal de audio con la escala sustituida 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 que es recibida por la interfaz receptora 1310 está alterada.

30

**[0433]** De acuerdo con otra realización, el procesador de muestras 1340 puede estar, por ejemplo, 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 que es recibida por la interfaz receptora 1310 está alterada.

35

**[0434]** En una realización, el selector de muestras 1330 puede estar, por ejemplo, configurado para obtener las muestras de señal de audio reconstruida sustituyendo la escala de las muestras de señal de audio seleccionadas dependiendo de una ganancia modificada, en la que la ganancia modificada se define de acuerdo con la fórmula:

40

$$\text{gain} = \text{gain\_past} * \text{damping};$$

en la que *gain* es la ganancia modificada, en la que el selector de muestras 1330 puede estar, por ejemplo, configurado para ajustar *gain\_past* a la ganancia después de *gain* y ha sido calculado, y en la que *damping* es un número real.

45

**[0435]** De acuerdo con una realización, el selector de muestras 1330 puede estar, por ejemplo, configurado para calcular la ganancia modificada.

50

**[0436]** En una realización, *damping* puede definirse, por ejemplo, de acuerdo con:  $0 < \text{damping} < 1$ .

**[0437]** De acuerdo con una realización, la ganancia modificada *gain* puede ajustarse, por ejemplo, a cero, si al menos un número predeterminado de tramas no ha sido recibido por la interfaz receptora 1310 desde que una trama ha sido recibida la última por la interfaz receptora 1310.

55

**[0438]** En lo sucesivo, se considera la velocidad de desvanecimiento. Hay varios módulos de ocultamiento que aplican cierta clase de desvanecimiento. Aunque la velocidad de este desvanecimiento podría seleccionarse de forma 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:

Para ACELP, debe usarse la misma velocidad de desvanecimiento, en particular, para el libro de códigos adaptativo (alterando la ganancia), y/o para la señal del libro de códigos innovador (alterando la ganancia).

- 5
- [0439]** Además, para TCX, debe usarse la misma velocidad de desvanecimiento, en particular, para señal de dominio temporal, y/o para la ganancia de LTP (desvanecimiento a cero), y/o para la ponderación de LPC (desvanecimiento a cero), y/o para los coeficientes de LP (desvanecimiento a forma espectral de fondo), y/o para el desvanecimiento gradual a ruido blanco.
- 10
- [0440]** Podría ser preferible además usar también la misma velocidad de desvanecimiento para ACELP y TCX pero, debido a la diferente naturaleza de los núcleos, también podría elegirse usar diferentes velocidades de desvanecimiento.
- 15
- [0441]** Esta velocidad de desvanecimiento podría ser estática, pero es preferentemente adaptativa a las características de señal. Por ejemplo, la velocidad de desvanecimiento puede depender, por ejemplo, del factor de estabilidad de LPC (TCX) y/o de una clasificación, y/o de un número de tramas perdidas consecutivamente.
- 20
- [0442]** La velocidad de desvanecimiento puede determinarse, por ejemplo, dependiendo del factor de atenuación, que podría proporcionarse absoluta o relativamente, y que también podría cambiar a lo largo del tiempo durante cierto desvanecimiento.
- [0443]** En realizaciones, se usa la misma velocidad de desvanecimiento para desvanecimiento de la ganancia de LTP que para el desvanecimiento de ruido blanco.
- 25
- [0444]** Se ha proporcionado un aparato, procedimiento y programa informático para generar una señal de ruido de confort tal como se ha descrito anteriormente.
- [0445]** Aunque algunos aspectos se han descrito en el contexto de un aparato, es evidente que estos aspectos también representan una descripción del procedimiento correspondiente, donde un bloque o dispositivo corresponde a una etapa del procedimiento o una característica de una etapa del procedimiento. Análogamente, los aspectos descritos en el contexto de una etapa del procedimiento también representan una descripción de un bloque o elemento o característica correspondiente de un aparato correspondiente.
- 30
- 35 **[0446]** La señal descompuesta de la invención puede ser almacenada en un medio de almacenamiento digital o puede ser transmitida sobre un medio de transmisión tal como un medio de transmisión inalámbrico o un medio de transmisión por cable, tal como Internet.
- [0447]** Dependiendo de ciertos requisitos de implementación, las realizaciones de la invención pueden implementarse en hardware o en 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 tienen señales de control legibles electrónicamente almacenadas en el mismo, que cooperan (o son capaces de cooperar) con un sistema informático programable de modo que se lleve a cabo el procedimiento respectivo.
- 40
- 45 **[0448]** Algunas realizaciones de acuerdo con la invención comprenden un soporte de datos no transitorio con señales de control legibles electrónicamente, que son capaces de cooperar con un sistema informático programable, de modo que se lleve a cabo uno de los procedimientos descritos en el presente documento.
- 50 **[0449]** Generalmente, las realizaciones de la presente invención pueden implementarse como un producto de programa informático con un código de programa, siendo el código de programa operativo para realizar uno de los procedimientos, cuando el producto de programa informático se ejecuta en un ordenador. El código de programa puede ser almacenado en un soporte legible por máquina, por ejemplo.
- 55 **[0450]** Otras realizaciones comprenden el programa informático para realizar uno de los procedimientos descritos en el presente documento, almacenado en un soporte legible por máquina.
- [0451]** En otras palabras, una realización del procedimiento de la invención es, por lo tanto, un programa informático que tiene un código de programa para realizar uno de los procedimientos descritos en el presente

documento, cuando el programa informático es ejecutado en un ordenador.

**[0452]** Una realización adicional de los procedimientos de la invención es, por lo tanto, un soporte de datos (o un medio de almacenamiento digital, o un medio legible por ordenador) que comprende, registrado en el mismo, el programa informático para realizar uno de los procedimientos descritos en el presente documento.

**[0453]** Una realización adicional del procedimiento de la invención es, por lo tanto, un flujo de datos o una secuencia de señales que representan el programa informático para realizar uno de los procedimientos descritos en el presente documento. El flujo de datos o la secuencia de señales pueden, por ejemplo, estar configurados para ser transferidos a través de una conexión de comunicación de datos, por ejemplo, a través de Internet.

**[0454]** Una realización adicional comprende un medio de procesamiento, por ejemplo, un ordenador, o un dispositivo lógico programable, configurado para o adaptado para realizar uno de los procedimientos descritos en el presente documento.

**[0455]** Una realización comprende además un ordenador que tiene instalado en el mismo el programa informático para realizar uno de los procedimientos descritos en el presente documento.

**[0456]** En algunas realizaciones, puede usarse un dispositivo lógico programable (por ejemplo, una matriz de puertas programables in situ) para realizar algunas o todas las funcionalidades de los procedimientos descritos en el presente documento. En algunas realizaciones, una matriz de puertas programables in situ podrá cooperar con un microprocesador con el fin de realizar uno de los procedimientos descritos en el presente documento. Generalmente, los procedimientos son realizados preferentemente por cualquier aparato de hardware.

**[0457]** Las realizaciones descritas anteriormente son meramente 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 materia. La intención es, por lo tanto, estar limitada sólo 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.

## Referencias

### **[0458]**

**[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.

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

**[3GP09c]** vSpeech 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, septiembre de 2012.

**[3GP12b]** Adaptive multi-rate (AMR) speech codec; transcoding functions (release 11), 3GPP TS 26.090, 3rd Generation Partnership Project, septiembre de 2012.

**[3GP12c]**, ANSI-C code for the adaptive multi-rate - wideband (AMR-WB) speech codec, 3GPP TS 26.173, 3rd Generation Partnership Project, septiembre de 2012.

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

**[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, septiembre de 2012.

5

[BJH06] I. Batina, J. Jensen, y 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.

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

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

15

[CPK08] Choong Sang Cho, Nam In Park y Hong Kook Kim, A packet loss concealment algorithm robust to burst packet loss for celp- type speech coders, Tech. report, Korea Electronics Technology Institute, Gwang Institute of Science and Technology, 2008, The 23rd International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC 2008).

20

[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, mayo de 2010.

25

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

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

30

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

35

[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.

40

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

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

45

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

50

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

[ISO09] ISO/IEC JTC1/SC29/WG11, Information technology - coding of audiovisual objects - part 3: Audio, ISO/IEC IS 14496-3, International Organization for Standardization, 2009.

55

[ITU03] ITU-T, Wideband coding of speech at around 16 kbit/s using adaptive multirate wideband (amr-wb), Recommendation ITU-T G.722.2, Telecommunication Standardization Sector of ITU, julio de 2003.

- [ITU05] Low-complexity coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss, Recommendation ITU-T G.722.1, Telecommunication Standardization Sector of ITU, mayo de 2005.
- 5 [ITU06a] G.722 Appendix III: A high-complexity algorithm for packet loss concealment for G. 722, ITU-T Recommendation, ITU-T, noviembre de 2006.
- [ITU06b] G.729.1: G.729-based embedded variable bit-rate coder: An 8-32 kbit/s scalable wideband coder bitstream interoperable with g.729, Recommendation ITU-T G.729.1, Telecommunication Standardization Sector of ITU, mayo  
10 de 2006.
- [ITU07] G.722 Appendix IV: A low-complexity algorithm for packet loss concealment with G.722, ITU-T Recommendation, ITU-T, agosto de 2007.
- 15 [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, junio de 2008.
- [ITU08b] G.719: Low-complexity, full-band audio coding for high-quality, conversational applications,  
20 Recommendation ITU-T G.719, Telecommunication Standardization Sector of ITU, junio de 2008.
- [ITU12] G. 729: Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear prediction (cs-acelp), Recommendation ITU-T G.729, Telecommunication Standardization Sector of ITU, junio de 2012.
- 25 [LS01] Pierre Lauber y Ralph Sperschneider, Error concealment for compressed digital audio, Audio Engineering Society Convention 111, no. 5460, septiembre de 2001.
- [Mar01] Rainer Martin, Noise power spectral density estimation based on optimal smoothing and minimum statistics,  
30 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, septiembre de 2003.
- [MC99] R. Martin y R. Cox, New speech enhancement techniques for low bit rate speech coding, in Proc. IEEE  
35 Workshop on Speech Coding (1999), 165-167.
- [MCA99] D. Malah, R. V. Cox, y A. J. Accardi, Tracking speech-presence uncertainty to improve speech enhancement in nonstationary noise environments, Proc. IEEE Int. Conf. on Acoustics Speech and Signal Processing (1999), 789-792.  
40
- [MEP01] Nikolaus Meine, Bernd Edler y Heiko Purnhagen, Error protection and concealment for HILN MPEG-4 parametric audio coding, Audio Engineering Society Convention 110, no. 5300, mayo de 2001.
- [MPC89] Y. Mahieux, J.-P. Petit y A. Charbonnier, Transform coding of audio signals using correlation between  
45 successive transform blocks, Acoustics, Speech, and Signal Processing, 1989. ICASSP-89., 1989 International Conference on, 1989, págs. 2021-2024 vol.3.
- [NMR+12] Max Neuendorf, Markus Multrus, Nikolaus Rettelbach, Guillaume Fuchs, Julien Robilliard, Jérémie Lecomte, Stephan Wilde, Stefan Bayer, Sascha Disch, Christian Helmrich, Roch Lefebvre, Philippe Gournay, Bruno  
50 Bessette, Jimmy Lapierre, Kristopher Kjörling, Heiko Purnhagen, Lars Villemoes, Werner Oomen, Erik Schuijers, Kei Kikuri, Toru Chinen, Takeshi Norimatsu, Chong Kok Seng, Eunmi Oh, Miyoung Kim, Schuyler Quackenbush y Berndhard Grill, MPEG Unified Speech and Audio Coding - The ISO / MPEG Standard for High-Efficiency Audio Coding of all Content Types, Convention Paper 8654, AES, abril de 2012, Presentado en la 132ª Convención Budapest, Hungría.  
55
- [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.

[QD03] Schuyler Quackenbush y Peter F. Driessen, Error mitigation in MPEG-4 audio packet communication systems, Audio Engineering Society Convention 115, no. 5981, octubre de 2003.

[RL06] S. Rangachari y P. C. Loizou, A noise-estimation algorithm for highly non-stationary environments, *Speech Commun.* 48 (2006), 220-231.

[SFB00] V. Stahl, A. Fischer y R. Bippus, Quantile based noise estimation for spectral subtraction and wiener filtering, in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process.* (2000), 1875-1878.

10 [SS98] J. Sohn y W. Sung, A voice activity detector employing soft decision based noise spectrum adaptation, *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, no. págs. 365-368, IEEE, 1998.

[Yu09] Rongshan Yu, A low-complexity noise estimation algorithm based on smoothing of noise power estimation and estimation bias correction, *Acoustics, Speech and Signal Processing*, 2009. ICASSP 2009. IEEE International  
15 Conference on, abril de 2009, págs. 4421-4424.



## REIVINDICACIONES

1. Un aparato para decodificar una señal de audio codificada para obtener una señal de audio reconstruida, en el que el aparato comprende:
- 5 una interfaz receptora (1210) para recibir una o más tramas que comprenden información sobre una pluralidad de muestras de señal de audio de un espectro de señal de audio de la señal de audio codificada, y un procesador (1220) para generar la señal de audio reconstruida, en el que el procesador (1220) está configurado para generar la señal de audio reconstruida desvaneciendo un espectro modificado a un espectro diana, 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 espectro modificado comprende una pluralidad de muestras de señal modificadas, en el que, para cada una de las muestras de señal 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 señal de audio, y
- 10 en el que el procesador (1220) está configurado para no desvanecer el espectro modificado al espectro diana, si la trama actual de las una o más tramas es recibida por la interfaz receptora (1210) y si la trama actual que es recibida por la interfaz receptora (1210) no está alterada.
2. Un aparato de acuerdo con la reivindicación 1, en el que el espectro diana es un espectro de tipo ruido.
3. Un aparato de acuerdo con la reivindicación 2, en el que el espectro de tipo ruido representa ruido blanco.
- 25 4. Un aparato de acuerdo con la reivindicación 2 o 3, en el que el espectro de tipo ruido está conformado.
5. Un aparato de acuerdo con la reivindicación 4, en el que la forma del espectro de tipo ruido depende de un espectro de señal de audio de una señal recibida previamente.
- 30 6. Un aparato de acuerdo con la reivindicación 4 o 5, en el que el espectro de tipo ruido está conformado dependiendo de la forma del espectro de señal de audio.
7. Un aparato de acuerdo con una de las reivindicaciones 4 a 6, en el que el procesador (1220) emplea un factor de inclinación para conformar el espectro de tipo ruido.
- 35 8. Un aparato de acuerdo con la reivindicación 7, en el que el procesador (1220) emplea la fórmula
- $$\text{shaped\_noise}[i] = \text{noise} * \text{power}(\text{tilt\_factor}, i/N)$$
- 40 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  $\text{power}$  es una función de potencia.
- 45 9. Un aparato de acuerdo con una de las reivindicaciones anteriores, en el que el procesador (1220) está configurado para generar el espectro modificado, cambiando un signo de una o más de las muestras de señal de audio del espectro de señal de audio, si la trama actual no es recibida por la interfaz receptora (1210) o si la trama actual que es recibida por la interfaz receptora (1210) está alterada.
- 50 10. Un aparato de acuerdo con una de las reivindicaciones anteriores, en el que cada una de las muestras de señal de audio del espectro de señal de audio está representada por un número real pero no por un número imaginario.
11. Un aparato de acuerdo con una de las reivindicaciones anteriores, en el que las muestras de señal de audio del espectro de señal de audio están representadas en un dominio de transformada de coseno discreta modificada.
- 55 12. Un aparato de acuerdo con una de las reivindicaciones 1 a 10, en el que las muestras de señal de audio del espectro de señal de audio están representadas en un dominio de transformada de seno discreta

modificada.

13. Un aparato de acuerdo con la reivindicación 9, en el que el procesador (1220) está configurado para generar el espectro modificado empleando una función de signo aleatorio que emite de forma aleatoria o pseudoaleatoria bien un primer o bien un segundo valor.

14. Un aparato de acuerdo con una de las reivindicaciones anteriores, en el que el procesador (1220) está configurado para desvanecer el espectro modificado al espectro diana disminuyendo posteriormente un factor de atenuación.

15. Un aparato de acuerdo con una de las reivindicaciones 1 a 3, en el que el procesador (1220) está configurado para desvanecer el espectro modificado al espectro diana aumentando posteriormente un factor de atenuación.

16. Un aparato de acuerdo con una de las reivindicaciones 1 a 12, en el que, si la trama actual no es recibida por la interfaz receptora (1210) o si la trama actual que es recibida por la interfaz receptora (1210) está alterada, el procesador (1220) está configurado para generar la señal de audio reconstruida empleando la fórmula:

$$x[i] = (1 - cum\_damping) * noise[i] + cum\_damping * random\_sign() * x\_old[i]$$

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 señal de audio de la señal de audio codificada,  
 en la que  $random\_sign()$  retorna 1 o -1, y  
 en la que  $noise$  es un vector aleatorio que indica el espectro diana.

17. Un aparato de acuerdo con la reivindicación 16, en el que dicho vector aleatorio  $noise$  es convertido a escala de modo que su media cuadrática sea similar a la media cuadrática del espectro de la señal de audio codificada que está compuesta por una de las tramas que han sido recibidas por la interfaz receptora (1210).

18. Un aparato de acuerdo con una de las reivindicaciones 1 a 15, en el que el procesador (1220) está configurado para generar la señal de audio reconstruida, empleando un vector aleatorio que es convertido a escala de modo que su media cuadrática sea similar a la media cuadrática del espectro de la señal de audio codificada que está compuesta por una de las tramas que han sido recibidas por la interfaz receptora (1210).

19. Un procedimiento para decodificar una señal de audio codificada para obtener una señal de audio reconstruida, en el que el procedimiento comprende:

recibir una o más tramas que comprenden información sobre una pluralidad de muestras de señal de audio de un espectro de señal de audio de la señal de audio codificada, y generar la señal de audio reconstruida,  
 en el que generar la señal de audio reconstruida se lleva a cabo desvaneciendo un espectro modificado a un espectro diana, 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 señal modificadas, en la que, para cada una de las muestras de señal 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 señal de audio, y en el que generar la señal de audio reconstruida se lleva a cabo no desvaneciendo el espectro modificado al espectro diana, si la trama actual de las una o más tramas es recibida y si la trama actual que es recibida no está alterada.

20. Un programa informático adaptado para implementar el procedimiento de la reivindicación 19, cuando es ejecutado en un ordenador o un procesador de señales.

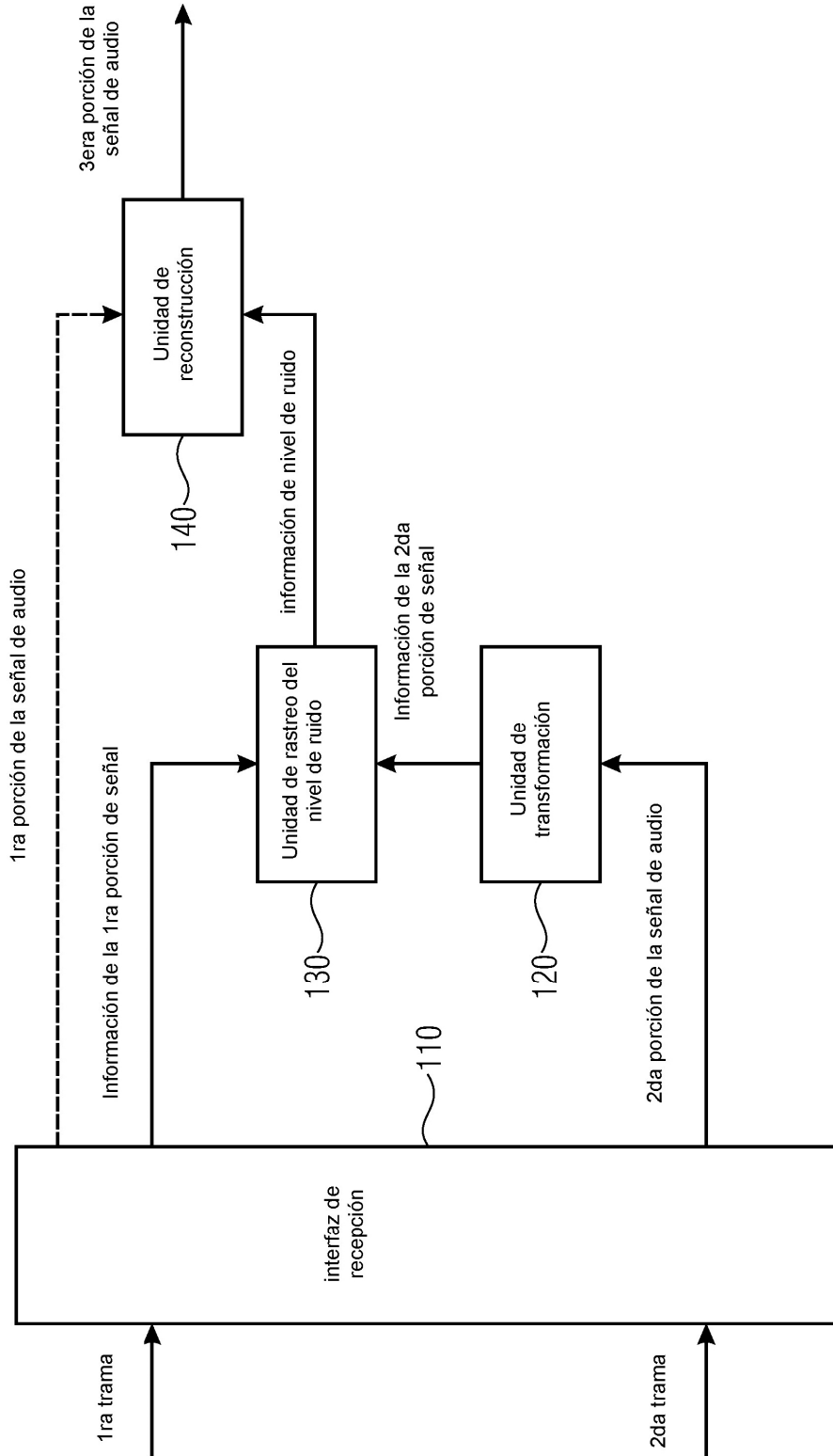


FIG 1A

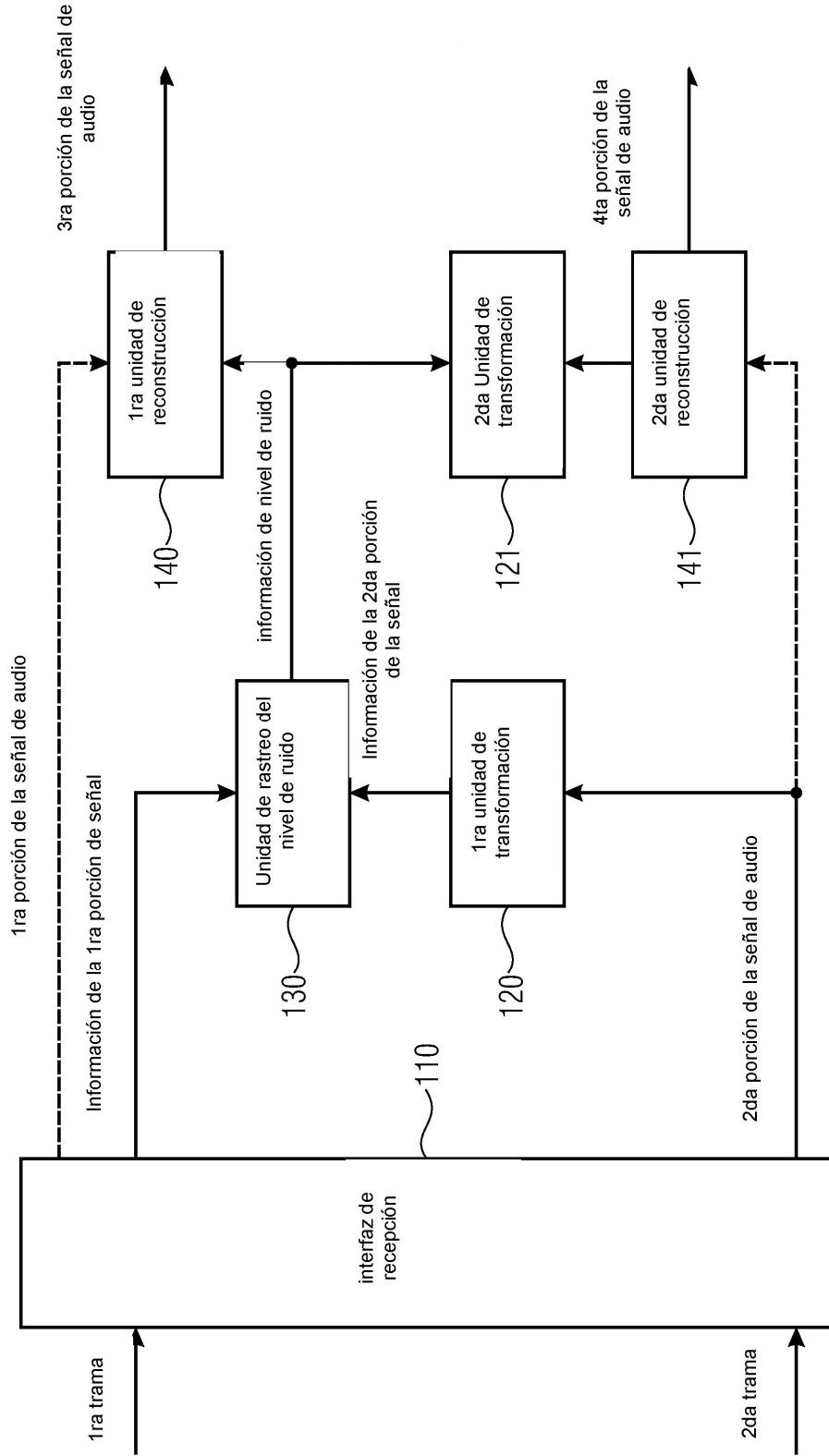


FIG 1B

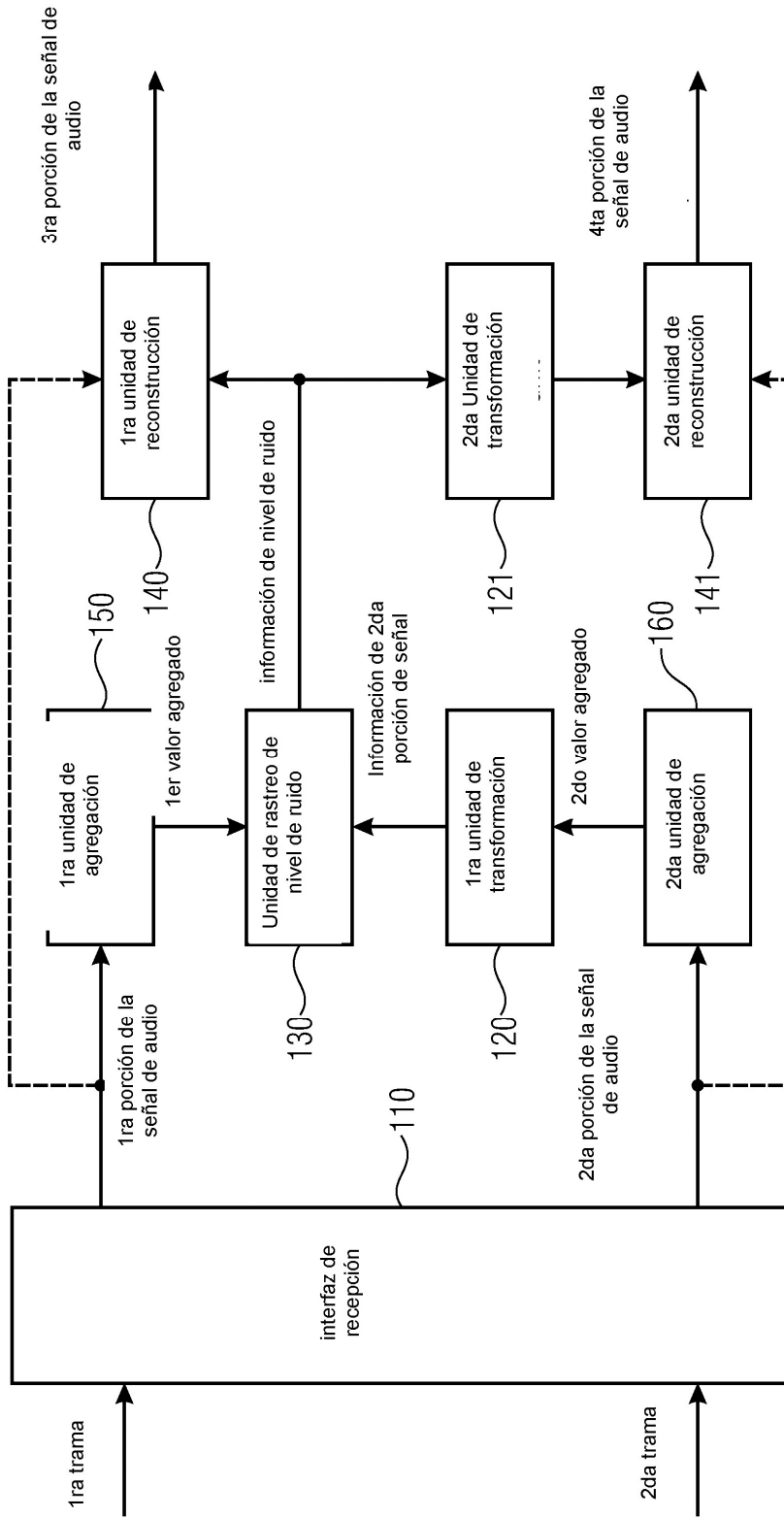


FIG 1C

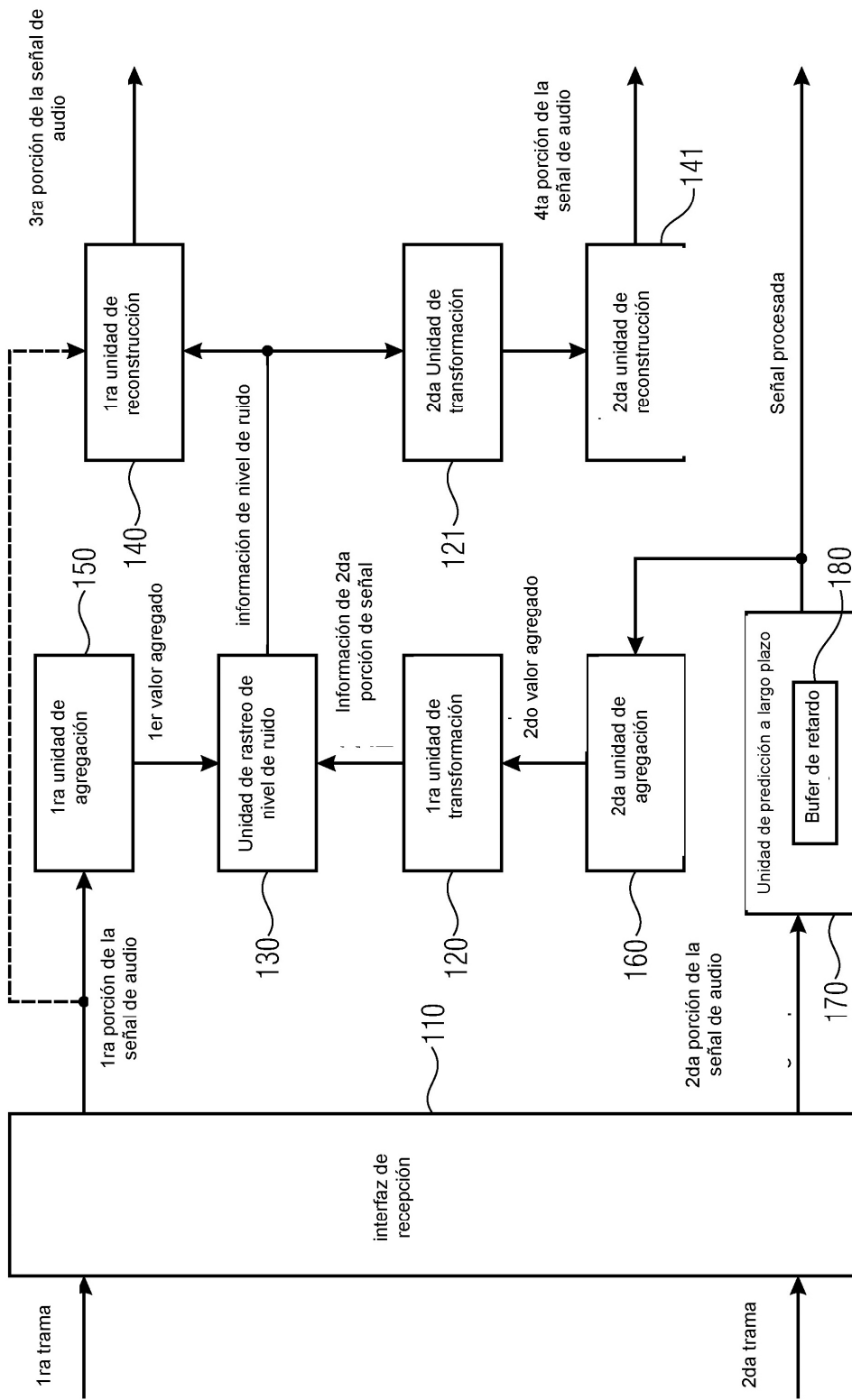


FIG 1D

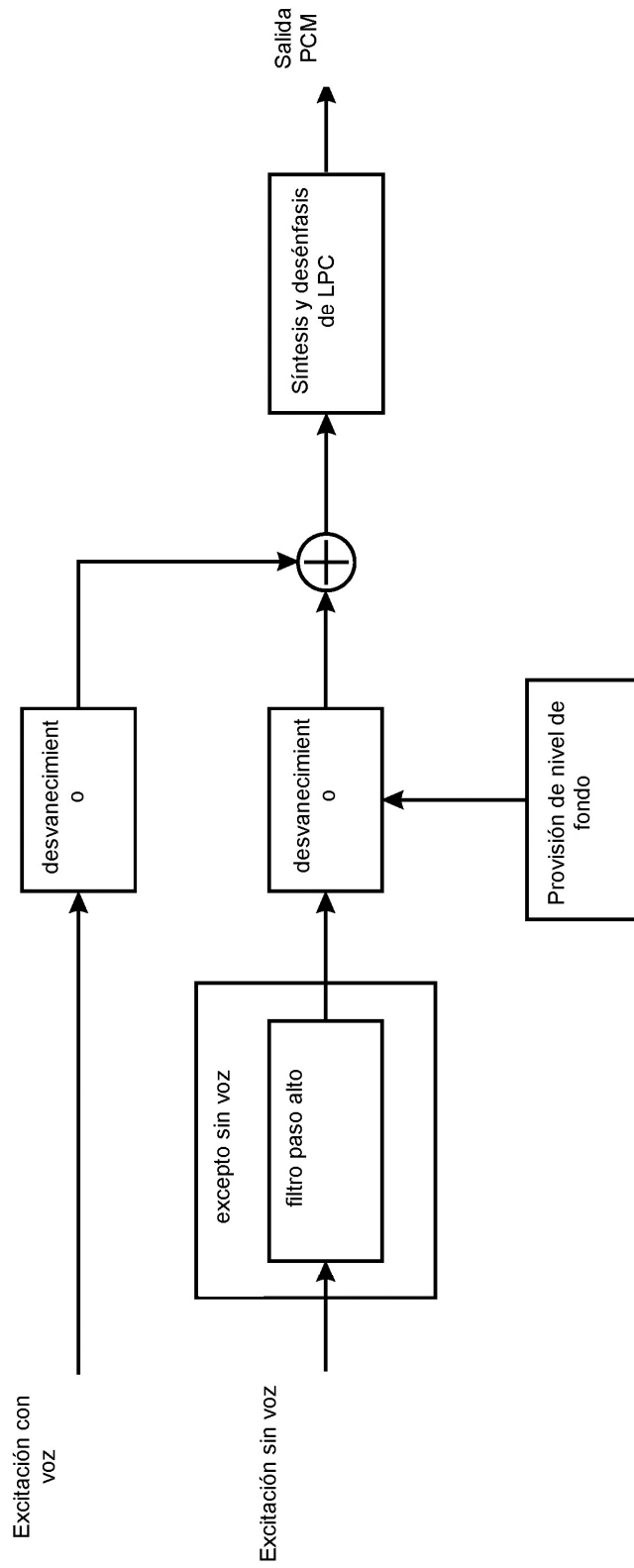
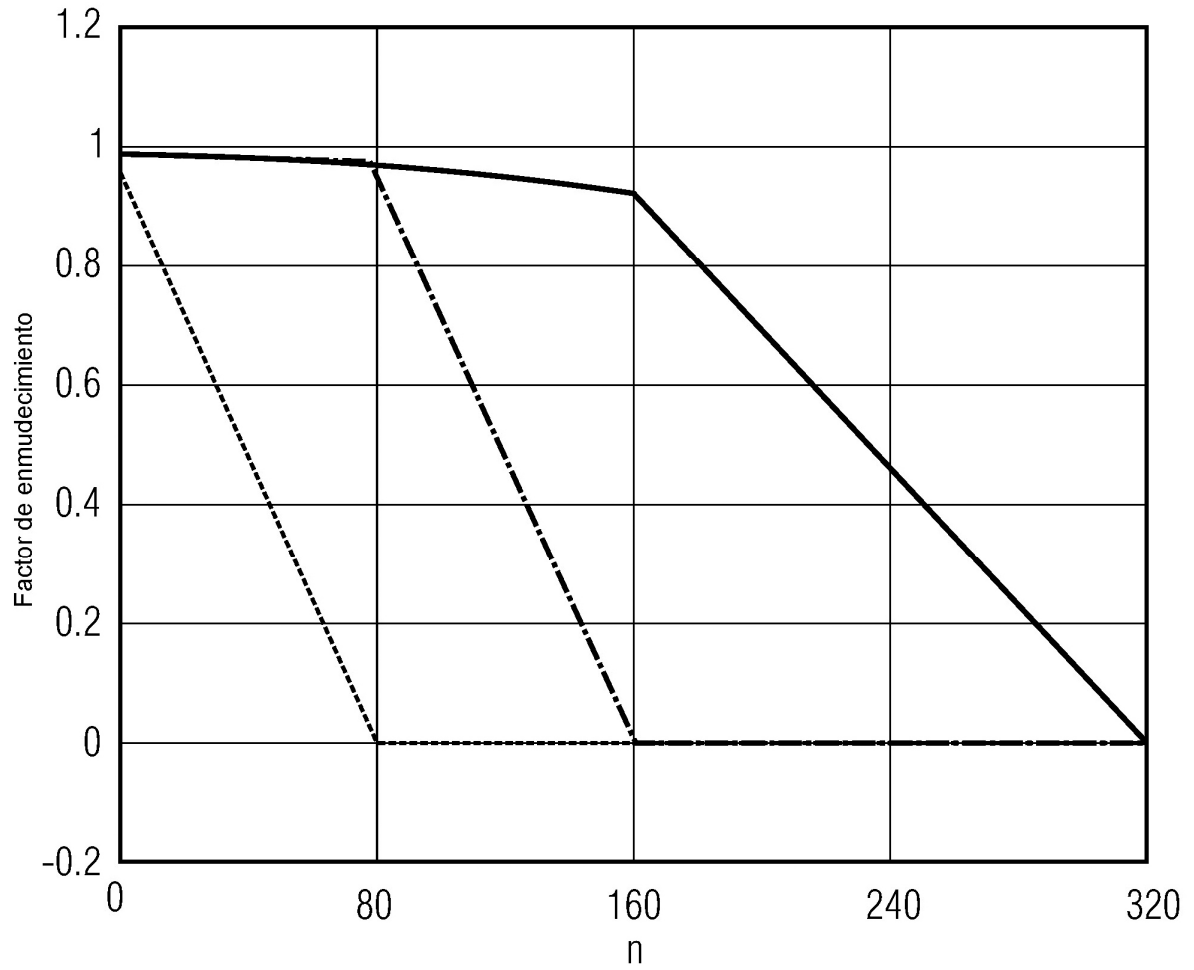


FIG 2



G.722APP/IV\_F07

- ..... transitorio
- · - · - VUV\_TRANSITION
- Otras clases

FIG 3



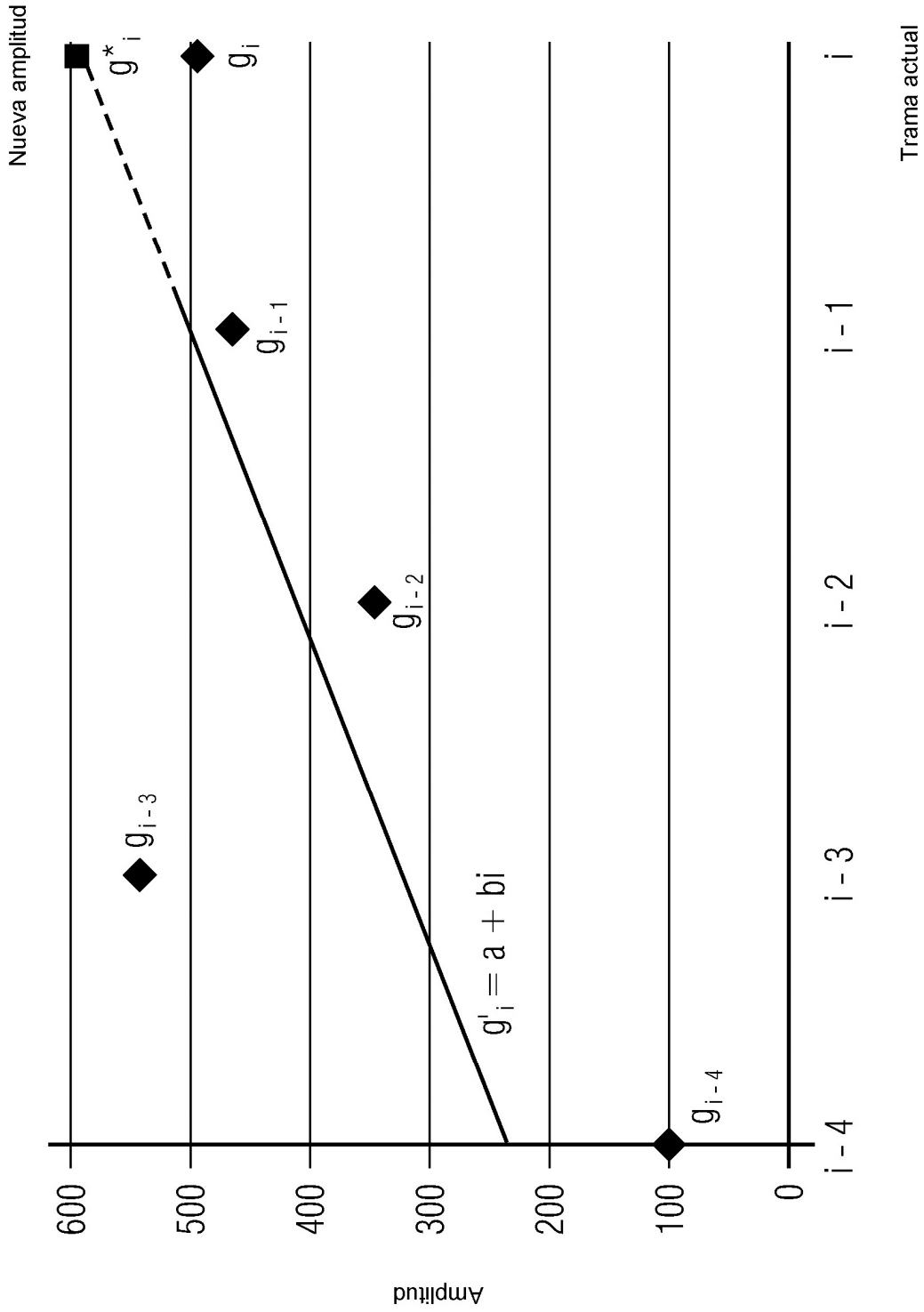


FIG 4

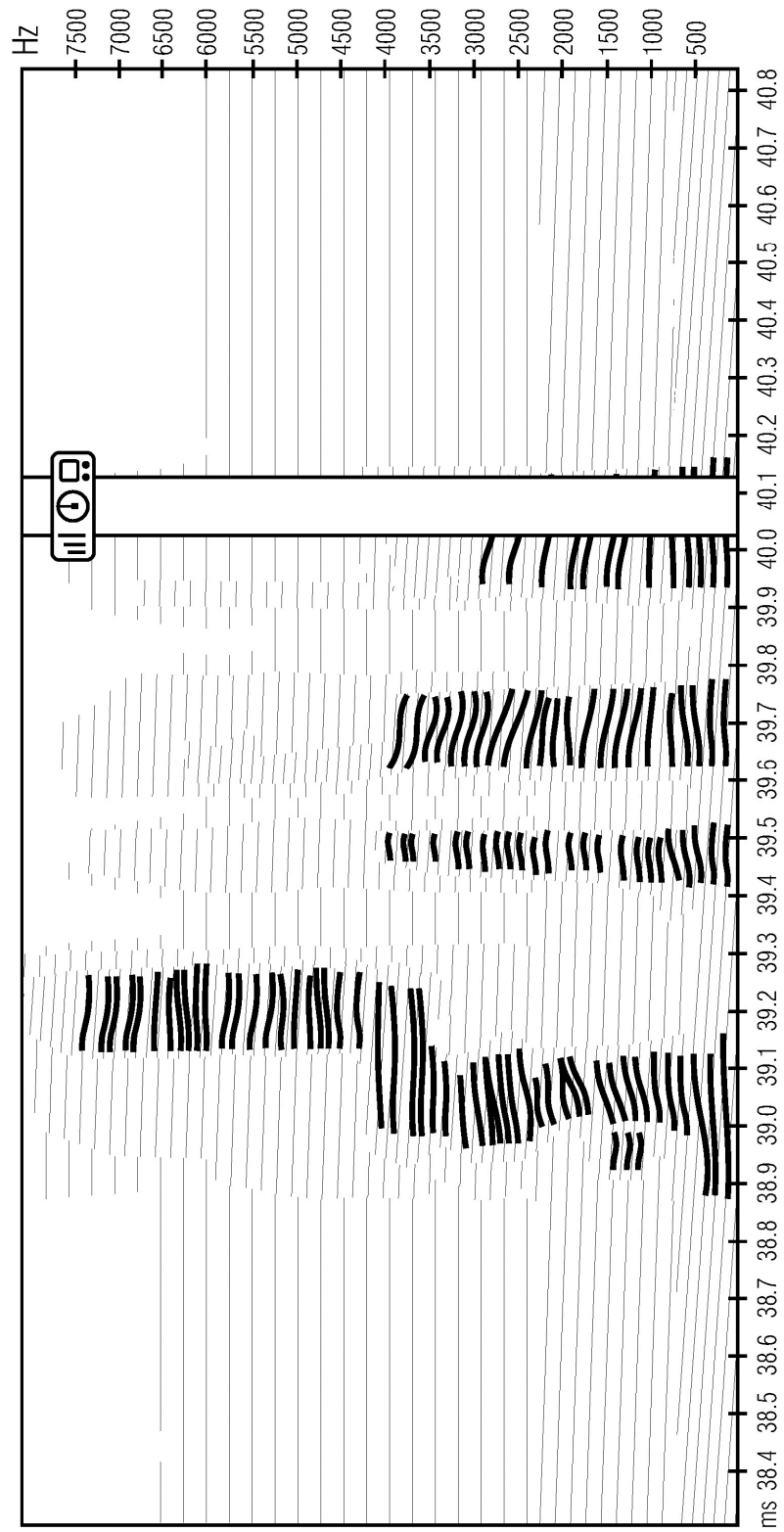


FIG 5

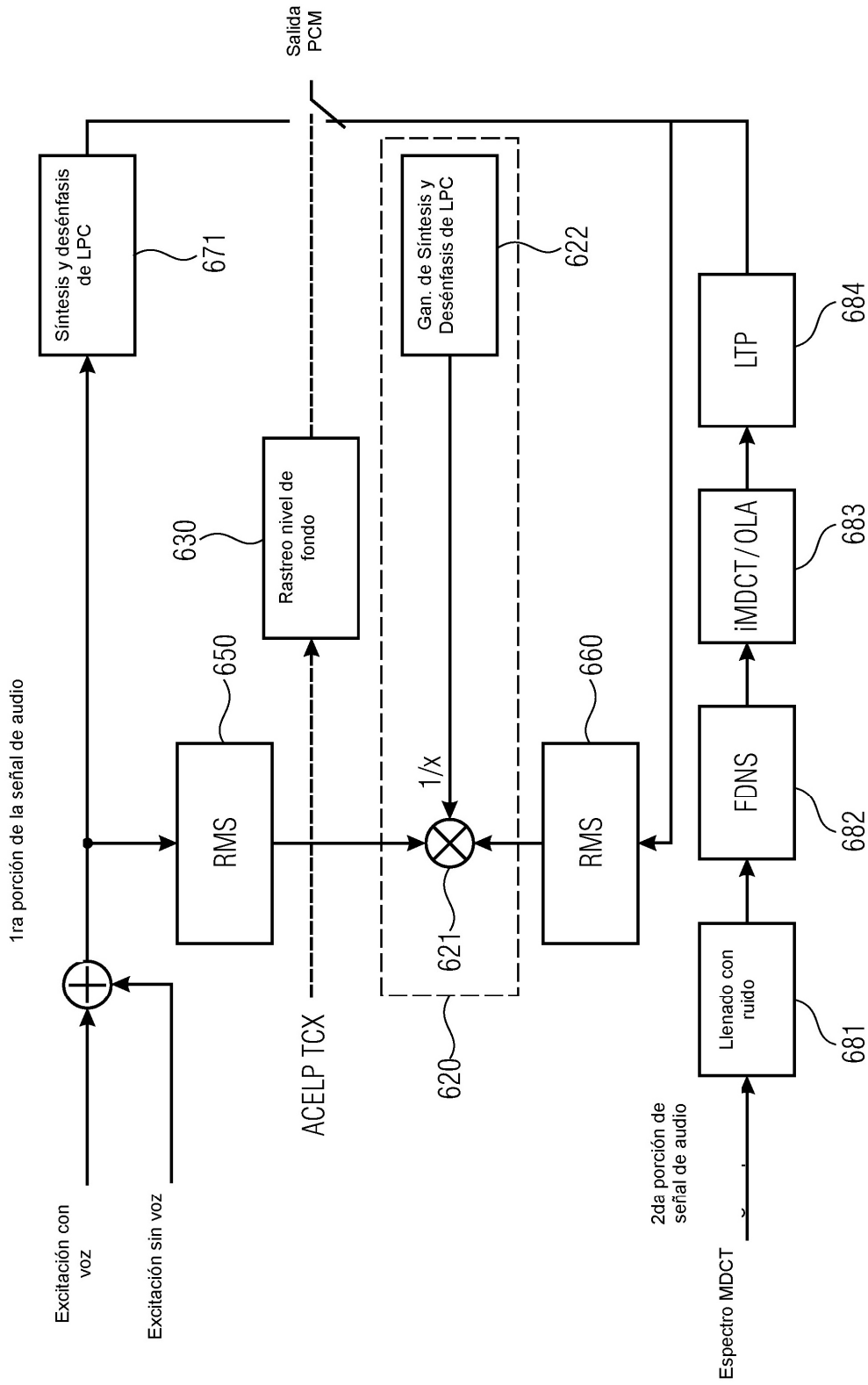


FIG 6

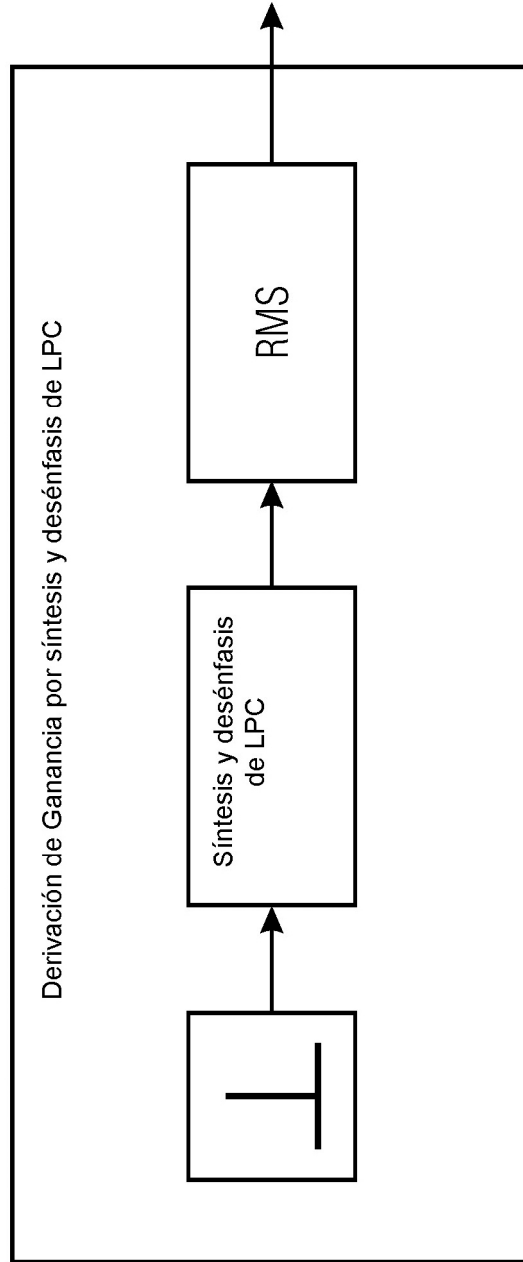


FIG 7

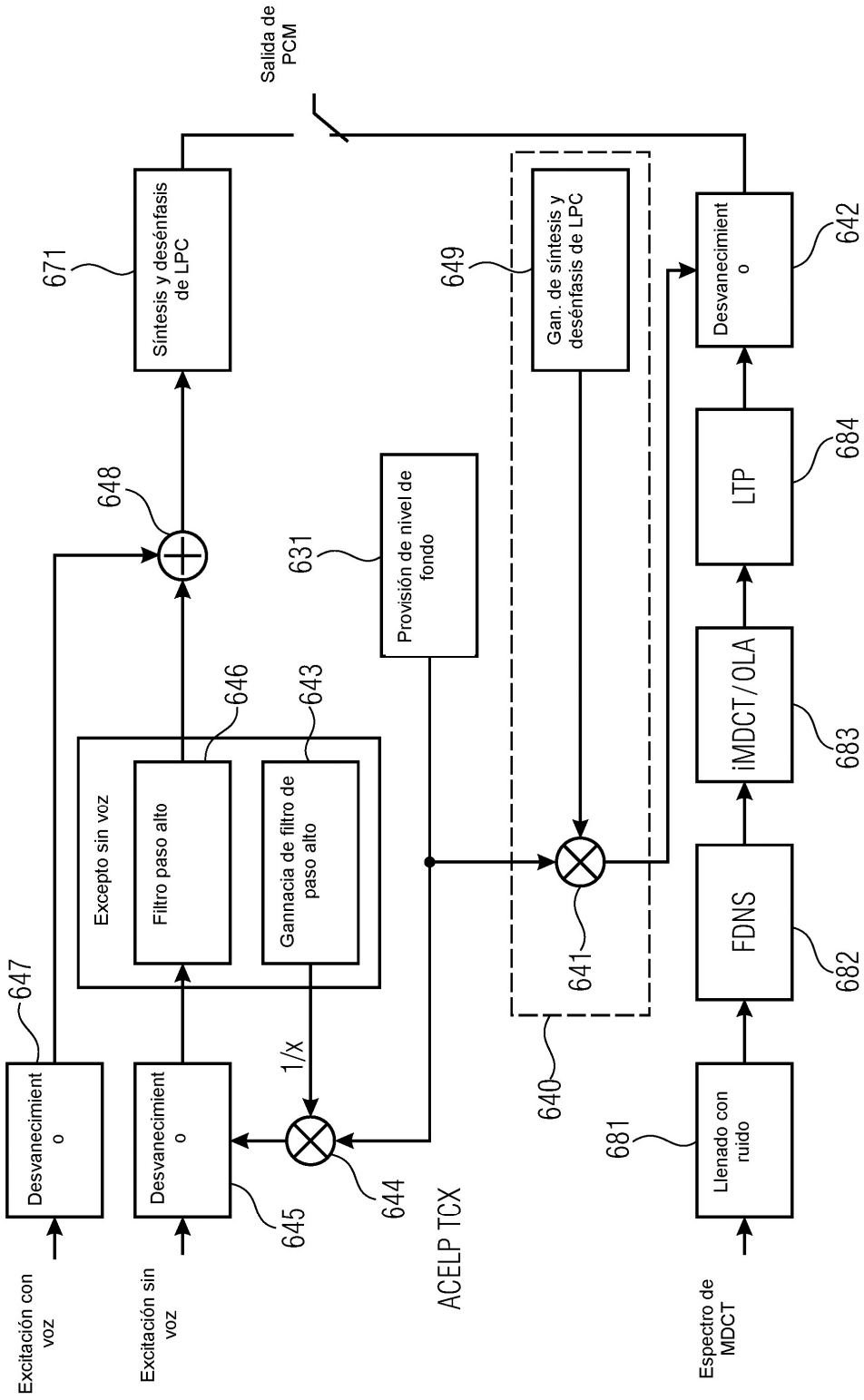


FIG 8

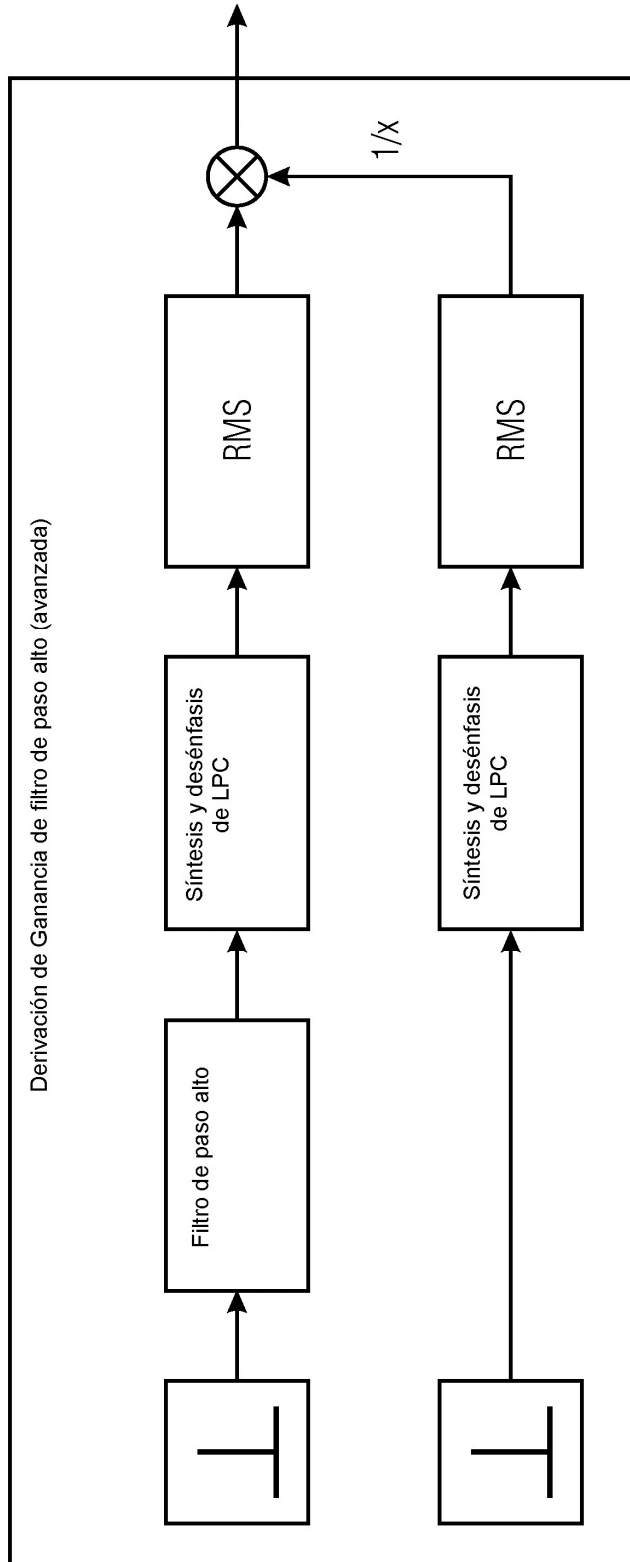


FIG 9

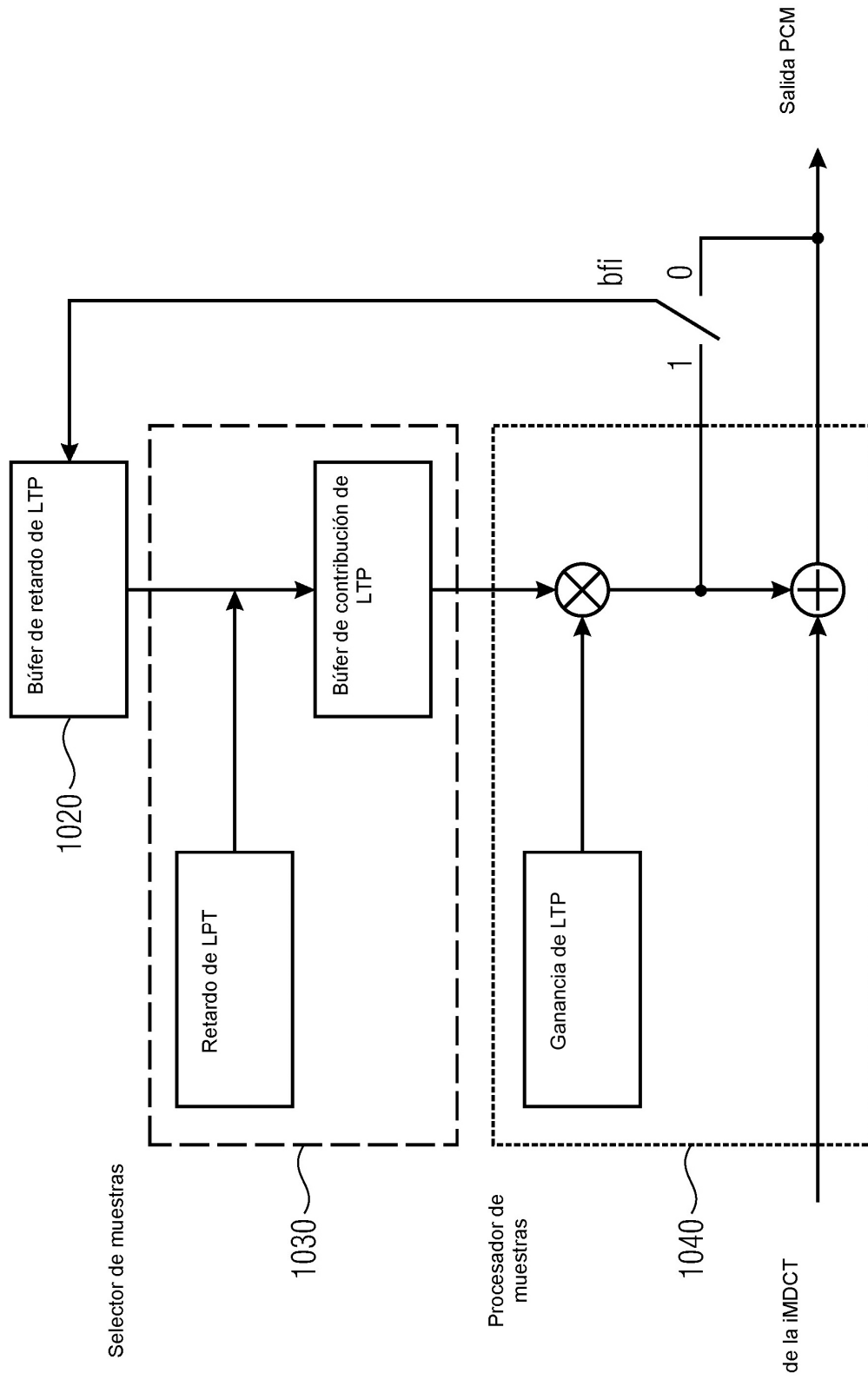


FIG 10

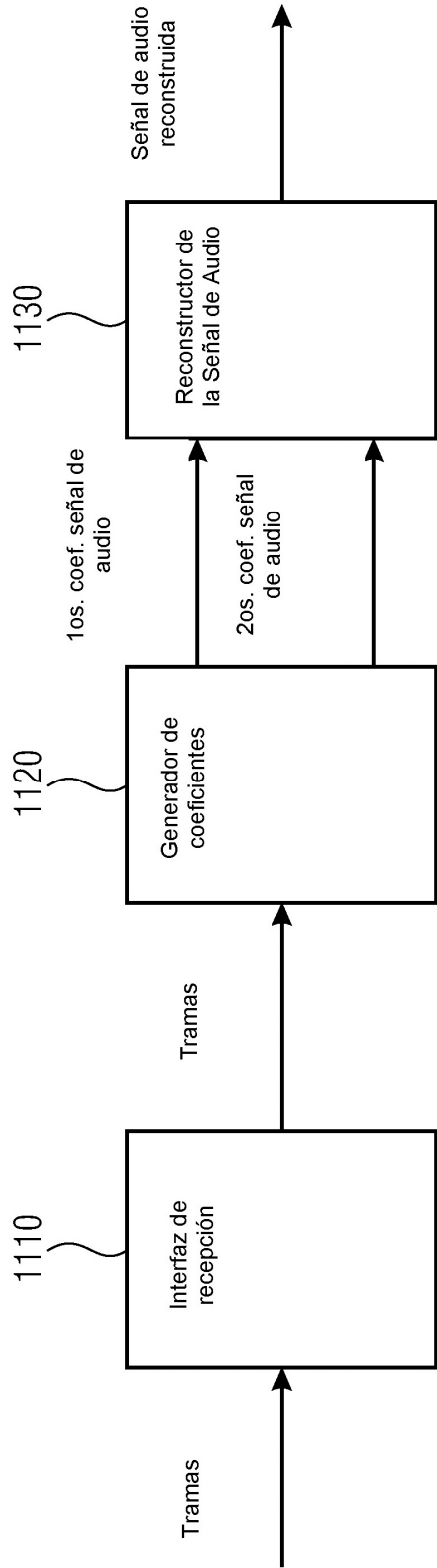


FIG 11

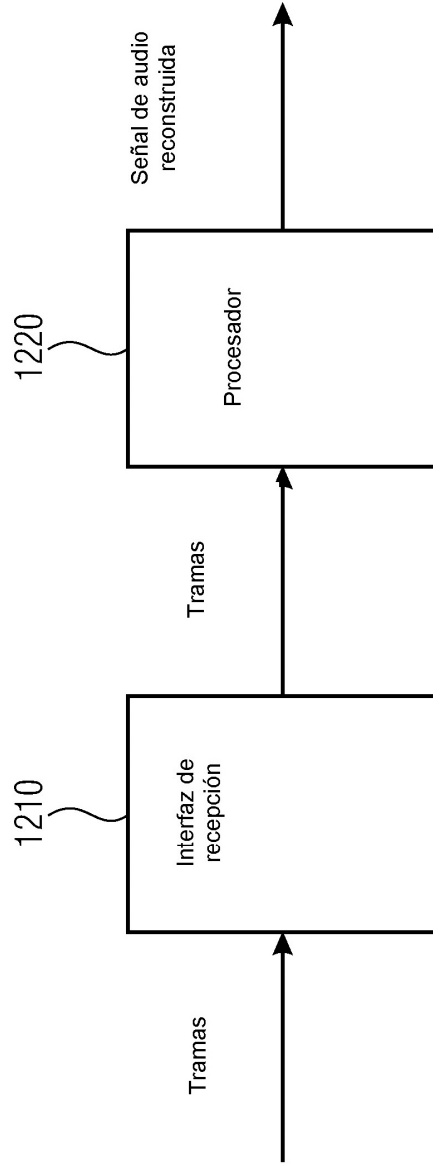


FIG 12



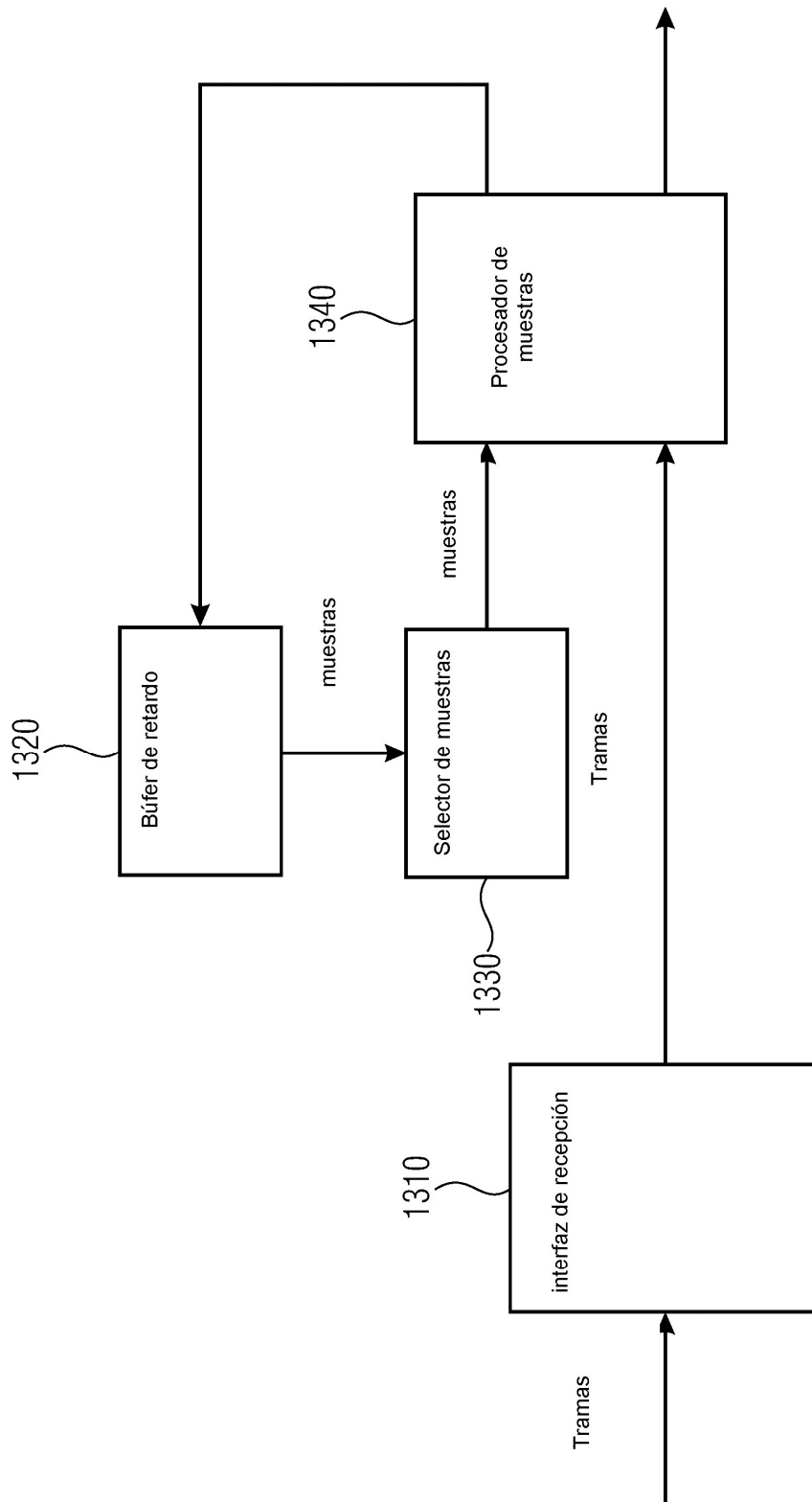


FIG 13

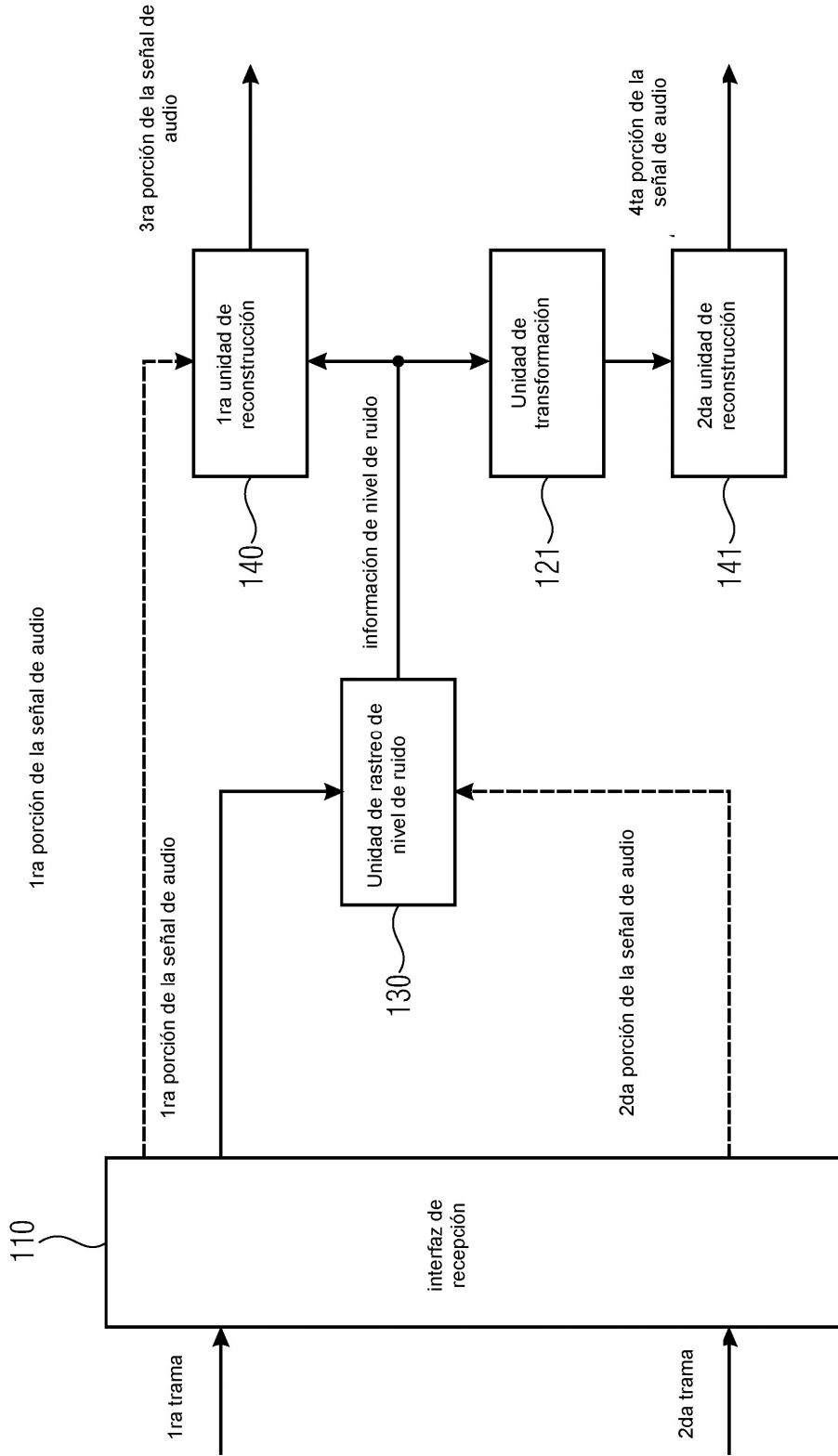


FIG 14