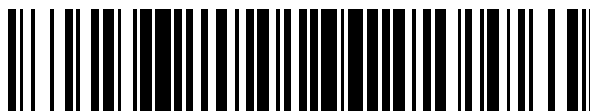


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 625 895**

51 Int. Cl.:

**G10L 19/005** (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **30.05.2003 PCT/CA2003/00830**

87 Fecha y número de publicación internacional: **11.12.2003 WO 2003/102921**

96 Fecha de presentación y número de la solicitud europea: **30.05.2003 E 03727094 (9)**

97 Fecha y número de publicación de la concesión europea: **12.04.2017 EP 1509903**

54 Título: **Método y dispositivo para la ocultación eficiente del borrado de tramas en códecs de voz basados en la predicción lineal**

30 Prioridad:

**31.05.2002 CA 2388439**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**20.07.2017**

73 Titular/es:

**VOICEAGE CORPORATION (100.0%)  
Suite 250 750, chemin Lucerne, Ville Mont-Royal  
Quebec H3R 2H6 , CA**

72 Inventor/es:

**JELINEK, MILAN y  
GOURNAY, PHILIPPE**

74 Agente/Representante:

**ELZABURU SLP, .**

**ES 2 625 895 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

Método y dispositivo para la ocultación eficiente del borrado de tramas en códecs de voz basados en la predicción lineal

**Campo de la invención**

5 La presente invención se refiere a una técnica para codificar digitalmente una señal de sonido, en particular, aunque no de manera exclusiva, una señal de voz, con vistas a la transmisión y/o síntesis de esta señal de sonido. Más específicamente, la presente invención se refiere a la codificación y descodificación robustas de señales de sonido con el fin de mantener un buen rendimiento en caso de una(s) trama(s) borrada(s) debido, por ejemplo, a errores de los canales en sistemas inalámbricos o a paquetes perdidos en aplicaciones de voz a través de redes por paquetes.

**10 Antecedentes de la invención**

La demanda de técnicas eficientes de codificación digital de la voz de banda estrecha y banda ancha con un buen compromiso entre la calidad subjetiva y la velocidad de bits se está incrementando en varias áreas de aplicación, tales como teleconferencias, multimedia y comunicaciones inalámbricas. Hasta hace poco, en aplicaciones de codificación de voz se ha usado principalmente un ancho de banda telefónico limitado a un intervalo de entre 200 y 3.400 Hz. No obstante, las aplicaciones de voz de banda ancha proporcionan un aumento de la inteligibilidad y la naturalidad en la comunicación en comparación con el ancho de banda telefónico convencional. Se ha observado que un ancho de banda en el intervalo de 50 a 7.000 Hz es suficiente para aportar una buena calidad dando la impresión de una comunicación presencial. Para señales de audio genéricas, este ancho de banda aporta una calidad subjetiva aceptable, aunque la misma sigue siendo menor que la calidad de radio FM o del CD que funcionan, respectivamente, en intervalos de 20 a 16.000 Hz y de 20 a 20.000 Hz.

Un codificador de voz convierte una señal de voz en un flujo continuo de bits digital el cual se transmite a través de un canal de comunicaciones o se almacena en un soporte de almacenamiento. La señal de voz se digitaliza, es decir, se muestrea y se cuantifica, habitualmente con 16 bits por muestra. El codificador de voz tiene la función de representar estas muestras digitales con un número de bits menor, aunque manteniendo una buena calidad subjetiva de la voz. El descodificador o sintetizador de voz actúa sobre el flujo continuo de bits transmitido o almacenado, y lo convierte de nuevo a una señal de sonido.

La codificación de *Predicción Lineal con Excitación por Código* (CELP) es una de las mejores técnicas disponibles con el fin de lograr un buen compromiso entre la calidad subjetiva y la velocidad de bits. Esta técnica de codificación es uno de los fundamentos de varias normativas de codificación de la voz, en aplicaciones tanto inalámbricas como por cable. En la codificación CELP, la señal de voz muestreada se procesa en bloques sucesivos de  $L$  muestras, denominados habitualmente *tramas*, donde  $L$  es un número predeterminado que se corresponde típicamente con entre 10 y 30 ms. En cada trama se calcula y transmite un filtro de predicción lineal (LP). El cálculo del filtro de LP requiere típicamente un *anticipo*, un segmento de voz de 5 a 15 ms de la trama sucesiva. La trama de  $L$  muestras se divide en bloques de menor tamaño denominados *subtramas*. Habitualmente, el número de subtramas es tres o cuatro, lo cual da como resultado subtramas de 4 a 10 ms. En cada subtrama, se obtiene habitualmente una señal de excitación a partir de dos componentes, la excitación pasada y la excitación innovadora, con libro de códigos fijo. A la componente formada a partir de la excitación pasada se le hace referencia, normalmente, como excitación de altura tonal o de libro de códigos adaptativo. Los parámetros que caracterizan la señal de excitación se codifican y se transmiten hacia el descodificador, donde la señal de excitación reconstruida se usa como entrada del filtro de LP.

En la medida en la que las aplicaciones principales de codificación de voz de baja velocidad de bits son sistemas inalámbricos de comunicaciones móviles y voz a través de redes por paquetes, entonces, el aumento de la robustez de los códecs de voz en caso de borrado de tramas resulta significativamente importante. En sistemas celulares inalámbricos, la energía de la señal recibida puede presentar desvanecimientos severos y frecuentes que dan como resultado tasas elevadas de errores de bit, y esto se hace más evidente en los límites de las células. En este caso, el descodificador de canales no consigue corregir los errores en la trama recibida y, como consecuencia, el detector de errores usado habitualmente después del descodificador de canales declarará la trama como borrada. En aplicaciones de voz a través de redes por paquetes, la señal de voz se empaqueta, colocándose, habitualmente, una trama de 20 ms en cada paquete. En comunicaciones por conmutación de paquetes, en un router puede producirse un descarte de paquetes si el número de los mismos se hace muy grande, o el paquete puede llegar al receptor después de un retardo elevado y se debe declarar como perdido si su retardo es mayor que la longitud de una memoria intermedia antifluctuaciones en el lado del receptor. En estos sistemas, el códec está sometido típicamente a tasas de borrado de tramas de entre el 3 y el 5%. Además, el uso de codificación de voz de banda ancha es un recurso importante para estos sistemas con el fin de permitirles competir con la PSTN (red telefónica pública conmutada) tradicional que hace uso de las señales heredadas de voz de banda estrecha.

El libro de códigos adaptativo, o el predictor de altura tonal, en la CELP juega un papel importante en el mantenimiento de una calidad de voz elevada con bajas velocidades de bits. No obstante, puesto que el contenido del libro de códigos adaptativo se basa en la señal de tramas pasadas, esto hace que el modelo del códec sea

sensible a la pérdida de tramas. En el caso de tramas borradas o perdidas, el contenido del libro de códigos adaptativo en el descodificador resulta diferente con respecto a su contenido en el codificador. Así, después de ocultar una trama perdida y recibir tramas buenas consiguientes, la señal sintetizada en las tramas buenas recibidas es diferente de la señal de síntesis deseada puesto que se ha cambiado la contribución del libro de códigos adaptativo. El impacto de una trama perdida depende de la naturaleza del segmento de voz en el cual se produjo el borrado. Si el borrado se produce en un segmento estacionario de la señal, entonces puede llevarse a cabo una ocultación eficiente del borrado de la trama y puede reducirse al mínimo el impacto sobre tramas buenas consiguientes. Por otro lado, si el borrado se produce en un inicio de voz o una transición, el efecto del borrado se puede propagar a través de varias tramas. Por ejemplo, si se pierde el comienzo de un segmento sonoro, entonces, en el contenido del libro de códigos adaptativo, faltará el primer periodo de altura tonal. Esto tendrá un efecto importante sobre el predictor de altura tonal en tramas buenas consiguientes, dando como resultado un tiempo prolongado antes de que la señal de síntesis converja en la deseada en el codificador.

El documento WO 01/086637 describe un método para reducir la probabilidad de que se borre una trama de voz durante la transmisión, el cual se basa en el uso de técnicas de corrección directa de errores (FEC). El documento WO 01/086637 apunta más específicamente a la mejora de la calidad de voz producida con el uso de las técnicas de FEC.

### Sumario de la invención

El objetivo de la presente invención se logra por medio de las reivindicaciones independientes. En las reivindicaciones dependientes se definen realizaciones específicas.

La presente invención se refiere a un método de ocultación del borrado de tramas provocado por tramas de una señal de sonido codificada que se han borrado durante la transmisión desde un codificador a un descodificador, y para acelerar la recuperación del descodificador después de que se hayan recibido tramas no borradas de la señal de sonido codificada, según la reivindicación 1.

La presente invención se refiere también a un método para la ocultación del borrado de tramas provocado por tramas borradas durante la transmisión de una señal de sonido codificada bajo la forma de parámetros de codificación de la señal, desde un codificador a un descodificador, y para acelerar la recuperación del descodificador después de que se hayan recibido tramas no borradas de la señal de sonido codificada, según la reivindicación 39.

De acuerdo con la presente invención, se proporciona también un dispositivo para llevar a cabo la ocultación del borrado de tramas provocado por tramas de una señal de sonido codificada, borradas durante la transmisión desde un codificador a un descodificador, y para acelerar la recuperación del descodificador después de que se hayan recibido tramas no borradas de la señal de sonido codificada, según la reivindicación 54.

Según la invención, se proporciona además un dispositivo para la ocultación del borrado de tramas provocado por tramas borradas durante la transmisión de una señal de sonido codificada bajo la forma de parámetros de codificación de la señal, desde un codificador a un descodificador, y para acelerar la recuperación del descodificador después de que se hayan recibido tramas no borradas de la señal de sonido codificada, según la reivindicación 75.

Los anteriores objetivos, ventajas y características, y otros, de la presente invención se pondrán más claramente de manifiesto al leer la siguiente descripción no limitativa de realizaciones ilustrativas de la misma, proporcionadas únicamente a título de ejemplo, en referencia a los dibujos adjuntos.

### Breve descripción de los dibujos

La Figura 1 es un diagrama de bloques esquemático de un sistema de comunicación de voz que ilustra una aplicación de dispositivos de codificación y descodificación de voz según la presente invención;

la Figura 2 es un diagrama de bloques esquemático de un ejemplo de dispositivo de codificación de banda ancha (codificador de AMR-WB);

la Figura 3 es un diagrama de bloques esquemático de un ejemplo de dispositivo de descodificación de banda ancha (descodificador de AMR-WB);

la Figura 4 es un diagrama de bloques simplificado del codificador de AMR-WB de la Figura 2, en donde el módulo diezmador, el módulo de filtro paso-alto y el módulo de filtro de pre-énfasis se han agrupado en un único módulo de pre-procesado, y en donde el módulo de búsqueda de altura tonal en bucle cerrado, el módulo de cálculo de la respuesta a entrada cero, el módulo generador de la respuesta a impulsos, el módulo de búsqueda de excitación innovadora y el módulo de actualización de la memoria se han agrupado en un único módulo de búsqueda de altura tonal en bucle cerrado y de libro de códigos innovador;

la Figura 5 es una extensión del diagrama de bloques de la Figura 4, en la que se han añadido módulos relacionados con una realización ilustrativa de la presente invención;

la Figura 6 es un diagrama de bloques que explica la situación cuando se construye un inicio artificial; y

la Figura 7 es un diagrama esquemático que muestra una realización ilustrativa de una máquina de estados de clasificación de tramas para la ocultación del borrado.

**Descripción detallada de las realizaciones ilustrativas**

5 Aunque las realizaciones ilustrativas de la presente invención se describirán en la siguiente descripción en relación con una señal de voz, debe tenerse en mente que los conceptos de la presente invención se aplican igualmente a otros tipos de señal, en particular, aunque no de forma exclusiva, a otros tipos de señales de sonido.

10 La Figura 1 ilustra un sistema 100 de comunicación de voz que representa el uso de la codificación y la descodificación de voz en el contexto de la presente invención. El sistema 100 de comunicación de voz de la Figura 1 presta soporte a la transmisión de una señal de voz a través de un canal 101 de comunicaciones. Aunque el canal 101 de comunicaciones puede comprender, por ejemplo, un cable, un enlace óptico o un enlace de fibra, el mismo comprende, típicamente, al menos en parte, un enlace de radiofrecuencia. Normalmente, el enlace de radiofrecuencia presta soporte a múltiples comunicaciones de voz simultáneas que requieren recursos compartidos de ancho de banda, tal como puede observarse con los sistemas de telefonía celular. Aunque no se muestra, el canal 101 de comunicaciones se puede sustituir por un dispositivo de almacenamiento en una realización de un solo dispositivo del sistema 100 que graba y almacena la señal de voz codificada, con vistas a una posterior reproducción.

15 En el sistema 100 de comunicación de voz de la Figura 1, un micrófono 102 produce una señal 103 de voz analógica que se suministra a un conversor analógico-a-digital (A/D) 104 para convertirla en una señal 105 de voz digital. Un codificador 106 de voz codifica la señal 105 de voz digital para producir un conjunto de parámetros 107 de codificación de señales los cuales se codifican en formato binario y se entregan a un codificador 108 de canales. El codificador 108 de canales opcional añade redundancia a la representación binaria de los parámetros 107 de codificación de señales, antes de transmitirlos a través del canal 101 de comunicaciones.

20 En el receptor, un descodificador 109 de canales utiliza dicha información redundante en el flujo continuo 111 de bits recibido para detectar y corregir errores de canal que se hayan producido durante la transmisión. Un descodificador 110 de voz convierte el flujo continuo 112 de bits recibido desde el descodificador 109 de canales, de nuevo en un conjunto de parámetros de codificación de señales, y crea, a partir de los parámetros de codificación de señales recuperados, una señal 113 de voz sintetizada digital. La señal 113 de voz sintetizada digital, reconstruida en el descodificador 110 de voz, se convierte a un formato analógico 114 por medio de un conversor digital-a-analógico (D/A) 115 y se reproduce a través de una unidad 116 de altavoz.

25 La realización ilustrativa del método de ocultación eficiente del borrado de tramas que se da a conocer en la presente memoria descriptiva se puede usar con códecs basados en la predicción lineal o bien de banda estrecha o bien de banda ancha. La presente realización ilustrativa se da a conocer en relación con un códec de voz de banda ancha que ha sido normalizado por la Unión Internacional de Telecomunicaciones (ITU) en forma de la Recomendación G.722.2, y que se conoce como el códec de AMR-WB (Códec de Banda Ancha con Multi-Velocidad Adaptativa) [Recomendación G-722.2 de la ITU-T "Wideband coding of speech at around 16 kbit/s using Adaptive Multi-Rate Wideband (AMR-WB)", Ginebra, 2002]. Este códec ha sido seleccionado también por el proyecto de asociación de tercera generación (3GPP) para la telefonía de banda ancha en sistemas inalámbricos de la tercera generación [3GPP TS 26.190, "AMR Wideband Speech Codec: Transcoding Functions", Especificación Técnica del 3GPP]. El AMR-WB puede funcionar con 9 velocidades de bits que van desde 6,6 a 23,85 kbit/s. Para ilustrar la presente invención se usa la velocidad de bits de 12,65 kbit/s.

30 En este caso, debe entenderse que la realización ilustrativa del método de ocultación eficiente del borrado de tramas se podría aplicar a otros tipos de códecs.

35 En las siguientes secciones, se ofrecerá, en primer lugar, una visión general del codificador y el descodificador de AMR-WB. A continuación, se dará a conocer la realización ilustrativa del planteamiento novedoso para mejorar la robustez del códec.

*Visión general del codificador de AMR-WB*

40 La señal de voz muestreada se codifica bloque a bloque por medio del dispositivo 200 de codificación de la Figura 2, el cual se divide en once módulos numerados del 201 al 211.

45 Por lo tanto, la señal 212 de voz de entrada se procesa bloque a bloque, es decir, en los anteriormente mencionados bloques de *L* muestras, denominados tramas.

50 En referencia a la Figura 2, la señal 212 de voz de entrada muestreada se diezma en un módulo diezizador 201. La señal se diezma desde 16 kHz bajándola hasta 12,8 kHz, con el uso de técnicas bien conocidas para aquellos con conocimientos habituales en la materia. El diezmo hace que aumente la eficiencia de codificación, ya que se codifica un ancho de banda de frecuencia menor. Esto reduce también la complejidad algorítmica puesto que se reduce el número de muestras en una trama. Después del diezmo, la trama de 320 muestras de 20 ms se reduce a una trama de 256 muestras (relación de diezmo de 4/5).

A continuación, la trama de entrada se suministra al módulo 202 de pre-procesado, opcional. El módulo 202 de pre-procesado puede constar de un filtro paso-alto con una frecuencia de corte de 50 Hz. El filtro paso-alto 202 elimina los componentes de sonido no deseados por debajo de 50 Hz.

5 La señal pre-procesada, diezmada, se indica con  $s_p(n)$ ,  $n=0, 1, 2, \dots, L-1$ , donde  $L$  es la longitud de la trama (256 con una frecuencia de muestreo de 12,8 kHz). En una realización ilustrativa del filtro 203 de pre-énfasis, la señal  $s_p(n)$  se somete a un pre-énfasis utilizando un filtro que presenta la siguiente función de transferencia:

$$P(z) = 1 - \mu z^{-1}$$

10 donde  $\mu$  es un factor de pre-énfasis con un valor situado entre 0 y 1 (un valor típico es  $\mu = 0,7$ ). La función del filtro 203 de pre-énfasis es potenciar el contenido de frecuencias altas de la señal de voz de entrada. Reduce también el rango dinámico de la señal de voz de entrada, lo que la hace más adecuada para la implementación en coma fija. El pre-énfasis juega también un papel importante en la obtención de una ponderación perceptual total apropiada del error de cuantificación, lo cual contribuye a una mejora de la calidad del sonido. Esto se explicará de forma más detallada posteriormente en la presente.

15 La salida del filtro 203 de pre-énfasis se indica como  $s(n)$ . Esta señal se usa para llevar a cabo el análisis de LP en el módulo 204. El análisis de LP es una técnica bien conocida para aquellos con conocimientos habituales en la materia. En esta implementación ilustrativa, se usa el planteamiento de autocorrelación. En el planteamiento de autocorrelación, en primer lugar la señal  $s(n)$  se enventana usando, típicamente, una ventana Hamming con una longitud del orden de entre 30 y 40 ms. A partir de la señal enventanada se calculan las autocorrelaciones, y se usa la recursión de Levinson-Durbin para calcular coeficientes del filtro de LP,  $a_i$ , donde  $i = 1, \dots, p$ , y en donde  $p$  es el  
20 orden de la LP, el cual, típicamente, es 16 en la codificación de banda ancha. Los parámetros  $a_i$  son los coeficientes de la función de transferencia  $A(z)$  del filtro de LP, que viene dada por la siguiente relación:

$$A(z) = 1 + \sum_{i=1}^p a_i z^{-i}$$

25 El análisis de LP se lleva a cabo en el módulo 204, el cual realiza también la cuantificación y la interpolación de los coeficientes del filtro de LP. Los coeficientes del filtro de LP se transforman, en primer lugar, a otro dominio equivalente más adecuado con fines relativos a la cuantificación y la interpolación. Los dominios de los pares espectrales de líneas (LSP) y los pares espectrales de inmitancia (ISP) son dos dominios en los cuales pueden llevarse a cabo de manera eficiente la cuantificación y la interpolación. Los 16 coeficientes del filtro de LP,  $a_i$ , se pueden cuantificar con una magnitud del orden de entre 30 y 50 bits usando una cuantificación fraccionada o de múltiples fases, o una combinación de las mismas. La finalidad de la interpolación es permitir la actualización de los  
30 coeficientes del filtro de LP cada subtrama al mismo tiempo que se transmiten una vez cada trama, lo cual hace que mejore el rendimiento del codificador sin que aumente la velocidad de bits. Por otro lado, se cree que la cuantificación y la interpolación de los coeficientes del filtro de LP son bien conocidas para aquellos con conocimientos habituales en la materia y, por consiguiente, no se detallarán más en la presente memoria descriptiva.

35 Los siguientes párrafos describirán el resto de las operaciones de codificación llevadas a cabo sobre la base de cada subtrama. En esta implementación ilustrativa, la trama de entrada se divide en 4 subtramas de 5 ms (64 muestras a la frecuencia de muestreo de 12,8 kHz). En la siguiente descripción, el filtro  $A(z)$  indica el filtro de LP interpolado, sin cuantificar, de la subtrama, y el filtro  $\hat{A}(z)$  indica el filtro de LP interpolado cuantificado de la subtrama. El filtro  $\hat{A}(z)$  se suministra cada subtrama a un multiplexor 213 para su transmisión a través de un canal de comunicaciones.

40 En codificadores de análisis-por-síntesis, los parámetros óptimos de altura tonal e innovación se buscan minimizando el error cuadrático medio entre la señal 212 de voz de entrada y una señal de voz sintetizada en un dominio ponderado perceptualmente. La señal ponderada  $s_w(n)$  se calcula en un filtro 205 de ponderación perceptual como respuesta a la señal  $s(n)$  del filtro 203 de pre-énfasis. Se usa un filtro 205 de ponderación perceptual con denominador fijo, adecuado para señales de banda ancha. Un ejemplo de función de transferencia para el filtro 205 de ponderación perceptual viene dado por la siguiente relación:

$$W(z) = A(z\gamma_1) / (1 - \gamma_2 z^{-1}) \quad \text{donde} \quad 0 < \gamma_2 < \gamma_1 \leq 1$$

50 Con el fin de simplificar el análisis de la altura tonal, en primer lugar se estima un retardo de altura tonal en bucle abierto  $T_{OL}$ , en un módulo 206 de búsqueda de altura tonal en bucle abierto, a partir de la señal de voz ponderada  $s_w(n)$ . A continuación, el análisis de altura tonal en bucle cerrado, que se lleva a cabo en un módulo 207 de búsqueda de altura tonal en bucle cerrado, sobre la base de cada subtrama, se restringe en torno al retardo de altura tonal en bucle abierto  $T_{OL}$ , lo cual reduce significativamente la complejidad de la búsqueda de los parámetros de LTP  $T$  (retardo de altura tonal) y  $b$  (ganancia de altura tonal). El análisis de la altura tonal en bucle abierto se lleva a cabo habitualmente en el módulo 206 una vez cada 10 ms (dos subtramas) usando técnicas bien conocidas para aquellos con conocimientos habituales en la materia.

5 En primer lugar, se calcula el vector objetivo  $\mathbf{x}$  para el análisis de LTP (Predicción a Largo Plazo). Habitualmente, esto se lleva a cabo restando, de la señal de voz ponderada  $\mathbf{s}_w(n)$ , la respuesta a entrada cero  $\mathbf{s}_0$  del filtro de síntesis ponderado  $W(z)/\hat{A}(z)$ . Esta respuesta a entrada cero  $\mathbf{s}_0$  se calcula por medio de un módulo 208 de cálculo de respuestas a entrada cero como respuesta al filtro de LP de interpolación cuantificado  $\hat{A}(z)$  del módulo 204 de análisis, cuantificación e interpolación de LP y a los estados iniciales del filtro de síntesis ponderado  $W(z)/\hat{A}(z)$  almacenados en el módulo 211 de actualización de memoria como respuesta a los filtros de LP  $A(z)$  y  $\hat{A}(z)$ , y el vector de excitación  $\mathbf{u}$ . Esta operación es bien conocida para aquellos con conocimientos habituales en la materia y, por consiguiente, no se detallará más en la presente memoria descriptiva.

10 En el generador 209 de respuesta a impulsos se calcula un vector de respuesta a impulsos N-dimensional  $\mathbf{h}$  del filtro de síntesis ponderado  $W(z)/\hat{A}(z)$  usando los coeficientes del filtro de LP  $A(z)$  y  $\hat{A}(z)$  del módulo 204. Nuevamente, esta operación es bien conocida para aquellos con conocimientos habituales en la materia y, por consiguiente, no se detallará más en la presente memoria descriptiva.

15 Los parámetros de altura tonal (o de libro de códigos de altura tonal) en bucle cerrado  $b$ ,  $T$  y  $j$  se calculan en el módulo 207 de búsqueda de altura tonal en bucle cerrado, que usa, como entradas, el vector objetivo  $\mathbf{x}$ , el vector de respuesta a impulsos  $\mathbf{h}$  y el retardo de altura tonal en bucle abierto  $T_{OL}$ .

La búsqueda de la altura tonal consiste en hallar el retardo  $T$  y la ganancia  $b$  de altura tonal óptimos que minimizan un error ponderado cuadrático medio de predicción de la altura tonal, por ejemplo,

$$e^j = \|\mathbf{x} - b^j \mathbf{y}^j\|^2 \quad \text{donde } j=1, 2, \dots, k$$

entre el vector objetivo  $\mathbf{x}$  y una versión filtrada escalada de la excitación pasada.

20 Más específicamente, en la presente implementación ilustrativa, la búsqueda de altura tonal (libro de códigos de altura tonal) está compuesta por tres fases.

25 En la primera fase, se estima un retardo de altura tonal en bucle abierto  $T_{OL}$ , en el módulo 206 de búsqueda de altura tonal en bucle abierto, como respuesta a la señal de voz ponderada  $\mathbf{s}_w(n)$ . Tal como se indica en la descripción anterior, este análisis de altura tonal en bucle abierto se lleva a cabo, habitualmente, una vez cada 10 ms (dos subtramas) usando técnicas bien conocidas para aquellos con conocimientos habituales en la materia.

30 En la segunda fase, se busca un criterio de búsqueda  $C$  en el módulo 207 de búsqueda de altura tonal en bucle cerrado, para retardos enteros de altura tonal en torno al retardo estimado de altura tonal en bucle abierto  $T_{OL}$  (habitualmente  $\pm 5$ ), lo cual simplifica significativamente el procedimiento de búsqueda. Para actualizar el vector de código filtrado  $\mathbf{y}_T$  (este vector se define en la siguiente descripción) sin necesidad de calcular la convolución para cada retardo de altura tonal, se usa un procedimiento sencillo. Un ejemplo de criterio de búsqueda  $C$  viene dado por:

$$C = \frac{\mathbf{x}^t \mathbf{y}_T}{\sqrt{\mathbf{y}_T^t \mathbf{y}_T}} \quad \text{donde } t \text{ indica transposición del vector}$$

35 Una vez que se ha hallado un retardo de altura tonal entero óptimo en la segunda fase, una tercera fase de la búsqueda (módulo 207) somete a prueba, por medio del criterio de búsqueda  $C$ , las fracciones en torno a ese retardo de altura tonal entero óptimo. Por ejemplo, la normativa de AMR-WB usa una resolución de submuestreo de  $\frac{1}{4}$  y  $\frac{1}{2}$ .

40 En señales de banda ancha, la estructura armónica existe solamente hasta una cierta frecuencia, en función del segmento de voz. Así, con el fin de lograr una representación eficiente de la contribución de la altura tonal en segmentos sonoros de una señal de voz de banda ancha, se requiere flexibilidad para variar la cantidad de periodicidad sobre el espectro de banda ancha. Esto se logra procesando el vector código de altura tonal a través de una pluralidad de filtros de conformación de frecuencia (por ejemplo, filtros paso-bajo o pasa-banda). Se selecciona el filtro de conformación de frecuencia que minimiza el error ponderado cuadrático medio  $e^j$ . El filtro seleccionado de conformación de frecuencias se identifica con un índice  $j$ .

45 El índice de libro de códigos de altura tonal  $T$  se codifica y se transmite al multiplexor 213 para su transmisión a través de un canal de comunicaciones. La ganancia de altura tonal  $b$  se cuantifica y se transmite al multiplexor 213. Se usa un bit adicional para codificar el índice  $j$ , suministrándose también este bit adicional al multiplexor 213.

Una vez que se han determinado la altura tonal, o los parámetros de LTP (Predicción a Largo Plazo)  $b$ ,  $T$ , y  $j$ , la siguiente etapa consiste en buscar la excitación innovadora óptima por medio del módulo 210 de búsqueda de excitación innovadora de la Figura 2. En primer lugar, el vector objetivo  $\mathbf{x}$  se actualiza restando la contribución de LTP:

$$x' = x - by_T$$

donde  $b$  es la ganancia de altura tonal e  $y_T$  es el vector de libro de códigos de altura tonal filtrado (la excitación pasada en el retardo  $T$  filtrada con el filtro de conformación de frecuencia seleccionado (índice  $j$ ) y convolucionada con la respuesta a impulsos  $h$ ).

- 5 El procedimiento de búsqueda de la excitación innovadora en la CELP se lleva a cabo en un libro de códigos de innovación para encontrar el vector código de excitación  $c_k$  y la ganancia  $g$  óptimos que minimizan el error cuadrático medio  $E$  entre el vector objetivo  $x'$  y una versión filtrada escalada del vector código  $c_k$ , por ejemplo:

$$E = \|x' - gHc_k\|^2$$

- 10 donde  $H$  es una matriz de convolución triangular inferior obtenida a partir del vector de respuesta a impulsos  $h$ . El índice  $k$  del libro de códigos de innovación correspondiente al vector código  $c_k$  y la ganancia  $g$  óptimos hallados se suministran al multiplexor 213 para su transmisión a través de un canal de comunicaciones.

- 15 Debe indicarse que el libro de códigos de innovación usado es un libro de códigos dinámico que consta de un libro de códigos algebraico seguido por un pre-filtro adaptativo  $F(z)$  que potencia componentes espectrales especiales con el fin de mejorar la calidad de la voz de síntesis, según la patente US 5.444.816 concedida a Adoul et al., el 22 de agosto de 1995. En esta implementación ilustrativa, la búsqueda en el libro de códigos innovador se lleva a cabo en el módulo 210 por medio de un libro de códigos algebraico tal como se describe en las patentes US n.º: 5.444.816 (Adoul et al.) emitida el 22 de agosto de 1995; 5.699.482 concedida a Adoul et al., el 17 de diciembre de 1997; 5.754.976 concedida a Adoul et al., el 19 de mayo de 1998; y 5.701.392 (Adoul et al.) con fecha del 23 de diciembre de 1997.

20 *Visión general del descodificador de AMR-WB*

El descodificador 300 de voz de la Figura 3 ilustra las diversas etapas llevadas a cabo entre la entrada digital 322 (flujo continuo de bits de entrada hacia el demultiplexor 317) y la señal 323 de voz de salida muestreada (salida del sumador 321).

- 25 El demultiplexor 317 extrae los parámetros del modelo de síntesis a partir de la información binaria (flujo continuo 322 de bits de entrada) recibida desde un canal de entrada digital. De cada trama binaria recibida, los parámetros extraídos son:

- los coeficientes de LP interpolados, cuantificados,  $\hat{A}(z)$ , denominados también parámetros de predicción a corto plazo (STP) producidos una vez por cada trama;
- los parámetros de predicción a largo plazo (LTP)  $T$ ,  $b$  y  $j$  (para cada subtrama); y
- 30 • el índice de libro de códigos de innovación  $k$  y la ganancia  $g$  (para cada subtrama).

La señal de voz actual se sintetiza sobre la base de estos parámetros, tal como se explicará más adelante en la presente.

- 35 El libro 318 de códigos de innovación es sensible al índice  $k$  para producir el vector código de innovación  $c_k$ , que se escala según el factor de ganancia descodificado  $g$  a través de un amplificador 324. En la implementación ilustrativa, para producir el vector código innovador  $c_k$  se usa un libro de códigos de innovación, según se describe en las patentes US antes mencionadas n.º 5.444.816; 5.699.482; 5.754.976; y 5.701.392.

El vector código escalado, generado, en la salida del amplificador 324 se procesa a través de un potenciador 305 de la altura tonal, dependiente de la frecuencia.

- 40 La potenciación de la periodicidad de la señal de excitación  $u$  hace que mejore la calidad de los segmentos sonoros. La potenciación de la periodicidad se logra filtrando el vector código innovador  $c_k$  del libro de códigos (fijo) de innovación a través de un filtro de innovación  $F(z)$  (potenciador 305 de la altura tonal), cuya respuesta en frecuencia enfatiza las frecuencias superiores más que las frecuencias inferiores. Los coeficientes del filtro de innovación  $F(z)$  están relacionados con la cantidad de periodicidad en la señal de excitación  $u$ .

- 45 Una forma eficiente, ilustrativa, de obtener los coeficientes del filtro de innovación  $F(z)$  es relacionarlos con la cantidad de contribución de altura tonal en la señal de excitación total  $u$ . Esto da como resultado una respuesta en frecuencia que depende de la periodicidad de las subtramas, en donde las frecuencias superiores tienen un mayor énfasis (mayor pendiente total) para ganancias más altas de la altura tonal. El filtro 305 de innovación tiene el efecto de reducir la energía del vector código de innovación  $c_k$  a frecuencias inferiores cuando la señal de excitación  $u$  es más periódica, lo cual potencia la periodicidad de la señal de excitación  $u$  a frecuencias inferiores más que a las
- 50 frecuencias superiores. Una de las formas sugeridas para el filtro 305 de innovación es la siguiente:

$$F(z) = -\alpha z + 1 - \alpha z^{-1}$$

donde  $\alpha$  es un factor de periodicidad obtenido a partir del nivel de periodicidad de la señal de excitación  $u$ . El factor de periodicidad  $\alpha$  se calcula en el generador 304 de factores de sonoridad. En primer lugar, se calcula un factor de sonoridad  $r_v$  en el generador 304 de factores de sonoridad con:

5

$$r_v = (E_v - E_c) / (E_v + E_c)$$

donde  $E_v$  es la energía del vector código de altura tonal escalado  $bv_T$ , y  $E_c$  es la energía del vector código innovador escalado  $gc_k$ . Es decir:

$$E_v = b^2 v_T^t v_T = b^2 \sum_{n=0}^{N-1} v_T^2(n)$$

y

$$E_c = g^2 c_k^t c_k = g^2 \sum_{n=0}^{N-1} c_k^2(n)$$

10

Obsérvese que el valor de  $r_v$  se sitúa entre -1 y 1 (1 se corresponde con señales puramente sordas y -1 se corresponde con señales puramente sordas).

15

El vector código de altura tonal escalado  $bv_T$  ante mencionado se produce aplicando el retardo de altura tonal  $T$  a un libro 301 de códigos de altura tonal para producir un vector código de altura tonal. A continuación, el vector código de altura tonal se procesa a través de un filtro paso-bajo 302 cuya frecuencia de corte se selecciona en relación con el índice  $j$  del demultiplexor 317, para producir el vector código de altura tonal filtrado  $v_T$ . A continuación, el vector código de altura tonal filtrado  $v_T$  se amplifica entonces según la ganancia de altura tonal  $b$  por medio de un amplificador 326, para producir el vector código de altura tonal escalado  $bv_T$ .

20

En esta implementación ilustrativa, el factor  $\alpha$  se calcula a continuación en el generador 304 de factores de sonoridad con:

$$\alpha = 0.125 (1 + r_v)$$

que se corresponde con un valor de 0 para señales puramente sordas, y 0,25 para señales puramente sonoras.

Por lo tanto, la señal potenciada  $c_f$  se calcula filtrando el vector código innovador escalado  $gc_k$  a través del filtro 305 de innovación ( $F(z)$ ).

25

La señal de excitación potenciada  $u'$  se calcula con el sumador 320 como:

$$u' = c_f + bv_T$$

30

Debe indicarse que este proceso no se lleva a cabo en el codificador 200. Así, resulta esencial actualizar el contenido del libro 301 de códigos de altura tonal usando el valor pasado de la señal de excitación  $u$ , sin potenciación, y almacenado en la memoria 303, para mantener la sincronización entre el codificador 200 y el descodificador 300. Por lo tanto, la señal de excitación  $u$  se usa para actualizar la memoria 303 del libro 301 de códigos de altura tonal, y la señal de excitación potenciada  $u'$  se usa en la entrada del filtro 306 de síntesis de LP.

35

La señal sintetizada  $s'$  se calcula filtrando la señal de excitación potenciada  $u'$  a través del filtro 306 de síntesis de LP que tiene la forma  $1/\hat{A}(z)$ , donde  $\hat{A}(z)$  es el filtro de LP interpolado, cuantificado, en la subtrama actual. Tal como puede observarse en la Figura 3, los coeficientes de LP interpolados, cuantificados,  $\hat{A}(z)$  en la línea 325 del demultiplexor 317 se suministran al filtro 306 de síntesis de LP para ajustar de manera correspondiente los parámetros del filtro 306 de síntesis de LP. El filtro 307 de desénfasis es el inverso del filtro 203 de pre-énfasis de la Figura 2. La función de transferencia del filtro 307 de desénfasis viene dada por

$$D(z) = 1 / (1 - \mu z^{-1})$$

40

donde  $\mu$  es un factor de pre-énfasis con un valor situado entre 0 y 1 (un valor típico es  $\mu = 0,7$ ). También podría usarse un filtro de orden mayor.

El vector  $s'$  se filtra a través del filtro de desénfasis  $D(z)$  307 para obtener el vector  $s_d$ , el cual se procesa a través del filtro paso-alto 308 para eliminar las frecuencias no deseadas por debajo de 50 Hz, y obtener adicionalmente  $s_n$ .



5 El sobremuestreador 309 lleva a cabo el proceso inverso del diezmadador 201 de la Figura 2. En esta realización ilustrativa, el sobremuestreo convierte la frecuencia de muestreo de 12,8 kHz de nuevo en la velocidad de muestreo original de 16 kHz, usando técnicas bien conocidas para aquellos con conocimientos habituales en la materia. La señal de síntesis sobremuestreada se indica como  $\hat{s}$ . A la señal  $\hat{s}$  se le hace referencia también como señal intermedia sintetizada de banda ancha.

10 La señal de síntesis sobremuestreada  $\hat{s}$  no contiene los componentes de frecuencia superiores que se perdieron durante el proceso de diezmadado (módulo 201 de la Figura 2) en el codificador 200. Esto confiere una percepción de paso-bajo a la señal de voz sintetizada. Para restablecer la banda completa de la señal original, en el módulo 310 se lleva a cabo un procedimiento de generación de altas frecuencias, y el mismo requiere una entrada proveniente del generador 304 de factores de sonoridad (Figura 3).

La secuencia de ruido filtrada por pasa-banda, resultante,  $z$ , del módulo 310 de generación de altas frecuencias es sumada por el sumador 321 a la señal de voz sintetizada, sobremuestreada,  $\hat{s}$  para obtener la señal de voz de salida, reconstruida, final,  $s_{out}$  en la salida 323. En la solicitud de patente PCT internacional publicada con el número WO 00/25305 el 4 de mayo de 2000, se describe un ejemplo de proceso de regeneración de altas frecuencias.

15 En la Tabla 1 se proporciona la asignación de bits del códec de AMR-WB a 12,65 kbit/s.

Tabla 1. Asignación de bits en el modo de 12,65 kbit/s

Parámetro	Bits/trama
Parámetros de LP	46
Retardo de Altura Tonal	30 = 9 + 6 + 9 + 6
Filtrado de Altura Tonal	4 = 1 + 1 + 1 + 1
Ganancias	28 = 7 + 7 + 7 + 7
Libro de Códigos Algebraico	144 = 36 + 36 + 36 + 36
Bit de Modo	1
Total	253 bits = 12,65 kbit/s

*Ocultación robusta del borrado de tramas*

20 El borrado de tramas tiene un efecto importante sobre la calidad de la voz sintetizada en sistemas digitales de comunicación de voz, especialmente cuando funciona en entornos inalámbricos y redes por conmutación de paquetes. En los sistemas celulares inalámbricos, la energía de la señal recibida puede presentar frecuentes desvanecimientos severos que dan como resultado tasas elevadas de errores de bit, y esto se hace más evidente en los límites de las células. En este caso, el decodificador de canales no consigue corregir los errores en la trama recibida y, como consecuencia, el detector de errores usado habitualmente después del decodificador de canales  
 25 declarará la trama como borrada. En aplicaciones de voz a través de redes por paquetes, tales como el Protocolo de Voz por Internet (VoIP), la señal de voz se empaqueta, de manera que habitualmente se ubica una trama de 20 ms en cada paquete. En comunicaciones por conmutación de paquetes, en un router puede producirse un descarte de paquetes si el número de paquetes se hace muy grande, o el paquete puede llegar al receptor después de un retardo prolongado y debería ser declarado como perdido si su retardo es mayor que la longitud de una memoria intermedia antifuertuaciones en el lado del receptor. En estos sistemas, el códec está sometido típicamente a tasas de borrado de tramas del 3 al 5%.

Básicamente, el problema del procesado de borrado de tramas (FER) es doble. En primer lugar, cuando llega un indicador de trama borrada, la trama que falta se debe generar usando la información enviada en la trama previa y estimando la evolución de la señal en la trama ausente. El éxito de la estimación depende, no solamente de la estrategia de ocultación, sino también del lugar en el que se produce el borrado en la señal de voz. En segundo lugar, debe garantizarse una transición suave cuando se recupera el funcionamiento normal, es decir, cuando llega la primera trama buena después de un bloque de tramas borradas (una o más). Esta no es una tarea trivial en la medida en la que la síntesis real y la síntesis estimada pueden evolucionar de manera diferente. Por tanto, cuando  
 35 llega la primera trama buena, el decodificador está desincronizado con respecto al codificador. El motivo principal es que los codificadores de baja velocidad de bits se basan en la predicción de la altura tonal y, durante las tramas borradas, la memoria del predictor de altura tonal ya no es la misma que la del codificador. El problema se amplifica cuando se borran muchas tramas consecutivas. En cuanto a la ocultación, la dificultad de la recuperación del procesado normal depende del tipo de señal de voz en la que se produjo el borrado.

45 El efecto negativo del borrado de tramas se puede reducir significativamente adaptando la ocultación y la recuperación del procesado normal (recuperación posterior) al tipo de la señal de voz en el que se produce el borrado. Con este fin, es necesario clasificar cada trama de voz. Esta clasificación se puede realizar en el codificador y la misma es transmitida. Alternativamente, se puede estimar en el decodificador.

Para obtener la ocultación y la recuperación óptimas, hay algunas características críticas de la señal de voz que se deben controlar de forma cuidadosa. Estas características críticas son la energía de la señal o la amplitud, la cantidad de periodicidad, la envolvente espectral y el periodo de la altura tonal. En el caso de una recuperación de voz sonora, puede lograrse una mejora adicional con un control de fase. Con un ligero aumento en la velocidad de bits, se pueden cuantificar y transmitir unos pocos parámetros suplementarios para obtener un mejor control. Si no hay disponible ningún ancho de banda adicional, los parámetros se pueden estimar en el descodificador. Con estos parámetros controlados, la ocultación y la recuperación del borrado de tramas se pueden mejorar significativamente, especialmente mejorando la convergencia de la señal descodificada hacia la señal real en el codificador, y aliviando el efecto de disparidad entre el codificador y el descodificador cuando se recupera el procesado normal.

5 En la presente realización ilustrativa de la presente invención, se dan a conocer métodos para una ocultación eficiente del borrado de tramas, y métodos para extraer y transmitir parámetros que mejorarán el rendimiento y la convergencia en el descodificador, en las tramas que suceden a una trama borrada. Estos parámetros incluyen dos o más de los siguientes: clasificación de las tramas, energía, información de sonoridad e información de fase. Además, se dan a conocer métodos para extraer dichos parámetros en el descodificador en caso de que no sea posible la transmisión de bits adicionales. Finalmente, se dan a conocer también métodos para mejorar la convergencia del descodificador en tramas buenas que suceden a una trama borrada.

10 Las técnicas de ocultación del borrado de tramas de acuerdo con la presente realización ilustrativa se han aplicado al códec de AMR-WB antes descrito. Este códec servirá como escenario de ejemplo para la implementación de los métodos de ocultación del FER en la siguiente descripción. Tal como se ha explicado anteriormente, la señal 212 de voz de entrada hacia el códec presenta una frecuencia de muestreo de 16 kHz, aunque se diezma a una frecuencia de muestreo de 12,8 kHz antes del procesado adicional. En la presente realización ilustrativa, el procesado del FER se realiza sobre la señal diezmada.

15 La Figura 4 proporciona un diagrama de bloques simplificado del codificador 400 de AMR-WB. En este diagrama de bloques simplificado, el diezgador 201, el filtro paso-alto 202 y el filtro 203 de pre-énfasis se agrupan juntos en el módulo 401 de preprocesado. Además, el módulo 207 de búsqueda en bucle cerrado, el módulo 208 de cálculo de respuesta a entrada cero, el módulo 209 de cálculo de la respuesta a impulsos, el módulo 210 de búsqueda de la excitación innovadora, y el módulo 211 de actualización de la memoria se agrupan en un módulo 402 de búsqueda de libro de códigos de innovación y de la altura tonal en bucle cerrado. Esta agrupación se lleva a cabo para simplificar la introducción de los nuevos módulos relacionados con la realización ilustrativa de la presente invención.

20 La Figura 5 es una extensión del diagrama de bloques de la Figura 4, en la que se han añadido módulos relacionados con la realización ilustrativa de la presente invención. En estos módulos añadidos 500 a 507, se calculan parámetros adicionales, y los mismos se cuantifican y se transmiten con la finalidad de mejorar la ocultación del FER y la convergencia y la recuperación del descodificador después de tramas borradas. En la presente realización ilustrativa, estos parámetros incluyen clasificación de la señal, energía, e información de fase (la posición estimada del primer pulso glotal en una trama).

25 En las siguientes secciones, se aportarán de forma detallada el cálculo y la cuantificación de estos parámetros adicionales, y los mismos se pondrán más claramente de manifiesto en referencia a la Figura 5. Entre estos parámetros, se tratará más detalladamente la clasificación de las señales. En las secciones sucesivas, se explicará la ocultación eficiente del FER usando estos parámetros adicionales para mejorar la convergencia.

#### 30 *Clasificación de las señales para la ocultación y la recuperación del FER*

35 La idea básica que subyace tras el uso de una clasificación de la voz para la reconstrucción de una señal en presencia de tramas borradas, consiste en el hecho de que la estrategia de ocultación ideal es diferente para segmentos de voz cuasi-estacionarios y para segmentos de voz con características que varían rápidamente. Mientras el procesado óptimo de tramas borradas en segmentos de voz no estacionarios se puede resumir como una convergencia rápida de los parámetros de codificación de la voz a las características del ruido ambiente, en el caso de una señal cuasi-estacionaria, los parámetros de codificación de la voz no varían drásticamente y se pueden mantener casi sin variaciones durante varias tramas borradas adyacentes antes de desvanecerse. Además, el método óptimo para una recuperación de la señal tras un bloque borrado de tramas varía con la clasificación de la señal de voz.

40 La señal de voz se puede clasificar aproximadamente como sonora, sorda y pausas. La voz sonora contiene una cantidad importante de componentes periódicos y se puede dividir adicionalmente en las siguientes categorías: inicios sonoros, segmentos sonoros, transiciones sonoras y finales sonoros. Un inicio sonoro se define como un comienzo de un segmento de voz sonoro después de una pausa o un segmento sordo. Durante los segmentos sonoros, los parámetros de la señal de voz (envolvente espectral, periodo de altura tonal, relación de componentes periódicos y no periódicos, energía) varían lentamente de una trama a otra. Una transición sonora se caracteriza por variaciones rápidas de una voz sonora, tales como una transición entre vocales. Los finales sonoros se caracterizan por una disminución gradual de energía y sonoridad en la terminación de segmentos sonoros.

45 Las partes sordas de la señal se caracterizan por carecer del componente periódico y se pueden dividir

adicionalmente en tramas inestables, en las que la energía y el espectro cambian rápidamente, y tramas estables en las que estas características permanecen relativamente estables. Las tramas restantes se clasifican como silencio. Las tramas de silencio comprenden todas las tramas sin voz activa, es decir, también tramas con solamente ruido en caso de que haya presencia de ruido de fondo.

- 5 No todas las clases antes mencionadas requieren un procesado aparte. Por tanto, con fines relativos a las técnicas de ocultación de los errores, algunas de las clases de la señal se agrupan entre sí.

#### *Clasificación en el codificador*

10 Cuando en el flujo continuo de bits hay un ancho de banda disponible para incluir la información de clasificación, la clasificación se puede realizar en el codificador. Esto presenta varias ventajas. La más importante es que, en los codificadores de voz, se dispone normalmente de un anticipo. El anticipo permite estimar la evolución de la señal en la trama sucesiva, y, consecuentemente, la clasificación se puede realizar teniendo en cuenta el comportamiento futuro de la señal. En general, cuanto más prolongado sea el anticipo, mejor podrá ser la clasificación. Otra de las ventajas es una reducción de la complejidad, en la medida en la que la mayor parte del procesado de la señal necesario para la ocultación del borrado de tramas es necesaria de todos modos para la codificación de la voz. Finalmente, existe también la ventaja de trabajar con la señal original en lugar de la señal sintetizada.

15 La clasificación de las tramas se realiza teniendo en mente la estrategia de ocultación y recuperación. En otras palabras, cualquier trama se clasifica de tal manera que la ocultación puede ser óptima si falta la trama sucesiva, o la recuperación puede resultar óptima si se hubiese perdido la trama previa. Algunas de las clases usadas para el procesado del FER no necesitan ser transmitidas, ya que se pueden deducir sin ambigüedades en el descodificador. En la presente realización ilustrativa, se usan cinco (5) clases distintas, y estas se definen de la manera siguiente:

- La clase SORDA comprende todas las tramas de voz sordas y todas las tramas sin voz activa. Una trama de final sonora también se puede clasificar como SORDA si su terminación tiende a ser sorda, y la ocultación diseñada para tramas sordas se puede usar para la trama sucesiva en caso de que esta se pierda.
- 25 • La clase TRANSICIÓN SORDA comprende tramas sordas con un posible inicio sonoro en su terminación. No obstante, el inicio sigue siendo demasiado corto o sigue sin estar construido suficientemente bien como para usar la ocultación diseñada para tramas sonoras. La clase TRANSICIÓN SORDA puede suceder únicamente a una trama clasificada como SORDA o TRANSICIÓN SORDA.
- La clase TRANSICIÓN SONORA comprende tramas sonoras con características sonoras relativamente débiles. Son típicamente tramas sonoras con características que varían rápidamente (transiciones entre vocales) 30 o finales sonoros que duran la trama completa. La clase TRANSICIÓN SONORA puede suceder únicamente a una trama clasificada como TRANSICIÓN SONORA, SONORA o INICIO.
- La clase SONORA comprende tramas sonoras con características estables. Esta clase puede suceder únicamente a una trama clasificada como TRANSICIÓN SONORA, SONORA o INICIO.
- 35 • La clase INICIO comprende todas las tramas sonoras con características estables que suceden a una trama clasificada como SORDA o TRANSICIÓN SORDA. Las tramas clasificadas como INICIO se corresponden con tramas de inicio sonoras en donde el inicio ya está suficientemente bien construido con vistas al uso de la ocultación diseñada para tramas sonoras perdidas. Las técnicas de ocultación usadas para un borrado de tramas tras la clase INICIO, son las mismas que tras la clase SONORA. La diferencia se encuentra en la estrategia de recuperación. Si se ha perdido una trama de clase INICIO (es decir, una trama buena SONORA llega después de un borrado, pero la última trama buena antes del borrado era SORDA), puede usarse una técnica especial para reconstruir artificialmente el inicio perdido. Este escenario puede observarse en la Figura 6. En la siguiente exposición se describirán más detalladamente las técnicas de reconstrucción artificial del inicio. Por otro lado, si una trama buena de INICIO llega después de un borrado, y la última trama buena antes del borrado era SORDA, este procesado especial no es necesario, ya que el inicio no se ha perdido (no se ha encontrado en la trama 45 perdida).

El diagrama de estados de la clasificación se expone en líneas generales en la Figura 7. Si el ancho de banda disponible es suficiente, la clasificación se realiza en el codificador y se transmite usando 2 bits. Tal como puede observarse a partir de la Figura 7, la clase TRANSICIÓN SORDA y la clase TRANSICIÓN SONORA se pueden agrupar entre sí en la medida en la que se pueden diferenciar sin ambigüedades en el descodificador (la 50 TRANSICIÓN SORDA puede suceder únicamente a tramas SORDAS o de TRANSICIÓN SORDA, la TRANSICIÓN SONORA puede suceder únicamente a tramas de INICIO, SONORAS, o de TRANSICIÓN SONORA). Para la clasificación se usan los siguientes parámetros: una correlación normalizada  $r_x$ , una medición de la inclinación espectral  $ef$ , una relación señal/ruido  $snr$ , un medidor de la estabilidad de la altura tonal  $pc$ , una energía de trama relativa de la señal en la terminación de la trama actual  $E_s$  y un contador de cruces por cero  $zc$ . Tal como puede observarse en el siguiente análisis detallado, el cálculo de estos parámetros hace uso del anticipo disponible al 55 máximo posible para tener en cuenta el comportamiento de la señal de voz también en la siguiente trama.

La correlación normalizada  $r_x$  se calcula como parte del módulo 206 de búsqueda de la altura tonal en bucle abierto

de la Figura 5. Este módulo 206 habitualmente da salida a la estimación de la altura tonal en bucle abierto cada 10 ms (dos veces por trama). En este caso, se usa también para dar salida a las mediciones de la correlación normalizada. Estas correlaciones normalizadas se calculan sobre la señal de voz ponderada actual  $s_w(n)$  y la señal de voz ponderada pasada, en el retardo de la altura tonal en bucle abierto. Para reducir la complejidad, la señal de voz ponderada  $s_w(n)$  se diezma con un factor de 2 antes del análisis de la altura tonal en bucle abierto, bajándola hasta la frecuencia de muestreo de 6.400 Hz [3 GPP TS 26.190, "AMR Wideband Speech Codec: Transcoding Functions", Especificación Técnica del 3GPP]. La correlación promedio  $r_x$  se define como

$$\bar{r}_x = 0.5(r_x(1) + r_x(2)) \quad (1)$$

donde  $r_x(1)$ ,  $r_x(2)$  son respectivamente la correlación normalizada de la segunda mitad de la trama actual y del anticipo. En esta realización ilustrativa, se usa un anticipo de 13 ms a diferencia de la normativa de AMR-WB que hace uso de 5 ms. La correlación normalizada  $r_x(k)$  se calcula de la manera siguiente:

$$r_x(k) = \frac{r_{xy}}{\sqrt{r_{xx} \cdot r_{yy}}} \quad (2)$$

donde

$$r_{xy} = \sum_{i=0}^{L_k-1} x(t_k+i) \cdot x(t_k+i-p_k)$$

$$r_{xx} = \sum_{i=0}^{L_k-1} x^2(t_k+i)$$

$$r_{yy} = \sum_{i=0}^{L_k-1} x^2(t_k+i-p_k)$$

Las correlaciones  $r_x(k)$  se calculan usando la señal de voz ponderada  $s_w(n)$ . Los instantes  $t_k$  están relacionados con el comienzo de la trama actual y son iguales, respectivamente, a 64 y 128 muestras a la velocidad o frecuencia de muestreo de 6,4 kHz (10 y 20 ms). Los valores  $p_k = T_{OL}$  son las estimaciones seleccionadas de la altura tonal en bucle abierto. La longitud del cálculo de autocorrelación  $L_k$  depende del periodo de la altura tonal. A continuación se resumen los valores de  $L_k$  (para la frecuencia de muestreo de 6,4 kHz):

$$L_k = 40 \text{ muestras para } p_k \leq 31 \text{ muestras}$$

$$L_k = 62 \text{ muestras para } p_k \leq 61 \text{ muestras}$$

$$L_k = 115 \text{ muestras para } p_k > 61 \text{ muestras}$$

Estas longitudes garantizan que la longitud del vector correlacionado comprenda por lo menos un periodo de altura tonal que ayude a una detección robusta de la altura tonal en bucle abierto. Para periodos de altura tonal prolongados ( $p_1 > 61$  muestras),  $r_x(1)$  y  $r_x(2)$  son idénticas, es decir, únicamente se calcula una correlación puesto que los vectores correlacionados son suficientemente largos para que el análisis sobre el anticipo ya no sea necesario.

El parámetro de inclinación espectral  $e_t$  contiene la información sobre la distribución de energía en frecuencia. En la presente realización ilustrativa, la inclinación espectral se estima como una relación entre la energía concentrada en frecuencias bajas y la energía concentrada en frecuencias altas. No obstante, también se puede estimar de diferentes maneras, tales como una relación entre los dos primeros coeficientes de autocorrelación de la señal de voz.

La Transformada de Fourier discreta se usa para llevar a cabo el análisis espectral en el módulo 500 de análisis espectral y de estimación de energía espectral de la Figura 5. El análisis de frecuencia y el cálculo de la inclinación se realizan dos veces por cada trama. Se usa una Transformada Rápida de Fourier (FFT) de 256 puntos con un solapamiento del 50 por ciento. Las ventanas del análisis se sitúan de manera que se aproveche la totalidad del anticipo. En esta realización ilustrativa, el comienzo de la primera ventana se sitúa 24 muestras después del comienzo de la trama actual. La segunda ventana se sitúa 128 muestras más allá. Pueden utilizarse ventanas diferentes para ponderar la señal de entrada para el análisis de frecuencia. En la presente realización ilustrativa se ha usado una raíz cuadrada de una ventana Hamming (que es equivalente a una ventana sinusoidal). Esta ventana resulta particularmente bien adecuada para métodos de suma con solapamiento. Por lo tanto, este análisis espectral particular se puede usar en un algoritmo opcional de supresión de ruido, basado en el análisis/síntesis de suma con solapamiento y sustracción espectral.

La energía en altas frecuencias y en bajas frecuencias se calcula en el módulo 500 de la Figura 5 siguiendo las bandas críticas perceptivas. En la presente realización ilustrativa, cada banda crítica se considera hasta el siguiente número [J. D. Johnston, "Transform Coding of Audio Signals Using Perceptual Noise Criteria", *IEEE Jour. on Selected Areas in Communications*, vol. 6, n.º 2, págs. 314 a 323]:

- 5 Bandas críticas = {100,0; 200,0; 300,0; 400,0; 510,0; 630,0; 770,0; 920,0; 1.080,0; 1.270,0; 1.480,0; 1.720,0; 2.000,0; 2.320,0; 2.700,0; 3.150,0; 3.700,0; 4.400,0; 5.300,0; 6.350,0} Hz.

La energía en frecuencias superiores se calcula en el módulo 500 como el promedio de las energías de las dos últimas bandas críticas:

$$\bar{E}_h = 0.5(e(18) + e(19)) \quad (3)$$

- 10 donde las energías de las bandas críticas  $e(i)$  se calculan como una suma de las energías de los compartimentos (*bins*) dentro de la banda crítica, promediadas por el número de los compartimentos.

- 15 La energía en frecuencias inferiores se calcula como el promedio de las energías en las 10 primeras bandas críticas. Las bandas críticas centrales se han excluido del cálculo, para mejorar la discriminación entre frecuencias con alta concentración de energía en bajas frecuencias (generalmente sonoras) y con alta concentración de energía en altas frecuencias (generalmente sordas). En medio, el contenido de energía no es característico de ninguna de las clases e incrementaría la confusión a la hora de tomar decisiones.

- 20 En el módulo 500, la energía en bajas frecuencias se calcula de manera diferente para periodos largos de la altura tonal y periodos cortos de altura tonal. Para segmentos sonoros de voz femenina, puede aprovecharse la estructura armónica del espectro para incrementar la discriminación sonoridad-sordez. Así, para periodos cortos de la altura tonal,  $\bar{E}_l$  se calcula a nivel de compartimentos y, en el sumatorio, se tienen en cuenta únicamente compartimentos de frecuencia suficiente próximos a los armónicos de la voz, es decir

$$\bar{E}_l = \frac{1}{cnt} \cdot \sum_{i=0}^{24} e_b(i) \quad (4)$$

- 25 donde  $e_b(i)$  son las energías de compartimento en los primeros 25 compartimentos de frecuencia (no se considera el componente DC). Obsérvese que estos 25 compartimentos se corresponden con las primeras 10 bandas críticas. En el sumatorio anterior, únicamente los términos relacionados con los compartimentos cuya proximidad a los armónicos más cercanos es mayor que un cierto umbral de frecuencia, son diferentes de cero. El contador *cnt* equivale al número de esos términos diferentes de cero. El umbral para que un compartimento se incluya en la suma se ha fijado en 50 Hz, es decir, únicamente se tienen en cuenta compartimentos más próximos que 50 Hz a los armónicos más cercanos. Por lo tanto, si la estructura es armónica en frecuencias bajas, en la suma se incluirá solamente un término de alta energía. Por otro lado, si la estructura no es armónica, la selección de los términos será aleatoria y la suma será menor. De este modo, pueden detectarse incluso sonidos sordos con un alto contenido de energía en bajas frecuencias. Este procesado no se puede realizar para periodos más largos de la altura tonal, en la medida en la que la resolución frecuencial no es suficiente. El valor de altura tonal de umbral es 128 muestras, correspondiente a 100 Hz. Esto significa que, para periodos de altura tonal mayores que 128 muestras y, también, para sonidos sordos a priori (es decir, cuando  $\bar{r}_x + r_e < 0,6$ ), la estimación de la energía de bajas frecuencias se realiza por cada banda crítica y se calcula como

$$\bar{E}_l = \frac{1}{10} \cdot \sum_{i=0}^9 e(i) \quad (5)$$

- 40 El valor  $r_e$ , calculado en un módulo 501 de corrección de la correlación normalizada y de estimación de ruido, es una corrección añadida a la correlación normalizada en presencia de ruido de fondo por el siguiente motivo. En presencia de ruido del fondo, la correlación normalizada promedio se reduce. No obstante, a efectos de la clasificación de las señales, esta reducción no debería afectar a la decisión sobre sonoridad-sordez. Se ha observado que la dependencia entre esta reducción  $r_e$  y la energía de ruido de fondo total en dB es aproximadamente exponencial, y se puede expresar usando la siguiente relación

$$r_e = 2.4492 \cdot 10^{-d} \cdot e^{0.1696 \cdot N_{dB}} - 0.022$$

- 45 donde  $N_{dB}$  significa

$$N_{dB} = 10 \cdot \log_{10} \left( \frac{1}{20} \sum_{l=0}^{19} n(l) \right) - g_{dB}$$

En este caso,  $n(i)$  son las estimaciones de la energía de ruido para cada banda crítica normalizadas de la misma manera que  $e(i)$ , y  $g_{dB}$  es el nivel de supresión de ruido máximo en dB permitido para la rutina de reducción de ruido. No se permite que el valor  $r_e$  sea negativo. Debe indicarse que, cuando se usa un buen algoritmo de reducción de ruido y  $g_{dB}$  es suficientemente alto,  $r_e$  es prácticamente igual a cero. Únicamente es relevante cuando se deshabilita la reducción de ruido o si el nivel de ruido de fondo es significativamente mayor que la reducción permitida máxima. la influencia de  $r_e$  se puede precisar multiplicando este término por una constante.

Finalmente, las energías resultantes de frecuencias inferiores y superiores se obtienen restando una energía de ruido estimada, con respecto a los valores  $\bar{E}_l$  y  $\bar{E}_h$  calculados anteriormente. Es decir

$$E_h = \bar{E}_h - f_c \cdot N_h \quad (6)$$

$$E_l = \bar{E}_l - f_c \cdot N_l \quad (7)$$

donde  $N_h$  y  $N_l$  son las energías de ruido promedias en las dos (2) últimas bandas críticas y las diez (10) primeras bandas críticas, respectivamente, calculadas usando ecuaciones similares a las Ecuaciones (3) y (5), y  $f_c$  es un factor de corrección precisado de manera que estas mediciones permanezcan próximas a una constante con la variación del nivel de ruido de fondo. En esta realización ilustrativa, el valor de  $f_c$  se ha fijado a 3.

La inclinación espectral  $e_i$  se calcula en el módulo 503 de estimación de inclinación espectral usando la relación:

$$e_i = \frac{E_l}{E_h} \quad (8)$$

y se promedia en el dominio de los dB para los dos (2) análisis de frecuencia llevados a cabo por cada trama:

$$e_i = 10 \cdot \log_{10}(e_i(0) \cdot e_i(1))$$

La medida de la relación señal/ruido (SNR) aprovecha el hecho de que, para un codificador por comparación de formas de onda, general, la SNR es mucho mayor para sonidos sonoros. La estimación del parámetro de  $snr$  se debe realizar al final del bucle de subtramas del codificador, y se calcula en el módulo 504 de cálculo de la SNR usando la relación:

$$snr = \frac{E_{sw}}{E_e} \quad (9)$$

donde  $E_{sw}$  es la energía de la señal de voz ponderada  $s_w(n)$  de la trama actual, proveniente del filtro 205 de ponderación perceptual, y  $E_e$  es la energía del error entre esta señal de voz ponderada y la señal de síntesis ponderada de la trama actual, proveniente del filtro 205' de ponderación perceptual.

El medidor de estabilidad de altura tonal  $pc$  evalúa la variación del periodo de la altura tonal. Se calcula dentro del módulo 505 de clasificación de señales como respuesta a las estimaciones de la altura tonal en bucle abierto, de la manera siguiente:

$$pc = |p_1 - p_0| + |p_2 - p_1| \quad (10)$$

Los valores  $p_0$ ,  $p_1$ ,  $p_2$  se corresponden con las estimaciones de la altura tonal en bucle abierto calculadas por el módulo 206 de búsqueda de altura tonal en bucle abierto a partir de la primera mitad de la trama actual, la segunda mitad de la trama actual y el anticipo, respectivamente.

La energía de trama relativa  $E_s$  es calculada por el módulo 500 como una diferencia entre la energía de trama actual en dB y su promedio a largo plazo

$$E_s = \bar{E}_t - E_t$$

donde la energía de las tramas  $\bar{E}_t$  se obtiene como un sumatorio de las energías de las bandas críticas, promediado para el análisis espectral llevado a cabo en cada trama:

$$E_t = 10 \log_{10}(0.5E_t(0) + E_t(1))$$

$$E_r(j) = \sum_{i=0}^{19} e(i)$$

La energía promediada a largo plazo se actualiza en tramas de voz activas usando la siguiente relación:

$$E_{it} = 0.99E_{it} + 0.01E_r$$

5 El último parámetro es el parámetro de cruce por cero  $z_c$  calculado sobre una trama de la señal de voz por el módulo 508 de cálculo de cruce por cero. La trama comienza en medio de la trama actual y usa dos (2) subtramas del anticipo. En esta realización ilustrativa, el contador de cruces por cero  $z_c$  cuenta el número de veces que cambia el signo de la señal de positivo a negativo durante ese intervalo.

10 Para conseguir que la clasificación resulte más robusta, los parámetros de clasificación se consideran conjuntamente de manera que forman una función de mérito  $f_m$ . Con ese fin, los parámetros de clasificación en primer lugar se escalan entre 0 y 1, de modo que el valor de cada parámetro típico para una señal sorda se traduce en 0 y el valor de cada parámetro típico para una señal sonora se traduce en 1. Entre ellos se usa una función lineal. Considérese un parámetro  $p_x$ , su versión escalada se obtiene usando:

$$p^s = k_p \cdot p_x + c_p$$

15 y se recorta entre 0 y 1. Los coeficientes de la función  $k_p$  y  $c_p$  se han hallado experimentalmente para cada uno de los parámetros, de manera que la distorsión de la señal debida a las técnicas de ocultación y recuperación usadas en presencia de FERs es mínima. Los valores usados en esta implementación ilustrativa se resumen en la Tabla 2:

Tabla 2. Parámetros de Clasificación de la señal y los coeficientes de sus funciones de escalado respectivas

Parámetro	Significado	$k_p$	$c_p$
$\bar{r}_x$	Correlación Normalizada	2,857	-1,286
$\bar{\theta}_t$	Inclinación Espectral	0,04167	0
$snr$	Relación Señal/Ruido	0,1111	-0,3333
$pc$	Medidor de Estabilidad de la Altura	-0,07143	1,857
$E_s$	Energía de Trama Relativa	0,05	0,45
$z_c$	Contador de Cruces por Cero	-0,04	2,4

La función de mérito se ha definido como:

20 
$$f_m = \frac{1}{7}(2 \cdot \bar{r}_x^s + \bar{\theta}_t^s + snr^s + pc^s + E_s^s + zc^s)$$

donde el superíndice  $s$  indica la versión escalada de los parámetros.

A continuación, la clasificación se realiza usando la función de mérito  $f_m$  y siguiendo las reglas que se resumen en la Tabla 3:

Tabla 3. Reglas de Clasificación de la Señal en el Codificador

Clase de la Trama Previa	Regla	Clase de la Trama Actual
INICIO SONORA TRANSICIÓN SONORA	$f_m = 0,66$	SONORA
	$0,66 > f_m = 0,49$	TRANSICIÓN SONORA
	$f_m < 0,49$	SORDA
TRANSICIÓN SORDA SORDA	$f_m > 0,63$	INICIO
	$0,63 = f_m > 0,585$	TRANSICIÓN SORDA
	$f_m = 0,585$	SORDA

5 En caso de un codificador de velocidad de bits variable (VBR) controlado por la fuente, la clasificación de la señal es inherente al funcionamiento del códec. El códec funciona con varias velocidades de bits, y se usa un módulo de selección de velocidad para determinar la velocidad de bits utilizada para codificar cada trama de voz sobre la base de la naturaleza de la trama de voz (por ejemplo, las tramas sonoras, sordas, transitorias, de ruido de fondo se codifican, cada una de ellas, con un algoritmo de codificación especial). La información sobre el modo de codificación y, por lo tanto, sobre la clase de voz ya es una parte implícita del flujo continuo de bits, y no es necesario transmitirla explícitamente para el procesamiento del FER. A continuación, esta información de clase se puede  
10 usar para corregir la decisión de clasificación antes descrita.

En la aplicación de ejemplo para el códec de AMR WB, la única selección de velocidad controlada por la fuente representa la detección de actividad vocal (VAD). Esta bandera de VAD es igual a 1 para voz activa, y 0 para silencio. Este parámetro es útil para la clasificación, ya que indica directamente que no es necesaria ninguna clasificación adicional si su valor es 0 (es decir, la trama se clasifica directamente como SORDA). Este parámetro es  
15 la salida del módulo 402 de detección de actividad vocal (VAD). En la bibliografía existen diferentes algoritmos de VAD, y, a efectos de la presente invención, puede utilizarse cualquier algoritmo. Por ejemplo, puede usarse el algoritmo de VAD que forma parte de la norma G.722.2 [Recomendación de la ITU-T G.722.2 “Wideband coding of speech at around 16 kbit/s using Adaptive Multi-Rate Wideband (AMR-WB)”, Ginebra, 2002]. En este caso, el algoritmo de VAD se basa en la salida del análisis espectral del módulo 500 (sobre la base de la relación señal/ruido por banda crítica). La VAD usada con fines clasificatorios, difiere con respecto a la utilizada con fines de codificación con respecto al tiempo de mantenimiento (*hangover*). En codificadores de voz que usan una generación de ruido de confort (CNG) para segmentos sin voz activa (silencio o ruido solamente), con frecuencia se añade un tiempo de mantenimiento después de las ráfagas de voz (la CNG en la normativa de AMR-WB es un ejemplo [3GPP TS 26.192, “AMR Wideband Speech Codec: Comfort Noise Aspects”, Especificación Técnica del 3GPP]). Durante el  
20 tiempo de mantenimiento, el codificador de voz continúa usándose, y el sistema conmuta a la CNG únicamente después de que haya acabado el periodo de mantenimiento. A efectos de la clasificación para la ocultación del FER, esta elevada seguridad no es necesaria. Consecuentemente, la bandera de VAD para la clasificación será igual a 0 también durante el periodo de mantenimiento.

En esta realización ilustrativa, la clasificación se lleva a cabo en el módulo 505 sobre la base de los parámetros antes descritos; concretamente, correlaciones normalizadas (o información de sonoridad),  $r_x$ , inclinación espectral  $e_s$ ,  $snr$ , medidor de la estabilidad de la altura tonal  $pc$ , energía de tramas relativa  $E_s$ , frecuencia de cruces por cero  $zc$ , y bandera de VAD.

*Clasificación en el descodificador*

35 Si la aplicación no permite la transmisión de la información de clase (no se pueden transportar bits adicionales), la clasificación todavía puede llevarse a cabo en el descodificador. Tal como ya se ha indicado, la principal desventaja en este caso es que en general no hay ningún anticipo disponible en los descodificadores de voz. Además, normalmente existe la necesidad de mantener limitada la complejidad del descodificador.

Puede realizarse una clasificación simple estimando la sonoridad de la señal sintetizada. Si se considera el caso de un codificador de tipo CELP, puede usarse la estimación de la sonoridad  $r_v$  calculada como en la Ecuación (1). Es decir:  
40

$$r_v = (E_v - E_c) / (E_v + E_c)$$

donde  $E_v$  es la energía del vector código de altura tonal escalado  $bv_T$  y  $E_c$  es la energía del vector código innovador escalado  $gc_k$ . Teóricamente, para una señal puramente sonora  $r_v = 1$  y, para una señal puramente sorda,  $r_v = -1$ . La clasificación real se realiza promediando valores de  $r_v$  cada 4 subtramas. El factor resultante  $f_{rv}$  (promedio de valores de  $r_v$  de cada cuatro subtramas) se usa de la manera siguiente  
45



Tabla 4. Reglas de Clasificación de la Señal en el Descodificador

Clase de la Trama Previa	Regla	Clase de la Trama Actual
INICIO SONORA TRANSICIÓN SONORA	$f_{rv} > -0,1$	SONORA
	$-0,1 = f_{rv} = -0,5$	TRANSICIÓN SONORA
TRANSICIÓN SORDA SORDA	$f_{rv} < -0,5$ $f_{rv} > -0,1$	SORDA INICIO
	$-0,1 = f_{rv} = -0,5$	TRANSICIÓN SORDA
	$f_{rv} < -0,5$	SORDA

De forma similar a la clasificación en el codificador, en el decodificador pueden usarse otros parámetros para ayudar a la clasificación, en calidad de parámetros del filtro de LP o de la estabilidad de la altura tonal.

- 5 En el caso de un codificador de velocidad de bits variable controlado por la fuente, la información sobre el modo de codificación ya forma parte del flujo continuo de bits. Por tanto, si, por ejemplo, se usa un modo de codificación puramente sordo, la trama puede clasificarse automáticamente como SORDA. De manera similar, si se usa un modo de codificación puramente sonoro, la trama se clasifica como SONORA.

*Parámetros de la voz para el procesado del FER*

- 10 Existen unos pocos parámetros críticos que se deben controlar cuidadosamente para evitar artefactos molestos cuando se producen FERs. Si pueden transmitirse unos pocos bits adicionales, entonces estos parámetros se pueden estimar en el codificador, pueden cuantificarse y transmitirse. Si no, algunos de ellos se pueden estimar en el decodificador. Estos parámetros incluyen clasificación de la señal, información de energía, información de fase e información de sonoridad. Lo más importante es un control preciso de la energía de la voz. La fase y la periodicidad de la voz también se pueden controlar para mejorar adicionalmente la ocultación y la recuperación del FER.

15 La importancia del control de la energía se manifiesta principalmente cuando se recupera un funcionamiento normal después de un bloque borrado de tramas. Puesto que la mayoría de codificadores de voz hace uso de una predicción, en el decodificador no puede estimarse apropiadamente la energía correcta. En segmentos de voz sonoros, la energía incorrecta puede persistir durante varias tramas consecutivas, lo cual es muy molesto especialmente cuando esta energía incorrecta se incrementa.

20 Aunque el control de energía es lo más importante para la voz sonora debido a la predicción a largo plazo (predicción de la altura tonal), también es importante para la voz sorda. El motivo en este caso es la predicción del cuantificador de ganancia de innovación usado frecuentemente en codificadores de tipo CELP. La energía errónea durante segmentos sordos puede provocar una fluctuación molesta de alta frecuencia.

- 25 El control de fase se puede realizar de varias maneras, principalmente en función del ancho de banda disponible. En nuestra implementación, se logra un control de fase sencillo durante inicios sonoros perdidos buscando la información aproximada sobre la posición del pulso glotal.

30 Por tanto, aparte de la información de clasificación de la señal descrita en la sección previa, la información más importante a enviar es la información sobre la energía de la señal y la posición del primer pulso glotal en una trama (información de fase). Si hay disponible un ancho de banda suficiente, también puede enviarse una información de sonoridad.

*Información de energía*

35 La información de energía se puede estimar y se puede enviar o bien en el dominio residual de LP o bien en el dominio de la señal de voz. El envío de la información en el dominio residual presenta la desventaja de no tener en cuenta la influencia del filtro de síntesis de LP. Esto puede resultar particularmente delicado en el caso de una recuperación sonora después de varias tramas sonoras perdidas (cuando el FER se produce durante un segmento de voz sonoro). Cuando llega un FER después de una trama sonora, la excitación de la última trama buena se usa típicamente durante la ocultación con alguna estrategia de atenuación. Cuando un filtro de síntesis de LP nuevo llega con la primera trama buena después del borrado, puede producirse una disparidad entre la energía de excitación y la ganancia del filtro de síntesis de LP. El filtro de síntesis nuevo puede producir una señal de síntesis con una energía notablemente diferente con respecto a la energía de la última trama borrada, sintetizada, y, también, con respecto a la energía original de la señal. Por este motivo, la energía se calcula y se cuantifica en el dominio de la señal.

40 La energía  $E_q$  se calcula y cuantifica en el módulo 506 de estimación y cuantificación de energía. Se ha observado

que 6 bits son suficientes para transmitir la energía. No obstante, el número de bits se puede reducir sin ningún efecto significativo, si no hay disponibles bits suficientes. En esta realización preferida, se utiliza un cuantificador uniforme de 6 bits en el intervalo de -15 dB a 83 dB con un paso de 1,58 dB. El índice de cuantificación viene dado por la parte entera de:

$$i = \frac{10 \log_{10}(E + 0.001) + 15}{1.58} \quad (15)$$

donde  $E$  es el máximo de la energía de la señal para tramas clasificadas como SONORAS o de INICIO, o la energía promedio por muestra para otras tramas. Para tramas SORDAS o de INICIO, el máximo de la energía de la señal se calcula de manera sincronizada con respecto a la altura tonal, en la terminación de la trama, de la manera siguiente:

$$E = \max_{i=L-t_E}^{L-1} (s^2(i)) \quad (16)$$

donde  $L$  es la longitud de la trama, y la señal  $s(i)$  significa señal de voz (o la señal de voz sin ruido en caso de que se use una supresión de ruido). En esta realización ilustrativa,  $s(i)$  significa la señal de entrada después de un diezmado a 12,8 kHz y de un pre-procesado. Si el retardo de altura tonal es mayor de 63 muestras,  $t_E$  es igual al retardo de altura tonal en bucle cerrado, redondeado, de la última subtrama. Si el retardo de la altura tonal es menor de 64 muestras, entonces  $t_E$  se fija a dos veces el retardo de altura tonal en bucle cerrado, redondeado, de la última subtrama.

Para otras clases,  $E$  es la energía promedio por muestra de la segunda mitad de la trama actual, es decir,  $t_E$  se fija a  $L/2$  y la  $E$  se calcula como:

$$E = \frac{1}{t_E} \sum_{j=L-t_E}^{L-1} s^2(j) \quad (17)$$

#### Información de control de fase

El control de fase es particularmente importante mientras se lleva a cabo la recuperación después de un segmento perdido de voz sonora, por motivos similares a los descritos en la sección previa. Después de un bloque de tramas borradas, las memorias del decodificador se desincronizan con las memorias del codificador. Para volver a sincronizar el decodificador, puede enviarse cierta información de fase en función del ancho de banda disponible. En la implementación ilustrativa descrita, se envía una posición aproximada del primer pulso glotal en la trama. A continuación, esta información se usa para la recuperación después de inicios sonoros perdidos tal como se describirá posteriormente.

Sea  $T_0$  el retardo de altura tonal en bucle cerrado, redondeado, para la primera subtrama. En primer lugar, el módulo 507 de cuantificación y búsqueda del pulso glotal busca la posición del primer pulso glotal  $\tau$  entre las  $T_0$  primeras muestras de la trama mediante la búsqueda de la muestra con la máxima amplitud. Se obtienen resultados óptimos cuando la posición del primer pulso glotal se mide sobre la señal residual filtrada por paso-bajo.

La posición del primer pulso glotal se codifica usando 6 bits de la siguiente manera. La precisión usada para codificar la posición del primer pulso glotal depende del valor de altura tonal en bucle cerrado para la primera subtrama  $T_0$ . Esto es posible debido a que este valor es conocido tanto por el codificador como por el decodificador, y no está sujeto a propagación de errores después de una o varias pérdidas de tramas. Cuando  $T_0$  es menor de 64, la posición del primer pulso glotal con respecto al comienzo de la trama se codifica directamente con una precisión de una muestra. Cuando  $64 = T_0 < 128$ , la posición del primer pulso glotal con respecto al comienzo de la trama se codifica con una precisión de dos muestras usando una simple división entera, es decir,  $\tau/2$ . Cuando  $T_0 = 128$ , la posición del primer pulso glotal con respecto al comienzo de la trama se codifica con una precisión de cuatro muestras dividiendo adicionalmente  $\tau$  por 2. El procedimiento inverso se realiza en el decodificador. Si  $T_0 < 64$ , la posición cuantificada recibida se usa tal como está. Si  $64 = T_0 < 128$ , la posición cuantificada recibida se multiplica por 2 y se incrementa en 1. Si  $T_0 = 128$ , la posición cuantificada recibida se multiplica por 4 y se incrementa en 2 (el incremento en 2 da como resultado un error de cuantificación uniformemente distribuido).

Según otra realización de la invención en la que se codifica la forma del primer pulso glotal, la posición del primer pulso glotal se determina mediante un análisis de correlación entre la señal residual y las posibles formas, signos (positivo o negativo) y posiciones del pulso. La forma del pulso se puede tomar de un libro de códigos de formas de pulso conocido tanto en el codificador como en el decodificador, conociéndose este método como cuantificación vectorial por parte de aquellos con conocimientos habituales en la materia. A continuación, la forma, el signo y la amplitud del primer pulso glotal se codifican y se transmiten al decodificador.

*Información de periodicidad*

5 En caso de que haya suficiente ancho de banda, puede calcularse y transmitirse una información de periodicidad, o información de sonoridad, y la misma se puede usar en el decodificador para mejorar la ocultación del borrado de tramas. La información de sonoridad se estima basándose en la correlación normalizada. Se puede codificar de manera bastante precisa con 4 bits, aunque, si fuera necesario, bastaría con 3 ó incluso 2 bits. La información de sonoridad es necesaria, en general, únicamente para tramas con ciertos componentes periódicos, y, para tramas de alta sonoridad, es necesaria una mejor resolución de sonoridad. La correlación normalizada se proporciona en la Ecuación (2), y se usa como indicador para la información de sonoridad. Se cuantifica en el módulo 507 de cuantificación y búsqueda del primer pulso glotal. En esta realización ilustrativa, se ha usado un cuantificador lineal por tramos, para codificar la información de sonoridad de la manera siguiente:

$$i = \frac{r_x(2) - 0.65}{0.03} + 0.5, \text{ para } r_x(2) < 0.92 \quad (18)$$

$$i = 9 + \frac{r_x(2) - 0.92}{0.01} + 0.5, \text{ para } r_x(2) \geq 0.92 \quad (19)$$

15 Nuevamente, se codifica y se transmite la parte entera de  $i$ . La correlación  $r_x(2)$  tiene el mismo significado que en la Ecuación (1). En la Ecuación (18), la sonoridad se cuantifica linealmente entre 0,65 y 0,89 con el paso de 0,03. En la Ecuación (19), la sonoridad se cuantifica linealmente entre 0,92 y 0,98 con el paso de 0,01.

Si es necesario un intervalo de cuantificación mayor, puede usarse la siguiente cuantificación lineal:

$$i = \frac{\bar{r}_x - 0.4}{0.04} + 0.5 \quad (20)$$

Esta ecuación cuantifica la sonoridad en el intervalo de 0,4 a 1, con el paso de 0,04. La correlación  $\bar{r}_x$  se define en la Ecuación (2a).

20 A continuación, las ecuaciones (18) y (19) o la ecuación (20) se usan en el decodificador para calcular  $r_x(2)$  o  $\bar{r}_x$ . Denomínese  $r_q$  a esta correlación normalizada cuantificada. Si la sonoridad no puede transmitirse, la misma puede estimarse usando el factor de sonoridad de la Ecuación (2a) mapeándolo en el intervalo de 0 a 1.

$$r_q = 0.5 \cdot (f + 1) \quad (21)$$

*Procesado de tramas borradas*

25 Las técnicas de ocultación del FER en esta realización ilustrativa se muestran claramente sobre codificadores de tipo ACELP. No obstante, las mismas se pueden aplicar fácilmente a cualquier códec de voz en el que la señal de síntesis se genere filtrando una señal de excitación a través de un filtro de síntesis de LP. La estrategia de ocultación se puede resumir como una convergencia de la energía de la señal y la envolvente espectral hacia 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 clase de la última trama recibida buena y del número de tramas borradas consecutivas, y se controla por medio de un factor de atenuación  $\alpha$ . El factor  $\alpha$  depende además de la estabilidad del filtro de LP para tramas SORDAS. En general, la convergencia es lenta si la última trama recibida buena está en un segmento estable, y es rápida si la trama se encuentra en un segmento de transición. Los valores de  $\alpha$  se resumen en la Tabla 5.

35 Tabla 5. Valores del factor de atenuación  $\alpha$  de la ocultación del FER

Última Trama Recibida Buena	Número de tramas borradas sucesivas	$\alpha$
INICIO ARTIFICIAL		0,6
INICIO, SONORA	= 3	1,0
	> 3	0,4
TRANSICIÓN SONORA		0,4
TRANSICIÓN SORDA		0,8
SORDA	= 1	0,6 $\theta$ + 0,4
	> 1	0,4

5

Un factor de estabilidad  $\theta$  se calcula basándose en una medición de la distancia entre los filtros de LP adyacentes. En este caso, el factor  $\theta$  está relacionado con la medición de la distancia de ISF (Frecuencias Espectrales de Inmitancia) y está acotado por  $0 \leq \theta \leq 1$ , de manera que valores mayores de  $\theta$  se corresponden con señales más estables. Esto da como resultado fluctuaciones menores de energía y de la envolvente espectral cuando se produce un borrado de trama aislado dentro de un segmento sordo estable.

10

La clase de la señal permanece invariable durante el procesado de tramas borradas, es decir, la clase sigue siendo la misma que en la última trama recibida buena.

15

*Construcción de la parte periódica de la excitación*

Para una ocultación de tramas borradas tras una trama SORDA recibida correctamente, no se genera parte periódica de la señal de excitación. Para una ocultación de tramas borradas tras una trama recibida correctamente que no sea SORDA, la parte periódica de la señal de excitación se construye repitiendo el último periodo de altura tonal de la trama previa. Si se trata de la 1ª trama borrada después de una trama buena, este impulso de altura tonal en primer lugar se filtra por paso-bajo. El filtro usado es un filtro FIR de fase lineal y de 3 coeficientes, simple, con los coeficientes del filtro iguales a 0,18; 0,64 y 0,18. Si hay disponible una información de sonoridad, el filtro también se puede seleccionar dinámicamente con una frecuencia de corte dependiente de la sonoridad.

20

El periodo de la altura tonal  $T_c$  usado para seleccionar el último impulso de altura tonal y, por lo tanto, usado durante la ocultación, se define de manera que puedan evitarse, o reducirse, múltiplos o sub-múltiplos de la altura tonal. Se usa la siguiente lógica en la determinación del periodo de la altura tonal  $T_c$ .

25

si  $((T_3 < 1,8 T_s) \text{ AND } (T_3 > 0,6 T_s)) \text{ OR } (T_{cnt} = 30)$ , entonces  $T_c = T_3$ , si no  $T_c = T_s$ .

En este caso,  $T_3$  es el periodo de altura tonal redondeado de la 4ª subtrama de la última trama buena recibida, y  $T_s$  es el periodo de altura tonal redondeado de la 4ª subtrama de la última trama sonora, estable, buena, con estimaciones coherentes de la altura tonal. Una trama sonora estable se define en este caso como una trama SONORA precedida por una trama de tipo sonoro (TRANSICIÓN SONORA, SONORA, INICIO), la coherencia de la altura tonal se verifica en esta implementación examinando si las estimaciones de la altura tonal en bucle cerrado son razonablemente parecidas, es decir, si las relaciones entre la altura tonal de la última subtrama, la altura tonal de la 2ª subtrama y la altura tonal de la última subtrama de la trama previa se sitúan dentro del intervalo (0,7; 1,4).

30

Esta determinación del periodo de la altura tonal  $T_c$  significa que, si la altura tonal en la terminación de la última trama buena y la altura tonal de la última trama estable están próximas entre sí, se usa la altura tonal de la última trama buena. Si no, esta altura tonal se considera como no fiable, y se utiliza, en cambio, la altura tonal de la última trama estable, para evitar el impacto de estimaciones erróneas de la altura tonal en inicios sonoros. No obstante, esta lógica tiene sentido únicamente si el último segmento estable no está demasiado alejado en el pasado. Por tanto, se define un contador  $T_{cnt}$  que limita el alcance de la influencia del último segmento estable. Si  $T_{cnt}$  es superior o igual a 30, es decir, si se han producido por lo menos 30 tramas desde la última actualización de  $T_s$ , se usa sistemáticamente la altura tonal de la última trama buena.  $T_{cnt}$  se reinicializa a 0 cada vez que se detecta un segmento estable y se actualiza  $T_s$ . A continuación, el periodo  $T_c$  se mantiene constante durante la ocultación para el bloque borrado completo.

35

40

Puesto que el último impulso de la excitación de la trama previa se usa para la construcción de la parte periódica, su ganancia es aproximadamente correcta en el comienzo de la trama ocultada y se puede fijar a 1. A continuación, la ganancia se atenúa linealmente durante toda la trama, muestra a muestra, para lograr el valor de  $\alpha$  en la terminación de la trama.

45

Los valores de  $\alpha$  se corresponden con la Tabla 5, con la excepción de que se modifican para borrados que suceden a tramas SONORAS y de INICIO, con el fin de tener en cuenta la evolución de la energía de segmentos sonoros. Esta evolución se puede extrapolar en cierta medida usando los valores de ganancia de excitación de la altura tonal de cada subtrama de la última trama buena. En general, si estas ganancias son superiores a 1, la energía de la señal es creciente, si son inferiores a 1, la energía es decreciente. Así,  $\alpha$  se multiplica por un factor de corrección  $f_b$  calculado de la manera siguiente:

50

$$f_b = \sqrt{0.1b(0) + 0.2b(1) + 0.3b(2) + 0.4b(3)} \quad (23)$$

donde  $b(0)$ ,  $b(1)$ ,  $b(2)$  y  $b(3)$  son las ganancias de altura tonal de las cuatro subtramas de la última trama recibida correctamente. El valor de  $f_b$  se recorta entre 0,98 y 0,85 antes de usarse para escalar la parte periódica de la excitación. De esta manera, se evitan aumentos y reducciones fuertes de la energía.

- 5 Para tramas borradas que suceden a una trama recibida correctamente y que no sea SORDA, la memoria intermedia de excitación se actualiza solamente con esta parte periódica de la excitación. Esta actualización se usará para construir la excitación de libro de códigos de altura tonal en la siguiente trama.

*Construcción de la parte aleatoria de la excitación*

- 10 La parte de innovación (no periódica) de la señal de excitación se genera aleatoriamente. Se puede generar en forma de un ruido aleatorio o utilizando el libro de códigos de innovación de CELP con índices vectoriales generados de forma aleatoria. En la presente realización ilustrativa, se ha usado un simple generador aleatorio con una distribución aproximadamente uniforme. Antes de ajustar la ganancia de innovación, la innovación generada aleatoriamente se escala a algún valor de referencia, fijado en este caso a la energía unitaria por muestra.

- 15 En el 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:

$$g_s = 0.1g(0) + 0.2g(1) + 0.3g(2) + 0.4g(3) \quad (23a)$$

- 20 donde  $g(0)$ ,  $g(1)$ ,  $g(2)$  y  $g(3)$  son las ganancias de libro de códigos fijo, o de innovación, de las cuatro (4) subtramas de la última trama recibida correctamente. La estrategia de atenuación de la parte aleatoria de la excitación es algo diferente con respecto a la atenuación de la excitación de la altura tonal. El motivo es que la excitación de la altura tonal (y, por lo tanto, la periodicidad de la excitación) converge a 0, mientras que la excitación aleatoria converge a la energía de excitación de la generación de ruido de confort (CNG). La atenuación de la ganancia de innovación se lleva a cabo en forma de:

$$g_s^1 = \alpha \cdot g_s^0 + (1 - \alpha) \cdot g_n \quad (24)$$

- 25 donde  $g_s^1$  es la ganancia de innovación en el comienzo de la siguiente trama,  $g_s^0$  es la ganancia innovadora en el comienzo de la trama actual,  $g_n$  es la ganancia de la excitación usada durante la generación del ruido de confort y  $\alpha$  es tal como se define en la Tabla 5. Por lo tanto, de manera similar a la atenuación de la excitación periódica, la ganancia se atenúa linealmente durante toda la trama, muestra a muestra, comenzando con  $g_s^0$  y yendo hacia el valor de  $g_s^1$  que se alcanzaría en el comienzo de la siguiente trama.

- 30 Finalmente, si la última trama recibida buena (recibida correctamente o no borrada) es diferente de SORDA, la excitación de innovación se filtra a través de un filtro paso-alto FIR de fase lineal con coeficientes -0,0125; -0,109; 0,7813; -0,109; -0,0125. Para reducir la cantidad de componentes con ruido durante segmentos sonoros, estos coeficientes del filtro se multiplican por un factor adaptativo igual a  $(0,75 - 0,25 r_i)$ , siendo  $r_i$  el factor de sonoridad que se definió en la Ecuación (1). A continuación, la parte aleatoria de la excitación se añade a la excitación adaptativa para formar la señal de excitación total.

- 35 Si la última trama buena es SORDA, se usa únicamente la excitación de innovación, y la misma se atenúa adicionalmente con un factor de 0,8. En este caso, la memoria intermedia de excitación pasada se actualiza con la excitación de innovación, ya que no hay disponible ninguna parte periódica de la excitación.

*Ocultación, síntesis y actualizaciones de la envolvente espectral*

- 40 Para sintetizar la voz descodificada, deben obtenerse los parámetros del filtro de LP. La envolvente espectral se mueve gradualmente a la envolvente estimada del ruido ambiente. En este caso, se usa la representación en ISF de los parámetros de LP:

$$l^1(j) = \alpha l^0(j) + (1 - \alpha) l_n(j), \quad j=0, \dots, p-1 \quad (25)$$

En la ecuación (25),  $l^1(j)$  es el valor de la ISF  $f^{\text{ésima}}$  de la trama actual,  $l^0(j)$  es el valor de la ISF  $f^{\text{ésima}}$  de la trama previa,  $l_n(j)$  es el valor de la ISF  $f^{\text{ésima}}$  de la envolvente estimada del ruido de confort y  $p$  es el orden del filtro de LP.

- 45 La voz sintetizada se obtiene filtrando la señal de excitación a través del filtro de síntesis de LP. Los coeficientes del filtro se calculan a partir de la representación en ISF y se interpolan para cada subtrama (cuatro (4) veces por trama) como durante el funcionamiento normal del codificador.

En la medida en la que tanto el cuantificador de la ganancia de innovación como el cuantificador de ISF usan una predicción, su memoria no estará al día después de que se restablezca el funcionamiento normal. Para reducir este efecto, las memorias de los cuantificadores se estiman y se actualizan en la terminación de cada trama borrada.

*Recuperación del funcionamiento normal después del borrado*

5 El problema de la recuperación después de un bloque borrado de tramas es básicamente debido a la fuerte predicción usada prácticamente en todos los codificadores de voz actuales. En particular, los codificadores de voz de tipo CELP logran su elevada relación señal/ruido para voz sonora, debido al hecho de que usan la señal de excitación pasada para codificar la excitación de la trama actual (predicción a largo plazo o de altura tonal). Además, la mayor parte de los cuantificadores (cuantificadores de LP, cuantificadores de ganancia) hacen uso de una predicción.

*Construcción del inicio artificial*

15 La situación más complicada en relación con el uso de la predicción a largo plazo en codificadores de CELP se produce cuando se pierde un inicio sonoro. Inicio perdido significa que el inicio de la voz sonora se produjo en algún lugar durante el bloque borrado. En este caso, la última trama recibida buena era sorda y, por lo tanto, en la memoria intermedia de excitación no se encuentra ninguna excitación periódica. No obstante, la primera trama buena después del bloque borrado es sonora, la memoria intermedia de excitación en el codificador es altamente periódica y la excitación adaptativa se ha codificado usando esta excitación pasada periódica. Puesto que esta parte periódica de la excitación falta por completo en el descodificador, puede tardarse varias tramas en la recuperación de esta pérdida.

20 Si se pierde una trama de INICIO (es decir, una trama buena SONORA llega después de un borrado, pero la última trama buena antes del borrado era SORDA, tal como se muestra en la Figura 6), se usa la técnica especial para reconstruir artificialmente el inicio perdido y activar la síntesis sonora. En el comienzo de la 1ª trama buena después de un inicio perdido, la parte periódica de la excitación se construye artificialmente en forma de un tren periódico, filtrado por paso-bajo, de impulsos separados por un periodo de altura tonal. En la presente realización ilustrativa, el filtro paso-bajo es un simple filtro FIR de fase lineal con la respuesta a impulsos  $h_{low} = \{-0,0125; 0,109; 0,7813; 0,109; -0,0125\}$ . No obstante, el filtro también se podría seleccionar dinámicamente con una frecuencia de corte correspondiente a la información de sonoridad en caso de que esta información esté disponible. La parte innovadora de la excitación se construye usando una descodificación de CELP normal. Las entradas del libro de códigos de innovación también se podrían seleccionar aleatoriamente (o la propia innovación se podría generar aleatoriamente), ya que, de todos modos, se ha perdido la sincronización con la señal original.

35 En la práctica, la longitud del inicio artificial está limitada de manera que por lo menos un periodo de altura tonal completo se construye con este método, y se continúa con el método hasta la terminación de la subtrama actual. después de esto, se reanuda un procesado de ACELP regular. El periodo de altura tonal considerado es el promedio redondeado de los periodos de altura tonal descodificados de todas las subtramas en las que se usa la reconstrucción artificial del inicio. El tren de impulsos filtrado por paso-bajo se obtiene colocando las respuestas a impulsos del filtro paso-bajo en la memoria intermedia de excitación adaptativa (inicializada previamente a cero). La primera respuesta a impulsos estará centrada en la posición cuantificada  $\tau_q$  (transmitida dentro del flujo continuo de bits) con respecto al comienzo de la trama, y los impulsos restantes se colocarán con la distancia de la altura tonal promediada hasta la terminación de la última subtrama afectada por la construcción artificial del inicio. Si el ancho de banda disponible no es suficiente para transmitir la posición del primer pulso glotal, la primera respuesta a impulsos se puede colocar arbitrariamente en torno a la mitad del periodo de altura tonal después del comienzo de la trama actual.

45 Como ejemplo, para la longitud de subtrama de 64 muestras, considérese que los periodos de la altura tonal en la primera y la segunda subtrama son  $p(0)=70,75$  y  $p(1)=71$ . Puesto que este valor es mayor que el tamaño de subtrama de 64, entonces el inicio artificial se construirá durante las dos primeras subtramas, y el periodo de la altura tonal será igual al promedio de la altura tonal de las dos subtramas redondeado al entero más próximo, es decir, 71. Las dos últimas subtramas serán procesadas por el descodificador de CELP normal.

50 A continuación, la energía de la parte periódica de la excitación del inicio artificial se escala según la ganancia correspondiente a la energía cuantificada y transmitida para la ocultación del FER (según se define en las Ecuaciones 16 y 17) y se divide por la ganancia del filtro de síntesis de LP. La ganancia del filtro de síntesis de LP se calcula como:

$$g_{LP} = \sqrt{\sum_{l=0}^{63} h^2(l)} \quad (31)$$

55 donde  $h(l)$  es la respuesta a impulsos del filtro de síntesis de LP. Finalmente, la ganancia del inicio artificial se reduce multiplicando la parte periódica por 0,96. Alternativamente, este valor podría corresponderse con la sonoridad, si hubiera un ancho de banda disponible para transmitir también la información de sonoridad.

Alternativamente, sin desviarse con respecto a la esencia de esta invención, el inicio artificial también se puede construir en la memoria intermedia de excitaciones pasadas antes de entrar en el bucle de subtramas del descodificador. Esto presentaría la ventaja de evitar que el procesado especial construya la parte periódica del inicio artificial, y podría usarse, en su lugar, la descodificación de CELP regular.

- 5 El filtro de LP para la síntesis de voz de salida no se interpola en el caso de una construcción de inicio artificial. En cambio, los parámetros de LP recibidos se usan para la síntesis de la trama completa.

*Control de energía*

La tarea más importante en la recuperación después de un bloque borrado de tramas es controlar apropiadamente la energía de la señal de voz sintetizada. El control de la energía de síntesis es necesario debido a la fuerte predicción utilizada habitualmente en los codificadores de voz actuales. El control de energía adquiere la mayor importancia cuando se produce un bloque de tramas borradas durante un segmento sonoro. Cuando un borrado de trama llega después de una trama sonora, la excitación de la última trama buena se usa típicamente durante la ocultación con cierta estrategia de atenuación. Cuando un filtro de LP nuevo llega con la primera trama buena después del borrado, puede producirse una disparidad entre la energía de excitación y la ganancia del nuevo filtro de síntesis de LP. El nuevo filtro de síntesis puede producir una señal de síntesis con una energía considerablemente diferente con respecto a la energía de la última trama borrada sintetizada, y también con respecto a la energía de la señal original.

El control de energía durante la primera trama buena después de una trama borrada se puede resumir de la manera siguiente. La señal sintetizada se escala de manera que su energía sea similar a la energía de la señal de voz sintetizada en la terminación de la última trama borrada en el comienzo de la primera trama buena, y de manera que converja a la energía transmitida, hacia la terminación de la trama con la evitación de un incremento de energía demasiado importante.

El control de energía se realiza en el dominio de la señal de voz sintetizada. Incluso si la energía se controla en el dominio de la voz, la señal de excitación se debe escalar ya que sirve como memoria de predicción a largo plazo para las siguientes tramas. A continuación, la síntesis se vuelve a realizar para suavizar las transiciones. Considérese que  $g_0$  indica la ganancia usada para escalar la primera muestra en la trama actual y  $g_1$  la ganancia usada en la terminación de la trama. La señal de excitación se escala entonces de la manera siguiente:

$$u_s(i) = g_{AGC}(i) \cdot u(i), \quad i=0, \dots, L-1 \quad (32)$$

donde  $u_s(i)$  es la excitación escalada,  $u(i)$  es la excitación antes del escalado,  $L$  es la longitud de la trama, y  $g_{AGC}(i)$  es la ganancia comenzando desde  $g_0$  y convergiendo exponencialmente a  $g_1$ :

$$g_{AGC}(i) = f_{AGC} g_{AGC}(i-1) + (1-f_{AGC}) g_1 \quad i=0, \dots, L-1$$

con la inicialización de  $g_{AGC}(-1) = g_0$ , donde  $f_{AGC}$  es el factor de atenuación fijado en esta implementación al valor de 0,98. Este valor se ha hallado experimentalmente como un compromiso entre, por un lado, tener una transición suave desde la trama previa (borrada), y, por otro lado, escalar el último periodo de altura tonal de la trama actual lo máximo posible al valor correcto (transmitido). Esto es importante debido a que el valor de la energía transmitida se estima de manera sincronizada con respecto a la altura tonal, en la terminación de la trama. Las ganancias  $g_0$  y  $g_1$  se definen como:

$$g_0 = \sqrt{E_{-1} / E_0} \quad (33a)$$

$$g_1 = \sqrt{E_q / E_1} \quad (33b)$$

donde  $E_{-1}$  es la energía calculada en la terminación de la trama previa (borrada),  $E_0$  es la energía en el comienzo de la trama actual (recuperada),  $E_1$  es la energía en la terminación de la trama actual, y  $E_q$  es la información de energía transmitida cuantificada en la terminación de la trama actual, según el cálculo realizado en el codificador a partir de las Ecuaciones (16, 17). El cálculo de  $E_{-1}$  y  $E_1$  es similar con la excepción de que el mismo se realiza sobre la señal de voz sintetizada  $s'$ .  $E_{-1}$  se calcula de manera sincronizada con respecto a la altura tonal usando el periodo de la altura tonal de ocultación  $T_c$ , y  $E_1$  usa la altura tonal redondeada de la última subtrama  $T_3$ .  $E_0$  se calcula de manera similar utilizando el valor redondeado de la altura tonal  $T_0$  de la primera subtrama, modificándose las ecuaciones (16, 17) para obtener:

$$E = \max_{i=0}^{t_E} (s'^2(i))$$

para tramas SONORAS y de INICIO.  $t_E$  es igual al retardo redondeado de la altura tonal o dos veces esa longitud si

la altura tonal es menor de 64 muestras. Para otras tramas,

$$E = \frac{1}{t_0} \sum_{i=0}^{t_E} s^{i^2} (i)$$

con  $t_E$  igual a la mitad de la longitud de la trama. Las ganancias  $g_0$  y  $g_1$  se limitan adicionalmente a un valor permitido máximo, para evitar una energía de alto valor. En la presente implementación ilustrativa, este valor se ha fijado a 1,2.

La ejecución de la ocultación del borrado de tramas y la recuperación del descodificador comprende, cuando una ganancia de un filtro de LP de una primera trama no borrada recibida tras el borrado de la trama es mayor que una ganancia de un filtro de LP de una última trama borrada durante dicho borrado de trama, ajustar la energía de una señal de excitación del filtro de LP producida en el descodificador durante la primera trama no borrada recibida, a una ganancia del filtro de LP de dicha primera trama no borrada recibida, usando la siguiente relación:

Si  $E_q$  no se puede transmitir,  $E_q$  se fija a  $E_1$ . No obstante, si el borrado se produce durante un segmento de voz sonora (es decir, la última trama buena antes del borrado y la primera trama buena después del borrado se clasifican como TRANSICIÓN SONORA, SONORA o INICIO), deben tomarse precauciones adicionales debido a la posible disparidad entre la energía de la señal de excitación y la ganancia del filtro de LP, mencionada previamente. Surge una situación particularmente peligrosa cuando la ganancia del filtro de LP de una primera trama no borrada, recibida tras el borrado de la trama, es mayor que la ganancia del filtro de LP de una última trama borrada durante ese borrado de trama. En ese caso particular, la energía de la señal de excitación del filtro de LP producida en el descodificador durante la primera trama no borrada, recibida, se ajusta a una ganancia del filtro de LP de la primera trama no borrada, recibida, usando la siguiente relación:

$$E_q = E_1 \frac{E_{LP0}}{E_{LP1}}$$

donde  $E_{LP0}$  es la energía de la respuesta a impulsos del filtro de LP de la última trama buena antes del borrado, y  $E_{LP1}$  es la energía del filtro de LP de la primera trama buena antes del borrado. En esta implementación, se usan los filtros de LP de las últimas subtramas en una trama. Finalmente, el valor de  $E_q$  se limita al valor de  $E_1$  en este caso (transmitiéndose información del borrado de segmento sonoro sin  $E_q$ ).

Las siguientes excepciones, relacionadas todas ellas con transiciones en la señal de voz, corrigen adicionalmente el cálculo de  $g_0$ . Si se usa un inicio artificial en la trama actual,  $g_0$  se fija a 0,5  $g_1$ , para hacer que la energía del inicio aumente gradualmente.

En el caso de una primera trama buena después de un borrado clasificada como INICIO, se evita que la ganancia  $g_0$  sea mayor que  $g_1$ . Esta precaución se toma para evitar que un ajuste de ganancia positivo en el comienzo de la trama (el cual, probablemente, todavía es sordo, al menos de forma parcial) amplifique el inicio sonoro (en la terminación de la trama).

Finalmente, durante una transición de sonoridad a sordez (es decir, que la última trama buena esté clasificada como TRANSICIÓN SONORA, SONORA o INICIO, y la trama actual esté clasificada como SORDA), o durante una transición de un periodo de voz no activa a un periodo de voz activa (la última trama recibida buena codificada como ruido de confort, y la trama actual codificada como voz activa), la  $g_0$  se fija a  $g_1$ .

En el caso de un borrado de segmento sonoro, el problema de la energía errónea también puede manifestarse en tramas que sucedan a la primera trama buena después del borrado. Esto puede ocurrir incluso si la energía de la primera trama buena se ha ajustado tal como se ha descrito anteriormente. Para atenuar este problema, puede continuarse con el control de energía hasta la terminación del segmento sonoro.

Aunque, en la descripción anterior, la presente invención se ha expuesto en relación con una realización ilustrativa de la misma, esta realización ilustrativa se puede modificar a voluntad. El alcance de protección queda definido en las reivindicaciones adjuntas.



**REIVINDICACIONES**

1. Método de ocultación del borrado de tramas provocado por tramas de una señal de sonido codificada, borradas durante su transmisión desde un codificador a un descodificador, y para acelerar la recuperación del descodificador después de que se hayan recibido tramas no borradas de la señal de sonido codificada, que comprende:
  - 5 determinar, en el codificador, parámetros de ocultación/recuperación que comprenden por lo menos dos parámetros seleccionados del grupo que consiste en un parámetro de clasificación de la señal, un parámetro de información de energía, un parámetro de información de sonoridad y un parámetro de información de fase;
 

cuantificar los parámetros de ocultación/recuperación; y

10 transmitir al descodificador los parámetros de ocultación/recuperación cuantificados, determinados en el codificador;

en donde:

los parámetros de ocultación/recuperación son utilizables para mejorar la ocultación del borrado de tramas y la recuperación del descodificador después de un borrado de tramas;

la señal de sonido es una señal de voz;

15 caracterizado por que:

la determinación, en el codificador, de los parámetros de ocultación/recuperación comprende clasificar tramas sucesivas de la señal de sonido codificada, como sorda, transición sorda, transición sonora, sonora, o inicio; y

20 la determinación de los parámetros de ocultación/recuperación comprende calcular el parámetro de información de energía en relación con un máximo de una energía de señal para tramas clasificadas como sonoras o inicio, y calcular el parámetro de información de energía en relación con una energía promedio por muestra, para otras tramas.
  2. Método según la reivindicación 1, en el que la determinación del parámetro de información de fase comprende determinar una posición de un primer pulso glotal en una trama de la señal de sonido codificada.
  3. Método según la reivindicación 2, en el que la determinación del parámetro de información de fase comprende
 

25 codificar, en el codificador, una forma, un signo y una amplitud del primer pulso glotal, y transmitir la forma, el signo y la amplitud codificados, desde el codificador al descodificador.
  4. Método según la reivindicación 2, en el que la determinación de la posición del primer pulso glotal comprende:
 

medir una muestra de amplitud máxima dentro de un periodo de altura tonal como primer pulso glotal; y

cuantificar una posición de la muestra de amplitud máxima dentro del periodo de la altura tonal.
  - 30 5. Método según la reivindicación 1, en el que la clasificación de las tramas sucesivas comprende clasificar como sorda, toda trama que sea una trama sorda, toda trama sin voz activa, y toda trama de final sonora que tenga una terminación que tienda a ser sorda.
  6. Método según la reivindicación 1, en el que la clasificación de las tramas sucesivas comprende clasificar como
 

35 transición sorda, toda trama sorda que tenga una terminación con un posible inicio sonoro que sea demasiado corto o que no esté construido suficientemente bien para ser procesado como trama sonora.
  7. Método según la reivindicación 1, en el que la clasificación de las tramas sucesivas comprende clasificar como transición sonora, toda trama sonora con características sonoras relativamente débiles, incluyendo tramas sonoras con características que varían rápidamente y finales sonoros que duran la trama completa, en donde una trama clasificada como transición sonora sucede únicamente a tramas clasificadas como transición sonora, sonora o inicio.
  - 40 8. Método según la reivindicación 1, en el que la clasificación de las tramas sucesivas comprende clasificar como sonora, toda trama sonora con características estables, en donde una trama clasificada como sonora sucede únicamente a tramas clasificadas como transición sonora, sonora o inicio.
  9. Método según la reivindicación 1, en el que la clasificación de las tramas sucesivas comprende clasificar como
 

45 inicio toda trama sonora con características estables que suceda a una trama clasificada como sorda o transición sorda.
  10. Método según la reivindicación 1, que comprende determinar la clasificación de las tramas sucesivas de la señal de sonido codificada, sobre la base de al menos una parte de los siguientes parámetros: un parámetro de correlación normalizada, un parámetro de inclinación espectral, un parámetro de relación señal/ruido, un parámetro de estabilidad de altura tonal, un parámetro de energía de trama relativa, y un parámetro de cruce por cero.

11. Método según la reivindicación 10, en el que la determinación de la clasificación de las tramas sucesivas comprende:
- 5            calcular una figura de mérito basándose en el parámetro de correlación normalizada, el parámetro de inclinación espectral, el parámetro de relación señal/ruido, el parámetro de estabilidad de altura tonal, el parámetro de energía de trama relativa, y el parámetro de cruce por cero; y
- comparar la figura de mérito con umbrales para determinar la clasificación.
12. Método según la reivindicación 10, que comprende calcular el parámetro de correlación normalizada sobre la base de una versión ponderada actual de la señal de voz y una versión ponderada pasada de dicha señal de voz.
- 10 13. Método según la reivindicación 10, que comprende estimar el parámetro de inclinación espectral como una relación entre una energía concentrada en frecuencias bajas y una energía concentrada en frecuencias altas.
14. Método según la reivindicación 10, que comprende estimar el parámetro de relación de señal/ruido como una relación entre una energía de una versión ponderada de la señal de voz de una trama actual y una energía de un error entre dicha versión ponderada de la señal de voz de la trama actual y una versión ponderada de una señal de voz sintetizada de dicha trama actual.
- 15 15. Método según la reivindicación 10, que comprende calcular el parámetro de estabilidad de altura tonal como respuesta a estimaciones de la altura tonal en bucle abierto para una primera mitad de una trama actual, una segunda mitad de la trama actual y un anticipo.
16. Método según la reivindicación 10, que comprende calcular el parámetro de energía de trama relativa como una diferencia entre una energía de una trama actual y un promedio a largo plazo de una energía de tramas de voz activa.
- 20 17. Método según la reivindicación 10, que comprende determinar el parámetro de cruce por cero como el número de veces que cambia el signo de la señal de voz de una primera polaridad a una segunda polaridad.
18. Método según la reivindicación 10, que comprende calcular por lo menos uno de entre el parámetro de correlación normalizada, el parámetro de inclinación espectral, el parámetro de relación señal/ruido, el parámetro de estabilidad de altura tonal, el parámetro de energía de trama relativa, y el parámetro de cruce por cero, usando un anticipo disponible, para tener en cuenta el comportamiento de la señal de voz en una trama sucesiva.
- 25 19. Método según la reivindicación 10, que comprende determinar la clasificación de las tramas sucesivas de la señal de sonido codificada, también sobre la base de una bandera de detección de actividad vocal.
20. Método según la reivindicación 1, en el que la determinación, en el codificador, de parámetros de ocultación/recuperación comprende calcular el parámetro de información de sonoridad.
- 30 21. Método según la reivindicación 20, en el que:
- dicho método comprende determinar la clasificación de las tramas sucesivas de la señal de sonido codificada, sobre la base de un parámetro de correlación normalizada; y
- el cálculo del parámetro de información de sonoridad comprende estimar dicho parámetro de información de sonoridad sobre la base de la correlación normalizada.
- 35 22. Método según la reivindicación 1, en el que la ocultación del borrado de tramas y la recuperación del descodificador comprende:
- tras la recepción de una trama sorda no borrada, después de un borrado de trama, no se genera ninguna parte periódica de una señal de excitación del filtro de LP;
- 40            tras la recepción, después de un borrado de trama, de una trama no borrada que no sea sorda, construir una parte periódica de la señal de excitación del filtro de LP repitiendo un último periodo de altura tonal de una trama previa.
23. Método según la reivindicación 22, en el que la construcción de la parte periódica de la señal de excitación del filtro de LP comprende filtrar el último periodo de altura tonal repetido de la trama previa a través de un filtro paso-bajo.
- 45 24. Método según la reivindicación 23, en el que:
- la determinación de parámetros de ocultación/recuperación comprende calcular el parámetro de información de sonoridad;
- el filtro paso-bajo tiene una frecuencia de corte; y

la construcción de la parte periódica de la señal de excitación comprende ajustar dinámicamente la frecuencia de corte en relación con el parámetro de información de sonoridad.

5 25. Método según la reivindicación 1, en el que la ocultación del borrado de tramas y la recuperación del descodificador comprenden generar aleatoriamente una parte de innovación, no periódica, de una señal de excitación del filtro de LP.

26. Método según la reivindicación 25, en el que la generación aleatoria de la parte de innovación, no periódica, de la señal de excitación del filtro de LP comprende generar un ruido aleatorio.

10 27. Método según la reivindicación 25, en el que la generación aleatoria de la parte de innovación, no periódica, de la señal de excitación del filtro de LP, comprende generar aleatoriamente índices de los vectores de un libro de códigos de innovación.

28. Método según la reivindicación 25, en el que:

la generación aleatoria de la parte de innovación, no periódica, de la señal de excitación del filtro de LP comprende:

- 15
- si una última trama recibida correctamente no es sorda, filtrar la parte de innovación de la señal de excitación a través de un filtro paso-alto; y
  - si la última trama recibida correctamente es sorda, usar solamente la parte de innovación de la señal de excitación.

29. Método según la reivindicación 1, en el que:

20 la ocultación del borrado de tramas y la recuperación del descodificador comprende, cuando se pierde una trama de inicio lo cual se indica por la presencia de una trama sonora tras el borrado de la trama y una trama sorda antes del borrado de la trama, reconstruir artificialmente la trama de inicio perdida construyendo una parte periódica de una señal de excitación en forma de un tren periódico, filtrado por paso-bajo, de impulsos separados por un periodo de altura tonal.

25 30. Método según la reivindicación 29, en el que la ocultación del borrado de tramas y la recuperación del descodificador comprenden construir una parte de innovación de la señal de excitación por medio de descodificación normal.

31. Método según la reivindicación 30, en el que la construcción de una parte de innovación de la señal de excitación comprende seleccionar aleatoriamente entradas de un libro de códigos de innovación.

30 32. Método según la reivindicación 29, en el que la reconstrucción artificial de la trama de inicio perdida comprende limitar una longitud del inicio reconstruido artificialmente, de manera que se construye al menos un periodo de altura tonal completo por medio de la reconstrucción artificial del inicio, continuándose con dicha reconstrucción hasta la terminación de una subtrama actual.

35 33. Método según la reivindicación 32, en el que la ocultación del borrado de tramas y la recuperación del descodificador comprenden, después de la reconstrucción artificial del inicio perdido, reanudar un procesado de CELP regular en el que el periodo de la altura tonal es un promedio redondeado de periodos de altura tonal descodificados de subtramas donde se usa la reconstrucción de inicio artificial.

34. Método según la reivindicación 1, en el que la ocultación del borrado de tramas y la recuperación del descodificador comprenden:

40 controlar una energía de una señal de sonido sintetizada, producida por el descodificador, de manera que el control de la energía de la señal de sonido sintetizada comprende escalar la señal de sonido sintetizada para reproducir una energía de dicha señal de sonido sintetizada, en el comienzo de una primera trama no borrada, recibida tras el borrado de trama, similar a una energía de dicha señal de sonido sintetizada, en la terminación de una última trama borrada durante dicho borrado de trama; y

45 hacer converger la energía de la señal de sonido sintetizada, en la primera trama no borrada, recibida, a una energía correspondiente al parámetro recibido de información de energía, hacia la terminación de dicha primera trama no borrada recibida mientras se limita el aumento de energía.

35. Método según la reivindicación 1, en el que:

el parámetro de información de energía no se transmite desde el codificador al descodificador; y

50 la ocultación del borrado de tramas y la recuperación del descodificador comprenden, cuando la ganancia de un filtro de LP de una primera trama no borrada recibida tras un borrado de trama es mayor que la ganancia de

un filtro de LP de una última trama borrada durante dicho borrado de trama, ajustar la energía de una señal de excitación del filtro de LP producida en el descodificador durante la primera trama no borrada recibida, a una ganancia del filtro de LP de dicha primera trama no borrada recibida.

36. Método según la reivindicación 35, en el que:

5 el ajuste de la energía de la señal de excitación del filtro de LP, producida en el descodificador durante la primera trama no borrada recibida, a una ganancia del filtro de LP de dicha primera trama no borrada recibida, comprende usar la siguiente relación:

$$E_q = E_1 \frac{E_{LP0}}{E_{LP1}}$$

10 donde  $E_1$  es la energía en la terminación de la trama actual,  $E_{LP0}$  es la energía de la respuesta a impulsos del filtro de LP de una última trama no borrada recibida antes del borrado de trama, y  $E_{LP1}$  es la energía de la respuesta a impulsos del filtro de LP de la primera trama no borrada recibida, tras el borrado de trama.

37. Método según la reivindicación 34, en el que:

15 cuando la primera trama no borrada recibida después de un borrado de trama se clasifica como inicio, la ocultación del borrado de trama y la recuperación del descodificador comprenden limitar a un valor dado una ganancia usada para escalar la señal de sonido sintetizada.

38. Método según la reivindicación 34,

que comprende hacer que la ganancia usada para escalar la señal de sonido sintetizada en el comienzo de la primera trama no borrada recibida después del borrado de trama, sea igual a la ganancia usada en la terminación de dicha primera trama no borrada recibida:

20 - durante una transición de una trama sonora a una trama sorda, en el caso de una última trama no borrada recibida antes del borrado de trama, clasificada como transición sonora, sonora o inicio, y una primera trama no borrada recibida después del borrado de trama, clasificada como sorda; y

25 - durante una transición de un periodo de voz no activa a un periodo de voz activa, cuando la última trama no borrada recibida antes del borrado de trama se codifica como ruido de confort y la primera trama no borrada recibida después del borrado de trama se codifica como voz activa.

39. Método para la ocultación del borrado de tramas provocado por tramas borradas durante la transmisión de una señal de sonido codificada bajo la forma de parámetros de codificación de la señal, desde un codificador a un descodificador, y para acelerar la recuperación del descodificador después de que se hayan recibido tramas no borradas de la señal de sonido codificada, que comprende:

30 determinar, en el descodificador, parámetros de ocultación/recuperación a partir de los parámetros de codificación de la señal, comprendiendo los parámetros de ocultación/recuperación por lo menos dos parámetros seleccionados del grupo que consiste en un parámetro de clasificación de la señal, un parámetro de información de energía, un parámetro de información de sonoridad y un parámetro de información de fase; y

35 en el descodificador, ejecutar la ocultación de la trama borrada y la recuperación del descodificador como respuesta a los parámetros de ocultación/recuperación determinados en el descodificador;

en donde:

la señal de sonido es una señal de voz;

caracterizado por que:

40 la determinación, en el descodificador, de los parámetros de ocultación/recuperación comprende clasificar tramas sucesivas de la señal de sonido codificada, como sorda, transición sorda, transición sonora, sonora, o inicio; y

45 la determinación de los parámetros de ocultación/recuperación comprende calcular el parámetro de información de energía en relación con un máximo de una energía de señal para tramas clasificadas como sonoras o inicio, y calcular el parámetro de información de energía en relación con una energía promedio por muestra, para otras tramas.

40. Método según la reivindicación 39, en el que la determinación, en el descodificador, de parámetros de ocultación/recuperación comprende calcular el parámetro de información de sonoridad.

41. Método según la reivindicación 39, en el que la ejecución de la ocultación del borrado de tramas y la

recuperación del descodificador comprende:

tras la recepción de una trama sorda no borrada, después de un borrado de trama, no se genera ninguna parte periódica de una señal de excitación del filtro de LP;

5 tras la recepción, después de un borrado de trama, de una trama no borrada que no sea sorda, construir una parte periódica de la señal de excitación del filtro de LP repitiendo un último periodo de altura tonal de una trama previa.

42. Método según la reivindicación 41, en el que la construcción de la parte periódica de la señal de excitación comprende filtrar el último periodo de altura tonal repetido de la trama previa a través de un filtro paso-bajo.

43. Método según la reivindicación 42, en el que:

10 la determinación, en el descodificador, de parámetros de ocultación/recuperación comprende calcular el parámetro de información de sonoridad;

el filtro paso-bajo tiene una frecuencia de corte; y

la construcción de la parte periódica de la señal de excitación del filtro de LP comprende ajustar dinámicamente la frecuencia de corte en relación con el parámetro de información de sonoridad.

15 44. Método según la reivindicación 39, en el que la ejecución de la ocultación del borrado de tramas y la recuperación del descodificador comprende generar aleatoriamente una parte de innovación, no periódica, de una señal de excitación del filtro de LP.

45. Método según la reivindicación 44, en el que la generación aleatoria de la parte de innovación, no periódica, de la señal de excitación del filtro de LP comprende generar un ruido aleatorio.

20 46. Método según la reivindicación 44, en el que la generación aleatoria de la parte de innovación, no periódica, de la señal de excitación del filtro de LP, comprende generar aleatoriamente índices de los vectores de un libro de códigos de innovación.

47. Método según la reivindicación 44, en el que:

25 la generación aleatoria de la parte de innovación, no periódica, de la señal de excitación del filtro de LP comprende:

- si una última trama no borrada recibida no es sorda, filtrar la parte de innovación de la señal de excitación del filtro de LP a través de un filtro paso-alto; y
- si la última trama no borrada recibida es sorda, usar solamente la parte de innovación de la señal de excitación del filtro de LP.

30 48. Método según la reivindicación 39, en el que:

la ejecución de la ocultación del borrado de tramas y la recuperación del descodificador comprende, cuando se pierde una trama de inicio lo cual se indica por la presencia de una trama sonora tras el borrado de la trama y una trama sorda antes del borrado de la trama, reconstruir artificialmente la trama de inicio perdida construyendo una parte periódica de una señal de excitación en forma de un tren periódico, filtrado por paso-bajo, de impulsos separados por un periodo de altura tonal.

35 49. Método según la reivindicación 48, en el que la ejecución de la ocultación del borrado de tramas y la recuperación del descodificador comprende construir una parte de innovación de la señal de excitación por medio de decodificación normal.

40 50. Método según la reivindicación 48, en el que la ejecución de la ocultación del borrado de tramas y de la recuperación del descodificador comprende construir una parte de innovación de la señal de excitación seleccionando aleatoriamente entradas de un libro de códigos de innovación.

45 51. Método según la reivindicación 48, en el que la reconstrucción artificial de la trama de inicio perdida comprende limitar una longitud del inicio reconstruido artificialmente, de manera que se construye al menos un periodo de altura tonal completo por medio de la reconstrucción artificial del inicio, continuándose con dicha reconstrucción hasta la terminación de una subtrama actual.

52. Método según la reivindicación 51, en el que la ejecución de la ocultación del borrado de tramas y de la recuperación del descodificador comprende, después de la reconstrucción artificial del inicio perdido, reanudar un procesado de CELP regular en el que el periodo de la altura tonal es un promedio redondeado de periodos de altura tonal descodificados de subtramas donde se usa la reconstrucción de inicio artificial.

53. Método según la reivindicación 39, en el que:

el parámetro de información de energía no se transmite desde el codificador al descodificador; y

5 la ejecución de la ocultación del borrado de tramas y de la recuperación del descodificador comprende, cuando la ganancia de un filtro de LP de una primera trama no borrada recibida tras un borrado de trama es mayor que la ganancia de un filtro de LP de una última trama borrada durante dicho borrado de trama, ajustar la energía de una señal de excitación del filtro de LP producida en el descodificador durante la primera trama no borrada recibida, a una ganancia del filtro de LP de dicha primera trama no borrada recibida, usando la siguiente relación:

$$E_q = E_1 \frac{E_{LP0}}{E_{LP1}}$$

10 donde  $E_1$  es la energía en la terminación de la trama actual,  $E_{LP0}$  es la energía de la respuesta a impulsos del filtro de LP de una última trama no borrada recibida antes del borrado de trama, y  $E_{LP1}$  es la energía de la respuesta a impulsos del filtro de LP de la primera trama no borrada recibida, tras el borrado de trama.

54. Dispositivo para ejecutar la ocultación del borrado de tramas provocado por tramas de una señal de sonido codificada, borradas durante su transmisión desde un codificador a un descodificador, y para acelerar la recuperación del descodificador después de que se hayan recibido tramas no borradas de la señal de sonido codificada, que comprende:

15 medios para determinar, en el codificador, parámetros de ocultación/recuperación que comprenden por lo menos dos parámetros seleccionados del grupo que consiste en un parámetro de clasificación de la señal, un parámetro de información de energía, un parámetro de información de sonoridad y un parámetro de información de fase;

20 medios para cuantificar los parámetros de ocultación/recuperación; y

medios para transmitir al descodificador los parámetros de ocultación/recuperación cuantificados, determinados en el codificador;

en donde:

25 los parámetros de ocultación/recuperación son utilizables para mejorar la ocultación del borrado de tramas y la recuperación del descodificador después de un borrado de tramas; y

la señal de sonido es una señal de voz;

caracterizado por que:

30 los medios para determinar, en el codificador, los parámetros de ocultación/recuperación comprenden medios para clasificar tramas sucesivas de la señal de sonido codificada, como sorda, transición sorda, transición sonora, sonora, o inicio; y

35 los medios para determinar los parámetros de ocultación/recuperación comprenden medios para calcular el parámetro de información de energía en relación con un máximo de una energía de señal para tramas clasificadas como sonoras o inicio, y medios para calcular el parámetro de información de energía en relación con una energía promedio por muestra, para otras tramas.

55. Dispositivo según la reivindicación 54, en el que los medios para determinar el parámetro de información de fase comprenden medios para determinar una posición de un primer pulso glotal en una trama de la señal de sonido codificada.

56. Dispositivo según la reivindicación 55, en el que los medios para determinar el parámetro de información de fase comprenden además medios para codificar, en el codificador, una forma, un signo y una amplitud del primer pulso glotal, y medios para transmitir la forma, el signo y la amplitud codificados, desde el codificador al descodificador.

57. Dispositivo según la reivindicación 55, en el que los medios para determinar la posición del primer pulso glotal comprenden:

45 medios para medir una muestra de amplitud máxima dentro de un periodo de altura tonal como primer pulso glotal; y

medios para cuantificar la posición de la muestra de amplitud máxima dentro del periodo de la altura tonal.

58. Dispositivo según la reivindicación 54, en el que los medios para clasificar las tramas sucesivas comprenden

medios para clasificar como sorda, toda trama que sea una trama sorda, toda trama sin voz activa, y toda trama de final sonora que tenga una terminación que tienda a ser sorda.

59. Dispositivo según la reivindicación 54, en el que los medios para clasificar las tramas sucesivas comprenden medios para clasificar como transición sorda, toda trama sorda que tenga una terminación con un posible inicio sonoro que sea demasiado corto o que no esté construido suficientemente bien para ser procesado como trama sonora.
60. Dispositivo según la reivindicación 54, en el que los medios para clasificar las tramas sucesivas comprenden medios para clasificar como transición sonora, toda trama sonora con características sonoras relativamente débiles, incluyendo tramas sonoras con características que varían rápidamente y finales sonoros que duran la trama completa, en donde una trama clasificada como transición sonora sucede únicamente a tramas clasificadas como transición sonora, sonora o inicio.
61. Dispositivo según la reivindicación 54, en el que los medios para clasificar las tramas sucesivas comprenden medios para clasificar como sonora, toda trama sonora con características estables, en donde una trama clasificada como sonora sucede únicamente a tramas clasificadas como transición sonora, sonora o inicio.
62. Dispositivo según la reivindicación 54, en el que los medios para clasificar las tramas sucesivas comprenden medios para clasificar como inicio toda trama sonora con características estables que suceda a una trama clasificada como sorda o transición sorda.
63. Dispositivo según la reivindicación 54, que comprende medios para determinar la clasificación de las tramas sucesivas de la señal de sonido codificada, sobre la base de al menos una parte de los siguientes parámetros: un parámetro de correlación normalizada, un parámetro de inclinación espectral, un parámetro de relación señal/ruido, un parámetro de estabilidad de altura tonal, un parámetro de energía de trama relativa, y un parámetro de cruce por cero.
64. Dispositivo según la reivindicación 63, en el que los medios para determinar la clasificación de las tramas sucesivas comprenden:
- medios para calcular una figura de mérito basándose en el parámetro de correlación normalizada, el parámetro de inclinación espectral, el parámetro de relación señal/ruido, el parámetro de estabilidad de altura tonal, el parámetro de energía de trama relativa, y el parámetro de cruce por cero; y
- medios para comparar la figura de mérito con umbrales con el fin de determinar la clasificación.
65. Dispositivo según la reivindicación 63, que comprende medios para calcular el parámetro de correlación normalizada sobre la base de una versión ponderada actual de la señal de voz y una versión ponderada pasada de dicha señal de voz.
66. Dispositivo según la reivindicación 63, que comprende medios para estimar el parámetro de inclinación espectral como una relación entre una energía concentrada en frecuencias bajas y una energía concentrada en frecuencias altas.
67. Dispositivo según la reivindicación 63, que comprende medios para estimar el parámetro de relación de señal/ruido como una relación entre una energía de una versión ponderada de la señal de voz de una trama actual y una energía de un error entre dicha versión ponderada de la señal de voz de la trama actual y una versión ponderada de una señal de voz sintetizada de dicha trama actual.
68. Dispositivo según la reivindicación 63, que comprende medios para calcular el parámetro de estabilidad de altura tonal como respuesta a estimaciones de la altura tonal en bucle abierto para una primera mitad de una trama actual, una segunda mitad de la trama actual y un anticipo.
69. Dispositivo según la reivindicación 63, que comprende medios para calcular el parámetro de energía de trama relativa como una diferencia entre una energía de una trama actual y un promedio a largo plazo de una energía de tramas de voz activa.
70. Dispositivo según la reivindicación 63, que comprende medios para determinar el parámetro de cruce por cero como el número de veces que cambia el signo de la señal de voz de una primera polaridad a una segunda polaridad.
71. Dispositivo según la reivindicación 63, que comprende medios para calcular por lo menos uno de entre el parámetro de correlación normalizada, el parámetro de inclinación espectral, el parámetro de relación señal/ruido, el parámetro de estabilidad de altura tonal, el parámetro de energía de trama relativa, y el parámetro de cruce por cero, usando un anticipo disponible, para tener en cuenta el comportamiento de la señal de voz en una trama sucesiva.
72. Dispositivo según la reivindicación 63, que además comprende medios para determinar la clasificación de las tramas sucesivas de la señal de sonido codificada, también sobre la base de una bandera de detección de actividad vocal.

73. Dispositivo según la reivindicación 63, en el que los medios para determinar, en el codificador, parámetros de ocultación/recuperación comprenden medios para calcular el parámetro de información de sonoridad.
74. Dispositivo según la reivindicación 73, en el que:
- 5 dicho dispositivo comprende medios para determinar la clasificación de las tramas sucesivas de la señal de sonido codificada, sobre la base de un parámetro de correlación normalizada; y
- los medios para calcular el parámetro de información de sonoridad comprenden medios para estimar dicho parámetro de información de sonoridad sobre la base de la correlación normalizada.
75. Dispositivo para la ocultación del borrado de tramas provocado por tramas borradas durante la transmisión de una señal de sonido codificada bajo la forma de parámetros de codificación de la señal, desde un codificador a un descodificador, y para acelerar la recuperación del descodificador después de que se hayan recibido tramas no borradas de la señal de sonido codificada, que comprende:
- 10 medios para determinar, en el descodificador, parámetros de ocultación/recuperación a partir de los parámetros de codificación de la señal, comprendiendo los parámetros de ocultación/recuperación por lo menos dos parámetros seleccionados del grupo que consiste en un parámetro de clasificación de la señal, un parámetro de información de energía, un parámetro de información de sonoridad y un parámetro de información de fase;
- 15 en el descodificador, medios para ejecutar la ocultación de la trama borrada y la recuperación del descodificador como respuesta a los parámetros de ocultación/recuperación determinados por los medios de determinación;
- en donde:
- 20 la señal de sonido es una señal de voz;
- caracterizado por que:
- los medios para determinar, en el descodificador, los parámetros de ocultación/recuperación comprenden medios para clasificar tramas sucesivas de la señal de sonido codificada, como sorda, transición sorda, transición sonora, sonora, o inicio; y
- 25 los medios para determinar los parámetros de ocultación/recuperación comprenden medios para calcular el parámetro de información de energía en relación con un máximo de una energía de señal para tramas clasificadas como sonoras o inicio, y medios para calcular el parámetro de información de energía en relación con una energía promedio por muestra, para otras tramas.
76. Dispositivo según la reivindicación 75, en el que los medios para determinar, en el descodificador, parámetros de ocultación/recuperación comprenden medios para calcular el parámetro de información de sonoridad.
- 30 77. Dispositivo según la reivindicación 75, en el que los medios para ejecutar la ocultación del borrado de tramas y la recuperación del descodificador comprenden:
- tras la recepción de una trama sorda no borrada, después de un borrado de trama, medios de generación que no generan ninguna parte periódica de una señal de excitación del filtro de LP;
- 35 tras la recepción, después de un borrado de trama, de una trama no borrada que no sea sorda, medios para construir una parte periódica de la señal de excitación del filtro de LP repitiendo un último periodo de altura tonal de una trama previa.
78. Dispositivo según la reivindicación 77, en el que los medios para construir la parte periódica de la señal de excitación comprenden un filtro paso-bajo para filtrar el último periodo de altura tonal repetido de la trama previa.
- 40 79. Dispositivo según la reivindicación 78, en el que:
- los medios para determinar, en el descodificador, parámetros de ocultación/recuperación comprenden medios para calcular el parámetro de información de sonoridad;
- el filtro paso-bajo tiene una frecuencia de corte; y
- 45 los medios para construir la parte periódica de la señal de excitación del filtro de LP comprenden medios para ajustar dinámicamente la frecuencia de corte en relación con el parámetro de información de sonoridad.
80. Dispositivo según la reivindicación 75, en el que los medios para ejecutar la ocultación del borrado de tramas y la recuperación del descodificador comprenden medios para generar aleatoriamente una parte de innovación, no periódica, de una señal de excitación del filtro de LP.



81. Dispositivo según la reivindicación 80, en el que los medios para generar aleatoriamente la parte de innovación, no periódica, de la señal de excitación del filtro de LP comprenden medios para generar un ruido aleatorio.

82. Dispositivo según la reivindicación 80, en el que los medios para generar aleatoriamente la parte de innovación, no periódica, de la señal de excitación del filtro de LP, comprenden medios para generar aleatoriamente índices de los vectores de un libro de códigos de innovación.

83. Dispositivo según la reivindicación 80, en el que:

los medios para generar aleatoriamente la parte de innovación, no periódica, de la señal de excitación del filtro de LP comprenden:

- si una última trama no borrada, recibida, no es sorda, un filtro paso-alto para filtrar la parte de innovación de la señal de excitación del filtro de LP; y
- si la última trama no borrada, recibida es sorda, medios para usar solamente la parte de innovación de la señal de excitación del filtro de LP.

84. Dispositivo según la reivindicación 75, en el que:

los medios para ejecutar la ocultación del borrado de tramas y la recuperación del descodificador comprenden, cuando se pierde una trama de inicio lo cual se indica por la presencia de una trama sonora tras el borrado de la trama y una trama sorda antes del borrado de la trama, medios para reconstruir artificialmente el inicio perdido construyendo una parte periódica de una señal de excitación en forma de un tren periódico, filtrado por paso-bajo, de impulsos separados por un periodo de altura tonal.

85. Dispositivo según la reivindicación 83, en el que los medios para ejecutar la ocultación del borrado de tramas y la recuperación del descodificador comprenden además medios para construir una parte de innovación de la señal de excitación del filtro de LP por medio de descodificación normal.

86. Dispositivo según la reivindicación 85, en el que los medios para construir una parte de innovación de la señal de excitación del filtro de LP comprenden medios para seleccionar aleatoriamente entradas de un libro de códigos de innovación.

87. Dispositivo según la reivindicación 84, en el que los medios para reconstruir artificialmente el inicio perdido comprenden medios para limitar una longitud del inicio reconstruido artificialmente, de manera que se construye al menos un periodo de altura tonal completo por medio de la reconstrucción artificial del inicio, continuándose con dicha reconstrucción hasta la terminación de una subtrama actual.

88. Dispositivo según la reivindicación 87, en el que los medios para ejecutar la ocultación del borrado de tramas y la recuperación del descodificador comprenden además, después de la reconstrucción artificial del inicio perdido, medios para reanudar un procesado de CELP regular en el que el periodo de la altura tonal es un promedio redondeado de periodos de altura tonal descodificados de subtramas donde se usa la reconstrucción de inicio artificial.

89. Dispositivo según la reivindicación 75, en el que:

el parámetro de información de energía no se transmite desde el codificador al descodificador; y

los medios para ejecutar la ocultación del borrado de tramas y la recuperación del descodificador comprenden, cuando la ganancia de un filtro de LP de una primera trama no borrada recibida tras un borrado de trama es mayor que la ganancia de un filtro de LP de una última trama borrada durante dicho borrado de trama, medios para ajustar la energía de una señal de excitación del filtro de LP producida en el descodificador durante la primera trama no borrada recibida, a una ganancia del filtro de LP de dicha primera trama no borrada recibida, usando la siguiente relación:

$$E_q = E_1 \frac{E_{LP0}}{E_{LP1}}$$

donde  $E_1$  es la energía en la terminación de la trama actual,  $E_{LP0}$  es la energía de la respuesta a impulsos del filtro de LP de una última trama no borrada recibida antes del borrado de trama, y  $E_{LP1}$  es la energía de la respuesta a impulsos del filtro de LP para la primera trama no borrada recibida, tras el borrado de trama.

90. Descodificador para descodificar una señal de sonido codificada, que comprende:

medios sensibles a la señal de sonido codificada, para recuperar, a partir de dicha señal de sonido codificada, un conjunto de parámetros de codificación de la señal;

medios para sintetizar la señal de sonido como respuesta al conjunto de parámetros de codificación de la señal; y

5 un dispositivo según se cita en una cualquiera de las reivindicaciones 75 a 89, para ocultar el borrado de tramas provocado por tramas de la señal de sonido codificada, borradas durante la transmisión desde un codificador al decodificador.

91. Codificador para codificar una señal de sonido, que comprende:

medios sensibles a la señal de sonido para producir un conjunto de parámetros de codificación de la señal;

medios para transmitir el conjunto de parámetros de codificación de la señal a un decodificador sensible a los parámetros de codificación de la señal para recuperar la señal de sonido; y

10 un dispositivo según se cita en cualquiera de las reivindicaciones 54 a 74, para ejecutar la ocultación del borrado de tramas provocado por tramas borradas durante la transmisión de los parámetros de codificación de la señal, desde el codificador al decodificador.

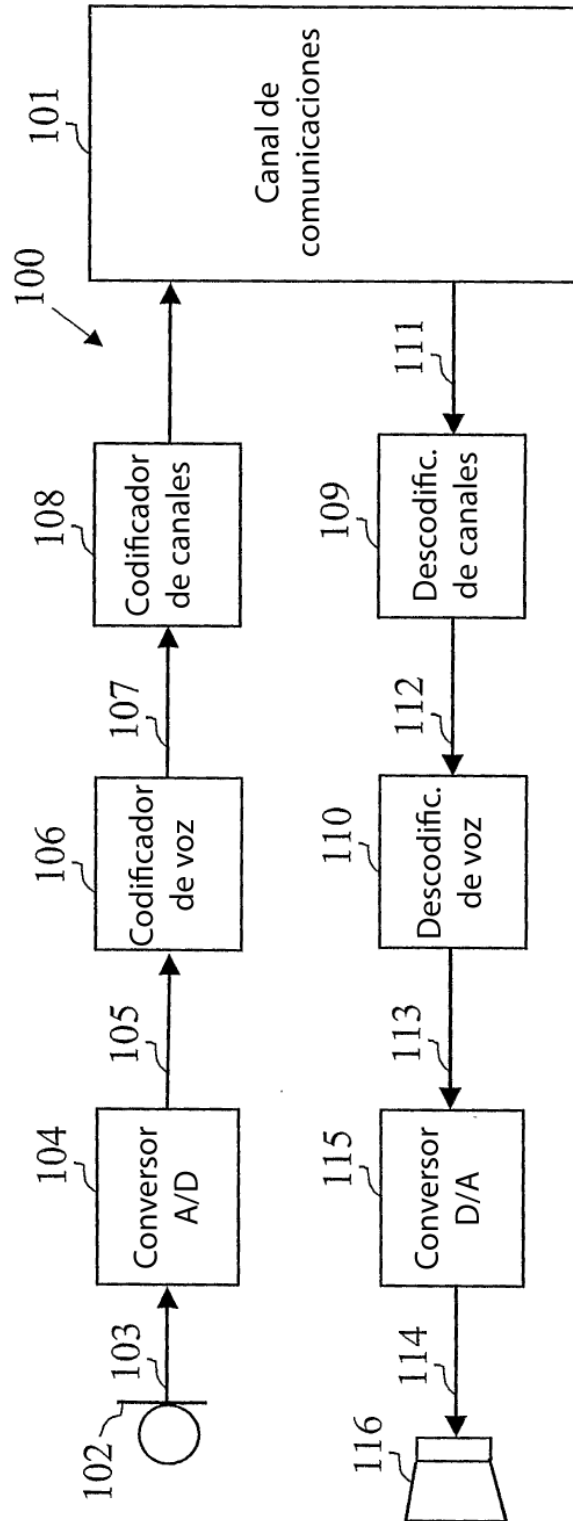


FIG. 1

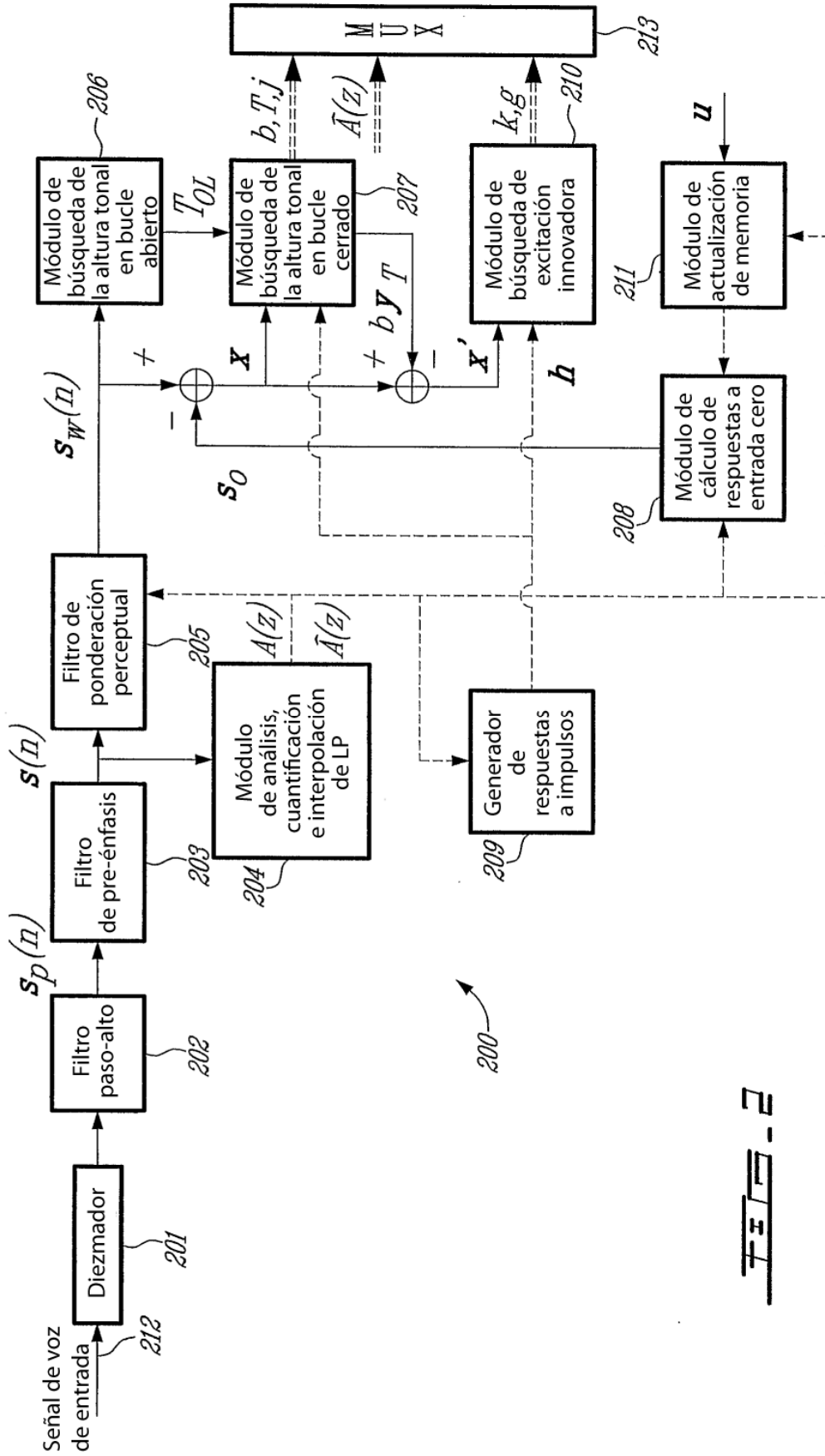
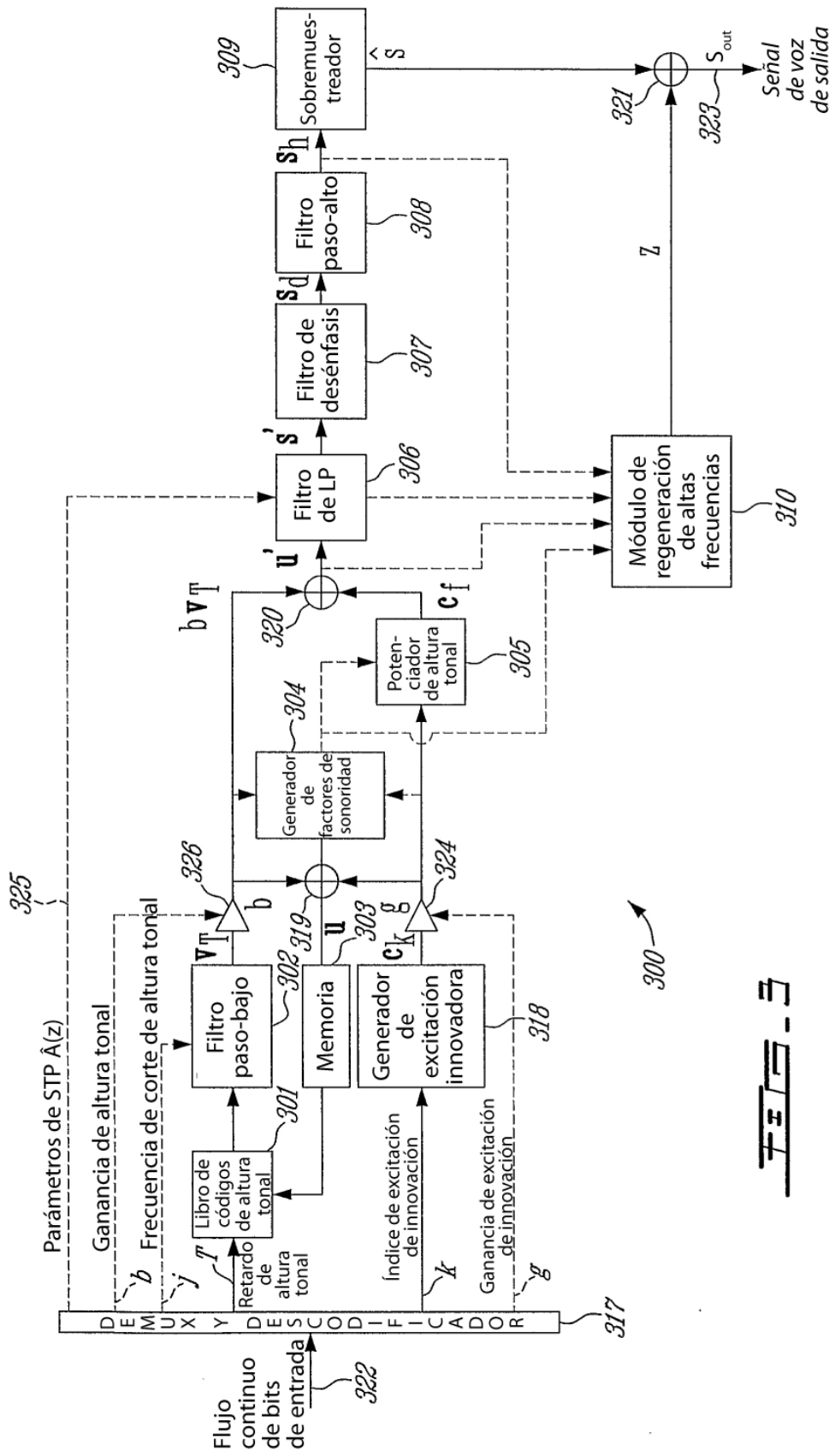
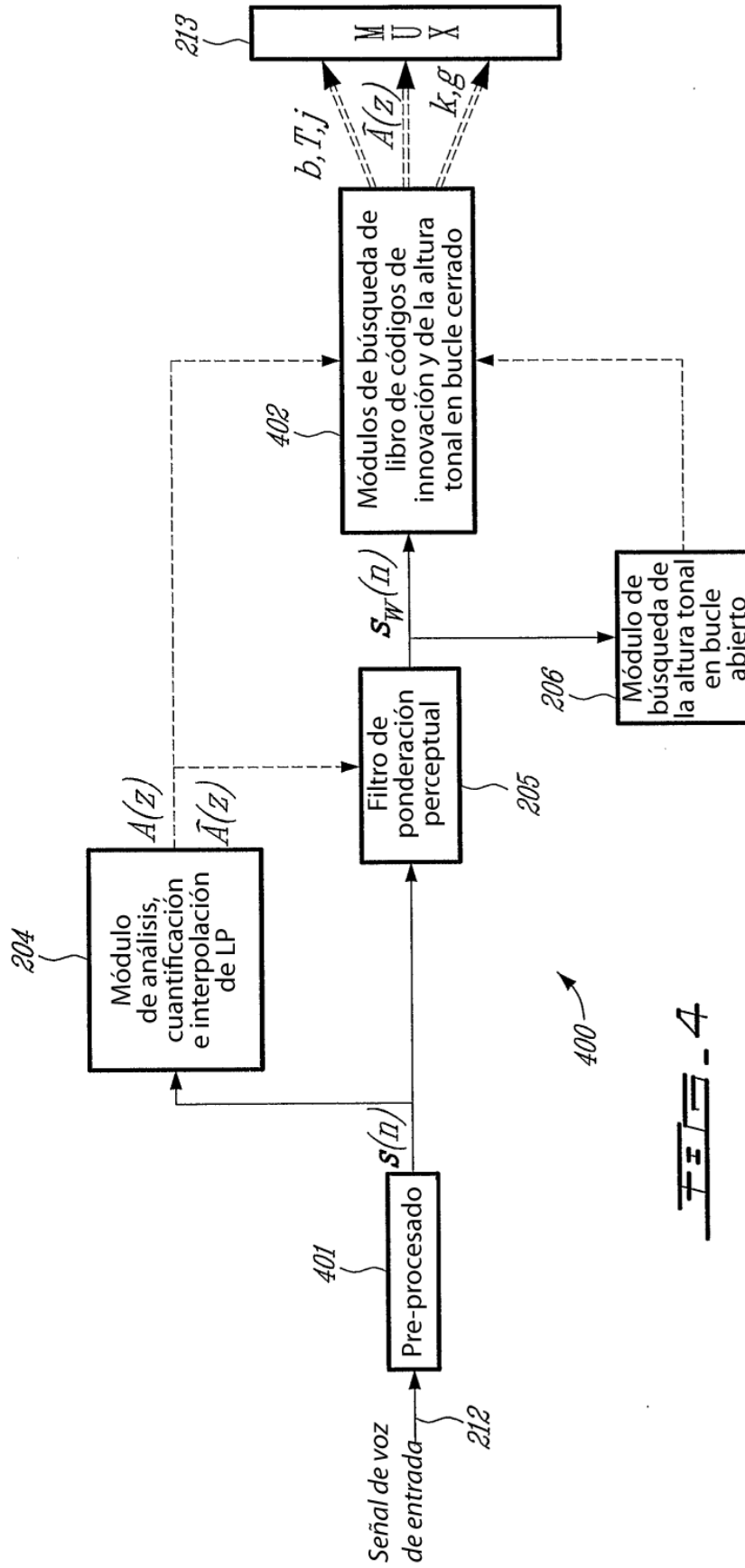


FIG. 2

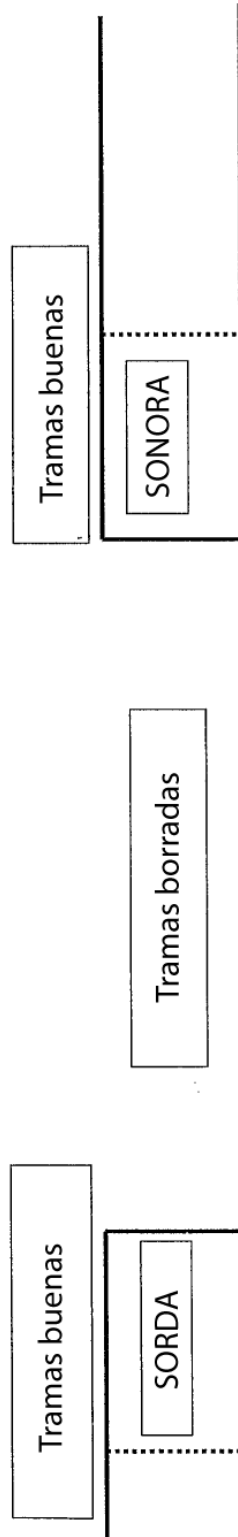


FEPA - 3



**FIG. 4**





*FECS.B*



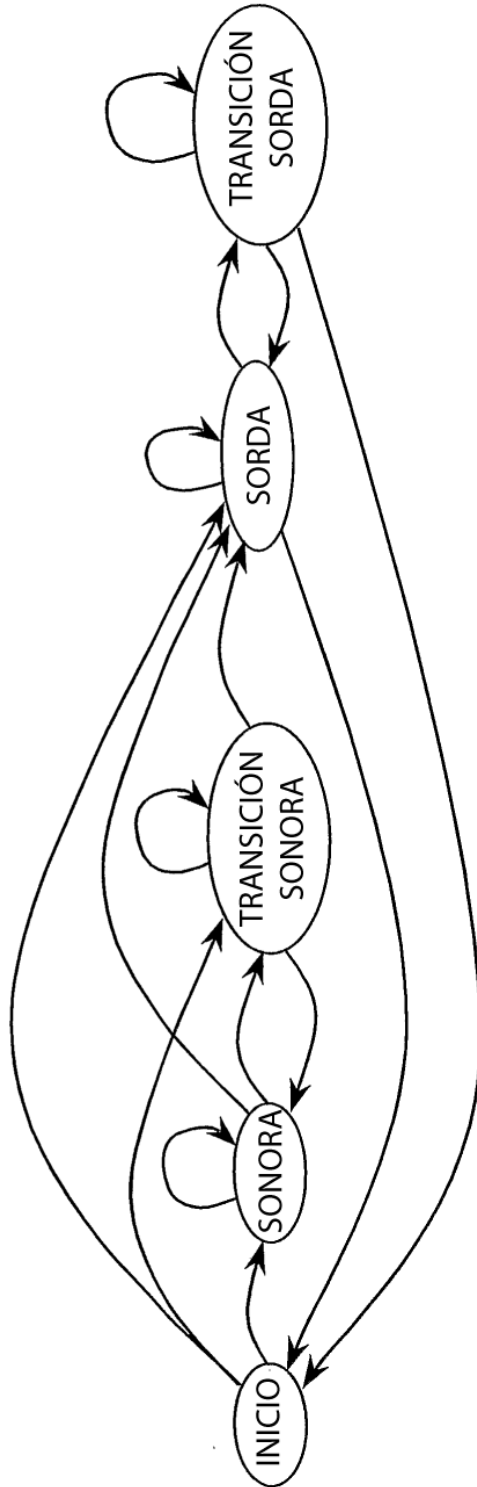


Fig. 7