

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 434 947**

51 Int. Cl.:

**G10L 19/00** (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **28.12.2006 E 06840572 (9)**

97 Fecha y número de publicación de la concesión europea: **09.10.2013 EP 1979895**

54 Título: **Procedimiento y dispositivo para la ocultación eficiente de un borrado de trama en códecs de voz**

30 Prioridad:

**28.12.2005 US 754187 P**

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

**18.12.2013**

73 Titular/es:

**VOICEAGE CORPORATION (100.0%)  
Suite 200, 750, chemin Lucerne  
Ville Mont-Royal Quebec H3R 2H6, CA**

72 Inventor/es:

**VAILLANCOURT, TOMMY;  
JELINEK, MILAN;  
GOURNAY, PHILIPPE y  
SALAMI, REDWAN**

74 Agente/Representante:

**DE ELZABURU MÁRQUEZ, Alberto**

**ES 2 434 947 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

Procedimiento y dispositivo para la ocultación eficiente de un borrado de trama en códecs de voz

**Campo de la invención**

5 La presente invención se refiere a una técnica para codificar digitalmente una señal de sonido, en particular, pero no exclusivamente, una señal de voz, para transmitir y/o sintetizar esta señal de sonido. Más específicamente, la presente invención se refiere a una codificación y decodificación robusta de señales de sonido para mantener un buen rendimiento en el caso de la trama o tramas borradas, debido, por ejemplo, a errores de canal en los sistemas inalámbricos o paquetes perdidos en aplicaciones de voz sobre red por paquetes.

**Antecedentes de la invención**

10 La demanda de técnicas digitales eficaces de codificación de voz de banda estrecha y banda ancha con un buen compromiso entre la calidad subjetiva y la tasa de bits está aumentando en varias áreas de aplicación, tales como teleconferencia, multimedia y comunicaciones inalámbricas. Hasta hace poco, se ha usado principalmente un ancho de banda telefónico restringido a un intervalo de 200-3.400 Hz para aplicaciones de codificación de voz. Sin embargo, las aplicaciones de voz de banda ancha proporcionan una mayor inteligibilidad y naturalidad en la comunicación en comparación con el ancho de banda telefónico convencional. Se ha encontrado que un ancho de banda en el intervalo de 50-7.000 Hz es suficiente para proporcionar una buena calidad dando una impresión de una comunicación cara a cara. Para las señales de audio en general, este ancho de banda proporciona una calidad subjetiva aceptable, pero es todavía inferior a la calidad de la radio FM o del CD que funcionan en intervalos de 20-16.000 Hz y 20-20.000 Hz, respectivamente.

20 Un codificador de voz convierte una señal de voz a un flujo de bits digital que es transmitido a través de un canal de comunicación o es almacenado en un medio de almacenamiento. La señal de voz es digitalizada, es decir, muestreada y cuantificada normalmente con 16 bits por muestra. El codificador de voz tiene el papel de representar estas muestras digitales con un menor número de bits, mientras mantiene una buena calidad de voz subjetiva. El decodificador de voz o sintetizador opera sobre el flujo de bits transmitido o almacenado y lo convierte de nuevo a una señal de sonido.

25 La codificación Predicción Lineal con Excitación por Código (Code-Excited Linear Prediction, CELP) es una de las mejores técnicas disponibles para conseguir un buen compromiso entre la calidad subjetiva y la tasa de bits. Esta técnica de codificación es una base de diversos estándares de codificación de voz, tanto en aplicaciones inalámbricas como por cable. En la codificación CELP, la señal de voz muestreada es procesada en bloques sucesivos de L muestras denominados, normalmente tramas, donde L es un número predeterminado que corresponde, típicamente, a 10-30 ms de la señal de voz. Un filtro de predicción lineal (LP) es calculado y transmitido cada trama. Típicamente, el cálculo del filtro LP necesita una anticipación, un segmento de voz de 5-15 ms de la trama subsiguiente. La trama de L muestras es dividida en bloques más pequeños denominados subtramas. Normalmente, el número de subtramas es de tres o cuatro, lo que da como resultado subtramas de 4-10 ms. En cada subtrama, generalmente se obtiene una señal de excitación a partir de dos componentes, la excitación pasada y la excitación innovadora, con un libro de códigos fijo. El componente formado a partir de la excitación pasada se conoce frecuentemente como el libro de códigos adaptativo o excitación de frecuencia fundamental. Los parámetros que caracterizan la señal de excitación se codifican y transmiten al decodificador, donde la señal de excitación reconstruida se usa como la entrada del filtro LP.

30 Debido a que las principales aplicaciones de codificación de voz de baja tasa de bits son sistemas de comunicaciones móviles inalámbricas y de voz a través de redes de paquetes, entonces, el aumento de la robustez de los códecs de voz en caso de borrado de tramas adquiere una importancia significativa. En los sistemas celulares inalámbricos, la energía de la señal recibida puede exhibir desvanecimientos graves frecuentes que resultan en altas tasas de error de bit y esto se hace más evidente en los límites de las celdas. En este caso, el decodificador de canal no consigue corregir los errores en la trama recibida y, como consecuencia, el detector de errores usado normalmente después del decodificador de canal declarará la trama como borrada. En aplicaciones de voz sobre redes de paquetes, la señal de voz es empaquetada de manera que, normalmente, cada paquete corresponde a 20-40 ms de la señal de sonido. En las comunicaciones de conmutación de paquetes, puede producirse una eliminación de paquetes en un router si el número de paquetes se hace muy grande, o el paquete puede llegar al receptor después de un largo retardo y debe ser declarado como perdido si su retardo es mayor que la longitud de una memoria intermedia anti-fluctuaciones en el lado del receptor. En estos sistemas, típicamente el códec es sometido a tasas de borrado de trama del 3 al 5%. Además, el uso de codificación de voz de banda ancha es un activo de estos sistemas para permitir que puedan competir con la PSTN (red telefónica pública conmutada) tradicional que usa las señales de voz de banda estrecha heredadas.

55 En CLP el libro de códigos adaptativo o el predictor de frecuencia fundamental desempeña un papel en el

mantenimiento de una alta calidad de voz a bajas tasas de bits. Sin embargo, debido a que el contenido del libro de códigos adaptativo está basado en la señal de las tramas pasadas, esto hace que el modelo de 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 decodificador es diferente de su contenido en el codificador. De esta manera, después de que una trama perdida es ocultada y se reciben tramas buenas subsiguientes, la señal sintetizada en las tramas buenas recibidas es diferente de la señal de síntesis deseada ya que la contribución del libro de códigos adaptativo ha sido modificada. El impacto de una trama perdida depende de la naturaleza del segmento de voz en el que ocurrió el borrado. Si el borrado ocurre en un segmento estacionario de la señal, entonces puede realizarse una ocultación eficiente del borrado de trama y puede minimizarse el Impacto sobre las tramas buenas subsiguientes. Por otro lado, si el borrado ocurre en un inicio o una transición de la voz, el efecto del borrado puede propagarse a lo largo de varias tramas. Por ejemplo, si se pierde el principio de un segmento sonoro, entonces faltará el primer periodo fundamental del contenido del libro de códigos adaptativo. Esto tendrá un severo efecto sobre el predictor de frecuencia fundamental en las tramas buenas subsiguientes, resultando en un tiempo más largo antes de que la señal de síntesis converja a la señal deseada en el codificador.

5

10

### 15 **Sumario de la invención**

Más específicamente, según un primer aspecto de la presente invención, se proporciona un procedimiento para ocultar borrados de tramas causados por tramas de una señal de sonido codificada que han sido borradas durante la transmisión desde un codificador a un decodificador y para la recuperación del decodificador después de los borrados de tramas, según la reivindicación 1.

20

Según un segundo aspecto de la presente invención, se proporciona un procedimiento alternativo para ocultar borrados de tramas causados por tramas de una señal de sonido codificada que han sido borradas durante la transmisión desde un codificador a un decodificador y para la recuperación del decodificador después de los borrados según la reivindicación 24.

25

Según un tercer aspecto de la presente invención, se proporciona un dispositivo para ocultar borrados de tramas causados por tramas de una señal de sonido codificada que han sido borradas durante la transmisión desde un codificador a un decodificador y para la recuperación del decodificador después de los borrados de tramas según la reivindicación 34.

30

Según un cuarto aspecto de la presente invención, se proporciona un dispositivo alternativo para ocultar borrados de tramas causados por las tramas de una señal de sonido codificada que han sido borradas durante la transmisión desde un codificador a un decodificador y para la recuperación del decodificador después de los borrados de tramas según la reivindicación 48.

Lo indicado anteriormente y otros objetos, ventajas y características de la presente invención serán más evidentes tras la lectura de la siguiente descripción no restrictiva de una realización ilustrativa de la misma, proporcionada a modo de ejemplo solamente, con referencia a los dibujos adjuntos.

35

### **Breve descripción de los dibujos**

En los dibujos adjuntos:

La Figura 1 es un diagrama de bloques esquemático de un sistema de comunicación de voz que ilustra un ejemplo de aplicación de dispositivos de codificación y decodificación de voz;

La Figura 2 es un diagrama de bloques esquemático de un ejemplo de un dispositivo de codificación CELP;

40

La Figura 3 es un diagrama de bloques esquemático de un ejemplo de un dispositivo de decodificación CELP;

La Figura 4 es un diagrama de bloques esquemático de un codificador integrado basado en el núcleo G.729 (G.729 se refiere a la Recomendación UIT-T G.729);

La Figura 5 es un diagrama de bloques esquemático de un decodificador integrado basado en núcleo G.729;

45

La Figura 6 es un diagrama de bloques simplificado del dispositivo de codificación CELP de la Figura 2, en el que el módulo de búsqueda de frecuencia fundamental, de bucle cerrado, el módulo calculador de respuesta de entrada cero, el módulo generador de respuesta a impulso, el módulo de búsqueda de excitación innovadora y el módulo de actualización de memoria han sido agrupados en un único módulo de búsqueda de frecuencia fundamental, de bucle cerrado, y búsqueda de libro de códigos innovador.

50

La Figura 7 es una extensión del diagrama de bloques de la Figura 4, en la que se han añadido los módulos relacionados con los parámetros para mejorar la ocultación/recuperación;

La Figura 8 es un diagrama esquemático que muestra un ejemplo de máquina de estados de clasificación de tramas para la ocultación de borrados;

La Figura 9 es un diagrama de flujo que muestra un procedimiento de ocultación de la parte periódica de la excitación según la realización ilustrativa, no restrictiva, de la presente invención;

5 La Figura 10 es un diagrama de flujo que muestra un procedimiento de sincronización de la parte periódica de la excitación según la realización ilustrativa, no restrictiva, de la presente invención;

La Figura 11 muestra ejemplos típicos de la señal de excitación con y sin el procedimiento de sincronización;

La Figura 12 muestra ejemplos de la señal de voz reconstruida usando las señales de excitación mostradas en la Figura 11; y

10 La Figura 13 es un diagrama de bloques que ilustra un ejemplo de un caso cuando se pierde una trama de inicio.

### Descripción detallada

Aunque la realización ilustrativa de la presente invención se describirá en la siguiente descripción con relación a una señal de voz, debería recordarse que los conceptos de la presente invención se aplican igualmente a otros tipos de señal, en particular, pero no exclusivamente, a otros tipos de señales de sonido.

15 La Figura 1 ilustra un sistema 100 de comunicación de voz que representa el uso de codificación y decodificación de voz en un contexto ilustrativo de la presente invención. El sistema 100 de comunicación de voz de la Figura 1 soporta la transmisión de una señal de voz a través de un canal 101 de comunicación. Aunque puede comprender por ejemplo un cable, un enlace óptico o un enlace de fibra, el canal 101 de comunicación comprende, típicamente, al menos en parte, un enlace de radiofrecuencia. Frecuentemente, dicho un enlace de radiofrecuencia soporta  
20 múltiples comunicaciones de voz simultáneas, que requieren recursos de ancho de banda compartidos, tal como puede encontrarse con los sistemas de telefonía celular. Aunque no se muestra, el canal 101 de comunicación puede ser remplazado por un dispositivo de almacenamiento en una realización de un único dispositivo del sistema 100, para registrar y almacenar la señal de voz codificada para su posterior reproducción.

25 En el sistema 100 de comunicación de voz de la Figura 1, un micrófono 102 produce una señal 103 de voz analógica que es suministrada a un convertidor 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 señal que son codificados en forma binaria y son suministrados a un codificador 108 de canal. El codificador 108 de canal opcional añade redundancia a la representación binaria de los parámetros 107 de codificación de señal, antes de transmitirlos por el canal 101 de comunicación.

30 En el receptor, un decodificador 109 de canal utiliza dicha información redundante en el flujo 111 de bits recibido para detectar y corregir los errores de canal que se han producido durante la transmisión. A continuación, un decodificador 110 de voz convierte el flujo 112 de bits recibido desde el decodificador 109 de canal de nuevo a un conjunto de parámetros de codificación de señal y crea, a partir de los parámetros de codificación de señal recuperados, una señal 113 digital de voz sintetizada. La señal 113 digital de voz sintetizada reconstruida en el  
35 decodificador 110 de voz es convertida a una forma 114 analógica por un convertidor digital-a-analógico (D/A) 115 y es reproducida a través de una unidad 116 de altavoz.

La realización ilustrativa no restrictiva de un procedimiento eficiente de ocultación de borrado de tramas descrito en la presente memoria descriptiva puede ser usada con códecs basados en predicción lineal de banda estrecha o de banda ancha. Además, esta realización ilustrativa se describe con relación a un códec integrado basado en la  
40 Recomendación G.729 estandarizada por la Unión Internacional de Telecomunicaciones UIT (International Telecommunications Union, ITU) [Recomendación UIT-T G.6729 "Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP)", Ginebra, 1996).

El códec G.729 integrado basado en G.729 ha sido estandarizado por la ITU-T en 2006 y es conocido como la Recomendación G.729.1 [Recomendación UIT-T G.729.1 "G.729 based Embedded Variable bit-rate coder: An 8-32  
45 kbit/s scalable wideband coder bitstream interoperable with G.729", Ginebra, 2006]. Las técnicas descritas en la presente memoria han sido implementadas en la Recomendación UIT-T G.729.1.

Aquí, debería entenderse que la realización ilustrativa del procedimiento eficiente de ocultación de borrado de tramas podría ser aplicada a otros tipos de códecs. Por ejemplo, la realización ilustrativa del procedimiento eficiente de ocultación de borrado de tramas presentada en la presente memoria descriptiva es usada en un algoritmo  
50 candidato para la normalización por la UIT-T de un códec integrado de tasa de bits variable. En el algoritmo candidato, la capa de núcleo está basada en una técnica de codificación de banda ancha similar a AMR-WB (Recomendación UIT-T G.722.2).

En las secciones siguientes, se proporcionará, en primer lugar, una descripción general de CELP y el codificador y decodificador incorporado basado en G.729. A continuación, se describirá la realización ilustrativa del enfoque innovador para mejorar la robustez del códec.

**Descripción general del codificador ACELP**

5 La señal de voz muestreada es codificada, bloque a bloque, por el dispositivo 200 de codificación de la Figura 2, que se divide en once módulos con la numeración 201-211.

Por lo tanto, la señal 212 de voz de entrada es procesada, bloque a bloque, es decir, en los bloques de L muestras indicados anteriormente, denominados tramas.

10 Con referencia a la Figura 2, la señal 212 de voz de entrada muestreada es suministrada al módulo 201 de pre-procesamiento opcional. El módulo 201 de pre-procesamiento puede consistir en un filtro de paso alto con una frecuencia de corte de 200 Hz para las señales de banda estrecha y frecuencia de corte de 50 Hz para las señales de banda ancha.

La señal pre-procesada se denota mediante  $s(n)$ ,  $n = 0, 1, 2 \dots L-1$ , donde L es la longitud de la trama que es, típicamente, de 20 ms (160 muestras a una frecuencia de muestreo de 8 kHz).

15 La señal  $s(n)$  es usada para realizar un análisis LP en el módulo 204. El análisis LP es una técnica bien conocida por las personas con conocimientos ordinarios en la materia. En esta implementación ilustrativa, se usa el enfoque de autocorrelación. En el enfoque de autocorrelación, la señal  $s(n)$  es ventanizada en primer lugar usando, típicamente, una ventana Hamming que tiene una longitud del orden de 30-40 ms. Las autocorrelaciones se calculan a partir de la señal ventanizada, y se usa una recursión de Levinson-Durbin para calcular los coeficientes  $a_i$  del filtro LP, donde  $i = 1, \dots, p$ , y donde p es el orden de LP, que es típicamente 10 en codificación de banda estrecha y es 16 en codificación de banda ancha. Los parámetros  $a_i$  son los coeficientes de la función de transferencia  $A(z)$  del filtro LP, que viene determinada por la relación siguiente:

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

25 Por otra parte, se cree que el análisis LP es bien conocido para las personas con conocimientos ordinarios en la materia y, en consecuencia, no se describirá adicionalmente en la presente memoria descriptiva.

El módulo 204 realiza también la cuantificación y la interpolación de los coeficientes del filtro LP. En primer lugar, los coeficientes del filtro LP son transformados a otro dominio equivalente, más adecuado para los propósitos de cuantificación e interpolación. Los dominios de par espectral de línea (Line Spectral Pair, LSP) y de par espectral de inmitancia (Immitance Spectral Pair, ISP) son dos dominios en los cuales puede llevarse a cabo, de manera eficiente, una cuantificación y una interpolación. En la codificación de banda estrecha, los 10 coeficientes  $a_i$  del filtro LP pueden ser cuantificados en el orden de 18 a 30 bits usando cuantificación dividida o multi-etapa, o una combinación de las mismas. El propósito de la interpolación es permitir la actualización de los coeficientes del filtro LP en cada subtrama, mientras son transmitidos una vez para cada trama, lo que mejora el rendimiento del codificador sin aumentar la tasa de bits. Se cree que la cuantificación y la interpolación de los coeficientes de filtro LP son bien conocidas por las personas con conocimientos ordinarios en la materia y, en consecuencia, no se describirán más detalladamente en la presente memoria descriptiva.

40 Los párrafos siguientes describirán el resto de las operaciones de codificación realizadas subtrama a subtrama. En esta implementación ilustrativa, la trama de entrada de 20 ms es dividida en 4 subtramas de 5 ms (40 muestras a la frecuencia de muestreo de 8 kHz). En la descripción siguiente, el filtro  $A(z)$  denota el filtro LP interpolado, no cuantificado, de la subtrama, y el filtro  $\hat{A}(z)$  denota el filtro LP interpolado cuantificado de la subtrama. El filtro  $\hat{A}(z)$  es suministrado cada subtrama a un multiplexor 213 para su transmisión a través de un canal de comunicación (no mostrado).

45 En los codificadores de análisis por síntesis, la búsqueda de los parámetros óptimos de frecuencia fundamental y de innovación se realiza 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  $s_w(n)$  ponderada es calculada en un filtro 205 de ponderación perceptual en respuesta a la señal  $s(n)$ . Un ejemplo de función de transferencia para el filtro 205 de ponderación perceptual viene dado por la relación siguiente:

$$W(z) = A(z/\gamma_1)/A(z/\gamma_2) \quad \text{en la que} \quad 0 < \gamma_2 < \gamma_1 \leq 1$$

50

Con el propósito de simplificar el análisis de frecuencia fundamental, en primer lugar, se estima un retardo  $T_{OL}$  de frecuencia fundamental, de bucle abierto, en un módulo 206 de búsqueda de frecuencia fundamental, de bucle abierto, a partir de la señal  $s_w(n)$  de voz ponderada. A continuación, el análisis de frecuencia fundamental, de bucle cerrado, que se realiza en un módulo 207 de búsqueda de frecuencia fundamental, de bucle cerrado, subtrama a subtrama, es restringido alrededor del retardo  $T_{OL}$  de frecuencia fundamental, de bucle abierto, lo que reduce significativamente la complejidad de la búsqueda de los parámetros T (retardo de frecuencia fundamental) y b (ganancia de frecuencia fundamental) LTP (Long Term Prediction, Predicción a Largo Plazo). El análisis de frecuencia fundamental de bucle abierto es realizado, normalmente, en el módulo 206 una vez cada 10 ms (dos subtramas) usando técnicas bien conocidas por las personas con conocimientos ordinarios en la materia.

5

En primer lugar, se calcula el vector  $x$  objetivo para el análisis de LTP (Long Term Prediction, predicción a largo plazo). Normalmente, esto se realiza restando la respuesta  $s_0$  de entrada cero del filtro de síntesis  $W(z)/\hat{A}(z)$  ponderado a partir de la señal  $s_w(n)$  de voz ponderada. Esta respuesta  $s_0$  de entrada cero es calculada por un calculador 208 de respuesta a entrada cero en respuesta al filtro  $\hat{A}(z)$  interpolado cuantificado a partir del análisis LP, el módulo 204 de cuantificación y de interpolación y a los estados iniciales del filtro  $W(z)/\hat{A}(z)$  de síntesis ponderado almacenado en el módulo 211 de actualización de memoria en respuesta a los filtros  $A(z)$  y  $\hat{A}(z)$  LP, y el vector  $u$  de excitación. Esta operación es bien conocida por las personas con conocimientos ordinarios en la materia y, en consecuencia, no se describirá más detalladamente en la presente memoria descriptiva.

10

15

Un vector  $h$  de respuesta a impulso  $N$  dimensional del filtro  $W(z)/\hat{A}(z)$  de síntesis ponderado es calculado en el generador 209 de respuesta a impulso usando los coeficientes del filtro LP  $A(z)$  y  $\hat{A}(z)$  del módulo 204. Una vez más, esta operación es bien conocida por las personas con conocimientos ordinarios en la materia y, en consecuencia, no se describirá adicionalmente en la presente memoria descriptiva.

20

Los parámetros  $b$  y  $T$  de frecuencia fundamental, de bucle cerrado (o libro de códigos de frecuencia fundamental) se calculan en el módulo 207 de búsqueda de frecuencia fundamental, de bucle cerrado, que usa el vector  $x$  objetivo, el vector  $h$  de respuesta a impulso y el retardo  $T_{OL}$  de frecuencia fundamental, de bucle abierto, como entradas.

25

La búsqueda de frecuencia fundamental consiste en encontrar los mejores valores de retardo  $T$  de frecuencia fundamental y de ganancia  $b$  que minimizan un error de predicción de frecuencia fundamental ponderado cuadrático medio, por ejemplo,

$$e = \|x - b y\|^2$$

30

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

Más específicamente, en la presente implementación ilustrativa, la búsqueda de frecuencia fundamental (libro de códigos de frecuencia fundamental o libro de códigos adaptativo) está compuesta por tres (3) etapas.

En la primera etapa, se estima un retardo  $T_{OL}$  de frecuencia fundamental, de bucle abierto, en el módulo 206 de búsqueda de frecuencia fundamental, de bucle abierto, en respuesta a la señal  $s_w(n)$  de voz ponderada. Tal como se indica en la descripción anterior, este análisis de frecuencia fundamental de bucle abierto se realiza, normalmente, una vez cada 10 ms (dos subtramas) usando técnicas bien conocidas por las personas con conocimientos ordinarios en la materia.

35

En la segunda etapa, se realiza una búsqueda de un criterio  $C$  de búsqueda en el módulo 207 de búsqueda de frecuencia fundamental, de bucle cerrado, para los retardos enteros de frecuencia fundamental alrededor del retardo  $T_{OL}$  de frecuencia fundamental, de bucle abierto, estimado (normalmente  $\pm 5$ ), lo que simplifica considerablemente el procedimiento de búsqueda. Un ejemplo del criterio de búsqueda  $C$  viene dado por:

40

$$C = \frac{x^t y_T}{\sqrt{y_T^t y_T}}$$

45

donde  $t$  denota la traspuesta del vector.

Una vez encontrado un retardo de frecuencia fundamental entero óptimo en la segunda etapa, una tercera etapa de la búsqueda comprueba (módulo 207), por medio del criterio  $C$  de búsqueda, las fracciones alrededor de ese retardo de frecuencia fundamental entero óptimo. Por ejemplo, la Recomendación UIT-T G.729 usa una resolución de 1/3 de sub-muestra.

50

El índice  $T$  de libro de códigos de frecuencia fundamental es codificado y transmitido al multiplexor 213 para su transmisión a través de un canal de comunicación (no mostrado). La ganancia  $b$  de frecuencia fundamental es cuantificada y transmitida al multiplexor 213.

5 Una vez determinados la frecuencia fundamental, o los parámetros  $b$  y  $T$  de LTP (Predicción a Largo Plazo), la etapa siguiente es la búsqueda de la excitación innovadora óptima por medio del módulo 210 de búsqueda de excitación innovadora de la Figura 2. En primer lugar, el vector  $x$  objetivo es actualizado restando la contribución LTP:

$$x' = x - by_T$$

10 donde  $b$  es la ganancia de frecuencia fundamental e  $y_T$  es el vector de libro de códigos de frecuencia fundamental filtrado (la excitación pasada en el retardo  $T$  convolucionada con la respuesta a impulso  $h$ ).

El procedimiento de búsqueda de excitación innovadora en CELP es realizado en un libro de códigos innovador para encontrar el vector de código  $c_k$  de excitación óptimo y la ganancia  $g$  que minimizan el error  $E$  cuadrático medio entre el vector  $x'$  objetivo y una versión filtrada escalada del vector de códigos  $c_k$ , por ejemplo:

15

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

donde  $H$  es una matriz de convolución triangular inferior derivada del vector  $h$  de respuesta a impulso. El índice  $k$  del libro de códigos de innovación correspondiente al vector de códigos  $c_k$  óptimo encontrado y la ganancia  $g$  son suministrados al multiplexor 213 para su transmisión a través de un canal de comunicación.

20 En una implementación ilustrativa, el libro de códigos de innovación usado es un libro de código dinámico que comprende un libro de códigos algebraico seguido de un pre-filtro adaptativo  $F(z)$  que mejora los componentes espectrales especiales con el fin de mejorar la calidad de la síntesis de voz, según la patente US N° 5.444.816 concedida a Adoul et al., el 22 de Agosto de 1995. En esta implementación ilustrativa, la búsqueda del libro de códigos innovador es realizada en el módulo 210 por medio de un libro de códigos algebraico tal como se describe en las patentes US N°: 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, 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.

### Descripción general del decodificador ACELP

30 El decodificador 300 de voz de la Figura 3 ilustra las diversas etapas llevadas a cabo entre la entrada 322 digital (flujo de bits de entrada al demultiplexor 317) y la señal  $s_{salida}$  de voz de salida muestreada.

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

- 35
- los coeficientes  $\hat{A}(z)$  LP interpolados, cuantificados, denominados también parámetros de predicción a corto plazo (STP) producidos una vez por trama;
  - los parámetros de predicción a largo plazo (LTP)  $T$  y  $b$  (para cada subtrama); y
  - el índice  $k$  de libro de códigos de innovación y la ganancia  $g$  (para cada subtrama).

La señal de voz actual es sintetizada en base a estos parámetros, tal como se explicará más adelante.

40 El libro 318 de códigos de innovación es sensible al índice  $k$  para producir el vector  $c_k$  de código de innovación, que es escalado por la ganancia  $g$  decodificada mediante un amplificador 324. En la implementación ilustrativa, un libro de códigos de innovación, tal como se describe en las patentes US 5.444.816; 5.699.482; 5.754.976 y 5.701, 392, indicadas anteriormente, es usado para producir el vector  $c_k$  de código innovador.

45 El vector  $by_T$  de código es producido aplicando el retardo  $T$  de frecuencia fundamental a un libro 301 de códigos de frecuencia fundamental para producir un vector de código de frecuencia fundamental. A continuación, el vector  $v_T$  de código de frecuencia fundamental es amplificado por la ganancia  $b$  de frecuencia fundamental por un amplificador 326 para producir el vector  $by_T$  de código de frecuencia fundamental escalado.

La señal  $u$  de excitación es calculada por el sumador 320 como:

$$u = gc_k + bv_T$$

5 El contenido del libro 301 de códigos de frecuencia fundamental es actualizado usando el valor pasado de la señal  $u$  de excitación almacenado en la memoria 303 para mantener el sincronismo entre el codificador 200 y el decodificador 300.

La señal  $s'$  sintetizada es calculada filtrando la señal  $u$  de excitación mediante el filtro 306 de síntesis LP, que tiene la forma  $1/\hat{A}(z)$ , donde  $\hat{A}(z)$  es el filtro LP interpolado cuantificado de la subtrama actual. Como puede verse en la Figura 3, los coeficientes  $A(z)$  LP interpolados cuantificados en la línea 325 desde el demultiplexor 317 se suministran al filtro 306 de síntesis LP para ajustar los parámetros del filtro 306 de síntesis LP en consecuencia.

10 El vector  $s'$  es filtrado a través del post-procesador 307 para obtener la señal  $s_{salida}$  de voz de salida muestreada. Típicamente, el post-procesamiento consiste en un potsfiltrado de corto plazo, postfiltrado de largo plazo y escalado de la ganancia. También puede consistir en un filtro de paso alto para eliminar las bajas frecuencias no deseadas. Por otra parte, el postfiltrado es bien conocido por las personas con conocimientos ordinarios en la materia.

**Descripción general de la codificación integrada basada en G.729**

15 El códec G.729 se basa en el paradigma de codificación CELP algebraico (ACELP) explicado anteriormente. La asignación de bits del códec G.729 a 8 kbit/s se proporciona en la Tabla 1.

Tabla 1. Asignación de bits en G.729 a 8 kbit/s

Parámetro	Trama bits / 10 ms
Parámetros LP	18
Retardo de frecuencia fundamental	13 = 8 + 5
Paridad de frecuencia fundamental	1
Ganancias	14 = 7 + 7
Libro de códigos algebraico	34 = 17 + 17
Total	80 bits/10 ms = 8 kbits/s

20 La recomendación UIT-T G.729 opera en tramas de 10 ms (80 muestras a una frecuencia de muestreo de 8 kHz). Los parámetros LP se cuantifican y se transmiten una vez por trama. La trama G.729 es dividida en dos subtramas de 5 ms. El retardo de frecuencia fundamental (o índice de libro de códigos adaptativo) es cuantificado con 8 bits en la primera subtrama y 5 bits en la segunda sub-trama (en relación al retardo de la primera subtrama). Las ganancias de frecuencia fundamental y libro de códigos algebraicos se cuantifican conjuntamente usando 7 bits por subtrama. Se usa un libro de códigos algebraico de 17 bits para representar la excitación de libro de códigos fijo o de innovación.

25 El códec integrado está construido en base al códec de núcleo G.729. La codificación integrada, o codificación por capas, consiste en una capa de núcleo y capas adicionales para una mayor calidad o un mayor ancho de banda codificada. El flujo de bits correspondiente a las capas superiores puede ser descartado por la red según sea necesario (en caso de congestión o en situación de multidifusión, donde algunos enlaces tienen una menor tasa de bits disponible). El decodificador puede reconstruir la señal en base a las capas que recibe.

30 En esta realización ilustrativa, la capa L1 de núcleo consiste en G.729 a 8 kbit/s. La segunda capa L2 proporciona 4 kbit/s adicionales para mejorar la calidad de banda estrecha (a la tasa de bits  $R2 = L1 + L2 = 12$  kbit/s). Cada una de las 10 capas superiores de 2 kbit/s es usada para obtener una señal codificada de banda ancha. Las 10 capas L3 a L12 corresponden a tasas de bits de 14, 16, ... y 32 kbit/s, respectivamente. De esta manera, el codificador integrado funciona como un codificador de banda ancha para tasas de bits de 14 kbit/s y superiores.

35 Por ejemplo, el codificador usa codificación predictiva (CELP) en las dos primeras capas (G.729 modificado mediante la adición de un segundo libro de códigos algebraico) y, a continuación, cuantifica en el dominio de la frecuencia el error de codificación de las primeras capas. Se usa una MDCT (Modified Discrete Cosine Transform, Transformada Discreta de Coseno Modificada) para asignar la señal al dominio de la frecuencia. Los coeficientes MDCT se cuantifican usando cuantificación vectorial algebraica escalable. Para aumentar el ancho de banda de



audio, se aplica una codificación paramétrica a las altas frecuencias.

El codificador funciona sobre tramas de 20 ms, y necesita 5 ms de anticipación para la ventana de análisis LP. MDCT con una superposición del 50% requiere 20 ms adicionales de anticipación que podrían ser aplicados tanto en el codificador como en el decodificador. Por ejemplo, la anticipación en MDCT es usada en el decodificador, lo que resulta en una mejor ocultación de borrado de tramas, tal como se explicará a continuación. El codificador produce una salida en 32 kbps, lo que se traduce en tramas de 20 ms que contienen 640 bits cada una. Los bits en cada trama están dispuestos en capas integradas. La Capa 1 tiene 160 bits que representan 20 ms de la norma G.729 a 8 kbps (correspondiente a dos tramas G.729). La Capa 2 tiene 80 bits, lo que representa 4 kbps adicionales. A continuación, cada capa adicional (Capas 3 a 12) añade 2 kbps, hasta 32 kbps.

En la Figura 4 se muestra un diagrama de bloques de un ejemplo de codificador integrado.

La señal de banda ancha original  $x$  (401), muestreada a 16 kHz, es dividida primero en dos bandas: 0-4.000 Hz y 4.000-8.000 Hz en el módulo 402. En el ejemplo de la Figura 4, la división de banda se realiza usando un banco de filtros QMF (Quadrature Mirror Filter, filtro espejo en cuadratura) con 64 coeficientes. Esta operación es bien conocida por las personas con conocimientos ordinarios en la materia. Después de la división de banda, se obtienen dos señales, una que cubre la banda de 0-4.000 Hz (banda baja) y la otra que cubre la banda de 4.000-8.000 (banda alta). La frecuencia de muestreo de las señales en cada una de estas dos bandas se reduce en un factor de 2 en el módulo 402. Esto proporciona 2 señales a una frecuencia de muestreo de 8 kHz:  $X_{LF}$  para la banda baja (403), y  $X_{HF}$  para la banda alta (404).

La señal  $X_{LF}$  de banda baja es suministrada a una versión modificada del codificador 405 de G.729. Esta versión 405 modificada produce primero el flujo de bits estándar G.729 a 8 kbps, que constituye los bits para la Capa 1. Obsérvese que el codificador opera sobre tramas de 20 ms, por lo tanto, los bits de la Capa 1 corresponden a dos tramas G.729.

A continuación, el codificador 405 de G.729 es modificado para incluir un segundo libro de códigos algebraico innovador para mejorar la señal de banda baja. Este segundo libro de códigos es idéntico al libro de códigos innovador en G.729, y requiere 17 bits por subtrama de 5 ms para codificar los pulsos del libro de códigos (68 bits por trama de 20 ms). Las ganancias del segundo libro de códigos algebraico se cuantifican con respecto a la ganancia del primer libro de códigos usando 3 bits en la primera y la tercera subtramas y 2 bits en la segunda y la cuarta subtramas (10 bits por trama). Se usan dos bits para enviar información de clasificación para mejorar la ocultación en el decodificador. Esto produce  $68 + 10 + 2 = 80$  bits para la Capa 2. La señal objetivo usada para este libro de códigos innovador de la segunda etapa se obtiene restando la contribución del libro de códigos G.729 innovador en el dominio de voz ponderada.

La señal  $\hat{x}_{LF}$  de síntesis del codificador 405 de G.729 modificado es obtenida mediante la adición de la excitación de la norma G.729 (adición de vectores de código innovador y adaptativo escalados) y la excitación innovadora del libro de códigos innovador adicional, y pasando esta excitación mejorada a través del filtro de síntesis G.729 habitual. Esta es la señal de síntesis que el decodificador producirá si recibe sólo la Capa 1 y la Capa 2 desde el flujo de bits. Obsérvese que el contenido del libro de códigos (o frecuencia fundamental) adaptativo es actualizado sólo usando la excitación G.729.

La Capa 3 extiende el ancho de banda desde la calidad de la banda estrecha a la calidad de la banda ancha. Esto se realiza aplicando codificación paramétrica (módulo 407) a la componente  $x_{HF}$  de alta frecuencia. Sólo se calculan y transmiten la envolvente espectral y la envolvente de dominio en el tiempo de  $x_{HF}$  para esta capa. La extensión de ancho de banda requiere 33 bits. Los 7 bits restantes en esta capa se usan para transmitir información de fase (posición de pulso glótico) para mejorar la ocultación de borrado de tramas en el decodificador según la presente invención. Esto se explicará más detalladamente en la descripción siguiente.

A continuación, a partir de la Figura 4, el error de codificación del sumador 406 ( $x_{LF} - \hat{x}_{LF}$ ) junto con la señal  $x_{HF}$  de alta frecuencia se asignan ambos en el dominio de la frecuencia en el módulo 408. La MDCT, con un solapamiento del 50%, se usa para esta asignación de tiempo-frecuencia. Esto puede realizarse mediante el uso de dos MDCTs, una para cada banda. La señal de banda alta puede ser multiplicada espectralmente, en primer lugar, antes de la MDCT por el operador  $(-1)^n$ , de manera que los coeficientes MDCT de ambas transformaciones pueden ser unidos en un vector para propósitos de cuantificación. A continuación, los coeficientes MDCT son cuantificados en el módulo 409 usando cuantificación vectorial algebraica escalable en una manera similar a la cuantificación de los coeficientes de FFT (Fast Fourier Transform, Transformada Rápida de Fourier) en el codificador de audio 3GPP AMR-WB+ (3GPP TS 26.290). Por supuesto, pueden aplicarse otras formas de cuantificación. La tasa de bits total para esta cuantificación espectral es de 18 kbps, lo que equivale a una capacidad de bits de 360 bits por trama de 20 ms. Después de la cuantificación, los bits correspondientes se distribuyen en capas en etapas de 2 kbps en el

módulo 410 para formar las capas 4 a 12. De esta manera, cada capa de 2 kbps contiene 40 bits por cada trama de 20 ms. En una realización ilustrativa, pueden reservarse 5 bits en la Capa 4 para transmitir información de energía para mejorar la ocultación y la convergencia del decodificador en el caso de borrado de tramas.

5 Las extensiones algorítmicas, en comparación con el codificador de núcleo G.729, puede ser resumidas de la siguiente manera: 1) el libro de códigos innovador de G.729 es repetido una segunda vez (Capa 2); 2) la codificación paramétrica es aplicada para extender el ancho de banda, donde sólo se calculan y cuantifican (Capa 3) la envolvente espectral y la envolvente en el dominio del tiempo (la información de ganancia), 3) se calcula una MDCT cada 20 ms, y sus coeficientes espectrales se cuantifican en bloques de 8 dimensiones usando VQ (Vector Quantization, Cuantificación vectorial) algebraica escalable, y 4) se aplica una rutina de capas de bits para dar  
10 formato al flujo de 18 kbps a partir de la VQ algebraica en capas de 2 kbps cada una (capas 4 a 12). En una realización, pueden transmitirse 14 bits de ocultación de información y de convergencia en la Capa 2 (2 bits), la Capa 3 (7 bits) y la Capa 4 (5 bits).

15 La Figura 5 es un diagrama de bloques de un ejemplo de decodificador 500 integrado. En cada trama de 20 ms, el decodificador 500 puede recibir cualquiera de las tasas de bits soportadas, de 8 kbps a 32 kbps. Esto significa que la operación del decodificador depende del número de bits, o capas, recibidas en cada trama. En la Figura 5, se supone que al menos las Capas 1, 2, 3 y 4 han sido recibidas en el decodificador. A continuación, se describirán los casos de las tasas de bits más bajas.

20 En el decodificador de la Figura 5, el flujo 501 de bits recibido es separado, en primer lugar, en Capas de bits según es producido por el codificador (módulo 502). Las Capas 1 y 2 forman la entrada al decodificador 503 G.729 modificado, que produce una señal  $\hat{x}_{LF}$  de síntesis para la banda inferior (0-4.000 Hz, muestreada a 8 kHz). Recuérdese que la Capa 2 contiene esencialmente los bits para un segundo libro de códigos innovador con la misma estructura que el libro de códigos G.729 innovador.

25 A continuación, los bits de la Capa 3 forman la entrada al decodificador 506 paramétrico. Los bits de la Capa 3 proporcionan una descripción paramétrica de la banda alta (4.000-8.000 Hz, muestreada a 8 kHz). Específicamente, los bits de la Capa 3 describen la envolvente espectral de banda alta de la trama de 20 ms, junto con la envolvente en el dominio del tiempo (o información de ganancia). El resultado de la decodificación paramétrica es una aproximación paramétrica de la señal de banda alta, llamada  $\bar{x}_{HF}$  en la Figura 5.

30 A continuación, los bits de la Capa 4 y superiores forman la entrada del cuantificador 504 inverso ( $Q^{-1}$ ). La salida del cuantificador 504 inverso es un conjunto de coeficientes espectrales cuantificados. Estos coeficientes cuantificados forman la entrada del módulo 505 de transformada inversa ( $T^{-1}$ ), específicamente una MDCT inversa con una superposición del 50%. La salida de la MDCT inversa es la señal  $\hat{x}_D$ . Esta señal  $\hat{x}_D$  puede considerarse como el error de codificación cuantificado del codificador G.729 modificado en la banda baja, junto con la alta banda cuantificada si cualquier bit es asignado a la banda alta en la trama determinada. El módulo 505 de transformada inversa ( $T^{-1}$ ) es implementado como dos MDCTs inversas, entonces  $\hat{x}_D$  consistirá en dos componentes,  $\hat{x}_{D1}$ , que  
35 representa el componente de baja frecuencia, y  $\hat{x}_{D2}$ , que representa el componente de alta frecuencia.

40 La componente  $\hat{x}_{D1}$  que forma el error de codificación cuantificado del codificador G.729 modificado es combinada, a continuación, con  $\hat{x}_{LF}$  en el combinador 507 para formar la síntesis  $\hat{s}_{LF}$  de banda baja. De la misma manera, la componente  $\hat{x}_{D2}$  que forma la banda alta cuantificada es combinada con la aproximación paramétrica de  $\bar{x}_{HF}$  de banda alta en el combinador 508 para formar la síntesis  $\hat{s}_{HF}$  de banda alta. Las señales  $\hat{s}_{LF}$  y  $\hat{s}_{HF}$  son procesadas mediante el banco 509 de filtros QMF de síntesis para formar la señal  $\hat{s}$  de síntesis total a una frecuencia de muestreo de 16 kHz.

45 En el caso en el que no se reciben las Capas 4 y superiores, entonces,  $\hat{x}_D$  es cero, y las salidas de los combinadores 507 y 508 son iguales a su entrada, concretamente,  $\hat{x}_{LF}$  y  $\bar{x}_{HF}$ . Si se reciben sólo las Capas 1 y 2, entonces el decodificador sólo tiene que aplicar el decodificador G.729 modificado para producir la señal  $\hat{x}_{LF}$ . La componente de banda alta será cero, y la señal muestreada a frecuencias superiores a 16 kHz (si es necesario) tendrá contenido sólo en la banda baja. Si sólo se recibe la Capa 1, entonces el decodificador sólo tiene que aplicar el decodificador G.729 para producir la señal  $\hat{x}_{LF}$ .

### Ocultación robusta de borrado de tramas

El borrado de tramas tiene un efecto importante sobre la calidad de voz sintetizada en sistemas de comunicaciones de voz digitales, especialmente cuando funcionan en entornos inalámbricos y redes de conmutación de paquetes. En los sistemas celulares inalámbricos, la energía de la señal recibida puede exhibir frecuentes desvanecimientos graves que resultan en altas tasas de error de bit y esto se hace más evidente en los límites de las celdas. En este caso, el decodificador de canal no puede corregir los errores en la trama recibida y, como consecuencia, el detector de errores usado normalmente después del decodificador de canal declarará la trama como borrada. En aplicaciones de voz sobre redes de paquetes, tales como Voz sobre Protocolo de Internet (Voice over Internet Protocol, VoIP), la señal de voz es empaquetada, colocando normalmente una trama de 20 ms en cada paquete. En las comunicaciones de conmutación de paquetes, puede producirse una eliminación de paquetes en un enrutador si el número de paquetes se hace muy grande, o el paquete puede llegar al receptor después de un largo retardo y debería ser declarado como perdido si su retardo es mayor que la longitud de una memoria intermedia anti-fluctuaciones en el lado del receptor. En estos sistemas, el códec podría ser sometido a típicamente a tasas del 3 al 5% de borrado de tramas.

Básicamente, el problema del procesamiento de borrado de tramas (Frame Erasure, FER) es doble. En primer lugar, cuando llega un indicador de trama borrada, la trama ausente debe ser generada usando la información enviada en la trama anterior y estimando la evolución de la señal en la trama ausente. El éxito de la estimación no sólo depende de la estrategia de ocultación, sino también en el lugar en la señal de voz donde sucede el borrado. En segundo lugar, debe asegurarse 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 ya que la síntesis real y la síntesis estimada pueden evolucionar de manera diferente. Cuando llega la primera trama buena, el decodificador está desincronizado con respecto al codificador. La razón principal es que los codificadores de baja tasa de bits dependen de la predicción de frecuencia fundamental y, durante las tramas borradas, la memoria del predictor de frecuencia fundamental (o el libro de códigos adaptativo) ya no es la misma que la del codificador. El problema se amplifica cuando se borran muchas tramas consecutivas. Con respecto a la ocultación, la dificultad de la recuperación del procesamiento normal depende del tipo de señal, por ejemplo señal de voz, donde se produjo el borrado.

El efecto negativo de los borrados de tramas puede reducirse significativamente adaptando la ocultación y la recuperación del procesamiento normal (recuperación adicional) al tipo de la señal de voz donde se produce el borrado. Para este propósito, es necesario clasificar cada trama de voz. Esta clasificación puede realizarse en el codificador y puede ser transmitida. De manera alternativa, puede ser estimada en el decodificador.

Para la mejor ocultación y recuperación, hay pocas características críticas de la señal de voz que deben ser controladas cuidadosamente. Estas características críticas son la energía o la amplitud de la señal, la cantidad de periodicidad, la envolvente espectral y el periodo fundamental. En caso de una recuperación de voz sonora, puede conseguirse una mejora adicional mediante un control de fase. Con un ligero aumento en la tasa de bits, unos pocos parámetros suplementarios pueden ser cuantificados y transmitidos para un mejor control. Si no hay ancho de banda adicional disponible, los parámetros pueden ser estimados en el decodificador. Con estos parámetros controlados, la ocultación y la recuperación del borrado de tramas pueden mejorarse de manera significativa, especialmente mejorando la convergencia de la señal decodificada a la señal real en el codificador y aliviando el efecto de la discrepancia entre el codificador y el decodificador cuando se recupera el procesamiento normal.

Estas ideas han sido descritas en la solicitud de patente PCT en la referencia [1]. Según la realización ilustrativa no limitativa de la presente invención, la ocultación y la convergencia se mejoran adicionalmente mediante una mejor sincronización del pulso glótico en el libro de códigos de frecuencia fundamental (o libro de códigos adaptativo) tal como se describe más adelante, en la presente memoria. Esto puede realizarse con o sin la información de fase recibida, que corresponde, por ejemplo, a la posición del pulso de frecuencia fundamental o pulso glótico.

En la realización ilustrativa de la presente invención, se describen procedimientos para una ocultación eficiente de borrado de tramas, y procedimientos para mejorar la convergencia en el decodificador en las tramas siguientes a una trama borrada.

Las técnicas de ocultación de borrado de tramas según la realización ilustrativa han sido aplicadas al códec basado en G.729 integrado descrito anteriormente. Este códec servirá como una estructura ejemplar para la aplicación de los procedimientos de ocultación FER en la descripción siguiente.

La Figura 6 proporciona un diagrama de bloques simplificado de las Capas 1 y 2 de un codificador 600 integrado, basado en el modelo de codificador CELP de la Figura 2. En este diagrama de bloques simplificado, el módulo 207 de búsqueda de frecuencia fundamental, de bucle cerrado, el calculador 208 de respuesta de entrada cero, el calculador 209 de respuesta a impulso, el módulo 210 de búsqueda de excitación innovadora y el módulo 211 de

actualización de memoria se agrupan en los módulos 602 de búsqueda de libro de códigos de innovación y búsqueda de frecuencia fundamental de bucle cerrado. Además, la segunda etapa de la búsqueda de libro de códigos en la Capa 2 está incluida también en los módulos 602. Esta agrupación se realiza para simplificar la introducción de los módulos relacionados con la realización ilustrativa de la presente invención.

5 La Figura 7 es una extensión del diagrama de bloques de la Figura 6, donde se han añadido los módulos relacionados con la realización ilustrativa no limitativa de la presente invención. En estos módulos 702 a 707 añadidos, se calculan, cuantifican y transmiten parámetros adicionales con el objetivo de mejorar la ocultación FER y la convergencia y la recuperación del decodificador después de las tramas borradas. En esta realización ilustrativa, estos parámetros de ocultación/recuperación incluyen información de clasificación, energía y fase de la  
10 señal (por ejemplo, la posición estimada del último pulso glótico en la trama o tramas anteriores).

En la descripción siguiente, el cálculo y la cuantificación de estos parámetros de ocultación/recuperación adicionales se proporcionarán más detalladamente y serán más evidentes con referencia a la Figura 7. Entre estos parámetros, la clasificación de la señal será tratada más detalladamente. En las secciones siguientes, se explicará la ocultación FER eficiente usando estos parámetros de ocultación/recuperación adicionales para mejorar la convergencia.  
15

### **Clasificación de señal para ocultación y recuperación FER**

La idea básica detrás del uso de una clasificación de la voz para una reconstrucción de la 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 cambian rápidamente. Aunque el mejor procesamiento de tramas borradas en los segmentos de voz no estacionarios puede resumirse como una rápida convergencia de los parámetros de codificación de voz a las características de ruido ambiental, en el caso de señal cuasi-estacionaria, los parámetros de codificación de voz no varían dramáticamente y pueden mantenerse prácticamente inalterados durante varias tramas borradas contiguas antes de ser amortiguados. Además, el procedimiento óptimo para una recuperación de señal después de un bloque borrado de tramas varía con la clasificación de la señal de voz.  
20  
25

La señal de voz puede ser clasificada, en términos generales, como sonora, sorda y pausas.

La voz sonora contiene una cantidad de componentes periódicos y puede ser dividida adicionalmente en las categorías siguientes: inicios sonoros, segmentos sonoros, transiciones sonoras y compensaciones sonoras. 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 fundamental, relación de los componentes periódicos y no periódicos, energía) varían lentamente de trama a trama. Una transición de voz se caracteriza por variaciones rápidas de una voz sonora, tal como una transición entre vocales. Las compensaciones sonoras se caracterizan por una disminución gradual de la energía y la sonoridad al final de los segmentos sonoros.  
30

35 Las partes sordas de la señal se caracterizan por la ausencia del componente periódico y pueden dividirse 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 de sólo ruido, si hay presente un ruido de fondo.

40 No todas las clases indicadas anteriormente necesitan un procesamiento por separado. Por lo tanto, para los propósitos de las técnicas de ocultación de errores, algunas de las clases de señal se agrupan entre sí.

### **Clasificación en el codificador**

Cuando hay un ancho de banda disponible en el flujo de bits para incluir la información de clasificación, la clasificación puede ser realizada en el codificador. Esto tiene diversas ventajas. Una ventaja es que frecuentemente hay una anticipación en los codificadores de voz. La anticipación permite estimar la evolución de la señal en la trama siguiente y, en consecuencia, la clasificación puede ser realizada teniendo en cuenta el comportamiento futuro de la señal. En general, cuanto mayor es la anticipación, mejor puede ser la clasificación. Una ventaja adicional es una reducción de la complejidad, ya que la mayoría del procesamiento necesario para la ocultación del borrado de tramas es necesario de todos modos para la codificación de voz. Por último, también existe la ventaja de trabajar con la señal original en lugar de la señal sintetizada.  
45  
50

La clasificación de tramas se realiza teniendo en cuenta la estrategia de ocultación y recuperación. En otras palabras, cualquier trama es clasificada de manera que la ocultación puede ser óptima si se pierde la siguiente

trama, o de manera que la recuperación puede ser óptima si se pierde la trama anterior. Algunas de las clases usadas para el procesamiento FER no necesitan ser transmitidas, ya que pueden ser deducidas sin ambigüedad en el decodificador. En la presente realización ilustrativa, se usan cinco (5) clases distintas, y se definen como se indica a continuación:

- 5           • La clase SORDA comprende todas las tramas de voz sordas y todas las tramas sin voz activa. Una trama de compensación sonora puede ser clasificada también como SORDA si su final tiende a ser SORDO y la ocultación diseñada para las tramas sordas puede ser usada para la trama siguiente en caso de que se pierda.
- 10          • La clase TRANSICIÓN SORDA comprende tramas sordas con un posible inicio sonoro al final. Sin embargo, el inicio es todavía demasiado corto o no está construido suficientemente bien para usar la ocultación diseñada para las tramas sonoras. La clase TRANSICIÓN SORDA puede seguir sólo a una trama clasificada como SORDA o TRANSICIÓN SORDA.
- 15          • La clase TRANSICIÓN SONORA comprende tramas sonoras con características sonoras relativamente débiles. Típicamente, esas son tramas sonoras con características que cambian rápidamente (transiciones entre vocales) o compensaciones sonoras que duran toda la trama. La clase TRANSICIÓN SONORA puede seguir sólo a una trama clasificada como TRANSICIÓN SONORA, SONORA O INICIO.
- 20          • La clase SONORA comprende tramas sonoras con características estables. Esta clase puede seguir sólo a una trama clasificada como TRANSICIÓN SONORA, SONORA O INICIO.
- 25          • La clase INICIO comprende todas las tramas sonoras con características estables después de una trama clasificada como SORDA o TRANSICIÓN SORDA. Las tramas clasificadas como INICIO corresponden a tramas de inicio sonoro en las que el inicio ya está suficientemente bien construido para el uso de la ocultación diseñada para las tramas sonoras perdidas. Las técnicas de ocultación usadas para un borrado de trama después de la clase INICIO son las mismas que después de la clase SONORA. La diferencia está en la estrategia de recuperación. Si se pierde una trama de clase INICIO (es decir, una trama SONORA buena llega después de un borrado, pero la última trama antes del borrado era SORDA), puede usarse una técnica especial para reconstruir artificialmente el inicio perdido. Este escenario puede observarse en la Figura 6. Las técnicas artificiales de reconstrucción de inicio se describirán más detalladamente en la descripción siguiente. Por otra parte, si llega una trama de INICIO buena después de un borrado y la última trama antes del borrado era SORDA, este tratamiento especial no es necesario, ya que el inicio no se ha perdido (no estaba en la trama perdida).

El diagrama de estados de clasificación se describe en la Figura 8. Si el ancho de banda disponible es suficiente, la clasificación se realiza en el codificador y es transmitida usando 2 bits. Tal como puede observarse en la Figura 8, TRANSICIÓN SORDA 804 y TRANSICIÓN SONORA 806 pueden agruparse, ya que pueden diferenciarse sin ambigüedad en el decodificador (las tramas TRANSICIÓN SORDA 804 pueden seguir sólo a las tramas SORDAS 802 o las tramas TRANSICIÓN SORDA 804, las tramas TRANSICIÓN SONORA 806 pueden seguir sólo a las tramas INICIO 810, SONORA 808 o TRANSICIÓN SONORA 806). En esta realización ilustrativa, la clasificación es realizada en el codificador y es cuantificada usando 2 bits que se transmiten en la Capa 2. De esta manera, si se recibe al menos la Capa 2, entonces la información de clasificación del decodificador es usada para una ocultación mejorada. Si se recibe sólo la Capa 1 de núcleo, entonces la clasificación es realizada en el decodificador.

40 Los parámetros siguientes se usan para la clasificación en el codificador: una correlación  $r_x$  normalizada, una medida  $e_t$  de inclinación espectral y, una relación  $snr$  señal-a-ruido, un contador  $pc$  de estabilidad de frecuencia fundamental, una energía de trama relativa de la señal al final de la trama  $E_s$  actual y un contador  $zc$  de cruce por cero.

El cálculo de estos parámetros, que se usan para clasificar la señal, se explica a continuación.

45 La correlación  $r_x$  normalizada se calcula como parte del módulo 206 de búsqueda de frecuencia fundamental, de bucle abierto, de la Figura 7. Este módulo 206 normalmente proporciona la estimación de frecuencia fundamental de bucle abierto cada 10 ms (dos veces por trama). Aquí, se usa también para proporcionar las medidas de correlación normalizada. Estas correlaciones normalizadas se calculan sobre la señal  $s_w(n)$  de voz ponderada actual y la señal de voz ponderada pasada con el retardo de frecuencia fundamental de bucle abierto. La correlación  $\bar{r}_x$  promedio se define como:

$$\bar{r}_x = 0,5(r_x(0) + r_x(1)) \quad (1)$$

en la que  $r_x(0)$ ,  $r_x(1)$  son, respectivamente, la correlación normalizada de la primera mitad de la trama y la segunda mitad de la trama. La correlación  $r_x(k)$  normalizada se calcula como sigue:

$$r_x(k) = \frac{\sum_{i=0}^{L-1} x(t_k + i)x(t_k + i - T_k)}{\sqrt{\sum_{i=0}^{L-1} x^2(t_k + i)\sum_{i=0}^{T-1} x^2(t_k + i - T_k)}} \quad (2)$$

Las correlaciones  $r_x(k)$  se calculan usando la señal  $s_w(n)$  de voz ponderada (como "x"). Los instantes  $t_k$  están relacionados con el principio de la semi-trama actual y son iguales a 0 y 80 muestras respectivamente. El valor  $T_k$  es el retardo de frecuencia fundamental en la semi-trama que maximiza la correlación cruzada

$$\sum_{i=0}^{L-1} x(t_k + i)x(t_k + i - T).$$

La longitud  $L'$  del cálculo de la autocorrelación es igual a 80 muestras. En otra realización para determinar el valor  $T_k$  en una semi-trama, se calcula la correlación cruzada

$$\sum_{i=0}^{L-1} x(\tau + i)x(\tau + i - T)$$

y se encuentran los valores de  $\tau$  correspondiente al máximo en las tres secciones de retardo 20-39, 40-79, 80-143. A continuación,  $T_k$  se establece al valor  $\tau$  que maximiza la correlación normalizada en la ecuación (2).

El parámetro  $e_i$  de inclinación espectral contiene la información acerca de la distribución de frecuencias de la energía. En la presente realización ilustrativa, la inclinación espectral es estimada en el módulo 703 como los primeros coeficientes de autocorrelación normalizados de la señal de voz (el primer coeficiente de reflexión obtenido durante el análisis LP).

Debido a que el análisis LP se realiza dos veces por trama (una vez cada trama G.729 de 10 ms), la inclinación espectral es calculada como el promedio del primer coeficiente de reflexión a partir de ambos análisis LP. Es decir

$$e_i = -0,5(k_1^{(1)} + k_1^{(2)}) \quad (3)$$

en la que  $k_1^{(j)}$  es el primer coeficiente de reflexión a partir del análisis LP en la semi-trama j.

La medición de la relación  $snr$  de señal-a-ruido (SNR) hace uso del hecho de que para un codificador de ajuste de forma de onda general, la SNR es mucho mayor para los sonidos sonoros. La estimación del parámetro  $snr$  debe realizarse al final del bucle de subtrama del codificador y es calculada para toda la trama en el módulo 704 de cálculo de SNR usando la relación:

$$snr = \frac{E_{sw}}{E_e} \quad (4)$$

en la que  $E_{sw}$  es la energía de la señal  $s(n)$  de voz de la trama actual y  $E_e$  es la energía del error entre la señal de voz y la señal de síntesis de la trama actual.

El contador pc de estabilidad de frecuencia fundamental evalúa la variación del periodo fundamental. Se calcula dentro del módulo 705 de clasificación de señal en respuesta a las estimaciones de frecuencia fundamental de bucle abierto de la manera siguiente:

$$pc = |\rho_3 - \rho_2| + |\rho_2 - \rho_1| \quad (5)$$

Los valores  $p_1$ ,  $p_2$  y  $p_3$  corresponden al retardo de frecuencia fundamental de bucle cerrado de las últimas 3 subtramas.

La energía  $E_s$  relativa de la trama es calculada por el módulo 705 como una diferencia entre la energía de la trama actual en dB y su promedio de largo plazo:

5

$$E_s = E_f - E_{ff} \quad (6)$$

en la que la energía  $E_f$  de la trama como la energía de la señal de entrada ventanizada en dB:

10

$$E_f = 10 \log_{10} \left( \frac{1}{L} \sum_{i=0}^{L-1} s^2(i) w_{\text{hanning}}(i) \right) \quad (7)$$

en la que  $L = 160$  es la longitud de la trama y  $w_{\text{hanning}}(i)$  es una ventana de Hanning de longitud  $L$ . La energía promedio a largo plazo es actualizada en las tramas de voz activas usando la relación siguiente:

$$E_{ff} = 0,99E_{ff} + 0,01E_f \quad (8)$$

15

El último parámetro es el parámetro  $z_c$  de cruce por cero calculado en una trama de la señal de voz por el módulo 702 de cálculo de cruce por cero. En esta realización ilustrativa, el contador  $z_c$  de cruce por cero cuenta el número de veces que el signo de la señal cambia de positivo a negativo durante ese intervalo.

20

Para hacer la clasificación más robusta, los parámetros de clasificación son tenidos en cuenta en el módulo 705 de clasificación de señal formando conjuntamente una función  $f_m$  de mérito. Para ello, en primer lugar, a los parámetros de clasificación se les aplica un factor de escala entre 0 y 1, de manera que el valor de cada parámetro típico para la señal sorda se traduce en un 0 y el valor de cada parámetro típico para la señal sonora se traduce en un 1. Se usa una función lineal entre estos valores. Considérese un parámetro  $p_x$ ; su versión escalada se obtiene usando:

25

$$p^s = k_p \cdot p_x + c_p \quad (9)$$

y está limitada entre 0 y 1 (excepto para la energía relativa que está limitada entre 0,5 y 1). Los coeficientes  $k_p$  y  $c_p$  de la función se han encontrado 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:

30

Tabla 2. Parámetros de clasificación de señales y los coeficientes de sus funciones de escala respectivas

Parámetro	Significado	$k_p$	$c_p$
$\bar{r}_x$	Correlación normalizada	0,91743	0,26606
$\bar{e}_t$	Inclinación espectral	2,5	-1,25
$snr$	Relación señal a ruido	0,09615	-0,25
$pc$	Contador de estabilidad de frecuencia fundamental	-0,1176f	2,0
$E_s$	Energía relativa de trama	0,05	0,45
$z_c$	Contador de cruce por cero	-0,067	2,613

La función de mérito se ha definido como:

$$f_m = \frac{1}{7} (2 \cdot \bar{r}_x^s + \bar{e}_t^s + 1,2snr^s + pc^s + E_s^s + zc^s) \quad (10)$$

en la que el superíndice s indica la versión a escala de los parámetros.

A continuación, la función de mérito es escalada por un factor 1,05 si la energía  $E_s^s$  relativa escalada es igual a 0,5 y es escalada por un factor 1,25 si  $E_s^s$  es mayor que 0,75. Además, la función de mérito es escalada también por un factor  $f_E$  derivado en base a una máquina de estados que comprueba la diferencia entre la variación instantánea de la energía relativa y la variación a largo plazo de la energía relativa. Este se añade para mejorar la clasificación de la señal en presencia de ruido de fondo.

Un parámetro  $E_{var}$  de variación de energía relativa es actualizado como:

$$E_{var} = 0,05(E_s - E_{prev}) + 0,95E_{var}$$

en la que  $E_{prev}$  es el valor de  $E_s$  de la trama previa.

Si  $(|E_s - E_{prev}| < (|E_{var}| + 6))$  Y  $(clase_{antigua} = SORDA)$   $f_E = 0,8$

Si no

Si  $((E_s - E_{prev}) > (E_{var} + 3))$  Y  $(clase_{antigua} = SORDA \text{ o } TRANSICIÓN)$   $f_E = 1,1$

Si no

Si  $((E_s - E_{prev}) < (E_{var} - 5))$  Y  $(clase_{antigua} = SONORA \text{ o } SORDA)$   $f_E = 0,6$ .

donde  $clase_{antigua}$  es la clase de la trama previa.

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

Tabla 3. Normas de Clasificación de señales en el codificador

Clase de trama previa	Regla	Clase de trama actual
INICIO SONORA TRANSICIÓN SONORA	$f_m \geq 0,68$	SONORA
	$0,56 \leq f_m < 0,68$	TRANSICIÓN SONORA
	$f_m < 0,56$	SORDA
TRANSICIÓN SORDA SORDA	$f_m > 0,64$	INICIO
	$0,64 \geq f_m > 0,58$	TRANSICIÓN SORDA
	$f_m \leq 0,58$	SORDA

En caso de que haya presente una función de detección de actividad de voz (Voice Activity Detection, VAD) en el codificador, el indicador VAD puede ser usado para la clasificación, ya que indica directamente que no se necesita ninguna clasificación adicional si su valor indica voz inactiva (es decir, la trama es clasificada directamente como SORDA). En esta realización ilustrativa, la trama es clasificada directamente como SORDA si la energía relativa es menor de 10 dB.

Clasificación en el decodificador



Si la aplicación no permite la transmisión de la información de clase (no pueden transportarse bits adicionales), la clasificación todavía puede realizarse en el decodificador. En esta realización ilustrativa, los bits de clasificación se transmiten en la Capa 2, por lo tanto, la clasificación es realizada también en el decodificador para el caso en el que sólo se recibe la Capa de núcleo 1.

- 5 Los parámetros siguientes se usan para la clasificación en el decodificador: una correlación  $r_x$  normalizada, una medida  $e_t$  de inclinación espectral y un contador  $pc$  de estabilidad de frecuencia fundamental, una energía relativa de trama de la señal al final de la trama  $E_s$  actual y un contador  $zc$  de cruce por cero.

El cálculo de estos parámetros, que se usan para clasificar las señales, se explica a continuación.

- 10 La correlación  $r_x$  normalizada es calculada al final de la trama en base a la señal de síntesis. Se usa el retardo de frecuencia fundamental de la última subtrama.

La correlación  $r_x$  normalizada es calculada de manera no sincronizada con la frecuencia fundamental, tal como se indica a continuación:

$$15 \quad r_x = \frac{\sum_{i=0}^{T-1} x(t+i)x(t+i-T)}{\sqrt{\sum_{i=0}^{T-1} x^2(t+i) \sum_{i=0}^{T-1} x^2(t+i-T)}} \quad (11)$$

- 20 en la que T es el retardo de frecuencia fundamental de la última subtrama y  $t = L-T$ , y L es el tamaño de la trama. Si el retardo de frecuencia fundamental de la última subtrama es mayor que  $3N/2$  (N es el tamaño de subtrama), T se establece al retardo promedio de frecuencia fundamental de las dos últimas subtramas.

La correlación  $r_x$  es calculada usando la señal  $s_{salida}(n)$  de voz de síntesis. Para retardos de frecuencia fundamental menores que el tamaño de subtrama (40 muestras) la correlación normalizada es calculada dos veces en los instantes  $t = L-T$  y  $t = L-2T$ , y  $r_x$  es proporcionada como la media de los dos cálculos.

- 25 El parámetro  $e_t$  de inclinación espectral contiene la información acerca de la distribución de frecuencias de la energía. En la presente realización ilustrativa, la inclinación espectral en el decodificador es estimada como el primer coeficiente de autocorrelación normalizado de la señal de síntesis. Se calcula en base a las 3 últimas subtramas como:

$$30 \quad e_t = \frac{\sum_{i=N}^{L-1} x(i)x(i-1)}{\sum_{i=N}^{L-1} x^2(i)} \quad (12)$$

en la que  $x(n) = s_{salida}(n)$  es la señal de síntesis, N es el tamaño de la subtrama y L es el tamaño de trama (N = 40 y L = 160 en esta realización ilustrativa).

- 35 El contador  $pc$  de estabilidad de frecuencia fundamental evalúa la variación del periodo fundamental. Se calcula en el decodificador de la siguiente manera:

$$pc = |p_3 + p_2 - p_1 - p_0| \quad (13)$$

Los valores  $p_0, p_1, p_2$  y  $p_3$  corresponden al retardo de frecuencia fundamental de bucle cerrado de las 4 subtramas.

- 40 La energía  $E_s$  relativa de la trama es calculada como la diferencia entre la energía de la trama actual en dB y su energía promedio a largo plazo:

$$E_s = \bar{E}_r - E_{ll} \quad (14)$$

en la que la energía  $\bar{E}_f$  de la trama es la energía de la señal de síntesis en dB calculada de manera síncrona con respecto a la frecuencia fundamental al final de la trama como:

$$E_f = 10 \log_{10} \left( \frac{1}{T} \sum_{i=0}^{T-1} s_{\text{ca}}^2(i+L-T) \right) \quad (15)$$

5 en la que L = 160 es la longitud de trama y T es el retardo promedio de frecuencia fundamental de las dos últimas subtramas. Si T es menor que el tamaño de subtrama, entonces T se establece a 2T (la energía es calculada usando dos periodos de frecuencia fundamental para retardos de frecuencia fundamental cortos).

La energía promediada de largo plazo es actualizada en las tramas de voz activas usando la relación siguiente:

$$E_n = 0,99E_n + 0,01E_f \quad (16)$$

10

El último parámetro es el parámetro  $z_c$  de cruce por cero calculado sobre una trama de la señal de síntesis. En esta realización ilustrativa, el contador  $z_c$  de cruce por cero cuenta el número de veces que el signo de la señal cambia de positivo a negativo durante ese intervalo.

15

Para hacer la clasificación más robusta, se tienen en cuenta los parámetros de clasificación, formando conjuntamente una función  $f_m$  de mérito. Con ese propósito, los parámetros de clasificación son escalados, en primer lugar, usando una función lineal. Considérese un parámetro  $p_x$ ; su versión escalada se obtiene usando:

$$p^s = k_p \cdot p_x + c_p \quad (17)$$

20

El parámetro de coherencia frecuencia fundamental escalada está limitado a un valor entre 0 y 1, el parámetro de correlación normalizada escalado es doble si es positivo. Los coeficientes  $k_p$  y  $c_p$  de la función se han encontrado 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 FER es mínima. Los valores usados en esta implementación ilustrativa se resumen en la Tabla 4:

25

Tabla 4. Parámetros de clasificación de señales en el decodificador y los coeficientes de sus respectivas funciones de escalado

Parámetro	Significado	$k_p$	$c_p$
$\bar{r}_x$	Correlación normalizada	2,857	-1,286
$\bar{e}_t$	Inclinación espectral	0,8333	0,2917
$pc$	Contador de estabilidad de frecuencia fundamental	-0,0588	1,6468
$E_s$	Energía relativa de trama	0,57143	0,85741
$z_c$	Contador de cruce por cero	-0,067	2,613

La función de la función de mérito se ha definido como:

$$f_m = \frac{1}{6} (2 \cdot \bar{r}_x^s + \bar{e}_t^s + pc^s + E_s^s + zc^s) \quad (18)$$

30

en la que el superíndice s indica la versión escalada de los parámetros.

A continuación, la clasificación se realiza usando la función  $f_m$  de mérito y siguiendo las reglas resumidas en la Tabla 5:

Tabla 5. Reglas de clasificación de señales en el decodificador

Clase de trama previa	Regla	Clase de trama actual
INICIO SONORA TRANSICIÓN SONORA INICIO ARTIFICIAL	$f_m \geq 0,63$	SONORA
	$0,39 \leq f_m < 0,63$	SONORA TRANSICIÓN
	$f_m < 0,39$	SORDA
TRANSICIÓN SORDA SORDA	$f_m > 0,56$	INICIO
	$0,56 \geq f_m > 0,45$	SORDA TRANSICIÓN
	$f_m \leq 0,45$	SORDA

**Parámetros de voz para procesamiento FER**

Hay unos pocos parámetros que se controlan cuidadosamente para evitar artefactos molestos cuando se producen FERs. Si pueden transmitirse unos pocos bits adicionales, entonces estos parámetros pueden ser estimados, cuantificados y transmitidos en el codificador. De lo contrario, algunos de ellos pueden ser estimados en el decodificador. Estos parámetros podrían incluir clasificación de señal, información de energía, información de fase e información de voz.

La importancia del control de energía se manifiesta principalmente cuando se recupera un funcionamiento normal después de un bloque de tramas borradas. Debido a que la mayoría de los codificadores de voz hacen uso de una predicción, la energía correcta no puede ser estimada de manera apropiada en el decodificador. En los segmentos de voz sonoros, la energía incorrecta puede persistir durante varias tramas consecutivas, lo cual es muy molesto, especialmente cuando esta energía incorrecta aumenta.

La energía no sólo es controlada para la voz sonora, debido a la predicción a largo plazo (predicción de frecuencia fundamental), también es controlada para la voz sorda. La razón aquí es la predicción del cuantificador de ganancia de innovación usado frecuentemente en los codificadores de tipo CELP. La energía incorrecta durante los segmentos sordos puede causar una molesta fluctuación alta frecuencia.

El control de fase es también una parte a tener en cuenta. Por ejemplo, la información de fase es enviada con relación a la posición de pulso glótico. En la solicitud de patente PCT en [1], la información de fase es transmitida como la posición del primer pulso glótico en la trama, y se usa para reconstruir los inicios sonoros perdidos. Un uso adicional de la información de fase es para volver a sincronizar el contenido del libro de códigos adaptativo. Esto mejora la convergencia del decodificador en la trama oculta y las tramas siguientes y mejora significativamente la calidad de la voz. El procedimiento para la resincronización del libro de códigos adaptativo (o excitación pasada) puede realizarse de diversas maneras, dependiendo de la información de fase recibida (recibida o no) y del retardo disponible en el decodificador.

**Información de energía**

La información de la energía puede ser estimada y enviada en el dominio residual LP o en el dominio de señal de voz. El envío de la información en el dominio residual tiene la desventaja de no tener en cuenta la influencia del filtro de síntesis LP. Esto puede ser particularmente difícil en el caso de una recuperación sonora después de varias tramas sonoras perdidas (cuando la FER se produce durante un segmento de voz sonoro). Cuando una FER llega después de una trama sonora, la excitación de la última trama buena es usada típicamente durante la ocultación con alguna estrategia de atenuación. Cuando un nuevo filtro de síntesis LP llega con la primera trama buena después del borrado, puede haber una discrepancia entre la energía de excitación y la ganancia del filtro de síntesis LP. El nuevo filtro de síntesis puede producir una señal de síntesis cuya energía es muy diferente de la energía de la última trama borrada sintetizada y también de la energía de la señal original. Por esta razón, la energía es calculada y cuantificada en el dominio de la señal.

La energía  $E_q$  se calcula y se cuantifica en el módulo 706 de estimación y cuantificación de energía de la Figura 7.

En esta realización ilustrativa no restrictiva, se usa un cuantificador uniforme de 5 bits en el intervalo de 0 dB a 96 dB con una etapa de 3,1 dB. El índice de cuantificación viene determinado por la parte entera de:

$$j = \frac{10 \log_{10}(E + 0,001)}{3,1} \quad (19)$$

5

en la que el índice está limitado a  $0 \leq i \leq 31$ .

$E$  es la energía máxima de muestra para las tramas clasificadas como SONORAS o INICIO, o la energía promedio por muestra para otras tramas. Para las tramas SONORAS o INICIO, la energía máxima de la muestra se calcula de manera síncrona con respecto a la frecuencia fundamental al final de la trama de la manera siguiente:

10

$$E = \max_{i=L-t_E}^{L-1} (s^2(i)) \quad (20)$$

15

en la que  $L$  es la longitud de la trama y la señal  $s(i)$  hace referencia a la señal de voz. Si el retardo de frecuencia fundamental es mayor que el tamaño de subtrama (40 muestras en esta realización ilustrativa),  $t_E$  es igual al retardo de frecuencia fundamental redondeado de bucle cerrado de la última subtrama. Si el retardo de frecuencia fundamental es más corto que 40 muestras, entonces  $t_E$  se establece a dos veces el retardo de frecuencia fundamental redondeado de bucle cerrado 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 establece a  $L/2$  y  $E$  se calcula como:

20

$$E = \frac{1}{t_E} \sum_{i=L-t_E}^{L-1} s^2(i) \quad (21)$$

En esta realización ilustrativa, se usa la señal de síntesis local en el codificador para calcular la información de la energía.

25

En esta realización ilustrativa, la información sobre la energía es transmitida en la Capa 4. De esta manera, si se recibe la Capa 4, esta información puede ser usada para mejorar la ocultación del borrado de tramas. Si no, la energía es estimada en el lado del decodificador.

#### Información de control de fase

30

Se usa un control de fase mientras se recupera después de un segmento de voz sonora perdido por razones similares a las descritas en la sección anterior. Después de un bloque de tramas borradas, las memorias del decodificador pierden sincronización con las memorias del codificador. Para volver a sincronizar el decodificador, puede transmitirse cierta información de fase. Como un ejemplo no limitativo, la posición y el signo del último pulso glótico en la trama anterior pueden ser enviados como la información de fase. A continuación, esta información de fase es usada para la recuperación después de inicios sonoros perdidos, tal como se describirá más adelante.

35

También, tal como se describirá más adelante, esta información se usa también para volver a sincronizar la señal de excitación de las tramas borradas con el fin de mejorar la convergencia en las tramas consecutivas recibidas correctamente (reduce el error propagado).

40

La información de fase puede corresponder al primer pulso glótico en la trama o al último pulso glótico en la trama anterior. La elección dependerá de si hay o no un retardo adicional disponible en el decodificador. En esta realización ilustrativa, hay un retardo de trama disponible en el decodificador para la operación de solapamiento y adición en la reconstrucción MDCT. De esta manera, cuando se borra una única trama, los parámetros de la trama futura están disponibles (debido al retardo de trama adicional). En este caso, la posición y el signo del pulso máximo al final de la trama borrada están disponibles a partir de la trama futura. Por lo tanto la excitación de frecuencia fundamental puede ser ocultada de manera que el último pulso máximo es alineado con la posición recibida en la trama futura. Esto se describirá más detalladamente a continuación.

45

Es posible que no haya disponible ningún retardo adicional. En este caso, la información de fase no se usa cuando se oculta la trama borrada. Sin embargo, en la trama buena recibida después de la trama borrada, la información de

fase es usada para realizar la sincronización de pulso glótico en la memoria del libro de códigos adaptativo. Esto mejorará el rendimiento en la reducción de la propagación de errores.

Supóngase que  $T_0$  es el retardo de frecuencia fundamental redondeado de bucle cerrado para la última subtrama. La búsqueda del pulso máximo se realiza sobre el residual LP filtrado de paso bajo. El residual filtrado de paso bajo viene dado por:

$$r_{LP}(n) = 0.25r(n-1) + 0.5r(n) + 0.25r(n+1) \quad (22)$$

El módulo 707 de búsqueda y cuantificación de pulsos glóticos busca la posición del último pulso  $\tau$  glótico entre las últimas  $T_0$  muestras del residual filtrado de paso bajo en la trama, buscando la muestra con la amplitud máxima absoluta ( $\tau$  es la posición con relación al final de la trama).

La posición del último pulso glótico es codificada usando 6 bits de la manera siguiente. La precisión usada para codificar la posición del primer pulso glótico depende del valor de frecuencia fundamental de bucle cerrado para la última subtrama  $T_0$ . Esto es posible debido a que este valor es conocido tanto por el codificador como por el decodificador, y no está sometido a la propagación de errores después de una o varias pérdidas de trama. Cuando  $T_0$  es menor de 64, la posición del último pulso glótico con respecto al final de la trama es codificada directamente con una precisión de una muestra. Cuando  $64 \leq T_0 < 128$ , la posición del último pulso glótico con respecto al final de la trama es codificada con una precisión de dos muestras usando una división de enteros simple, es decir,  $\tau/2$ . Cuando  $T_0 \geq 128$ , la posición del último pulso glótico con respecto al final de la trama es codificada con una precisión de cuatro muestras dividiendo  $\tau$  adicionalmente por 2. El procedimiento inverso se realiza en el decodificador. Si  $T_0 < 64$ , la posición cuantificada recibida es usada como tal. Si  $64 \leq T_0 < 128$ , la posición cuantificada recibida es multiplicada por 2 e incrementada en 1. Si  $T_0 \geq 128$ , la posición cuantificada recibida es multiplicada por 4 e incrementada en 2 (el incremento en 2 resulta en un error de cuantificación distribuido uniformemente).

También se cuantifica el signo de la amplitud máxima de pulso absoluto. Esto proporciona un total de 7 bits para la información de fase. El signo es usado para volver a sincronizar la fase ya que la forma del pulso glótico contiene, frecuentemente, dos pulsos grandes con signos opuestos. Haciendo caso omiso de la señal puede resultar en una pequeña desviación en la posición y puede reducir el rendimiento del procedimiento de resincronización.

Cabe señalar que pueden usarse procedimientos eficientes para la cuantificación de la información de fase. Por ejemplo, la posición del último pulso en la trama anterior puede ser cuantificada con relación a una posición estimada a partir del retardo de frecuencia fundamental de la primera subtrama en la trama actual (la posición puede ser estimada fácilmente a partir del primer pulso en la trama retrasada por el retardo de frecuencia fundamental).

En el caso en el que hay más bits disponibles, la forma del pulso glótico puede ser codificada. En este caso, la posición del primer pulso glótico puede ser determinada mediante un análisis de correlación entre la señal residual y las posibles formas de pulso, signos (positivo o negativo) y posiciones. La forma del pulso puede ser tomada de un libro de códigos de formas de pulso conocidas tanto en el codificador como en el decodificador, conociéndose este procedimiento como cuantificación vectorial por las personas con conocimientos ordinarios en la materia. A continuación, la forma, el signo y la amplitud del primer pulso glótico se codifican y se transmiten al decodificador.

#### Procesamiento de las tramas borradas

Las técnicas de ocultación de FER en esta realización ilustrativa se demuestran en los códecs de tipo ACELP. Sin embargo, pueden ser aplicadas fácilmente a cualquier códec de voz, donde la señal de síntesis es generada filtrando una señal de excitación a través de un filtro de síntesis LP. La estrategia de ocultación puede ser resumida como una convergencia de la energía de la señal y la envolvente espectral a los parámetros estimados del ruido de fondo. La periodicidad de la señal se hace converger a cero. La velocidad de la convergencia depende de los parámetros de la última clase de trama buena recibida y el número de tramas borradas consecutivas y es controlada por un factor  $\alpha$  de atenuación. El factor  $\alpha$  depende además de la estabilidad del filtro LP para tramas SORDAS. En general, la convergencia es lenta si la última trama buena recibida está en un segmento estable y es rápida si la trama está en un segmento de transición. Los valores de  $\alpha$  se resumen en la Tabla 6.

Tabla 6. Los valores del factor  $\alpha$  de atenuación de ocultación FER

Ultima trama buena recibida	Número de tramas sucesivas borradas	$\alpha$
SONORA, INICIO, INICIO ARTIFICIAL	1	$\beta$
	> 1	$\bar{g}_p$
TRANSICIÓN SONORA	$\leq 2$	0,8
	> 2	0,2
TRANSICIÓN SORDA		0,88
SORDA	= 1	0,95
	> 1	$0,5 \theta + 0,4$

En la Tabla 6,  $\bar{g}_p$  es una ganancia de frecuencia fundamental promedio por trama proporcionada por:

$$\bar{g}_p = 0,1g_p^{(0)} + 0,2g_p^{(1)} + 0,3g_p^{(2)} + 0,4g_p^{(3)} \quad (23)$$

5

en la que  $g_p^{(i)}$  es la ganancia de frecuencia fundamental en la subtrama i.

El valor de  $\beta$  viene determinado por

$$\beta = \sqrt{\bar{g}_p} \quad \text{limitado por } 0,85 \leq \beta \leq 0,98 \quad (24)$$

10

El valor  $\theta$  es un factor de estabilidad calculado en base a una medida de distancia entre los filtros LP contiguos. Aquí, el factor  $\theta$  está relacionado con la medida de distancia LSP (par espectral de línea) y está delimitado por  $0 \leq \theta \leq 1$ , con los valores mayores de  $\theta$  correspondientes a señales más estables. Esto resulta en una disminución de la energía y fluctuaciones de la envolvente espectral cuando se produce un borrado de trama aislado dentro de un segmento sordo estable. En esta realización ilustrativa, el factor  $\theta$  de estabilidad viene dado por:

15

$$\theta = 1,25 - \frac{1}{1,4} \sum_{i=0}^9 (LSP_i - LSP_{ant_i})^2 \quad \text{delimitado por } 0 \leq \theta \leq 1. \quad (25)$$

en la que  $LSP_i$  son las tramas LSPs actuales y  $LSP_{ant_i}$ , son las tramas LSPs pasadas. Obsérvese que los LSPs están en el dominio de coseno (de -1 a 1).

20

En el caso en el que la información de clasificación de la trama futura no está disponible, la clase se establece de manera que sea la misma que en la última trama buena recibida. Si la información de clase está disponible en la trama futura, la clase de la trama perdida es estimada en base a la clase en la trama futura y la clase de la última trama buena. En esta realización ilustrativa, la clase de la trama futura puede estar disponible si se recibe la Capa 2 de la trama futura (tasa de bits de la trama futura superior a 8 kbit/s y no perdida). Si el codificador funciona a una tasa de bits máxima de 12 kbit/s, entonces el retardo de trama adicional en el decodificador usado para el solapamiento y adición MDCT no es necesario y el implementador puede elegir reducir el retardo de decodificador. En este caso, la ocultación se realizará sólo sobre la información pasada. Este se denominará como modo de decodificador de bajo retardo.

25

Supóngase que  $clase_{antigua}$  denota la clase de la última trama buena, y  $clase_{nueva}$  denota la clase de la trama futura y  $clase_{perdida}$  es la clase de la trama perdida a estimar.

30

Inicialmente,  $clase_{perdida}$  se establece de manera que sea igual a  $clase_{antigua}$ . Si la trama futura está disponible,

entonces su información de clase es decodificada en  $clase_{nueva}$ . Entonces, el valor  $clase_{perdida}$  se actualiza como se indica a continuación:

- Si  $clase_{nueva}$  es SONORA y  $clase_{antigua}$  es INICIO, entonces la  $clase_{perdida}$  se establece a SONORA.
- 5     • Si  $clase_{nueva}$  es SONORA y la clase de la trama antes de la última trama buena es INICIO o SONORA, entonces  $clase_{perdida}$  se establece a SONORA.
- Si  $clase_{nueva}$  es SORDA y la  $clase_{antigua}$  es SONORA, entonces la  $clase_{perdida}$  se establece a TRANSICIÓN SORDA.
- Si  $clase_{nueva}$  es SONORA O INICIO y  $clase_{antigua}$  es SORDA, entonces  $clase_{perdida}$  se establece a INICIO SIN (reconstrucción de inicio).

## 10     **Construcción de la parte periódica de la excitación**

Para una ocultación de tramas borradas cuya clase se establece a SORDA o TRANSICIÓN SORDA, no se genera ninguna parte periódica de la señal de excitación. Para otras clases, la parte periódica de la señal de excitación es construida de la manera siguiente.

15     En primer lugar, el último ciclo de frecuencia fundamental de la trama anterior es copiado repetidamente. Si es el caso de la 1ª trama borrada después de una trama buena, a este ciclo de frecuencia fundamental se le aplica un filtro de paso bajo. El filtro usado es un filtro FIR (Finite Impulse Response, respuesta finita a pulso) de fase lineal de 3 etapas con coeficientes de filtro iguales a 0,18, 0,64 y 0,18.

20     El periodo  $T_c$  de la frecuencia fundamental usado para seleccionar el último ciclo de frecuencia fundamental y, por lo tanto, usado durante la ocultación, se define de manera que puedan evitarse o reducirse los múltiplos o submúltiplos de la frecuencia fundamental. Se usa la lógica siguiente para determinar el periodo fundamental

$$\text{Si } ((T_3 < 1,8 T_s) \text{ Y } (T_3 > 0,6 T_s)) \text{ O } (T_{cnt} \geq 30), \text{ entonces, } T_0 = T_3, \text{ si no } T_0 = T_5.$$

25     Aquí,  $T_3$  es el periodo fundamental redondeado de la 4ª subtrama de la última trama buena recibida y  $T_s$  es el periodo fundamental predicho redondeado de la 4ª subtrama de la última trama sonora estable con estimaciones de frecuencia fundamental coherentes. Una trama sonora estable se define aquí como una trama SONORA precedida de una trama de tipo sonora (TRANSICIÓN SONORA, SONORA o INICIO). La coherencia de la frecuencia fundamental es verificada en esta implementación examinando si las estimaciones de frecuencia fundamental de bucle cerrado son razonablemente cercanas, es decir, si las relaciones entre la frecuencia fundamental de la última subtrama, la frecuencia fundamental de la segunda subtrama y la frecuencia fundamental de la última subtrama de la trama anterior se encuentran dentro del intervalo (0,7, 1,4). De manera alternativa, si hay múltiples tramas perdidas,  $T_3$  es el periodo fundamental estimado redondeado de la 4ª subtrama de la última trama oculta.

30     Esta determinación del periodo  $T_c$  fundamental significa que si la frecuencia fundamental al final de la última trama buena y la frecuencia fundamental de la última trama estable son cercanas una a la otra, se usa la frecuencia fundamental de la última trama buena. De lo contrario, esta frecuencia fundamental se considera poco fiable y en su lugar, se usa la frecuencia fundamental de la última trama estable para evitar el impacto de las estimaciones erróneas de la frecuencia fundamental en los inicios sonoros. Sin embargo, esta lógica solo tiene sentido si el último segmento estable no está demasiado lejano en el pasado. Por lo tanto, se define un contador  $T_{cnt}$  que limita el alcance de la influencia del último segmento estable. Si  $T_{cnt}$  es mayor o igual a 30, es decir, si hay al menos 30 tramas desde la última actualización de  $T_s$ , se usa la frecuencia fundamental de la última trama buena, de manera sistemática.  $T_{cnt}$  se reinicializa cada vez que se detecta un segmento estable y se actualiza  $T_s$ . Entonces, el periodo  $T_c$  se mantiene constante durante la ocultación para todo el bloque borrado.

35     Para las tramas borradas después de una trama recibida correctamente que no sea SORDA, la memoria intermedia de excitación es actualizada con sólo esta parte periódica de la excitación. Esta actualización será usada para construir la excitación del libro de códigos de frecuencia fundamental en la trama siguiente.

45     El procedimiento descrito anteriormente puede resultar en una deriva en la posición del pulso glótico, ya que el periodo fundamental usado para construir la excitación puede ser diferente del periodo fundamental real en el codificador. Esto causará que la memoria intermedia del libro de códigos adaptativo (o memoria intermedia de excitación pasada) esté desincronizada con respecto a la memoria intermedia de excitación real. De esta manera, en caso de que se reciba una trama buena después de la trama borrada, la excitación de la frecuencia fundamental (o excitación del libro de códigos adaptativo) tendrá un error que puede persistir durante varias tramas y puede

50     afectar al rendimiento de las tramas recibidas correctamente.

La Figura 9 es un diagrama de flujo que muestra el procedimiento 900 de ocultación de la parte periódica de la

excitación descrita en la realización ilustrativa, y la Figura 10 es un diagrama de flujo que muestra el procedimiento 1000 de sincronización de la parte periódica de la excitación.

Para superar este problema y mejorar la convergencia en el decodificador, se describe un procedimiento (900 en la Figura 9) de resincronización que ajusta la posición del último pulso glótico en la trama oculta a ser sincronizada con la posición real del pulso glótico. En una primera implementación, este procedimiento de resincronización puede ser realizado en base a una información de fase con respecto a la posición real del último pulso glótico en la trama oculta que es transmitida en la trama futura. En una segunda implementación, la posición del último pulso glótico es estimada en el decodificador cuando la información a partir de la trama futura no está disponible.

Tal como se ha descrito anteriormente, la excitación de frecuencia fundamental de toda la trama perdida es construida repitiendo el último ciclo  $T_c$  de frecuencia fundamental de la trama anterior (operación 906 en la Figura 9), donde  $T_c$  se ha definido anteriormente. Para la primera trama borrada (detectada durante la operación 902 en la Figura 9), el ciclo de frecuencia fundamental es sometido, en primer lugar, a un filtrado paso bajo (operación 904 en la Figura 9) usando un filtro con coeficientes 0,18, 0,64 y 0,18. Esto se realiza de la manera siguiente:

$$\begin{aligned} u(n) &= 0,18u(n - T_c - 1) + 0,64u(n - T_c) + 0,18u(n - T_c + 1), & n = 0, \dots, T_c - 1 \\ u(n) &= u(n - T_c), & n = T_c, \dots, L + N - 1 \end{aligned} \quad (26)$$

en la que  $u(n)$  es la señal de excitación,  $L$  es el tamaño de trama y  $N$  es el tamaño de subtrama. Si esta no es la primera trama borrada, la excitación ocultada se construye simplemente como:

$$u(n) = u(n - T_c), \quad n = 0, \dots, L + N - 1 \quad (27)$$

Cabe señalar que la excitación ocultada es calculada también para una subtrama adicional para ayudar en la resincronización, tal como se mostrará a continuación.

Una vez encontrada la excitación ocultada, el procedimiento de resincronización se realiza como se indica a continuación. Si la trama futura está disponible (operación 908 en la Figura 9) y contiene la información del pulso glótico, entonces esta información es decodificada (operación 910 en la Figura 9). Tal como se ha descrito anteriormente, esta información consiste en la posición del pulso máximo absoluto desde el final de la trama y su signo. Denótese esta posición decodificada mediante  $P_0$ , entonces la posición real del pulso máximo absoluto viene determinada por:

$$P_{\text{último}} = L - P_0$$

A continuación, la posición del pulso máximo en la excitación ocultada desde el inicio de la trama con un signo similar a la información de signo decodificada es determinada en base a una excitación filtrada paso bajo (operación 912 en la Figura 9). Es decir, si la posición de pulso máxima decodificada es positiva, entonces se determina un pulso positivo máximo en la excitación ocultada desde el comienzo de la trama, de lo contrario, se determina el pulso negativo máximo. Denótese el primer pulso máximo en la excitación ocultada mediante  $T(0)$ . Las posiciones de los otros pulsos máximos vienen determinadas por (operación 914 en la Figura 9):

$$T(i) = T(0) + iT_c, \quad i = 1, \dots, N_p - 1 \quad (28)$$

en la que  $N_p$  es el número de pulsos (incluyendo el primer pulso en la trama futura).

El error en la posición de pulso del último pulso ocultado en la trama (operación 916 en la Figura 9) es encontrado buscando el pulso  $T(i)$  más cercano al pulso  $P_{\text{último}}$  real. Si el error viene dado por:

$$T_e = P_{\text{último}} - T(k),$$

En la que  $k$  es el índice del pulso más cercano a  $P_{\text{último}}$ .

Si  $T_e = 0$ , entonces no se requiere ninguna resincronización (operación 918 en la Figura 9). Si el valor de  $T_e$  es



positivo ( $T(k) < P_{\text{último}}$ ), entonces deben insertarse  $T_e$  muestras (operación 1002 en la Figura 10). Si  $T_e$  es negativo ( $T(k) > P_{\text{último}}$ ), entonces deben eliminarse  $T_e$  muestras (operación 1002 en la Figura 10). Además, solo se realiza la resincronización si  $T_e < N$  y  $T_e < N_p \times T_{\text{dif}}$ , donde  $N$  es el tamaño de la subtrama y  $T_{\text{dif}}$  es la diferencia absoluta entre  $T_c$  y el retardo de frecuencia fundamental de la primera subtrama en la trama futura (operación 918 en la Figura 9).

5 Las muestras que deben ser añadidas o suprimidas se distribuyen a lo largo de los ciclos de frecuencia fundamental en la trama. Se determinan las regiones de energía mínima en los diferentes ciclos de frecuencia fundamental y se realiza la eliminación o inserción de muestras en esas regiones. El número de pulsos de frecuencia fundamental en la trama es  $N_p$  en las posiciones  $T(i)$ ,  $i = 0, \dots, N_p-1$  respectivas. El número de regiones de energía mínima es  $N_{p-1}$ . Las regiones de energía mínima se determinan calculando la energía usando una ventana deslizante de 5 muestras (operación 1002 en la Figura 10). La posición de energía mínima se establece en el centro de la ventana en la que la energía es un mínimo (operación 1004 en la Figura 10). La búsqueda realizada entre dos pulsos de frecuencia fundamental en la posición  $T(i)$  y  $T(i+1)$  está restringida entre  $T(i) + T_c/4$  y  $T(i+1) - T_c/4$ .

10 Denótese las posiciones mínimas, determinadas tal como se ha descrito anteriormente, como  $T_{\text{min}}(i)$ ,  $i = 0, \dots, N_{\text{min}}-1$ , donde  $N_{\text{min}} = N_p - 1$  es el número de regiones de energía mínima. La eliminación o inserción de muestras se realiza alrededor de  $T_{\text{min}}(i)$ . Las muestras a añadir o eliminar se distribuyen a lo largo de los diferentes ciclos de frecuencia fundamental, tal como se describe a continuación.

Si  $N_{\text{min}} = 1$ , entonces sólo hay una región de energía mínima y todos los pulsos  $T_e$  se insertan o eliminan en  $T_{\text{min}}(0)$ .

20 Para  $N_{\text{min}} > 1$ , se usa un algoritmo simple para determinar el número de muestras a añadir o eliminar en cada ciclo de frecuencia fundamental, de manera que se añaden/eliminan menos muestras al inicio y más hacia el final de la trama (operación 1006 en la Figura 10). En esta realización ilustrativa, para los valores de número total de pulsos a eliminar/añadir  $T_e$  y número de regiones de mínimo mínima  $N_{\text{min}}$ , el número de muestras a eliminar/añadir por ciclo de frecuencia fundamental,  $R(i)$ ,  $i = 0 \dots N_{\text{min}}-1$ , se encuentra usando la relación recursiva siguiente (operación 1006 en la Figura 10):

$$25 \quad R(i) = \text{redondear} \left( \frac{(i+1)^2}{2} f - \sum_{k=0}^{i-1} R(k) \right) \quad (29)$$

en la que

$$30 \quad f = \frac{2|T_e|}{N_{\text{min}}^2}$$

Cabe señalar que, en cada etapa, se comprueba la condición  $R(i) < R(i-1)$  y, si se cumple, entonces se intercambian los valores de  $R(i)$  y  $R(i-1)$ .

35 Los valores  $R(i)$  corresponden a ciclos de frecuencia fundamental empezando desde el inicio de la trama.  $R(0)$  corresponde a  $T_{\text{min}}(0)$ ,  $R(1)$  corresponde a  $T_{\text{min}}(1)$ , ...,  $R(N_{\text{min}}-1)$  corresponde a  $T_{\text{min}}(N_{\text{min}}-1)$ . Debido a que los valores de  $R(i)$  están en orden creciente, se añaden/eliminan más muestras hacia los ciclos del final de la trama.

A modo de ejemplo, para el cálculo de  $R(i)$ , para  $T_e = 11$  o  $-11$   $N_{\text{min}} = 4$  (11 muestras a añadir/eliminar y 4 ciclos de frecuencia fundamental en la trama), se encuentran los siguientes valores de  $R(i)$ :

$$f = 2 \times 11/16 = 1,375$$

$$40 \quad R(0) = \text{redondear} (f/2) = 1$$

$$R(1) = \text{redondear} (2f - 1) = 2$$

$$R(2) = \text{redondear} (4,5f - 1 - 2) = 3$$

$$R(3) = \text{redondear} (8f - 1 - 2 - 3) = 5$$

45 De esta manera, se añade/elimina 1 muestra alrededor de la posición de energía mínima  $T_{\text{min}}(0)$ , se añaden/eliminan 2 muestras alrededor de la posición de energía mínima  $T_{\text{min}}(1)$ , se añaden/eliminan 3 muestras alrededor de la posición de energía mínima  $T_{\text{min}}(2)$ , y se añaden/eliminan 5 muestras alrededor de la posición de

energía mínima  $T_{\min}(3)$  (operación 1008 en la Figura 10).

La eliminación de muestras es sencilla. La adición de muestras (operación 1008 en la Figura 10) se realiza en esta realización ilustrativa copiando las últimas  $R(i)$  muestras después de dividir por 20 e invertir el signo. En el ejemplo anterior, en el que deben insertarse 5 muestras en la posición  $T_{\min}(3)$  se realiza lo siguiente:

$$5 \quad u(T_{\min}(3) + i) = -u(T_{\min}(3) + i - R(3)) / 20, \quad i=0, \dots, 4 \quad (30)$$

Usando el procedimiento descrito anteriormente, el último pulso máximo en la excitación ocultada es forzado a alinearse a la posición del pulso máximo real al final de la trama que es transmitida en la trama futura (operación 920 en la Figura 9 y operación 1010 en la Figura 10).

10 Si la información de la fase del pulso no está disponible, pero la trama futura está disponible, el valor de la frecuencia fundamental de la trama futura puede ser interpolado con el valor de la frecuencia fundamental pasada para encontrar los retardos de frecuencia fundamental estimados por subtrama. Si la trama futura no está disponible, el valor de frecuencia fundamental de la trama ausente puede ser estimado, a continuación, puede ser interpolado con el valor de la frecuencia fundamental pasado para encontrar los retardos estimados de la frecuencia fundamental por subtrama. A continuación, se calcula el retardo total de todos los ciclos de frecuencia fundamental en la trama ocultada, tanto para la última frecuencia fundamental usada en la ocultación como para los retardos de frecuencia fundamental estimados por subtrama. La diferencia entre estos dos retardos totales proporciona una estimación de la diferencia entre el último pulso máximo ocultado en la trama y el pulso estimado. A continuación, los pulsos pueden ser sincronizados de nuevo tal como se ha descrito anteriormente (operación 920 en la Figura 9 y operación 1010 en la Figura 10).

Si el decodificador no tiene ningún retardo adicional, la información de fase de pulso presente en la trama futura puede ser usada en la primera trama buena recibida para volver a sincronizar la memoria del libro de códigos adaptativo (la excitación pasada) y obtener el último pulso glótico máximo alineado con la posición transmitida en la trama actual antes de construir la excitación de la trama actual. En este caso, la sincronización se realizará exactamente como se ha descrito anteriormente, pero en la memoria de la excitación en lugar de ser realizada en la excitación actual. En este caso, la construcción de la excitación actual se iniciará con una memoria sincronizada.

25 Cuando no hay ningún retardo adicional disponible, también es posible enviar la posición del primer pulso máximo de la trama actual en lugar de la posición del último pulso glótico máximo de la última trama. Si este es el caso, la sincronización se consigue también en la memoria de la excitación antes de construir la excitación actual. Con esta configuración, la posición real del pulso máximo absoluto en la memoria de la excitación viene determinada por:

$$P_{\text{último}} = L + P_0 - T_{\text{nuevo}}$$

en la que  $T_{\text{nuevo}}$  es el primer ciclo de frecuencia fundamental de la nueva trama y  $P_0$  es la posición decodificada del primer pulso glótico máximo de la trama actual.

35 Debido a que el último pulso de la excitación de la trama anterior es usado para la construcción de la parte periódica, su ganancia es aproximadamente correcta al inicio de la trama oculta y puede ser ajustada a 1 (operación 922 en la Figura 9). A continuación, la ganancia es atenuada linealmente a lo largo de la trama, muestra a muestra, para obtener el valor de  $\alpha$  en el extremo de la trama (operación 924 en la Figura 9).

40 Los valores de  $\alpha$  (operación 922 en la Figura 9) corresponden a los valores de la Tabla 6, que tienen en cuenta la evolución de la energía de los segmentos de voz. Esta evolución puede ser extrapolada hasta cierto punto usando los valores de ganancia de excitación de frecuencia fundamental de cada subtrama de la última trama buena. En general, si estas ganancias son mayores que 1, la energía de la señal es creciente, si son menores que 1, la energía es decreciente. De esta manera,  $\alpha$  es fija a  $\beta = \sqrt{g_p}$ , tal como se ha descrito anteriormente. El valor de  $\beta$  está restringido entre 0,98 y 0,85 para evitar fuertes aumentos y disminuciones de energía.

45 Para tramas las borradas después de una trama recibida correctamente que no sea SORDA, la memoria intermedia de excitación es actualizada con sólo la parte periódica de la excitación (después de una resincronización y escalado de ganancia). Esta actualización será usada para construir la excitación de libro de códigos de frecuencia fundamental en la trama siguiente (operación 926 en la Figura 9).

50 La Figura 11 muestra ejemplos típicos de la señal de excitación con y sin el procedimiento de sincronización. La señal de excitación original sin borrado de tramas se muestra en la Figura 11b. La Figura 11c muestra la señal de

excitación ocultada cuando se borra la trama mostrada en la Figura 11a, sin usar el procedimiento de sincronización. Puede observarse claramente que el último pulso glótico en la trama ocultada no está alineado con la posición de pulso verdadera mostrada en la Figura 11b. Además, puede observarse que el efecto de ocultación del borrado de trama persiste en las tramas siguientes no borradas. La Figura 11d muestra la señal de excitación ocultada cuando se ha usado el procedimiento de sincronización según la realización ilustrativa descrita anteriormente de la invención. Puede observarse claramente que el último pulso glótico en la trama ocultada está correctamente alineado con la posición de pulso verdadera mostrada en la Figura 11b. Además, puede observarse que el efecto de la ocultación del borrado de trama en las siguientes tramas recibidas correctamente es menos problemático que el caso de la Figura 11c. Esta observación se confirma en las Figuras 11e y 11f. La Figura 11e muestra el error entre la excitación original y la excitación ocultada sin sincronización. La Figura 11-F muestra el error entre la excitación original y la excitación ocultada cuando se usa el procedimiento de sincronización.

La Figura 12 muestra ejemplos de la señal de voz reconstruida usando las señales de excitación mostradas en la Figura 11. La señal reconstruida sin borrado de trama se muestra en la Figura 12b. La Figura 12c muestra la señal de voz reconstruida cuando se borra la trama mostrada en la Figura 12a, sin usar el procedimiento de sincronización. La Figura 12d muestra la señal de voz reconstruida cuando se borra la trama mostrada en la Figura 12a, con el uso del procedimiento de sincronización tal como se ha descrito en la realización ilustrativa anterior de la presente invención. La Figura 12e muestra la relación de señal-a-ruido (SNR) por subtrama entre la señal original y la señal en la Figura 12c. Puede observar en la Figura 12e que la SNR se mantiene muy baja, incluso cuando se reciben tramas buenas (se mantiene por debajo de 0 dB para las próximas dos tramas buenas y se mantiene por debajo de 8 dB hasta la 7ª trama buena). La Figura 12f muestra la relación de señal-a-ruido (SNR) por subtrama entre la señal original y la señal en la Figura 12d. Puede observarse en la Figura 12d que la señal converge rápidamente a la señal reconstruida verdadera. La SNR crece rápidamente por encima de 10 dB después de dos tramas buenas.

**Construcción de la parte aleatoria de la excitación**

La parte de innovación (no periódica) de la señal de excitación es generada aleatoriamente. Puede ser generada como un ruido aleatorio o usando el libro de códigos de innovación CELP con índices de vector generados aleatoriamente. En la presente realización ilustrativa, se ha usado un generador aleatorio simple con una distribución aproximadamente uniforme. Antes de ajustar la ganancia de la innovación, la innovación generada aleatoriamente es escalada a un valor de referencia, fijado aquí a la energía unitaria para cada muestra.

En el inicio de un bloque borrado, la ganancia  $g_s$  de innovación es inicializada usando las ganancias de excitación de innovación de cada subtrama de la última trama buena:

$$g_s = 0,1g(0) + 0,2g(1) + 0,3g(2) + 0,4g(3) \tag{31}$$

en la que  $g(0)$ ,  $g(1)$ ,  $g(2)$  y  $g(3)$  son el libro de códigos fijo, o de innovación, las ganancias 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 de la atenuación de la excitación de la frecuencia fundamental. La razón es que la excitación de la frecuencia fundamental (y, de esta manera, a periodicidad de la excitación) converge a 0 mientras la excitación aleatoria converge a la energía de excitación de generación de ruido de confort (CNG). La atenuación de la ganancia de innovación se realiza como:

$$g_s^1 = \alpha \cdot g_s^0 + (1 - \alpha) \cdot g_n \tag{32}$$

en la que  $g_s^1$  es la ganancia de la innovación en el inicio de la siguiente trama, y  $g_s^0$  es la ganancia de la innovación en el inicio de la trama actual,  $g_n$  es la ganancia de la excitación usada durante la generación de ruido de confort y  $\alpha$  es tal como se ha definido en la Tabla 5. De manera similar a la atenuación de la excitación periódica, de esta manera, la ganancia es atenuada linealmente a lo largo de la trama, muestra a muestra, empezando con  $g_s^0$  y hasta el valor  $g_s^1$  que se conseguiría en el inicio de la trama siguiente.

Por último, si la última trama buena recibida (recibida correctamente o no borrada) es diferente de SORDA, la excitación de innovación es filtrada a través de un filtro FIR, paso alto, de fase lineal, con coeficientes -0,0125, -0,109, 0,7813, -0,109, -0,0125. Para reducir la cantidad de componentes ruidosos durante los segmentos de voz, estos coeficientes de filtro se multiplican por un factor de adaptación igual a  $(0,75 - 0,25 r_v)$ , siendo  $r_v$  un factor de voz en el intervalo de -1 a 1. A continuación la parte aleatoria de la excitación es añadida a la excitación adaptativa para formar la señal de excitación total.

Si la última trama buena es SORDA, sólo se usa la excitación de innovación y es atenuada adicionalmente en un factor de 0,8. En este caso, la memoria intermedia de excitación pasada es actualizada con la excitación de innovación ya que no hay disponible ninguna parte periódica de la excitación.

**Ocultación de la envolvente espectral, síntesis y actualizaciones**

5 Para sintetizar la voz decodificada, deben obtenerse los parámetros del filtro LP.

En caso de que la trama futuro no está disponible, la envolvente espectral es desplazada gradualmente a la envolvente estimada del ruido ambiente. Aquí, se usa la representación LSF de los parámetros LP:

$$l^1(j) = \alpha l^0(j) + (1 - \alpha) l_n(j), \quad j=0, \dots, p-1 \quad (33)$$

10 En la ecuación (33),  $l^1(j)$  es el valor de la  $j^{\text{a}}$  LSF de la trama actual,  $l^0(j)$  es el valor de la  $j^{\text{a}}$  LSF de la trama anterior,  $l^2(j)$  es el valor de la  $j^{\text{a}}$  LSF de la envolvente del ruido de confort estimada y p es el orden del filtro LP (obsérvese que las LSF están en el dominio de la frecuencia). De manera alternativa, los parámetros LSF de la trama borrada pueden ser establecidos simplemente a valores iguales a los parámetros de la última trama ( $l^1(j) = l^0(i)$ ).

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

20 En caso de que la trama futura está disponible, los parámetros de filtro LP por subtrama se obtienen interpolando los valores LSP en las tramas futura y anterior. Pueden usarse diversos procedimientos para encontrar los parámetros interpolados. En un procedimiento, los parámetros LSP se encuentran para toda la trama con la relación:

$$LSP^{(1)} = 0,4 LSP^{(0)} + 0,6 LSF^{(2)} \quad (34)$$

en la que  $LSP^{(1)}$  son los LSPs estimados de la trama borrada,  $LSP^{(0)}$  son los LSPs en la trama pasada y  $LSF^{(2)}$  son los LSPs en la trama futura.

25 Como un ejemplo no limitativo, los parámetros LSP son transmitidos dos veces por cada trama de 20 ms (centrados en las subtramas segunda y cuarta). De esta manera,  $LSP^{(0)}$  se centra en la cuarta subtrama de la trama pasada y  $LSP^{(2)}$  se centra en la segunda subtrama de la trama futura. De esta manera, los parámetros LSP interpolados pueden ser encontrados para cada subtrama en la trama borrada como:

$$LSP^{(1,i)} = ((5-i)LSP^{(0)} + (i+1) LSF^{(2)}) / 6, \quad i=0,\dots,3, \quad (35)$$

30

en la que i es el índice de subtrama. Los LSP están en el dominio coseno (-1 a 1).

Debido a que el cuantificador de ganancia de innovación y el cuantificador LSF usan una predicción, su memoria no estará actualizada después de recuperar el funcionamiento normal. Para reducir este efecto, la memoria de los cuantificadores se estiman y actualizan al final de cada trama borrada.

35 **Recuperación del funcionamiento normal después de un borrado**

40 El problema de la recuperación después de un bloque de tramas borradas es debido, básicamente, a la fuerte predicción usada prácticamente en todos los codificadores de voz modernos. En particular, los codificadores de voz de tipo CELP consiguen su alta relación de señal-a-ruido para la voz sonora debido al hecho de que los mismos usan la señal de excitación pasada para codificar la excitación de la trama actual (predicción de frecuencia fundamental o de largo plazo). Además, la mayoría de los cuantificadores (cuantificadores LP, cuantificadores de ganancia, etc.) hacen uso de una predicción.

**Construcción de inicio artificial**

45 La situación más complicada con relación al uso de la predicción a largo plazo en los codificadores CELP es cuando se pierde un inicio de voz. El inicio perdido significa que el inicio de voz sonora ocurrió en algún momento durante el bloque borrado. En este caso, la última trama buena recibida era sorda y, de esta manera, no se encuentra ninguna excitación periódica en la memoria intermedia de excitación. Sin embargo, 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 ha sido codificada usando esta excitación pasada periódica. Debido a que esta parte periódica de la excitación está completamente ausente en el decodificador, pueden necesitarse hasta varias tramas para recuperarse de esta pérdida.

5 Si se pierde una trama de INICIO (es decir, llega una trama SONORA buena después de un borrado, pero la última trama buena antes del borrado era SORDA, tal como se muestra en la Figura 13), se usa una técnica especial para reconstruir artificialmente el inicio perdido y desencadenar la síntesis de voz. En esta realización ilustrativa, la posición del último pulso glótico en la trama ocultada puede estar disponible a partir de la trama futura (la trama futura no está perdida y la información de fase relacionada con la trama anterior es recibida en la trama futura). En este caso, la ocultación de la trama borrada se realiza como de costumbre. Sin embargo, el último pulso glótico de la trama borrada es reconstruido artificialmente en base a la información de posición y de signo disponible en la trama futura. Esta información consiste en la posición del pulso máximo desde el final de la trama y su signo. De esta manera, el último pulso glótico en la trama borrada es construido artificialmente como un pulso filtrado con un filtro de paso bajo. En esta realización ilustrativa, si el signo del pulso es positivo, el filtro de paso bajo usado es un filtro FIR simple, de fase lineal, con la respuesta a impulso  $H_{bajo} = \{-0,0125, 0,109, 0,7813, 0,109, -0,0125\}$ . Si el signo del pulso es negativo, el filtro de paso bajo usado es un filtro FIR de fase lineal con la respuesta a impulso  $h_{bajo} = \{0,0125, -0,109, -0,7813, -0,109, 0,0125\}$ .

20 El periodo fundamental considerado es la última subtrama de la trama ocultada. El pulso sometido a filtrado de paso bajo se realiza colocando la respuesta a impulso del filtro de paso bajo en la memoria de la memoria intermedia de excitación adaptativa (previamente inicializada a cero). El pulso glótico sometido a filtrado de paso bajo (respuesta a impulso del filtro de paso bajo) se centrará en la posición  $P_{ultimo}$  decodificada (transmitida en el flujo de bits de la trama futura). En la decodificación de la siguiente trama buena, se reanuda la decodificación CELP normal. La colocación del pulso glótico sometido a filtrado de paso bajo en la posición apropiada al final de la trama ocultada mejora significativamente el rendimiento de las tramas buenas consecutivas y acelera la convergencia del decodificador a los estados reales del decodificador.

A continuación, la energía de la parte periódica de la excitación de inicio artificial es escalada por la ganancia correspondiente a la energía cuantificada y transmitida para la ocultación FER y es dividida por la ganancia del filtro de síntesis LP. La ganancia del filtro de síntesis LP se calcula como:

$$30 \quad g_{LP} = \sqrt{\sum_{i=0}^{40} h^2(i)} \quad (36)$$

en la que  $h(i)$  es la respuesta a impulso del filtro de síntesis LP. Por último, la ganancia de inicio artificial es reducida multiplicando la parte periódica por 0,96.

35 El filtro LP para la síntesis de voz de salida no se interpola en el caso de una construcción de inicio artificial. En su lugar, los parámetros LP recibidos se usan para la síntesis de toda la trama.

### Control de energía

40 Una tarea en la recuperación después de un bloque de tramas borradas es controlar, de manera apropiada, la energía de la señal de voz sintetizada. El control de energía de síntesis es necesario debido a la fuerte predicción usada, normalmente, en los codificadores de voz modernos. El control de energía se realiza también cuando se produce un bloque de tramas borradas durante un segmento sonoro. Cuando un borrado de trama llega después de una trama sonora, típicamente, se usa la excitación de la última trama buena durante la ocultación con alguna estrategia de atenuación. Cuando un nuevo filtro LP llega con la primera trama buena después del borrado, puede haber una discrepancia entre la energía de excitación y la ganancia del nuevo filtro de síntesis LP. El nuevo filtro de síntesis puede producir una señal de síntesis con una energía muy diferente de la energía de la última trama borrada sintetizada y también de la energía de señal original.

45 El control de energía durante la primera trama buena después de una trama borrada puede resumirse de la manera siguiente. La señal sintetizada es escalada de manera que su energía sea similar a la energía de la señal de voz sintetizada al final de la última trama borrada en el inicio de la primera trama buena y converge a la energía transmitida hacia el final de la trama para prevenir un aumento de energía demasiado alto.

50 El control de la energía se realiza en el dominio de señal de voz sintetizada. Incluso si la energía es controlada en el dominio de voz, la señal de excitación debe ser escalada, ya que sirve como memoria de predicción a largo plazo para las tramas siguientes. A continuación, la síntesis vuelve a realizarse para suavizar las transiciones. Denótese

mediante  $g_0$  la ganancia usada para escalar la 1ª muestra en la trama actual y  $g_1$  la ganancia usada al final de la trama. Entonces, la señal de excitación es escalada de la manera siguiente:

$$u_s(i) = g_{AGC}(i) \cdot u(i), \quad i=0, \dots, L-1 \quad (37)$$

5 en la que  $u_s(i)$  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 que comienza en  $g_0$  y converge exponencialmente a  $g_1$ :

$$g_{AGC}(i) = f_{AGC} g_{AGC}(i-1) + (1-f_{AGC}) g_1 \quad i=0, \dots, L-1 \quad (38)$$

10 con la inicialización de  $g_{AGC}(-1) = g_0$ , en la que  $f_{AGC}$  es el factor de atenuación establecido, en esta implementación, al valor de 0,98. Este valor se ha encontrado experimentalmente como un compromiso entre tener una transición suave desde la trama anterior (borrada), por un lado, y escalar el último periodo fundamental de la trama actual tanto como sea posible al valor correcto (transmitido), por otro lado. Esto se realiza debido a que el valor de energía transmitido es estimado de manera síncrona con la frecuencia fundamental al final de la trama. Las ganancias  $g_0$  y  $g_1$  se definen como:

$$15 \quad g_0 = \sqrt{E_{-1} / E_0} \quad (39)$$

$$g_1 = \sqrt{E_q / E_1} \quad (40)$$

20 en las que  $E_{-1}$  es la energía calculada al final de la trama anterior (borrada),  $E_0$  es la energía en el inicio de la trama actual (recuperada),  $E_1$  es la energía al final de la trama actual y  $E_q$  es la información de energía transmitida cuantificada al final de la trama actual, calculada en el codificador a partir de las Ecuaciones (20; 21).  $E_{-1}$  y  $E_1$  se calculan de manera similar, con la excepción de que se calculan sobre la señal  $s'$  de voz sintetizada.  $E_{-1}$  se calcula de manera síncrona con respecto a la frecuencia fundamental usando el periodo fundamental de ocultación  $T_c$  y  $E_1$  usa la frecuencia fundamental  $T_3$  redondeada de la última subtrama.  $E_0$  se calcula de manera similar, usando el valor  $T_c$  de frecuencia fundamental redondeado de la primera subtrama, modificándose las ecuaciones (20, 21) a:

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

para tramas SONORAS y de INICIO.  $t_E$  es igual al retardo de frecuencia fundamental redondeado o dos veces esa longitud si la frecuencia fundamental es más corta que 64 muestras. Para otras tramas,

$$30 \quad E = \frac{1}{t_E} \sum_{i=0}^{t_E} s'^2(i)$$

con  $t_E$  igual a la mitad de la longitud de la trama. Las ganancias  $g_0$  y  $g_1$  se limitan además a un valor máximo permitido, para prevenir una energía fuerte. Este valor se ha establecido a 1,2 en la presente implementación ilustrativa.

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

45 Si  $E_q$  no puede ser transmitida,  $E_q$  se establece a  $E_r$ . Sin embargo, si el borrado ocurre 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 discrepancia entre la energía de la señal de excitación y la ganancia del filtro LP, indicadas anteriormente. Una situación particularmente peligrosa se produce cuando la ganancia del filtro LP de una primera trama no borrada recibida después del borrado de trama es mayor que la ganancia del filtro LP de una última trama borrada



durante ese borrado de trama. En ese caso particular, la energía de la señal de excitación del filtro LP producida en el decodificador durante la primera trama recibida no borrada es ajustada a una ganancia del filtro LP de la primera trama no borrada recibida usando la relación siguiente:

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

5

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

10

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

En el caso de una primera trama buena después de un borrado clasificado como INICIO, se previene que la ganancia  $g_0$  sea mayor que  $g_1$ . Esta precaución se toma para prevenir que un ajuste positivo de ganancia en el inicio de la trama (que, probablemente, todavía es al menos parcialmente sorda) amplifique el inicio sonoro (al final de la trama).

15

Por último, durante una transición desde sonora a sorda (es decir, que la última trama es clasificada como TRANSICIÓN SONORA, SONORA o INICIO y la trama actual es clasificada como SORDA) o durante una transición desde un periodo de voz inactivo a un periodo de voz activo (donde la última trama buena recibida está codificada como ruido de confort y la trama actual está codificada como voz activa),  $g_0$  se establece a  $g_1$ .

20

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

25

#### **Aplicación de la ocultación descrita en un códec integrado con una capa de núcleo de banda ancha**

Tal como se ha indicado anteriormente, la realización ilustrativa descrita anteriormente de la presente invención ha sido usada también en un algoritmo candidato para la normalización de un códec integrado de tasa de bits variable por la UIT-T. En el algoritmo candidato, la capa de núcleo se basa en una técnica de codificación de banda ancha similar a AMR-WB (Recomendación UIT-T G.722.2). La capa de núcleo funciona a 8 kbit/s y codifica un ancho de banda de hasta 6.400 Hz con una frecuencia de muestreo interna de 12,8 kHz (similar a AMR-WB). Una segunda capa CELP de 4 kbit/s es usada para aumentar la tasa de bits hasta 12 kbit/s. Entonces, se usa MDCT para obtener las capas superiores de 16 a 32 kbit/s.

30

La ocultación es similar al procedimiento descrito anteriormente, con pocas diferencias debidas principalmente a la diferente velocidad de muestreo de la capa de núcleo. El tamaño de la trama es de 256 muestras a una frecuencia de muestreo de 12,8 kHz y el tamaño de la subtrama es de 64 muestras.

35

La información de fase se codifica con 8 bits, donde el signo se codifica con 1 bit y la posición se codifica con 7 bits de la manera siguiente.

La precisión usada para codificar la posición del primer pulso glótico depende del valor  $T_0$  de frecuencia fundamental de bucle cerrado para la primera subtrama en la trama futura. Cuando  $T_0$  es menor de 128, la posición del último pulso glótico con relación al final de la trama es codificada directamente con una precisión de una muestra. Cuando  $T_0 \geq 128$ , la posición del último pulso glótico con relación al final de la trama es codificada con una precisión de dos muestras usando una división simple de enteros, es decir,  $\tau/2$ . El procedimiento inverso se realiza en el decodificador. Si  $T_0 < 128$ , la posición cuantificada recibida es usada sin modificaciones. Si  $T_0 \geq 128$ , la posición cuantificada recibida es multiplicada por 2 e incrementada en 1.

40

Los parámetros de recuperación de ocultación consisten en información de fase de 8 bits, información de clasificación de 2 bits e información de energía de 6 bits. Estos parámetros son transmitidos en la tercera capa a 16 kbit/s.

Aunque la presente invención ha sido descrita en la descripción anterior con relación a una realización no restrictiva ilustrativa de la misma, esta realización puede ser modificada a voluntad, dentro del alcance de las reivindicaciones

50

adjuntas, sin apartarse del alcance de la presente invención.

**Referencias**

[1] Milan Jelinek y Philippe Goumay. solicitud de patente PCT WO03102921A1, "A method and device for efficient frame erasure concealment in linear predictive based speech codecs".



## REIVINDICACIONES

1. Un procedimiento para ocultar borrados de tramas causados por las tramas de una señal de sonido codificada borradas durante la transmisión desde un codificador (700) a un decodificador (300) y para recuperar el decodificador (300) después de los borrados de trama, comprendiendo el procedimiento:

5 en el codificador (700), la determinación (707) de los parámetros de ocultación/recuperación, incluyendo al menos información de fase relacionada con las tramas de la señal de sonido codificada, en el que la información de fase comprende una posición de un pulso ( $\tau_q$ ) glótico en cada trama de la señal de sonido codificada, determinada midiendo (707) el pulso ( $\tau_q$ ) glótico como un pulso de amplitud máxima en un ciclo de frecuencia fundamental predeterminado de la trama de la señal de sonido codificada y determinando (707) la posición del pulso de amplitud máxima;

10 transmitir (213) al decodificador (300) los parámetros de ocultación/recuperación determinados en el codificador (700); y

15 en el decodificador (300), realizar una ocultación de borrado de trama en respuesta a los parámetros de ocultación/recuperación recibidos, en el que la ocultación de borrado de tramas comprende resincronizar (900), en respuesta a la información de fase recibida, las tramas cuyo borrado se oculta con las tramas correspondientes de la señal de sonido codificada en el codificador (700);

caracterizado por que esa resincronización de una trama cuyo borrado se oculta con una trama correspondiente de la señal de sonido codificada comprende:

decodificar (910) la posición del pulso ( $\tau_q$ ) glótico de dicha trama de la señal de sonido codificada;

20 determinar (912), en la trama cuyo borrado se oculta, una posición de un pulso de amplitud máxima más cercano a la posición de dicho pulso ( $\tau_q$ ) glótico de dicha trama correspondiente de dicha señal de sonido codificada; y

alinear (920) la posición del pulso de amplitud máxima en la trama cuyo borrado se oculta con la posición del pulso ( $\tau_q$ ) glótico de la trama correspondiente de la señal de sonido codificada.

25 2. Procedimiento según se define en la reivindicación 1, en el que la determinación de los parámetros de ocultación/recuperación comprenden determinar (707), como la información de fase, una posición y un signo de un último pulso ( $\tau_q$ ) glótico en cada trama de la señal de sonido codificada.

3. Procedimiento según se define en la reivindicación 1, que comprende además cuantificar la posición del pulso glótico antes de transmitir la posición del pulso glótico al decodificador.

30 4. Procedimiento según se define en la reivindicación 2, que comprende además cuantificar (707) la posición y el signo del último pulso ( $\tau_q$ ) glótico antes de transmitir (213) la posición y el signo del último pulso glótico al decodificador (300).

5. Procedimiento según se define en la reivindicación 1, que comprende además codificar (707) una posición cuantificada del pulso glótico en un trama futura de la señal de sonido codificada.

35 6. Procedimiento según se define en la reivindicación 1, que comprende además determinar (707), como la información de fase, un signo del pulso ( $\tau_q$ ) glótico midiendo un signo del pulso de amplitud máxima.

7. Procedimiento según se define en la reivindicación 2, en el que la determinación de la posición del último pulso glótico comprende:

40 medir (707) el último pulso ( $\tau_q$ ) glótico como un pulso de amplitud máxima en cada trama de la señal de sonido codificada; y

determinar (707) la posición del pulso de amplitud máxima.

8. Procedimiento según se define en la reivindicación 7, en el que la determinación del signo del último pulso ( $\tau_q$ ) glótico comprende:

medir (707) una señal del pulso de amplitud máxima.

45 9. Procedimiento según se define en la reivindicación 8, en el que la resincronización (900) de una trama cuyo borrado se oculta con una trama correspondiente de la señal de sonido codificada comprende:

decodificar (910) la posición y el signo del último pulso ( $\tau_q$ ) glótico de dicha trama correspondiente de la señal de sonido codificada;

5 determinar (912), en la trama cuyo borrado se oculta una posición de un pulso de amplitud máxima que tiene un signo similar al signo del último pulso ( $\tau_q$ ) glótico de la trama correspondiente de la señal de sonido codificada, más cercana a la posición de dicho último pulso ( $\tau_q$ ) glótico de dicha trama correspondiente de dicha señal de sonido codificada; y

alinear (920) la posición del pulso de amplitud máxima en la trama cuyo borrado se oculta con la posición del último pulso ( $\tau_q$ ) glótico de la trama correspondiente de la señal de sonido codificada.

10 10. Procedimiento según se define en la reivindicación 1, en el que la alineación de la posición del pulso de amplitud máxima en la trama cuyo borrado se oculta con la posición del pulso ( $\tau_q$ ) glótico en la trama correspondiente de la señal de sonido codificada comprende:

determinar (916) un desplazamiento entre la posición del pulso de amplitud máxima en la trama cuyo borrado se oculta y la posición del pulso ( $\tau_q$ ) glótico en la trama correspondiente de la señal de sonido codificada; e

15 insertar/eliminar (1008), en la trama cuyo borrado se oculta, un número de muestras correspondientes al desplazamiento determinado.

11. Procedimiento según se define en la reivindicación 10, en el que la inserción/eliminación del número de muestras comprende:

determinar (1002; 1004) al menos una región de energía mínima en la trama cuyo borrado se oculta; y

20 distribuir (1006) el número de muestras a insertar/eliminar alrededor de la al menos una región de energía mínima.

12. Procedimiento según se define en la reivindicación 11, en el que la distribución del número de muestras a insertar/eliminar alrededor de la al menos una región de energía mínima comprende distribuir (1006) el número de muestras alrededor de la al menos una región de energía mínima usando la relación siguiente:

25 
$$R(i) = \text{redondear} \left( \frac{(i+1)^2}{2} f - \sum_{k=0}^{i-1} R(k) \right) \text{ para } i=0, \dots, N_{\min}-1 \text{ y } k=0, \dots, i-1 \text{ y } N_{\min} > 1$$

en la que

$$f = \frac{2|T_e|}{N_{\min}^2}$$

30

$N_{\min}$  es el número de regiones de energía mínima, y  $T_e$  es el desplazamiento entre la posición del pulso de amplitud máxima en la trama cuyo borrado se oculta y la posición del pulso ( $\tau_q$ ) glótico en la trama correspondiente de la señal de sonido codificada.

35 13. Procedimiento según se define en la reivindicación 12, en el que  $R(i)$  está en orden creciente, de manera que las muestras se añaden/eliminan (1008) principalmente hacia un final de la trama cuyo borrado se oculta.

14. Procedimiento según se define en la reivindicación 1, en el que la realización de la ocultación de borrado de tramas en respuesta a los parámetros de ocultación/recuperación recibidos comprende, para las tramas borradas sonoras:

40 construir una parte periódica de una señal de excitación en la trama cuyo borrado se oculta en respuesta a los parámetros de ocultación/recuperación recibidos; y

construir una parte innovadora aleatoria de la señal de excitación mediante la generación aleatoria de una señal innovadora, no periódica.

45 15. Procedimiento según se define en la reivindicación 1, en el que la realización de la ocultación del borrado de tramas en respuesta a los parámetros de ocultación/recuperación recibidos comprende, para las tramas borradas sordas, construir una parte innovadora aleatoria de una señal de excitación mediante la generación aleatoria de una

señal innovadora, no periódica.

16. Procedimiento según se define en la reivindicación 1, en el que los parámetros de ocultación/recuperación incluyen además la clasificación (705) de las señales.

5 17. Procedimiento según se define en la reivindicación 16, en el que la clasificación de la señal comprende clasificar (705) tramas sucesivas de la señal de sonido codificada como sorda, transición sorda, transición sonora, sonora o inicio.

18. Procedimiento según se define en la reivindicación 17, en el que la clasificación de una trama perdida es estimada en base a la clasificación de una trama futura y una última trama buena recibida.

10 19. Procedimiento según se define en la reivindicación 18, en el que la clasificación de la trama perdida se establece como sonora si la trama futura es sonora y la última trama buena recibida es de inicio.

20. Procedimiento según se define en la reivindicación 19, en el que la clasificación de la trama perdida se establece como transición sorda si la trama futura es sorda y la última trama buena recibida es sonora.

21. Procedimiento según se define en la reivindicación 1, en el que:

la señal de sonido es una señal de voz;

15 la determinación, en el codificador (700), de los parámetros de ocultación/recuperación incluye determinar (705; 707) la información de fase y clasificar la señal de las tramas sucesivas de la señal de sonido codificada;

20 la realización de la ocultación de borrado de tramas en respuesta a los parámetros de ocultación/recuperación comprende, cuando se pierde un trama de inicio, lo que se indica por la presencia de una trama sonora que sigue a un borrado de trama y una trama sorda antes del borrado de trama, reconstruir artificialmente la trama de inicio perdida; y

volver a sincronizar (900) la trama de inicio perdida, cuyo borrado se oculta, en respuesta a la información de fase con la trama de inicio correspondiente de la señal de sonido codificada.

25 22. Procedimiento según se define en la reivindicación 21, en el que la reconstrucción de manera artificial de la trama de inicio perdida comprende reconstruir, de manera artificial, un último pulso ( $\tau_q$ ) glótico en la trama de inicio perdida como un pulso con un filtrado de paso bajo.

23. Procedimiento según se define en la reivindicación 21, que comprende además escalar la trama de inicio perdida reconstruida por una ganancia.

30 24. Un procedimiento para ocultar borrados de tramas causados por tramas de una señal de sonido codificada borradas durante la transmisión desde un codificador (700) a un decodificador (300) y para la recuperación del decodificador (300) después de los borrados de tramas, en el que el procedimiento comprende, en el decodificador (300):

estimar una información de fase de cada trama de la señal de sonido codificada que ha sido borrada durante la transmisión desde el codificador (700) al decodificador (300); y

35 realizar la ocultación de borrado de tramas en respuesta a la información de fase estimada, en el que la ocultación de borrado de tramas comprende volver a sincronizar (900), en respuesta a la información de fase estimada, cada trama cuyo borrado se oculta con un trama correspondiente de la señal de sonido codificada en el codificador (700);

caracterizado por que

40 la información de fase estimada es una posición estimada de un pulso ( $\tau_q$ ) glótico de cada trama de la señal de sonido codificada que ha sido borrada;

la estimación de la posición del pulso glótico de cada trama de la señal de sonido codificada que ha sido borrada comprende un pulso ( $\tau_q$ ) glótico de un valor de frecuencia fundamental pasado;

45 la resincronización una trama cuyo borrado se oculta con la trama correspondiente de la señal de sonido codificada comprende determinar (912) un pulso de amplitud máxima en la trama cuyo borrado se oculta, y alinear (920) el pulso de amplitud máxima en la trama cuyo borrado se oculta con el pulso ( $\tau_q$ ) glótico estimado.

25. Procedimiento según se define en la reivindicación 24, en el que la estimación de la posición del último pulso glótico de cada trama de la señal de sonido codificada que ha sido borrada comprende:

interpolarse el pulso glótico estimado con el valor de frecuencia fundamental pasado para determinar los retardos de frecuencia fundamental estimados.

5 26. Procedimiento según se define en la reivindicación 25, en el que la alineación de la posición del pulso de amplitud máxima en la trama cuyo borrado se oculta con la posición estimada del pulso ( $\tau_q$ ) glótico comprende:

calcular los ciclos de frecuencia fundamental en la trama cuyo borrado se oculta;

determinar un desplazamiento entre los retardos de frecuencia fundamental estimados y los ciclos de frecuencia fundamental en la trama cuyo borrado se oculta, e

10 insertar/eliminar (1008) una serie de muestras correspondientes al desplazamiento determinado en la trama cuyo borrado se oculta.

27. Procedimiento según se define en la reivindicación 26, en el que la inserción/eliminación del número de muestras comprende:

determinar (1002; 1004) al menos una región de energía mínima en la trama cuyo borrado se oculta; y

15 distribuir (1006) el número de muestras a insertar/eliminar alrededor de la al menos una región de energía mínima.

28. Procedimiento según se define en la reivindicación 27, en el que la distribución del número de muestras a insertar/eliminar alrededor de la al menos una región de energía mínima comprende distribuir (1006) el número de muestras alrededor de la al menos una región de energía mínima usando la relación siguiente:

20 
$$R(i) = \text{redondear} \left( \frac{(i+1)^2}{2} f - \sum_{k=0}^{i-1} R(k) \right) \text{ para } i=0, \dots, N_{\min}-1 \text{ y } k=0, \dots, i-1 \text{ y } N_{\min} > 1$$

en la que

$$f = \frac{2|T_e|}{N_{\min}^2},$$

25

$N_{\min}$  es el número de regiones de energía mínima y  $T_e$  es el desplazamiento entre los retardos de frecuencia fundamental estimada y los ciclos de frecuencia fundamental en la trama cuyo borrado se oculta.

29. Procedimiento según se define en la reivindicación 28, en el que  $R(i)$  está en orden creciente, de manera que las muestras se insertan/eliminadas (1008), en su mayoría, hacia el final de la trama cuyo borrado se oculta.

30 30. Procedimiento según se define en la reivindicación 24, que comprende atenuar (924) una ganancia de cada trama cuyo borrado se oculta, de una manera lineal, desde el inicio hasta el final de la trama cuyo borrado se oculta.

35 31. Procedimiento según se define en la reivindicación 30, en el que la ganancia de cada trama cuyo borrado se oculta es atenuada (924) hasta que se alcanza el valor  $\alpha$ , en el que  $\alpha$  es un factor para controlar una velocidad de convergencia de la recuperación del decodificador (300) después del borrado de tramas.

32. Procedimiento según se define en la reivindicación 31, en el que el factor  $\alpha$  depende de la estabilidad de un filtro LP para tramas sordas.

33. Procedimiento según se define en la reivindicación 32, en el que el factor  $\alpha$  tiene en cuenta además una evolución de la energía de los segmentos sonoros.

40 34. Un dispositivo para ocultar borrados de tramas causados por tramas de una señal sonora codificada borradas durante la transmisión desde un codificador (700) a un decodificador (300) y para recuperar el decodificador (300) después de los borrados de tramas, en el que el dispositivo comprende:

en el codificador (700), medios (707) para determinar los parámetros de ocultación/recuperación, incluyendo al menos información de fase relacionada con tramas de la señal de sonido codificada, en el que la información

de fase comprende una posición de un pulso ( $\tau_q$ ) glótico en cada trama de la señal de sonido codificada, determinado por medios para medir (707) el pulso ( $\tau_q$ ) glótico como un pulso de amplitud máxima en un ciclo de frecuencia fundamental predeterminado de la trama de la señal de sonido codificada y por medios para determinar (707) la posición del pulso de amplitud máxima;

5 medios (213) para transmitir al decodificador (300) los parámetros de ocultación/recuperación determinados en el codificador (700); y

10 en el decodificador (300), medios para realizar la ocultación del borrado de tramas en respuesta a los parámetros de ocultación/recuperación recibidos, en el que los medios para realizar la ocultación del borrado de tramas comprenden medios (900) para volver a sincronizar, en respuesta a la información de fase recibida, las tramas de cuyo borrado se oculta con tramas correspondientes de la señal de sonido codificada en el codificador (700).

caracterizado por que los medios (900) para resincronizar las tramas cuyo borrado se oculta con tramas correspondientes de la señal de sonido codificada en el codificador (700) comprenden:

15 medios (912) para determinar en cada trama cuyo borrado se oculta, una posición de un pulso de amplitud máxima más cercano a la posición del pulso ( $\tau_q$ ) glótico en una trama correspondiente de la señal de sonido codificada; y

medios (920) para alinear la posición del pulso de amplitud máxima en la trama cuyo borrado se oculta con la posición del pulso ( $\tau_q$ ) glótico en la trama correspondiente de la señal de sonido codificada.

20 35. Dispositivo según se define en la reivindicación 34, en el que los medios para determinar los parámetros de ocultación/recuperación comprenden además medios (707) para determinar, como la información de fase, una posición y un signo de un último pulso ( $\tau_q$ ) glótico en cada trama de la señal de sonido codificada.

36. Dispositivo según se define en la reivindicación 34, que comprende además medios (707) para cuantificar la posición del pulso glótico antes de la transmisión de la posición del pulso glótico al decodificador, a través de los medios (213) de transmisión.

25 37. Dispositivo según se define en la reivindicación 35, que comprende además medios (707) para cuantificar la posición y el signo del último pulso ( $\tau_q$ ) glótico antes de la transmisión de la posición y el signo del último pulso ( $\tau_q$ ) glótico al decodificador (300), a través de los medios (213) de transmisión.

38. Dispositivo según se define en la reivindicación 36, que comprende además un codificador de la posición cuantificada del pulso glótico en una trama futura de la señal de sonido codificada.

30 39. Dispositivo según se define en la reivindicación 35, en el que los medios (707) para determinar la posición y el signo del último pulso glótico determinan, como la posición y el signo del último pulso ( $\tau_q$ ) glótico, una posición y un signo de un pulso de amplitud máxima en cada trama de la señal de sonido codificada.

35 40. Dispositivo según se define en la reivindicación 34, en el que los medios (707) para determinar la posición y el signo del pulso glótico determinan, como información de fase, un signo del pulso ( $\tau_q$ ) glótico, como un signo del pulso de amplitud máxima.

41. Dispositivo según se define en la reivindicación 34, en el que los medios (900) de resincronización comprenden:

medios para determinar (916) un desplazamiento entre la posición del pulso de amplitud máxima en cada trama cuyo borrado se oculta y la posición del pulso ( $\tau_q$ ) glótico en la trama correspondiente de la señal de sonido codificada; y

40 en el que los medios para alinear la posición del pulso de amplitud máxima en la trama cuyo borrado se oculta insertan/eliminan (1008) un número de muestras correspondiente al desplazamiento determinado en cada trama cuyo borrado se oculta.

42. Dispositivo según se define en la reivindicación 35, en el que los medios (900) de resincronización:

45 determinan (912), en cada trama cuyo borrado se oculta, una posición de un pulso de amplitud máxima que tiene un signo similar al signo del pulso ( $\tau_q$ ) glótico más cercano a la posición del último pulso glótico en una trama correspondiente de la señal de sonido codificada;

determinan (916) un desplazamiento entre la posición del pulso de amplitud máxima en cada trama cuyo borrado se oculta y la posición del último pulso ( $\tau_q$ ) glótico en la trama correspondiente de la señal de sonido

codificada; y

insertan/eliminan (1008) un número de muestras correspondientes al desplazamiento determinado en cada trama cuyo borrado se oculta para alinear la posición del pulso de amplitud máxima en la trama cuyo borrado se oculta con la posición del último pulso ( $\tau_q$ ) glótico en la trama correspondiente de la señal de sonido codificada.

5

43. Dispositivo según se define en la reivindicación 41, en el que, además, los medios (900) de resincronización:

determinan (1002; 1004) al menos una región de energía mínima en cada trama cuyo borrado se oculta usando una ventana deslizante; y

distribuyen (1006) el número de muestras a insertar/eliminar alrededor de al menos una región de energía mínima.

10

44. Dispositivo según se define en la reivindicación 43, en el que los medios (900) de resincronización usan la relación siguiente para distribuir (1006) el número de muestras a insertar/eliminar alrededor de la al menos una región de energía mínima:

$$R(i) = \text{redondear} \left( \frac{(i+1)^2}{2} f - \sum_{k=0}^{i-1} R(k) \right) \text{ para } i=0, \dots, N_{\min}-1 \text{ y } k=0, \dots, i-1 \text{ y } N_{\min} > 1$$

15

en la que

$$f = \frac{2|T_e|}{N_{\min}^2},$$

20

$N_{\min}$  es el número de regiones de energía mínima, y  $T_e$  es el desplazamiento entre la posición del pulso de amplitud máxima en la trama cuyo borrado se oculta y la posición del pulso ( $\tau_q$ ) glótico en la trama correspondiente de la señal de sonido codificada.

45. Dispositivo según se define en la reivindicación 44, en el que  $R(i)$  está en orden creciente, de manera que las muestras se insertan/eliminan (1008), en su mayoría, hacia un final de la trama cuyo borrado se oculta.

25

46. Dispositivo según se define en la reivindicación 34, en el que los medios para realizar la ocultación del borrado de tramas, a los que se suministran los parámetros de ocultación/recuperación recibidos, comprenden, para las tramas sonoras borradas:

medios para construir una parte periódica de una señal de excitación en cada trama cuyo borrado se oculta en respuesta a los parámetros de ocultación/recuperación recibidos; y

30

medios para construir una parte innovadora aleatoria, no periódica, de la señal de excitación.

47. Dispositivo según se define en la reivindicación 34, en el que los medios para realizar la ocultación de borrado de tramas, a los que se suministran los parámetros de ocultación/recuperación recibidos, comprenden medios para construir, para tramas borradas sordas, una parte innovadora aleatoria, no periódica, de una señal de excitación.

48. Un dispositivo para ocultar los borrados de tramas causados por las tramas de una señal sonora codificada borradas durante la transmisión desde un codificador (700) a un decodificador (300) y para recuperar el decodificador (300) después de los borrados de tramas, en el que el dispositivo comprende:

35

medios para estimar, en el decodificador (300), una información de fase de cada trama de la señal de sonido codificada que ha sido borrada durante la transmisión desde el codificador (700) al decodificador (300); y

medios para realizar la ocultación del borrado de tramas en respuesta a la información de fase estimada, en el que los medios para realizar la ocultación del borrado de tramas que comprenden medios para resincronizar (900), en respuesta a la información de fase estimada, cada trama cuyo borrado se oculta con una trama correspondiente de la señal de sonido codificada en el codificador (700);

40

caracterizado por que:

los medios para estimar la información de fase comprenden medios para estimar una posición de un pulso ( $\tau_q$ ) glótico en cada trama de la señal de sonido codificada;

los medios para estimar la posición del pulso glótico estiman el pulso glótico a partir de un valor de frecuencia fundamental pasado; y

5 los medios de resincronización comprenden medios para determinar un pulso de amplitud máxima en la trama cuyo borrado se oculta, y medios para alinear el pulso de amplitud máxima en la trama cuyo borrado se oculta con el pulso glótico estimado.

10 49. Dispositivo según se define en la reivindicación 48, en el que los medios para estimar la información de fase estiman, a partir de un valor de frecuencia fundamental pasado, una posición y un signo de un último pulso glótico en cada trama de la señal de sonido codificada, e interpolan el pulso glótico estimado con el valor de frecuencia fundamental pasado para determinar los retardos de frecuencia fundamental estimados.

50. Dispositivo según se define en la reivindicación 49, en el que los medios de resincronización comprenden:

medios para determinar los ciclos de frecuencia fundamental en cada trama cuyo borrado se oculta;

15 medios para determinar un desplazamiento entre los ciclos de frecuencia fundamental en cada trama cuyo borrado se oculta y los retardos de frecuencia fundamental estimados en la trama correspondiente de la señal de sonido codificada; y

20 en el que los medios para alinear la posición del pulso de amplitud máxima en la trama cuyo borrado se oculta insertan/eliminan (1008) un número de muestras correspondientes al desplazamiento determinado en cada trama cuyo borrado se oculta con el fin de alinear el pulso de amplitud máxima en la trama cuyo borrado se oculta con el último pulso glótico estimado.

51. Dispositivo según se define en la reivindicación 50, en el que los medios de resincronización además:

determinan (1002; 1004) al menos una región de energía mínima usando una ventana deslizante; y

distribuyen (1006) el número de muestras alrededor de la al menos una región de energía mínima.

25 52. Dispositivo según se define en la reivindicación 51, en el que los medios de resincronización usan la relación siguiente para distribuir (1006) el número de muestras alrededor de la al menos una región de energía mínima:

$$R(i) = \text{redondear} \left( \frac{(i+1)^2}{2} f - \sum_{k=0}^{i-1} R(k) \right) \text{ para } i=0, \dots, N_{\min}-1 \text{ y } k=0, \dots, i-1 \text{ y } N_{\min} > 1$$

en la que

30 
$$f = \frac{2|T_e|}{N_{\min}^2},$$

35  $N_{\min}$  es el número de regiones de energía mínima, y  $T_e$  es el desplazamiento entre los ciclos de frecuencia fundamental en cada trama cuyo borrado se oculta y los retardos de frecuencia fundamental estimados en la trama correspondiente de la señal de sonido codificada.

53. Dispositivo según se define en la reivindicación 52, en el que  $R(i)$  está en orden creciente, de manera que las muestras se insertan/eliminan (1008), en su mayoría, hacia un final de la trama cuyo borrado se oculta.

40 54. Dispositivo según se define en la reivindicación 49, que comprende además medios (924) para atenuar una ganancia de cada trama cuyo borrado se oculta, de una manera lineal, desde un inicio a un final de la trama cuyo borrado se oculta.

55. Dispositivo según se define en la reivindicación 54, en el que los medios (924) de atenuación atenúan la ganancia de cada trama cuyo borrado se oculta hasta  $\alpha$ , en el que  $\alpha$  es un factor para controlar una velocidad de convergencia de la recuperación del decodificador después del borrado de tramas.

56. Dispositivo según se define en la reivindicación 55, en el que el factor  $\alpha$  depende de la estabilidad de un filtro LP

para tramas sordas.

57. Dispositivo según se define en la reivindicación 56, en el que el factor  $\alpha$  tiene en cuenta además una evolución de la energía de los segmentos sonoros.

5 58. Procedimiento según se define en la reivindicación 1, que comprende, cuando la información de fase no está disponible en el momento de ocultar una trama borrada, actualizar el contenido de un libro de códigos adaptativo del decodificador con la información de fase cuando está disponible antes de decodificar una trama siguiente, no borrada, recibida.

59. Procedimiento según se define en la reivindicación 58, en el que:

10 la actualización del libro de códigos adaptativo comprende resincronizar el pulso glótico en el libro de códigos adaptativo.

60. Dispositivo según se define en la reivindicación 34, en el que el decodificador (300) actualiza, cuando la información de fase no está disponible en el momento de ocultar una trama borrada, el contenido de un libro de códigos adaptativo del decodificador con la información de fase cuando esté disponible antes de decodificar una trama siguiente, no borrada, recibida.

15 61. Dispositivo según se define en la reivindicación 60, en el que:

el decodificador, para actualizar el libro de códigos adaptativo, resincroniza el pulso glótico en el libro de códigos adaptativo.



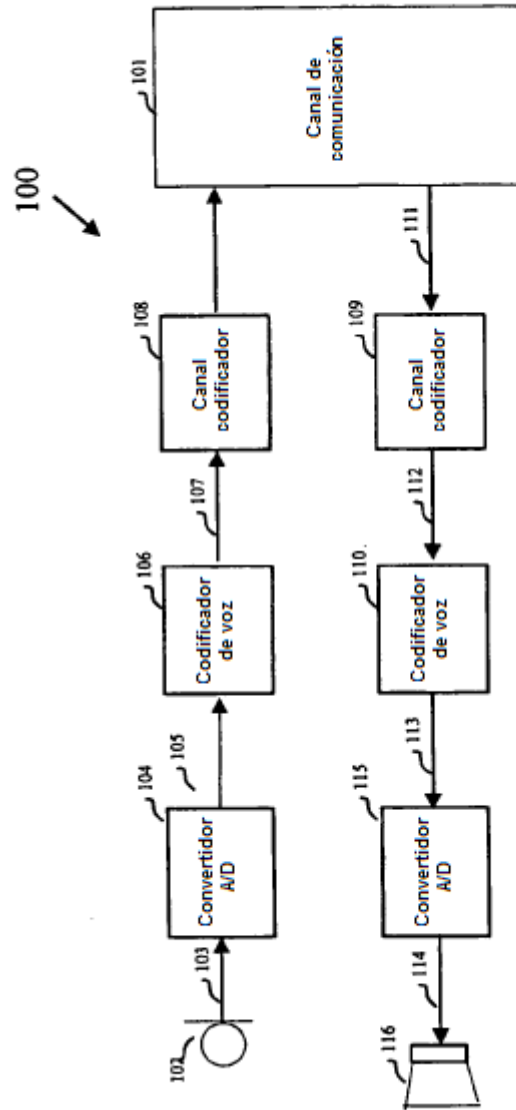


Figura 1

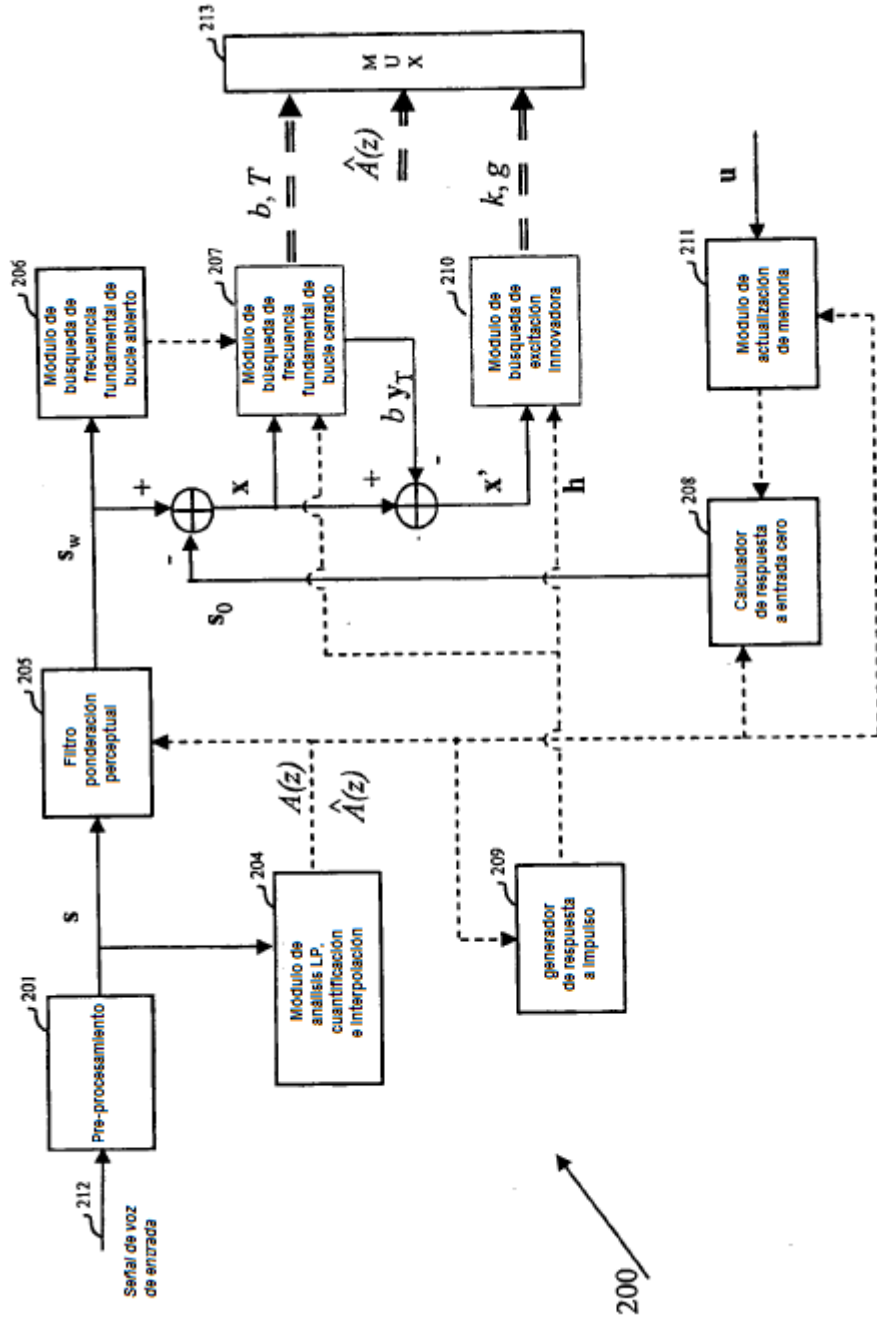


Figura 2

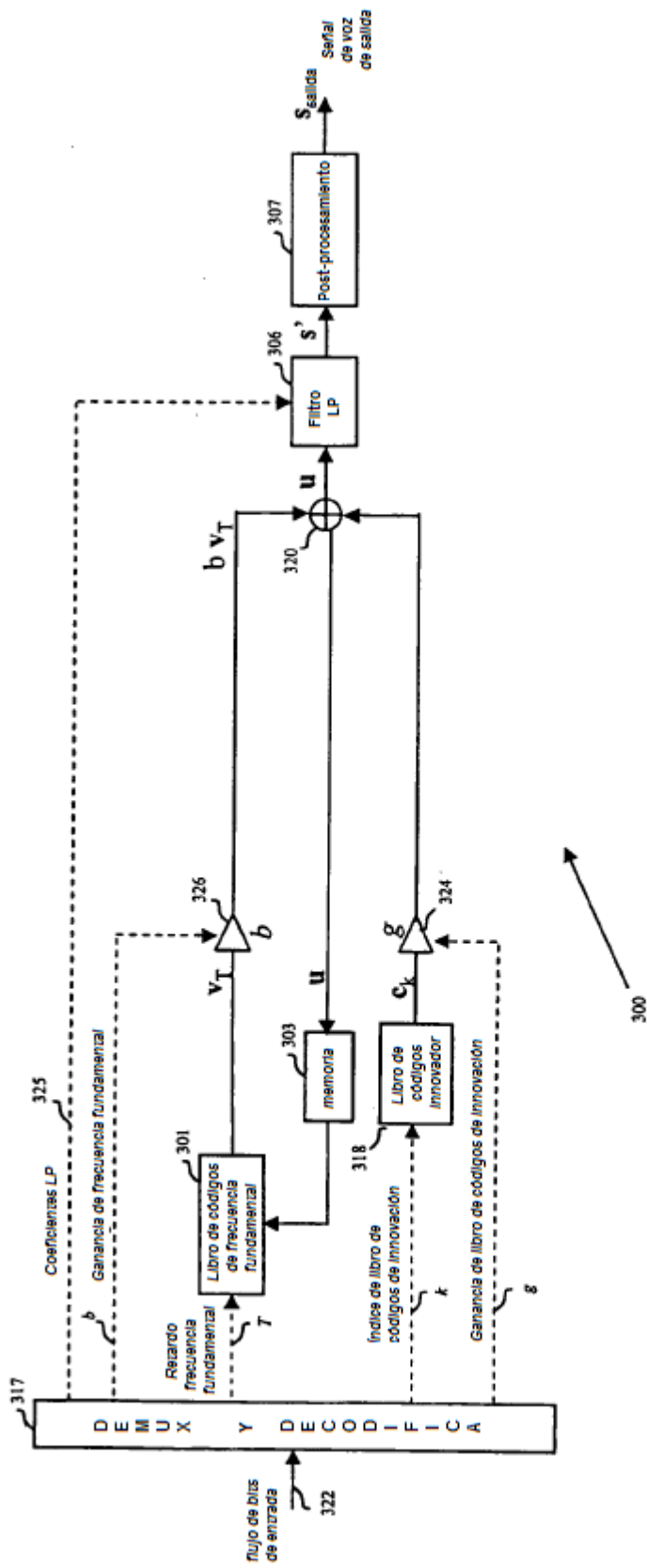


Figura 3

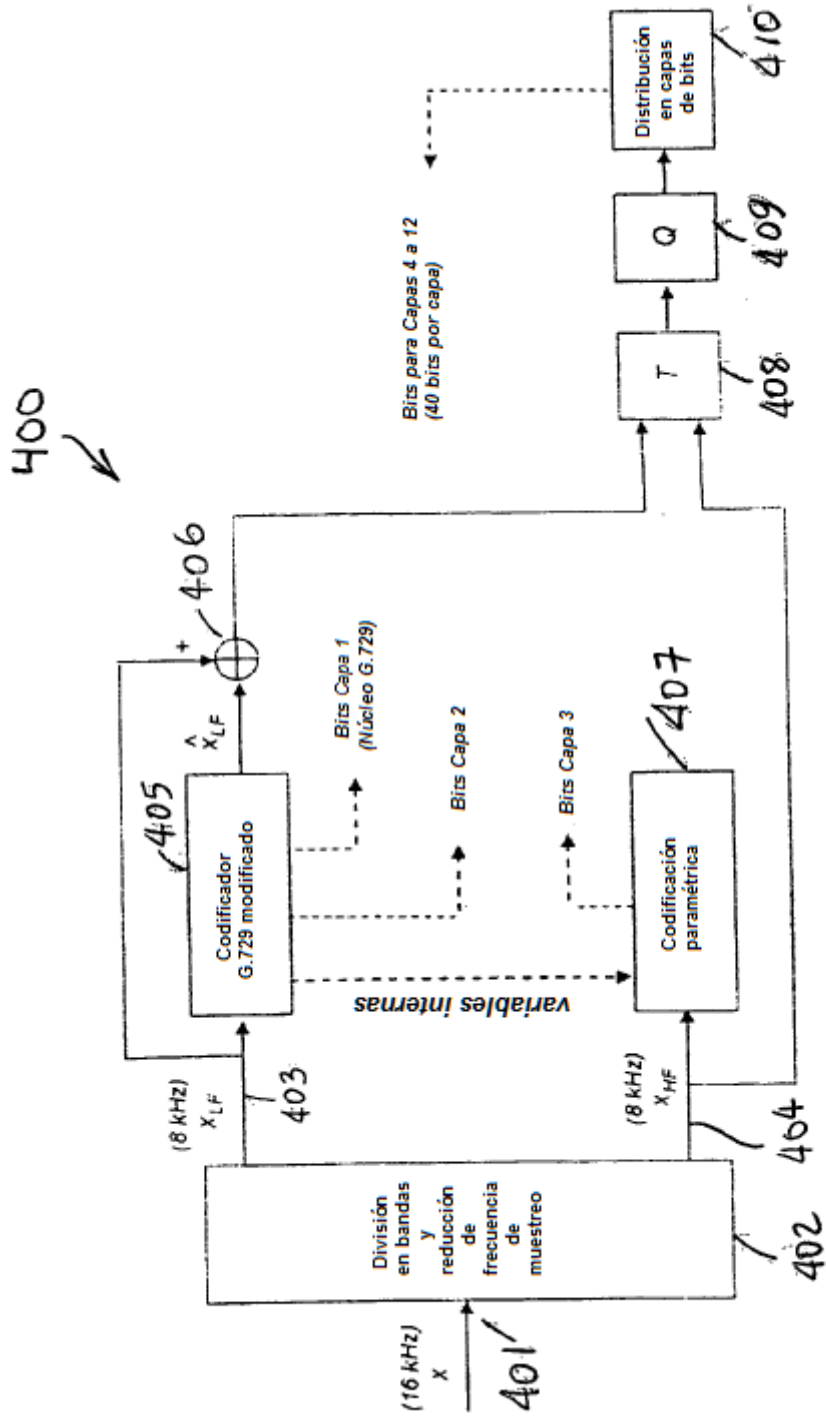


Figura 4

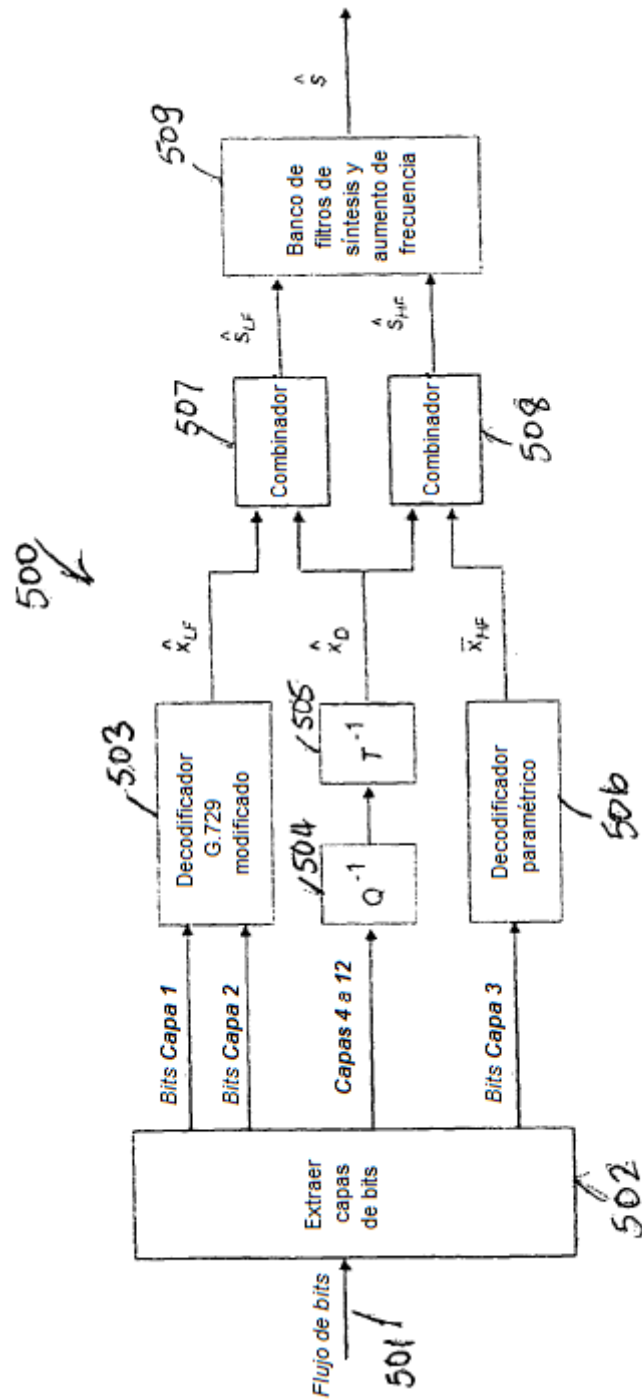


Figura 5

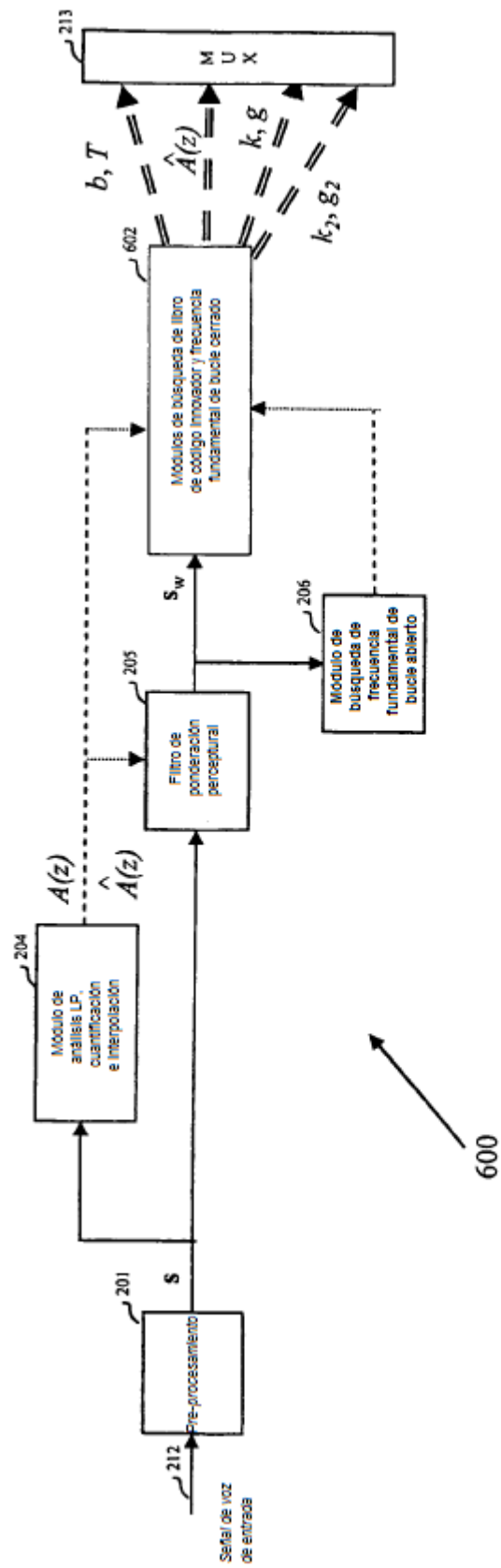


Figura 6

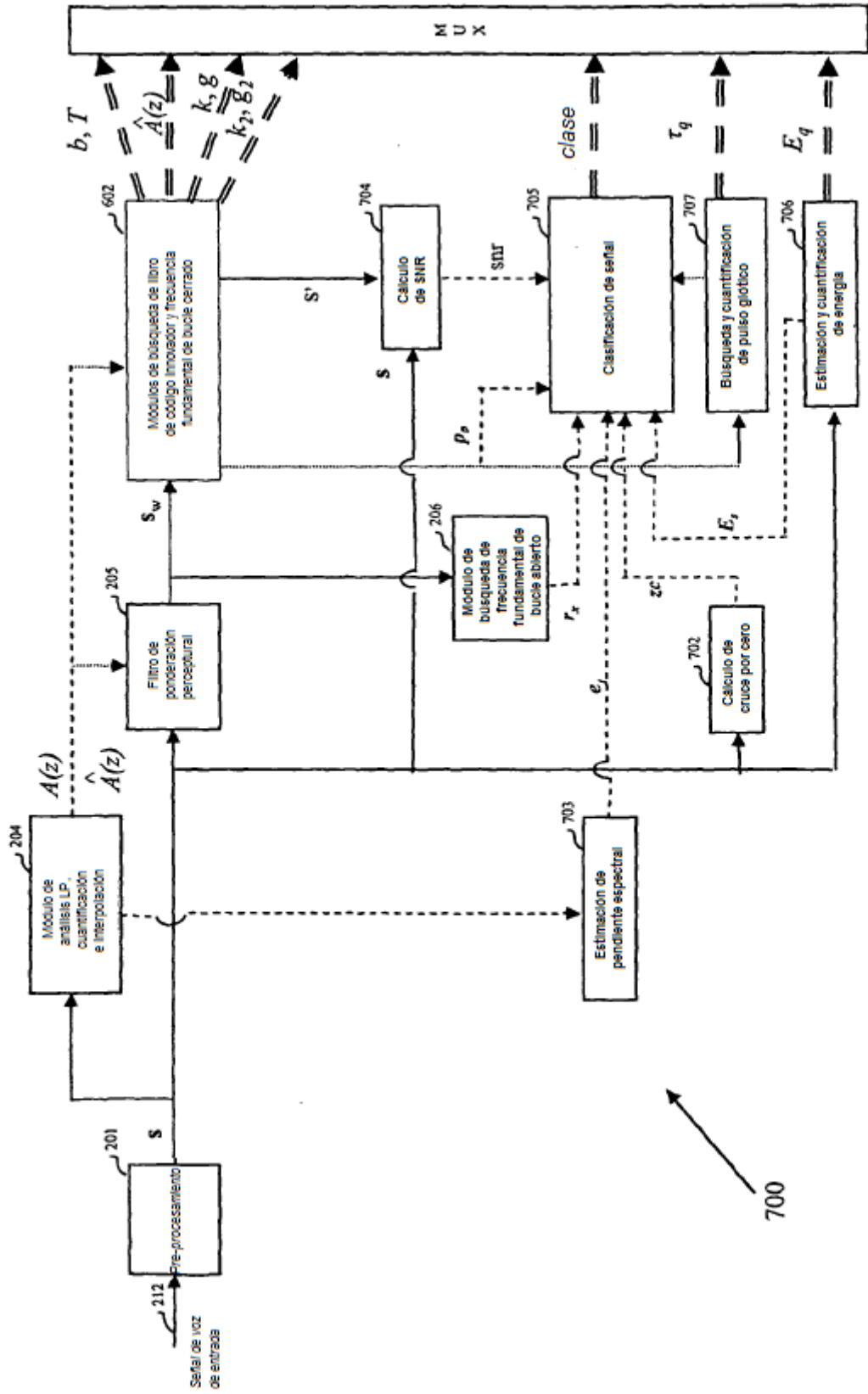


Figura 7

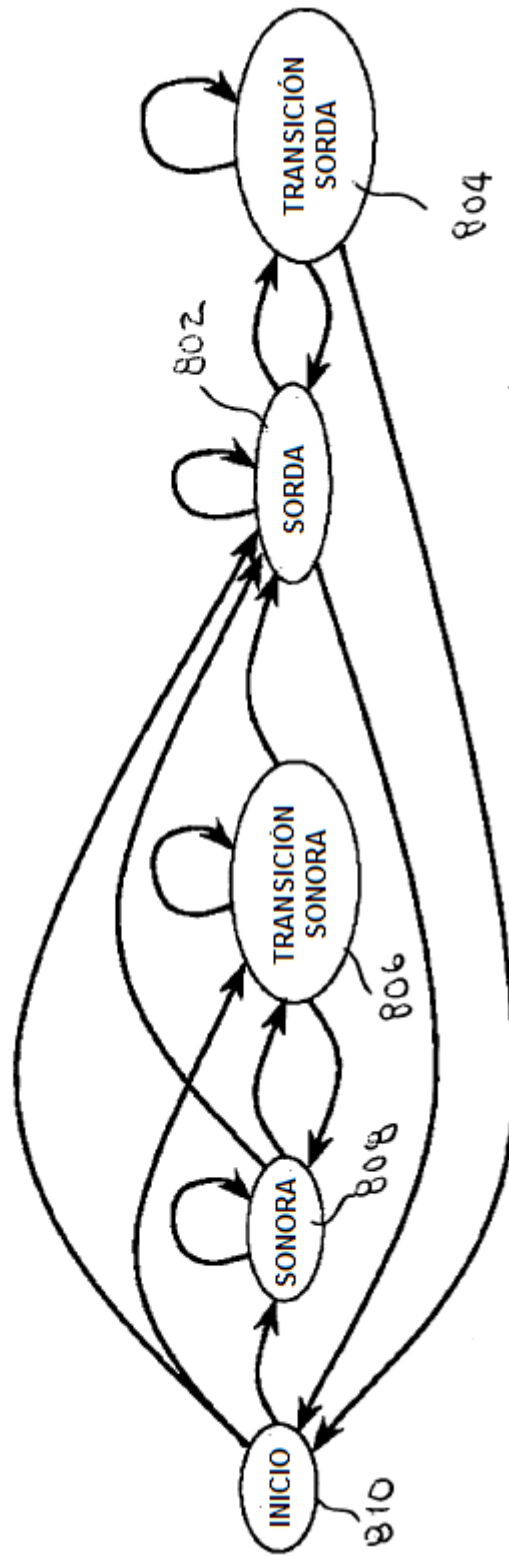


Figura 8



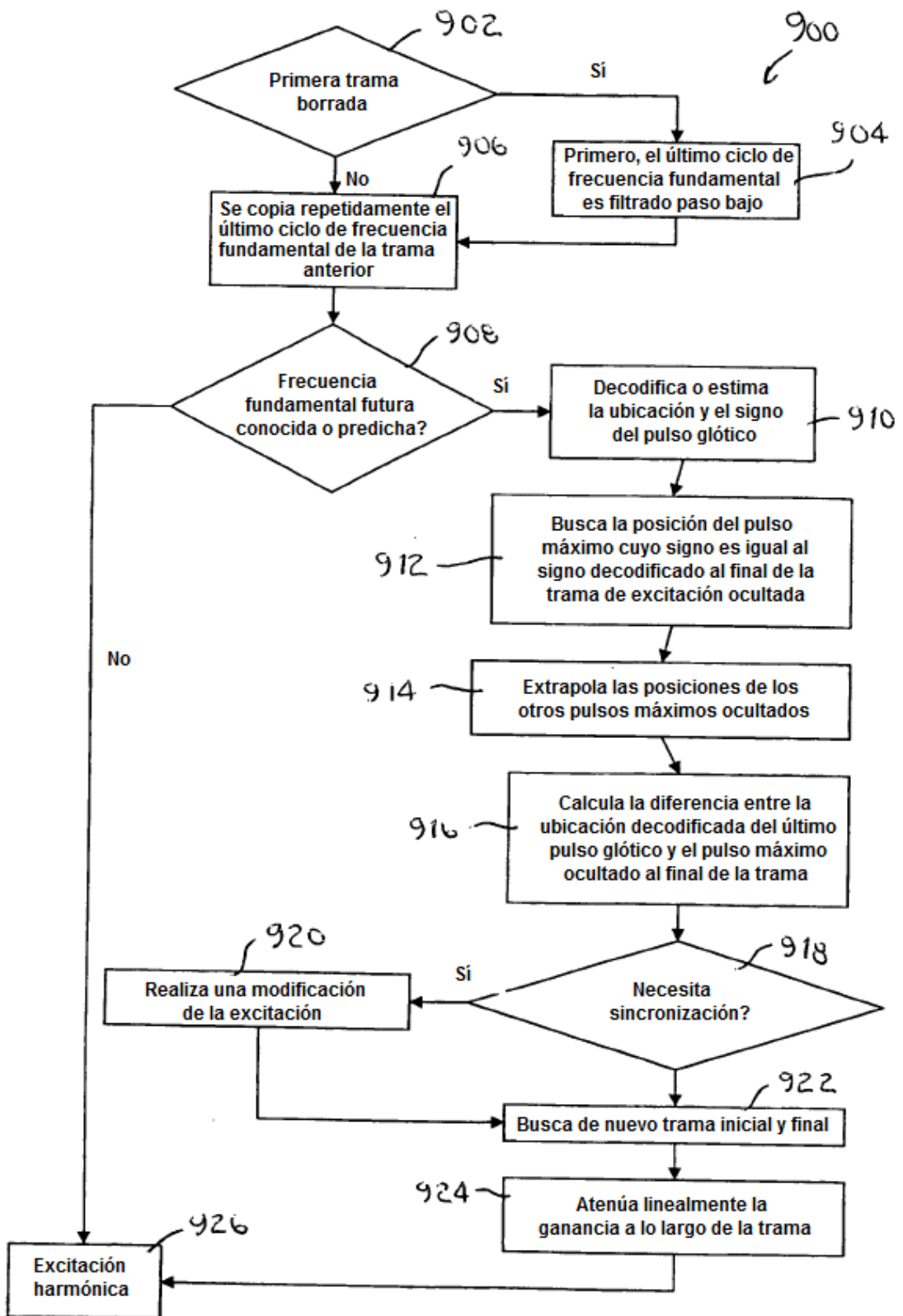


Figura 9

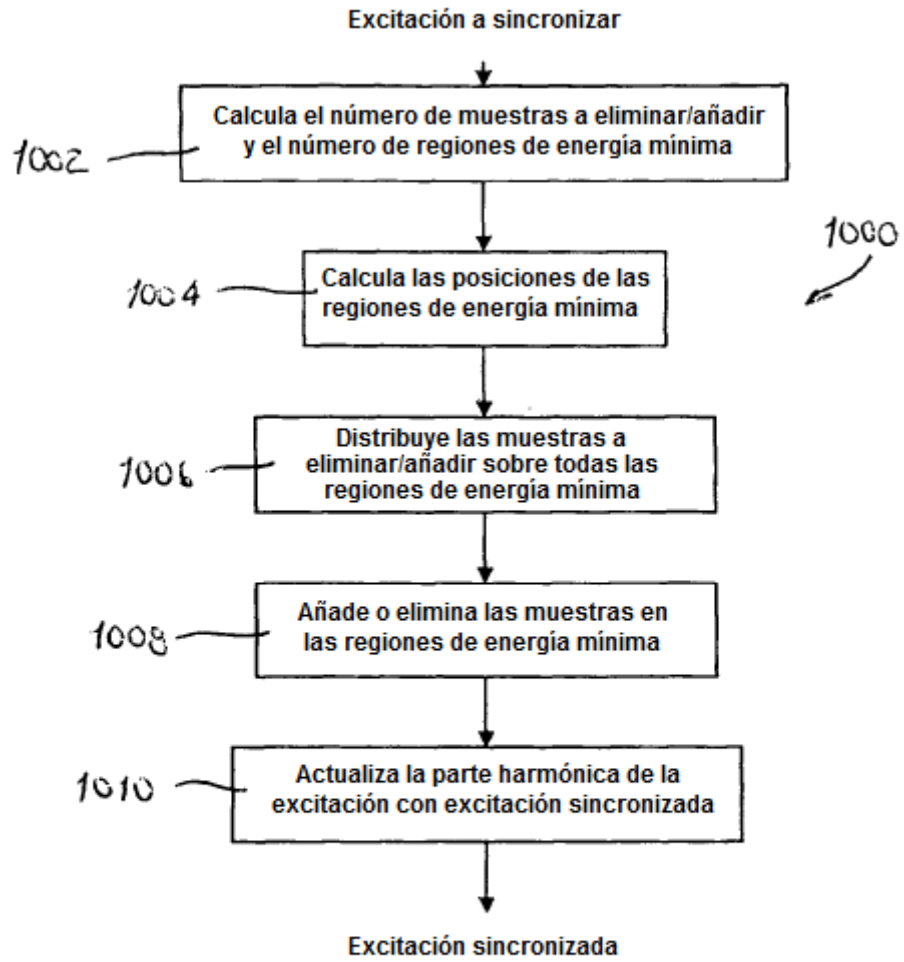


Figura 10

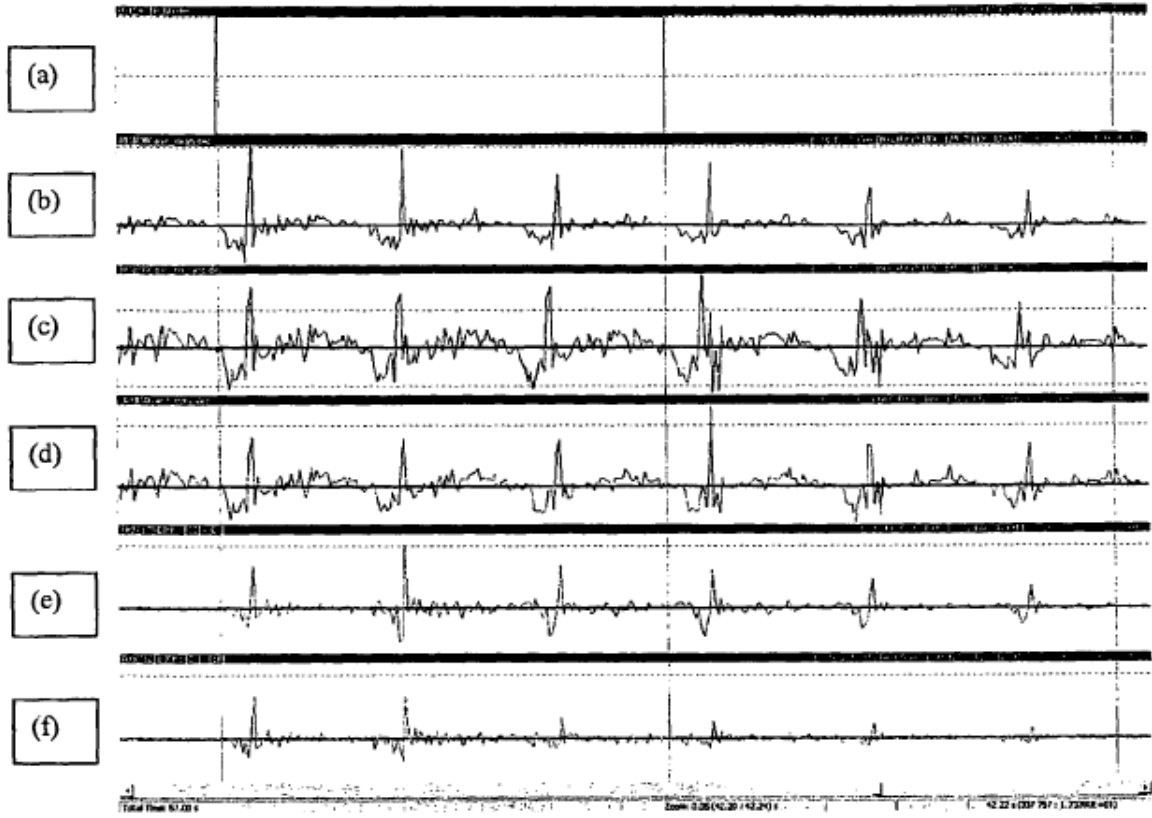


Fig 11

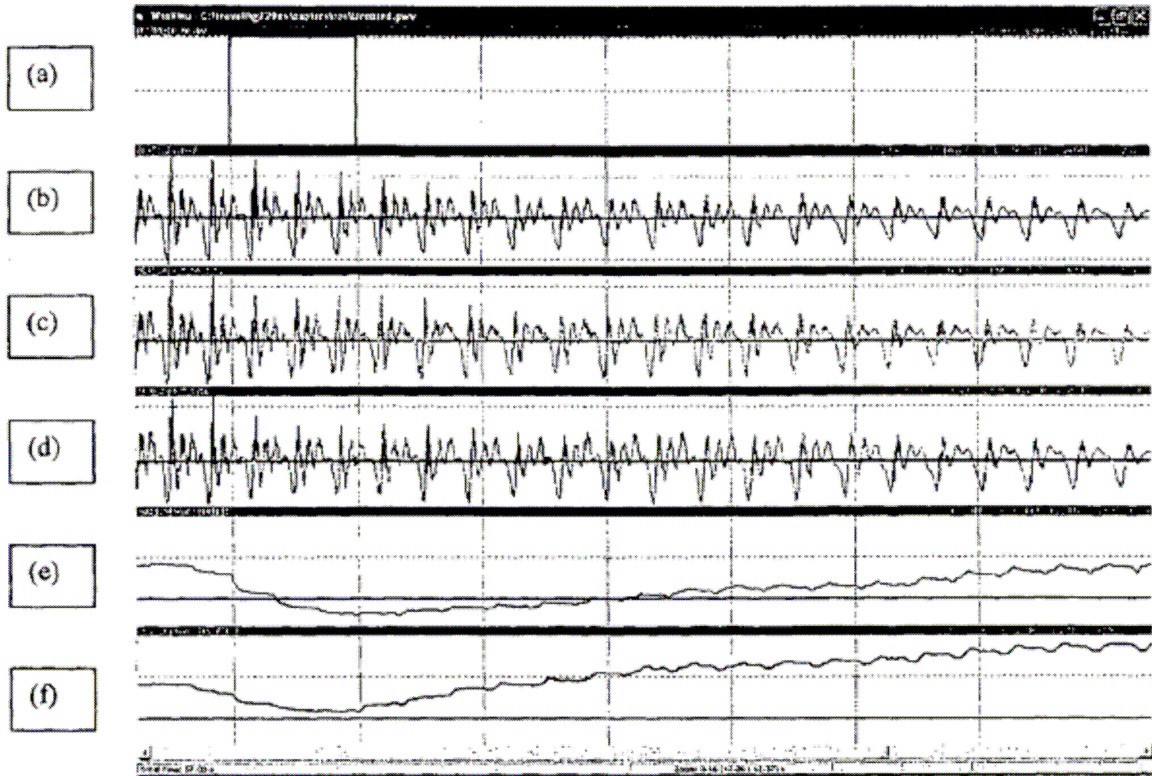


Fig 12

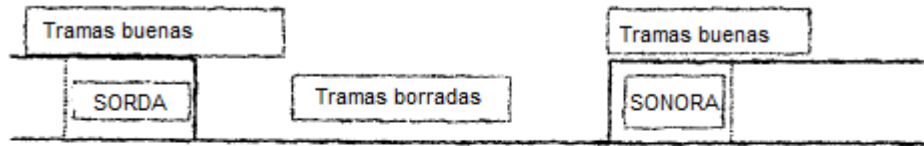


Figura 13