



19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 361 154**

51 Int. Cl.:  
**G10L 19/14** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **03769097 .1**

96 Fecha de presentación : **10.10.2003**

97 Número de publicación de la solicitud: **1554718**

97 Fecha de publicación de la solicitud: **20.07.2005**

54 Título: **Procedimiento para la interoperación entre codecs de la voz de banda ancha tasa múltiple adaptativo (AMR-WR) y de banda ancha, tasa de bits variable modo múltiple (VMR-WB).**

30 Prioridad: **11.10.2002 US 417667 P**

45 Fecha de publicación de la mención BOPI:  
**14.06.2011**

45 Fecha de la publicación del folleto de la patente:  
**14.06.2011**

73 Titular/es: **Nokia Corporation  
Keilalahdentie 4  
02150 Espoo, FI**

72 Inventor/es: **Jelinek, Milan y  
Salami, Redwan**

74 Agente: **López Bravo, Joaquín Ramón**

ES 2 361 154 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

Procedimientos para la interoperación entre códecs de la voz de banda ancha tasa múltiple adaptativo (AMR-WR) y de banda ancha, tasa de bits variable modo múltiple (VMR-WB)

5 La presente invención se refiere a la codificación digital de una señal de voz, con vistas a la transmisión y sintetizado de esta señal de sonido.

10 La demanda de técnicas digitales eficientes para la codificación de la voz en 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 la teleconferencia, multimedia y comunicaciones inalámbricas. Hasta recientemente, se ha usado principalmente el ancho de banda telefónico restringido a un intervalo de 200-3400 Hz en aplicaciones de codificaciones de la voz. Sin embargo, las aplicaciones de la voz de banda ancha proporcionan una inteligibilidad y naturalidad incrementada en la comunicación comparada con el ancho de banda telefónico convencional. Un ancho de banda en el intervalo 50-7000 Hz se ha hallado como suficiente para proporcionar una buena calidad dando una impresión de comunicación cara a cara. Para señales de audio generales, este ancho de banda da una calidad subjetiva aceptable pero aún es más bajo que la calidad de la radio FM o del CD que funcionan con intervalos de 20-16.000 Hz y 20-20.000 Hz, respectivamente.

15 Un codificador de la voz convierte una señal de voz en una transmisión digital de bits, que se transmite continuamente a través de un canal de comunicación o se almacena en un medio de almacenamiento. La señal de voz se digitaliza, esto es, se muestrea y cuantifica normalmente con 16 bits por muestra. El codificador de voz tiene el papel de representar estas muestras digitales con un número más pequeño de bits en tanto que mantiene una buena calidad subjetiva de la voz. El decodificador de la voz o sintetizador funciona sobre la transmisión continua de bits transmitida o almacenada y la convierte de vuelta a una señal de sonido.

20 La Codificación de Predicción Lineal con Excitación por Código (CELP) es una técnica bien conocida que permite conseguir un buen compromiso entre la calidad subjetiva y la tasa de bits. Esta técnica de codificación es una base para varias normas de codificación de la voz tanto en aplicaciones inalámbricas como por línea de cable. En la codificación CELP, la señal de voz muestreada se procesa en bloques sucesivos de L muestras normalmente denominadas tramas, en las que L es un número predeterminado que corresponde típicamente a 10-30 ms. Se calcula un filtro de predicción lineal (LP) y se transmite cada trama. El cálculo del filtro LP necesita típicamente una anticipación, un segmento de voz de 5-15 ms de la trama posterior. La trama de muestra L se divide en bloques más pequeños denominados subtramas. Normalmente el número de subtramas es tres o cuatro dando como resultado subtramas de 4-10 ms. En cada subtrama, se obtiene normalmente una señal de excitación a partir de dos componentes, la excitación pasada y la excitación innovadora, de libro de códigos fijo. El componente formado a partir de la excitación pasada se denomina a menudo como de libro de códigos adaptativo o excitación de tono. Los parámetros que caracterizan a la señal de excitación se codifican y transmiten al decodificador, en donde se usa la señal de excitación reconstruida como la entrada del filtro LP.

35 En sistemas inalámbricos que usan la tecnología de acceso múltiple por división de código (CDMA), el uso de una codificación de la voz de tasa de bits variable (VBR) controlada en origen mejora significativamente la capacidad del sistema. En la codificación VBR controlada en origen, el códec funciona a varias tasas de bits y se usa el módulo de selección de la tasa para determinar la tasa de bits utilizada para codificar cada trama de la voz en base a la naturaleza de la trama de voz (por ejemplo, sonoro, insonoro, transitorio, ruido de fondo). El objetivo es conseguir la mejor calidad de la voz con una tasa de bits media dada, también denominada, la tasa de datos media (ADR). El códec puede funcionar en diferentes modos mediante el ajuste del módulo de selección de tasa para conseguir diferentes ADR en los diferentes modos, en los que el rendimiento del códec se mejora con ADR incrementados. El modo de funcionamiento se impone por el sistema dependiendo de las condiciones del canal. Esto habilita al códec con un mecanismo de compromiso entre la calidad de la voz y la capacidad del sistema.

40 Típicamente, en la codificación VBR para sistemas CDMA, se usa un octavo de la tasa para codificar tramas sin actividad de voz (tramas de silencio o sólo con ruido). Cuando la trama es sonora de modo fijo o insonoro fijo, se usa media tasa o cuarto de tasa dependiendo del modo de funcionamiento. Si se puede usar media tasa, se usa el modelo CELP sin el libro de códigos de tono en el caso no sonorizado y se usa la modificación de la señal para mejorar la periodicidad y reducir el número de bits para los índices de tono en el caso sonorizado. Si el modo operativo impone un cuarto de tasa, no es posible normalmente un ajuste de la forma de onda dado que el número de bits es insuficiente y se aplica en general alguna codificación paramétrica. Se usa la tasa completa para las tramas de comienzo, transitorios y tramas sonoras de mezcla (se usa normalmente un modelo CELP típico). Además de la operación del códec controlado en origen en sistemas CDMA, el sistema puede limitar la tasa de bits máxima en alguna trama de la voz para enviar la información de señalización de la banda (denominada señalización dim-and-burst) o durante malas condiciones en el canal (tal como cerca de los límites de la célula) para mejorar la robustez del códec. Esta se denomina como la máxima media tasa. Cuando el módulo de selección de tasa elige la trama para ser codificada como una trama de tasa completa y el sistema impone por ejemplo una trama HR, el rendimiento de la voz se degrada dado que los modos de HR dedicados no son capaces de codificar eficientemente señales de comienzo y transitorias. Se puede proporcionar otro modelo de codificación HR (o de cuarto de tasa (QR)) para manejar estos casos especiales.

Como se puede ver a partir de la descripción anterior, la clasificación de la señal y la determinación de la tasa son muy esenciales para una codificación VBR eficiente. La selección de la tasa es una parte clave para conseguir la tasa de datos media más baja con la mejor calidad posible.

5 Se ha seleccionado recientemente un códec de voz de banda ancha de tasa múltiple adaptativo (AMR-WB) por la ITU-T (Unión de Telecomunicaciones Internacional - Sector de Normalización de las Telecomunicaciones) para varios servicios y telefonía de voz en banda ancha y por el 3GPP (proyecto de asociación para la tercera generación) para GSM y sistemas inalámbricos de tercera generación W-CDMA. El códec AMR-WB consiste en 9 tasas de bits, concretamente 6,6, 8,85, 12,65, 14,25, 15,85, 18,25, 19,85, 23,05 y 23,85 kbits/s. Es deseable por ello la interoperación entre los códecs CDMA-WB y AMR-WB.

10 El documento WO-A-01/22402 y la Recomendación G. 722.2 del ITU-T se refieren a la codificación de la voz.

Un objetivo de la presente invención es proporcionar unos procedimientos de clasificación de la señal y de selección de tasa mejorados para una codificación de la voz en banda ancha de tasa variable en general y en particular para proporcionar unos procedimientos de clasificación de la señal y de selección de tasa mejorados para una codificación de la voz en banda ancha modo múltiple tasa variable adecuados para sistemas CDMA. Otro objetivo es proporcionar técnicas para una interoperación eficiente entre el códec VBR de banda ancha para sistemas CDMA y el códec AMR-WB estándar.

La invención se define por las reivindicaciones.

20 Otros objetivos, ventajas y características de la presente invención serán más evidentes tras la lectura de la siguiente descripción no limitativa de las realizaciones ilustrativas de la misma, dada a modo de ejemplo solamente, con referencia a los dibujos adjuntos.

En los dibujos adjuntos:

la Figura 1 es un diagrama de bloques de un sistema de comunicaciones de voz que ilustra el uso de un dispositivo de codificación y decodificación de la voz de acuerdo con un primer aspecto de la presente invención;

25 la Figura 2 es un diagrama de flujo que ilustra un procedimiento para la codificación digital de una señal de sonido de acuerdo con una primera realización ilustrativa de un segundo aspecto de la presente invención;

la Figura 3 es un diagrama de flujo que ilustra un procedimiento para discriminar una trama insonora de acuerdo con una realización ilustrativa de un tercer aspecto de la presente invención;

30 la Figura 4 es un diagrama de flujo que ilustra un procedimiento para la discriminación de una trama sonora estable de acuerdo con una realización ilustrativa de un cuarto aspecto de la presente invención;

la Figura 5 es un diagrama de flujo que ilustra un procedimiento para la codificación digital de una señal de sonido en un modo Premium de acuerdo con una segunda realización ilustrativa del segundo aspecto de la presente invención;

35 la Figura 6 es un diagrama de flujo que ilustra un procedimiento para la codificación digital de una señal de sonido en el modo Estándar de acuerdo con una tercera realización ilustrativa del segundo aspecto de la presente invención;

la Figura 7 es un diagrama de flujo que ilustra un procedimiento para la codificación digital de una señal de sonido en el modo Económico de acuerdo con una cuarta realización ilustrativa del segundo aspecto de la presente invención;

40 la Figura 8 es un diagrama de flujo que ilustra un procedimiento para la codificación digital de una señal de sonido en el modo Interoperativo de acuerdo con una quinta realización ilustrativa del segundo aspecto de la presente invención;

45 la Figura 9 es un diagrama de flujo que ilustra un procedimiento para la codificación digital de una señal de sonido en el modo Premium o Estándar durante el máximo de media tasa de acuerdo con una sexta realización ilustrativa del segundo aspecto de la presente invención;

la Figura 10 es un diagrama de flujo que ilustra un procedimiento para la codificación digital de una señal de sonido en el modo económico durante un máximo de media tasa de acuerdo con una séptima realización ilustrativa del segundo aspecto de la presente invención

50 la Figura 11 es un diagrama de flujo que ilustra un procedimiento para la codificación digital de una señal de sonido en el modo el Interoperativo durante un máximo de media tasa de acuerdo con una octava realización ilustrativa del segundo aspecto de la presente invención y

la Figura 12 es un diagrama de flujo que ilustra un procedimiento para la codificación digital de una señal de sonido de modo que se permita la Interoperación entre los códecs VMR-WB y AMR-WB, de acuerdo con una realización ilustrativa de un quinto aspecto de la presente invención.

5 Volviendo ahora a la Figura 1 de los dibujos adjuntos, se ilustra un sistema de comunicaciones por la voz 10 que representa el uso de la codificación y decodificación de la voz de acuerdo con una realización ilustrativa del primer aspecto de la presente invención. El sistema de comunicación por la voz 10 soporta la transmisión y reproducción de una señal de voz a través de un canal de comunicación 12. El canal de comunicación 12 puede comprender por ejemplo un enlace cableado, óptico o de fibra o un enlace de radiofrecuencia. El canal de comunicación 12 puede ser también una combinación de diferentes medios de transmisión, por ejemplo en parte un enlace de fibra y en parte un enlace de radiofrecuencia. El enlace de radiofrecuencia puede permitir soportar múltiples comunicaciones de voz simultáneas lo que requiere recursos compartidos de ancho de banda tal como los que se pueden encontrar en la telefonía celular. Alternativamente, el canal de comunicación se puede sustituir por un dispositivo de almacenamiento (no mostrado) en una única realización de dispositivos del sistema de comunicación que registra y almacena la señal de voz codificada para reproducción posterior.

10 El sistema de comunicación 10 incluye un dispositivo codificador que se compone de un micrófono 14, un convertidor analógico a digital 16, un codificador de la voz 18 y un codificador del canal 20 en el lado emisor del canal de comunicación 12 y un decodificador de canal 22, un decodificador de la voz 24, un convertidor digital a analógico 26 y un altavoz 28 en el lado receptor.

20 El micrófono 14 produce una señal analógica de la voz que se conduce a un convertidor analógico a digital (A/D) 16 para convertirla en una forma digital. El codificador de la voz 18 codifica la señal de voz digitalizada produciendo un conjunto de parámetros que se codifican en una forma binaria y se suministran a un codificador del canal 20. El codificador del canal opcional 20 añade redundancia a la representación binaria de los parámetros de codificación antes de la transmisión a través del canal de comunicación 12. También, en algunas aplicaciones tales como las aplicaciones en paquetes de redes, las tramas codificadas se empaquetan antes de la transmisión.

25 En el lado receptor, un decodificador del canal 22 utiliza la información redundante recibida en el flujo de bits para detectar y corregir los errores del canal ocurridos en la transmisión. Un decodificador de la voz 24 convierte el flujo de bits recibidos desde el decodificador del canal 20 de nuevo en un conjunto de parámetros de codificación para la creación de una señal de voz sintetizada. La señal de voz sintetizada reconstruida en el decodificador de la voz 24 se convierte a una forma analógica en un convertidor digital a analógico (D/A) 26 y se reproduce en una unidad de altavoz 28.

El micrófono 14 y/o el convertidor A/D 16 se puede sustituir en algunas realizaciones por otras fuentes de voz para el codificador de la voz 18.

El codificador 20 y el decodificador 22 se configuran de modo que realicen un procedimiento para la codificación de una señal de voz de acuerdo con la presente invención como se describe a continuación en el presente documento.

### 35 **Clasificación de la señal**

Volviendo ahora a la Figura 2, se ilustra un procedimiento 100 para la codificación digital de una señal de voz de acuerdo con una primera realización ilustrativa de un primer aspecto de la presente invención. El procedimiento 100 incluye un procedimiento de clasificación de la señal de voz de acuerdo con una realización ilustrativa de un segundo aspecto de la presente invención. Se hace notar que la expresión señal de voz se refiere a señales de voz así como a cualquier señal multimedia que pueda incluir una parte de voz tal como audio con contenido de voz (voz en medio de música, voz con música de fondo, voz con efectos sonoros especiales, etc.).

40 Como se ilustra en la Figura 2, la clasificación de la señal se realiza en tres etapas 102, 106 y 110, discriminando cada una de ellas una clase de señal específica. Primero, en la etapa 102, un clasificador de primer nivel en la forma de un detector de actividad de voz (VAD) (no mostrado) discrimina entre tramas de la voz activas e inactivas. Si se detecta una trama de la voz inactiva entonces el procedimiento de codificación 100 acaba con la codificación de la trama actual con, por ejemplo, la generación de ruido de confort (CNG) (etapa 104). Si se detecta una trama de la voz activa en la etapa 102, la trama se somete a un clasificador de segundo nivel (no mostrado) configurado para discriminar tramas insonoras. En la etapa 106, si el clasificador clasifica la trama como señal de voz insonora, el procedimiento de codificación 100 acaba en la etapa 108, en la que la trama se codifica usando una técnica de codificación optimizada para señales insonoras. En otro caso, la trama de la voz se pasa en la etapa 110, a través de un clasificador de tercer nivel (no mostrado) en la forma de un módulo de clasificación "sonoro estable" (no mostrado). Si la trama actual se clasifica como una trama sonora estable, entonces la trama se codifica usando una técnica de codificación optimizada para señales sonoras estables (etapa 112). En otro caso, la trama es probable que contenga un segmento de voz no fija tal como un comienzo sonoro, una parte de señal de voz sonora que evoluciona rápidamente y la trama se codifica usando un codificador de la voz de propósito general con una alta tasa de bits permitiendo mantener una buena calidad subjetiva (etapa 114). Nótese que si la energía relativa de la trama es más baja que un cierto umbral entonces estas tramas se pueden codificar con un tipo de codificación genérica de tasa menor para reducir adicionalmente la tasa de datos media.

Los clasificadores y codificadores pueden tomar muchas formas desde un circuito electrónico a un chip de procesador.

Se explicará a continuación con más detalle la clasificación de los diferentes tipos de señal de voz y se desvelarán los procedimientos para la clasificación de la voz como sonora e insonora.

5 **Discriminación de tramas de voz inactivas (VAD)**

Las tramas de la voz inactivas se discriminan en la etapa 102 usando el detector de actividad de voz (VAD). El diseño del VAD es el conocido para un experto en la técnica y no se describirá en el presente documento con más detalle. Se describe un ejemplo de VAD en [5].

**Discriminación de tramas de la voz activas insonoras**

10 Las partes insonora de una señal de voz se caracterizan por la pérdida de periodicidad y se pueden dividir adicionalmente en tramas inestables, en las que la energía y el espectro cambian rápidamente y tramas estables en donde estas características se mantienen relativamente estables.

En la etapa 106, se discriminan las tramas insonoras usando al menos tres de entre los siguientes parámetros:

- una medida de la sonorización, que se puede calcular como una correlación normalizada promediada ( $\bar{r}_x$ );
- 15 • una medida de la tendencia espectral ( $e_t$ );
- una relación de energía de la señal ( $dE$ ) usada para evaluar la variación de energía de la trama dentro de la trama y por ello la estabilidad de la trama; y
- la energía relativa de la trama.

**Medida de la sonorización**

20 La Figura 3 ilustra un procedimiento 400 para la discriminación de una trama insonora de acuerdo con una realización ilustrativa de un tercer aspecto de la presente invención.

La correlación normalizada, usada para determinar la medición de voces, se calcula como parte del módulo de búsqueda del tono en bucle abierto 214. En la realización ilustrativa de la Figura 3, se usan tramas de 20 ms. El módulo de búsqueda del tono en bucle abierto produce normalmente la estimación del tono en bucle abierto p cada 25 10 ms (2 veces por trama). En el procedimiento 200, se usa también la salida de las mediciones de correlación normalizadas  $r_x$ . Estas correlaciones normalizadas se calculan sobre la voz ponderada y la voz ponderada pasada en el retardo de tono de bucle abierto. La señal de voz ponderada  $s_w(n)$  se calcula en un filtro de ponderación perceptual 212. En esta realización ilustrativa, se usa un filtro de ponderación perceptual 212 con denominador fijo, adecuado para señales de banda ancha. La siguiente relación da un ejemplo de la función de transferencia para el 30 filtro de ponderación perceptual 212:

$$W(z) = A(z / \gamma_1) / (1 - \gamma_2 z^{-1})$$

en el que  $0 < \gamma_2 < \gamma_1 \leq 1$

en el que  $A(z)$  es la función de transferencia del filtro de predicción lineal (LP) calculado en el módulo 218, que viene dado por la relación siguiente:

35 
$$A(z) = 1 + \sum_{i=1}^p a_i z^{-i}$$

La medición de sonorización se da por la correlación promedio  $\bar{r}_x$  que se define como

$$\bar{r}_x = \frac{1}{3} (r_x(0) + r_x(1) + r_x(2)) \quad (1)$$

en la que  $r_x(0)$ ,  $r_x(1)$  y  $r_x(2)$  son respectivamente la correlación normalizada de la primera mitad de la trama actual, la correlación normalizada de la segunda mitad de la trama actual y la correlación normalizada de la anticipación (comienzo de la siguiente trama).

40

Se puede añadir un factor de correlación  $r_e$  a la correlación normalizada en la Ecuación (1) para tener en cuenta la presencia de ruido de fondo. En presencia de ruido de fondo, la correlación normalizada promedio disminuye. Sin embargo, para la finalidad de clasificación de la señal, esta disminución no debería afectar a la decisión sonora-insonora, de modo que se compensa por la adición de  $r_e$ . Se debería tener en cuenta que cuando se usa un buen

algoritmo de reducción de ruido,  $r_e$  es prácticamente cero.

En el procedimiento 200, se usa una anticipación de 13 ms. La correlación normalizada  $r_x(k)$  se calcula como sigue

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

en la que

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

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

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

En el procedimiento 200, el cálculo de las correlaciones es como sigue: las correlaciones  $r_x(k)$  se calculan sobre la señal de voz ponderada  $s_w(n)$ . Los instantes  $t_k$  se relacionan con el comienzo de la media trama actual y son iguales a 0, 128 y 256 muestras respectivamente para  $k = 0, 1$  y  $2$ , a 12800 Hz de tasa de muestreo. Los valores  $p_k = T_{OL}$  son las estimaciones del tono en bucle abierto seleccionadas estimadas para las medias tramas. La longitud del cálculo de autocorrelación  $L_k$  depende del período del tono. En una primera realización, los valores de  $L_k$  se resumen a continuación (para la tasa de muestreo de 12,8 kHz):

$L_k = 80$  muestras para  $p_k \leq 62$  muestras

$L_k = 124$  muestras para  $62 < p_k \leq 122$  muestras

$L_k = 230$  muestras para  $p_k > 122$  muestras

estas longitudes aseguran que la longitud del vector correlacionado comprende al menos un período del tono, lo que ayuda para una detección del tono en bucle abierto robusta. Para períodos de tono largos ( $p_1 > 122$  muestras),  $r_x(1)$  y  $r_x(2)$  son idénticos, es decir sólo se calcula una correlación dado que los vectores correlacionados son suficientemente largos para que el análisis sobre la anticipación ya no sea necesario.

Alternativamente, la señal de voz ponderada se puede diezmar en 2 para simplificar la búsqueda del tono en bucle abierto. La señal de voz ponderada se puede filtrar en paso bajo antes del diezmo. En este caso, los valores de  $L_k$  se serán por

$L_k = 40$  muestras para  $p_k \leq 31$  muestras

$L_k = 62$  muestras para  $62 < p_k \leq 61$  muestras

$L_k = 115$  muestras para  $p_k > 61$  muestras

se pueden usar otros procedimientos para calcular las correlaciones. Por ejemplo, se puede calcular solamente un valor de correlación normalizado para la totalidad de la trama en lugar de promediar varias correlaciones normalizadas. Adicionalmente, las correlaciones se puede calcular sobre las señales distintas a las de la voz ponderada tal como la señal residual, la señal de voz o una señal filtrada en paso bajo residual, de la voz o de la voz ponderada.

### **Tendencia espectral**

El parámetro de la tendencia espectral contiene la información sobre la distribución en frecuencia de la energía. En el procedimiento 200, la tendencia espectral se estima en el dominio de la frecuencia como una relación entre la energía concentrada en las bajas frecuencias y la energía concentrada en las altas frecuencias. Sin embargo, se puede estimar también de formas diferentes tal como una relación entre los dos primeros coeficientes de autocorrelación de la señal de voz.

En el procedimiento 200, se usa la Transformada de Fourier discreta para realizar el análisis espectral en el módulo 210 de la Figura 10. El análisis de frecuencia y el cálculo de tendencia se realizan dos veces por trama. Se usa una Transformada de Fourier Rápida (FFT) de 256 puntos con un solape del 50 por ciento. Las ventanas de análisis se sitúan de modo que se explote la anticipación completa. El comienzo de la primera ventana se sitúa 24 muestras después del comienzo de la trama actual. La segunda ventana se coloca 128 muestras más adelante. Se pueden usar diferentes ventanas para ponderar la señal de entrada para el análisis de frecuencia. Se usa una raíz cuadrada

de una ventana de Hamming (que es equivalente a una ventana senoidal). Esta ventana se adecua particularmente bien para procedimientos de solape-adición, por lo tanto este análisis espectral particular se puede usar en un algoritmo de supresión de ruido opcional en base a la sustracción espectral y al análisis/síntesis de solape-adición. Estos algoritmos de supresión de ruido se cree que son bien conocidos en la técnica, no serán descritos en el presente documento con más detalle.

La energía en altas frecuencias y en bajas frecuencias se calcula siguiendo las bandas críticas preceptuales [6]:

Bandas críticas = {100,0, 200,0, 300,0, 400,0, 510,0, 630,0, 770,0, 920,0, 1080,0, 1270,0, 1480,0, 1720,0, 2000,0, 2320,0, 2700,0, 3150,0, 3700,0, 4400,0, 5300,0, 6350,0} Hz.

La energía en altas frecuencias se calcula como el promedio de las energías de las dos últimas bandas críticas

$$\bar{E}_h = 0,5 (E_{CB}(18) + E_{CB}(19))$$

en la que  $E_{CB}(i)$  son las energías promedio para bandas críticas calculadas como

$$E_{CB}(i) = \frac{1}{N_{CB}(i)} \sum_{k=0}^{N_{CB}(i)-1} (x_R^2(k + j_i) + x_I^2(k + j_i)),$$

$i = 0, \dots, 19$

en la que  $N_{CB}(i)$  es el número de contenidos de frecuencia en la banda  $i$ -ésima y  $X_R(k)$  y  $X_I(k)$  son, respectivamente, la parte real e imaginaria del  $k$ -ésimo contenido de frecuencia y  $j_i$  es el índice del primer contenido en la  $i$ -ésima banda crítica.

La energía en bajas frecuencias se calcula como el promedio de las energías en las primeras 10 bandas críticas. Las bandas críticas medias se han excluido del cálculo para mejorar la discriminación entre tramas con alta concentración de energía en bajas frecuencias (generalmente sonora) y con alta concentración de energía en altas frecuencias (generalmente insonora). Entre medias, el contenido de energía no es característico para ninguna de las clases y aumenta la confusión de la decisión.

La energía en bajas frecuencias se calcula de modo diferente para períodos de tono largo y períodos de tono corto. Para segmentos de voz femenina sonoros, se explota la estructura armónica del espectro para aumentar la discriminación sonora-insonora. Por ello se calculan los períodos de tono cortos,  $E_i$  en relación a los contenidos y solamente los contenidos de frecuencia suficientemente cercanos a los armónicos de la voz se tienen en cuenta en la suma. Esto es

$$\bar{E}_h = \frac{1}{cnt} \sum_{k=0}^{24} E_{BIN}(k) w_h(k)$$

en la que  $E_{BIN}(k)$  son las energías de contenidos en las primeras 25 contenidos de energía (no se considera el componente en corriente continua). Nótese que estos 25 contenidos corresponden a las primeras 10 bandas críticas.

En la suma anterior, solamente se consideran los términos relacionados con los contenidos cercanos a los armónicos del tono, de modo que  $w_h(k)$  se establece en 1 si la distancia entre el contenido y el armónico más cercano no es más grande que un cierto umbral de frecuencia (50 Hz) y se establece en 0 en caso contrario. El contador  $cnt$  es el número de términos no cero en la suma. Sólo se tienen en cuenta los contenidos más cerca que 50 Hz a los armónicos más cercanos. Por ello, si la estructura es armónica en bajas frecuencias, sólo los términos de alta energía serán incluidos en la suma. Por otro lado, si la estructura no es armónica, la selección de términos será aleatoria y la suma será más pequeña. Por ello se pueden detectar incluso sonidos insonoros con alto contenido en energía en bajas frecuencias. Este procesamiento no se puede realizar para períodos de tono largos, dado que la resolución de frecuencia no es suficiente. Para valores de tono más grandes que 128 o para sonidos a priori insonoros la energía de baja frecuencia se calcula por banda crítica como

$$\bar{E}_l = \frac{1}{10} \sum_{k=0}^9 E_{CB}(k)$$

Se determinan a priori unos sonidos insonoros cuando  $r_x(0) + r_x(1) + r_e < 0,6$ , en el que el valor  $r_e$  es una correlación añadida a la correlación normalizada como se ha descrito anteriormente.

Las energías en baja y alta frecuencia resultantes se obtienen mediante la resta de la energía de ruido estimada de los valores  $\bar{E}_l$  y  $\bar{E}_h$  calculados anteriormente. Esto es

$$E_h = \bar{E}_h - N_h$$

$$E_l = \bar{E}_l - N_l$$

en las que  $N_h$  y  $N_l$  son las energía de ruido promedio en las 2 últimas bandas críticas y las primeras 10 bandas críticas respectivamente. Las energías de ruido estimadas se han añadido al cálculo de tendencia para tener en cuenta la presencia del ruido de fondo.

5 Finalmente, la tendencia espectral viene dada por

$$e_{tend}(i) = \frac{E_l}{E_h}$$

nótese que el cálculo de la tendencia espectral se realiza dos veces por trama para obtener  $e_{tend}(0)$  y  $e_{tend}(1)$  correspondiendo a ambos análisis espectrales por trama. La tendencia espectral promedio usada en la clasificación de una trama insonora viene dada por

10 
$$e_t = \frac{1}{3} (e_{ant} + e_{tend}(0) + e_{tend}(1))$$

en la que  $e_{ant}$  es la tendencia desde el segundo análisis espectral de la trama previo.

**Variación de energía dE**

15 La variación de energía dE se evalúa sobre la señal de voz sin ruido  $s(n)$ , en la que  $n = 0$  corresponde al comienzo de la trama actual. La energía de la señal se evalúa dos veces por subtrama, es decir 8 veces por trama, en base a segmentos de tiempo corto de 32 muestras de longitud. Adicionalmente, se calculan también las energías a corto plazo de las últimas 32 muestras de la trama previa y de las primeras 32 muestras de la siguiente trama. Las energías máximas a corto plazo se calculan como

$$E_{st}^{(1)}(j) = \max_{i=0}^{31} (s^2(i + 32j)),$$

$j = -1, \dots, 8$

20 en la que  $j = -1$  y  $j = 8$  corresponden al final de la trama previa y el comienzo de la trama siguiente. Se calculan otro conjunto de 9 energías máximas desplazando los índices de la voz en 16 muestras. Esto es

$$E_{st}^{(2)}(j) = \max_{i=0}^{31} (s^2(i + 32j - 16)),$$

$j = 0, \dots, 8$

25 la variación de energía dE máxima entre segmentos de corto plazo consecutivos se calcula como la máxima de las siguientes:

$$\begin{aligned} & \text{si} \\ & E_{st}^{(1)}(0) / E_{st}^{(1)}(-1) > E_{st}^{(1)}(0) > E_{st}^{(1)}(-1), \\ & \text{si} \\ & E_{st}^{(1)}(7) / E_{st}^{(1)}(8) > E_{st}^{(1)}(7) > E_{st}^{(1)}(8), \\ & \text{para } j=1 \text{ a } 7 \\ & \frac{\max(E_{st}^{(1)}(j), E_{st}^{(1)}(j-1))}{\min(E_{st}^{(1)}(j), E_{st}^{(1)}(j-1))} \\ & \text{para } j=1 \text{ a } 8 \\ & \frac{\max(E_{st}^{(2)}(j), E_{st}^{(2)}(j-1))}{\min(E_{st}^{(2)}(j), E_{st}^{(2)}(j-1))} \end{aligned}$$

Alternativamente, se pueden usar otros procedimientos para evaluar la variación de energía en la trama.



**Energía relativa  $E_{rel}$**

La energía relativa de la trama viene dada por la diferencia entre la energía de la trama en dB y la energía promedio a largo plazo. La energía de la trama se calcula como

$$E_t = 10 \log \left( \sum_{i=0}^{19} E_{CB}(i) \right), \text{ en dB}$$

5 en la que  $E_{CB}(i)$  son las energías promedio para bandas críticas como se ha descrito anteriormente. La energía de la trama promedio a largo plazo viene dada por

$$\bar{E}_t = 0,99 \bar{E}_t + 0,01 E_t$$

con un valor inicial  $\bar{E}_t = 45$  dB.

Por ello la energía relativa de la trama viene dada por

10 
$$E_{rel} = E_t - \bar{E}_t$$

La energía de la trama relativa se usa para identificar tramas de baja energía que no se han clasificado como tramas de ruido de fondo o tramas insonoras. Estas tramas se pueden codificar con un codificador HR genérico para reducir el ADR.

**Clasificación de la voz insonora**

15 La clasificación de la voz insonora se hace en base a los parámetros descritos anteriormente, concretamente: la medida de la sonorización  $\bar{r}_x$ , la tendencia espectral  $e_t$ , la variación de energía dentro de una trama  $dE$  y la energía de la trama relativa  $E_{rel}$ . La decisión se toma en base a al menos tres de estos parámetros. Los umbrales de decisión se ajustan en base al modo de operación (la tasa de datos promedio requerida). Básicamente para modos de operación con bajas tasas de datos deseadas, los umbrales se ajustan para favorecer más clasificaciones insonoras (dado que se usará una codificación de media tasa o de cuarto de tasa para codificar la trama). Las tramas insonoras se codifican usualmente con codificadores de HR insonora. Sin embargo, en el caso del modo económico, se puede usar también QR insonora para reducir adicionalmente el ADR si se satisfacen ciertas condiciones adicionales.

20

En el modo Premium, se codifica la trama como de HR insonora si se satisface la siguiente condición

25 
$$(\bar{r}_x < th_1) \text{ Y } (e_t < th_2) \text{ Y } (dE < th_3)$$

en la que  $th_1 = 0,5$ ,  $th_2 = 1$  y

$$th_3 = \begin{cases} -4 & \text{para } \bar{E}_t > 34 \\ 0 & \text{para } 21 < \bar{E}_t \leq 34 \\ 4 & \text{en caso contrario} \end{cases}$$

30 En la decisión de la actividad de voz, se usa una decisión de persistencia. Por ello, después de períodos de voz activa, cuando el algoritmo decide que la trama es una trama de voz inactiva, se fija un VAD local en cero pero el marcador VAD real se fija a cero solamente después de un cierto número de tramas transcurridas (el período de persistencia). Esto evita el recorte de la voz desplazada. En ambos modos Estándar y Económico, si el VAD local es cero, la trama se clasifica como una trama insonora.

En el modo Estándar, la trama se codifica como HR insonora si el VAD local = 0 o si se satisface la siguiente condición:

35 
$$(\bar{r}_x < th_4) \text{ Y } (e_t < th_5) \text{ Y } ((dE < th_6) \text{ O } (E_{rel} < th_7))$$

en la que  $th_4 = 0,695$ ,  $th_5 = 4$ ,  $th_6 = 40$  y  $th_7 = -14$ .

En el modo Económico, la trama se declara como una trama insonora si el VAD local = 0, O si se satisface la siguiente condición:

$$(\bar{r}_x < th_8) \text{ Y } (e_t < th_9) \text{ Y } ((dE < th_{10}) \text{ O } (E_{rel} < th_{11}))$$

en la que  $th_8 = 0,695$ ,  $th_9 = 4$ ,  $th_{10} = 60$  y  $th_{11} = -14$ .

En el modo Económico, las tramas insonora se codifican normalmente como de HR insonora. Sin embargo, se pueden codificar también como de QR insonora si se satisfacen también las siguientes condiciones adicionales: si la última trama es o bien una trama insonora o bien una trama de ruido de fondo y si al final de la trama la energía se concentra en las altas frecuencias y no se detecta el comienzo sonoro potencial en la anticipación cuando la trama se codifica como de QR insonora. Las dos últimas condiciones se detectan como:

$$(r_x(2) < th_{12}) \text{ Y } (e_{tend}(1) < th_{13}) \text{ en la que } th_{12} = 0,73, th_{13} = 3.$$

Nótese que  $r_x(2)$  es la correlación normalizada en la anticipación y  $e_{tend}(1)$  es la tendencia en el segundo análisis espectral que abarca el final de la trama y la anticipación.

Naturalmente, se pueden usar otros procedimientos distintos al procedimiento 200 para la discriminación de una trama insonora.

### **Discriminación de tramas de voz sonora estables**

En el caso de los modos Estándar y Económico, las tramas sonoras estables se pueden codificar usando el tipo de codificación HR sonora.

El tipo de codificación HR sonora hace uso de la modificación de la señal para una codificación eficiente de tramas sonoras estables.

Las técnicas de modificación de la señal ajustan el tono de la señal a un contorno de retardo predeterminado. La predicción a largo plazo mapea la señal de excitación pasada con la subtrama presente usando este contorno de retardo y escalado mediante un parámetro de ganancia. El contorno de retardo se obtiene directamente mediante la interpolación entre dos estimaciones de tono en bucle abierto, la primera obtenida en la trama previa y la segunda en la trama actual. La interpolación da un valor de retardo para cada instante de tiempo de la trama. Después de que está disponible el contorno de retardo, se ajusta el tono en la subtrama a ser codificada en la actualidad para seguir este contorno artificial mediante deformación, cambiando la escala de tiempo de la señal. En una deformación discontinua [1, 4, 5], se desplaza un segmento de señal bien a la izquierda o bien a la derecha sin alterar la longitud del segmento. La deformación discontinua requiere un procedimiento para el manejo del solape resultante o la pérdida de partes de la señal. Para reducir las artificiosidades en estas operaciones, el cambio tolerado en la escala de tiempo se mantiene pequeño. Más aún, la deformación se realiza típicamente usando la señal residual de LP o la señal de voz ponderada para reducir las distorsiones resultantes. El uso de estas señales en lugar de la señal de voz facilita también la detección de los pulsos de tono y regiones de baja potencia entre medias de ellos y por ello la determinación de los segmentos de señal para la deformación. La señal de voz modificada real se genera mediante filtrado inverso.

Después de que se realiza la modificación de la señal para la subtrama presente, se puede proceder a la codificación en una forma convencional excepto en que la excitación del libro de códigos adaptativo se genera usando el contorno de retardo predeterminado.

En la presente realización ilustrativa, la modificación de la señal se realiza sincronizadamente en tono y trama, esto es, adaptando cada vez un segmento del ciclo de tono en la trama actual de modo que una trama de voz posterior comience en perfecta alineación de tiempo con la señal original. Los segmentos del ciclo de tono están limitados por los límites de la trama. Esto impide que el desplazamiento de tiempo se traslade a través de los límites de la trama simplificando la implementación del codificador y reduciendo un riesgo de artificiosidades en la señal de voz modificada. Esto simplifica también la operación con tasa de bits variable entre tipos de codificación con la modificación de la señal habilitada o inhabilitada, dado que cada nueva trama comienza en alineación de tiempo con la señal original.

Como se ilustra en la Figura 2, si una trama no se clasifica como trama de voz inactiva ni es una trama insonora entonces se comprueba si es una trama sonora estable (etapa 110). La clasificación de tramas sonoras estables se realiza usando un enfoque de bucle cerrado en conjunto con el procedimiento de modificación de señal usado para la codificación de tramas sonoras estables.

La Figura 4 ilustra un procedimiento 300 para la discriminación de tramas sonoras estables de acuerdo con una realización ilustrativa de un cuarto aspecto de la presente invención.

Los subprocedimientos en la modificación de la señal producen indicadores que cuantifican el rendimiento alcanzable para una previsión a largo plazo en la trama actual. Si cualquiera de estos indicadores está fuera de sus límites permitidos, el procedimiento de modificación de la señal se finaliza por uno de los bloques lógicos. En este

caso, la señal original se preserva intacta y la trama no se clasifica como trama sonora estable. Esta lógica integrada permite maximizar la calidad de la señal de voz modificada después de la modificación de la señal y la codificación a una tasa de bits baja.

- 5 El procedimiento de búsqueda del pulso de tono de la etapa 302 produce varios indicadores sobre la periodicidad de la trama actual. De ahí que el bloque lógico que le sigue es un componente importante de la lógica de clasificación. Se observa la evolución de la longitud del ciclo de tono. El bloque lógico compara la distancia de las posiciones del pulso de tono detectado contra la estimación del tono en bucle abierto interpolado, así como contra la distancia de los pulsos de tono detectados previamente. El procedimiento de modificación de la señal se finaliza si la diferencia de la estimación del tono en bucle abierto o las longitudes del ciclo del tono previas son demasiado largas.
- 10 La selección del contorno de retardo en la etapa 304 da información adicional sobre la evolución de los ciclos de tono y la periodicidad de la trama de voz actual. El procedimiento de modificación de la señal se continúa desde este bloque si se completa la condición  $|d_n - d_{n-1}| < 0,2 d_n$ , en la que  $d_n$  y  $d_{n-1}$  son retardos de tono en las tramas presente y pasada. Esto significa esencialmente que sólo se tolera un cambio de retardo pequeño para la clasificación de la trama presente como sonora estable.
- 15 Cuando las tramas sometidas a la modificación de la señal se codifican con una tasa de bits baja, la forma de los elementos del ciclo de tono se mantiene similar a través de la trama para permitir un modelizado de la señal fidedigno mediante previsión a largo plazo y por ello la codificación con una tasa de bits baja sin degradar la calidad subjetiva. En la etapa de modificación de la señal 306, se puede cuantificar la similitud de segmentos sucesivos mediante la correlación normalizada entre el segmento actual y la señal objetivo en el desplazamiento óptimo. El desplazamiento de los segmentos del ciclo de tono que maximizan su correlación con la señal objetivo mejora la periodicidad y produce una alta ganancia de previsión a largo plazo si la modificación de la señal es útil. El éxito del procedimiento se garantiza por el requisito de que todos los valores de correlación deben ser más grandes que un umbral predefinido. Si esta condición no se cumple para todos los segmentos, el procedimiento de modificación de la señal se finaliza y la señal original se mantiene intacta. En general, se puede permitir un intervalo de umbral de ganancia ligeramente más bajo en voces masculinas con igual rendimiento de la codificación. Los umbrales de ganancia se pueden cambiar en diferentes modos operativos del códec VBR para ajustar el uso de los modos de codificación que aplican la modificación de la señal y por ello cambiar la tasa de bits promedio objetivo.
- 20
- 25

Como se ha descrito anteriormente en el presente documento, la lógica de selección de tasa completa de acuerdo con el procedimiento 100 comprende tres etapas, discriminando cada una de ellas una clase de señal específica.

30 Una de las etapas incluye el algoritmo de modificación de la señal como su parte integral. Primero, un VAD discrimina entre tramas de voz activa e inactiva. Si se detecta una trama de voz inactiva, el procedimiento de clasificación acaba dado que la trama se considera como ruido de fondo y se codifica, por ejemplo, con un generador de ruido de confort. Si se detecta una trama de voz activa, la trama se somete a la segunda etapa dedicada a discriminar tramas insonoras. Si la trama se clasifica como una señal de voz insonora, la cadena de clasificación finaliza y la trama se codifica con un modo dedicado a tramas insonoras. Como última etapa, la trama de voz se procesa a través del procedimiento de modificación de la señal propuesto que habilita la modificación si se verifican las condiciones descritas anteriormente en esta sección. En este caso, la trama se clasifica como trama sonora estable, el tono de la señal original se ajusta a un contorno de retardo artificial, bien definido y la trama se codifica usando un modo específico utilizado para estos tipos de tramas. En caso contrario, la trama es probable que contenga un segmento de voz no fija tal como un comienzo sonoro o señal de voz sonora que evoluciona rápidamente. Estas tramas requieren típicamente un modelo de codificación más genérico. Estas tramas se codifican usualmente con un tipo de codificación FR Genérica. Sin embargo, si la energía relativa de la trama es más baja que un cierto umbral entonces estas tramas se pueden codificar con un tipo de codificación HR Genérica para reducir adicionalmente el ADR.

35

40

#### 45 **Codificación de la voz y selección de la tasa para sistemas de VBR de modo múltiple CDMA**

Los procedimientos para la selección de la tasa y codificación digital del sonido con sistemas de VBR de modo múltiple CDMA que pueden operar en el Conjunto de Tasas II se describirán ahora de acuerdo con las realizaciones ilustradas de la presente invención.

50 El códec descrito se basa en el códec de voz de banda ancha tasa múltiple adaptativo (AMR-WB) que se seleccionó recientemente por la ITU-T (Unión Internacional de Telecomunicaciones - Sector de Normalización de Telecomunicaciones) para varios servicios de voz de banda ancha y por el 3GPP (proyecto de asociación para la tercera generación) para sistemas inalámbricos GSM y W- CDMA de tercera generación. El códec AMR-WB consiste en 9 tasas de bits, concretamente 6,6, 8,85, 12,65, 14,25, 15,85, 18,25, 19,85, 23,05 y 23,85 kbit/s. Un códec VBR controlado en origen basado en AMR-WB para sistemas CDMA permite la habilitación de la interoperación entre

55 CDMA y otros sistemas que usan el códec AMR-WB. La tasa de bits del AMR-WB de 12,65 kbit/s, que es la tasa más cercana que se puede acomodar en los 13,3 kbit/s de tasa completa del Conjunto de Tasas II se puede usar como la tasa común entre el códec VBR y el AMR-WB que habilita la interoperabilidad sin la necesidad de transcodificación (lo que degrada la calidad de la voz). Se proporcionan tipos de codificación de tasas más bajas específicamente para la solución de banda ancha VBR de CDMA para habilitar la operación eficiente en la trama de trabajo del Conjunto de Tasas II. El códec puede operar entonces en unos pocos modos específicos de CDMA

60

usando todas las tasas pero habrá un modo que habilita la interoperabilidad con sistemas que usan el códec AMR-WB.

Los procedimientos de codificación de acuerdo con las realizaciones de la presente invención se resumen en la Tabla 1 y se denominan generalmente tipos de codificación.

5 **Tabla 1.** Tipos de codificación usados en las realizaciones ilustrativas con las correspondientes tasas de bit.

Tipo de Codificación	Tasa de Bit [kbit/s]	Bits / trama de 20 ms
FR Genérico	13,3	266
FR Interoperable	13,3	266
HR sonora	6,2	124
HR insonora	6,2	124
HR Interoperable	6,2	124
HR Genérico	6,2	124
QR insonora	2,7	54
QR de CNG	2,7	54
ER de CNG	1,0	20

Los tipos de codificación a tasa completa (FR) se basan en el códec estándar AMR-WB a 12,65 kbit/s. El uso de la tasa de 12,65 kbit/s del códec AMR-WB habilita el diseño de un códec de tasa de bit variable para el sistema CDMA capaz de interoperar con otros sistemas usando la norma del códec AMR-WB. Se añaden 13 bits extra por trama para acomodarse a los 13,3 kbit/s de tasa completa del Conjunto de Tasas II del CDMA. Estos bits se usan para mejorar la robustez del códec en caso de tramas borradas y marcan esencialmente la diferencia entre los tipos de códigos FR Genérico y FR Interoperable (no se usan en el FR Interoperable). Los tipos de codificación FR se basan en el modelo de previsión lineal con excitación por código algebraica (ACELP) optimizado para señales de voz de banda ancha generales. Opera sobre las tramas de voz de 20 ms con una frecuencia de muestreo de 16 kHz. Antes de un procesamiento adicional, se rebaja el muestreo de la señal de entrada a 12,8 kHz de frecuencia de muestreo y se preprocesa. Los parámetros del filtro LP se codifican una vez por trama usando 46 bits. A continuación la trama se divide en cuatro subtramas en las que se codifican una vez por subtrama los índices y ganancias del libro de códigos adaptativo y fijo. El libro de códigos fijo se construye usando una estructura del libro de códigos algebraico en el que las 64 posiciones en una subtrama se dividen en 4 pistas de posiciones intercaladas y en las que se colocan en cada pista 2 pulsos con signo. Los dos pulsos por pista se codifican usando 9 bits dando un total de 36 bits por subtrama. Se pueden hallar más detalles sobre el códec AMR-WB en la referencia [1]. Las asignaciones de bits para los tipos de codificación FR se dan en la Tabla 2.

Tabla 2. Asignación de bits del Conjunto de Tasas II de CDMA2000 de tasa completa Genérica e Interoperable basada en la norma AMR-WB a 12,65 kbit/s.

Parámetro	Bits por trama	
	FR Genérico	FR Interoperable
Información de clase	-	-
Bit VAD	-	1
Parámetros LP	46	46
Retardo de tono	30	30
Filtrado de tono	4	4
Ganancias	28	28
Libro de códigos algebraico	144	144
Bits de protección FER	14	-
Bits no usados	-	13
Total	266	266

En el caso de tramas sonoras estables, se usa la codificación de media tasa sonora. La asignación de bits de la media tasa sonora se da en la Tabla 3. Dado que las tramas a codificar en este modo de comunicación son característicamente muy periódicas, una tasa de bits más baja es sustancialmente suficiente para mantener una buena calidad subjetiva comparada por ejemplo con las tramas de transición. Se usa la modificación de la señal lo que permite una codificación eficiente de la información de retardo usando sólo 9 bits por cada trama de 20 ms ahorrando una proporción considerable del presupuesto de bits para otros parámetros de codificación de la señal. En la modificación de la señal, la señal se fuerza a seguir un cierto contorno de tono que se puede transmitir con 9 bits por trama. El buen rendimiento de la previsión a largo plazo permite usar sólo 12 bits por subtrama de 5 ms para la excitación del libro de códigos fijo sin sacrificar la calidad de la voz subjetiva. El libro de códigos fijo es un libro de códigos algebraico y comprende dos pistas con un pulso en cada una, mientras que cada pista tiene 32 posiciones posibles.

Tabla 3. Asignación de bits de tasa media Genérico, Sonora, Insonora de acuerdo con el Conjunto de Tasas II de CDMA2000

Parámetro	Bits por trama			
	HR Genérica	HR Sonora	HR Insonora	HR Interoperable
Información de clase	1	3	2	3
Bit VAD	-	-	-	1
Parámetros LP	36	36	46	46
Retardo de tono	13	9	-	30
Filtrado de tono	-	2	-	4
Ganancias	26	26	24	28
Libro de códigos algebraico	48	48	52	-
Bits de protección FER	-	-	-	-
Bits no usados	-	-	-	12
Total	124	124	124	124

En el caso de tramas insonoras, el libro de códigos adaptativo (o libro de códigos de tono) no se usa. Se usa un libro de códigos gaussiano de 13 bits en cada subtrama en el que la ganancia del libro de códigos se codifica con 6 bits por subtrama. Se ha de tener en cuenta que en casos en los que la tasa de bits media necesita reducirse adicionalmente, se puede usar cuarto de tasa insonora en caso de tramas insonoras estables.

Se usa un modo de tasa media genérica para segmentos de baja energía. Este modo HR genérico puede usarse también en operación a media tasa máxima como se explicará posteriormente. La asignación de bits del HR genérico se muestra en la Tabla 3 anterior.

Como ejemplo, para información de clasificación para los diferentes codificadores HR, en caso del HR Genérico, se usa 1 bit para indicar si la trama es un HR Genérico u otro HR. En caso de HR Insonoro, se usan 2 bits para la clasificación: el primer bit para indicar que la trama no es un HR Genérico y el segundo bit para indicar que es una HR Insonora y no una HR Sonora o una HR Interoperable (lo que se explicará posteriormente). En el caso de HR Sonora, se usan 3 bits. Los 2 primeros bits indican que la trama no es HR Genérica o Insonora y el tercer bit indica si la trama es HR Insonora o Interoperable.

En el modo Económico, la mayor parte de las tramas insonoras se pueden codificar usando un codificador QR Insonora. En este caso, los índices del libro de códigos gaussiano se generan aleatoriamente y la ganancia se codifica con solamente 5 bits por subtrama. También, se cuantifican los coeficientes del filtro LP con una tasa de bits más baja. Se usa 1 bit para la discriminación entre los dos tipos de codificación de cuarto de tasa: QR Insonora y QR de CNG. La asignación de bits para tipos de codificación insonoros se da en 6.

El tipo de codificación HR Interoperable permite abordar las situaciones en las que el sistema CDMA impone el HR como la tasa máxima para una trama particular mientras que la trama se ha clasificado como de tasa completa. El HR se deriva directamente del codificador de tasa completa mediante la retirada de los índices del libro de códigos fijo después de que la trama se haya codificado como una trama de tasa completa (Tabla 4). En el lado del decodificador, los índices del libro de código fijo se pueden generar aleatoriamente y el decodificador funcionará como si está en tasa completa. Este diseño tiene la ventaja de que minimiza el impacto del modo de tasa media forzado durante una operación libre en tándem entre el sistema CDMA y otros sistemas que usan la norma AMR-WB (tal como el sistema GSM de móviles o el sistema inalámbrico de tercera generación W-CDMA). Como se ha mencionado anteriormente, el tipo de codificación FR Interoperable o QR de CNG se usan para una operación libre en tándem (TFO) con AMR-WB. El enlace en la dirección desde CDMA 2000 a un sistema que usa el códec AMR-WB, cuando la subcapa del multiplexor indica una solicitud para un modo de tasa media, el códec VMR-WB usará el tipo de codificación HR Interoperable. En la interfaz del sistema, cuando se recibe una trama HR Interoperable, los índices del libro de códigos algebraico generados aleatoriamente se añaden al flujo de bits para sacar una tasa de 12,65 kbit/s. El decodificador AMR-WB en el lado del receptor interpretará esto como una trama de 12,65 kbit/s ordinario. En la otra dirección, es decir en un enlace desde un sistema que usa el códec AMR-WB a CDMA2000, si en la interfaz del sistema se recibe la solicitud para tasa media, entonces los índices del libro de códigos algebraico se retiran y se añaden los bits de modo que indiquen un tipo de trama HR Interoperable. El decodificador en el lado del CDMA 2000 opera como un tipo de codificación HR Interoperable, que es parte de la solución de codificación VMR-WB. Sin el HR Interoperable, un modo de tasa media forzado se interpretaría como un borrado de la trama.

La técnica de Generación de Ruido de Confort (CNG) se usa para procesar las tramas de voz inactivas. El tipo de codificación de octavo de tasa (ER) de CNG se usa para codificar tramas de voz inactivas cuando se opera dentro del sistema CDMA. En una llamada en la que se requiere la interoperación con la norma de codificación de la voz AMR-WB, el ER de CNG no se puede usar siempre dado que su tasa de bits es más baja que la tasa de bits necesaria para transmitir la información actualizada para el decodificador del CNG en AMR-WB [3]. En este caso, se usa el QR de CNG. Sin embargo, el códec AMR-WB opera a menudo en un modo de transmisión discontinua (DTX). Durante la transmisión discontinua, la información de ruido de fondo no se actualiza en cada trama. Típicamente sólo

5 se transmite una trama de cada 8 tramas de voz inactivas consecutivas. Este trama de actualización se denomina como el Descriptor del Silencio (SID) [4]. La operación de DTX no se usa en el sistema CDMA en el que se codifica cada trama. Consecuentemente, sólo se necesita codificar las tramas SID con el QR de CNG en el lado del CDMA y las tramas restantes se pueden codificar aún con el ER de CNG para disminuir la ADR ya que no se usarán por la otra parte del AMR-WB. En la codificación del CNG, sólo se codifican una vez por trama los parámetros del filtro LP y una ganancia. La asignación de bits para el QR de CNG se da en la Tabla 4 y el del ER de CNG se da en la Tabla 5.

Tabla 4. Asignación de bits para los tipos de codificación QR Insonora y QR de CNG

Parámetro	QR insonora	QR de CNG
Bits de selección	1	1
Parámetros LP	32	28
Ganancias	20	6
Bits no usados	1	19
Total	54	54

10 Tabla 5. Asignación de bits para el ER de CNG

parámetros	Bits/trama del ER de CNG
Parámetros LP	14
Ganancia	6
No usados	-
Total	20

**Clasificación de la señal y selección de la tasa en el Modo Premium**

15 Se ilustra en la Figura 5 un procedimiento 400 para la codificación digital de una señal de sonido de acuerdo con una segunda realización ilustrativa del segundo aspecto de la presente invención. Se ha de tener en cuenta que el procedimiento 400 es una aplicación específica del procedimiento 100 en el modo Premium, que se proporciona para una calidad de la voz sintetizada máxima dada la tasa de bits disponible (se debería notar que el caso en el que el sistema limita la tasa disponible máxima para una trama particular se describirá en una subsección separada). Consecuentemente, la mayor parte de las tramas de voz activas se codifican con tasa completa, es decir a 13,3 kbit/s.

20 De modo similar al procedimiento 100 ilustrado en la Figura 2, un detector de actividad de voz (VAD), discrimina entre tramas de voz activas e inactivas (etapa 102). El algoritmo VAD puede ser idéntico para todos los modos de operación. Si se detecta una trama de voz inactiva (señal de ruido de fondo) entonces el procedimiento de clasificación se detiene y la trama se codifica con el tipo de codificación ER de CNG a 1,0 kbit/s de acuerdo con el Conjunto de Tasas II del CDMA (etapa 402). Si se detecta una trama de voz activa, la trama se somete a un  
 25 segundo clasificador dedicado a discriminar tramas insonoras (etapa 404). Dado que el Modo Premium está indicado para la mejor calidad posible, la discriminación de la trama insonora es muy severa y solamente se seleccionan tramas insonoras altamente estacionarias. Las reglas de clasificación de insonoridad y los umbrales de decisión son como los dados anteriormente. Si el segundo clasificador clasifica la trama como una señal de voz insonora, el procedimiento de clasificación se detiene y la trama se codifica usando el tipo de codificación HR insonora (etapa  
 30 408) optimizado para señales insonoras (6,2 kbit/s de acuerdo con el Conjunto de Tasas II del CDMA). Todas las otras tramas se procesan con el tipo de codificación FR genérico, basándose en la norma AMR-WB a 12,65 kbit/s (etapa 406).

**Clasificación de la señal y selección de la tasa en el Modo Estándar**

35 Se ilustra en la Figura 6 un procedimiento 500 para la codificación digital de una señal de sonido de acuerdo con una tercera realización ilustrativa del segundo aspecto de la presente invención. El procedimiento 500 permite la clasificación de una señal de voz y su codificación en el modo Estándar.

40 En la etapa 102, un VAD discrimina entre tramas activas e inactivas de voz. Si se detecta una trama de voz inactiva entonces el procedimiento de clasificación se detiene y la trama se codifica como una trama ER de CNG (etapa 510). Si se detecta una trama de voz activa, la trama se somete a un clasificador de segundo nivel dedicado a discriminar tramas insonoras (etapa 404). Las reglas de clasificación y los umbrales de decisión de insonoridad se han descrito anteriormente. Si el clasificador de segundo nivel clasifica la trama como una señal de voz insonora, el procedimiento de clasificación se detiene y la trama se codifica con un tipo de codificación HR insonora (etapa 508). En caso contrario, la trama de la voz se pasa a través del módulo de clasificación de "sonora estable" (etapa 502). La discriminación de las tramas sonoras es una característica inherente al algoritmo de modificación de la señal  
 45 como se ha descrito anteriormente en el presente documento. Si la trama es adecuada para modificación de la señal, se clasifica como trama sonora estable y se codifica con el tipo de codificación HR sonora (etapa 506) en un módulo optimizado para las señales sonoras estables (6,2 kbit/s de acuerdo con el Conjunto de Tasas II del CDMA). En caso contrario la trama es probable que contenga un segmento de voz no estacionario tal como una señal de voz

de inicio sonora o sonora evolucionando rápidamente. Esas tramas requieren típicamente una alta tasa de bits para mantener una buena calidad subjetiva. Sin embargo, si la energía de la trama es más baja que un cierto umbral entonces la trama se pueden codificar con un tipo de codificación HR genérico. Por ello, si en la etapa 512, el clasificador de cuarto nivel detecta una señal de baja energía la trama se codifica usando el HR genérico (etapa 514). El otro caso, la trama de la voz se codifica como una trama de FR Genérica (13,3 kbit/s de acuerdo con el Conjunto de Tasas II del CDMA) (etapa 504).

### **Clasificación de la señal y selección de la tasa en el Modo Económico**

Se ilustra en la Figura 6 un procedimiento 600 para la codificación digital de una señal de sonido de acuerdo con una cuarta realización ilustrativa del primer aspecto de la presente invención. El procedimiento 600 que es un procedimiento de clasificación de cuarto nivel, permite la clasificación de una señal de voz y su codificación en el modo Económico.

El Modo Económico permite la máxima capacidad del sistema mientras que sigue produciendo un ancho de banda de voz de alta calidad. La lógica de determinación de la tasa es similar al modo Estándar con la sección de que también se usa el tipo de codificación QR Insonora y se reduce el uso del FR Genérico.

Primero, en la etapa 102, un VAD discrimina entre tramas activas e inactivas de voz. Si se detecta una trama de voz inactiva entonces el procedimiento de clasificación se detiene y la trama se codifica como una trama ER de CNG (etapa 402). Si se detecta una trama de voz activa, la trama se somete a un segundo clasificador dedicado a discriminar todas las tramas insonoras (etapa 106). Las reglas de clasificación de insonoridad y los umbrales de decisión se han descrito anteriormente. Si el segundo clasificador clasifica la trama como una señal de voz insonora, la trama de voz se pasa a un clasificador de tercer nivel (etapa 602). El clasificador de tercer nivel comprueba si la trama está en una transición de sonora a insonora usando las reglas descritas anteriormente. En particular, este clasificador de tercer nivel comprueba si la última trama es o bien una trama insonora o de ruido de fondo y si al final de la trama la energía se concentra en las altas frecuencias y no se detecta un inicio sonoro potencial en la anticipación. Como se ha explicado anteriormente, las últimas dos condiciones se detectan como:

$$(r_x(2) < th_{12}) \text{ Y } (e_{tend}(1) < th_{13}) \text{ con } th_{12} = 0,73, th_{13} = 3,$$

en la que  $r_x(2)$  es la correlación en la anticipación y  $e_{tend}(1)$  es la tendencia en el segundo análisis espectral que abarca el final de la trama y la anticipación.

Si la trama contiene una transición de sonora a insonora, la trama se codifica en la etapa 508 con el tipo de codificación HR Insonora. En caso contrario, la trama de voz se codifica con el tipo de codificación QR Insonora (etapa 604). Las tramas no clasificados como insonoras se pasan a través de un módulo de clasificación "sonoro estable", que es un segundo clasificador de tercer nivel (etapa 110). La discriminación de las tramas sonoras es una característica inherente al algoritmo de modificación de la señal como se ha explicado antes. Si la trama es adecuada para modificación de la señal, se clasifica como trama sonora estable y se codifica con HR sonora en la etapa 506. De modo similar al modo estándar, las tramas restantes (no clasificadas como insonoras o sonoras estables) se comprueban con respecto al contenido de energía bajo, si se detecta una señal de energía baja en la etapa 512, la trama se codifica en la etapa 514 usando el HR Genérico. En otro caso, la trama de voz se codifica como una trama de FR Genérico (13,3 kbit/s de acuerdo con el Conjunto de Tasas II del CDMA) (etapa 504).

### **Clasificación de la señal y selección de la tasa en el Modo Interoperable**

Se ilustra en la Figura 8 un procedimiento 700 para la codificación digital de una señal de sonido de acuerdo con una quinta realización ilustrativa del segundo aspecto de la presente invención. El procedimiento 700 permite la clasificación de una señal de voz y la codificación en el modo Interoperable.

El modo Interoperable permite una operación libre en tándem entre el sistema CDMA y otros sistemas que usan la norma AMR-WB a 12,65 kbit/s (o tasas menores). En ausencia de la limitación de tasa impuesta por el sistema CDMA solamente se usan el FR Interoperable y los Generadores de Ruido de Confort.

Primero, en la etapa 102, un VAD discrimina entre tramas activas e inactivas de voz. Si se detecta una trama de voz inactiva, se toma una decisión en la etapa 702 si se debería codificar la trama como una trama SID. Como se ha mencionado anteriormente, la trama SID sirve para actualizar los parámetros del CNG en el lado del AMR-WB durante la operación DTX [4]. Típicamente, sólo una de cada 8 tramas de voz inactivas se codifica durante los periodos de silencio. Sin embargo, después de un segmento de voz activa, la actualización SID se debe enviar ya en la 4ª trama (véase la referencia [4] para más detalles). Dado que el ER no es suficiente para codificar una trama SID, las tramas SID se codifican con QR de CNG en la etapa 704. Las tramas distintas a las SID inactivas se codifican con el ER de CNG en la etapa 402. En el enlace con dirección desde VMR-WB de CDMA a AMR-WB en una operación libre en tándem (TFO), las tramas ER de CNG se descartan en la interfaz del sistema dado que el AMR-WB no pueda hacer uso de ellas. En la dirección opuesta, esas tramas no están disponibles (el AMR-WB está generando solamente tramas SID) y se declaran como tramas borradas. Todas las tramas de voz activas se procesan con el tipo de codificación FR Interoperable (etapa 706), que es esencialmente la norma de codificación AMR-WB a 12,65 kbit/s.

**Clasificación de la señal y selección de la tasa en operación en Máximo Media Tasa**

Se ilustra en la Figura 9 un procedimiento 800 para la codificación digital de una señal de sonido de acuerdo con una sexta realización ilustrativa del segundo aspecto de la presente invención. El procedimiento 800 permite la clasificación de una señal de voz y la codificación en operación en Máximo Media Tasa para los modos Premium y Estándar.

Como se ha explicado en el presente documento anteriormente, el sistema CDMA impone una tasa de bits máxima para una trama particular. Más a menudo, la tasa de bits máxima impuesta por el sistema se limita a HR. Sin embargo, el sistema puede imponer también tasas más bajas.

Todos las tramas de voz activa que se clasificarían convencionalmente como FR durante la operación normal se codifican ahora usando tipo de codificación HR. El mecanismo de clasificación y selección de tasa clasifica entonces a todos del tipo de tramas sonoras usando HR Sonora (codificado en la etapa 506) y todos los tipos de tramas insonoras usando HR Insonora (codificado en la etapa 408). Todas las tramas restantes se clasificarían como FR durante la operación normal y se codifican usando el tipo de codificación HR genérico en la etapa 514 excepto en el modo Interoperable en el que se usa el tipo de código HR Interoperable (etapa 908 en la Figura 10).

Como se puede ver en la Figura 9, la clasificación de la señal y mecanismo de codificación es similar al de la operación normal en el modo estándar. Sin embargo, se usa el HR genérico (etapa 514) el lugar de la codificación FR genérica (etapa 406 en la Figura 5) y los umbrales usados para discriminar tramas insonora y sonora son más relajados para permitir que se codifiquen tantas tramas como sea posible usando los tipos de codificación HR insonora y HR sonora. Básicamente, los umbrales para el modo económico se usan en el caso de la operación de máxima tasa media en el modo Premium o Estándar.

Se ilustra en la Figura 10 un procedimiento 900 para la codificación digital de una señal de sonido de acuerdo con una séptima realización ilustrativa del primer aspecto de la presente invención. El procedimiento 900 permite la clasificación de una señal de voz y la codificación en operación en Máximo Media Tasa para el modo Económico. El procedimiento 900 en la Figura 10 es similar al procedimiento 600 en la Figura 7 con la excepción de que todas las tramas que se habrían codificado con el FR Genérico se codifican ahora con el HR Genérico (no hay necesidad de clasificación de la trama en energía baja en la operación en máximo media tasa). Se ilustra en la Figura 11 un procedimiento 920 para la codificación digital de una señal de sonido de acuerdo con una octava realización ilustrativa del primer aspecto de la presente invención. El procedimiento 920 permite la clasificación de una señal de voz y la determinación de la tasa en el modo Interoperable durante la operación en máximo media tasa. Dado que el procedimiento 920 es muy similar al procedimiento 700 de la Figura 8, sólo se describirán en el presente documento las diferencias entre los dos procedimientos.

En el caso del procedimiento 920, no se pueden usar tipos de codificación específicos de la señal (HR insonora y HR sonora) dado que no serían comprensibles por la otra parte del AMR-WB y tampoco se puede usar la codificación HR Genérica. En consecuencia, todas las tramas de voz activas en la operación en máximo media tasa se codifican usando el tipo de codificación HR Interoperable.

Si el sistema impone una tasa de bits máxima más baja que la HR, no se proporciona ningún tipo de codificación general para manejarse con esos casos, esencialmente porque esos casos son extremadamente raros y tales tramas se pueden declarar como tramas borradas. Sin embargo, si la tasa de bits máxima está limitada a QR por el sistema y la señal se clasifica como insonora, entonces se puede usar el QR Insonora. Esto es posible sin embargo solamente en modos específicos de CDMA (Premium, Estándar, Económico), dado que la otra parte del AMR-WB es incapaz de interpretar las tramas QR.

**Interoperación eficiente entre AMR-WB y el códec VMR-WB del Conjunto de Tasas II**

Se describirá ahora un procedimiento 1000 para la codificación de una señal de voz para la Interoperación entre los códec AMR-WB y VMR-WB de acuerdo con una realización ilustrativa de un cuarto aspecto de la presente invención con referencia a la Figura 12.

Más específicamente, el procedimiento 1000 habilita la operación libre en tándem entre el códec AMR-WB estándar y el códec VBR controlado en origen diseñado, por ejemplo, para sistemas CDMA2000 (denominado en el presente documento como códec VMR-WB). En un modo interoperable permitido por el procedimiento 1000, el códec VMR-WB hace uso de tasas de bits que se pueden interpretar por el códec AMR-WB y acomodarse aún dentro de las tasas de bits del Conjunto de Tasas II usadas en un códec CDMA, por ejemplo.

Dado que las tasas de bits del Conjunto de Tasas II son la FR 13,3, HR 6,2, QR 2,7 y ER 1,0 kbit/s, entonces las tasas de bits del códec AMR-WB que se pueden usar son 12,65, 8,85 ó 6,6 en la tasa completa y las tramas SID a 1,75 kbit/s en el cuarto de tasa. El AMR-WB a 12,65 kbit/s es la tasa de bits más cercana al FR 13,3 kbit/s del CDMA2000 y se usa como el códec FR en esta realización ilustrativa. Sin embargo, cuando se usa el AMR-WB en sistemas GSM el algoritmo de adaptación del enlace puede disminuir la tasa de bits a 8,85 ó 6,6 kbit/s dependiendo de las condiciones del canal (para asignar más bits a la codificación del canal). Por ello, las tasas de bits de 8,85 y 6,6 kbit/s del AMR-WB pueden ser parte del modo Interoperable que se pueden usar en el receptor CDMA2000 en el



caso de que el sistema GSM decida usar cualquiera de estas tasas de bits. En la realización ilustrativa de la Figura 12, se usan tres tipos de I-FR que corresponden a las tasas AMR-WB a 12,65, 8,85 y 6,6 kbit/s y se indicarán por I-FR-12, I-FR-8 y I-FR-6, respectivamente. En el I-FR-12 hay 13 bits sin utilizar. Los primeros 8 bits se usan para distinguir las tramas I-FR y las tramas FR Genéricas (que usan los bits extra para mejorar la ocultación del borrado de tramas). Los otros 5 bits se usan para señalar los tres tipos de tramas I-FR. En operación ordinaria, se usa del I-FR-12 y se usan las tasas más bajas si se requiere por la adaptación del enlace GSM.

En el sistema CDMA 2000, la tasa de datos media del códec de voz se relaciona directamente con la capacidad del sistema. Por lo tanto alcanzar el ADR más bajo posible con la mínima pérdida en la calidad de la voz es de una importancia significativa. El códec AMR-WB se diseñó principalmente para sistemas celulares GSM e inalámbricos de tercera generación basados en la evolución del GSM. Por ello un modo Interoperable para el sistema CDMA2000 debe dar como resultado un ADR más alto comparado con el códec VBR diseñado específicamente para sistemas CDMA2000. Las razones principales son:

- la falta de un modo de tasa media a 6,2 kbit/s en AMR-WB;
- la tasa de bits del SID en AMR-WB es de 1,75 kbit/s que no se acomoda en el octavo de tasa (ER) del Conjunto de Tasas II;
- la operación VAD/DTX del AMR-WB usa varias tramas de persistencia (codificadas como tramas de voz) para computar el SID\_FIRST.

Un procedimiento para la codificación de una señal de voz para Interoperación entre los códec AMR-WB y VMR-WB permite superar las limitaciones mencionadas anteriormente y da como resultado un ADR reducido del modo Interoperable de modo que es equivalente a los modos específicos CDMA2000 con una calidad de voz comparable. Se describen a continuación los procedimientos para ambas direcciones de operación: codificación VMR-WB - decodificación AMR-WB y codificación AMR-WB - decodificación VMR-WB.

#### **Codificación VMR-WB - decodificación AMR-WB**

Cuando se codifica en el lado del códec VMR-WB de CDMA, no se requiere la operación VAD/DTX/CNG de la norma. El VAD es propio del códec VMR-WB y trabaja exactamente de la misma manera que en los otros modos específicos de CDMA2000, es decir la persistencia VAD usada es justamente tan larga como la necesaria para no perder paradas insonoras y siempre que el marcador VAD = 0 (clasificado como ruido de fondo) la codificación CNG es operativa.

La operación VAD/CNG se realiza para que sea tan próxima como sea posible a la operación DTX AMR. La operación VAD/DTX/CNG en el códec AMR-WB trabaja como sigue. Las siete tramas de ruido de fondo tras un periodo de voz activa se codifican como tramas de voz pero el bit VAD se establece a cero (persistencia DTX). Entonces se envía una trama SID\_FIRST. En una trama SID\_FIRST la señal no se codifica y los parámetros CNG se deducen de la persistencia DTX (las 7 tramas de voz) en el decodificador. Se ha de notar que el AMR-WB no usa la persistencia DTX después de periodos de voz activa que sean más cortos que 24 tramas para reducir la sobrecarga de la persistencia DTX. Después de una trama SID\_FIRST, se envían dos tramas como tramas NO\_DATA (DTX), seguidos por una trama SID\_UPDATE (1,75 kbit/s). Después de eso, se envían 7 tramas NO\_DATA seguidas por una trama SID\_UPDATE y así sucesivamente. Esto continúa hasta que se detecta una trama de voz activa (marcador VAD = 1). [4]

En la realización ilustrativa de la Figura 12, el VAD en el códec VMR-WB no usa la persistencia DTX. La primer trama de ruido de fondo después de un periodo de voz activa se codifica a 1,75 kbit/s y se envía en QR entonces hay 2 tramas codificadas a 1 kbit/S (octavo de tasa) y se envía a continuación otra trama a 1,75 kbit/s en QR. Después de eso, se envían 7 tramas en ER seguidos por una trama en QR y así sucesivamente. Esto corresponde aproximadamente a una operación DTX AMR-WB con la excepción de que no se usa la persistencia DTX para reducir el ADR.

Aunque la operación VAD/CNG en el códec VMR-WB descrito en esta realización ilustrativa está cercana a la operación DTX AMR-WB, se pueden usar otros procedimientos que puedan reducir adicionalmente el ADR. Por ejemplo, las tramas QR de CNG se pueden enviar menos frecuentemente, por ejemplo una vez cada 12 tramas. Adicionalmente, las variaciones de ruido se pueden evaluar en el decodificador y se pueden enviar solamente tramas QR de CNG cuando cambian las características de ruido (no una vez cada 8 ó 12 tramas).

Para superar la limitación de la no existencia de una tasa media a 6,2 kbit/s en el codificador AMR-WB, se proporciona una tasa media Interoperable (I-HR) que incluye la codificación de la trama como una trama de tasa completa y la retirada de los bits que corresponden a los índices del libro de código algebraico (144 bits por trama en AMR-WB a 12,65 kbit/s). Esto reduce la tasa de bits a 5,45 kbit/s que se acomoda en la tasa media del Conjunto de Tasas II de CDMA 2000. Antes de la decodificación, los bits retirados se pueden generar o bien aleatoriamente (por ejemplo usando un generador aleatorio) o pseudo aleatoriamente (por ejemplo mediante la repetición de parte del flujo de bits existente) o en alguna forma predeterminada. Se puede usar el I-HR cuando se señala la solicitud dim-and-burst o en máximo media tasa por el sistema CDMA 2000. Esto evita la declaración de la trama de voz como

una trama perdida. El I-HR se puede usar también por el códec VMR-WB en un modo Interoperable para codificar tramas insonoras o tramas en las que la contribución del libro de códigos algebraico a la calidad de la voz sintetizada es mínima. Esto da como resultado un ADR reducido. Se debería notar que en este caso, el codificador puede elegir tramas para ser codificadas en el modo I-HR y por ello minimizar la degradación en la calidad de la voz producida por el uso de tales tramas.

Como se ilustra en la Figura 12, en la dirección codificación VMR-WB / decodificación AMR-WB, las tramas de voz se codifican con el modo Interoperable del codificador VMR-WB 1002. Que saca una de las tasas de bits posibles siguientes: I-FR para tramas de voz activas (I-FR-12, I-FR-8 o I-FR-6), I-HR en caso de señalización dim-and-burst o, como una opción, para codificar algunas tramas insonoras o tramas en donde la contribución del libro de códigos algebraico a la calidad de la voz sintetizada es mínima, el QR de CNG para codificar tramas de ruido de fondo relevantes (una de cada ocho tramas de ruido de fondo como se ha descrito anteriormente o cuando se detecta una variación en la característica del ruido) y tramas ER de CNG para la mayor parte de las tramas de ruido de fondo (tramas de ruido de fondo no codificados como tramas QR de CNG). En la interfaz del sistema, que está en la forma de una pasarela, se realizan las siguientes operaciones:

Primero, se comprueba la validez de la trama recibida por la pasarela desde el codificador VMR-WB. Si no es una trama válida VMR-WB en el modo Interoperable entonces se envía como un borrado (tipo de pérdida de voz del AMR-WB). La trama se considera inválida por ejemplo si sucede una de las siguientes condiciones:

- Si se recibe la trama todo ceros (usado por la red en casos de blank and burst) entonces la trama se borra;
- En el caso de tramas FR, si los 13 bits del preámbulo no se corresponden a I-FR-12, I-FR-8 o I-FR-6 o si los bits no usados no son cero, entonces la trama se borra. También, el I-FR establece el bit VAD a 1 tal modo que si el bit VAD de la trama recibido no es 1 de la trama se borra;
- En caso de tramas HR, de modo similar al FR, si los bits del preámbulo no se corresponden a I-FR-12, I-FR-8 o I-FR-6 o si los bits no usados no son cero, entonces la trama se borra. Lo mismo para el bit VAD;
- En el caso de tramas QR, si los bits del preámbulo no se corresponden a QR de CNG entonces la trama se borra. Adicionalmente, el codificador VMR-WB fija el bit de SID\_UPDATE a 1 y el modo solicita bits a 0010. Si este no es el caso entonces la trama se borra;
- En el caso de tramas ER, si se recibe la trama ER todo unos entonces la trama se borra. Adicionalmente, el codificador VMR-WB usa el patrón de bits ISF de todos ceros (primeros 14 bits) para señalar tramas en blanco. Si se recibe este patrón entonces la trama se borra.

Si la trama recibida es una trama del modo Interoperable válida se realizan las siguientes operaciones:

- Se envían tramas I-FR al decodificador AMR-WB como tramas de 12,65, 8,8 ó 6,6 kbit/s dependiendo del tipo del I-FR;
- Se envían tramas QR de CNG al decodificador AMR-WB como tramas SID\_UPDATE;
- Se envían tramas ER de CNG al decodificador AMR-WB como tramas NO\_DATA; y
- Las tramas I-HR se traducen a tramas de 12,65, 8,8 ó 6,6 kbit/s (dependiendo del tipo de tramas) mediante la generación de los índices del libro de códigos algebraico perdido en la etapa 1010. Los índices se pueden generar aleatoriamente o mediante la repetición de parte de los bits de codificación existentes o en alguna manera predeterminada. También descarta los bits que indican el tipo de I-HR (los bits usados para distinguir diferentes tipos de medias tasas en el códec VMR-WB).

#### Codificación AMR-WB - decodificación VMR-WB

En esta dirección, el procedimiento 1000 está limitado por la operación DTX AMR-WB. Sin embargo, durante la codificación de la voz activa, hay un bit en el flujo de bits (el 1er bit de datos) que indica el marcador VAD (0 para el período de persistencia DTX, 1 para la voz activa). Por tanto la operación de la pasarela se puede resumir como sigue:

- Las tramas SID\_UPDATE se envían como tramas QR de CNG;
- Las tramas SID\_FIRST y las tramas NO\_DATA se envían como tramas en blanco ER;
- Las tramas borradas (pérdida de voz) se envían como tramas borrados ER;
- La primera trama después de la voz activa con el marcador VAD = 0 (verificado en la etapa 1012) se mantiene como trama FR pero las siguientes tramas con el marcador VAD = 0 se envían como tramas en

blanco ER;

- Si la pasarela recibe en la etapa 1014 una solicitud para una operación en máximo media tasa (señalización a nivel de tramas) mientras se reciben tramas FR, entonces la trama se traduce en una trama I-HR. Esto consiste en la retirada de los bits que corresponden a los índices del libro de códigos algebraico y la adición de los bits de modo que indican el tipo de trama I-HR.

En esta realización ilustrativa, en tramas ER en blanco, los primeros dos bytes se fijan a 0x00 y en las tramas borradas en ER los dos primeros bytes se fijan en 0x04. Básicamente, los primeros 14 bits corresponden a los índices ISF y se reservan dos patrones para indicar tramas en blanco (todos cero) o tramas borradas (todos ceros excepto el bit 14<sup>o</sup> fijado a 1, que es 0x04 en hexadecimal). En el decodificador VMR-WB 1004, cuando se detectan tramas ER en blanco, son procesados por el decodificador CNG mediante el uso de los últimos parámetros CNG recibidos bien. Una excepción es el caso de la primera trama ER en blanco recibida (inicialización del decodificador CNG; no son conocidos aún parámetros CNG antiguos). Dado que la primera trama con el marcador VAD = 0 se transmite como FR, los parámetros desde esta trama así como los últimos parámetros CNG se usan para inicializar la operación CNG. En caso de las tramas borradas ER, el decodificador usa el procedimiento de ocultación usado para tramas borradas.

Nótese que en la realización ilustrada mostrada en la Figura 12, se usan 12,65 kbit/s para tramas FR. Sin embargo, se podrían ir usar igualmente 8,85 y 6,6 kbit/s de acuerdo con un algoritmo de adaptación del enlace que requiere el uso de tasas más bajas en el caso de malas condiciones en el canal. Por ejemplo, para Interoperación entre sistemas CDMA2000 y GSM, el módulo de adaptación del enlace en el sistema GSM puede decidir bajar la tasa de bits a 8,85 ó 6,6 kbit/s en caso de malas condiciones en el canal. En este caso, estas tasas de bits más bajas necesitan ser incluidas en la solución VMR-WB de CDMA.

**Códec VMR-WB de CDMA operando en Conjunto de Tasas I**

En Conjunto de Tasas I, las tasas de bit usadas son 8,55 kbit/s para FR, 4,0 kbit/s para HR, 2,0 kbit/s para QR y 800 bit/s para ER. En este caso solamente se puede usar el códec AMR-WB a 6,6 kbit/s en FR y las tramas CNG se pueden enviar tanto en QR (SID\_UPDATE) como en ER para otras tramas de fondo (similar a la operación del Conjunto de Tasas II descrita anteriormente). Para superar la limitación de la baja calidad de la tasa de 6,6 kbit/s, se proporciona una tasa de 8,55 kbit/s que es interoperativa con la tasa de bits de 8,85 kbit/s del códec AMR-WB. Será denominada como Conjunto de Tasas I Interoperable FR (I-FR-I). La asignación de bits de la tasa de 8,85 kbit/s y dos configuraciones posibles del I-FR-I se muestran en la Tabla 6.

Tabla 6. Asignación de bits de los tipos de codificación I-FR-I en la configuración Conjunto de Tasas I.

Parámetro	AMR-WB a 8,85 kbit/s Bits/Trama	I-FR-I a 8,55 kbit/s (configuración 1) Bits/Trama	I-FR-I a 8,55 kbit/s (configuración 2) Bits/Trama
Bits en modo Tasa media	-	-	-
Marcador VAD	1	0	0
LP	46	41	46
Parámetros			
Retardo de tono	26 = 8 + 5 + 8 + 5	26	26
Ganancias	24 = 6 + 6 + 6 + 6	24	24
Libro de códigos algebraico	80 = 20 + 20 + 20 + 20	80	75
Total	177	171	171

En el I-FR-I, el bit del marcador VAD y 5 bits adicionales se retiran para tener una tasa de 8,55 kbit/s. Los bits retirados se pueden introducir fácilmente en el decodificador o interfaz del sistema de modo que se puedan usar los 8,85 kbit/s del decodificador. Se pueden usar varios procedimientos para retirar los 5 bits en una forma que produzcan un pequeño impacto sobre la calidad de la voz. En la Configuración 1 mostrada en la Tabla 6, los 5 bits se retiran de la cuantificación del parámetro de previsión lineal (LP). En AMR-WB, se usan 46 bits para cuantificar los parámetros LP en el dominio ISP (par en el espectro de admisión) (usando la retirada de la media y trasladando la previsión promedio). El vector residual (tras la predicción) ISP de 16 dimensiones se cuantifica usando la cuantificación del vector multietapa con división. El resto se divide en 2 subvectores de dimensiones 9 y 7, respectivamente. Los 2 subvectores se cuantifican en dos etapas. En la primera etapa cada subvector se cuantifica con 8 bits. Los vectores de error de cuantificación se reparten en la segunda etapa en 3 y 2 subvectores, respectivamente. Los subvectores de la segunda etapa son de dimensión 3, 3, 3, 3 y 4 y se cuantifican con 6, 7, 7, 5 y 5 bits, respectivamente. En el modo I-FR-I propuesto, los 5 bits de los últimos subvectores de la segunda etapa se retiran. Estos tienen el menor impacto dado que corresponden a la parte de alta frecuencia del espectro. La retirada de estos 5 bits se realiza en la práctica mediante la fijación del índice del último subvector de la segunda etapa en un cierto valor que no necesita transmitirse. El hecho de que se fije este índice de 5 bits es tenido en cuenta fácilmente durante la cuantificación en el decodificador VMR-WB. El índice fijo se añade o bien en la interfaz del sistema (es decir durante la operación codificador VMR-WB/decodificador AMR-WB) o en el decodificador (es decir durante la

operación codificador AMR-WB/decodificador VMR-WB). En esta forma se usa el decodificador AMR-WB a 8,85 kbit/s para decodificar la trama en Conjunto de Tasas II-FR.

En una segunda configuración de la realización ilustrada los 5 bits se retiran de los índices del libro de códigos algebraico. En el AMR-WB a 8,85 kbit/s, una trama se divide en cuatro subtramas de 64 muestras. El libro de códigos de excitación algebraica consiste en la división de la subtrama en 4 pistas de 16 posiciones y la colocación de un pulso con signo en cada pista. Cada pulso se codifica con 5 bits: 4 bits para la posición y 1 bit para el signo. Por ello, para cada subtrama, se usa un libro de códigos algebraico de 20 bits. Una forma de retirar los cinco bits es retirar un pulso a partir de una cierta subtrama. Por ejemplo, el 4º pulso en la 4ª posición de pista en la 4ª subtrama. En el codificador VMR-WB, este pulso se puede fijar en un valor predeterminado (posición y signo) durante la búsqueda en el libro de códigos. Este índice de pulso conocido se puede añadir entonces en la interfaz del sistema y enviar al decodificador AMR-WB. En la otra dirección, el índice de este pulso se retira en la interfaz del sistema y, en el decodificador VMR-WB del CDMA, el índice del pulso se puede generar aleatoriamente. Se pueden usar también otros procedimientos para retirar estos bits.

Para manejar una solicitud dim-and-burst o con máximo media tasa por parte del sistema CDMA2000, se proporciona un modo HR Interoperable también para el códec Conjunto de Tasas I (I-HR-I). De modo similar al caso del Conjunto de Tasas II, se pueden retirar algunos bits en la interfaz del sistema durante la operación de codificación AMR-WB/decodificación VMR-WB o generarse en la interfaz del sistema durante la codificación VMR-WB/decodificación AMR-WB. Se muestra en la Tabla 7 una asignación de bits de la tasa de 8,85 kbit/s y un ejemplo de la configuración del I-HR-I.

Tabla 7. Asignación de bits de ejemplo del tipo de codificación I-HR-I en la configuración Conjunto de Tasas I.

Parámetro	AMR-WB a 8,85 kbit/s Bits/Trama	I-HR-I a 4,0 Bits/Trama
Bits del modo tasa media	-	-
Marcador VAD	1	0
Parámetros LP	46	36
Retardo de tono	26 = 8 + 5 + 8 + 5	20
Ganancias	24 = 6 + 6 + 6 + 6	24
Libro de códigos algebraico	80 = 20 + 20 + 20 + 20	0
Total	177	80

En el modo I-HR-I propuesto, los 10 bits de los últimos 2 subvectores de la segunda etapa en la cuantificación de los parámetros del filtro LP se retiran o generan en la interfaz del sistema en una forma similar al Conjunto de Tasas II descrita anteriormente. El retardo de tono se codifica sólo con resolución entera y con una asignación de bits de 7, 3, 7, 3 bits en 4 subtramas. Esto se traduce en la operación de codificación AMR-WB/decodificación VMR-WB para retirar la parte fraccional del tono en la interfaz del sistema y para recortar el retardo diferencial a 3 bits para la 2ª y la 4ª subtrama. Los índices del libro de códigos algebraico se retiran totalmente de modo similar a la solución I-HR del Conjunto de Tasas II. La información de energía de la señal se mantiene intacta.

El resto de la operación del modo Interoperable Conjunto de Tasas I es similar a la operación del modo Conjunto de Tasas II explicado anteriormente en la Figura 12 (en términos de la operación VAD/DTX/CNG) y no se describirá en el presente documento con más detalle.

Aunque la presente invención se ha descrito en el presente documento anteriormente por medio de las realizaciones ilustrativas de la misma, se pueden modificar sin separarse de la invención, como se define en las reivindicaciones adjuntas.

## 35 REFERENCIAS

[1] ITU-T Recomendación G.722.2 "Wideband coding of speech at around 16 kbit/s using Adaptive Multi-Rate Wideband (AMR-WB)", Ginebra, 2002.

[2] 3GPP TS 26.190, "AMR Wideband Speech Codec; Transcoding Functions", Especificación Técnica 3GPP.

[3] 3GPP TS 26.192, "AMR Wideband Speech Codec; Comfort Noise Aspects", Especificación Técnica 3GPP.

40 [4] 3GPP TS 26.193: "AMR Wideband Speech Codec; Source Controlled Rate operation", Especificación Técnica 3GPP.

[5] M. Jelinek y F. Labonté, "Robust Signal/Noise Discrimination for Wideband Speech and Audio Coding," Proc. IEEE Workshop on Speech Coding, págs. 151-153, Delavan, Wisconsin, EE.UU., septiembre de 2000.

45 [6] J. D. Johnston, "Transform Coding of Audio Signals Using Perceptual Noise Criteria", IEEE Jour. on Selected Areas in Communications, vol. 6, nº 2, págs. 314-323.

[7] 3GPP2 C.S0030-0, "Selectable Mode Vocoder Service Option for Wideband Spread Spectrum Communication Systems", Especificación Técnica 3GPP.

[8] 3GPP2 C.S0014-0, "Enhanced Variable Rate Codec (EVRC)", Especificación Técnica 3GPP2.

5 [9] TIA/EIA/IS-733, "High Rate Speech Service option 17 for Wideband Spread Spectrum Communication Systems". También la Especificación Técnica 3GPP2 C.S0020-0.

## REIVINDICACIONES

1. Un procedimiento para la codificación de una señal de voz usando un códec de banda ancha en modo múltiple tasa de bits variable controlado el origen para la interoperación con un códec de banda ancha de tasa múltiple adaptativo, comprendiendo el procedimiento:
- (i) la recepción de una señal de voz de entrada;
  - 5 (ii) la aplicación de una trama de voz derivada de la señal de voz de entrada a una función de detección de la actividad de voz para determinar si la trama de voz es una trama de voz activa que contiene voz activa o una trama de voz inactiva que no contiene voz activa;
  - (iii) cuando se determina que la trama de voz de entrada es una trama de voz inactiva, la determinación de si la trama de voz inactiva se ha de codificar como una trama descriptora del silencio para su uso por el códec de banda ancha tasa múltiple adaptativo;
  - 10 (iv) cuando se determina que la trama de voz inactiva se ha de codificar como una trama descriptora del silencio, la codificación de la trama de voz inactiva usando un algoritmo de codificación generador de ruido de confort en cuarto de tasa permitiendo de ese modo a la trama de voz inactiva que se determina que ha de ser codificada como una trama descriptora del silencio sea decodificado por el códec de banda ancha tasa múltiple adaptativo;
  - 15 (v) cuando se determina que la trama de voz inactiva no se ha de codificar como una trama descriptora del silencio, la codificación de la trama de voz inactiva usando un algoritmo de codificación generador de ruido de confort en octavo de tasa por lo que se disminuye una tasa de datos promedio.
2. Un procedimiento de acuerdo con la reivindicación 1, en el que el algoritmo de codificación generador de ruido de confort en cuarto de tasa se dispone para producir tramas de voz codificadas en cuarto de tasa que tienen un cierto primer número predeterminado de bits menor que el número de bits usados para representar una trama descriptora del silencio para su uso por dicho códec de banda ancha tasa múltiple adaptativo y en el que el algoritmo de codificación del generador de ruido de confort en octavo de tasa se dispone para producir tramas de voz codificadas en octavo de tasa que tienen un cierto segundo número predeterminado de bits menor que el número de bits usados para representar una trama descriptora del silencio para su uso por dicho códec de banda ancha tasa múltiple adaptativo y cuando se determina que la trama de voz inactiva se ha de codificar como una trama descriptora del silencio, la trama de voz inactiva se codifica con un número de bits compatible con una trama descriptora del silencio para su uso por el códec de banda ancha tasa múltiple adaptativa y se transmite como una trama de voz codificada en cuarto de tasa.
3. Un procedimiento de acuerdo con la reivindicación 2, en el que el códec de banda ancha modo múltiple tasa de bits variable comprende un modo de codificación de tasa completa dispuesto para producir tramas de voz codificadas en tasa completa que comprenden un tercer número de bits y un modo de codificación en tasa media dispuesto para producir tramas de voz codificadas en tasa media que tienen un cuarto número de bits menor que dicho tercer número de bits en el que dicho primer número de bits es menor que dicho cuarto número de bits y el segundo número de bits es menor que dicho primer número de bits.
4. Un procedimiento de acuerdo con cualquiera de las reivindicaciones precedentes, en el que el códec de banda ancha en modo múltiple tasa de bits variable se ajusta o bien al conjunto de tasas 1 de CDMA o bien al conjunto de tasas 2 de CDMA.
5. Un procedimiento de acuerdo con cualquiera de las reivindicaciones 1 a 3, en el que el códec de banda ancha en modo múltiple tasa de bits variable se ajusta al conjunto de tasas 1 de CDMA.
6. Un procedimiento de acuerdo con cualquiera de las reivindicaciones precedentes, en el que el códec de banda ancha en modo múltiple tasa de bits variable se define de acuerdo con la norma de codificación de voz VMR-WB y el códec de banda ancha tasa múltiple adaptativo se define de acuerdo con la norma de codificación de voz AMR-WB.
7. Un procedimiento de acuerdo con la reivindicación 2, en el que dicho primer número predeterminado de bits es 54 y dicho segundo número predeterminado de bits es 20.
8. Un procedimiento de acuerdo con la reivindicación 3, en el que dicho primer número de bits es 54, dicho segundo número de bits es 20, dicho tercer número de bits es 266 y dicho cuarto número de bits es 124.
9. Un procedimiento de acuerdo con la reivindicación 7, en el que dicho primer número predeterminado de bits corresponde a una tasa de bits de 2,7 kbit/s y dicho segundo número predeterminado de bits corresponde a una tasa de bits de 1,0 kbit/s.
10. Un procedimiento de acuerdo con la reivindicación 3, en el que dicho primer número de bits corresponde a una tasa de bits de 2,7 kbit/s, dicho segundo número de bits corresponde a una tasa de bits de 1,0 kbit/s, dicho tercer número de bits corresponde a una tasa de bits de 13,3 kbit/s y dicho cuarto número de bits corresponde a una tasa de bits de 6,2 kbit/s.
11. Un procedimiento de acuerdo con la reivindicación 7 o la reivindicación 8, en el que cuando se determina que la

trama de voz inactiva se ha de codificar como una trama descriptora del silencio, la trama de voz inactiva se codifica con 35 bits, dejando 19 bits de dicha trama de voz codificada en cuarto de tasa sin usar.

5 12. Un procedimiento de acuerdo con la reivindicación 2 o la reivindicación 3, en el que el número de bits usado para representar una trama descriptora del silencio para su uso por el códec de banda ancha tasa múltiple adaptativo se corresponde a 1,75 kbit/s.

10 13. Un procedimiento de acuerdo con cualquiera de las reivindicaciones 1 a 12, en el que, cuando tramas de voz de entrada consecutivas que siguen a un periodo de voz activa se determina que son tramas de voz inactivas, formando por lo tanto una secuencia de tramas de voz inactivas, dicha determinación de si la trama de voz inactiva se ha de codificar como una trama descriptora del silencio especificando que la primera trama de voz inactiva de dicha secuencia, la cuarta trama de voz inactiva y posteriormente cada octava trama de voz inactiva de dicha secuencia se ha de codificar como una trama descriptora del silencio para su uso por el códec de banda ancha tasa múltiple adaptativo.

15 14. Un procedimiento de acuerdo con cualquiera de las reivindicaciones 1 a 12, en el que cuando tramas de voz de entrada consecutivas que siguen a un periodo de voz activo se determina que son tramas de voz inactivas, formando por lo tanto una secuencia de tramas de voz inactivas, dicha determinación de si la trama de voz inactiva se ha de codificar como una trama descriptora del silencio especificando que:

- 20 a) la primera trama de voz inactiva de dicha secuencia se ha de codificar como una trama descriptora del silencio para su uso por el códec de banda ancha tasa múltiple adaptativo;
- b) las siguientes dos tramas de voz inactivas de dicha secuencia se han de codificar usando el algoritmo de codificación generador de ruido de confort en octavo de tasa;
- c) la cuarta trama de voz inactiva de dicha secuencia se ha de codificar como una trama descriptora del silencio para su uso por el códec de banda ancha tasa múltiple adaptativo;
- 25 d) las siguientes siete tramas de voz inactivas se han de codificar usando dicho algoritmo de codificación generador de ruido de confort en octavo de tasa y la siguiente trama de voz inactiva se ha de codificar como una trama descriptora del silencio y
- d) se ha de repetir hasta que se detecte una trama de voz activa.

30 15. Un procedimiento de acuerdo con cualquiera de las reivindicaciones 1 a 12, en el que cuando tramas de voz de entrada consecutivas que siguen a un periodo de voz activa se determina que son tramas de voz inactivas, formando por lo tanto una secuencia de tramas de voz inactivas, dicha determinación de si la trama de voz inactiva se ha de codificar como una trama descriptora del silencio especificando que la primera trama de voz inactiva de dicha secuencia se ha de codificar como una trama descriptora del silencio para su uso por el códec de banda ancha tasa múltiple adaptativo y posteriormente cada octava trama de voz inactiva de dicha secuencia se ha de codificar como una trama descriptora del silencio para su uso por el códec de banda ancha tasa múltiple adaptativo.

35 16. Un procedimiento de acuerdo con cualquiera de las reivindicaciones 1 a 12, en el que cuando tramas de voz de entrada consecutivas se determina que son tramas de voz inactivas, formando por lo tanto una secuencia de tramas de voz inactivas, dicha determinación de si la trama de voz inactiva se ha de codificar como una trama descriptora del silencio especificando que:

- 40 a) la primera trama de voz inactiva de dicha secuencia sea de codificar como una trama descriptora del silencio para su uso por el códec de banda ancha tasa múltiple adaptativo; y
- b) las siguientes k tramas de voz inactivas de dicha secuencia se han de codificar usando el algoritmo de codificación de generación de ruido de confort en octavo de tasa y la siguiente trama de voz inactiva se ha de codificar como una trama descriptora del silencio para su uso por el códec de banda ancha tasa múltiple adaptativo; y
- b) se ha de repetir hasta que se detecte una trama de voz activa.

45 17. Un procedimiento de acuerdo con la reivindicación 16, en el que k es igual a 7.

50 18. Un procedimiento de acuerdo con cualquiera de las reivindicaciones 1 a 12, en el que cuando tramas de voz de entrada consecutivas que siguen a un periodo de voz activa se determina que son tramas de voz inactivas, formando por lo tanto una secuencia de tramas de voz inactivas, dicha determinación de si la trama de voz inactiva se ha de codificar como una trama descriptora del silencio para su uso por el códec de banda ancha tasa múltiple adaptativo especificando que una trama de voz inactiva se ha de codificar como una trama descriptora del silencio para su uso por el códec de banda ancha tasa múltiple adaptativo cuando cambian las características del ruido.

55 19. Un aparato para la codificación de una señal de voz usando un códec de banda ancha en modo múltiple tasa de bit variable controlado en origen para la interoperación con un códec de banda ancha de tasa múltiple adaptativo, comprendiendo el aparato:

- una entrada configurada para recibir una señal de voz;
- un detector de actividad de voz configurado para determinar si una trama de voz derivada de dicha señal de voz se puede clasificar como una trama de voz activa que contiene voz activa o una trama de voz inactiva que

no contiene voz activa;

- una unidad de procesamiento de la trama de voz inactiva configurada para determinar si la trama de voz inactiva se ha de codificar como una trama descriptora del silencio para su uso por el códec de banda ancha tasa múltiple adaptativo; y

5 - una unidad de codificación que responde a la operación de determinación realizada por dicha unidad de procesamiento de la trama inactiva, configurada para codificar la trama de voz inactiva usando un algoritmo de codificación generador del ruido de confort en cuarto de tasa permitiendo de este modo a la trama de voz inactiva que se determina que se ha de codificar como una trama descriptora del silencio que sea codificada por el códec de banda ancha tasa múltiple adaptativo y configurado para codificar la trama de voz inactiva  
10 usando un algoritmo de codificación generador de ruido de confort en octavo de tasa cuando se determina que la trama de voz inactiva no se ha de codificar como una trama descriptora del silencio para de ese modo disminuir la tasa de datos promedio.

20. Un aparato de acuerdo con la reivindicación 19, en el que el algoritmo de codificación generador de ruido de confort en cuarto de tasa se configura para producir tramas de voz codificadas en cuarto de tasa que comprenden un cierto primer número predeterminado de bits mayor que el número de bits configurados para representar una trama descriptora del silencio para su uso por el códec de banda ancha tasa múltiple adaptativo y en el que el algoritmo de codificación de generador de ruido de confort en octavo de tasa se configura para producir tramas de voz codificadas en octavo de tasa que comprenden un cierto segundo número predeterminado de bits menor que el número de bits configurado para representar una trama descriptora del silencio para su uso por el códec de banda ancha tasa múltiple adaptativo, y  
15  
20

la unidad de codificación se configura, en respuesta a una determinación de que la trama de voz inactiva se ha de codificar como una trama descriptora del silencio, para codificar la trama de voz inactiva con un número de bits compatible con una trama descriptora del silencio para su uso por el códec de banda ancha tasa múltiple adaptativo y para codificar la trama de voz inactiva para ser transmitida como una trama de voz codificada en cuarto de tasa.

25 21. Un aparato de acuerdo con una cualquiera de las reivindicaciones 19 y 21, en el que el códec de banda ancha en modo múltiple tasa de bits variable se define de acuerdo con la norma de codificación de la voz VMR-WB y el códec de banda ancha tasa múltiple adaptativo se define de acuerdo con la norma de codificación de la voz AMR-WB.

30 22. Un aparato de acuerdo con cualquiera de las reivindicaciones 19 a 20, en el que la unidad de procesamiento de la trama de voz inactiva se configura además para determinar que la primera trama de voz inactiva de una secuencia de tramas de voz inactivas, la cuarta trama de voz inactiva y posteriormente cada octava trama de voz inactiva de la secuencia se ha de codificar como una trama descriptora del silencio para su uso por el códec de banda ancha tasa múltiple adaptativo.

35 23. Un aparato de acuerdo con cualquiera de las reivindicaciones 19 a 21, en el que la unidad de procesamiento de la trama de voz inactiva se configura además, en respuesta a una determinación de que una secuencia de tramas de voz de entrada consecutivas que siguen a un período de voz activa son tramas de voz inactivas, para realizar una operación de determinación sobre una trama de voz inactiva de la secuencia para especificar que la trama de voz inactiva se ha de codificar de acuerdo con la siguiente regla:

40 a) una primera trama de voz inactiva de la secuencia se ha de codificar como una trama descriptora del silencio para su uso por el códec de banda ancha tasa múltiple adaptativo;

b) las siguientes dos tramas de voz inactivas de la secuencia se han de codificar usando el algoritmo de codificación generador de ruido de confort en octavo de tasa;

c) una cuarta trama de voz inactiva de la secuencia se ha de codificar como una trama descriptora del silencio para su uso por el códec de banda ancha tasa múltiple adaptativo;

45 d) las siete siguientes tramas de voz inactivas se han de codificar usando el algoritmo de codificación generador de ruido de confort en octavo de tasa y la trama de voz inactiva siguiente se ha de codificar como una trama descriptora del silencio para su uso por el códec de banda ancha tasa múltiple adaptativo; y  
d) se ha de repetir hasta que se detecte una trama de voz activa.

50 24. Un aparato de acuerdo con cualquiera de las reivindicaciones 19 a 21, en el que la unidad de procesamiento de la trama de voz inactiva se configura además, en respuesta a una determinación de que una secuencia de tramas de voz de entrada consecutivas que siguen a un período de voz activa se determina que son tramas de voz inactivas, para realizar una operación de determinación sobre una trama de voz inactiva de la secuencia para especificar que la trama de voz inactiva se ha de codificar de acuerdo con la siguiente regla:

55 la primera trama de voz inactiva de la secuencia se ha de codificar como una trama descriptora del silencio para su uso por el códec de banda ancha tasa múltiple adaptativo y posteriormente cada octava trama de voz inactiva de la secuencia se ha de codificar como una trama descriptora del silencio para su uso por el códec de banda ancha tasa múltiple adaptativo.



25. Un aparato de acuerdo con cualquiera de las reivindicaciones 19 a 21, en el que la unidad de procesamiento de la trama de voz inactiva se configura además, en respuesta a una determinación de que una secuencia de tramas de voz de entrada consecutivas se determina que son tramas de voz inactivas, para realizar una operación de determinación sobre una trama de voz inactiva de la secuencia para especificar que la trama de voz inactiva se ha de codificar de acuerdo con la siguiente regla:

- 5
- a) una primera trama de voz inactiva de la secuencia se ha de codificar como una trama descriptora del silencio para su uso por el códec de banda ancha tasa múltiple adaptativo;
  - b) las siguientes k tramas de voz inactivas de la secuencia se han de codificar usando el algoritmo de codificación generador de ruido de confort en octavo de tasa y la siguiente trama de voz inactiva se ha de codificar como una trama descriptora del silencio para su uso por el códec de banda ancha tasa múltiple adaptativo; y
  - b) se ha de repetir hasta que se detecte una trama de voz activa.
- 10

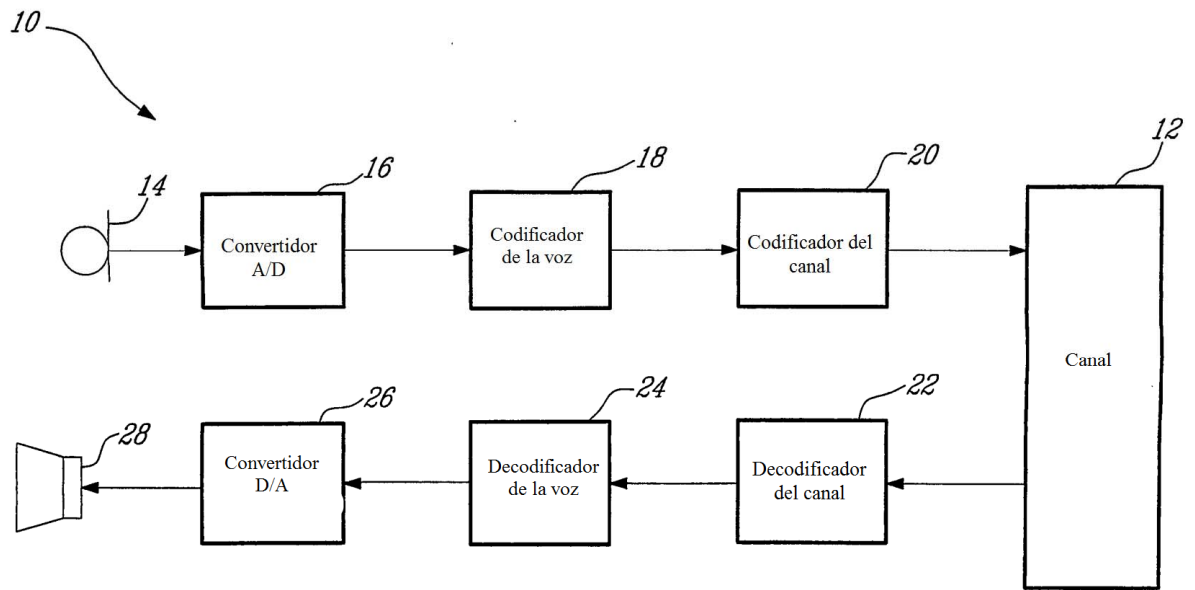


Fig. 1

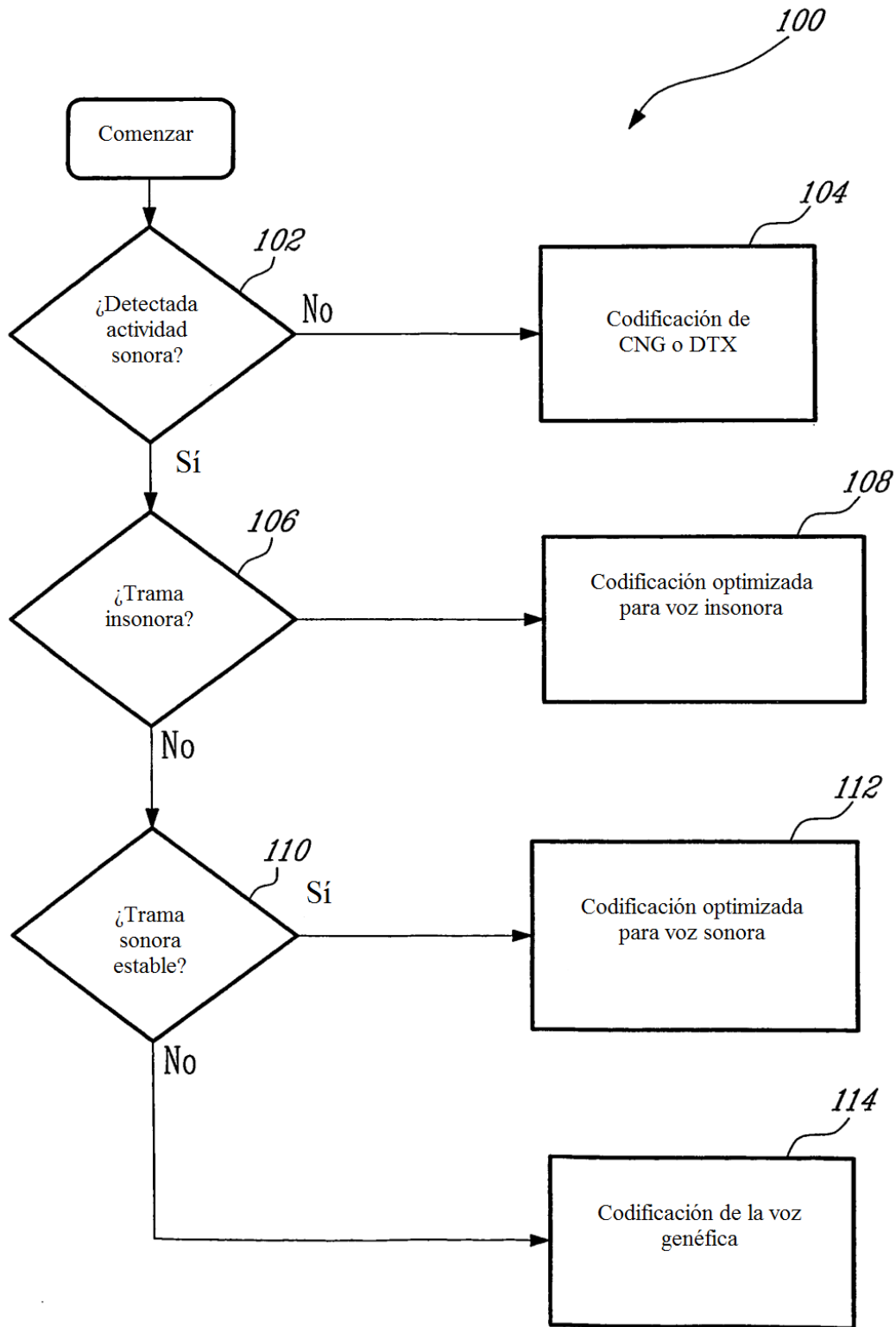
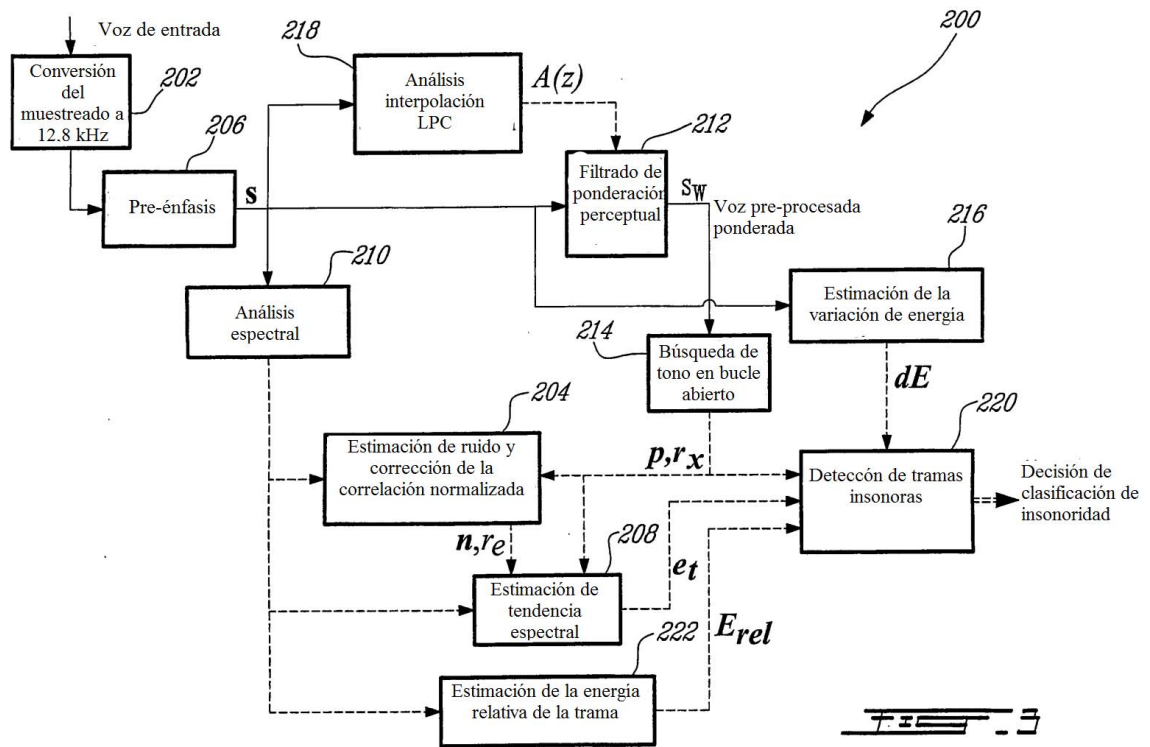


Fig. 2



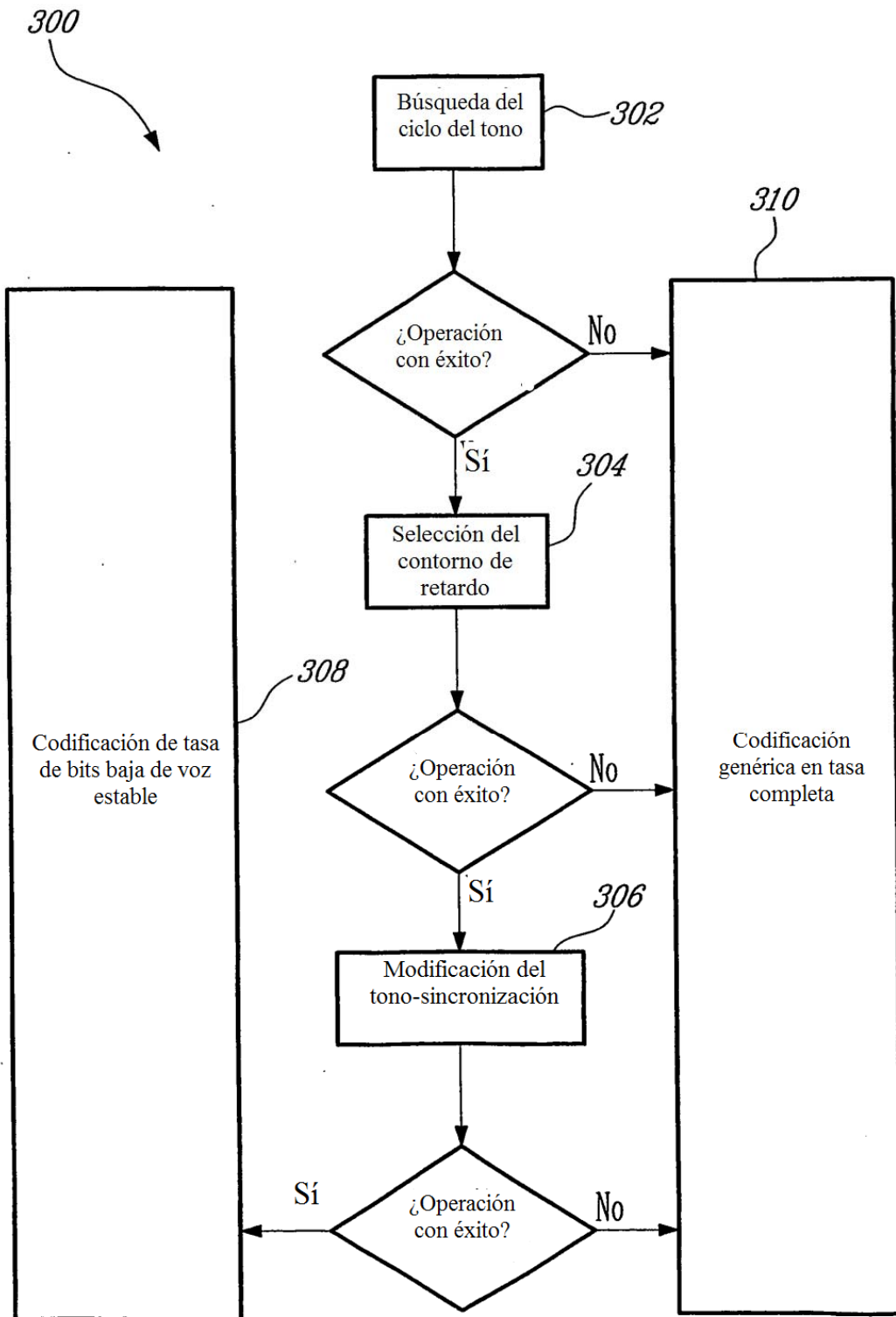
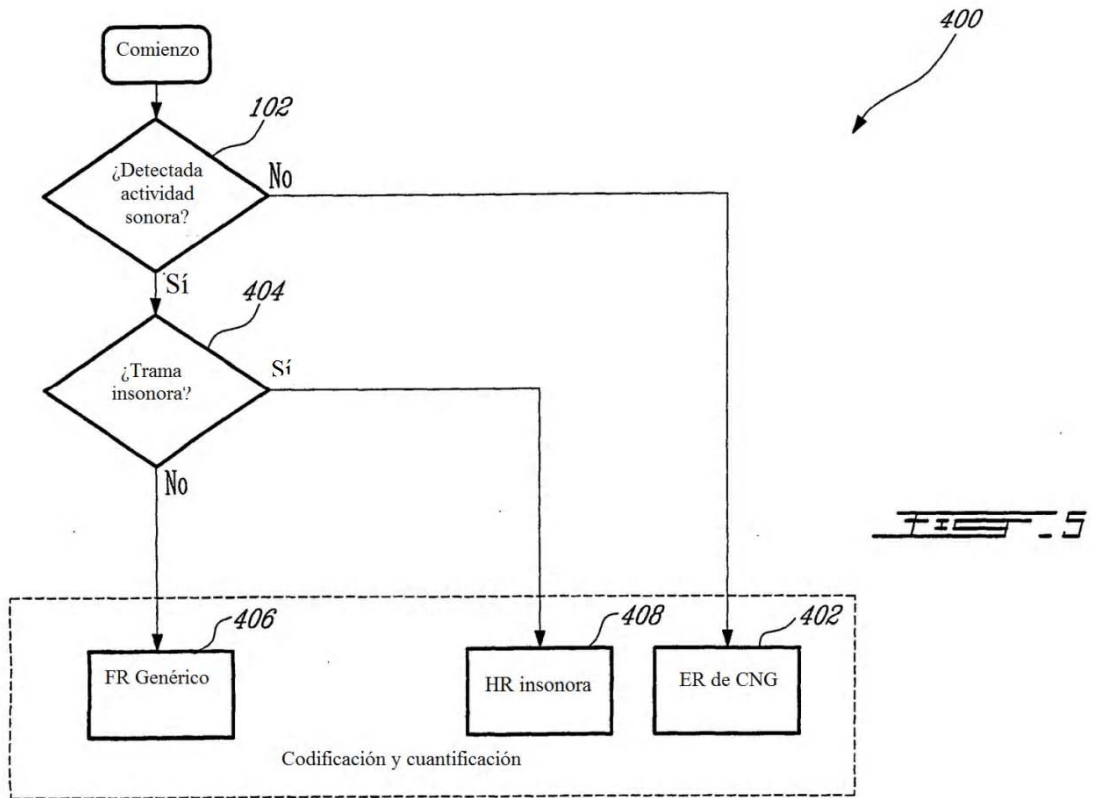


Fig. 4



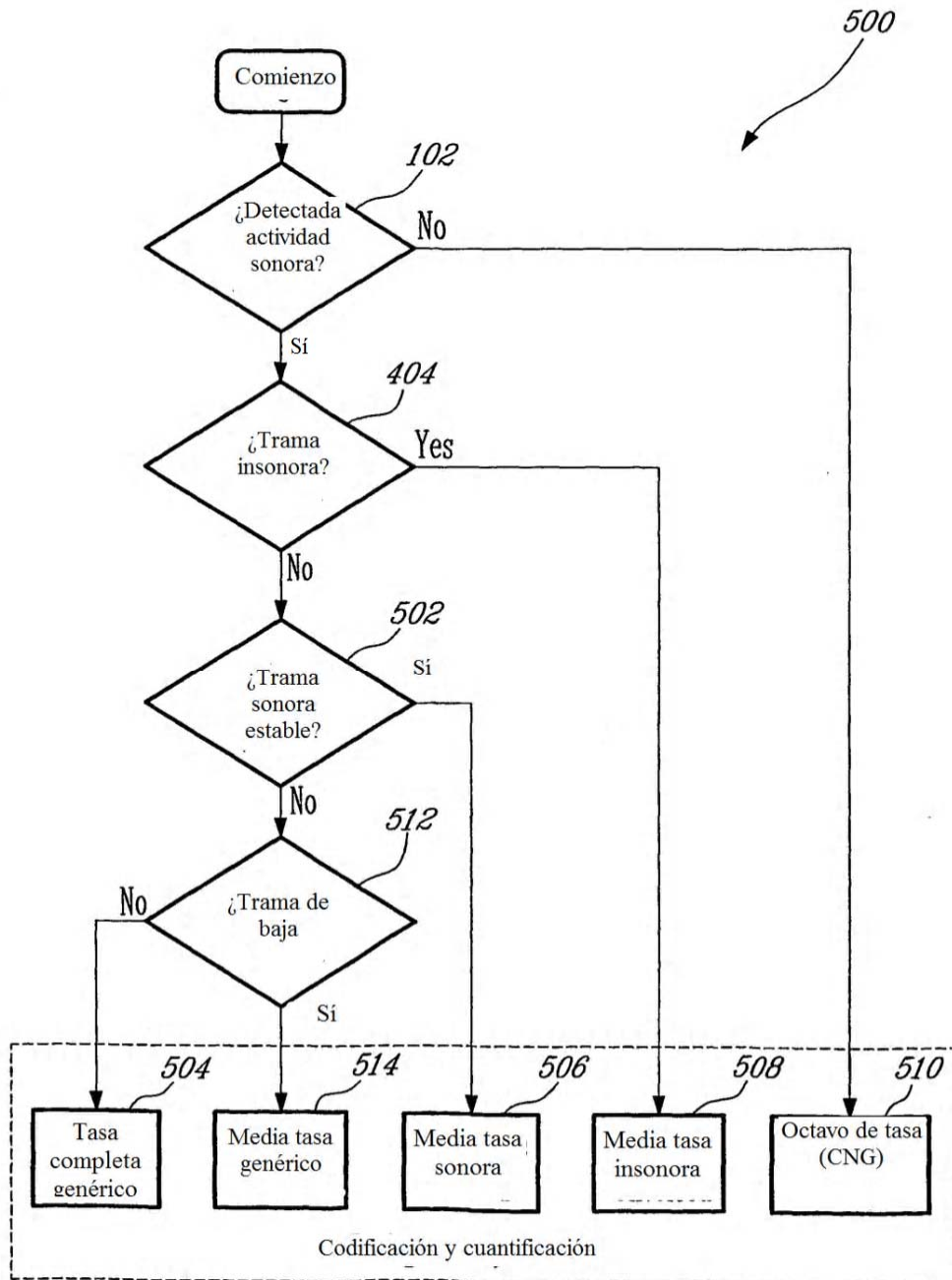


FIG. 6

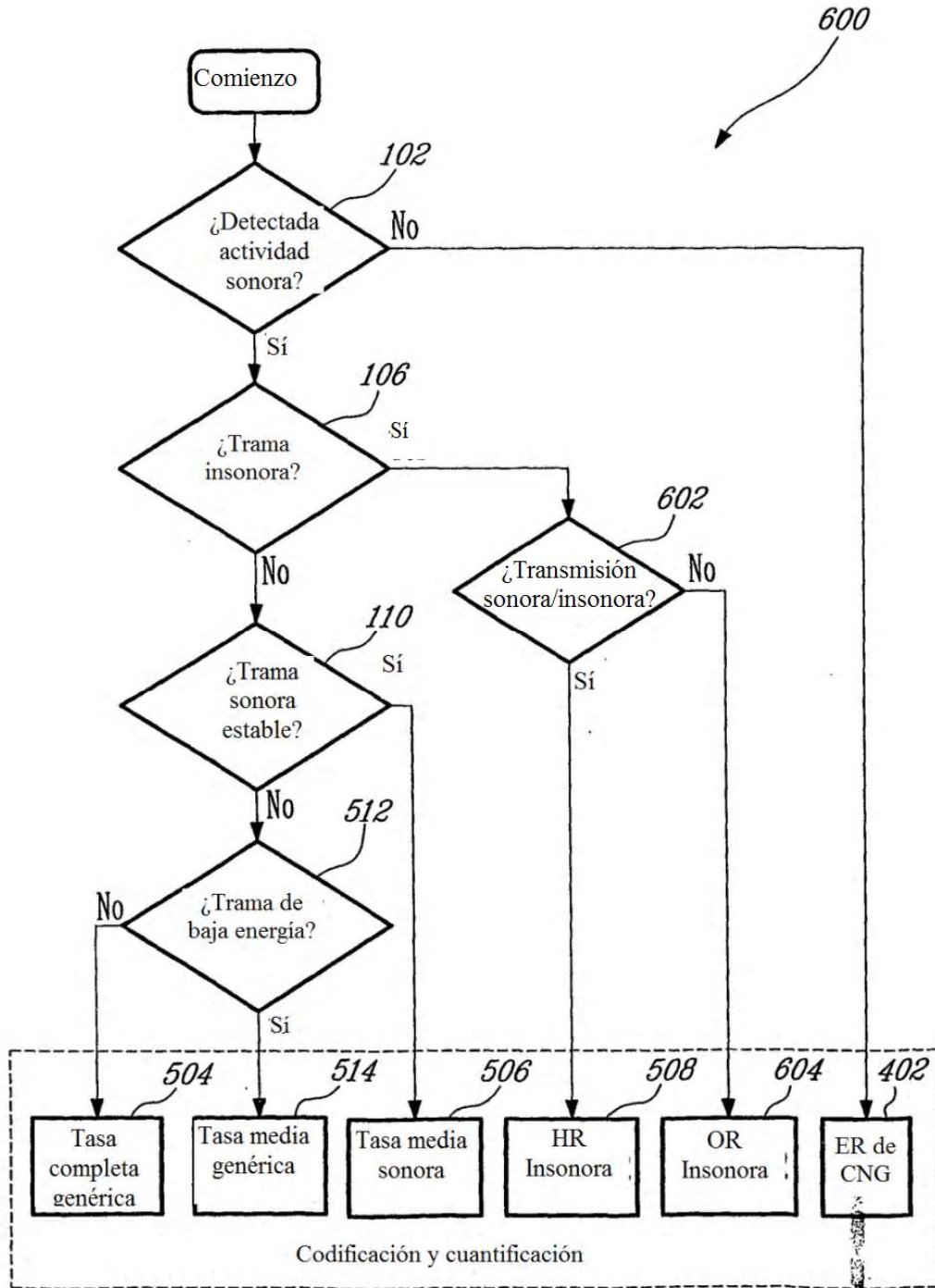
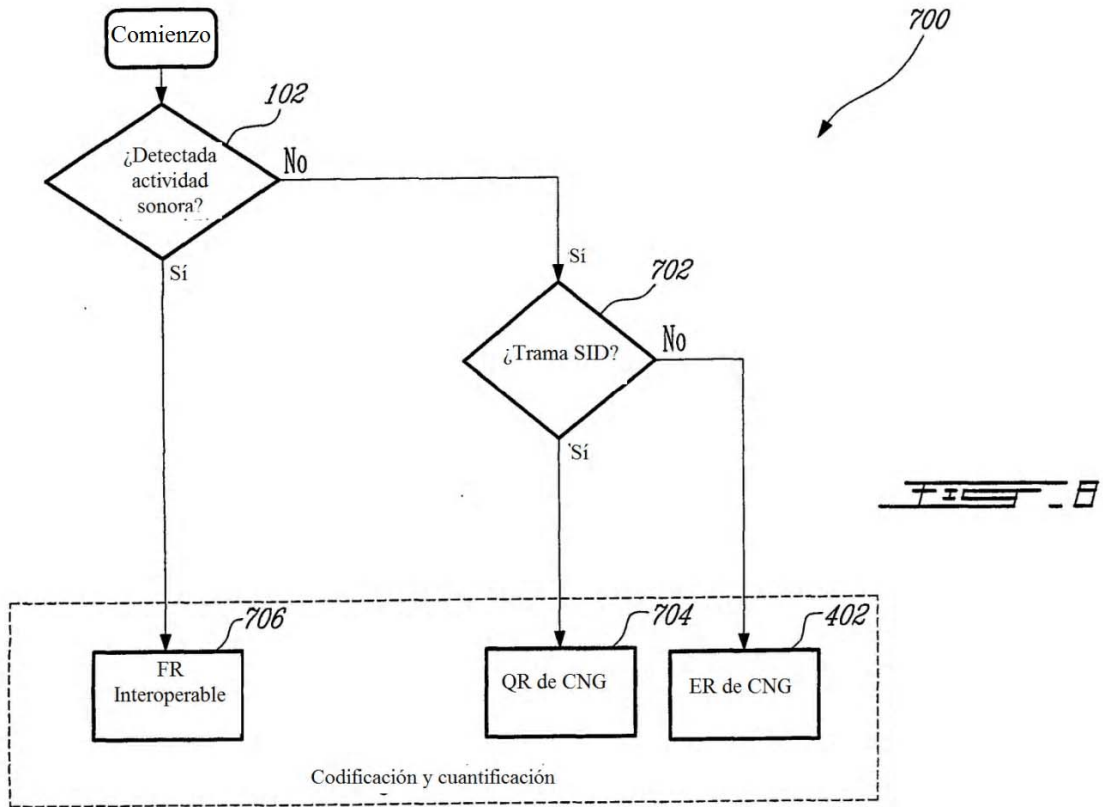


FIG. 7





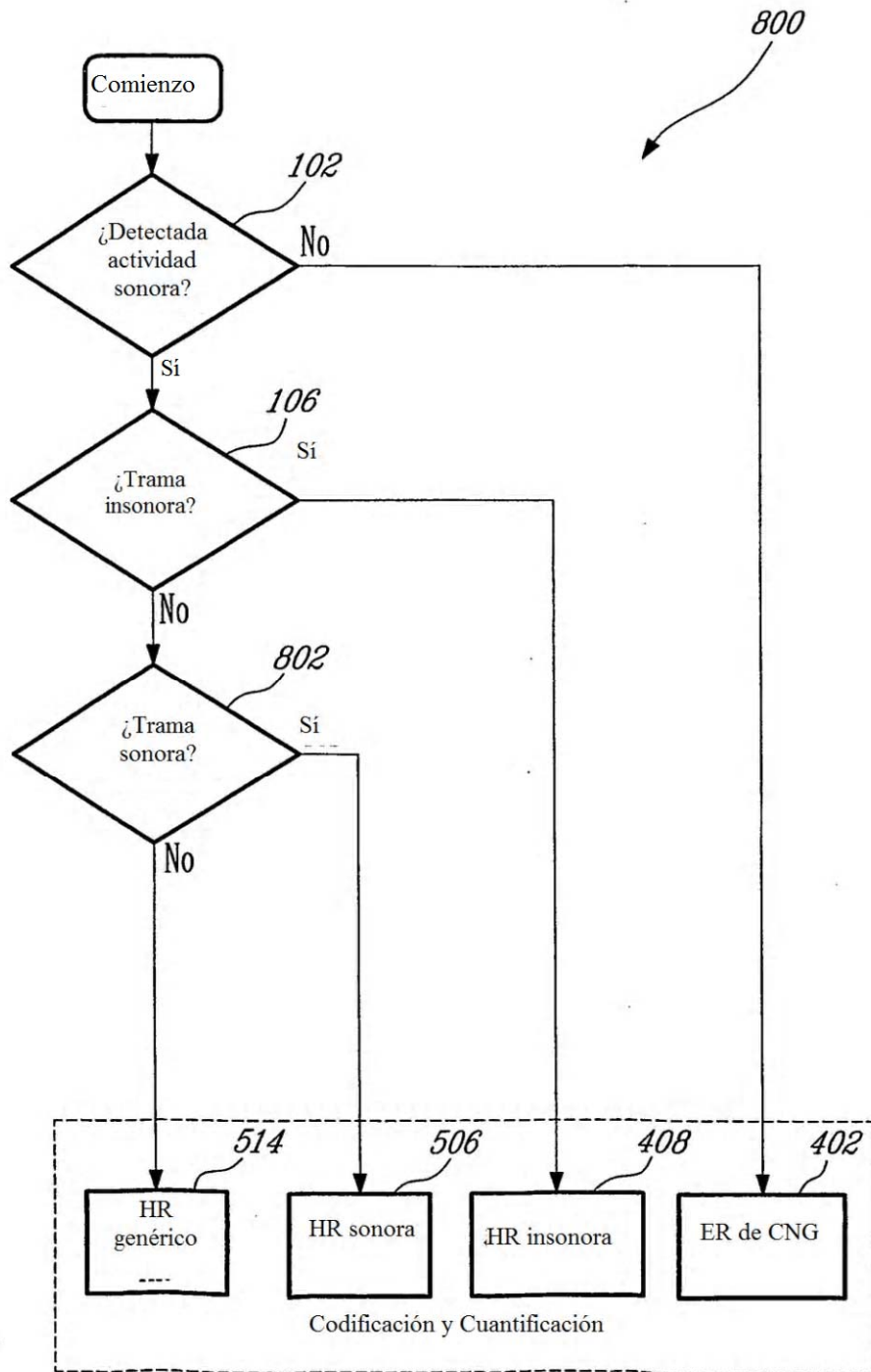


FIG. 9

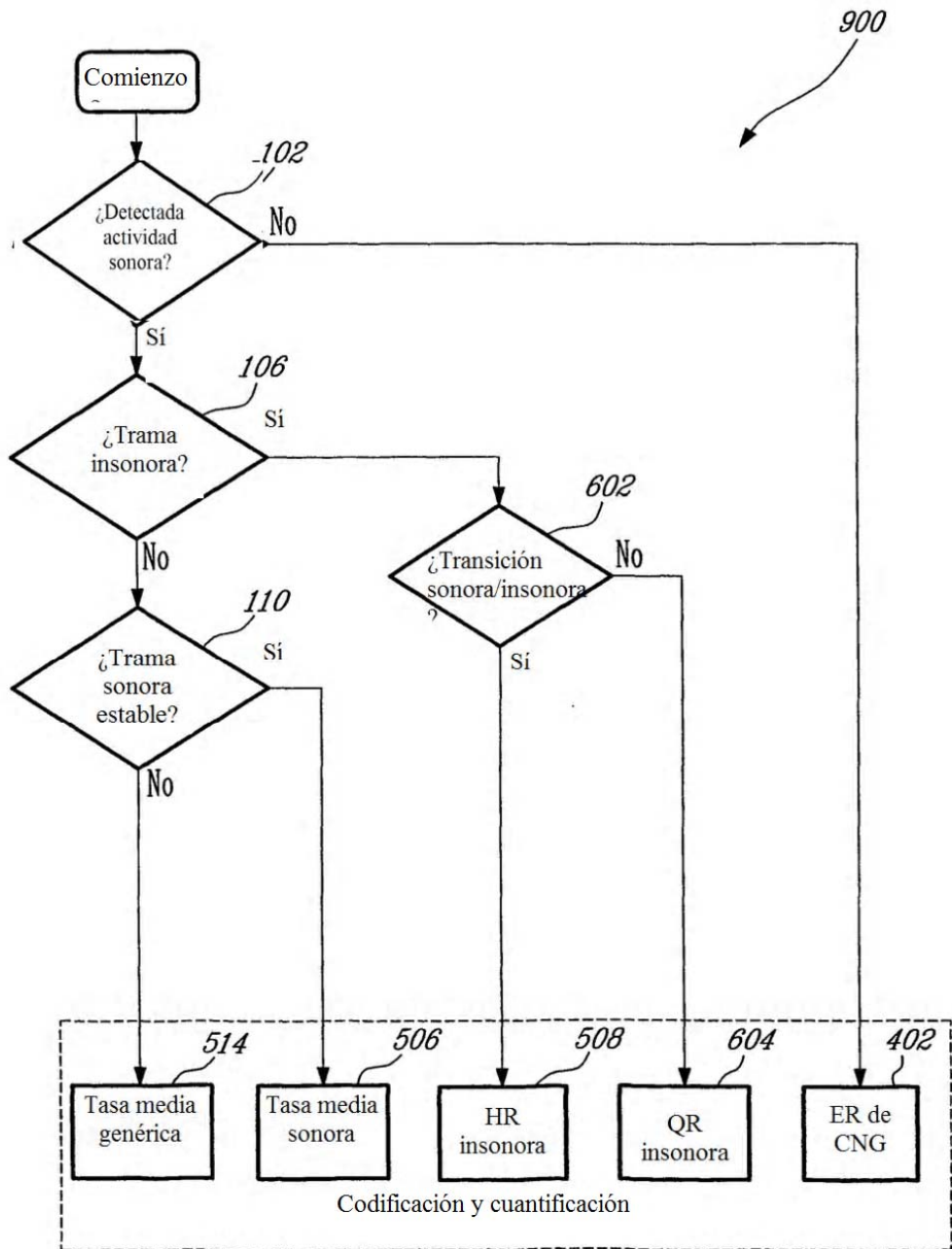


FIG. 10

