



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS  
ESPAÑA



⑪ Número de publicación: **2 624 718**

⑮ Int. Cl.:

**G10L 19/08** (2013.01)  
**G10L 19/12** (2013.01)

⑫

TRADUCCIÓN DE PATENTE EUROPEA

T3

⑥ Fecha de presentación y número de la solicitud internacional: **24.10.2007 PCT/CA2007/001896**

⑦ Fecha y número de publicación internacional: **02.05.2008 WO08049221**

⑨ Fecha de presentación y número de la solicitud europea: **24.10.2007 E 07816046 (2)**

⑩ Fecha y número de publicación de la concesión europea: **22.03.2017 EP 2102619**

⑭ Título: **Método y dispositivo para la codificación de tramas de transición en señales de voz**

⑩ Prioridad:

**24.10.2006 US 853749 P**

⑮ Fecha de publicación y mención en BOPI de la traducción de la patente:

**17.07.2017**

⑬ Titular/es:

**VOICEAGE CORPORATION (100.0%)  
SUITE 250 750 LUCERNE ROAD CITY OF MOUNT  
ROYAL  
QUEBEC H3R 2H6, CA**

⑭ Inventor/es:

**EKSLER, VACLAV;  
JELINEK, MILAN y  
SALAMI, REDWAN**

⑭ Agente/Representante:

**DE ELZABURU MÁRQUEZ, Alberto**

**ES 2 624 718 T3**

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

## DESCRIPCIÓN

Método y dispositivo para la codificación de tramas de transición en señales de voz

### Campo de la invención

La presente invención se refiere a una técnica para codificar digitalmente una señal de sonido, por ejemplo, una señal de voz o de audio, con vistas a transmitir y sintetizar esta señal de sonido.

Más específicamente, pero no exclusivamente, la presente invención se refiere a un método y un dispositivo para codificar tramas de transición y tramas siguientes a la transición en una señal de sonido, por ejemplo, una señal de voz o de audio, con el fin de reducir la propagación de errores en el decodificador en caso de borrado de trama, y/o mejorar la eficiencia de la codificación principalmente al comienzo de los segmentos sonoros (tramas de inicio). En particular, el método y el dispositivo reemplazan el libro de códigos adaptativo utilizado típicamente en los codificadores predictivos por un libro de códigos de formas de impulso glotal en las tramas de transición y en las tramas siguientes a la transición. El libro de códigos de forma glotal puede ser un libro de códigos fijo independiente de la excitación anterior, de modo que, una vez que el borrado de la trama ha terminado, el codificador y el decodificador utilizan la misma excitación para que la convergencia a la síntesis de canal limpio sea bastante rápida. En la codificación de la trama de inicio en el CELP tradicional, la memoria temporal de excitación anterior se actualiza utilizando la excitación de tipo ruido de la trama anterior no sonora o inactiva que es muy diferente de la excitación actual. Por otro lado, la técnica propuesta puede construir la parte periódica de la excitación con mucha precisión.

### Antecedentes

Un codificador de voz convierte una señal de voz en una secuencia de digital de bits que se transmite a través de un canal de comunicación o está almacenada en un medio de almacenamiento. La señal de voz se digitaliza, que se muestrea y cuantifica habitualmente con 16 bits por muestra. El codificador de voz tiene la función de representar estas muestras digitales con un menor número de bits, manteniendo al mismo tiempo una buena calidad de voz subjetiva. El decodificador o sintetizador de voz funciona en la secuencia de bits transmitida o almacenada y la convierte de nuevo en una señal de voz.

La codificación de predicción lineal excitada por código (CELP – Code-Excited Linear Prediction, en inglés) es una de las mejores técnicas de la técnica anterior para lograr un buen compromiso entre la calidad subjetiva y la tasa de bits. Esta técnica de codificación constituye la base de varios estándares de codificación de voz, tanto en aplicaciones inalámbricas como en líneas fijas. En la codificación CELP, la señal de voz muestreada se procesa en bloques sucesivos de M muestras usualmente habitualmente tramas, donde M es un número predeterminado que corresponde típicamente a 10 a 30 ms. Un filtro de predicción lineal (LP – Linear Prediction, en inglés) se calcula y transmite cada trama. El cálculo del filtro LP normalmente necesita una anticipación, un segmento de voz de 5 a 15 ms de la siguiente trama. La trama de la muestra M se divide en bloques más pequeños llamados subtramas. Por lo general, el número de subtramas es de tres o cuatro, lo que resulta en subtramas de 4 a 10 ms. En cada subtrama, normalmente se obtiene una señal de excitación a partir de dos componentes, la excitación anterior y la excitación innovadora de libro de códigos fijo. El componente formado a partir de la excitación anterior a menudo se denomina libro de códigos adaptativo o excitación de tono. Los parámetros que caracterizan la señal de excitación son codificados y transmitidos al decodificador, donde la señal de excitación reconstruida se utiliza como entrada del filtro LP.

Los códices de voz de tipo CELP se basan en gran medida en la predicción para conseguir su alto rendimiento. La predicción utilizada puede ser de diferentes clases, pero, por lo general, comprende la utilización de un libro de códigos adaptativo que contiene una señal de excitación seleccionada en tramas anteriores. Un codificador CELP aprovecha la casi periodicidad de la señal de voz vocal mediante la búsqueda en la excitación anterior del segmento más similar al segmento que se está codificando actualmente. La misma señal de excitación anterior se mantiene también en el decodificador. Por consiguiente, es suficiente que el codificador envíe un parámetro de retardo y una ganancia para que el decodificador reconstruya la misma señal de excitación que se utiliza en el codificador. La evolución (diferencia) entre el segmento de voz previo y el segmento de voz actualmente codificado se modeliza además utilizando una innovación seleccionada de un libro de códigos fijo. La tecnología CELP se describirá con más detalle a continuación en la presente memoria.

Un problema de fuerte predicción inherente en los codificadores de voz basados en CELP aparece en presencia de errores de transmisión (tramas o paquetes borrados) cuando el estado del codificador y del decodificador se desincronizan. Debido a la predicción, el efecto de una trama borrada no se limita por ello a la trama borrada, sino que continúa propagándose después del borrado, a menudo durante varias tramas siguientes. Naturalmente, el impacto perceptual puede ser muy molesto.

Las transiciones desde el segmento de voz no sonoro hablada (por ejemplo, la transición entre una consonante o un período de conversación inactiva, y una vocal) o las transiciones entre dos segmentos sonoros diferentes (por ejemplo, las transiciones entre dos vocales) son los casos más problemáticos para la ocultación del borrado de tramas. Cuando se pierde una transición desde un segmento de voz no sonoro a un segmento de voz sonoro (inicio

sonoro), la trama justo antes de la trama de inicio sonora es no sonora o inactiva y, por lo tanto, no se encuentra excitación periódica significativa en la memoria temporal de la excitación anterior (libro de códigos adaptativo). En el codificador, la excitación periódica anterior se acumula en el libro de códigos adaptativo durante la trama de inicio, y la siguiente trama sonora se codifica utilizando esta última excitación periódica. La mayoría de las técnicas de 5 ocultación de errores de trama utilizan la información de la última trama correctamente recibida para ocultar la trama que falta. Cuando se pierde la trama de inicio, la memoria temporal de excitación anterior del decodificador se 10 actualizará, de este modo, utilizando la excitación de tipo ruido de la trama anterior (trama no sonora o inactiva). Por lo tanto, la parte periódica de la excitación está completamente ausente en el libro de códigos adaptativo en el decodificador después de un inicio sonoro perdido, y pueden pasar hasta varias tramas para que el decodificador se 15 recupere de esta pérdida.

Una situación similar ocurre en el caso de pérdida de voz a transición de voz. En ese caso, la excitación almacenada en el libro de códigos adaptativo antes de la trama de transición tiene características típicamente muy diferentes de la excitación almacenada en el libro de códigos adaptativo después de la transición. De nuevo, dado que el decodificador suele ocultar la trama perdida con la utilización de la información de la trama anterior, el estado del codificador y el decodificador serán muy diferentes y la señal sintetizada puede sufrir distorsiones importantes.

El documento US 6.223.550 describe un códec CELP en el que el libro de códigos adaptativo es reemplazado para tramas transitorias por un libro de códigos que comprende entradas de excitación de múltiples impulsos.

### **Objetivos de la invención**

Un objetivo de la presente invención es, por lo tanto, proporcionar un método y un dispositivo para codificar tramas 20 de transición en un codificador predictivo de voz y/o de audio para mejorar la robustez del codificador frente a pérdida de tramas y/o mejorar la eficiencia de la codificación.

Otro objetivo de la presente invención es eliminar la propagación de errores y aumentar la eficiencia de codificación 25 en códecs basados en CELP reemplazando la búsqueda en el libro de códigos adaptativo dependiente de intertramas mediante una búsqueda no predictiva en el libro de códigos, por ejemplo, de forma glotal. Esta técnica requiere un nulo retraso adicional, una complejidad adicional despreciable y un nulo aumento en la tasa de bits en comparación con la codificación CELP tradicional.

Estos objetivos se consiguen mediante las reivindicaciones independientes.

### **Compendio de la invención**

Más específicamente, de acuerdo con un aspecto de la presente invención, se proporciona un método de modo de transición para su uso en un códec de señal de sonido de tipo predictivo para producir una excitación de modo de transición que reemplaza a una excitación del libro de códigos adaptativo en una trama de transición y/o una trama 30 siguiente a la transición en la señal de sonido, que comprende: proporcionar un libro de códigos de modo de transición para generar un conjunto de vectores de código independientes de la excitación anterior; suministrar un índice del libro de códigos al libro de códigos de modo de transición; y generar, por medio del libro de códigos de modo de transición, y en respuesta al índice del libro de códigos, uno de los vectores de código del conjunto 35 correspondiente a la excitación de modo de transición.

De acuerdo con un segundo aspecto de la presente invención, se proporciona un dispositivo de modo de transición para su uso en un códec de señal de sonido de tipo predictivo para producir una excitación de modo de transición que reemplaza a una excitación del libro de códigos adaptativo en una trama de transición y/o una trama siguiente a la transición en la señal de sonido, que comprende una entrada para recibir un índice del libro de códigos y un libro 40 de códigos de modo de transición para generar un conjunto de vectores de código independientes de la excitación anterior. El libro de códigos de modo de transición es sensible al índice para generar, en la trama de transición y/o la trama siguiente a la transición, uno de los vectores de código del conjunto correspondiente a dicha excitación de modo de transición.

De acuerdo con un tercer aspecto de la presente invención, se proporciona un método de codificación para generar 45 una excitación de modo de transición que reemplaza a una excitación de libro de códigos adaptativo en una trama de transición y/o una trama siguiente a la transición en una señal de sonido, que comprende: generar una señal objetivo de búsqueda en el libro de códigos; proporcionar un libro de códigos de modo de transición para generar un conjunto de vectores de código independientes de la excitación anterior, correspondiendo cada uno de los vectores de código del conjunto a una respectiva excitación de modo de transición; realizar una búsqueda en el libro de códigos de modo de transición para encontrar el vector de código del conjunto correspondiente a una excitación de modo de transición que corresponde óptimamente a la señal objetivo de búsqueda en el libro de códigos.

De acuerdo con un cuarto aspecto de la presente invención, se proporciona un dispositivo codificador para generar 55 una excitación de modo de transición que reemplaza a una excitación de libro de códigos adaptativo en una trama de transición y/o una trama siguiente a la transición en una señal de sonido, que comprende: un generador de señal objetivo de búsqueda en el libro de códigos; un libro de códigos de modo de transición para generar un conjunto de vectores de código independientes de la excitación anterior, correspondiendo cada uno de los vectores de código del

conjunto a una respectiva excitación de modo de transición; y un buscador del libro de códigos de modo de transición para encontrar el vector de código del conjunto correspondiente a una excitación de modo de transición óptimamente correspondiente a la señal objetivo de búsqueda en el libro de códigos.

5 De acuerdo con un quinto aspecto de la presente invención, se proporciona un método de decodificación para generar una excitación de modo de transición que reemplaza a una excitación de libro de códigos adaptativo en una trama de transición y/o una trama siguiente a la transición en una señal de sonido, que comprende: recibir un índice del libro de códigos; suministrar el índice del libro de códigos a un libro de códigos de modo de transición para generar un conjunto de vectores de código independientes de la excitación anterior; y generar, por medio del libro de códigos de modo de transición y en respuesta al índice del libro de códigos, uno de los vectores de código del conjunto correspondiente a la excitación de modo de transición.

10 De acuerdo con un sexto aspecto de la presente invención, se proporciona un dispositivo decodificador para generar una excitación de modo de transición que reemplaza a una excitación de libro de códigos adaptativo en una trama de transición y/o una trama siguiente a la transición en una señal de sonido, que comprende una entrada para recibir un índice del libro de códigos y un libro de códigos de modo de transición para generar un conjunto de vectores de código independientes de la excitación anterior. El libro de códigos de modo de transición es sensible al índice para generar en la trama de transición y/o la trama siguiente a la transición uno de los vectores de código del conjunto correspondiente a la excitación de modo de transición.

15 Los objetivos, ventajas y características anteriores y otros de la presente invención se harán más evidentes con la lectura de la siguiente descripción no restrictiva de un modo de realización ilustrativo de la misma, dado solamente a modo de ejemplo con referencia a los dibujos que se acompañan.

#### Breve descripción de los dibujos

En los dibujos adjuntos:

la figura 1a es un diagrama de bloques esquemático de un codificador basado en CELP;

la figura 1b es un diagrama de bloques esquemático de un codificador basado en CELP;

20 25 la figura 2 es un diagrama de bloques esquemático de una máquina de estados de clasificación de tramas para ocultación de borrado;

la figura 3 es un ejemplo de segmento de una señal de voz con una trama de transición de voz y una trama de inicio;

30 35 la figura 4 es un diagrama de bloques funcional que ilustra una regla de clasificación para seleccionar tramas TM (modo de transición – Transition Mode, en inglés) en inicios de voz, donde N\_TM\_FRAMES representa un número de tramas consecutivas para evitar la utilización de una técnica de codificación de trama de transición, 'clase' representa una clase de trama, y SONORA\_TYPE significa clases de inicio, de voz y de transición de voz;

la figura 5a es una ilustración esquemática de un ejemplo de trama de una señal de voz dividida en cuatro (4) subtramas, que muestra la señal de voz en el dominio del tiempo;

35 40 la figura 5b es una ilustración esquemática de un ejemplo de trama de una señal de voz dividida en cuatro (4) subtramas, que muestra una señal residual LP;

la figura 5c es una ilustración esquemática de un ejemplo de trama de una señal de voz dividida en cuatro (4) subtramas, que muestra una señal de excitación de la primera etapa construida utilizando la técnica de codificación TM en el codificador;

45 50 la figura 6 muestra gráficos que ilustran ocho impulsos glotales con una longitud de 17 muestras utilizados para la construcción del libro de códigos de forma glotal, en las que el eje x indica un índice de tiempo discontinuo y el eje y una amplitud del impulso;

la figura 7 es un diagrama de bloques esquemático de un ejemplo de parte de TM de un codificador CELP, en el que K' representa un índice del libro de códigos de forma glotal y G (z) es un filtro de conformación;

la figura 8 es una representación gráfica del cálculo de  $Ck'$ , la raíz cuadrada del numerador en el criterio de la ecuación (16), en la que las porciones sombreadas del vector / matriz son distintas de cero;

la figura 9 es una representación gráfica del cálculo de  $EK'$ , el denominador del criterio de la ecuación (16)), en el que las porciones sombreadas del vector / matriz son distintas de cero;

la figura 10 es una representación gráfica del cálculo de la matriz de convolución  $Z^T$ ; en este ejemplo, el filtro de conformación G (z) tiene solo tres (3) coeficientes distintos de cero ( $L_{1/2} = 1$ );

50 la figura 11 es un diagrama de bloques esquemático de un ejemplo de la parte de TM de un decodificador CELP;

- la figura 12a es un diagrama de bloques esquemático de un ejemplo de estructura del filtro Q (z);
- la figura 12b es un gráfico de un ejemplo de modificación de vectores de código de forma glotal, en el que el impulso repetido está punteado;
- 5 la figura 13 es un diagrama de bloques esquemático de la parte de TM de un codificador CELP que incluye el filtro Q (z);
- la figura 14 es un gráfico que ilustra un vector de código de la forma glotal con construcción de dos impulsos cuando se utiliza una búsqueda en el libro de códigos adaptativo en una parte de la subtrama con una búsqueda en el libro de códigos de forma glotal;
- 10 la figura 15 es un gráfico que ilustra una construcción del vector de código de la forma glotal en el caso en el que el segundo impulso glotal aparece en las primeras posiciones  $L_{1/2}$  de la subtrama siguiente;
- la figura 16 es un diagrama de bloques esquemático de la parte de TM de un codificador utilizado en una implementación en un códec EV-VBR (Tasa de bits variable incorporada - Embedded Variable Bit Rate, en inglés);
- la figura 17a es un gráfico que muestra un ejemplo de señal de voz en el dominio del tiempo;
- la figura 17b es un gráfico que muestra una señal LP residual que corresponde a la señal de voz de la figura 17a;
- 15 la figura 17c es un gráfico que muestra una señal de excitación de la primera etapa en situación sin error;
- las figuras 18a a 18c son gráficos que ilustran un ejemplo de comparación de construcción de inicio, en los que el gráfico de la figura 18a representa la señal de voz de entrada, el gráfico de la figura 18b representa la voz sintetizada de salida de un códec EV-VBR sin la técnica de codificación TM, y el gráfico de la figura 18c representa la voz sintetizada de salida de un códec EV-VBR con la técnica de codificación TM;
- 20 las figuras 19a a 19c son gráficos que ilustran un ejemplo del efecto de la técnica de codificación TM en el caso de borrado de trama, en el que el gráfico de la figura 19a representa la señal de voz de entrada, el gráfico de la figura 19b representa la voz sintetizada de salida de un códec EV-VBR sin la técnica de codificación TM, y el gráfico de la figura 19c representa la voz sintetizada de salida de un códec EV-VBR con la técnica de codificación TM;
- 25 la figura 20 es un gráfico que ilustra un ejemplo de la señal de excitación de la primera etapa en una trama de la configuración TRANSITION\_1\_1;
- la figura 21 es un gráfico que ilustra un ejemplo de la señal de excitación de la primera etapa en una trama de la configuración TRANSITION\_1\_2;
- la figura 22 es un gráfico que ilustra un ejemplo de la señal de excitación de la primera etapa en una trama de la configuración TRANSITION\_1\_3;
- 30 la figura 23 es un gráfico que ilustra un ejemplo de la señal de excitación de la primera etapa en una trama de la configuración TRANSITION\_1\_4;
- la figura 24 es un gráfico que ilustra un ejemplo de la señal de excitación de la primera etapa en una trama de la configuración TRANSITION\_2;
- 35 la figura 25 es un gráfico que ilustra un ejemplo de la señal de excitación de la primera etapa en una trama de la configuración TRANSITION\_3;
- la figura 26 es un gráfico que ilustra un ejemplo de la señal de excitación de la primera etapa en una trama de la configuración TRANSITION\_4; y
- la figura 27 es un diagrama de bloques esquemático de un sistema de comunicación de voz que ilustra la utilización de dispositivos de codificación y decodificación de voz.

#### 40 **Descripción detallada**

El modo de realización ilustrativo no restrictivo de la presente invención se refiere a un método y dispositivo cuyo propósito es superar la propagación de los errores en las situaciones descritas anteriormente y aumentar la eficiencia de codificación.

Más específicamente, el método y el dispositivo de acuerdo con el modo de realización ilustrativo no restrictivo de la presente invención implementan una codificación especial, denominada técnica de codificación de modo de transición (TM – Transition Mode, en inglés), de tramas de transición y de tramas siguientes a la transición en una señal de sonido, por ejemplo, una señal de voz o de audio. La técnica de codificación TM reemplaza el libro de códigos adaptativo del códec CELP por un nuevo libro de códigos de formas de impulso glotal, denominado en lo sucesivo libro de códigos de forma glotal, en las tramas de transición y en las tramas siguientes a la transición. El

libro de códigos de forma glotal es un libro de códigos fijos independiente de la excitación anterior. Por consiguiente, una vez que el borrado de tramas ha finalizado, el codificador y el decodificador utilizan la misma excitación, por lo que la convergencia a la síntesis de canal limpio es bastante rápida.

5 Si bien la utilización de la técnica de codificación TM en tramas siguientes a una transición ayuda a prevenir la propagación de errores en el caso de que se pierda la trama de transición, otro propósito de utilizar la técnica de codificación TM también en la trama de transición es mejorar la eficiencia de la codificación. Por ejemplo, justo antes de un inicio sonoro, el libro de códigos adaptativo normalmente contiene una señal de tipo ruido no muy eficiente para codificar el comienzo de un segmento sonoro. Por lo tanto, la idea detrás de la técnica de codificación TM es complementar el libro de códigos adaptativo con un mejor libro de códigos que contiene versiones cuantificadas simplificadas de impulsos glotales para codificar los inicios sonoros.

10 La técnica de codificación TM propuesta se puede utilizar en cualquier códec de tipo CELP o códec predictivo. Como ejemplo, la técnica de codificación TM se implementa en un códec candidato en la actividad de estandarización ITU-T para un códec de tasa de bits variable incorporado al que se hará referencia en el resto del texto como códec EV-VBR. Aunque el modo de realización ilustrativo no restrictivo de la presente invención se describirá junto con el 15 marco del códec EV-VBR, se debe tener en cuenta que los principios y conceptos de la presente invención no se limitan a una aplicación al códec EV-VBR, sino a cualquier otro códec que utilice codificación predictiva. Asimismo, aunque el modo de realización ilustrativo no limitativo de la presente invención se describirá junto con una señal de voz, se debe tener en cuenta que la presente invención no se limita a una aplicación para señales de voz, sino que 20 sus principios y conceptos se puede aplicar a cualquier otro tipo de señales de sonido que incluyan señales de audio.

Una trama de voz se puede clasificar aproximadamente en una de las cuatro (4) siguientes clases de voz (esto se 25 explicará con mayor detalle en la siguiente descripción):

- tramas inactivas, caracterizadas por la ausencia de actividad de voz;
- tramas de voz no sonoras caracterizadas por una estructura aperiódica y una concentración de energía hacia las 25 frecuencias más altas;
- tramas de voz sonoras que tienen una clara naturaleza casi periódica con la energía concentrada principalmente en las bajas frecuencias; y
- cualquier otra trama clasificada como una transición que tenga características rápidamente variables.

30 En el códec EV-VBR, se ha diseñado un modo de codificación especializado para cada una de las clases. Se puede afirmar en general que las tramas inactivas se procesan mediante la generación de ruido de confort, las tramas de voz no sonoras, mediante un modo de codificación optimizada no sonora, las tramas de voz sonoras, mediante un modo optimizado de codificación sonora, y todas las demás tramas son procesadas con una tecnología CELP algebraica (ACELP – Algebraic CELP, en inglés) genérica. En el marco del códec EV-VBR, la técnica de codificación 35 TM se introduce, así, como otro modo de codificación en el esquema de codificación EV-VBR para codificar tramas de transición y tramas siguientes a la transición.

40 La figura 27 es un diagrama de bloques esquemático de un sistema de comunicación de voz que representa la utilización de codificación y decodificación de voz. El sistema de comunicación de voz soporta la transmisión y la reproducción de una señal de voz a través de un canal de comunicación 905. Aunque puede comprender, por ejemplo, un enlace por cable, óptico o por fibra, el canal de comunicación 905 comprende típicamente, al menos en parte, un enlace de radiofrecuencia. El enlace de radiofrecuencia a menudo soporta múltiples comunicaciones de voz simultáneas, que requieren recursos de ancho de banda compartidos como los que se pueden encontrar con la telefonía celular. Aunque no se muestra, el canal de comunicación 905 puede ser reemplazado por un dispositivo de almacenamiento en un único modo de realización del dispositivo del sistema de comunicación que graba y almacena la señal de voz codificada para su posterior reproducción.

45 Haciendo referencia todavía a la figura 27, un micrófono 901 produce una señal analógica de voz que es suministrada a un convertidor de analógico a digital (A/D) 902 para convertirla en una forma digital. Un codificador de voz 903 codifica la señal digital de voz produciendo de este modo un conjunto de parámetros de codificación que son codificados en una forma binaria y proporcionados a un codificador de canal 904. El codificador de canal opcional añade redundancia a la representación binaria de los parámetros de codificación antes de transmitirlos 50 sobre el canal de comunicación 905. En el lado del receptor, un decodificador de canal 906 utiliza la información redundante mencionada anteriormente en la secuencia de bits recibida para detectar y corregir los errores de canal que se han producido en la transmisión. Un decodificador de voz 907 convierte la secuencia de bits recibida desde el decodificador de canal 906 de nuevo en un conjunto de parámetros de codificación para crear una señal digital de voz sintetizada. La señal digital de voz sintetizada reconstruida en el decodificador de voz 907 es convertida en una 55 forma analógica en un convertidor de digital a analógico (D/A) 908 y reproducida en una unidad de hablante 909.

### Breve resumen de los antecedentes de CELP

Un códec de voz consta de dos partes básicas: un codificador y un decodificador. El codificador digitaliza la señal de audio, elige un número limitado de parámetros de codificación que representan la señal de voz y convierte estos parámetros en una secuencia digital de bits que se transmite al decodificador a través de un canal de comunicación.

5 El decodificador reconstruye la señal de voz para que sea lo más similar posible a la señal de voz original. Actualmente, una técnica de codificación de voz muy extendida se basa en la predicción lineal (LP - Linear Prediction, en inglés) y, más específicamente en la tecnología CELP. En la codificación basada en LP, la señal de voz es sintetizada filtrando una señal de excitación a través de un filtro de síntesis de todos los polos 1/A (z). En CELP, la excitación se compone típicamente de dos partes, se selecciona una señal de excitación de la primera etapa de un libro de códigos adaptativo y se selecciona una señal de excitación de la segunda etapa de un libro de códigos fijo. En general, la excitación de libro de códigos adaptativo modeliza la parte periódica de la excitación y la excitación del libro de códigos fija se añade para modelizar la evolución de la señal de voz.

10 La voz se procesa normalmente por tramas de típicamente 20 ms, y los coeficientes del filtro LP se transmiten una vez por trama. En CELP, cada trama está dividida en varias subtramas para codificar la señal de excitación. La longitud de subtrama es típicamente de 5 ms.

15 Haciendo referencia a las figuras 1a y 1b, el principio más importante en el que se basa el CELP se denomina análisis mediante síntesis, donde las posibles salidas del decodificador se prueban (síntesis) ya durante el proceso de codificación (análisis) y, a continuación, se comparan con la señal de voz original. La búsqueda minimiza un error cuadrático medio entre la señal de voz de entrada  $s(n)$  y la señal de voz sintetizada  $s'(n)$  en un dominio ponderado perceptualmente, donde el índice de tiempo discontinuo  $n = 0, 1, \dots, N-1$  y  $N$  es la longitud de la subtrama. El filtro de ponderación perceptual  $W(z)$  aprovecha el efecto de enmascaramiento de frecuencia y se deriva típicamente del filtro LP. Un ejemplo de filtro de ponderación perceptual  $W(z)$  se da en la siguiente ecuación (1):

$$20 W(z) = \frac{A(z/\gamma_1)}{A(z/\gamma_2)}, \quad (1)$$

25 en la que los factores  $\gamma_1$  y  $\gamma_2$  controlan la cantidad de ponderación perceptual y mantienen la relación  $0 < \gamma_2 < \gamma_1 \leq 1$ . Este filtro de ponderación perceptual tradicional funciona bien para señales de NB (banda estrecha - ancho de banda de 200 a 3400 Hz). En la Referencia [1], se puede encontrar un ejemplo de filtro de ponderación perceptual para las señales de WB (banda ancha - ancho de banda de 50 a 7000 Hz)

30 La secuencia de bits transmitida al decodificador contiene para las tramas sonoras los siguientes parámetros de codificación: los parámetros cuantificados del filtro de síntesis de LP, los índices del libro de códigos adaptativo y fijo y las ganancias de las partes adaptativa y fija.

### Búsqueda en el libro de códigos adaptativo

35 La búsqueda en el libro de códigos adaptativo en códecs basados en CELP se realiza en el dominio ponderado de voz para determinar el retardo (periodo de tono)  $t$  y la ganancia del tono  $g_p$  y para construir la parte casi periódica de la señal de excitación, denominada vector de código adaptativo  $v(n)$ . El periodo del tono depende en gran medida del hablante particular, y su determinación precisa influye de manera crítica en la calidad de la voz sintetizada.

40 En un códec EV-VBR, se utiliza un procedimiento de tres etapas para determinar el periodo y la ganancia de tono. En la primera etapa, se calculan tres estimaciones del tono de bucle abierto  $T_{op}$  (Open loop, en inglés) para cada trama -una estimación para cada semi-trama de 10 ms y una para una trama de anticipación de 10 ms- utilizando la señal de voz ponderada perceptualmente  $s_w(n)$  y el cálculo de la correlación normalizada. En la segunda etapa, se realiza una búsqueda del tono en bucle cerrado para períodos enteros alrededor de los períodos estimados del tono de bucle abierto  $T_{op}$  para cada subtrama. Una vez que se encuentra un período del tono entero óptimo, una tercera etapa de búsqueda examina las fracciones alrededor de ese valor entero óptimo. La búsqueda del tono en bucle cerrado se realiza minimizando el error ponderado cuadrático medio entre la voz original y la sintetizada. Esto se logra mediante la maximización del término

$$45 \quad \mathfrak{I}_t = \frac{\left( \sum_{n=0}^{N-1} x_1(n) y_1(n) \right)^2}{\sum_{n=0}^{N-1} y_1(n) y_1(n)}, \quad (2)$$

en el que  $x_1(n)$  es la señal objetivo y la señal de contribución de la primera etapa (también llamada vector de código adaptativo filtrado),  $y_1(n)$  se calcula mediante la convolución de la señal de excitación anterior  $v(n)$  en el periodo  $t$  con la respuesta de impulso  $h(n)$  del filtro de síntesis ponderada  $H(z)$

$$y_1(n) = v(n) * h(n) . \quad (3)$$

5 La señal de voz de entrada ponderada perceptualmente  $s_w(n)$  se obtiene procesando la señal de voz de entrada  $s(n)$  a través del filtro de ponderación perceptual  $W(z)$ . El filtro  $H(z)$  está formado por la cascada del filtro de síntesis de LP 1/A  $(z)$  y el filtro de ponderación perceptual  $W(z)$ . La señal objetivo  $x_1(n)$  corresponde a la señal de voz de entrada ponderada perceptualmente  $s_w(n)$  tras sustraer de ella la respuesta de entrada cero del filtro  $H(z)$ .

La ganancia del tono se encuentra minimizando el error cuadrático medio entre la señal  $x_1(n)$  y la señal de contribución de la primera etapa  $y_1(n)$ . La ganancia del tono se expresa mediante la siguiente ecuación:

$$g_p = \frac{\sum_{n=0}^{N-1} x_1(n)y_1(n)}{\sum_{n=0}^{N-1} y_1(n)y_1(n)} . \quad (4)$$

10 La ganancia del tono está entonces limitada por  $0 \leq g_p \leq 1,2$  y, típicamente, se cuantifica conjuntamente con la ganancia del libro de códigos fijo una vez que se encuentra la innovación.

15 En códecs basados en CELP, de este modo, la señal de excitación al inicio de la trama procesada actualmente se reconstruye a partir de la señal de excitación de la trama anterior. Este mecanismo es muy eficiente para segmentos sonoros de la señal de voz en los que la señal es casi periódica y en ausencia de errores de transmisión. En caso de borrado de tramas, la señal de excitación de la trama anterior se pierde y los respectivos libros de códigos adaptativos del codificador y del decodificador ya no son los mismos. En las tramas que siguen al borrado, el decodificador continúa sintetizando la voz utilizando el libro de códigos adaptativo con contenido incorrecto. Por consiguiente, un borrado de tramas degrada la calidad de la voz sintetizada no solo durante la trama borrada, sino que también puede degradar la calidad de la voz sintetizada durante varias tramas posteriores. Las técnicas de ocultación tradicionales a menudo se basan en repetir la forma de onda de la trama transmitida correctamente anterior, pero estas técnicas funcionan eficientemente solo en las partes de señal en las que las características de la señal de voz son casi estacionarias, por ejemplo, en segmentos de voz estables. En este caso, la diferencia entre los respectivos libros de códigos adaptativos del codificador y del decodificador es a menudo muy pequeña y la calidad de la señal sintetizada no resulta muy afectada. Sin embargo, si el borrado cae en una trama de transición, la eficiencia de estas técnicas es muy limitada. En sistemas de comunicación que utilizan códecs basados en CELP, en 20 los que la tasa de borrado de tramas (FER – Frame Erasure Rate, en inglés) es típicamente del 3% al 5%, la calidad de la voz sintetizada disminuye entonces de forma significativa.

25

30 Incluso en la transmisión de canal limpio, la eficiencia del libro de códigos adaptativo está limitada en las tramas de transición; el codificador CELP hace uso del libro de códigos adaptativo para aprovechar la periodicidad en la voz que es baja o inexistente durante las transiciones, por lo que la eficiencia de la codificación cae. Este es el caso de los inicios sonoros en particular, en los que la señal de excitación anterior y la señal de excitación óptima para la trama actual están correlacionadas muy débilmente o no lo están en absoluto.

### Búsqueda en un libro de códigos fijo

35 El objetivo de la contribución de la búsqueda en el libro de códigos fijo (innovación) (FCB – Fixed CodeBook, en inglés) en códecs basados en CELP es minimizar el error residual después de la utilización del libro de códigos adaptativo, es decir

$$\min_k \left\{ \sum_{n=0}^{N-1} [x_2(n) - g_c \cdot y_2^{(k)}(n)]^2 \right\} , \quad (5)$$

40 donde  $g_c$  es la ganancia del libro de códigos fijo, y la señal de contribución de la segunda etapa (llamada también vector de código fijo filtrado)  $y_2^{(k)}(n)$  es el vector de libro de códigos fijo  $c_k(n)$  convolucionado con  $h(n)$ . La señal objetivo  $x_1(n)$  se actualiza restando la contribución del libro de códigos adaptativo del objetivo de libro de códigos adaptativo para obtener:

$$x_2(n) = x_1(n) - g_p y_1(n) . \quad (6)$$

El libro de códigos fijo se puede realizar, por ejemplo, utilizando un libro de códigos algebraico tal como se describe en la Referencia [2]. Si  $c_k$  denota el vector de código algebraico en el índice  $k$ , entonces se realiza una búsqueda en el libro de códigos algebraico maximizando el siguiente criterio:

$$\mathfrak{I}_k = \frac{(\mathbf{x}_2^T \mathbf{y}_2)^2}{\mathbf{y}_2^T \mathbf{y}_2} = \frac{(\mathbf{x}_2^T \mathbf{H} \mathbf{c}_k)^2}{\mathbf{c}_k^T \mathbf{H}^T \mathbf{H} \mathbf{c}_k} = \frac{(\mathbf{d}^T \mathbf{c}_k)^2}{\mathbf{c}_k^T \Phi \mathbf{c}_k} = \frac{(C_k)^2}{E_k}, \quad (7)$$

donde  $\mathbf{H}$  es la matriz de convolución de Toeplitz inferior con diagonal  $h(0)$  y diagonales inferiores  $h(1), \dots, h(N-1)$ . El vector  $\mathbf{d} = \mathbf{H}^T \mathbf{x}_2$  es la correlación entre la señal objetivo actualizada  $\mathbf{x}_2(n)$  y  $h(n)$  (también conocido como vector objetivo filtrado hacia atrás), y la matriz  $\Phi = \mathbf{H}^T \mathbf{H}$  es la matriz de correlaciones de  $h(n)$ . El superíndice  $T$  denota la matriz o el vector traspuesta o traspuesto. Tanto  $\mathbf{d}$  como  $\Phi$  se calculan habitualmente antes de la búsqueda en el libro de códigos fijo. La Referencia [1] analiza que, si la estructura algebraica del libro de códigos fijo contiene solo unos pocos elementos distintos de cero, un cálculo del criterio de maximización para todos los posibles índices  $k$  es muy rápido. Un procedimiento similar se utiliza en la técnica de codificación de modo de transición (TM), como se verá a continuación.

10 Se cree que CELP es de otra manera bien conocida por las personas de habilidad ordinaria en la técnica y, por esa razón, no se describirán con más detalle en la presente memoria descriptiva.

#### Clasificación de las tramas en el códec EV-VBR

15 La clasificación de las tramas en el códec EV-VBR se basa en la clasificación de VMR-WB (Variable Rate Multi-Mode WideBand), tal como se describe en la Referencia [3]. La clasificación de VMR-WB se realiza con la consideración de la estrategia de ocultación y recuperación. En otras palabras, cualquier trama se clasifica de tal manera que la ocultación puede ser óptima si falta la siguiente trama, o que la recuperación puede ser óptima si se pierde la trama anterior. Algunas de las clases utilizadas para el procesamiento de ocultación de borrado de trama no necesitan ser transmitidas, ya que pueden deducirse sin ambigüedad en el decodificador. Se utilizan cinco clases distintas y se definen como sigue:

20 - La clase NO SONORA comprende todas las tramas de voz no sonoras y todas las tramas sin voz activa. Una trama desplazada sonora también se puede clasificar como NO SONORA si su final tiende a ser no sonoro, y la ocultación diseñada para las tramas no sonoras se puede utilizar para la siguiente trama en caso de que se pierda.

25 - La clase de TRANSICIÓN NO SONORA comprende tramas no sonoras con un posible inicio sonoro al final. El inicio sonoro es, sin embargo, todavía demasiado corto o no construido lo suficientemente bien como para utilizarla ocultación diseñada para las tramas sonoras. Una trama de TRANSICIÓN NO SONORA solo puede seguir a una trama clasificada como de TRANSICIÓN NO SONORA o NO SONORA.

30 - La clase de TRANSICIÓN SONORA comprende tramas sonoras con características sonoras relativamente débiles. Estas son típicamente tramas sonoras con características rápidamente cambiantes (transiciones entre vocales) o desplazamientos sonoros que duran toda la trama. Una trama de TRANSICIÓN SONORA puede seguir solamente a una trama clasificada como de TRANSICIÓN SONORA, SONORA o DE INICIO.

35 - La clase SONORA comprende las tramas sonoras con características estables. Una trama SONORA puede seguir solamente a una trama clasificada como de TRANSICIÓN SONORA, SONORA o DE INICIO.

40 - La clase DE INICIO comprende todas las tramas sonoras con características estables que siguen a una trama clasificada como NO SONORA o de TRANSICIÓN NO SONORA. Las tramas clasificadas como DE INICIO corresponden a tramas de inicio sonoras en las que el inicio ya está suficientemente bien construido para la utilización de la ocultación diseñada para tramas sonoras perdidas. Las técnicas de ocultación utilizadas para un borrado de trama siguiendo una trama clasificada como DE INICIO están en los códecs tradicionales basados en CELP lo mismo que siguiendo a una trama clasificada como SONORA, estando la diferencia en la estrategia de recuperación, cuando se puede utilizar una técnica especial para reconstruir artificialmente el inicio perdido. De acuerdo con el modo de realización ilustrativo no restrictivo de la presente invención, la técnica de codificación TM se utiliza con éxito en este caso.

45 El diagrama de estado de clasificación se describe en la figura 2. La información de clasificación se transmite utilizando 2 bits. Como puede verse en la figura 2, la clase de TRANSICIÓN NO SONORA y la clase de TRANSICIÓN SONORA se pueden agrupar, ya que se pueden diferenciar inequívocamente en el decodificador (una trama de TRANSICIÓN NO SONORA puede seguir solamente a tramas NO SONORAS o de TRANSICIÓN NO SONORA, una trama de TRANSICIÓN SONORA puede seguir solo a tramas DE INICIO, SONORAS o de TRANSICIÓN SONORA).

50 Para la clasificación se utilizan los siguientes parámetros: una correlación normalizada  $\bar{R}_{xy}$ , una medida de inclinación espectral  $e_t$ , un contador de estabilidad del tono  $pc$ , una energía de trama relativa a la señal de voz al final de la trama  $E_{rel}$  y un contador de cruces por cero  $zc$ . Tal como se puede ver en el siguiente análisis detallado, el cálculo de estos parámetros utiliza una anticipación. La anticipación permite estimar la evolución de la señal de voz

en la siguiente trama a estimar y, por consiguiente, la clasificación se puede hacer teniendo en cuenta el comportamiento futuro de la señal de voz.

La correlación normalizada media  $\bar{R}'_{xy}$  se calcula como una media de la correlación normalizada máxima de la segunda semitrama y la anticipación, utilizando la siguiente ecuación:

$$5 \quad \bar{R}'_{xy} = 0.5(C_{norm}(d_1) + C_{norm}(d_2)). \quad (8)$$

Las correlaciones normalizadas máximas  $C_{norm}$  se calculan como una parte de la búsqueda del tono de bucle abierto y corresponden a las correlaciones normalizadas maximizadas de dos períodos del tono adyacentes de la señal de voz ponderada.

10 El parámetro de inclinación espectral  $e'_t$  contiene la información sobre la distribución de frecuencia de la energía. La inclinación espectral para un análisis espectral se estima como una relación entre la energía concentrada en las bajas frecuencias y la energía concentrada en las altas frecuencias. En la presente memoria, la medida de inclinación utilizada es la media en el dominio logarítmico de las medidas de inclinación espectral  $e_{inclinación}(0)$  y  $e_{inclinación}(1)$ .

$$e'_t = 10 \log(e_{inclinación}(0) e_{inclinación}(1)). \quad (9)$$

15 El contador de estabilidad del tono  $pc$  evalúa la variación del periodo de tono. Se calcula como sigue:

$$pc = |T_{op1} - T_{op0}| + |T_{op2} - T_{op0}|. \quad (10)$$

Los valores  $T_{op0}$ ,  $T_{op1}$  y  $T_{op2}$  corresponden a las estimaciones del tono de bucle abierto de la primera mitad de la trama actual, de la segunda mitad de la trama actual y de la anticipación, respectivamente.

20 La energía de trama relativa  $E_{rel}$  se calcula como una diferencia en dB entre la energía de trama actual y la media de la energía de voz activa a largo plazo.

El último parámetro es el parámetro de cruces cero  $zc$  calculado en un segmento de 20 ms de la señal de voz. El segmento comienza en el centro de la trama actual y utiliza dos subtramas de la anticipación. En la presente memoria, el contador de cruces por cero  $zc$  cuenta el número de veces que el signo de la señal de voz cambia de positivo a negativo durante ese intervalo.

25 Para que la clasificación sea más robusta, los parámetros de clasificación se consideran conjuntamente formando una función del mérito  $f_m$ . Para este propósito, los parámetros de clasificación se escalan primero entre 0 y 1, de tal modo que el valor del parámetro típico para la señal de voz no sonora se traduce en 0 y el valor de cada parámetro típico para la señal de voz sonora se traduce en 1. Se utiliza una función lineal entre ellos. La versión escalada  $p^s$  de un determinado parámetro  $p_x$ , se obtiene utilizando la ecuación:

$$30 \quad p^s = k_p p_x + c_p \text{ restringida por } 0 \leq p^s \leq 1. \quad (11)$$

Los coeficientes de la función  $k_p$  y  $c_p$  han sido encontrados experimentalmente para cada uno de los parámetros, de manera que la distorsión de la señal debida a las técnicas de recogida y recuperación utilizadas en presencia de errores de trama es mínima. Los valores utilizados se resumen en la Tabla 1.

Tabla 1 - Parámetros de clasificación de señal y los coeficientes de sus respectivas funciones de escalado

Parámetro	Significado	$k_p$	$c_p$
$\bar{R}'_{xy}$	Correlación normalizada	2,857	-1,286
$e'_t$	Inclinación espectral	0,04167	0
$pc$	Contador de estabilidad de tono	-0,07143	1,857
$E_{rel}$	Energía de trama relativa	0,05	0,45
$zc$	Contador de cruces por cero	-0,04	2,4

35

Entonces, la función de mérito  $f_m$  se ha definido como:

$$f_m = \frac{1}{6} (2 \bar{R}_{xy}^s + e_t^s + pc^s + E_{rel}^s + zc_s), \quad (12)$$

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

Se toma una primera decisión de clasificación para la clase NO SONORA como sigue:

$$\text{Si (local\_VAD} = 0) \text{ o (} E_{rel} < -8 \text{) entonces clase} = \text{NO SONORA.} \quad (13)$$

- 5 donde local\_VAD representa la detección de actividad de voz local.

Si no se cumple la condición anterior (13), entonces la clasificación continua utilizando la función de mérito  $f_m$  y siguiendo las reglas resumidas en la Tabla 2.

Tabla 2 - Reglas de clasificación de señal en el codificador

Clase de la trama anterior	Regla	Clase de la trama actual
DE INICIO SONORA DE TRANSICIÓN SONORA	$f_m \geq 0,66$	SONORA
	$0,66 > f_m \geq 0,49$	DE TRANSICIÓN SONORA
	$f_m < 0,49$	NO SONORA
DE TRANSICIÓN NO SONORA NO SONORA	$f_m > 0,63$	DE INICIO
	$0,63 \geq f_m > 0,585$	DE TRANSICIÓN NO SONORA
	$f_m \leq 0,585$	NO SONORA

- 10 La información de clase se codifica con dos bits, tal como se ha explicado anteriormente en la presente memoria. A pesar de que la información suplementaria, que mejora la ocultación del borrado de trama, se transmite solamente en tramas genéricas, la clasificación se realiza para cada trama. Esto es necesario para mantener actualizada la máquina de estados de clasificación, ya que utiliza la información relativa a la clase de la trama anterior. Sin embargo, la clasificación es sencilla para los tipos de codificación dedicados a las tramas NO SONORA o SONORA.  
 15 Por ello, las tramas sonoras siempre se clasifican como SONORAS y las tramas no sonoras siempre se clasifican como NO SONORAS.

#### Selección de trama para codificación TM

- 20 Tal como se describió anteriormente, la técnica que se describe reemplaza el libro de códigos adaptativo en codificadores basados en CELP por un libro de códigos de forma glotal, para mejorar la robustez a los borrados de trama y para mejorar la eficiencia de la codificación cuando se procesan tramas de voz no estacionarias. Esto significa que esta técnica no construye la señal de excitación de la primera etapa con la utilización de la excitación anterior, sino que selecciona la señal de excitación de la primera etapa de del libro de códigos de forma glotal. La señal de excitación de la segunda etapa (la parte de innovación de la excitación total) se selecciona todavía del libro de códigos fijo CELP tradicional. Cualquiera de estos libros de códigos no utiliza ninguna información de las tramas de voz anteriores (previamente transmitidas), eliminando así la razón principal de la propagación de errores de trama inherente a los codificadores basados en CELP.

- 25 La utilización sistemática de la técnica de codificación TM (para codificar todas las tramas) limitaría en gran medida la propagación de errores, pero la eficiencia de la codificación y la calidad de la voz sintetizada caería en una situación sin errores. Como compromiso entre el rendimiento de canal limpio del códec y su robustez frente a errores de canal, la técnica de codificación TM se puede aplicar solamente a las tramas de transición y a varias tramas siguientes a cada trama de transición. Para la robustez del borrado de trama, la técnica de codificación TM se puede utilizar para tramas de voz sonoras siguientes a transiciones. Tal como se ha introducido anteriormente, estas transiciones comprenden básicamente los inicios sonoros y las transiciones entre dos sonidos sonoros diferentes. Para seleccionar las tramas pertinentes a codificar mediante la técnica de codificación TM, se detectan transiciones.  
 30 Aunque se puede utilizar cualquier detector de transiciones, el modo de realización ilustrativo no restrictivo utiliza la clasificación del marco EV-VBR tal como se ha descrito anteriormente en la presente memoria.

- 35 La técnica de codificación TM se puede aplicar para codificar tramas de transición (inicio sonoro o transición entre dos sonidos sonoros diferentes) tal como se describió anteriormente y varias tramas posteriores. El número de tramas TM (tramas codificadas utilizando la técnica de codificación TM) es una cuestión de compromiso entre el rendimiento del códec en condiciones de canal limpio y en condiciones con errores de canal. Si solo se codifican las tramas de transición (inicio sonoro o transición entre dos sonidos sonoros diferentes) utilizando la técnica de

codificación TM, la eficiencia de la codificación aumenta. Este aumento se puede medir por el aumento de la relación de señal a ruido (SNR – Signal-to-Noise Ratio, en inglés) del segmento, por ejemplo. La SNR se calcula utilizando la siguiente ecuación:

$$SNR = \frac{E_{sd}}{E_e}, \quad (14)$$

- 5 donde  $E_{sd}$  es la energía de la señal de voz de entrada de la trama actual y  $E_e$  es la energía del error entre esta señal de voz de entrada y la señal de voz de síntesis de la trama actual.

Sin embargo, la utilización de la técnica de codificación TM para codificar solo las tramas de transición no ayuda demasiado a la robustez frente a errores; si se pierde la trama de transición (inicio sonoro o transición entre dos sonidos sonoros diferentes), el error se propagará, ya que las siguientes tramas se codificarán utilizando el procedimiento CELP estándar. Por otra parte, si se pierde la trama que precede a la transición (inicio sonoro o transición entre dos sonidos sonoros diferentes), el efecto de esta trama precedente perdida en el rendimiento no es crítico, incluso sin la utilización de la técnica de codificación TM. En el caso de las transiciones de inicio sonoras, es probable que la trama que precede al inicio sea no sonora y la contribución del libro de códigos adaptativo no sea muy importante. En el caso de una transición entre dos sonidos sonoros, la trama antes de la transición es generalmente bastante estacionaria y los estados del libro de códigos adaptativo en el codificador y en el decodificador son a menudo similares después del borrado de trama.

20 Para aumentar la robustez, las tramas siguientes a la transición (inicio sonoro o transición entre dos sonidos sonoros diferentes) se pueden codificar utilizando la técnica de codificación TM. Si la mejora del rendimiento del canal limpio no es importante, la técnica de codificación TM solo se puede utilizar en las tramas siguientes a las tramas de transición. Básicamente, el número de tramas TM consecutivas depende del número de borrados consecutivos de trama que se deseé considerar para protección. Si solo se consideran borrados aislados (es decir, un borrado de tramas aisladas a la vez), basta con codificar solamente la trama siguiente a la transición (inicio sonoro o transición entre dos sonidos sonoros diferentes). Si se pierde la trama de transición (inicio sonoro o transición entre dos sonidos sonoros diferentes), se codifica la siguiente trama sin la utilización de la señal de excitación anterior y se interrumpe la propagación del error. Es preciso señalar, sin embargo, que si la trama de transición (inicio sonoro o transición entre dos sonidos sonoros diferentes) se transmite correctamente, pero se pierde la siguiente trama, la propagación de errores no se evitaría, ya que la siguiente trama ya está utilizando la codificación clásica de CELP. Sin embargo, la distorsión probablemente será limitada si al menos un periodo del tono ya está bien construido al final de la transición (inicio sonoro o transición entre dos sonidos sonoros diferentes), tal como se muestra en la figura 3.

30 Cuando se realiza una implementación de la técnica de codificación TM en algún códec existente y se conocen la clase de la trama actual y el modo de codificación, se puede utilizar el siguiente esquema para establecer el inicio y las tramas siguientes para la codificación TM. Un estado de parámetro, que es un contador de las tramas TM consecutivas utilizadas previamente, se almacena en la memoria de estado del codificador. Si el valor de este estado del parámetro es negativo, no se puede utilizar la codificación TM. Si el valor de este estado del parámetro no es negativo, sino menor o igual que número de borrados consecutivos de tramas a proteger, y la clase de la trama es DE INICIO, SONORA o DE TRANSICIÓN SONORA, la trama se denomina trama TM (véase la figura 4 para más detalles). En otras palabras, la trama se denota como trama TM si  $N_{TM\_FRAMES} \geq \text{estado} > 0$ , donde  $N_{TM\_FRAMES}$  es un número de tramas consecutivas para evitar la utilización de la técnica de codificación TM.

40 Si se espera que las características del canal de comunicación sean tales que más de una trama aislada a menudo se borren a la vez, es decir, que los borrados de trama tengan la tendencia a aparecer en paquetes, la mejor solución podría ser utilizar la técnica de codificación TM para proteger dos o incluso más borrados consecutivos de tramas. Sin embargo, la eficiencia de la codificación en condiciones de canal limpio disminuirá. Si se dispone de una retroalimentación sobre el canal en el codificador, el número de tramas TM consecutivas se puede adaptar a las condiciones de transmisión. En el modo de realización ilustrativo no limitativo de la presente invención, se consideran hasta dos tramas TM siguientes a la transición (inicio sonoro o transición entre dos sonidos sonoros diferentes), lo que corresponde a un diseño capaz de hacer frente hasta a dos borrados consecutivos de trama.

50 La decisión descrita anteriormente utiliza básicamente un número fijo (ya sea que este número esté fijo antes de la transmisión o sea dependiente de las condiciones de transmisión del canal) de tramas TM siguientes a la transición (inicio sonoro o transición entre dos sonidos sonoros diferentes). El compromiso entre el rendimiento de canal limpio y la robustez frente a errores de trama también se puede basar en una clasificación de bucle cerrado. Más específicamente, en la trama que se quiere proteger frente al borrado de trama anterior o se quiere decidir si se trata de la trama de inicio, se realiza un cálculo en paralelo de los dos posibles modos de codificación; la trama se procesa utilizando tanto el modo de codificación genérico (CELP) como la técnica de codificación TM. El rendimiento de ambos planteamientos se compara a continuación utilizando una medida SNR, por ejemplo; para obtener más detalles, consulte la sección siguiente titulada "Rendimiento de la técnica de codificación TM en el códec EV-VBR". Cuando la diferencia entre la SNR para el modo de codificación genérico (CELP) y la SNR para la técnica de

codificación TM es superior a un umbral dado, se aplica el modo de codificación genérico (CELP). Si la diferencia entre la SNR para el modo de codificación genérico (CELP) y la SNR para la técnica de codificación TM es inferior al umbral dado, se aplica la técnica de codificación TM. El valor del umbral se elige dependiendo de lo fuerte se necesite que sea la protección frente al borrado de tramas y la determinación de la codificación de inicio.

5 **Selección de subtramas para la búsqueda en libro de códigos de forma glotal**

En la sección anterior, se describieron las razones y mecanismos para la selección de tramas para codificación utilizando la técnica de codificación TM. A continuación, se mostrará que generalmente es más eficiente no utilizar el libro de códigos de forma glotal en todas las subtramas para conseguir el mejor compromiso entre el rendimiento de canal limpio a una tasa de bits dada y el rendimiento en presencia de un borrado en las tramas que preceden a las tramas TM. En primer lugar, la búsqueda en el libro de códigos de forma glotal es importante solo en el primer periodo del tono en una trama. Los siguientes periodos del tono se pueden codificar utilizando la búsqueda en libro de códigos adaptivo estándar más eficiente, dado que ya no utilizan la excitación de la trama anterior (cuando se realiza una búsqueda en el libro de códigos adaptativo, se busca la excitación hasta aproximadamente un periodo del tono en el pasado). Por consiguiente, no hay razón para emplear la búsqueda en el libro de códigos de forma glotal en subtramas que no contienen ninguna parte del primer periodo del tono de una trama.

De manera similar, cuando se utiliza la búsqueda en el libro de códigos de forma glotal para aumentar la eficiencia de la codificación en tramas de inicio sonoras, esta búsqueda en el libro de códigos de forma glotal se utiliza en el primer periodo del tono del segmento de voz inicial. La razón es que, para el primer periodo de tono, el libro de códigos adaptativo contiene una señal similar al ruido (el segmento anterior fue no sonoro), y reemplazarla por un impulso glotal cuantificado a menudo aumenta la eficiencia de la codificación. Sin embargo, para los siguientes periodos de tono, la excitación periódica ya se ha acumulado en el libro de códigos adaptativo y la utilización de este libro de códigos dará mejores resultados. Por esta razón, la información relativa a la posición del inicio sonoro está disponible al menos con la resolución de subtrama.

Otra optimización de la asignación de bits se refiere a tramas con periodos del tono más largos que la longitud de la subtrama. Dado que el libro de códigos de forma glotal contiene formas cuantificadas del impulso glotal, el libro de códigos es más adecuado para ser utilizado en subtramas que contienen el impulso glotal. En otras subtramas, su eficiencia es baja. Dado que la tasa de bits está a menudo bastante limitada en las aplicaciones de codificación de voz, y que la codificación del libro de códigos de forma glotal requiere un número relativamente mayor de bits para codificación de voz de tasa de bits baja, se eligió en el modo de realización no restrictivo, ilustrativo, una asignación de bits en la que se utiliza el libro de códigos de forma glotal y en el que se realiza una búsqueda solo en una subtrama por cada trama.

Para elegir la subtrama a codificar con el libro de códigos de forma glotal se busca el primer impulso glotal en la señal residual LP. Se puede utilizar el siguiente procedimiento sencillo. La muestra máxima en la señal residual LP se busca en el rango  $[0, 0 + T_{op} + 2]$ , donde  $T_{op}$  es el periodo del tono de bucle abierto para la primera semitrama, y 0 corresponde al inicio de la trama. En el caso de tramas de inicio sonoras, y si el comienzo del inicio se puede determinar de manera fiable, 0 denota el comienzo de la subtrama en la que se encuentra el comienzo del inicio. El libro de códigos de forma glotal se empleará entonces en la subtrama con la máxima energía de señal residual. Además, la posición del máximo proporciona información acerca de dónde puede estar situada aproximadamente la posición del impulso glotal, y esta situación se puede aprovechar para reducir la complejidad, tal como se analizará a continuación. Obsérvese que cuando la búsqueda en el libro de códigos de forma glotal reemplaza solamente la búsqueda en el libro de códigos adaptativo, se realiza una búsqueda en el libro de códigos fijo en cada subtrama de una trama TM.

Las otras subtramas (no codificadas con la utilización del libro de códigos de forma glotal) se procesarán de la siguiente manera. Si la subtrama que utiliza la búsqueda en el libro de códigos de forma glotal no es la primera subtrama en la trama, la señal de excitación en subtrama precedente o las subtramas precedentes de la trama se codifica utilizando solo el libro de códigos fijo de CELP; esto significa que la señal de excitación de la primera etapa es cero. Si la subtrama de libro de códigos de forma glotal no es la última subtrama en la trama, se procesa la subtrama siguiente o las subtramas siguientes de la trama utilizando codificación CELP estándar (es decir, utilizando la búsqueda en los libros de códigos adaptativo y fijo). En las figuras 5a a 5c, se muestra la situación para el caso en el que el primer impulso glotal emerge en la segunda subtrama. En la figura 5b,  $u(n)$  es la señal residual LP. La señal de excitación de la primera etapa se denomina  $q_k(n)$  cuando se construye utilizando el libro de códigos de forma glotal, o  $v(n)$  cuando se construye utilizando el libro de códigos adaptativo. En este ejemplo (figura 5c), la señal de excitación de la primera etapa es cero en la primera subtrama, es un vector de código de la forma glotal en la segunda subtrama y un vector de libro de códigos adaptativo en las dos últimas subtramas.

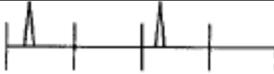
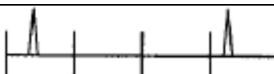
55 Con el fin de aumentar aún más la eficiencia de la codificación y optimizar la asignación de bits, se utilizan diferentes procesamientos en subtramas particulares de una trama TM dependiente del periodo del tono. Cuando se elige la primera subtrama como subtrama TM, se determina la subtrama con el 2º impulso glotal en la señal residual LP. Esta determinación se basa en el valor del periodo del tono y en las siguientes cuatro situaciones pueden ocurrir. En la primera situación, el 2º impulso glotal está en la 1ª subtrama, y las subtramas 2ª, 3ª y 4ª se procesan utilizando codificación CELP estándar (búsqueda en los libros de códigos adaptativo y fijo). En la segunda situación, el 2º

impulso glotal está en la 2<sup>a</sup> subtrama, y las 2<sup>a</sup>, 3<sup>a</sup> y 4<sup>a</sup> se procesan utilizando nuevamente la codificación CELP estándar. En la tercera situación, el 2º impulso glotal está en la 3<sup>a</sup> subtrama. La 2<sup>a</sup> subtrama se procesa utilizando solo búsqueda en el libro de códigos fijo, ya que no existe impulso glotal en la 2<sup>a</sup> subtrama de la señal residual LP en la que se va a buscar para utilizar el libro de códigos adaptativo. Las subtramas 3<sup>a</sup> y 4<sup>a</sup> se procesan utilizando la codificación CELP estándar. En la última (cuarta) situación, el 2º impulso glotal está en la 4<sup>a</sup> subtrama (o en la trama siguiente), las subtramas 2<sup>a</sup> y 3<sup>a</sup> se procesan utilizando solo búsqueda en el libro de códigos fijo, y la 4<sup>a</sup> subtrama se procesa utilizando codificación CELP estándar. Un análisis más detallado se proporciona en una implementación de ejemplo más adelante en lo que sigue.

5 La Tabla 3 muestra nombres de las posibles configuraciones de codificación y sus estadísticas de ocurrencia. En otras palabras, la Tabla 3 proporciona la distribución de la primera y segunda ocurrencia de impulso glotal en cada subtrama para tramas procesadas con la técnica de codificación TM. La Tabla 3 corresponde al escenario en el que se utiliza la técnica de codificación TM para codificar solamente la trama de inicio sonora y una trama posterior. La longitud de la trama de la señal de voz en este experimento fue de 20 ms, la longitud de la subtrama, de 5 ms, y el experimento se realizó utilizando voces de 32 hombres y 32 mujeres (si no se menciona otra cosa, la misma base de datos de voz se utilizó también en todos los demás experimentos mencionados en la siguiente descripción).

10 15 La Tabla 3 muestra nombres de las posibles configuraciones de codificación y sus estadísticas de ocurrencia. En otras palabras, la Tabla 3 proporciona la distribución de la primera y segunda ocurrencia de impulso glotal en cada subtrama para tramas procesadas con la técnica de codificación TM. La Tabla 3 corresponde al escenario en el que se utiliza la técnica de codificación TM para codificar solamente la trama de inicio sonora y una trama posterior. La longitud de la trama de la señal de voz en este experimento fue de 20 ms, la longitud de la subtrama, de 5 ms, y el experimento se realizó utilizando voces de 32 hombres y 32 mujeres (si no se menciona otra cosa, la misma base de datos de voz se utilizó también en todos los demás experimentos mencionados en la siguiente descripción).

Tabla 3 - Configuraciones del modo de codificación para TM y su ocurrencia cuando se procesa una señal de voz.

Configuración de codificación	Posición(s) del primer (y del segundo, si es relevante) impulso(s) glotal	Tipo de libro de códigos utilizado (GS = de forma glotal, A = Adaptativo, F = fijo)				Cantidad [%]
		1 <sup>a</sup> subtr.	2 <sup>a</sup> subtr.	3 <sup>a</sup> subtr.	4 <sup>a</sup> subtr.	
TRANSICIÓN_1_1		GS+F	A+F	A+F	A+F	25,5
TRANSICIÓN_1_2		GS+F	A+F	A+F	A+F	28,4
TRANSICIÓN_1_3		GS+F	F	A+F	A+F	16,3
TRANSICIÓN_1_4		GS+F	F	F	A+F	3,0
TRANSICIÓN_2		F	GS+F	A+F	A+F	21,2
TRANSICIÓN_3		F	F	GS+F	A+F	4,6
TRANSICIÓN_4		F	F	F	GS+F	1,0

#### Libro de códigos de forma glotal

20 En principio, el libro de códigos de forma glotal consiste en formas numéricas cuantificadas de los impulsos glotales colocados en una posición específica. En consecuencia, la búsqueda en el libro de códigos consiste tanto en la selección de la mejor forma como en la determinación de su mejor posición en una subtrama particular. En su forma más simple, la forma del impulso glotal se puede representar por un impulso unitario y no necesita ser cuantificada. En ese caso, solo se determina su posición en la subtrama. Sin embargo, el rendimiento de un libro de códigos tan simple es muy limitado.

25 Por otra parte, la mejor representación se obtendría probablemente si la longitud L de las entradas del libro de códigos de forma glotal correspondiese a la longitud del periodo de tono, y si se representan un gran número de

formas de impulso glotal. Dado que la longitud y la forma de los impulsos glotales varían de hablante a hablante y de trama a trama, la complejidad y los requisitos de memoria para realizar una búsqueda en dicho libro de códigos y para almacenar sería demasiado extensa. Como solución de compromiso, la longitud de los impulsos glotales, así como su número debe ser limitada.

5 En el modo de realización ilustrativo no restrictivo, el libro de códigos de forma glotal está compuesto por ocho (8) formas diferentes de impulso glotal y la longitud de cada impulso glotal es  $L = 17$  muestras. Las formas cuantificadas han sido seleccionadas de tal manera que el máximo absoluto está alrededor de la mitad de esta longitud. Durante la búsqueda en el libro de códigos de forma glotal, este centro está alineado con el índice  $k'$ , que representa la posición del impulso glotal en la subtrama actual y se elige entre el intervalo  $[0, N - 1]$ , siendo  $N$  la longitud de la subtrama. Dado que la longitud de 17 muestras de las entradas del libro de códigos es

10 más corta que la longitud de la subtrama, las muestras restantes se ponen a cero.

El libro de códigos de forma glotal está diseñado para representar tantos impulsos glotales existentes como sea posible. Se utilizó un proceso de entrenamiento basado en el algoritmo k-means [4]; el libro de códigos de forma glotal fue entrenado utilizando más de tres (3) horas de señal de voz compuestas por enunciados de muchos hablantes diferentes hablando en varios idiomas diferentes. A partir de esta base de datos, los impulsos glotales han sido extraídos de la señal residual LP y truncados a 17 muestras alrededor del valor absoluto máximo. A partir de las dieciséis (16) formas seleccionadas por el algoritmo k-means, el número de formas se ha reducido a ocho (8) formas experimentalmente utilizando una medida de calidad SNR segmentada. El libro de códigos de forma glotal seleccionado se muestra en la figura 6. Obviamente, se pueden utilizar otros medios para diseñar el libro de códigos de forma glotal.

## 20 Búsqueda en el libro de códigos de forma glotal

El modo de realización real del libro de códigos de forma glotal se puede realizar de varias maneras. Por ejemplo, la búsqueda se puede realizar de manera similar a la búsqueda en el libro de códigos fijos en CELP. En este caso, el libro de códigos se construye colocando el centro de las formas de impulso glotal en todas las posiciones posibles en la subtrama. Por ejemplo, para una longitud de subtrama de sesenta y cuatro (64) muestras y ocho (8) formas de impulso glotales, se obtiene un libro de códigos de forma glotal de tamaño  $64 \times 8 = 512$  vectores de código. De acuerdo con otro ejemplo, de manera similar a la búsqueda en el libro de códigos adaptivo, las entradas del libro de códigos se pueden colocar sucesivamente en todas las posiciones potenciales en la excitación anterior, y se puede seleccionar la mejor combinación de forma / posición de manera similar a la utilizada en la búsqueda en el libro de códigos adaptativo. En la última realización, toda la repetición del ciclo del tono se realiza automáticamente a través del filtro CELP a largo plazo, y los impulsos glotales se representan con formas de tamaño total (en contraste con la primera realización, en la que el truncamiento de la forma glotal es necesario en casos fronterizos, tal como se discutirá más adelante).

35 El modo de realización ilustrativo no restrictivo utiliza la configuración en la que la búsqueda en el libro de códigos es similar a la búsqueda en el libro de códigos fijo en CELP algebraico (ACELP). En este planteamiento, para cada una de las formas candidatas, la forma se representa como una respuesta de impulso de un filtro de conformación  $G(z)$ . Por lo tanto, los vectores de código correspondientes a las formas de impulso glotal centradas en diferentes posiciones pueden ser representados mediante vectores de código que contienen solo un elemento distinto de cero filtrado a través del filtro de conformación  $G(z)$  (para un tamaño de subtrama  $N$  existen  $N$  vectores de impulso único para potenciales posiciones del impulso glotal  $k'$ ).

40 Debido a que la posición  $k'$  del impulso glotal está en el centro de la forma glotal con una longitud impar de  $L$  muestras y  $k'$  pertenece al rango  $[0, N-1]$ , la forma glotal debe ser truncada para las muestras primera y última  $L_{1/2} = (L - 1) / 2$  muestras. Esto se tendrá en cuenta durante la búsqueda del impulso glotal, puesto que hace que el filtro de conformación  $G(z)$  sea un filtro no causal.

45 La configuración de la parte de TM se muestra en la figura 7 para el codificador, y en la figura 11 para el decodificador. Tal como ya se ha mencionado, la parte de TM reemplaza a la parte del libro de códigos adaptativo del codificador / decodificador. Durante la búsqueda, la respuesta de impulso del filtro de conformación  $G(z)$  se puede integrar en la respuesta de impulso del filtro  $H(z)$ .

50 A continuación, se describirá un procedimiento y un buscador de libro de códigos correspondiente para buscar la posición central óptima del impulso glotal  $k'$  para una cierta forma del impulso glotal proporcionada por el filtro de conformación  $G(z)$ . Debido a que la forma del filtro  $G(z)$  se elige entre varias formas candidatas (se utilizan ocho (8) formas en el modo de realización ilustrativo no restrictivo tal como se ilustra en la figura 6), el procedimiento de búsqueda se debe repetir para cada forma glotal del libro de códigos, con el fin de encontrar la forma y posición óptimas del impulso.

55 Para determinar los parámetros de la codificación TM, la búsqueda determina el error medio cuadrático entre el vector objetivo  $x_1$  y el vector de código de la forma glotal centrado en la posición  $k'$  que se filtra a través del filtro de síntesis ponderado  $H(z)$ . De manera similar a CELP, la búsqueda se puede realizar encontrando el máximo de un criterio en la forma:

$$\mathfrak{I}_{k'} = \frac{(\mathbf{x}_1^T \mathbf{y}_1)^2}{\mathbf{y}_1^T \mathbf{y}_1}, \quad (15)$$

donde  $\mathbf{y}_1$  es el vector de código de la forma glotal filtrado. Sea  $\mathbf{q}_k$  el vector de código de la forma glotal centrado en la posición  $k'$ , y  $\mathbf{p}_k$ , un vector de código de posición con un (1) elemento distinto de cero que indica la posición  $k'$ , entonces  $\mathbf{q}_k$ , se puede escribir como  $\mathbf{q}_k = \mathbf{G} \cdot \mathbf{p}_k$ , donde  $\mathbf{G}$  es una matriz de Toeplitz que representa la forma del impulso glotal. Por lo tanto, de manera similar a la búsqueda en el libro de códigos fijo, se puede escribir la siguiente ecuación:

$$\begin{aligned} \mathfrak{I}_{k'} &= \frac{(\mathbf{x}_1^T \mathbf{y}_1)^2}{\mathbf{y}_1^T \mathbf{y}_1} = \frac{(\mathbf{x}_1^T \mathbf{H} \mathbf{q}_{k'})^2}{\mathbf{q}_{k'}^T \mathbf{H}^T \mathbf{H} \mathbf{q}_{k'}} = \frac{(\mathbf{x}_1^T \mathbf{H} \mathbf{G} \mathbf{p}_{k'})^2}{\mathbf{p}_{k'}^T \mathbf{G}^T \mathbf{H}^T \mathbf{H} \mathbf{G} \mathbf{p}_{k'}} = \\ &= \frac{(\mathbf{x}_1^T \mathbf{Z} \mathbf{p}_{k'})^2}{\mathbf{p}_{k'}^T \mathbf{Z}^T \mathbf{Z} \mathbf{p}_{k'}} = \frac{(\mathbf{d}_g^T \mathbf{p}_{k'})^2}{\mathbf{p}_{k'}^T \Phi_g \mathbf{p}_{k'}} = \frac{(C_{k'})^2}{E_{k'}}, \end{aligned} \quad (16)$$

donde  $\mathbf{H}$  es la matriz de convolución de Toeplitz triangular inferior del filtro de síntesis ponderado. Tal como se analizará más adelante, las filas de la matriz  $\mathbf{Z}^T$  corresponden a la versión desplazada filtrada de la forma de impulso glotal o su representación truncada. Se debe tener en cuenta que todos los vectores en este texto se suponen vectores de columna (matrices  $N \times 1$ .)

Un ejemplo de la matriz  $\mathbf{G}$  en forma traspuesta ( $\mathbf{G}^T$ ) para una longitud de impulso de tres (3) muestras y  $N = 4$  tendría la forma:

$$\mathbf{G}^T = \begin{bmatrix} g(0) & g(1) & 0 & 0 \\ g(-1) & g(0) & g(1) & 0 \\ 0 & g(-1) & g(0) & g(1) \\ 0 & 0 & g(-1) & g(0) \end{bmatrix}, \quad (17)$$

15 donde  $g(n)$  son los coeficientes de la respuesta de impulso del filtro de conformación no causal  $G(z)$ . En la siguiente descripción, los coeficientes del filtro de conformación no causal  $G(z)$  vienen dados por los valores  $g(n)$ , para  $n$  situado dentro del intervalo  $[-L_{1/2}, L_{1/2}]$ . Debido al hecho de que el vector de código de posición  $\mathbf{p}_k$  tiene solo un elemento distinto de cero, el cálculo del criterio (16) es muy simple y se puede expresar utilizando la siguiente ecuación:

$$\mathfrak{I}_{k'} = \frac{(d_g(k'))^2}{\Phi_g(k', k')}. \quad (18)$$

20 Tal como se puede ver a partir de la ecuación (18), solo es preciso calcular la diagonal de la matriz  $\Phi_g$ . En las figuras 8 y 9 se muestra una representación gráfica del cálculo del criterio (18) para un vector de código de la forma glotal. Tal como se ha mencionado ya, la ecuación (18) se utiliza típicamente en la búsqueda en el libro de códigos algebraico ACELP, calculando previamente el vector objetivo filtrado hacia atrás  $\mathbf{d}_g$  y la matriz de correlación  $\Phi_g$ . Sin embargo, dada la naturaleza no causal del filtro de conformación  $G(z)$ , esto no puede aplicarse directamente para las primeras  $L_{1/2}$  posiciones. En estas situaciones se utiliza una búsqueda más sofisticada, donde algunos valores calculados pueden ser reutilizados para mantener la complejidad en un nivel bajo. Esto se describirá a continuación.

30 Sea  $\mathbf{z}_k$ , la fila de orden  $(k' + 1)$  de la matriz  $\mathbf{Z}^T$ , donde la matriz  $\mathbf{Z}^T$  (figura 10) se calcula como sigue. Dada la naturaleza no causal del filtro de conformación  $G(z)$ , la matriz  $\mathbf{Z}^T$  se calcula en dos etapas, para minimizar la complejidad de cálculo. Primero se calculan las primeras  $L_{1/2} + 1$  filas de esta matriz. Para el resto de la matriz  $\mathbf{Z}^T$  (las últimas  $N - L_{1/2} - 1$  filas de la matriz  $\mathbf{Z}^T$ ), se utiliza el criterio (18) de manera similar a la búsqueda en el libro de códigos fijo ACELP.

A continuación, se describirá una descripción detallada de cómo calcular la matriz  $\mathbf{Z}^T$  y el criterio (18).

35 En la primera etapa, se calculan las primeras  $L_{1/2} + 1$  filas de la matriz  $\mathbf{Z}^T$  que corresponden a las posiciones  $k'$  dentro del intervalo  $[0, L_{1/2}]$ . Para estas posiciones se utiliza una forma glotal truncada diferente para cada posición  $k'$

dentro de este intervalo. En una primera operación, se calcula una convolución entre la respuesta de la forma glotal para la posición  $k' = 0$  y la respuesta de impulso  $h(n)$  utilizando la ecuación:

$$z_0(n) = \sum_{i=0}^n g(n-i)h(i), \quad (19)$$

5 donde se aprovecha el hecho de que el filtro de conformación  $G(z)$  tiene solo  $L_{1/2} + 1$  coeficientes distintos de cero, es decir  $g(0), g(1), \dots, g(L_{1/2})$  son coeficientes distintos de cero.

En una segunda operación, se calcula la convolución  $z_1(n)$  entre la respuesta de libro de códigos de forma glotal para la posición  $k' = 1$  y la respuesta de impulso  $H(z)$ , reutilizando los valores de  $z_0(n)$  tal como sigue (la matriz  $Z^T = G^T \cdot H^T$  es una matriz con algunas diagonales de pendiente negativa iguales a cero, pero esta matriz  $Z^T$  ya no es una matriz Toeplitz y triangular, tal como se muestra en la figura 10):

$$z_1(0) = g(-1)h(0) \quad (20)$$

$$z_1(n) = z_0(n-1) + g(-1)h(n) \text{ para } n = 1, \dots, N-1.$$

10 Para las siguientes filas se reutiliza la recursividad en la ecuación (21):

$$z_{k'}(0) = g(-k')h(0) \quad (21)$$

$$z_{k'}(n) = z_{k'-1}(n-1) + g(-k')h(n) \text{ para } n = 1, \dots, N-1$$

Se repite la recursividad (21) para todo  $k' \leq L_{1/2}$ . Para  $k' = L_{1/2}$  el filtro de modelizado  $G(z)$  tiene ya  $L$  coeficientes distintos de cero y la fila de orden  $(L_{1/2} + 1)$  de la matriz  $Z^T$  se obtiene de este modo mediante

$$z_{L_{1/2}}(0) = g(-L_{1/2})h(0) \quad (22)$$

$$z_{L_{1/2}}(n) = z_{L_{1/2}-1}(n-1) + g(-L_{1/2})h(n) \text{ para } n = 1, \dots, N-1.$$

15 En este punto, se han calculado las primeras  $L_{1/2} + 1$  filas de la matriz  $Z^T$ . Estas filas no contienen ningún coeficiente igual a cero (figura 10). Entonces, el criterio (18) se puede calcular para  $k'$  dentro del rango  $[0, L_{1/2}]$  utilizando la ecuación:

$$\mathfrak{I}_{k'} = \frac{\left( \sum_{i=0}^{N-1} z_{k'}(i) \cdot x_1(i) \right)^2}{\sum_{i=0}^{N-1} z_{k'}(i) \cdot z_{k'}(i)}. \quad (23)$$

20 En la segunda etapa se calcula el resto de la matriz  $Z^T$  y se evalúa el criterio (18) para las posiciones  $k'$  dentro del intervalo  $[L_{1/2} + 1, N - 1]$ . Se aprovecha el hecho de que las filas  $L_{1/2} + 1, \dots, N - 1$  de la matriz  $Z^T$  se construyen utilizando coeficientes de la convolución  $z_{L_{1/2}}(n)$  que ya han sido calculados tal como se describe mediante la ecuación (22). La diferencia es que solo se necesita una parte de los coeficientes para calcular estas filas. Es decir, cada fila corresponde a la fila anterior desplazada a la derecha por 1 y añadiendo un cero al principio:

$$z_{k'}(0) = 0 \quad (24)$$

$$z_{k'}(n) = z_{k'-1}(n-1) \text{ para } n = 1, \dots, N-1.$$

25 Esto se repite para  $k'$  dentro del intervalo  $[L_{1/2} + 1, N - 1]$ .

En esta segunda etapa, el criterio (18) se puede calcular de una manera similar a la descrita en la sección anterior. Búsqueda en el libro de códigos fijo, para reducir aún más la complejidad de cálculo. El criterio (18) se evalúa primero para la última posición  $k' = N - 1$  (esta es la última fila de la matriz  $Z^T$ ). Para  $k' = N - 1$ , el numerador y el denominador del criterio (18) son proporcionados por la siguiente ecuación

$$d_g(N-1) = \sum_{i=0}^{L_{1/2}} x(N-1 - L_{1/2} + i) z_{L_{1/2}}(i), \quad (25)$$

5

y

$$\Phi_g(N-1, N-1) = \sum_{i=0}^{L_{1/2}} z_{L_{1/2}}(i) z_{L_{1/2}}(i). \quad (26)$$

Dado que algunos de los coeficientes de la matriz  $Z^T$  son ceros (figura 10), solo se utilizan  $L_{1/2} + 1$  multiplicaciones (en lugar de las  $N$  multiplicaciones utilizadas en la ecuación (23)) para calcular el numerador y el denominador del criterio (18)

10 Cuando se utiliza el ejemplo de la figura 10 ( $L_{1/2} = 1$ ), el criterio (18), calculado utilizando las ecuaciones (25) y (26), se puede simplificar de la siguiente manera:

$$\mathfrak{I}_{N-1} = \frac{(d_g(N-1))^2}{\Phi_g(N-1, N-1)} = \frac{(x(N-2)z_{L_{1/2}}(0) + x(N-1)z_{L_{1/2}}(1))^2}{z_{L_{1/2}}(0)z_{L_{1/2}}(0) + z_{L_{1/2}}(1)z_{L_{1/2}}(1)}. \quad (27)$$

15 En las etapas siguientes algunos de los valores calculados previamente se pueden reutilizar para el cálculo del denominador. Para la posición  $N - 2$  se calcula el denominador del criterio (18) utilizando

$$\Phi_g(N-2, N-2) = \Phi_g(N-1, N-1) + z_{L_{1/2}}(L_{1/2} + 1)z_{L_{1/2}}(L_{1/2} + 1). \quad (28)$$

El numerador se calcula utilizando la ecuación (25) con el índice de sumación modificado:

$$d_g(N-2) = \sum_{i=0}^{L_{1/2}+1} x(N-2 - L_{1/2} + i) z_{L_{1/2}}(i). \quad (29)$$

20 De una manera similar, se calculan el numerador y el denominador del criterio (18) para todas las posiciones  $k' > L_{1/2}$ .

25 El procedimiento descrito anteriormente permite hallar el máximo del criterio (18) para los vectores de código que representan la primera forma a partir de los impulsos glotales. La búsqueda continuará utilizando el procedimiento descrito anteriormente para todas las demás formas de impulso glotal. La búsqueda del máximo del criterio (18) continúa como búsqueda en el libro de códigos de forma glotal para encontrar un valor máximo para el criterio (18) que corresponde a la forma glotal y a una posición  $k'$ , que constituyen el resultado de la búsqueda.

Asimismo, es posible utilizar la resolución de submuestra cuando se busca la posición central del impulso glotal  $k'$ ; sin embargo, esto dará lugar a una mayor complejidad. Más específicamente, esto requerirá un muestreo de las formas de impulso glotales para aumentar la resolución y extraer diferentes versiones desplazadas con diferentes resoluciones. Esto es equivalente a utilizar un libro de códigos de forma glotal mayor.

30 Idealmente, el criterio (18) se calcula para todas las posibles posiciones glotales del impulso  $k'$ . En el modo de realización ilustrativo no restrictivo, la búsqueda se realiza solamente en un intervalo restringido alrededor de la posición esperada de la posición  $k'$ , para reducir adicionalmente la complejidad de cálculo. Esta posición esperada está en el rango  $[k_{\min}, k_{\max}]$ ,  $0 \leq k_{\min} < k_{\max} < N$ , y se puede determinar para la primera forma glotal a partir del máximo de la señal residual LP encontrada tal como se ha descrito en la sección anterior Selección de subtrama para búsqueda en el libro de códigos de forma glotal. A continuación, se realiza una búsqueda en el libro de códigos de forma glotal y se encuentra la posición  $k'$  para la primera forma glotal. El nuevo rango  $[k_{\min}, k_{\max}]$  se establece para la segunda búsqueda de forma glotal de la siguiente manera:

$$k_{min} = k' - \Delta, \quad (30)$$

$$k_{max} = k' + \Delta.$$

Típicamente  $\Delta = 4$ . De manera similar, se utiliza la ecuación (30) para definir el rango de búsqueda para la tercera forma alrededor de la posición seleccionada de la segunda forma, y así sucesivamente.

5 En el siguiente ejemplo, se supone que el rango de búsqueda inicial es  $[N - 15, N - 7]$ ,  $L = 17$  y  $N = 64$ . La búsqueda comienza con el cálculo del valor  $z_{L_{1/2}}(n)$ . A continuación, se evalúa el criterio (18) para la posición  $k' = N - 7$  utilizando

$$\mathfrak{I}_{N-7} = \frac{(d_g(N-7))^2}{\Phi_g(N-7, N-7)} = \frac{\left( \sum_{i=0}^{L_{1/2}+6} x(N-7 - L_{1/2} + i) z_{L_{1/2}}(i) \right)^2}{\sum_{i=0}^{L_{1/2}+7} z_{L_{1/2}}(i) z_{L_{1/2}}(i)} \quad (31)$$

Para calcular el criterio para la posición  $k' = N - 8$ , el denominador se calcula recursivamente como:

$$\Phi_g(N-8, N-8) = \Phi_g(N-7, N-7) + z_{L_{1/2}}(L_{1/2} + 8) z_{L_{1/2}}(L_{1/2} + 8). \quad (32)$$

10 De la misma manera, el denominador se calcula para todas las posiciones restantes hasta que  $k' = N - 15$ . El numerador del criterio (18) se calcula para cada posición dentro del rango  $[N - 15, N - 7]$  separadamente, de una manera similar a la ecuación (29), utilizando:

$$d_g(k') = \sum_{i=0}^{N-k'+L_{1/2}-1} x(k' - L_{1/2} + i) z_{L_{1/2}}(i). \quad (33)$$

15 El último parámetro a determinar en la búsqueda en el libro de códigos de forma glotal es la ganancia  $g_p$ , que se puede calcular como en la ecuación (4) con la diferencia de que no está limitada como en la búsqueda en el libro de códigos adaptativo. La razón es que el vector de código de la forma glotal filtrado se construye utilizando formas glotales cuantificadas normalizadas con una energía muy diferente de la energía de los impulsos de la señal de excitación reales.

20 Los índices relacionados con la posición del impulso glotal y la forma glotal se transmiten al decodificador. La reconstrucción del vector de código de la forma glotal filtrada en el decodificador se muestra en la figura 11. Se debe observar que la longitud del periodo del tono ya no necesita ser transmitida en una subtrama de búsqueda en el libro de códigos de forma glotal con la excepción de cuando la subtrama contiene más de un impulso glotal, tal como se analizará a continuación en la presente memoria.

#### Más de un impulso glotal en una subtrama

25 Existen situaciones en las que el periodo del tono de la señal de voz es más corto que la longitud de la subtrama y, en este caso, la subtrama puede contener más de un impulso glotal (especialmente en la configuración TRANSITION\_1\_1). En este caso es necesario modelizar todos los impulsos glotales. Dadas las limitaciones de longitud del periodo del tono y la longitud de la subtrama, una subtrama no puede contener más de dos impulsos glotales en este modo de realización ilustrativo no restrictivo.

30 Estas situaciones se pueden resolver mediante dos planteamientos diferentes. El primero y más sencillo resuelve estas situaciones mediante un procedimiento similar al de la mejora de la periodicidad (afinación de tono) utilizada en AMR-WB (Banda ancha de multi-tasa adaptativa - Adaptive Multi-Rate Wideband, en inglés) tal como se describe en la Referencia [1], donde el impulso se repite básicamente con el periodo del tono utilizando un filtro lineal. Tal como se ilustra en la figura 12a, el vector de código de la forma glotal  $q_k(n)$  se procesa de este modo a través de un filtro adaptativo de repetición de la forma:

$$Q(z) = \frac{1}{1 - \alpha \cdot z^{-T_0}}. \quad (34)$$

El periodo del tono  $T_0$  se puede determinar, por ejemplo, mediante el procedimiento de búsqueda del tono de bucle cerrado. El parámetro  $\alpha$  impacta sobre la energía del segundo impulso y, en el modo de realización ilustrativo no restrictivo, se ha establecido para  $\alpha = 0,85$ . Esta técnica añade el impulso glotal faltante en la posición correcta en el vector de código de la forma glotal. Esto se ilustra como el impulso punteado en la figura 12b. Esta situación aparece

5 cuando la suma de la posición central del impulso glotal  $k'$  y del periodo de paso  $T_0$  es menor que la longitud  $N$  de la subtrama, es decir  $(k' + T_0) < N$ . Pero también en situaciones en las que la suma de la posición del impulso  $k'$  y el periodo del tono superan la longitud de la subtrama, el valor del periodo del tono también se utiliza para construir el vector de código fijo cuando se utiliza el ajuste del tono en el libro de códigos algebraico.

10 El filtro de repetición  $Q(z)$  se inserta en la parte de TM del códec entre los filtros  $G(z)$  y  $H(z)$ , tal como se muestra en el diagrama de bloques de la figura 13 para el codificador. El mismo cambio se hace en el decodificador. De manera similar, para ajustar el tono, la respuesta de impulso del filtro de repetición  $Q(z)$  se puede añadir a la respuesta de impulso de  $G(z)$  y  $H(z)$  antes de la búsqueda en el libro de códigos, para que ambos impulsos se tengan en cuenta durante la búsqueda manteniendo al mismo tiempo la complejidad de la búsqueda en un nivel bajo.

15 Otro planteamiento para construir el vector de código de la forma glotal con dos impulsos glotales en una subtrama es utilizar una búsqueda en el libro de códigos adaptativo en una parte de la subtrama. Las primeras  $T_0$  muestras del vector de código de la forma glotal  $q_k(n)$  se construyen utilizando la búsqueda en el libro de códigos de forma glotal y luego se construyen las otras muestras en el subtrama utilizando la búsqueda adaptativa, tal como se muestra en la figura 14. Este planteamiento es más complejo, pero más preciso.

20 Para aumentar aún más la eficiencia de la codificación, se puede utilizar el procedimiento descrito anteriormente incluso si el segundo impulso glotal aparece en una de las primeras posiciones  $L_{1/2}$  de la siguiente subtrama (figura 15). En esta situación, es decir, cuando  $k'$  y  $T_0$  cumplen  $N \leq (k' + T_0) < (N + L_{1/2})$ , solo se utilizan unas pocas muestras (menos de  $L_{1/2} + 1$ ) de la forma glotal al final de la subtrama actual. Este planteamiento se utiliza en el modo de realización ilustrativo no restrictivo. Este planteamiento tiene una limitación porque el valor del periodo del tono transmitido en estas situaciones está limitado a  $T_0 < N$  (esta es una cuestión de codificación efectiva), aunque idealmente su valor debería limitarse a  $T_0 \leq N + L_{1/2}$ . Por lo tanto, si el segundo impulso glotal aparece al comienzo de la subtrama siguiente, el procedimiento de repetición no se puede utilizar para algunas de las primeras posiciones  $k'$  de impulso glotal  $L_{1/2}$  del primer impulso glotal.

#### Implementación de la técnica de codificación TM en un códec EV-VBR

30 La técnica de codificación TM de acuerdo con el modo de realización no limitativo ilustrativo se ha implementado en el códec EV-VBR. EV-VBR utiliza la frecuencia de muestreo interna de 12,8 kHz y la longitud de la trama de 20 ms. Cada trama se divide en cuatro subtramas de  $N = 64$  muestras. El procedimiento de clasificación EV-VBR ha sido adaptado para seleccionar las tramas a codificar utilizando la técnica de codificación TM. En esta implementación, la ganancia de la contribución del libro de códigos de forma glotal se cuantifica en dos etapas, tal como se representa en la figura 16, en la que  $G(z)$  es el filtro de conformación,  $k'$  es la posición del centro de la forma glotal y  $g_m$  es una ganancia de TM, es decir, una energía aproximadamente cuantificada del vector de código de la forma glotal. La ganancia  $g_m$  de TM se encuentra de la misma manera que la ganancia del tono utilizando la ecuación (4) con la única diferencia de que no está limitada. A continuación, se cuantifica mediante un cuantificador escalar de 3 bits y se utiliza un bit para cada señal. El vector de código de la forma glotal se escala después utilizando esta ganancia  $g_m$ . Tras hallarse ambas contribuciones a la señal de excitación filtrada (señales de contribución de la primera y de la segunda etapa, es decir, la contribución del libro de códigos de forma glotal filtrada y la contribución del libro de códigos algebraico filtrada), la ganancia de la señal de excitación de la primera etapa se ajusta adicionalmente junto con cuantificación de la ganancia de la señal de excitación de la segunda etapa, utilizando la cuantización del vector (VQ – Vector Quantization, en inglés) de ganancia EV-VBR estándar. De esta manera, los libros de códigos de cuantificación de la ganancia de EV-VBR diseñados para los modos de codificación genéricos o sonoros se podrían utilizar asimismo en la codificación TM. Por supuesto, realizar la cuantificación de ganancia utilizando otros métodos diferentes está dentro del alcance de la presente invención.

50 La búsqueda de la posición central del impulso glotal  $k'$  debería hacerse teóricamente para todas las posiciones en una subtrama, es decir, dentro del intervalo  $[0, N-1]$ . Sin embargo, tal como ya se ha mencionado, esta búsqueda requiere mucho cálculo dado el número de formas glotales a ensayar y, en la práctica, solo puede realizarse en el intervalo de varias muestras alrededor de la posición del valor máximo absoluto en la señal residual LP. El intervalo de búsqueda puede ajustarse a  $\pm 4$  muestras alrededor de la posición del primer máximo de impulso glotal en la señal residual LP en la trama actual. De esta manera, la complejidad de procesamiento es aproximadamente la misma que para la codificación genérica de EV-VBR, utilizando la búsqueda en el libro de códigos adaptativo y fijo.

55 Los parámetros transmitidos relacionados con la técnica de codificación TM se enumeran en la Tabla 4 con el número correspondiente de bits. El parámetro  $T_0$ , que se utiliza para determinar el filtro  $Q(z)$  o realizar una búsqueda adaptativa para el segundo impulso glotal en caso de dos impulsos en una subtrama, se transmite cuando  $T_0 \leq N$ . Los parámetros restantes utilizados para una trama TM, pero comunes con el procesamiento ACELP genérico, no se muestran aquí (bits de identificación de trama, parámetros LP, retardo del tono para excitación adaptativa, excitación de libro de códigos fijo, ganancias de libro de códigos de 1<sup>a</sup> y 2<sup>a</sup> etapas). Cuando se añaden

parámetros TM a la secuencia de bits, el número de bits asignados originalmente a otros parámetros EV-VBR se reduce para mantener una tasa de bits constante. Estos bits se pueden reducir, por ejemplo, a partir de los bits de excitación del libro de códigos fijo, así como de la cuantificación de la ganancia.

Tabla 4 - Parámetros en la secuencia de bits transmitida para la subtrama codificada utilizando la TM

Etiqueta	Significado	Número de bits
ID	identificación de la configuración	1 a 4
forma	forma del impulso glotal	3
$k'$	posición del centro del impulso glotal	6
$g_m$	ganancia de TM	3
signo ( $g_m$ )	signo de la ganancia de TM	1
$T_0$	periodo del tono de bucle cerrado (si es aplicable)	5

5 Las tablas de asignación de bits utilizadas en EV-VBR se muestran a continuación en la presente memoria. Se debe recordar que cuando la búsqueda en el libro de códigos de forma glotal no se aplica a la primera subtrama, solo el libro de códigos fijo y su ganancia se transmiten para codificar la señal de excitación en subtramas que preceden a la subtrama del libro de códigos de forma glotal. Lo mismo ocurre con las configuraciones TRANSICIÓN\_1\_3 y 10 TRANSICIÓN\_1\_4. En esos casos es posible mantener el mismo tamaño o incluso el tamaño grande del libro de códigos fijo para todas las subtramas que en la codificación ACELP genérica original.

#### Rendimiento de la técnica TM en el códec EV-VBR

15 En esta sección se presentan algunos ejemplos del rendimiento de la técnica de codificación TM en el códec EV-VBR. En la figura 17 se muestra un ejemplo del impacto de la técnica de codificación TM para la situación de canal limpio. La figura 17a muestra la señal de voz de entrada, la figura 17b muestra la señal residual LP y la figura 17c muestra la señal de excitación de la primera etapa en la que se utiliza la técnica de codificación TM en las tres (3) primeras tramas. Como era de esperar, la diferencia entre la señal residual y la señal de excitación de la primera etapa es más importante en el comienzo de cada trama. Hacia el final de la trama, la señal de excitación de la primera etapa se corresponde más estrechamente con la señal residual porque se utiliza la búsqueda en el libro de 20 códigos adaptativo estándar.

Las Tablas 5 y 6 resumen algunos ejemplos del rendimiento de la técnica de codificación TM medida utilizando valores de SNR.

25 En el primer ejemplo (Tabla 5) se implementó una técnica TM en códec con una frecuencia de muestreo central (interna)  $F_s = 8$  kHz (es decir, una longitud de subtrama  $N = 40$  muestras), se utilizó un libro de códigos de forma glotal con diecisiete muestras (17) de dieciséis (16) formas de longitud, y se ensayaron señales de entrada de banda estrecha. En la Tabla 5 se puede observar que codificar las tramas de inicio sonoras utilizando la técnica de codificación TM mejora la calidad de la señal de voz de salida (ver los valores de segmento y de segmento ponderado SNR para 1 y 2 tramas TM). Se puede observar un incremento adicional de SNR si la trama de inicio sonora y una trama siguiente se codifican utilizando la técnica de codificación TM. Sin embargo, si más de una trama 30 que sigue a la trama de inicio sonora también se codifica utilizando la técnica de codificación TM, los valores de SNR disminuyen. La SNR ponderada es la SNR ponderada mediante la energía de la trama normalizada por la longitud de la trama, en dB.

Tabla 5 - Comparación de las mediciones SNR del impacto de la técnica de codificación TM en las señales NB

Número de tramas TM	De segmento ponderado	De segmento SNR [dB]	SNR [dB]
0 (no codificación TM)	10,85	10,20	12,05
1 (TM en la trama de inicio)	10,88	10,48	11,03
2 (TM en la trama de inicio + 1 trama)	10,90	10,49	11,04
3 (TM en la trama de inicio + 2 tramas)	10,80	10,41	10,92

5 La tabla 6 resume un ejemplo del rendimiento del códec EV-VBR con frecuencia de muestreo central (interior)  $F_s = 12,8$  kHz, señal de voz de entrada de WB y libro de códigos de forma glotal con diecisiete (17) muestras de ocho (8) formas de longitud. Principalmente debido a la mayor longitud de la subtrama N, los valores de SNR muestran cierta degradación para el canal limpio cuando se utiliza la técnica de codificación TM, incluso si se utiliza en una sola trama. Esto se debe principalmente a la limitada longitud de los impulsos de forma glotal. En comparación con el ejemplo de NB, se presentan más valores cero en la señal de excitación de la primera etapa en la subtrama. La ventaja de utilizar la técnica de codificación TM en este ejemplo está en la protección FE (Borrado de trama - Frame Erasure, en inglés).

10 Tabla 6 - Comparación de las mediciones SNR del impacto de la técnica de codificación TM en las señales WB

Número de tramas TM	De segmento SNR ponderado [dB]	De segmento SNR [dB]	SNR [dB]
0 (no codificación TM)	7,52	7,21	8,61
1 (TM en la trama de inicio)	7,51	7,21	8,59
1 (TM en la trama tras la trama de inicio)	7,49	7,19	8,55
2 (TM en la trama de inicio + 1 trama)	7,48	7,17	8,55
2 (TM en 2 tramas tras la trama de inicio)	7,38	7,10	8,35
3 (TM en la trama de inicio + 2 tramas)	7,36	7,08	8,31

15 Se debe observar asimismo que incluso cuando se utiliza la técnica de codificación TM en una trama después de la trama borrada, todavía existe alguna pequeña diferencia entre la voz sintetizada en canal limpio y canal ruidoso. Esto se debe a que el codificador y los estados internos del decodificador no dependen solamente de la señal de excitación anterior, sino también de muchos otros parámetros (por ejemplo, memorias de filtros, memorias de cuantificador ISF (Frecuencias espectrales de inmitancia - Immitance Spectral Frequencies, en inglés), ...). Por supuesto, es posible probar la variante cuando se utiliza una codificación TM optimizada de cuantificación de parámetros LP sin memoria y todos los estados internos se restablecen para tramas TM. De este modo, todas las memorias que el códec EV-VBR utiliza en el modo de codificación genérico estándar se restablecieron para asegurar que los estados internos del decodificador después de un borrado de trama son los mismos que sus estados en condiciones sin errores. Sin embargo, la calidad de la voz en condiciones sin errores disminuye significativamente para esta variante. En consecuencia, existe un compromiso entre el alto rendimiento en condiciones sin errores y la robustez frente a tramas o paquetes borrados cuando no se realizan restablecimientos de memoria adicionales.

20 25 La Tabla 7 resume el problema de la complejidad de cálculo de la técnica codificación TM. En el peor de los casos, la técnica de codificación TM incrementa la complejidad en el codificador en 1,8 WMOPS (Millones ponderados de operaciones por segundo – Weighted Millions of Operations, en inglés). La complejidad del decodificador sigue siendo aproximadamente la misma.

30 Tabla 7 - Complejidad de la técnica de codificación TM (peor caso y valores medios)

Configuración	Codificador WMOPS		Decodificador WMOPS	
	Max	Media	Max	Media
Original (no codificación TM)	36,531	34,699	7,053	5,278
Se utilizó la técnica de codificación TM	38,346	34,743	7,055	5,281

35 Las figuras siguientes ilustran el rendimiento de la técnica de codificación TM para el modelizado de trama de inicio sonora (figuras 18a a 18c) y para la mitigación de la propagación de errores de trama (figuras 19a a 19c). La técnica de codificación TM se utiliza solo en una trama cada vez en este ejemplo. Se muestran un segmento de la señal de voz de entrada (figuras 18a y 19a), la señal de voz sintetizada de salida correspondiente procesada por el decodificador EV-VBR sin la técnica de codificación TM tal como se ilustra en las figuras 18b y 19b, y la señal de voz sintetizada de salida procesada utilizando el estándar EV-VBR con la técnica de codificación TM (figuras 18c y 19c). Los beneficios de la técnica de codificación TM se pueden observar tanto en el modelizado de la trama de inicio

sonora (2<sup>a</sup> trama de la figura 18) como en la limitación de la propagación del error de trama (tramas 4<sup>a</sup> y 5<sup>a</sup> de la figura 19).

La técnica de ocultación de borrado de tramas utilizada en el decodificador EV-VBR se basa en la utilización de un retardo de decodificador adicional de 20 ms de longitud (correspondiente a una longitud de trama). Significa que, si falta una trama, se oculta con el conocimiento de los parámetros de la trama futura. Supongamos tres (3) tramas consecutivas que se denotan como  $m - 1$ ,  $m$  y  $m + 1$  y suponen además una situación en la que falta la trama  $m$ . A continuación, se puede calcular una interpolación de la última trama recibida correctamente  $m - 1$  y la siguiente trama recibida correctamente  $m + 1$  con vistas a determinar los parámetros del códec, incluyendo, en particular, pero no exclusivamente, los coeficientes de filtro LP (representados por ISF - Frecuencias Espectrales de Immitancia), periodo del tono de bucle cerrado  $T_0$  y ganancias de libro de códigos fijo. La interpolación ayuda a estimar con mayor precisión los parámetros de trama perdidos para segmentos de voz estables. Sin embargo, a menudo falla en los segmentos de transición cuando los parámetros del códec varían rápidamente. Para hacer frente a este problema, el valor absoluto del periodo del tono puede ser transmitido en cada trama TM incluso en el caso de que no se utilice para la construcción de la excitación de la primera etapa en la trama  $m + 1$  actual. Esto es válido especialmente para las configuraciones TRANSICIÓN\_1\_4 y TRANSICIÓN\_4.

Otros parámetros transmitidos en una trama TM son los ISF de la trama anterior. En los codificadores de tipo CELP, los parámetros ISF generalmente se interpolan entre ISF de las tramas anteriores y los ISF de las tramas actuales para cada subtrama. Esto asegura una evolución continua del filtro de síntesis LP de una subtrama a otra. En el caso de un borrado de trama, los ISF de la trama que precede al borrado de trama se usan habitualmente para la interpolación en la trama que sigue al borrado, en lugar de los ISF de las tramas borradas. Sin embargo, durante los segmentos de transición, los ISF varían rápidamente y los últimos ISF de las tramas buenas pueden ser muy diferentes de los ISF de la trama desaparecida borrada. Reemplazar los ISF de la trama faltante por los ISF de la trama anterior puede provocar por ello aberraciones importantes. Si se pueden transmitir los últimos ISF de la trama anterior, se pueden utilizar para la interpolación de ISF en la trama TM en caso de que se borre la trama anterior. Posteriormente, se describirán diferentes estimaciones de los coeficientes de LP utilizados para la interpolación de ISF cuando la trama que precede a una trama TM falte.

La implementación final de la técnica de codificación TM para el códec EV-VBR supone que solo una trama después del inicio / trama de transición se codifica utilizando TM. De esta manera, se seleccionan aproximadamente el 6,3% de las tramas de voz activas para codificación y decodificación TM.

Otra categoría de pruebas se centró en el aumento de la eficiencia de la codificación. La clasificación se realizó en la búsqueda en bucle cerrado cuando se calcularon dos variantes -con y sin la técnica de codificación TM- en el codificador y se eligió una variante con una SNR más alta como señal de salida

Los resultados para el códec EV-VBR con tasa de bits de 8 kbps se resumen en la Tabla 8. En el caso de WB, el 28% de las tramas de voz activas se clasificó para codificación utilizando la técnica de codificación TM y se logró un incremento de 0,203 dB en la SNR de segmento. En el caso de NB, el 25% de las tramas de voz activas se clasificó para codificación utilizando la técnica de codificación TM y se logró un aumento de incluso 0,300 dB en la SNR de segmento. Desafortunadamente, este aumento objetivo de la prueba no fue confirmado por pruebas de escucha subjetiva que no informaron de ninguna preferencia entre el códec con y sin la técnica de codificación TM. Aunque no hay degradación de la calidad de la voz y el número total de tramas TM es cuatro (4) veces más alto en comparación con una clasificación en bucle abierto que resulta en una protección FE mucho mayor, esta clasificación y clasificaciones de resultados similares no se utilizan en una implementación de códec EV-VBR debido a la mayor complejidad.

Tabla 8 - Comparación de la medida de SNR y SNR de segmento entre el códec con y sin la técnica de codificación TM implementada cuando se utiliza la clasificación de bucle cerrado.

Número de tramas codificadas por TM	SNR de segmento [dB]	SNR [dB]
Códec sin TM, señal de WB	7,34	8,89
Códec con TM, señal de WB	7,54	9,04
Códec sin TM, señal de NB	7,58	10,62
Códec con TM, señal de NB	7,88	10,97

#### 45 Tablas de asignación de bits para la técnica de codificación TM en un códec EV-VBR

La técnica de codificación TM se implementó en un códec candidato de EV-VBR para la estandarización ITU-T. La Tabla 9 siguiente muestra las tablas de asignación de bits del modo genérico original y todas las configuraciones del modo de codificación TM que se han introducido anteriormente en este documento. Estas configuraciones se utilizan en el códec EV-VBR.

Tabla 9 – Tablas de asignación de bits para el modo de codificación genérico y para todas las configuraciones TM utilizadas en el códec EV-VBR (ID significa identificación de configuración, ISF significa frecuencias espectrales de Inmitancia y FCB significa libro de códigos fijo, subtr. es subrama)

5

a) GENÉRICO	B) TRANSICIÓN_1_1	C) TRANSICIÓN_1_2
Parámetro de # bits	Parámetro de # bits	Parámetro de # bits
2 tipo de codificador	2 tipo de codificador	2 tipo de codificador
1 NB / WB	1 NB / WB	1 NB / WB
36 ISF	36 ISF	36 ISF
3 estimación de energía	3 estimación de energía	3 estimación de energía
8 tono de la 1 <sup>a</sup> subtr.	1 ID de la subtr. TM	1 ID de la subtr. TM
5 ganancias de la 1 <sup>a</sup> subtr.	5 tono de la 1 <sup>a</sup> subtr.	1 ID de la subtr. TM
5 tono de la 2 <sup>a</sup> subtr.	3 forma de TM	3 forma de TM
5 ganancias de la 2 <sup>a</sup> subtr.	6 posición de TM	6 posición de TM
8 tono de la 3 <sup>a</sup> subtr.	1 signo de la ganancia de TM	1 signo de la ganancia de TM
5 ganancias de la 3 <sup>a</sup> subtr.	3 valor de la ganancia de TM	3 valor de la ganancia de TM
5 tono de la 4 <sup>a</sup> subtr.	5 ganancias de la 1 <sup>a</sup> subtr.	5 ganancias de la 1 <sup>a</sup> subtr.
5 ganancias de la 4 <sup>a</sup> subtr.	5 tono de la 2 <sup>a</sup> subtr.	1 ID2 de la subtr. TM
12 FCB de la 1 <sup>a</sup> subtr.	5 ganancias de la 2 <sup>a</sup> subtr.	1 ID2 de la subtr. TM
20 FCB de la 2 <sup>a</sup> subtr.	5 tono de la 3 <sup>a</sup> subtr.	7 tono de la 2 <sup>a</sup> subtr
20 FCB de la 3 <sup>a</sup> subtr.	5 ganancias de la 3 <sup>a</sup> subtr.	5 ganancias de la 2 <sup>a</sup> subtr.
20 FCB de la 4 <sup>a</sup> subtr.	5 tono de la 4 <sup>a</sup> subtr.	5 tono de la 3 <sup>a</sup> subtr.
	5 ganancias de la 4 <sup>a</sup> subtr.	5 ganancias de la 3 <sup>a</sup> subtr.
160 bits totales	20 FCB de la 1 <sup>a</sup> subtr.	5 tono de la 4 <sup>a</sup> subtr.
	20 FCB de la 2 <sup>a</sup> subtr.	5 ganancias de la 4 <sup>a</sup> subtr.
	12 FCB de la 3 <sup>a</sup> subtr.	20 FCB de la 1 <sup>a</sup> subtr.
	12 FCB de la 4 <sup>a</sup> subtr.	20 FCB de la 2 <sup>a</sup> subtr.
160 bits totales	160 bits totales	12 FCB de la 3 <sup>a</sup> subtr.
		12 FCB de la 4 <sup>a</sup> subtr.
		160 bits totales

d) TRANSICIÓN_1_3	e) TRANSICIÓN_1_4	f) TRANSICIÓN_2
Parámetro de # bits	Parámetro de # bits	Parámetro de # bits
2 tipo de codificador	2 tipo de codificador	2 tipo de codificador
1 NB / WB	1 NB / WB	1 NB / WB
36 ISF	36 ISF	36 ISF
3 estimación de energía	3 estimación de energía	3 estimación de energía
1 ID de la subtr. TM	1 ID de la subtr. TM	1 ID de la subtr. TM
1 ID de la subtr. TM	1 ID de la subtr. TM	1 ID de la subtr. TM
3 forma de TM	3 forma de TM	1 ID de la subtr. TM
6 posición de TM	6 posición de TM	2 ganancias de la 1 <sup>a</sup> subtr.
1 signo de la ganancia de TM	1 signo de la ganancia de TM	3 forma de TM
3 valor de la ganancia de TM	3 valor de la ganancia de TM	6 posición de TM
5 ganancias de la 1 <sup>a</sup> subtr.	5 ganancias de la 1 <sup>a</sup> subtr.	1 signo de la ganancia de TM
1 ID2 de la subtr. TM	1 ID2 de la subtr. TM	3 valor de la ganancia de TM
1 ID2 de la subtr. TM	3 ganancia de la 2 <sup>a</sup> subtr.	5 ganancias de la 2 <sup>a</sup> subtr.
3 ganancia de la 2 <sup>a</sup> subtr.	2 ganancia de la 3 <sup>a</sup> subtr.	8 tono de la 3 <sup>a</sup> subtr
7 tono de la 3 <sup>a</sup> subtr.	7 tono de la 4 <sup>a</sup> subtr.	5 ganancias de la 3 <sup>a</sup> subtr.
5 ganancias de la 3 <sup>a</sup> subtr.	5 ganancias de la 4 <sup>a</sup> subtr.	5 tono de la 4 <sup>a</sup> subtr.
4 tono de la 4 <sup>a</sup> subtr.	20 FCB de la 1 <sup>a</sup> subtr.	5 ganancias de la 4 <sup>a</sup> subtr
5 ganancias de la 4 <sup>a</sup> subtr.	20 FCB de la 2 <sup>a</sup> subtr.	20 FCB de la 1 <sup>a</sup> subtr.
20 FCB de la 1 <sup>a</sup> subtr.	20 FCB de la 3 <sup>a</sup> subtr.	20 FCB de la 2 <sup>a</sup> subtr.
12 FCB de la 2 <sup>a</sup> subtr.	20 FCB de la 4 <sup>a</sup> subtr.	12 FCB de la 3 <sup>a</sup> subtr.
20 FCB de la 3 <sup>a</sup> subtr.		20 FCB de la 4 <sup>a</sup> subtr.
20 FCB de la 4 <sup>a</sup> subtr.		
160 bits totales	160 bits totales	160 bits totales

g) TRANSICIÓN_3	h) TRANSICIÓN_4
Parámetro de # bits	Parámetro de # bits
2 tipo de codificador	2 tipo de codificador
1 NB / WB	1 NB / WB
36 ISF	36 ISF
3 estimación de energía	3 estimación de energía
1 ID de la subtr. TM	1 ID de la subtr. TM
1 ID de la subtr. TM	1 ID de la subtr. TM
1 ID de la subtr. TM	1 ID de la subtr. TM
1 ID de la subtr. TM	1 ID de la subtr. TM
3 ganancia de la 1 <sup>a</sup> subtr.	3 ganancia de la 1 <sup>a</sup> subtr.
3 ganancia de la 2 <sup>a</sup> subtr.	2 ganancia de la 2 <sup>a</sup> subtr.
5 tono de la 3 <sup>a</sup> subtr.	3 ganancia de la 3 <sup>a</sup> subtr.
3 forma de TM	8 tono de la 4 <sup>a</sup> subtr.
6 posición de TM	3 forma de TM
1 signo de la ganancia de TM	6 posición de TM
3 valor de la ganancia de TM	1 signo de la ganancia de TM
5 ganancias de la 3 <sup>a</sup> subtr.	3 valor de la ganancia de TM
8 tono de la 4 <sup>a</sup> subtr.	5 ganancias de la 4 <sup>a</sup> subtr.
5 ganancias de la 4 <sup>a</sup> subtr.	20 FCB de la 1 <sup>a</sup> subtr.
12 FCB de la 1 <sup>a</sup> subtr.	20 FCB de la 2 <sup>a</sup> subtr.
20 FCB de la 2 <sup>a</sup> subtr.	20 FCB de la 3 <sup>a</sup> subtr.
20 FCB de la 3 <sup>a</sup> subtr.	20 FCB de la 4 <sup>a</sup> subtr.
20 FCB de la 4 <sup>a</sup> subtr.	

160 bits totales

160 bits totales

- 5 Existe una excepción a la configuración TRANSICIÓN\_2 de la Tabla 9. Esta tabla de asignación de bits solo se puede utilizar en el caso en que se decida utilizar la técnica de codificación TM en las tramas que siguen únicamente a la trama de inicio sonora (la trama de inicio sonora se codifica utilizando el modo de codificación genérico y solo se codifica una trama siguiente a la trama de inicio sonora utilizando la técnica de codificación TM). En esta situación, el periodo del tono  $T_0$  es  $T_0 \geq N$  en la segunda subtrama y no es necesario transmitir este parámetro en la 2<sup>a</sup> subtrama. Pero si la técnica de codificación TM se utiliza también en la trama de inicio sonora, puede ocurrir la siguiente situación. El periodo del tono es menor que N, pero el inicio sonoro solo se puede iniciar en la 2<sup>a</sup> subtrama (por ejemplo, la primera subtrama que todavía contiene señal no sonora). En este caso, se debe transmitir el periodo del tono  $T_0$ . En esta situación se utiliza una tabla de asignación de bits diferente, el parámetro  $T_0$  se transmite en la 2<sup>a</sup> subtrama utilizando cinco (5) bits y en una subtrama se utiliza un libro de código fijo más corto (véase la Tabla 10). La misma situación aparece también para la configuración TRANSICIÓN\_3. Sin embargo, el periodo del tono se transmite aquí de todos modos en la implementación actual, no limitativa de (ya sea si la trama de inicio se codifica mediante la técnica de codificación TM o no), porque no existe un buen uso de los bits guardados para otro parámetro de codificación.

5 Se pueden utilizar otras asignaciones de bits en diferentes configuraciones de modo de transición. Por ejemplo, se pueden asignar más bits a los libros de códigos fijos en las subtramas que contienen impulsos glotales. Por ejemplo, en el modo TRANSICIÓN\_3, se puede utilizar un FCB con doce (12) bits en la segunda subtrama y veintiocho (28) bits en la tercera subtrama. Por supuesto, se pueden utilizar otros FCB distintos de los de 12 y 20 bits en diferentes implementaciones de codificadores.

Tabla 10 - Tabla de asignación de bits para la configuración TRANSICIÓN\_2 si se utiliza también TM en la trama de inicio

TRANSICIÓN_2a
Parámetro de # bits
2 tipo de codificador
1 NB / WB
36 ISF
3 estimación de energía
1 ID de la subtr. TM
1 ID de la subtr. TM
1 ID de la subtr. TM
3 ganancia de la 1 <sup>a</sup> subtr.
5 tono de la 2 <sup>a</sup> subtr.
3 forma de TM
6 posición de TM
1 signo de la ganancia de TM
3 valor de la ganancia de TM
5 ganancias de la 2 <sup>a</sup> subtr.
8 tono de la 3 <sup>a</sup> subtr.
5 ganancias de la 3 <sup>a</sup> subtr.
5 tono de la 4 <sup>a</sup> subtr.
5 ganancias de la 4 <sup>a</sup> subtr.
20 FCB de la 1 <sup>a</sup> subtr.
20 FCB de la 2 <sup>a</sup> subtr.
12 FCB de la 3 <sup>a</sup> subtr.
12 FCB de la 4 <sup>a</sup> subtr.
158 bits totales

10 Si existe ancho de banda disponible, se puede conseguir una mejora adicional transmitiendo más información para una mejor protección frente al borrado de trama (FE). El códec VMR-WB es un ejemplo de un códec que utiliza una parte de los bits de protección FE. Por ejemplo, se utilizan catorce (14) bits de protección por trama en el tipo de codificación de tasa total genérica en VMR-WB en Rate-Set II. Estos bits representan la clasificación de tramas (2 bits), la energía de voz sintetizada (6 bits) y la posición del impulso glotal (6 bits). El impulso glotal se inserta artificialmente en el decodificador cuando se pierde una trama de inicio sonora. Estos bits de protección FER no son muy importantes para la construcción de la excitación en una trama TM, porque la técnica de codificación TM no hace uso de la señal de excitación anterior; la técnica de codificación TM construye la señal de excitación utilizando parámetros transmitidos en la trama actual (TM). Sin embargo, estos bits se pueden emplear para la transmisión de otros parámetros. En un ejemplo de implementación, estos bits se pueden utilizar para transmitir en la trama TM actual los parámetros ISF de la trama anterior; sin embargo, están disponibles doce (12) bits en lugar de treinta y

seis (36) bits). Estos ISF se utilizan para la reconstrucción de coeficientes de filtro LP más precisos en caso de borrado de trama.

En el códec EV-VBR se calcula el conjunto de parámetros LP centrados en la cuarta subtrama, mientras que la primera, segunda y tercera subtramas utilizan una interpolación lineal de los parámetros de filtro LP entre la trama actual y la trama anterior. La interpolación se realiza en los ISP (Pares Espectrales de Inmitancia - Immittance Spectral Pairs, en inglés). Sea  $\mathbf{q}_4^{(m)}$  el vector ISP en la 4<sup>a</sup> subtrama de la trama y  $\mathbf{q}_4^{(m-1)}$  el vector ISP en la 4<sup>a</sup> subtrama de la trama anterior  $m - 1$ . Los vectores ISP interpolados en las subtramas 1<sup>a</sup>, 2<sup>a</sup> y 3<sup>a</sup> vienen dados por las ecuaciones:

$$\begin{aligned} \mathbf{q}_1^{(m)} &= 0.55\mathbf{q}_4^{(m-1)} + 0.45\mathbf{q}_4^{(m)}, \\ \mathbf{q}_2^{(m)} &= 0.2\mathbf{q}_4^{(m-1)} + 0.8\mathbf{q}_4^{(m)}, \\ \mathbf{q}_3^{(m)} &= 0.04\mathbf{q}_4^{(m-1)} + 0.96\mathbf{q}_4^{(m)}. \end{aligned} \quad (35)$$

Sin embargo, esta interpolación no es adecuada directamente para la técnica de codificación TM en el caso de borrado de la trama anterior. Cuando falta la trama que precede a la trama TM, se puede suponer que la última trama correctamente recibida no es sonora. En esta situación, es más eficiente reconstruir el vector ISF para la trama faltante con diferentes constantes de interpolación y no importa si tenemos o no alguna información ISF de los bits de protección FER disponibles. En general, la interpolación está utilizando más los ISP de la trama anterior. Los vectores ISP para la trama faltante  $m$  se puede dar en el decodificador, por ejemplo, utilizando las siguientes ecuaciones:

$$\begin{aligned} \mathbf{q}_1^{(m)} &= \mathbf{q}_4^{(m-1)}, \\ \mathbf{q}_2^{(m)} &= \mathbf{q}_4^{(m-1)}, \\ \mathbf{q}_3^{(m)} &= 0.7\mathbf{q}_4^{(m-1)} + 0.3\mathbf{q}_4^{(m)}, \\ \mathbf{q}_4^{(m)} &= 0.1\mathbf{q}_4^{(m-1)} + 0.9\mathbf{q}_4^{(m)}. \end{aligned} \quad (36)$$

Entonces, la siguiente trama TM  $m + 1$  correctamente recibida utiliza la interpolación de coeficientes LP descrita mediante las ecuaciones (35). Además, los coeficientes de interpolación en las ecuaciones (36) se dan como ejemplo no limitativo. Los coeficientes finales podrían ser diferentes y, además, es deseable utilizar un conjunto de coeficientes de interpolación cuando se dispone de alguna información ISF de la trama anterior y otra cuando la información ISF de la trama anterior no está disponible (es decir, no hay bits de protección frente al borrado de trama en la secuencia de bits).

## 25 Período del tono y codificación de la ganancia en tramas TM en el códec EV-VBR

El valor del período del tono  $T_0$  se transmite para cada subtrama en el modo de codificación genérico utilizado en el códec EV-VBR. En las tramas 1<sup>a</sup> y 3<sup>a</sup>, se utiliza una codificación de 8 bits mientras que el valor del período del tono se transfiere con resolución fraccionaria ( $\frac{1}{2}$  para  $T_0$  en el rango  $[T_{\min}, 91 \frac{1}{2}]$ ) o entero (para  $T_0$  en el rango  $[92, T_{\max}]$ ). En la 2<sup>a</sup> y 4<sup>a</sup> subtramas, se utiliza una búsqueda delta y el valor del período del tono siempre con resolución fraccionaria se codifica con cinco (5) bits. Búsqueda delta significa una búsqueda dentro del rango  $[T_{0p} - 8, T_{0p} + 7 \frac{1}{2}]$ , donde  $T_{0p}$  es el entero más cercano al período del tono fraccionario de la subtrama anterior (1<sup>a</sup> o 3<sup>a</sup>). Los valores del período del tono están limitados en el códec EV-VBR a valores dentro del intervalo  $[T_{\min}, T_{\max}]$ , donde  $T_{\min} = 34$  y  $T_{\max} = 231$ .

La ganancia del tono  $g_p$  y la ganancia de libro de códigos fijo  $g_c$ , se codifican en el códec EV-VBR en principio de la misma manera que en el códec AMR-WB + códec [5]. En primer lugar, se calcula una estimación de una energía del libro de códigos fijo, escalada, no predictiva, para todas las subtramas en una trama y se cuantifica con tres (3) bits una vez por trama (véase la estimación de energía del parámetro en la Tabla 9). A continuación, la ganancia del tono  $g_p$  y la ganancia  $g_c$  del libro de códigos fijo se cuantifican en vector y se codifican en una etapa utilizando cinco (5) bits para cada subtrama.

La energía del libro de códigos fijo estimada se calcula y cuantifica como sigue. En primer lugar, se calcula la energía residual LP en cada subtrama  $k$  utilizando la siguiente ecuación:

$$E_{res}(k) = 10 \log_{10} \left( \frac{1}{N} \sum_{n=0}^{N-1} u^2(n) \right), \quad (37)$$

donde  $u(n)$  es la señal residual LP. A continuación, se encuentra la energía residual media por subtrama mediante la siguiente ecuación:

$$\bar{E}_{res} = \frac{1}{4} \sum_{k=0}^3 E_{res}(k). \quad (38)$$

La energía del libro de códigos fijo se estima a partir de la energía residual mediante la eliminación de la contribución del libro de códigos adaptativo. Esto se realiza eliminando una energía relacionada con la correlación normalizada media obtenida a partir de los dos análisis del tono en bucle abierto realizados en la trama. Se utiliza la siguiente ecuación:

$$E_s = \bar{E}_{res} - 10 \bar{R}, \quad (39)$$

donde  $\bar{R}$  es la media de las correlaciones normalizadas del tono obtenidas a partir del análisis del tono en bucle abierto para cada semitrama de la trama actual. La energía del libro de códigos fijo escalada estimada no depende de la energía de trama previa y, por lo tanto, el principio de codificación de la ganancia es robusto frente a borrados de tramas.

Una vez obtenida la estimación de la energía del libro de códigos fijo, se calcula la ganancia del tono y la corrección de la ganancia del libro de códigos fijo: la energía del libro de códigos fijo escalada estimada se utiliza para calcular la ganancia del libro de códigos fijo estimada y el factor de corrección  $\gamma$  (relación entre las ganancias del libro de códigos fijo verdadera y estimada). El valor  $\gamma$  se cuantifica en vector junto con la ganancia del tono utilizando cinco (5) bits por subtrama. Para el diseño del cuantificador, se utiliza un método k-means modificado [4]. La ganancia del tono está restringida dentro del intervalo  $<0; 1,2>$  durante la inicialización del libro de códigos y  $<0; \infty>$  durante la mejora iterativa del libro de códigos. Asimismo, el factor de corrección  $\gamma$  está limitado por  $<0; 5>$  durante la inicialización y  $<0; \infty>$  durante la mejora del libro de códigos. El algoritmo k-means modificado busca minimizar el siguiente criterio:

$$E = g_p^2 \mathbf{y}_1^T \mathbf{y}_1 - 2g_p \mathbf{x}_1^T \mathbf{y}_1 + g_c^2 \mathbf{y}_2^T \mathbf{y}_2 - 2g_c \mathbf{x}_1^T \mathbf{y}_2 + 2g_p g_c \mathbf{y}_1^T \mathbf{y}_2. \quad (40)$$

Cuando se utiliza la técnica de codificación TM, la transmisión del periodo del tono y las ganancias tanto del tono como del libro de códigos fijo pueden no ser necesarias para subtramas en las que no hay impulso glotal importante, y solo se puede calcular la contribución del libro de códigos fijo.

La siguiente es una lista y descripción de todas las configuraciones TM:

Configuración TRANSICIÓN\_1\_1 (figura 20) - En esta configuración aparecen uno o dos primeros impulsos glotales en la primera subtrama que se procesa utilizando la búsqueda en el libro de códigos de forma glotal. Esto significa que el valor del periodo del tono en la primera subtrama puede tener un valor máximo menor que la longitud de la subtrama, es decir  $T_{min} < T_0 < N$ . Con la resolución entera se puede codificar con cinco (5) bits. Los períodos del tono en las subtramas siguientes se encuentran utilizando la búsqueda delta de 5 bits con una resolución fraccionaria.

Esta es la configuración más exigente en bits de la técnica de codificación TM, es decir, cuando el libro de códigos de forma glotal se utiliza en la primera subtrama y periodo del tono  $T_0$  se transmite para la determinación del filtro Q (z), o para la búsqueda en el libro de códigos adaptativo en la parte de la primera subtrama. Esta configuración utiliza en la primera subtrama el procedimiento descrito anteriormente. Esta configuración se utiliza en el códec EV-VBR también cuando solo aparece un impulso glotal en la primera subtrama. En este caso, el periodo del tono  $T_0$  cumple  $T_0 < N$  y se utiliza para mejorar la periodicidad [1] en la búsqueda en el libro de códigos fijo.

Configuración TRANSICIÓN\_1\_2 (figura 21) - Cuando se utiliza la configuración TRANSICIÓN\_1\_2, la primera subtrama se procesa utilizando la búsqueda en el libro de códigos de forma glotal. El periodo del tono no es necesario y todas las subtramas siguientes se procesan utilizando la búsqueda en el libro de códigos adaptativo. Debido a que se conoce que la segunda subtrama contiene el segundo impulso glotal, el valor máximo del periodo de paso cumple  $T_0 \leq 2 \cdot N - 1$ . Este valor máximo se puede reducir aún más gracias al conocimiento de la posición del impulso glotal  $k'$ . A continuación, se codifica el valor del periodo del tono en la segunda subtrama utilizando siete

(7) bits con una resolución fraccionaria en todo el intervalo. En la tercera y cuarta subtramas, se utiliza la búsqueda delta utilizando cinco (5) bits con una resolución fraccionaria.

- 5 Configuración TRANSICIÓN \_1\_3 (figura 22) - Cuando se utiliza la configuración TRANSICIÓN \_1\_3, se procesa la primera subtrama utilizando la búsqueda en el libro de códigos de forma glotal de nuevo sin utilizar el período del tono. Debido a que la segunda subtrama de la señal residual LP no contiene ningún impulso glotal y la búsqueda adaptativa es inútil, la señal de excitación de la primera etapa se reemplaza por ceros en la segunda subtrama. Los parámetros de libro de códigos adaptativos ( $T_0$  y  $g_p$ ) no se transmiten en la segunda subtrama, y los bits guardados se utilizan para el aumento de tamaño FCB en la tercera subtrama. Debido a que la segunda subtrama contiene un mínimo de la información útil, solo se utiliza el FCB de 12 bits y se utiliza el FCB de 20 bits en la cuarta subtrama. La señal de excitación de la primera etapa en la tercera subtrama se construye utilizando la búsqueda en el libro de códigos adaptativo con el valor máximo del período del tono ( $3 \cdot N - 1 - k'$ ) y el valor mínimo ( $2 \cdot N - k'$ ); por lo tanto, solo se utiliza una codificación de 7 bits del período del tono con resolución fraccionaria sobre todo el intervalo. La cuarta subtrama se procesa utilizando la búsqueda adaptativa de nuevo con una codificación de búsqueda de delta de 5 bits del valor de período de tono.
- 10 15 En la segunda subtrama solo se transmite la ganancia de libro de códigos fijo  $g_c$ . En consecuencia, solo se necesitan dos (2) o tres (3) bits para la cuantificación de la ganancia en lugar del cuantificador de 5 bits utilizado en la subtrama con codificación ACELP tradicional (es decir, cuando se transmiten las ganancias  $g_p$  y  $g_c$ ). Esto es válido también para todas las configuraciones siguientes. La decisión de si el cuantificador de la ganancia debe utilizar dos (2) o tres (3) bits se toma para ajustar el número de bits disponibles en la trama.
- 20 25 30 35 40 45 50 55 Configuración TRANSICIÓN \_1\_4 (figura 23) - Cuando se utiliza la configuración TRANSICIÓN \_1\_4, la primera subtrama se procesa utilizando la búsqueda en el libro de códigos de forma glotal. De nuevo, el período del tono no necesita ser transmitido. Pero debido a que la señal residual LP no contiene ningún impulso glotal en la segunda y tampoco en la tercera subtrama, la búsqueda en el libro de códigos adaptativo es inútil para estas dos subtramas. De nuevo, la señal de excitación de la primera etapa en estas subtramas se reemplaza por ceros, y los bits guardados se usan para el aumento del tamaño FCB de modo que todas las subtramas pueden aprovechar y utilizar los FCB de 20 bits. El valor del período del tono solo se transmite en la cuarta subtrama y su valor mínimo es ( $3 \cdot N - k'$ ). El valor máximo del período del tono está limitado por  $T_{max}$ . No importa si el segundo impulso glotal aparece o no en la cuarta subtrama (el segundo impulso glotal puede estar presente en la siguiente trama si  $k' + T_{max} \geq N$ ).
- El valor absoluto del período del tono se utiliza en el decodificador para la ocultación de la trama; por lo tanto, este valor absoluto del período del tono se transmite en la situación en que el segundo impulso glotal aparece en la siguiente trama. Cuando falta una trama  $m$  que precede a la trama TM  $m + 1$ , la correcta comprensión del valor del período del tono de las tramas  $m - 1$  y  $m + 1$  ayuda a reconstruir con éxito la parte que falta de la señal de síntesis en la trama  $m$ .
- Configuración TRANSICIÓN\_2 (figura 24) - Cuando el primer impulso glotal aparece en la segunda subtrama y solo las tramas siguientes a las tramas de inicio sonoras se codifican utilizando la técnica de codificación TM (es decir, las tramas de inicio sonoras se codifican con la codificación genérica heredada), el período del tono se transmite solo en la tercera y cuarta subtramas. En este caso, solo se transmiten parámetros de libro de códigos fijo en la primera subtrama.
- La trama mostrada en la figura 24 supone la configuración cuando no se utiliza TM en tramas de inicio sonoras. Si TM se utiliza también en las tramas de inicio sonoras, se utiliza la configuración TRANSICIÓN\_2a, en la que el período del tono  $T_0$  se transmite en la segunda subtrama para utilizar el procedimiento tal como se ha descrito anteriormente.
- Configuración TRANSICIÓN\_3 (figura 25) - Cuando el primer impulso glotal aparece en la tercera subtrama y solo las tramas siguientes a las tramas de inicio sonoras se codifican utilizando la técnica de codificación TM (es decir, las tramas de inicio sonoras están codificadas con la codificación genérica heredada), el período del tono es transmitida solo en la cuarta subtrama. En este caso, solo se transmiten parámetros de libro de códigos fijo en la primera y segunda subtramas.
- El período del tono todavía se transmite para la tercera subtrama en la secuencia de bits. Sin embargo, no es útil si la técnica de codificación TM no se utiliza para codificar las tramas de inicio sonoras. Este valor es útil solo cuando se codifican tramas de inicio sonoras utilizando la técnica de codificación TM.
- Configuración TRANSICIÓN\_4 (figura 26) - Cuando el primer impulso glotal aparece en la cuarta subtrama y solo las tramas después de que las tramas de inicio sonoras se codifican utilizando la técnica de codificación TM (es decir, las tramas de inicio sonoras están codificadas con la codificación genérica heredada), la información del valor del período del tono no se utiliza en este subtrama. Sin embargo, el valor del período del tono se utiliza en la ocultación de trama en el decodificador (este valor se utiliza para la reconstrucción de una trama faltante cuando falta la trama que precede a la trama TM). Por lo tanto, el valor del tono se transmite solamente en la cuarta subtrama y solo se transmiten parámetros del libro de códigos fijo en la primera, segunda y tercera subtramas (el tono de ganancia  $g_p$  no es necesario). Los bits guardados permiten que se utilice el FCB de 20 bits en cada subtrama.

Aunque la presente invención se ha descrito en la descripción anterior en conexión con un modo de realización ilustrativo, no restrictivo de la misma, este modo de realización ilustrativo, no restrictivo se puede modificar a voluntad, dentro del alcance de las reivindicaciones adjuntas, sin apartarse del alcance de la presente invención.

## Referencias

- [1] B. BESSETTE, R. SALAMI, R. LEFEBVRE, M. JELINEK, J. ROTOLA-PUKKILA, J. VAINIO, H. MIKKOLA, and K. JARVINEN, "The Adaptive Multi-Rate Wideband Speech Codec (AMR-WB)", Special Issue of IEEE Transactions on Speech and Audio Processing, Vol. 10, No. 8, pp. 620-636, November 2002.
- [2] R. SALAMI, C. LAFLAMME, J-P. ADOUL, and D. MASSALOUX, "A toll quality 8 kb/s speech codec for the personal communications system (PCS)", IEEE Trans. on Vehicular Technology, Vol. 43, No. 3, pp. 808-816, August 1994.
- [3] 3GPP2 Tech. Spec. C.S0052-A v1.0, "Source-Controlled Variable-Rate Multimode Wideband Speech Codec (VMR-WB), Service Options 62 and 63 for Spread Spectrum Systems," Apr. 2005; <http://www.3gpp2.org>
- [4] S. P. Lloyd, "Least squares quantization in PCM," IEEE Transactions on Information Theory, Vol. 28, No.2, pp. 129–136, March 1982.
- [5] 3GPP Tech. Spec. 26.290, "Adaptive Multi-Rate - Wideband (AMR-WB+) codec; Transcoding functions," June 2005.
- [6] "Extended high-level description of the Q9 EV-VBR baseline codec," ITU-T SG16 Tech. Cont. COM16-C199R1-E, June 2007.

## REIVINDICACIONES

1. Dispositivo de modo de transición para uso en un códec de señal de sonido de tipo predictivo para producir una excitación de modo de transición que reemplaza una excitación de libro de códigos adaptativo en una trama de transición y/o al menos una trama siguiente a la transición en la señal de sonido, que comprende:

- 5        una entrada para recibir un índice del libro de códigos; y  
       un libro de códigos de modo de transición para generar un conjunto de vectores de código independientes de la excitación anterior, siendo el libro de códigos de modo de transición sensible al índice del libro de códigos para generar, en la trama de transición y/o al menos en una trama siguiente a la transición, correspondiendo uno de los vectores de código del conjunto a dicha excitación de modo de transición;
- 10      en el que el libro de códigos de modo de transición comprende un libro de códigos de formas de impulso glotal.
2. Dispositivo de modo de transición de acuerdo con la reivindicación 1, en el que la señal de sonido comprende una señal de voz, y en la que la trama de transición se selecciona del grupo que consiste en una trama que comprende un inicio sonoro y una trama que comprende una transición entre dos sonidos sonoros diferentes.
- 15      3. Dispositivo de modo de transición de acuerdo con la reivindicación 1, en el que la trama de transición y/o la al menos una trama siguiente a la transición comprenden cada una una serie de subtramas, y en el que el libro de códigos de modo de transición se utiliza en una primera parte de las subtramas, y se utiliza un libro de códigos de tipo predictivo del códec de tipo predictivo en una segunda parte de las subtramas.
- 20      4. Dispositivo de modo de transición de acuerdo con la reivindicación 1, en el que la trama de transición y/o la al menos una trama siguiente a la transición comprenden cada una una serie de subtramas, y en el que el libro de códigos de modo de transición se utiliza solamente en la subtrama que contiene un primer impulso glotal de una trama actual.
- 25      5. Dispositivo de modo de transición de acuerdo con la reivindicación 4, que comprende medios para producir, en al menos una subtrama anterior a la subtrama que utiliza el libro de códigos de modo de transición, una señal de excitación global que comprende exclusivamente un componente del libro de códigos de innovación.
- 30      6. Dispositivo de modo de transición de acuerdo con la reivindicación 1, en el que el libro de códigos de formas de impulso glotal incluye un número predeterminado de formas diferentes de impulsos glotales, y en el que cada forma de impulso glotal está situada en una serie de posiciones diferentes en los vectores de código para formar una serie de vectores de código diferentes del libro de códigos de formas de impulso glotal.
- 35      7. Dispositivo de modo de transición de acuerdo con la reivindicación 6, en el que el libro de códigos de formas de impulso glotal comprende un generador de vectores de código que contienen solo un elemento distinto de cero y un filtro de conformación para procesar los vectores de código que contienen solo un elemento distinto de cero para producir vectores de código que representan formas de impulso glotal centradas en diferentes posiciones.
- 40      8. Dispositivo de modo de transición de acuerdo con la reivindicación 7, en el que la trama de transición y/o la al menos una trama siguiente a la transición comprenden cada una una serie de subtramas, el libro de códigos de formas de impulso glotal comprende además un filtro de repetición situado más abajo del filtro de conformación para repetir, cuando hay más de un impulso glotal por subtrama, la forma de impulso glotal después de que ha transcurrido un período del tono.
- 45      9. Dispositivo de modo de transición de acuerdo con la reivindicación 6, en el que las formas de impulso glotal comprenden muestras primera y última en las que se trunca un número predeterminado de la primera y última muestras.
10. Un dispositivo codificador para generar una excitación de modo de transición que reemplaza a una excitación de libro de códigos adaptativo en una trama de transición y/o al menos una trama siguiente a la transición en una señal de sonido, que comprende:
- 45      un generador de una señal objetivo de búsqueda en el libro de códigos;  
       un libro de códigos de modo de transición para generar un conjunto de vectores de código independientes de la excitación anterior, en el que los vectores de código de dicho conjunto corresponden cada uno a una excitación respectiva del modo de transición, y en el que el libro de códigos de modo de transición comprende un libro de códigos de formas de impulso glotal;
- 50      un buscador en el libro de códigos de modo de transición para encontrar el vector de código de dicho conjunto correspondiente a una excitación del modo de transición que corresponde óptimamente a la señal objetivo de búsqueda en el libro de códigos.

11. Dispositivo codificador de acuerdo con la reivindicación 10, en el que el buscador aplica un criterio dado a cada forma de impulso glotal del libro de códigos de formas de impulso glotal, y encuentra como vector de código que corresponde óptimamente a la señal objetivo de búsqueda en el libro de códigos, el vector de código del conjunto correspondiente a un valor máximo de dicho criterio.
- 5 12. Dispositivo codificador de acuerdo con la reivindicación 11, en el que el buscador identifica el vector de código encontrado por medio de parámetros de modo de transición seleccionados del grupo que consiste en una identificación de configuración del modo de transición, una forma de impulso glotal, una posición del centro de la forma de impulso glotal en el vector de código encontrado, una ganancia del modo de transición, un signo de la ganancia del modo de transición y un periodo de paso de bucle cerrado.
- 10 13. Dispositivo codificador de acuerdo con la reivindicación 10, en el que la señal de sonido comprende una señal de voz y en la que la trama de transición se selecciona del grupo que consiste en una trama que comprende un inicio sonoro y una trama que comprende una transición entre dos sonidos sonoros diferentes.
- 15 14. Dispositivo codificador de acuerdo con la reivindicación 10, en el que la trama de transición y/o la al menos una trama siguiente a la transición comprenden cada una una serie de subtramas, y en el que el buscador realiza una búsqueda en el libro de códigos de modo de transición en una primera parte de las subtramas y en un libro de códigos de tipo predictivo del dispositivo codificador en una segunda parte de las subtramas.
- 20 15. Dispositivo codificador de acuerdo con la reivindicación 10, en el que la trama de transición y/o la al menos una trama siguiente a la transición comprenden cada una una serie de subtramas, y en el que el libro de códigos de modo de transición se utiliza solamente en la subtrama que contiene un primer impulso glotal de una trama actual.
- 25 16. Dispositivo codificador de acuerdo con la reivindicación 15, que comprende medios para producir, en al menos una subtrama que precede a la subtrama que utiliza el libro de códigos de modo de transición, una señal de excitación global que comprende exclusivamente un componente de libro de códigos de innovación.
- 30 17. Dispositivo codificador de acuerdo con la reivindicación 10, en el que el libro de códigos de formas de impulso glotal incluye un número predeterminado de formas diferentes de impulsos glotales, y en el que cada forma de impulso glotal está situada en una serie de posiciones diferentes en los vectores de código para formar una serie de vectores de código diferentes del libro de códigos de formas de impulso glotal.
- 35 18. Dispositivo codificador de acuerdo con la reivindicación 17, en el que el libro de códigos de formas de impulso glotal comprende un generador de vectores de código que contienen solo un elemento distinto de cero y un filtro de conformación para procesar los vectores de código que contienen solo un elemento distinto de cero para producir vectores de código que representan formas de impulso glotal centradas en diferentes posiciones.
- 40 19. Dispositivo codificador de acuerdo con la reivindicación 10, que comprende, además:
- un generador de una señal objetivo de búsqueda en el libro de códigos de innovación;
  - un libro de códigos de innovación para generar un conjunto de vectores de código de innovación, correspondiendo cada uno a una excitación de innovación respectiva;
- 45 35 un buscador del libro de códigos de innovación para encontrar el vector de código de innovación de dicho conjunto correspondiente a una excitación de innovación que corresponde óptimamente a la señal objetivo de búsqueda en el libro de códigos de innovación; y
- un sumador de la excitación del modo de transición y la excitación de innovación para producir una excitación global para un filtro de síntesis de señal de sonido;
- 50 40 en el que la trama de transición y/o la al menos una trama siguiente a la transición comprenden cada una una serie de subtramas, y en donde, dependiendo de dónde se encuentran un impulso glotal o impulsos glotales en las subtramas, el dispositivo codificador comprende medios para codificar las subtramas utilizando al menos uno del libro de códigos de modo de transición, el libro de códigos adaptativo y el libro de códigos de innovación.
- 45 50 20. Dispositivo decodificador que comprende el dispositivo de modo de transición de acuerdo con una cualquiera de las reivindicaciones 1 a 9, para producir la excitación del modo de transición que reemplaza a la excitación de libro de códigos adaptativo en la trama de transición y/o en la al menos una trama después de la transición en la señal de sonido.
21. Dispositivo decodificador de acuerdo con la reivindicación 20, que comprende, además:
- 50 una entrada para recibir un índice del libro de códigos de innovación;
- un libro de códigos de innovación para generar un conjunto de vectores de código de innovación, siendo el libro de códigos de innovación sensible al índice del libro de códigos de innovación para generar en la trama

de transición y/o la al menos una trama siguiente a la transición de uno de los vectores de código de innovación del conjunto correspondiente a una excitación de innovación;

un sumador de la excitación del modo de transición y la excitación de innovación para producir una excitación global para un filtro de síntesis de señal de sonido.

5 22. Método de modo de transición para uso en un códec de señal de sonido de tipo predictivo para producir una excitación I modo de transición que reemplaza a una excitación de libro de códigos adaptativo en una trama de transición y/o al menos una trama siguiente a la transición en la señal de sonido, que comprende:

10 proporcionar un libro de códigos de modo de transición para generar un conjunto de vectores de código independientes de la excitación anterior, en el que el libro de códigos de modo de transición comprende un libro de códigos de formas de impulso glotal;

suministrar un índice del libro de códigos al libro de códigos de modo de transición;

y generar, por medio del libro de códigos de modo de transición y en respuesta al índice del libro de códigos, uno de los vectores de código del conjunto correspondiente a dicha excitación del modo de transición.

15 23. Método de modo de transición de acuerdo con la reivindicación 22, en el que la señal de sonido comprende una señal de voz, y dicho método comprende seleccionar la trama de transición del grupo que consiste en una trama que comprende un inicio sonoro y una trama que comprende una transición entre dos sonidos sonoros diferentes.

20 24. Método de modo de transición de acuerdo con la reivindicación 22, en el que la trama de transición y/o la al menos una trama siguiente a la transición comprenden cada una una serie de subtramas, y dicho método comprende utilizar el libro de códigos de modo de transición en una primera parte de las subtramas y un libro de códigos del códec de tipo predictivo en una segunda parte de las subtramas

25 25. Método de modo de transición de acuerdo con la reivindicación 22, en el que la trama de transición y/o la al menos una trama siguiente a la transición comprenden cada una una serie de subtramas y en el que dicho método comprende utilizar el libro de códigos de modo de transición en la subtrama que contiene un primer impulso glotal de una trama actual.

26. Método de modo de transición tal como se define en la reivindicación 25, que comprende producir, en al menos una subtrama anterior a la subtrama que utiliza el libro de códigos de modo de transición, una señal de excitación global que comprende exclusivamente un componente del libro de códigos de innovación.

30 27. Método de modo de transición de acuerdo con la reivindicación 22, en el que el libro de códigos de formas de impulso glotal incluye un número predeterminado de formas diferentes de impulsos glotales, y en el que el método de modo de transición comprende formar en el libro de códigos de formas de impulso glotal una serie de vectores de código diferentes situando cada forma de impulso glotal en una serie de posiciones diferentes en el vector de código.

35 28. Método de modo de transición de acuerdo con la reivindicación 27, que comprende, en el libro de códigos de formas de impulso glotal, generar vectores de código que contienen solo un elemento distinto de cero, y procesar a través de un filtro de conformación los vectores de código que contienen solo un elemento distinto de cero para producir vectores de código que representan formas de impulso glotal centradas en diferentes posiciones.

29. Método de codificación para generar una excitación de modo de transición que reemplaza a una excitación de libro de códigos adaptativo en una trama de transición y/o al menos una trama siguiente a la transición en una señal de sonido, que comprende:

40 generar una señal objetivo de búsqueda en el libro de códigos;

proporcionar un libro de códigos de modo de transición para generar un conjunto de vectores de código independientes de la excitación anterior, correspondiendo cada uno de los vectores de código de dicho conjunto a una excitación del modo de transición, en el que el libro de códigos de modo de transición comprende un libro de códigos de formas de impulso glotal;

45 realizar una búsqueda en el libro de códigos de modo de transición para encontrar el vector de código de dicho conjunto correspondiente a una excitación de modo de transición que corresponde óptimamente a la señal objetivo de búsqueda en el libro de códigos.

50 30. Método de codificación de acuerdo con la reivindicación 29, en el que realizar una búsqueda en el libro de códigos de modo de transición comprende aplicar un criterio dado a cada forma de impulso glotal del libro de códigos de formas de impulso glotal y encontrar como vector de código que corresponde óptimamente a la señal objetivo de búsqueda en el libro de códigos el vector de código del conjunto correspondiente a un valor máximo de dicho criterio.

31. Método de codificación de acuerdo con la reivindicación 29, en el que la señal de sonido comprende una señal de voz, y dicho método comprende además seleccionar la trama de transición del grupo que consiste en una trama que comprende un inicio sonoro y una trama que comprende una transición entre dos sonidos sonoros diferentes.
- 5 32. Método de codificación de acuerdo con la reivindicación 29, en el que la trama de transición y/o la al menos una trama siguiente a la transición comprenden cada una una serie de subtramas, y en el que la búsqueda en el libro de códigos de modo de transición comprende realizar una búsqueda en el libro de códigos de modo de transición en una primera parte de las subtramas, y realizar una búsqueda en un libro de códigos de tipo predictivo del dispositivo codificador en una segunda parte de las subtramas.
- 10 33. Método de codificación de acuerdo con la reivindicación 29, en el que la trama de transición y/o la al menos una trama siguiente a la transición comprenden cada una una serie de subtramas, y en el que dicho método comprende utilizar el libro de códigos de modo de transición en la subtrama que contiene un primer impulso glotal de una trama actual.
- 15 34. Método de codificación de acuerdo con la reivindicación 33, que comprende producir, en al menos una subtrama que precede a la subtrama que utiliza el libro de códigos de modo de transición, una señal de excitación global que comprende exclusivamente un componente del libro de códigos de innovación.
35. Método de codificación de acuerdo con la reivindicación 29, en el que el libro de códigos de formas de impulso glotal incluye un número predeterminado de formas diferentes de impulsos glotales, y el método de codificación comprende formar una serie de vectores de código diferentes del libro de códigos de formas de impulso glotal situando cada forma del impulso glotal en una serie de posiciones diferentes en los vectores de código.
- 20 36. Método de codificación de acuerdo con la reivindicación 35, en el que generar en el libro de códigos de formas de impulso glotal el conjunto de vectores de código independientes de la excitación anterior comprende generar vectores de código que contienen solo un elemento distinto de cero y procesar a través de un filtro de conformación los vectores de código que contienen solo un elemento distinto de cero para producir vectores de código que representan formas de impulso glotal centradas en diferentes posiciones.
- 25 37. Método de codificación de acuerdo con la reivindicación 29, que comprende, además:
- generar una señal objetivo de búsqueda en el libro de códigos de innovación;
- proporcionar un libro de códigos de innovación para generar un conjunto de vectores de código de innovación, cada uno correspondiente a una excitación de innovación respectiva;
- 30 realizar una búsqueda en el libro de códigos de innovación para encontrar el vector de código de innovación de dicho conjunto correspondiente a una excitación de innovación que corresponde óptimamente a la señal objetivo de búsqueda en el libro de códigos de innovación; y
- añadir la excitación del modo de transición y la excitación de innovación para producir una excitación global para un filtro de síntesis de la señal de sonido;
- 35 en el que la trama de transición y/o la a al menos una trama siguiente a la transición comprenden cada una una serie de subtramas, y en el que, dependiendo de dónde estén situados el impulso glotal o los impulsos glotales en las subtramas, el método de codificación comprende codificar las subtramas utilizando al menos uno del libro de códigos de modo de transición, el libro de códigos adaptativo y el libro de códigos de innovación.
- 40 38. Método de decodificación que comprende el método de transición de acuerdo con una cualquiera de las reivindicaciones 22 a 28 para producir la excitación del modo de transición que reemplaza a la excitación de libro de códigos adaptativo en la trama de transición y/o la al menos una trama siguiente a la transición en la señal de sonido.
- 45 39. Método de decodificación de acuerdo con la reivindicación 38, que comprende, además:
- proporcionar un libro de códigos de innovación para generar un conjunto de vectores de código de innovación;
- suministrar un índice del libro de códigos de innovación al libro de códigos de innovación;
- generar, mediante el libro de códigos de innovación y en respuesta al índice del libro de códigos de innovación, uno de los vectores de código de innovación del conjunto correspondiente a una excitación de innovación; y
- 50 añadir la excitación del modo de transición y la excitación de la innovación para producir una excitación global para un filtro de síntesis de la señal de sonido.

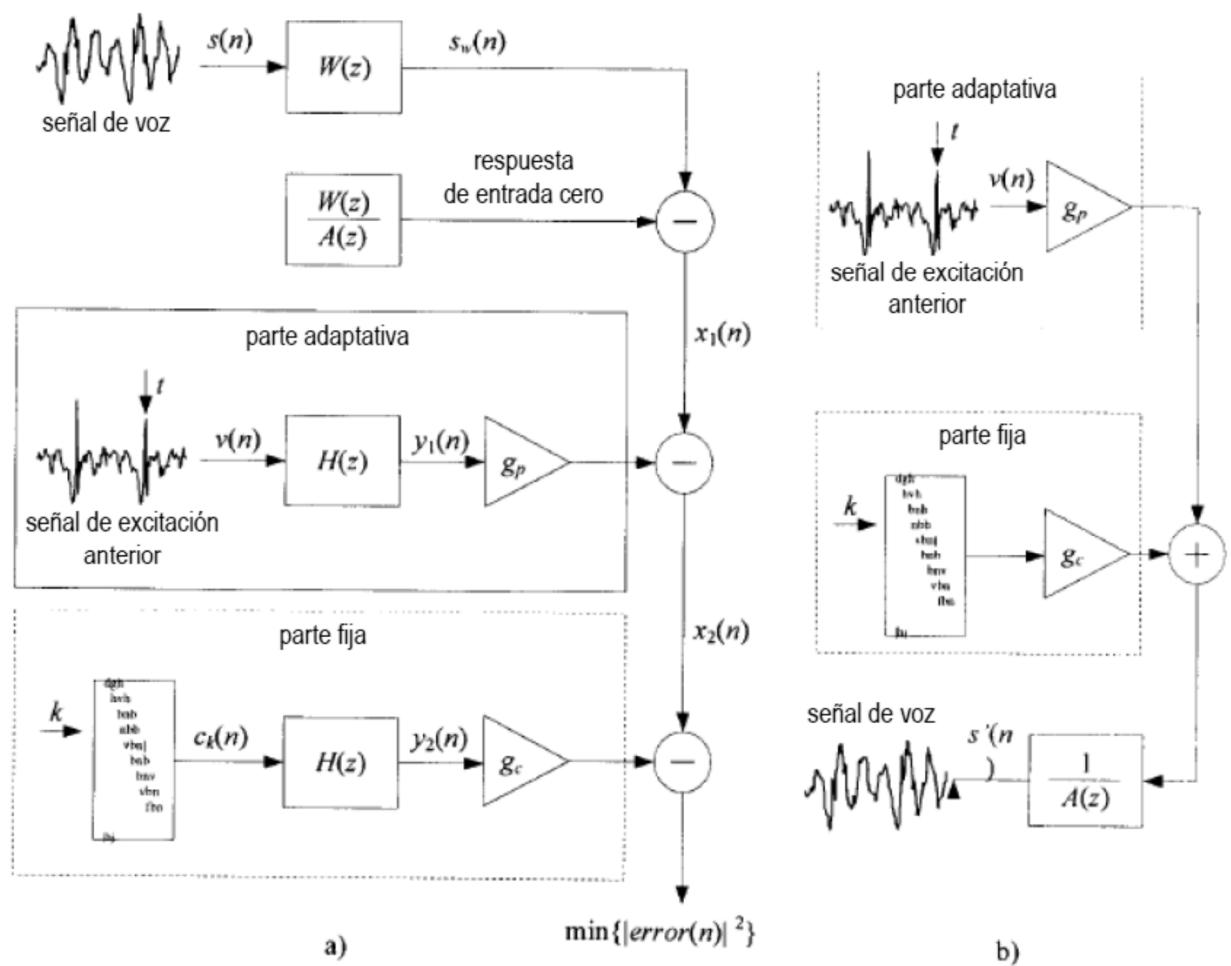


Figura 1

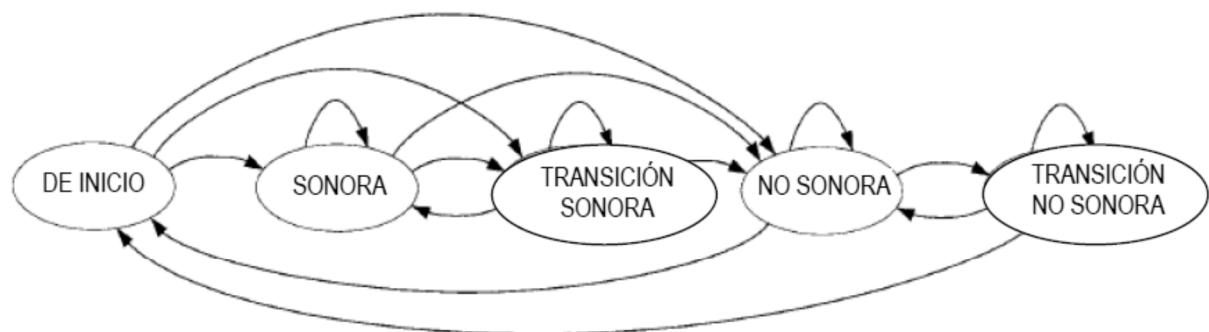


Figura 2

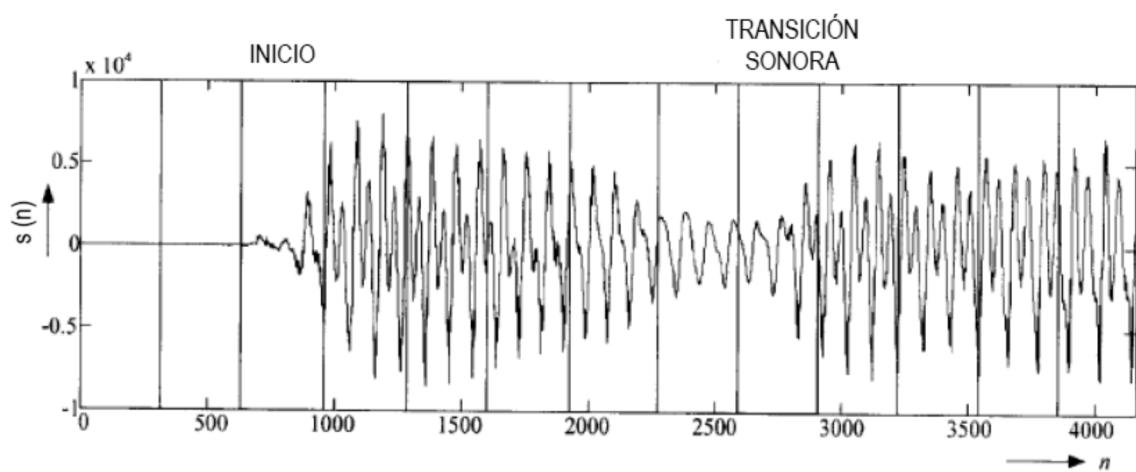


Figura 3

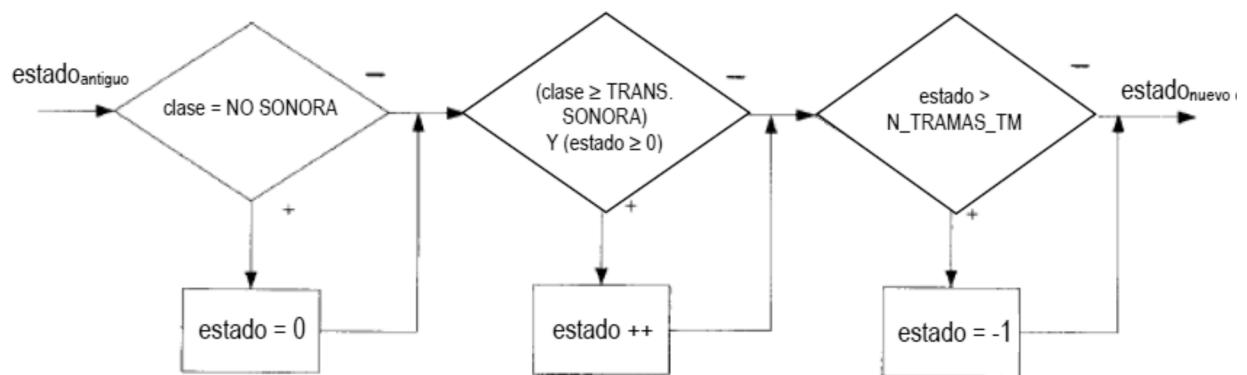


Figura 4

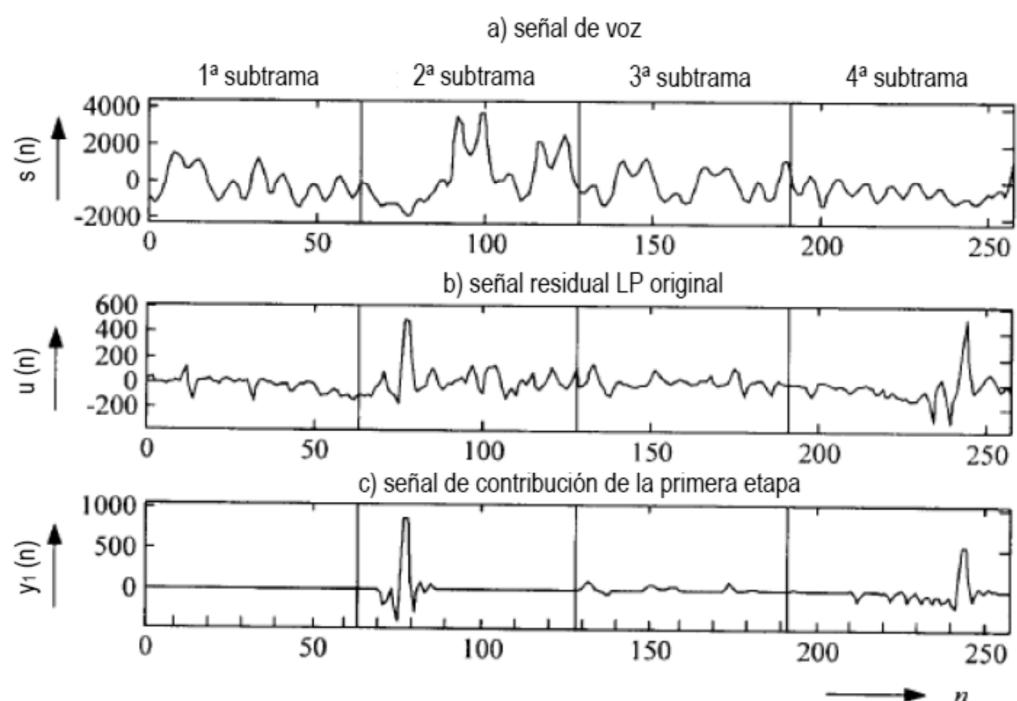


Figura 5

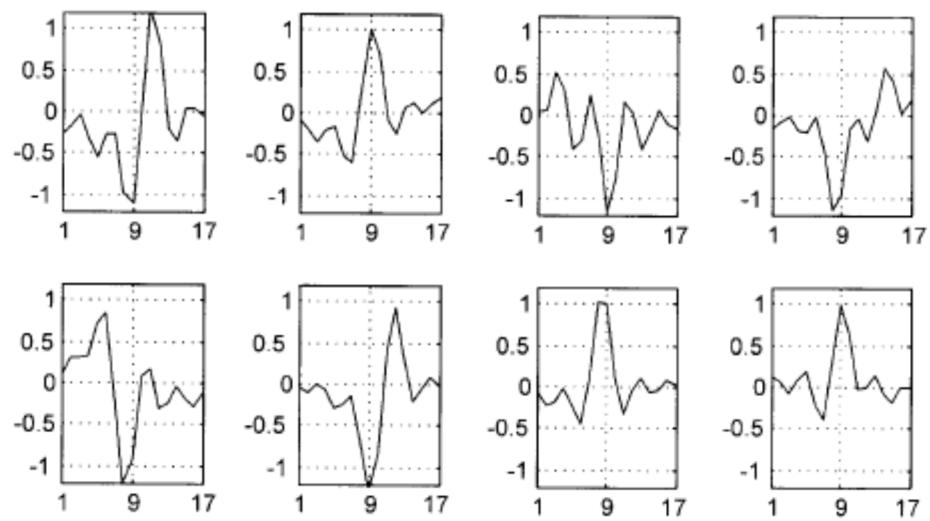


Figura 6

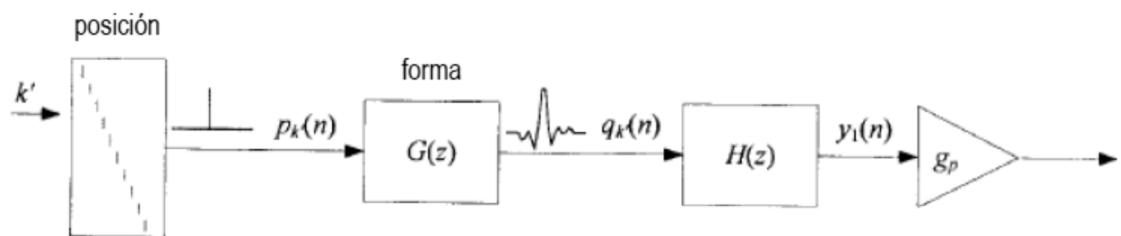


Figura 7

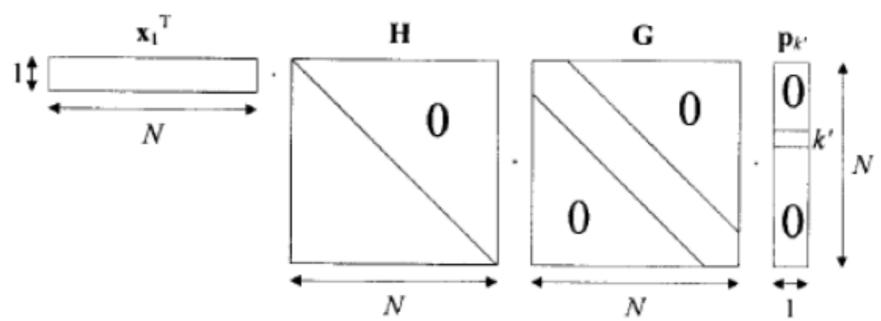


Figura 8

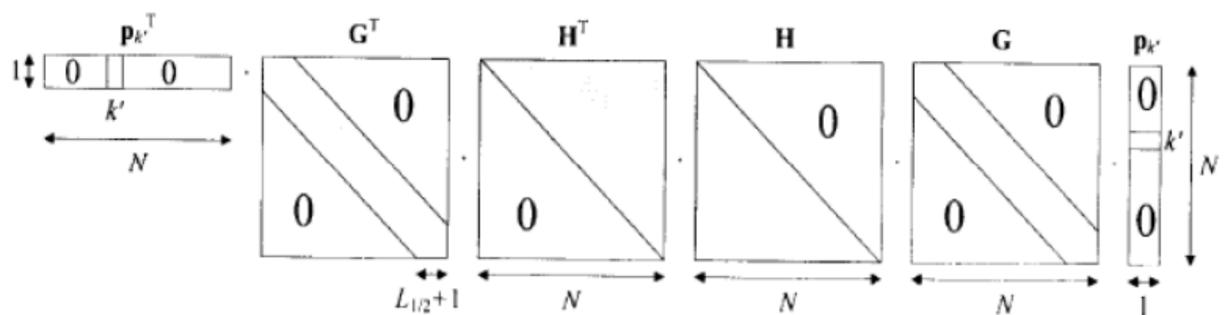


Figura 9

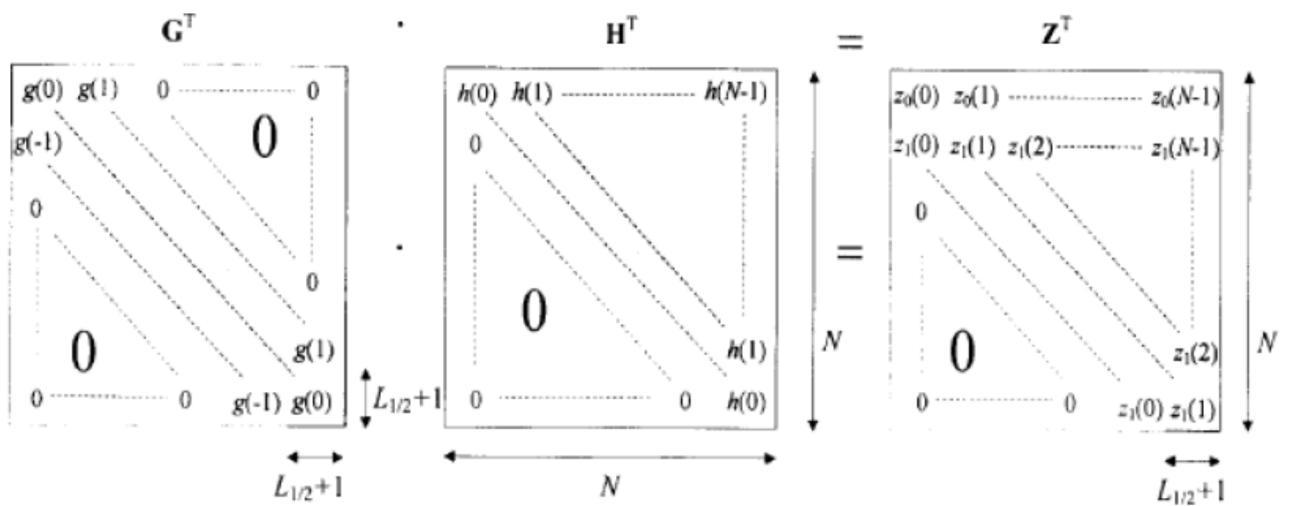


Figura 10

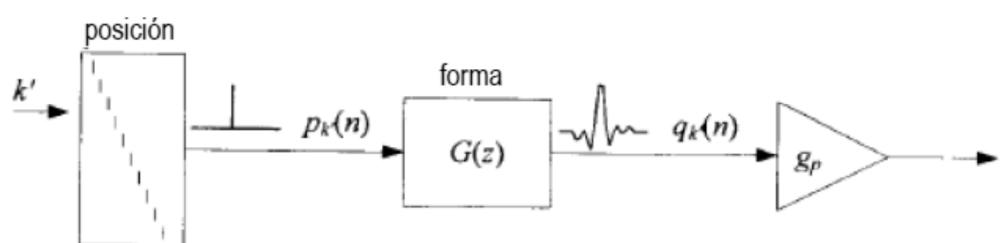


Figura 11

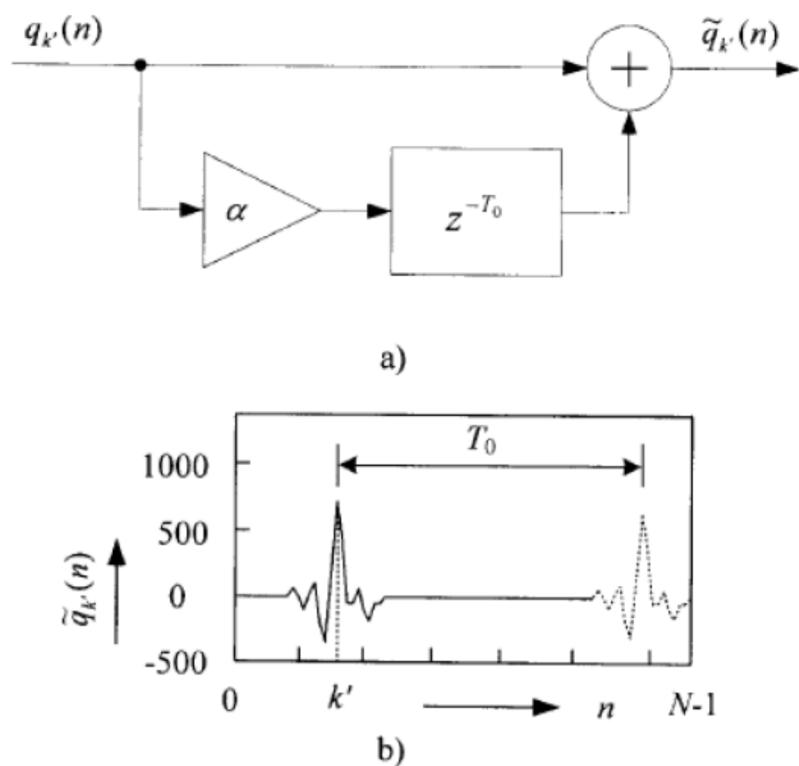


Figura 12

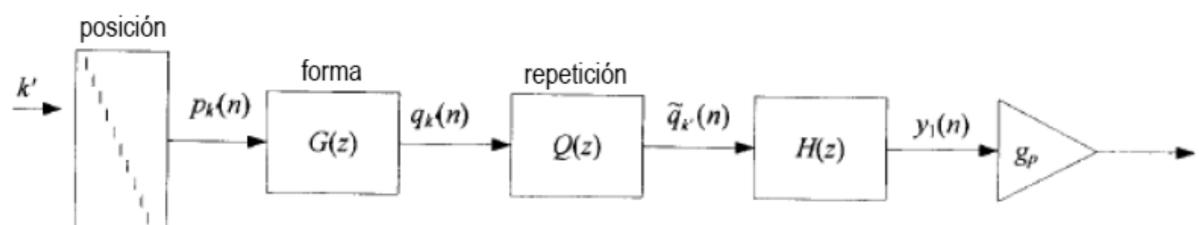


Figura 13

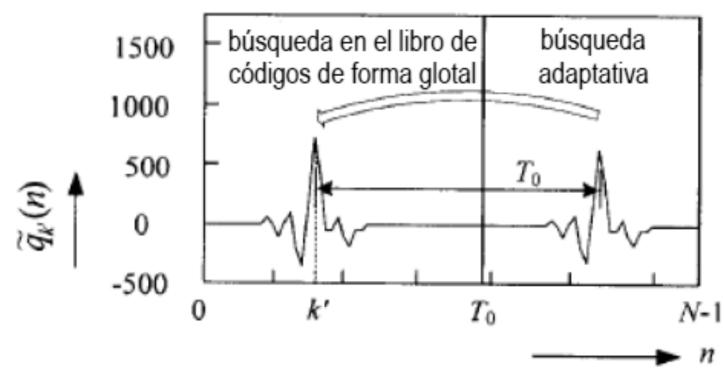


Figura 14

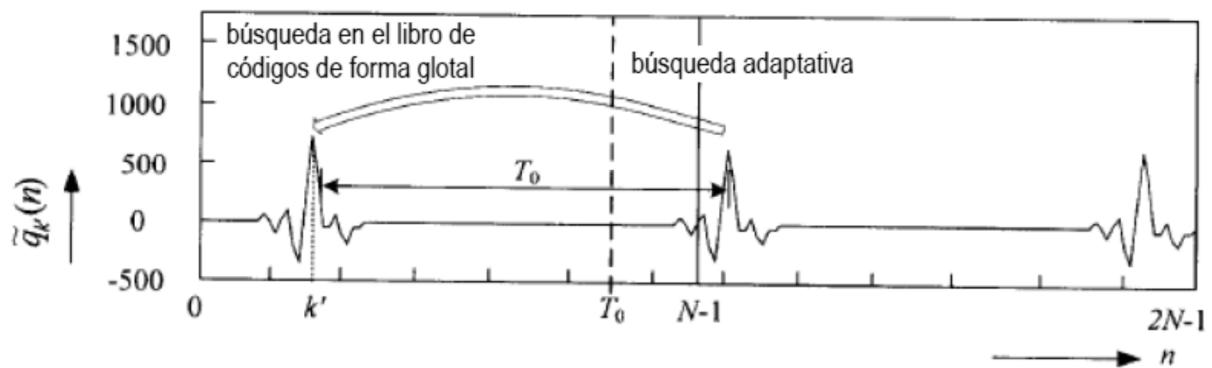


Figura 15

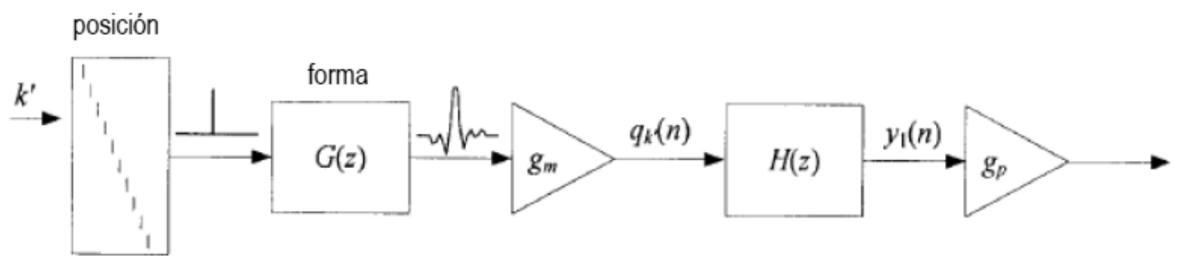


Figura 16

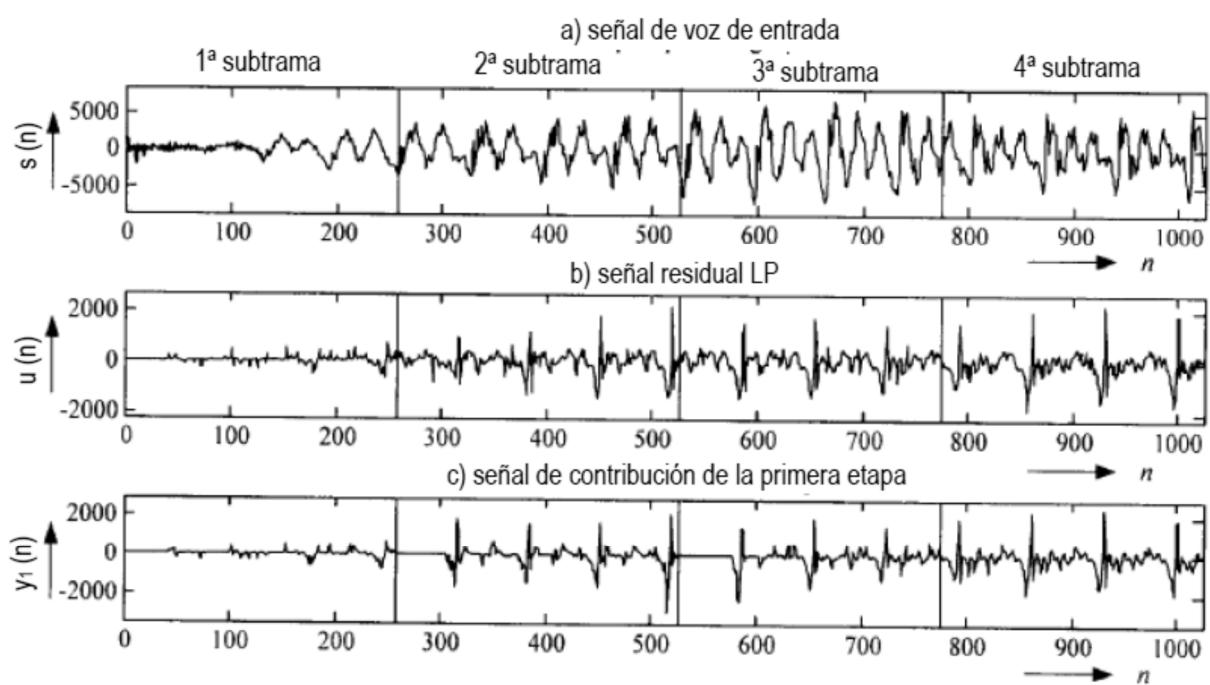


Figura 17

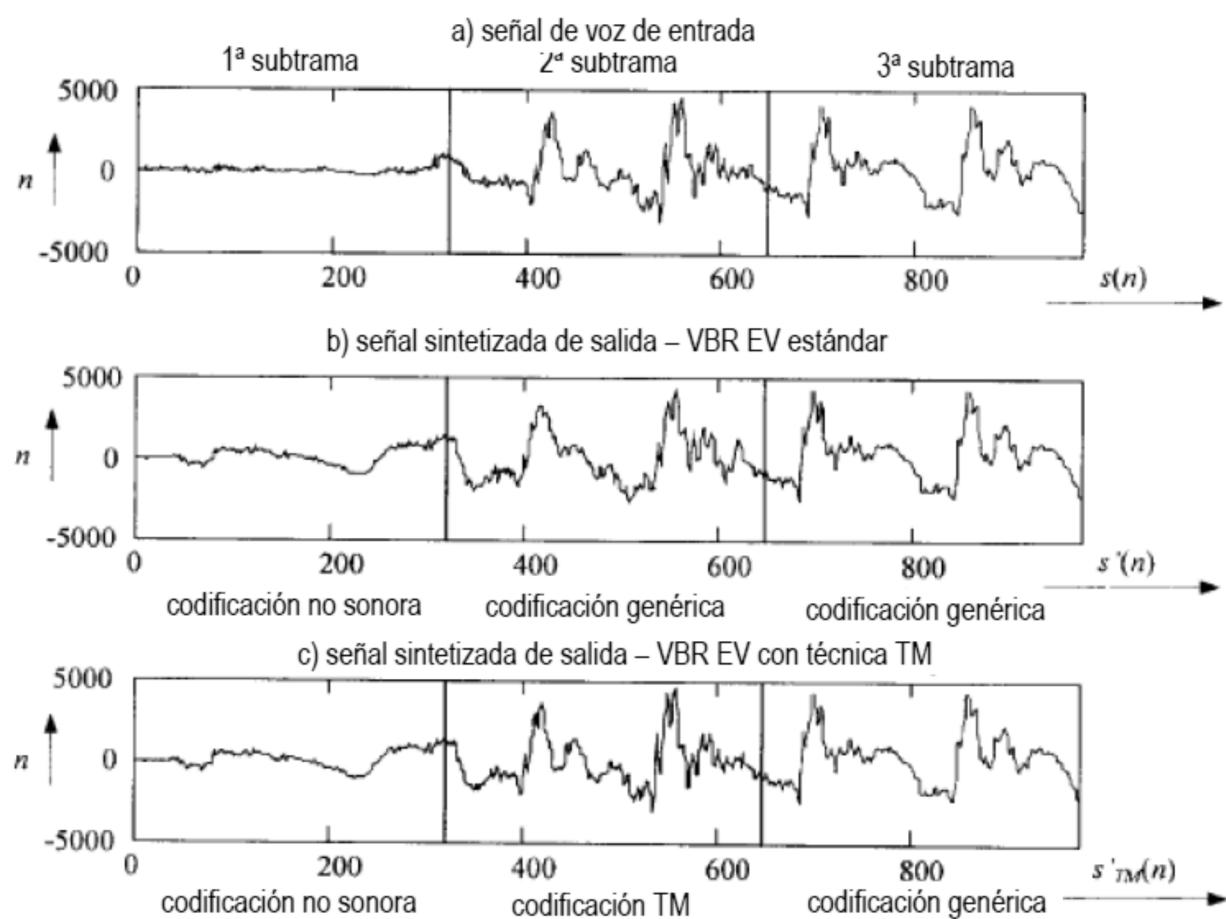


Figura 18

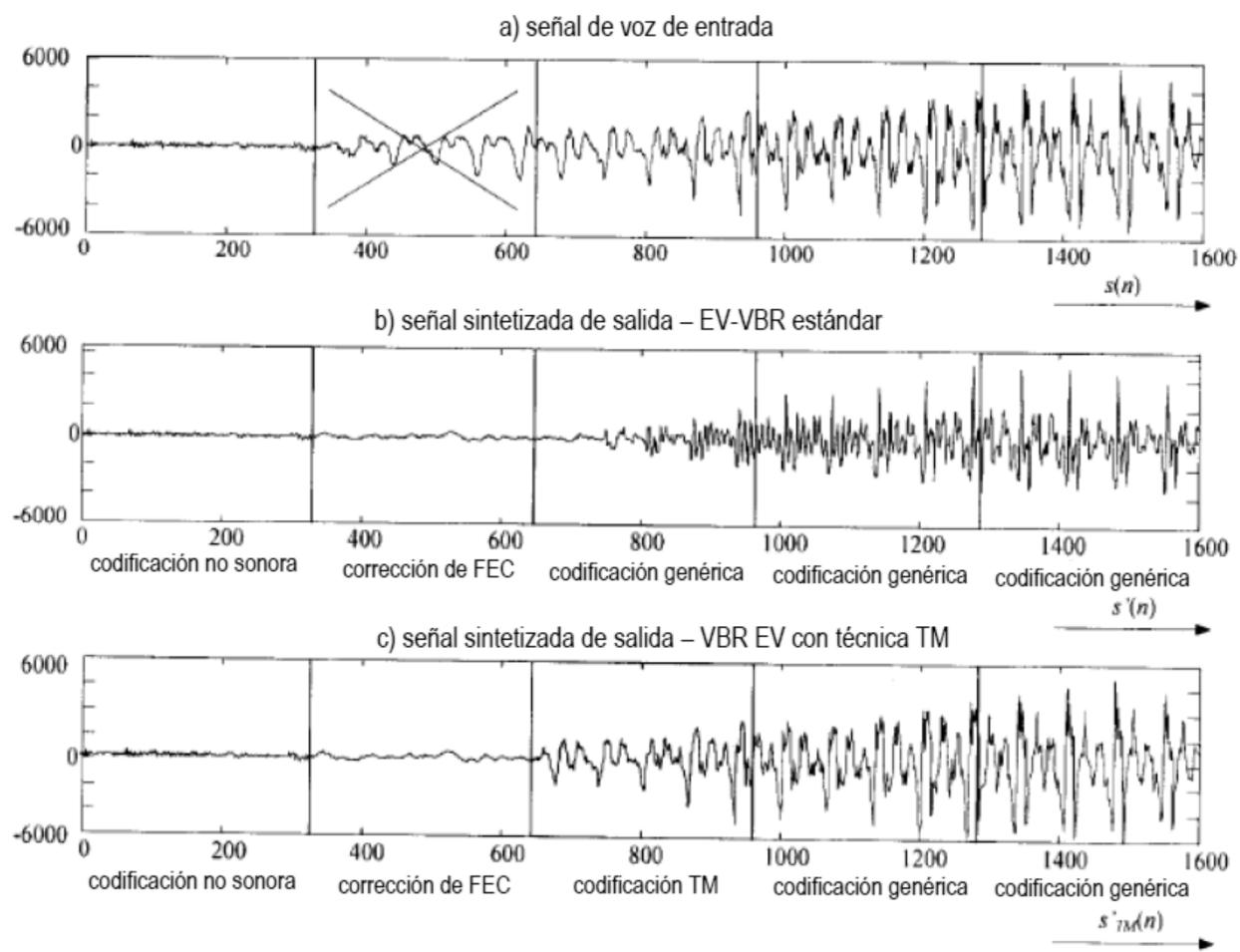


Figura 19

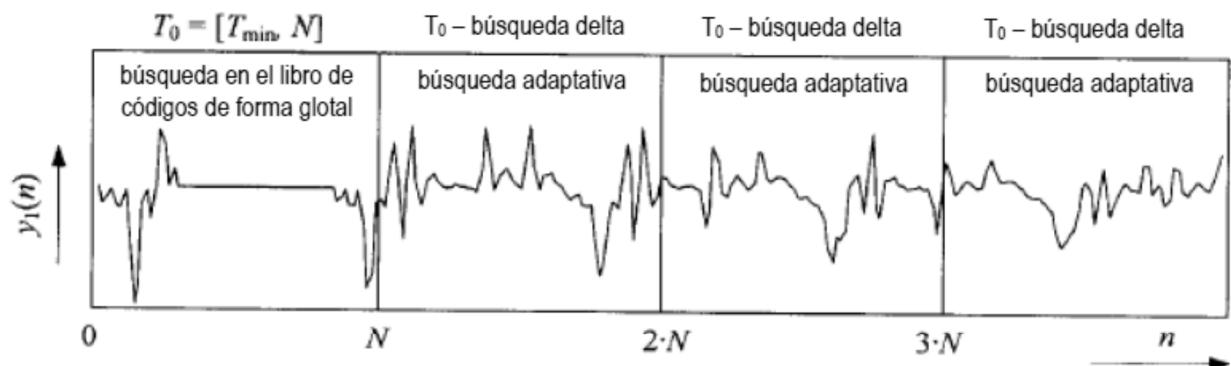


Figura 20

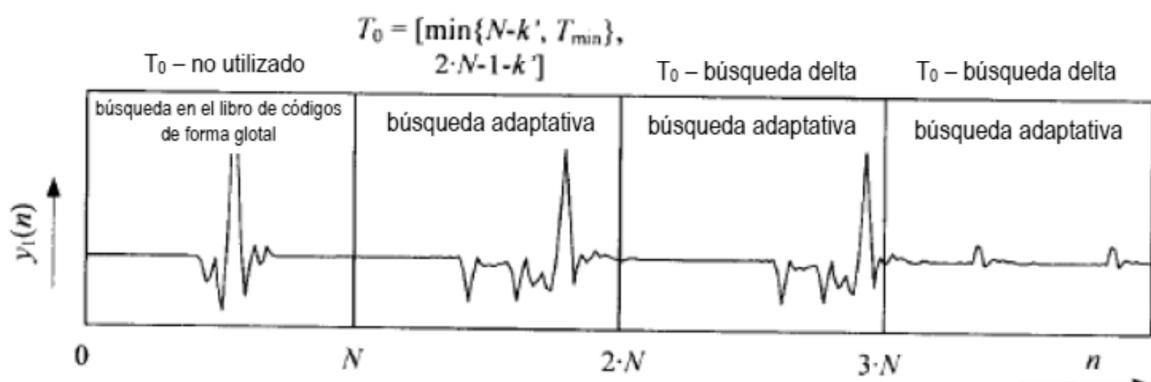


Figura 21

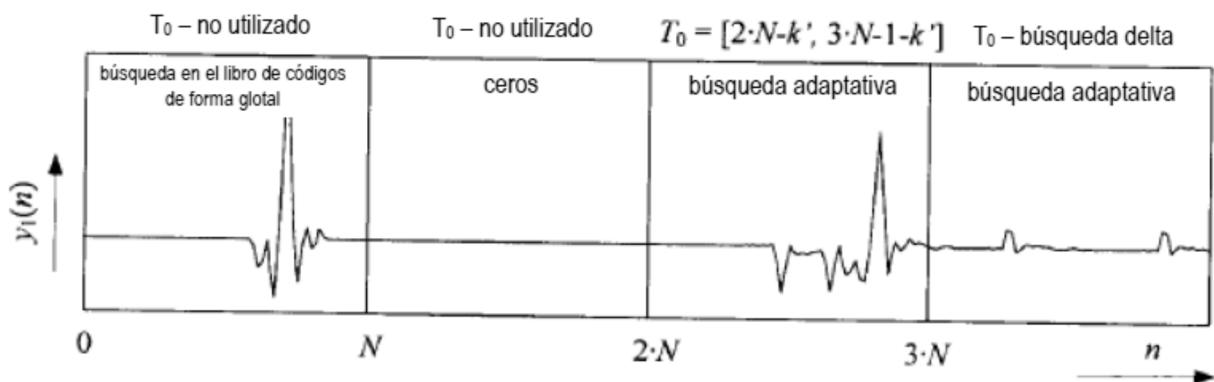


Figura 22

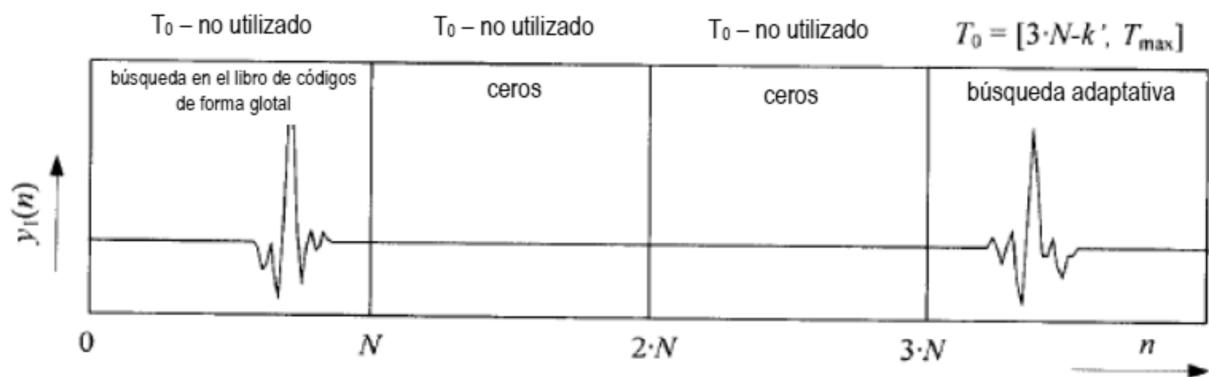


Figura 23



Figura 24

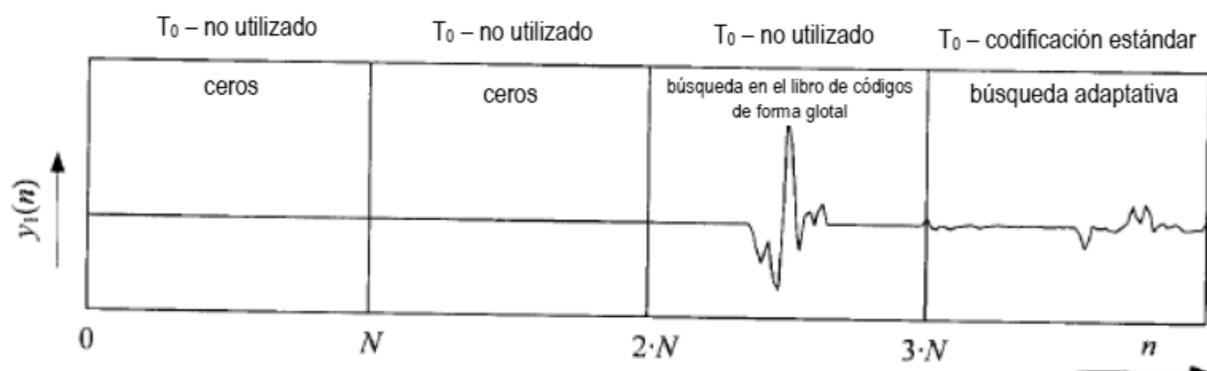


Figura 25

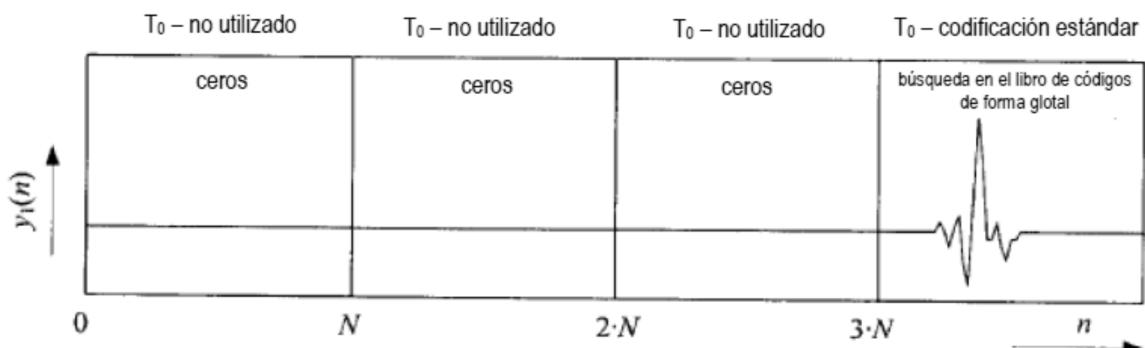


Figura 26

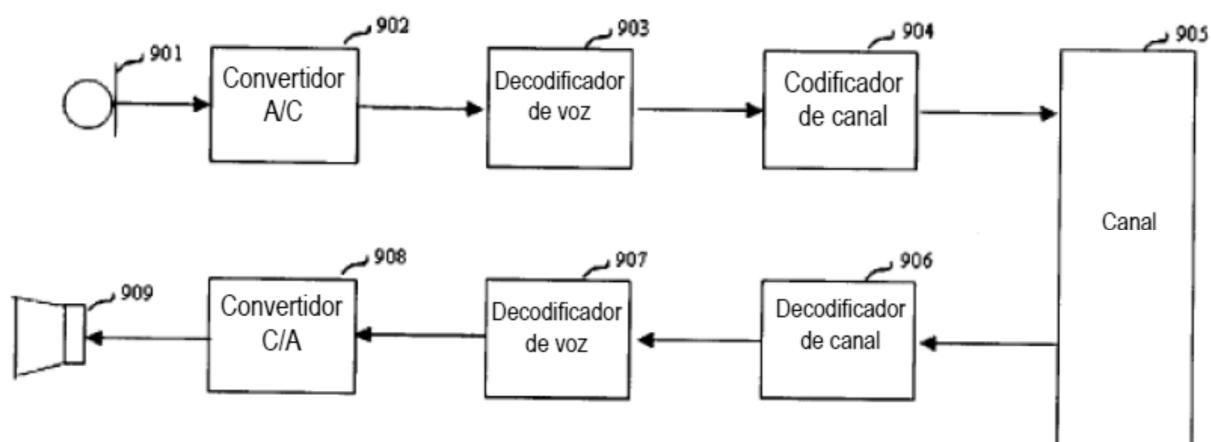


Figura 27