(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 721 789**

(51) Int. CI.:

***G10L 19/00*** (2013.01)
***G10L 19/002*** (2013.01)
***G10L 19/22*** (2013.01)
***G10L 19/125*** (2013.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **23.07.2015 PCT/CN2015/084931**

(87) Fecha y número de publicación internacional: **04.02.2016 WO16015591**

(96) Fecha de presentación y número de la solicitud europea: **23.07.2015 E 15828041 (2)**

(97) Fecha y número de publicación de la concesión europea: **13.02.2019 EP 3152755**

(54) Título: **Mejorar la clasificación entre codificación en el dominio del tiempo y codificación en el dominio de la frecuencia**

(30) Prioridad:

**26.07.2014 US 201462029437 P**
**10.10.2014 US 201414511943**

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
**05.08.2019**

(73) Titular/es:

**HUAWEI TECHNOLOGIES CO., LTD. (100.0%)**
**Huawei Administration Building Bantian Longgang District**
**Shenzhen, Guangdong 518129 CN**

(72) Inventor/es:

**GAO, YANG**

(74) Agente/Representante:

**LEHMANN NOVO, María Isabel**

## DESCRIPCIÓN

Mejorar la clasificación entre codificación en el dominio del tiempo y codificación en el dominio de la frecuencia

Campo técnico

La presente invención está, en general, en el campo de la codificación de señales. En particular, la presente invención está en el campo de mejorar la clasificación entre la codificación en el dominio del tiempo y la codificación en el dominio de la frecuencia.

Antecedentes

La codificación de voz se refiere a un proceso que reduce la tasa de bits de un archivo de voz. La codificación de voz es una aplicación de compresión de datos de señales de audio digital que contienen voz. La codificación de voz utiliza la estimación de parámetros específicos de la voz, utilizando técnicas de procesamiento de señales de audio para modelar la señal de voz, combinada con algoritmos de compresión de datos genéricos para representar los parámetros modelados resultantes en un flujo de bits compacto. El objetivo de la codificación de voz es lograr un ahorro en el espacio de almacenamiento de memoria, el ancho de banda de transmisión y la potencia de transmisión requeridos, al reducir el número de bits por muestra, de manera que la voz decodificada (descomprimida) es perceptivamente indistinguible de la voz original.

Sin embargo, los codificadores de voz son codificadores con pérdida, es decir, la señal decodificada es diferente de la original. Por lo tanto, uno de los objetivos de la codificación de voz es minimizar la distorsión (o pérdida perceptible) en una tasa de bits dada, o minimizar la tasa de bits para alcanzar una distorsión dada.

La codificación de voz difiere de otras formas de codificación de audio, en que la voz es una señal mucho más simple que la mayoría de las otras señales de audio y hay mucha más información estadística disponible sobre las propiedades de la voz. Como resultado, cierta información auditiva que es relevante en la codificación de audio puede ser innecesaria en el contexto de la codificación de voz. En la codificación de voz, el criterio más importante es la preservación de la inteligibilidad y la “afabilidad” de la voz, con una cantidad limitada de datos transmitidos.

La inteligibilidad de la voz incluye, además del contenido literal real, también, la identidad, las emociones, la entonación, el timbre, etc. del hablante, que son todas importantes para la inteligibilidad perfecta. El concepto más abstracto de la afabilidad de la voz degradada es una propiedad diferente de la inteligibilidad, ya que es posible que la voz degradada sea completamente inteligible, pero subjetivamente molesta para el oyente.

Tradicionalmente, todos los métodos de codificación de voz paramétrica hacen uso de la redundancia inherente a la señal de voz para reducir la cantidad de información que debe enviarse y para estimar los parámetros de las muestras de voz de una señal en intervalos cortos. Esta redundancia surge principalmente de la repetición de formas de onda de la voz en una frecuencia casi periódica y la envolvente espectral que cambia lenta de la señal de voz.

La redundancia de las formas de onda de la voz puede considerarse con respecto a varios tipos diferentes de señales de voz, tales como señales de voz sonora y no sonora. Los sonidos de voz, p. ej., ‘a’, ‘b’, se deben esencialmente a vibraciones de las cuerdas vocales y son oscilatorias. Por lo tanto, durante cortos períodos de tiempo, están bien modeladas mediante sumas de señales periódicas, tales como sinusoides. En otras palabras, para la voz sonora, la señal de voz es esencialmente periódica. Sin embargo, esta periodicidad puede ser variable durante la duración de un segmento de voz y la forma de la onda periódica, generalmente, cambia gradualmente de segmento a segmento. Una codificación de voz de tasa de bits baja podría beneficiarse enormemente de la exploración de dicha periodicidad. Una codificación de voz en el dominio del tiempo podría beneficiarse enormemente de la exploración de dicha periodicidad. El período de voz sonora también se llama paso y la predicción de paso, a menudo, se denomina Predicción a Largo Plazo (LTP). En contraste, los sonidos sin voz, tales como el ‘s’, ‘sh’, son más parecidos al ruido. Esto se debe a que la señal de voz no sonora es más como un ruido aleatorio y tiene una menor cantidad de predictibilidad.

En cualquier caso, la codificación paramétrica se puede utilizar para reducir la redundancia de los segmentos de voz al separar la componente de excitación de la señal de voz de la componente de envolvente espectral, que cambia a una tasa más lenta. La componente de envolvente espectral que cambia lentamente puede representarse mediante la Codificación de Predicción Lineal (LPC), también llamada Predicción a Corto Plazo (STP). Una codificación de voz de tasa de bits baja también podría beneficiarse mucho de la exploración de dicha Predicción a Corto Plazo. La ventaja de codificación surge de la tasa baja a la que cambian los parámetros. Sin embargo, es raro que los parámetros sean significativamente diferentes de los valores mantenidos dentro de unos pocos milisegundos.

En estándares bien conocidos más recientes, como el G.723.1, G.729, G.718, Tasa Completa Mejorada (EFR), Vocodificador de Modo Seleccionable (SMV), Multi-Tasa Adaptativa (AMR), Banda ancha Multimodo de Tasa Variable (VMR-WB), o Banda Ancha Multi-Tasa Adaptativa (AMR-WB), se ha adoptado la Técnica de Predicción Lineal con Excitación de Código ("CELP"). La CELP se entiende comúnmente como una combinación técnica de Excitación Codificada, Predicción a Largo Plazo y Predicción a Corto Plazo. La CELP se utiliza principalmente para codificar la señal de voz al beneficiarse de las características específicas de la voz humana o del modelo de producción de voz vocal humana. La Codificación de Voz de CELP es un principio de algoritmo muy popular en el área de compresión de voz, aunque los detalles de CELP para diferentes códecs podrían ser significativamente diferentes. Debido a su popularidad, el algoritmo de CELP se ha utilizado en diversos estándares de UIT-T, de MPEG, de 3GPP y de 3GPP2. Las variantes de CELP incluyen CELP algebraico, CELP relajado, CELP de bajo retardo y predicción lineal excitada de suma vectorial, y otros. CELP es un término genérico para una clase de algoritmos y no para un códec en particular.

El algoritmo de CELP se basa en cuatro ideas principales. Primero, se utiliza un modelo de filtro de fuente de producción de voz a través de predicción lineal (LP). El modelo de filtro de fuente de producción de voz, modela la voz como una combinación de una fuente de sonido, tal como las cuerdas vocales, y un filtro acústico lineal, el tracto vocal (y la característica de radiación). En la implementación del modelo de filtro de fuente de producción de voz, la fuente de sonido, o la señal de excitación, a menudo se modela como un tren de impulsos periódico, para voz sonora, o ruido blanco para voz no sonora. En segundo lugar, se utiliza un libro de códigos adaptativo y uno fijo como la entrada (excitación) del modelo de LP. En tercer lugar, se realiza una búsqueda en un circuito cerrado en un "dominio ponderado perceptivamente". En cuarto lugar, se aplica la cuantización vectorial (VQ).

El documento US 20140081629 A1 da a conocer que la calidad de las señales codificadas puede mejorarse reclasificando las señales de AUDIO que transportan datos de no voz como señales de VOZ cuando los parámetros de periodicidad de la señal satisfacen uno o más criterios. En algunas realizaciones, solo se consideran las señales de tasa de bits baja o media para la reclasificación. Los parámetros de periodicidad pueden incluir cualquier característica o conjunto de características indicativas de periodicidad. Por ejemplo, el parámetro de periodicidad puede incluir diferencias de paso entre subtramas en la señal de audio, una correlación de paso normalizada para una o más subtramas, una correlación de paso normalizada promedio para la señal de audio o combinaciones de las mismas. Las señales de audio que se reclasifican como señales SONORA pueden codificarse en el dominio del tiempo, mientras que las señales de audio que permanecen clasificadas como señales de AUDIO pueden codificarse en el dominio de la frecuencia.

Resumen

De acuerdo con una realización de la presente invención, un método para procesar señales de voz antes de codificar una señal digital que comprende datos de audio, incluye seleccionar la codificación en el dominio de la frecuencia o la codificación en el dominio del tiempo en base a una tasa de bits de codificación a ser utilizada para codificar la señal digital y una detección de retardo de paso corto de la señal digital.

De acuerdo con una realización alternativa de la presente invención, un método para procesar señales de voz antes de codificar una señal digital que comprende datos de audio, comprende seleccionar la codificación en el dominio de la frecuencia para codificar la señal digital, cuando una tasa de bits de codificación es mayor que un límite superior de tasa de bits. Alternativamente, el método selecciona la codificación en el dominio del tiempo para codificar la señal digital, cuando la tasa de bits de codificación es menor que un límite inferior de tasa de bits. La señal digital comprende una señal de paso corto, para la cual el retardo de paso es más corto que un límite de retardo de paso.

De acuerdo con una realización alternativa de la presente invención, un método para procesar señales de voz antes de codificar comprende seleccionar la codificación en el dominio del tiempo para codificar una señal digital que comprende datos de audio cuando la señal digital no comprende una señal de paso corto y la señal digital se clasifica como voz no sonora o voz normal. El método comprende, además, seleccionar la codificación en el dominio de la frecuencia para codificar la señal digital cuando la tasa de bits de codificación está entre medias de un límite inferior de tasa de bits y un límite superior de tasa de bits. La señal digital comprende una señal de paso corto y la periodicidad de entonación es baja. El método incluye, además, seleccionar la codificación en el dominio del tiempo para codificar la señal digital cuando la tasa de bits de codificación está entre medias y la señal digital comprende una señal de paso corto y una periodicidad de entonación es muy fuerte.

De acuerdo con una realización alternativa de la presente invención, un aparato para procesar señales de voz antes de codificar una señal digital que comprende datos de audio, comprende un selector de codificación configurado para seleccionar la codificación en el dominio de la frecuencia o la codificación en el dominio del tiempo, en base a una tasa de bits de codificación a ser utilizada para codificar la señal digital y a una detección de retardo de paso corto de la señal digital.

Breve descripción de los dibujos

Para una comprensión más completa de la presente invención y de sus ventajas, ahora se hace referencia a las siguientes descripciones tomadas junto con los dibujos adjuntos, en los cuales:

la Figura1 ilustra las operaciones realizadas durante la codificación de una voz original utilizando un codificador de CELP convencional;

la Figura2 ilustra las operaciones realizadas durante la decodificación de una voz original utilizando un decodificador de CELP;

la Figura 3 ilustra un codificador de CELP convencional;

la Figura 4 ilustra un decodificador de CELP básico, correspondiente al codificador en la Figura 3;

las Figuras 5 y 6 ilustran ejemplos de señales de voz esquemáticas y su relación con el tamaño de trama y el tamaño de subtrama en el dominio del tiempo;

la Figura 7 ilustra un ejemplo de un espectro de banda ancha de voz original;

la Figura 8 ilustra un espectro sonoro codificado del espectro de sonora original ilustrado en la Figura 7 utilizando duplicación de codificación de retardo de paso;

las Figuras 9A y 9B ilustran el esquema de un códec perceptivo típico en el dominio de la frecuencia, en donde la Figura 9A ilustra un codificador en el dominio de la frecuencia, mientras que la Figura 9B ilustra un decodificador en el dominio de la frecuencia;

la Figura10 ilustra un esquema de las operaciones en un codificador antes de codificar una señal de voz que comprende datos de audio, de acuerdo con realizaciones de la presente invención;

la Figura11 ilustra un sistema10 de comunicación, de acuerdo con una realización de la presente invención;

la Figura12 ilustra un diagrama de bloques de un sistema de procesamiento que puede utilizarse para implementar los dispositivos y métodos dados a conocer en el presente documento;

la Figura13 ilustra un diagrama de bloques de un aparato para procesar señales de voz antes de codificar una señal digital; y

la Figura14 ilustra un diagrama de bloques de otro aparato para procesar señales de voz antes de codificar una señal digital.

Descripción detallada de las realizaciones ilustrativas

En el sistema moderno de comunicación de señal digital de audio/voz, una señal digital se comprime en un codificador y, la información comprimida o el flujo de bits, se puede empaquetar y enviar a un decodificador trama a trama a través de un canal de comunicación. El decodificador recibe y decodifica la información comprimida para obtener la señal digital de audio/voz.

En el sistema moderno de comunicación de señal digital de audio/voz, una señal digital se comprime en un codificador y, la información comprimida o el flujo de bits, se puede empaquetar y enviar a un decodificador trama a trama a través de un canal de comunicación. El sistema del codificador y del decodificador juntos se llama códec. La compresión de voz/audio se puede utilizar para reducir el número de bits que representan la señal de voz/audio, reduciendo así el ancho de banda y/o la tasa de bits necesaria para la transmisión. En general, una mayor tasa de bits dará como resultado una mayor calidad de audio, mientras que una menor tasa de bits dará como resultado una menor calidad de audio.

La Figura1 ilustra las operaciones realizadas durante la codificación de una voz original utilizando un codificador de CELP convencional.

La Figura1 ilustra un codificador de CELP inicial convencional, donde un error 109 ponderado entre una voz 102 sintetizada y una voz 101 original se minimiza a menudo utilizando un enfoque de análisis por síntesis, lo que

significa que la codificación (análisis) se realiza optimizando perceptivamente la señal decodificada (síntesis) en un bucle cerrado.

El principio básico que explotan todos los codificadores de voz, es el hecho de que las señales de voz son formas de onda altamente correlacionadas. Como ilustración, la voz se puede representar utilizando un modelo autorregresivo (AR) como en la Ecuación (1) a continuación.

$$X_n = \sum_{i=1}^{P} a_i X_{n-1} + e_n \qquad (1)$$

En la ecuación (11), cada una de las muestras se representa como una combinación lineal de las *P* muestras anteriores más un ruido blanco. Los coeficientes $a_1, a_2, \ldots a_P$ de ponderación, se llaman Coeficientes de Predicción Lineal (LPC). Para cada una de la tramas, los coeficientes $a_1, a_2, \ldots a_P$ de ponderación, se eligen de modo que el espectro de $\{X_1, X_2, \ldots, X_N\}$, generado utilizando el modelo anterior, coincida estrechamente con el espectro de la trama de voz de entrada.

Alternativamente, las señales de voz también pueden representarse mediante una combinación de un modelo armónico y un modelo de ruido. La parte armónica del modelo es efectivamente una representación de la serie de Fourier de la componente periódica de la señal. En general, para las señales sonoras, el modelo de voz armónica más ruido se compone de una mezcla de armónicos y ruido. La proporción de armónicos y ruido en una voz sonora depende de una serie de factores que incluyen las características del hablante (p. ej., en qué medida la voz del hablante es normal o entrecortada); el carácter del segmento de voz (p. ej., hasta qué punto un segmento de voz es periódico) y en la frecuencia. Las frecuencias más altas de la voz sonora tienen una mayor proporción de componentes similares al ruido.

El modelo de predicción lineal y el modelo de ruido armónico son los dos métodos principales para modelar y codificar las señales de voz. El modelo de predicción lineal es particularmente bueno para modelar la envoltura espectral de la voz, mientras que el modelo de ruido armónico es bueno para modelar la estructura fina de la voz. Los dos métodos pueden combinarse para aprovechar sus fortalezas relativas.

Como se indicó anteriormente, antes de la codificación de CELP, la señal de entrada al micrófono del teléfono se filtra y se muestrea, por ejemplo, a una tasa de 8000 muestras por segundo. Cada una de las muestras se cuantiza, por ejemplo, con 13 bits por muestra. La voz muestreada se segmenta en segmentos o tramas de 20 ms (p. ej., en este caso, 160 muestras).

La señal de voz se analiza y se extrae su modelo de LP, sus señales de excitación y su paso. El modelo de LP representa la envoltura espectral de la voz. Se convierte en un conjunto de coeficientes de frecuencias espectrales de línea (LSF), que es una representación alternativa de los parámetros de predicción lineal, porque los coeficientes de LSF tienen buenas propiedades de cuantización. Los coeficientes de LSF pueden cuantizarse de forma escalar o, de manera más eficiente, pueden cuantizarse por vector utilizando libros de códigos de vector de LSF previamente entrenados.

El código de excitación incluye un libro de códigos que comprende vectores de códigos, que tienen componentes que se eligen todas de forma independiente, de modo que cada uno de los vectores de códigos puede tener un espectro aproximadamente “blanco”. Para cada una de las subtramas de voz de entrada, cada uno de los vectores de códigos se filtra a través del filtro 103 de predicción lineal a corto plazo y el filtro 105 de predicción a largo plazo, y la salida se compara con las muestras de voz. En cada una de las subtramas, el vector de códigos cuya salida coincide mejor con la voz de entrada (error minimizado) se elige para representar esa subtrama.

La excitación 108 codificada comprende, normalmente, una señal similar a pulso o una señal similar a ruido, que se construyen matemáticamente o se guardan en un libro de códigos. El libro de códigos está disponible tanto para el codificador como para el decodificador de recepción. La excitación 108 codificada, que puede ser un libro de códigos estocástico o fijo, puede ser un diccionario de cuantización vectorial que está codificado fijo (implícita o explícitamente) en el códec. Un libro de códigos fijo de este tipo puede ser una predicción lineal algebraica excitada por código o almacenarse explícitamente.

Un vector de códigos del libro de códigos se escala mediante una ganancia apropiada para hacer que la energía sea igual a la energía de la voz de entrada. En consecuencia, la salida de la excitación 108 codificada se escala por una ganancia $G_c$ 107 antes de pasar a través de los filtros lineales.

El filtro 103 de predicción lineal a corto plazo da forma al espectro 'blanco' del vector de códigos para parecerse al espectro de la voz de entrada. De manera equivalente, en el dominio del tiempo, el filtro 103 de predicción lineal a corto plazo incorpora correlaciones a corto plazo (correlación con muestras anteriores) en la secuencia blanca. El filtro que da forma a la excitación tiene un modelo de todos los polos de la forma 1 / A(z) (filtro 103 de predicción lineal a corto plazo), donde A(z) se denomina el filtro de predicción y se puede obtener utilizando la predicción lineal (p. ej., Algoritmo de Levinson-Durbin). En una o más realizaciones, se puede utilizar un filtro de todos los polos porque es una buena representación del tracto vocal humano y porque es fácil de calcular.

El filtro 103 de predicción lineal a corto plazo se obtiene al analizar la señal 101 original y se representa mediante un conjunto de coeficientes:

$$A(z) = \sum_{i-1}^{P} 1 + a_i \cdot z^{i} \quad , i = 1,2,....,P \qquad (2)$$

Como se describió anteriormente, las regiones de la voz sonora exhiben una periodicidad a largo plazo. Este período, conocido como paso, se introduce en el espectro sintetizado por el filtro de paso 1 / (B(z)). La salida del filtro 105 de predicción a largo plazo depende del paso y de la ganancia de paso. En una o más realizaciones, el paso puede estimarse a partir de la señal original, la señal residual o la señal original ponderada. En una realización, la función (*B(z)*) de predicción a largo plazo se puede expresar utilizando la Ecuación (3) de la siguiente manera.

$$B(z) = 1 - G_p \cdot z^{Paso} \qquad (3)$$

El filtro 110 de ponderación está relacionado con el filtro de predicción a corto plazo anterior. Uno de los filtros de ponderación típicos se puede representar como se describe en la Ecuación (4).

$$W(z) = \frac{A(z/\alpha)}{1 - \beta \cdot z^{-1}} \qquad (4)$$

donde $\beta < \alpha$, $0 < \beta < 1$, $0 < \alpha \leq 1$.

En otra realización, el filtro *W(z)* de ponderación puede derivarse a partir del filtro de LPC utilizando la expansión del ancho de banda, como se ilustra en una realización, en la Ecuación (5) a continuación.

$$W(z) = \frac{A(z/\gamma 1)}{A(z/\gamma 2)} \qquad (5),$$

En la ecuación (5), γ1 > γ2, son los factores con los cuales los polos se mueven hacia el origen.

En consecuencia, para cada una de las tramas de voz, se calculan los LPC y el paso, y se actualizan los filtros. Para todas las subtramas de voz, el vector de códigos que produce la "mejor" salida filtrada se elige para representar la subtrama. El valor de ganancia cuantizado correspondiente debe transmitirse al decodificador para la decodificación adecuada. Los LPC y los valores de paso también deben cuantizarse y enviarse cada una de las tramas para reconstruir los filtros en el decodificador. Por consiguiente, el índice de excitación codificado, el índice de ganancia cuantizado, el índice de parámetro de predicción a largo plazo cuantizado y el índice de parámetro de predicción a corto plazo cuantizado se transmiten al decodificador.

La Figura2 ilustra las operaciones realizadas durante la decodificación de una voz original utilizando un decodificador de CELP.

La señal de voz se reconstruye en el decodificador pasando los vectores de códigos recibidos a través de los filtros correspondientes. En consecuencia, todos los bloques, excepto el postprocesamiento, tiene la misma definición que se describe en el codificador de la Figura1.

El flujo de bits de CELP codificado se recibe y se desempaqueta 80 en un dispositivo de recepción. Para cada una de las subtramas recibida, el índice de excitación codificado, el índice de ganancia cuantizado, el índice de parámetro de predicción a largo plazo cuantizado y el índice de parámetro de predicción a corto plazo cuantizado recibidos, se utilizan para encontrar los parámetros correspondientes utilizando los decodificadores correspondientes, por ejemplo, el decodificador 81 de ganancia, el decodificador 82 de predicción a largo plazo y el decodificador 83 de predicción a corto plazo. Por ejemplo, las posiciones y los signos de amplitud de los pulsos de

excitación y del vector de códigos algebraico del código 402 de excitación, pueden determinarse a partir del índice de excitación codificado recibido.

Haciendo referencia a la Figura2, el decodificador es una combinación de varios bloques que incluye excitación 201 codificada, predicción 203 a largo plazo, predicción 205 a corto plazo. El decodificador inicial incluye, además, el bloque 207 de postprocesamiento después de una voz 206 sintetizada. El postprocesamiento puede comprender, además, el postprocesamiento a corto plazo y el postprocesamiento a largo plazo.

La Figura 3 ilustra un codificador de CELP convencional.

La Figura 3 ilustra un codificador de CELP básico que utiliza un libro de códigos adaptativo adicional para mejorar la predicción lineal a largo plazo. La excitación se produce sumando las contribuciones de un libro 307 de códigos adaptativo y un código 308 de excitación, que puede ser un libro de códigos estocástico o fijo, como se describió anteriormente. Las entradas en el libro de códigos adaptativo comprenden versiones retardadas de la excitación. Esto hace posible la codificación eficiente de señales periódicas, tal como los sonidos sonoros.

Haciendo referencia a la Figura 3, un libro 307 de códigos adaptativo comprende una excitación 304 sintetizada anterior o repetir el ciclo de paso de excitación anterior en el período de paso. El retardo de paso se puede codificar en valor entero cuando es grande o largo. El retardo de paso a menudo se codifica en un valor fraccional más preciso cuando es pequeño o corto. La información periódica del paso se emplea para generar la componente adaptativa de la excitación. Luego, esta componente de excitación se escala mediante una ganancia $G_p$ 305 (también llamada ganancia de paso).

La Predicción a Largo Plazo juega un papel muy importante para la codificación de voz sonora, porque la voz sonora tiene fuerte periodicidad. Los ciclos de paso adyacentes de la voz sonora son similares entre sí, lo que significa matemáticamente que la ganancia de paso $G_p$ en la siguiente expresión de excitación es alta o cercana a 1. La excitación resultante se puede expresar como en la Ecuación (6), como combinación de las excitaciones individuales.

$$e(n) = G_p \cdot e_p(n) + G_c \cdot e_c(n) \qquad (6)$$

donde, $e_p(n)$ es una subtrama de series de muestras indexadas por $n$, procedentes del libro 307 de códigos adaptativo que comprende la excitación 304 anterior a través del bucle de retroalimentación (Figura 3). $e_p(n)$ puede filtrarse en paso bajo adaptativamente, ya que el área de baja frecuencia es a menudo más periódica o más armónica que el área de frecuencia alta. $e_c(n)$ es del libro 308 de códigos de excitación codificado (también llamado libro de códigos fijo), que es una contribución de excitación actual. Además, $e_c(n)$ también puede mejorarse, tal como utilizando la mejora de filtrado de paso alto, la mejora de paso, la mejora de dispersión, la mejora de formantes y otras.

Para voz sonora, la contribución de $e_p(n)$ del libro 307 de códigos adaptativo puede ser dominante y la ganancia $G_p$ 305 de paso es aproximadamente un valor de 1. La excitación generalmente se actualiza para cada una de las subtramas. El tamaño de trama típico es de 20 milisegundos y el tamaño de subtrama típico es de 5 milisegundos.

Como se describe en la Figura1, la excitación 308 codificada fija se escala por una ganancia $G_c$ 306 antes de pasar a través de los filtros lineales. Las dos componentes de excitación escaladas de la excitación 108 codificada fija y el libro 307 de códigos adaptativo se suman antes de filtrarse a través del filtro 303 de predicción lineal a corto plazo. Las dos ganancias ($G_p$ y $G_c$) se cuantizan y se transmiten a un decodificador. Por consiguiente, el índice de excitación codificado, el índice de libro de códigos adaptativo, los índices de ganancia cuantizados y el índice de parámetros de predicción a corto plazo cuantizado se transmiten al dispositivo de audio receptor.

El flujo de bits de CELP codificado utilizando un dispositivo ilustrado en la Figura 3, se recibe en un dispositivo de recepción. La Figura 4 ilustra el decodificador correspondiente del dispositivo de recepción.

La Figura 4 ilustra un decodificador de CELP básico correspondiente al codificador de la Figura 3. La Figura 4 incluye un bloque 408 de postprocesamiento que recibe la voz 407 sintetizada del decodificador principal. Este decodificador es similar a la Figura 3, excepto el libro 307 de códigos adaptativo.

Para cada una de las subtramas recibida, el índice de excitación codificado, el índice de ganancia de excitación codificado cuantizado, el índice de paso cuantizado, el índice de ganancia del libro de códigos adaptativo cuantizado y el índice de parámetro de predicción a corto plazo cuantizado recibidos, se utilizan para encontrar los parámetros correspondientes utilizando los decodificadores correspondientes, por ejemplo, el decodificador 81 de ganancia, el

decodificador 84 de paso, el decodificador 85 de ganancia de libro de códigos adaptativo y decodificador 83 de predicción a corto plazo.

En diversas realizaciones, el decodificador de CELP es una combinación de varios bloques y comprende la excitación 402 codificada, el libro 401 de códigos adaptativo, la predicción 406 a corto plazo y el postprocesamiento 408. Todos los bloques, excepto el postprocesamiento, tienen la misma definición que la descrita en el codificador de la Figura 3. El postprocesamiento puede incluir, además, el postprocesamiento a corto plazo y el postprocesamiento a largo plazo.

El bloque de excitación de código (referenciado con la etiqueta 308 en la Figura 3 y 402 en la Figura 4) ilustra la ubicación del Libro de Códigos Fijo (FCB) para una codificación de CELP general. Un vector de códigos seleccionado del FCB se escala por una ganancia, a menudo indicada como $G_c$ 306.

Las Figuras 5 y 6 ilustran ejemplos de señales de voz esquemáticas y su relación con el tamaño de trama y el tamaño de subtrama en el dominio del tiempo. Las Figuras 5 y 6 ilustran una trama que incluye una pluralidad de subtramas.

Las muestras de la voz de entrada se dividen en bloques de muestras, cada uno, llamado trama, p. ej., 80-240 muestras o tramas. Cada una de las tramas se divide en bloques más pequeños de muestras, cada uno, llamado subtrama. En la tasa de muestreo de 8 kHz, 12,8 kHz o 16 kHz, el algoritmo de codificación de voz es tal que la duración nominal de la trama está en el rango de diez a treinta milisegundos y, típicamente, veinte milisegundos. En la Figura 5 ilustrada, la trama tiene un tamaño 1 de trama y un tamaño 2 de subtrama, en la que cada una de las tramas se divide en 4 subtramas.

Haciendo referencia a las partes más bajas o inferiores de las Figuras 5 y 6, las regiones sonoras en una voz parecen una señal casi periódica en la representación en el dominio del tiempo. La apertura y el cierre periódicos de las cuerdas vocales del hablante dan como resultado la estructura armónica en las señales de voz sonora. Por lo tanto, durante cortos períodos de tiempo, los segmentos de voz sonora pueden tratarse para ser periódicos para todo el análisis y procesamiento prácticos. La periodicidad asociada con dichos segmentos se define como “Período de Paso” o simplemente “paso” en el dominio del tiempo y “Frecuencia de paso o Frecuencia fundamental $f_0$” en el dominio de la frecuencia. La inversa del período de paso es la frecuencia fundamental de la voz. Los términos paso y frecuencia fundamental de la voz se utilizan frecuentemente de manera intercambiable.

Para la mayoría de las voces sonoras, una trama contiene más de dos ciclos de paso. La Figura 5 ilustra, además, un ejemplo que el período 3 de paso es más pequeño que el tamaño 2 de subtrama. En contraste, la Figura 6 ilustra un ejemplo en el que el período 4 de paso es mayor que el tamaño 2 de subtrama y menor que el tamaño de media trama.

Para codificar la señal de voz de manera más eficiente, la señal de voz se puede clasificar en diferentes clases y cada una de las clases se codifica de una manera diferente. Por ejemplo, en algunos estándares, tales como G.718, VMR-WB o AMR-WB, la señal de voz se clasifica en NO SONORA, TRANSICIÓN, GENÉRICA, SONORA y RUIDO.

Para cada una de las clases, el filtro de LPC o de STP siempre se utiliza para representar la envolvente espectral. Sin embargo, la excitación al filtro de LPC puede ser diferente. Las clases de NO SONORA y de RUIDO pueden codificarse con una excitación de ruido y alguna mejora de excitación. La clase de TRANSICIÓN se puede codificar con una excitación de pulso y alguna mejora de excitación sin utilizar un libro de códigos adaptativo o LTP.

La GENÉRICA puede codificarse con un enfoque de CELP tradicional, tal como el CELP algebraico utilizado en G.729 o en AMR-WB, en el que una trama de 20 ms contiene cuatro subtramas de 5 ms. Tanto la componente de excitación del libro de códigos adaptativo como la componente de excitación del libro de códigos fijo se producen con alguna mejora de excitación para cada una de las subtramas. Los retardos de paso para el libro de códigos adaptativo en la primera y la tercera subtramas se codifican en un rango completo de un límite *PIT_MIN* de paso mínimo hasta un límite *PIT_MAX* de paso máximo. Los retardos de paso para el libro de códigos adaptativo en las subtramas segunda y cuarta se codifican de manera diferente del retardo de paso codificado anterior.

Las clases SONORA pueden codificarse de tal manera que sean ligeramente diferentes de la clase GENÉRICA. Por ejemplo, el retardo de paso en la primera subtrama puede codificarse en un rango completo desde un límite *PIT_MIN* de paso mínimo hasta un límite *PIT_MAX* de paso máximo. Los retardos de paso en las otras subtramas pueden codificarse de manera diferente del retardo de paso codificado anterior. Como ilustración, suponiendo que la tasa de muestreo de excitación es 12,8 kHz, entonces el valor de *PIT_MIN* de ejemplo puede ser 34 y *PIT_MAX* puede ser 231.

Ahora se describirán realizaciones de la presente invención para mejorar la clasificación de la codificación en el dominio del tiempo y de la codificación en el dominio de la frecuencia.

En términos generales, es mejor utilizar la codificación en el dominio del tiempo para la señal de voz y la codificación en el dominio de la frecuencia para la señal de música con el fin de lograr la mejor calidad a una tasa de bits bastante alta (p. ej., 24 kbps <= tasa de bits <= 64 kbps). Sin embargo, para algunas señales de voz específicas, tal como la señal de paso corto, la señal de voz cantada o la señal de voz muy ruidosa, puede ser mejor utilizar la codificación en el dominio de la frecuencia. Para algunas señales de música específicas, tal como una señal muy periódica, puede ser mejor utilizar la codificación en el dominio del tiempo al beneficiarse de una ganancia de LTP muy alta. La tasa de bits es un parámetro importante para la clasificación. Generalmente, la codificación en el dominio del tiempo favorece la tasa de bits baja y la codificación en el dominio de la frecuencia favorece la tasa de bits alta. Debe decidirse cuidadosamente una mejor clasificación o selección entre la codificación en el dominio del tiempo y la codificación en el dominio de la frecuencia, considerando también el rango de tasa de bits y la característica de los algoritmos de codificación.

En las siguientes secciones, se describirá la detección de la señal de voz normal y de paso corto.

La voz normal es una señal de voz que excluye la señal de voz cantada, la señal de voz de paso corto o la señal mixta de voz/música. La voz normal también puede ser una señal de voz que cambia rápidamente, cuyo espectro y/o energía cambia más rápido que la mayoría de las señales de música. Normalmente, el algoritmo de codificación en el dominio del tiempo es mejor que el algoritmo de codificación en el dominio de la frecuencia para codificar la señal de voz normal. Lo siguiente es un ejemplo de algoritmo para detectar la señal de voz normal.

Para un candidato *P* de paso, la correlación de paso normalizada a menudo se define de forma matemática como en la Ecuación (8).

$$R(P) = \frac{\sum_n s_w(n) \cdot s_w(n-P)}{\sqrt{\sum_n \|s_w(n)\|^2 \cdot \sum_n \|s_w(n-P)\|^2}} \qquad (8)$$

En la ecuación (8), $s_w(n)$ es una señal de voz ponderada, el numerador es la correlación y el denominador es un factor de normalización de energía. Supongamos que *Entonación* señala el valor promedio de correlación del paso normalizado de las cuatro subtramas en la trama de voz actual, la Entonación se puede calcular como en la Ecuación (9) a continuación.

$$Entonación = [\, R_1(P_1) + R_2(P_2) + R_3(P_3) + R_4(P_4) \,] / 4 \qquad (9)$$

$R_1(P_1)$, $R_2(P_2)$, $R_3(P_3)$ y $R_4(P_4)$ son las cuatro correlaciones de paso normalizadas calculadas para cada una de las subtramas; $P_1$, $P_2$, $P_3$ y $P_4$ para cada una de las subtramas son los mejores candidatos de paso que se encuentran en el rango de paso de *P = PIT_MIN* a *P = PIT_MAX*. La correlación de paso suavizada desde la trama anterior a la trama actual se puede calcular como en la Ecuación (10).

$$\begin{aligned} &si\ (\ (Entonación > Entonación\_sm)\ y\ (clase\_voz \neq NO\ SONORA)\ ) \\ &\quad Entonación\_sm \Leftarrow (3 \cdot Entonación\_sm + Entonación) / 4 \\ &sino\ si\ (VAD = 1) \\ &\quad Entonación\_sm \Leftarrow (31 \cdot Entonación\_sm + Entonación) / 32 \end{aligned} \qquad (10)$$

En la ecuación (10), VAD es Detección de Actividad de Voz y VAD = 1 referencia que la señal de voz sale. Supongamos que $F_s$ es la frecuencia de muestreo, la energía máxima en la región [*0, $F_{MIN} = F_s / PIT\_MIN$*] (Hz) de muy baja frecuencia es *Energía0* (dB), la energía máxima en la región [$F_{MIN}$, 900] (Hz) de baja frecuencia es *Energía1* (dB), y la energía máxima en la región [*5000*, 5800] (Hz) de frecuencia alta es *Energía3* (dB), un parámetro *Inclinación* de inclinación espectral se define de la siguiente manera.

$$Inclinación = energía3 - max\{energía0, energía1\} \qquad (11)$$

Un parámetro de inclinación espectral suavizada se señala como en la ecuación (12).

$$Inclinación\_sm \Leftarrow (7 \cdot Inclinación\_sm + Inclinación) / 8 \qquad (12)$$

Una diferencia de inclinación espectral de la trama actual y de la trama anterior se puede dar como en la ecuación (13).

$$Diff\_inclinación = \left| inclinación - inclinación\_antigua \right| \qquad (13)$$

Una diferencia inclinación espectral suavizada se da como en la ecuación (14).

$$\begin{aligned} &si\ \ ( (Diff\_inclinación > Diff\_inclinación\_sm)\ y\ (clase\_voz \neq NO\ SONORA) ) \\ &\quad Diff\_inclinación\_sm \Leftarrow (3 \cdot Diff\_inclinación\_sm + Diff\_inclinación) / 4 \\ &sino\ si\ \ (VAD = 1) \\ &\quad Diff\_inclinación\_sm \Leftarrow (31 \cdot Diff\_inclinación\_sm + Diff\_inclinación) / 32 \end{aligned} \qquad (14)$$

Una diferencia de energía de baja frecuencia de la trama actual y de la trama anterior es

$$Diff\_energía1 = \left| energía1 - energía1\_antigua \right| \qquad (15)$$

Un diferencia energía suavizada viene dada por la ecuación (16).

$$\begin{aligned} &si\ \ ( (Diff\_energía1 > Diff\_energía1\_sm)\ y\ (clase\_voz \neq NO\ SONORA) ) \\ &\quad Diff\_energía\mathbf{1}\_sm \Leftarrow ( 3 \cdot Diff\_energía1\_sm + Diff\_energía1) / 4 \\ &sino\ si\ \ (VAD = \mathbf{1}) \\ &\quad Diff\_energía\mathbf{1}\_sm \Leftarrow (31 \cdot Diff\_energía1\_sm + Diff\_energía1) / 32 \end{aligned} \qquad (16)$$

Además, una bandera voz normal denotada como *Bandera_voz* se decide y se cambia durante el área sonora al considerar la variación *Diff_energía1_sm* de energía, variación *Entonación_sm* de entonación y la variación *Diff_entonación_sm* de inclinación espectral, como se proporcionan en la ecuación (17).

$$\begin{aligned} &si\ \ (clase\_voz \neq NO\ SONORA)\ \ \{ \\ &\quad Diff\_Sp = Diff\_energía1\_sm \cdot Entonación\_sm \cdot Diff\_entonación\_sm \\ &\quad si\ \ (Diff\_Sp > 800)\ Bandera\_voz = 1\ \ //cambiar\ a\ voz\ normal \\ &\quad si\ \ (Diff\_Sp < 100)\ Bandera\_voz = 0\ \ //cambiar\ a\ voz\ no\ normal \\ &\} \end{aligned} \qquad (17)$$

Se describirán realizaciones de la presente invención para detectar una señal de paso corto.

La mayoría de los códecs de CELP funcionan bien para las señales de voz normales. Sin embargo, los códecs de CELP de tasa de bits baja a menudo fallan para señales de música y/o señales de voz cantadas. Si el rango de codificación de paso es de *PIT_MIN* a *PIT_MAX* y el retardo de paso real es más pequeño que *PIT_MIN*, el rendimiento de la codificación de CELP puede ser perceptivamente malo debido al doble paso o triple paso. Por ejemplo, el rango de paso de *PIT_MIN = 34* a *PIT_MAX* = 231 para la frecuencia de muestreo $F_s$ = 12,8 kHz, se adapta a la mayoría de las voces humanas. Sin embargo, el retardo de paso real de la señal de música regular o de voz cantada puede ser mucho más corto que el límite mínimo *PIT_MIN = 34* definido en el algoritmo de CELP de ejemplo anterior.

Cuando el retardo de paso real es *P*, la correspondiente frecuencia fundamental normalizada (o primer armónico) es $f_0 = F_s / P$, donde $F_s$ es la frecuencia de muestreo y $f_0$ es la ubicación del primer pico de armónicos en el espectro. Entonces, para una frecuencia de muestreo dada, la limitación de paso mínima *PIT_MIN* define realmente la limitación $F_M = F_s / PIT\_MIN$ de frecuencia armónica fundamental máxima para el algoritmo de CELP.

La Figura 7 ilustra un ejemplo de un espectro sonoro original. La Figura 8 ilustra un espectro sonoro codificado del espectro de banda ancha de voz original ilustrado en la Figura 7 utilizando la duplicación de codificación de retardo de paso. En otras palabras, la Figura 7 ilustra un espectro antes de la codificación y la Figura 8 ilustra el espectro después de la codificación.

En el ejemplo mostrado en la Figura 7, el espectro está formado por los picos 701 de armónicos y la envolvente 702 espectral. La frecuencia armónica fundamental real (la ubicación del primer pico armónico) ya está más allá de la limitación $F_M$ de frecuencia armónica fundamental máxima, de modo que el retardo de paso transmitido para el algoritmo de CELP no puede ser igual al retardo de paso real y podría ser el doble o múltiplo del retardo de paso real.

El retardo de paso mal transmitido con múltiplo del retardo de paso real puede causar degradación de la calidad obvia. En otras palabras, cuando el retardo de paso real para la señal de música armónica o la señal de voz cantada es menor que la limitación *PIT_MIN* de retardo mínimo definido en el algoritmo de CELP, el retardo transmitido podría ser doble, triple o múltiple del retardo de paso real.

Como resultado, el espectro de la señal codificada con el retardo de paso transmitido puede ser como se muestra en la Figura 8. Como se ilustra en la Figura 8, además de incluir los picos 8011 armónicos y la envolvente 802 espectral, se pueden ver picos 803 pequeños no deseados entre los picos armónicos reales, mientras que el espectro correcto debería ser como el de la Figura 7. Esos picos de espectro pequeños en la Figura 8 podrían causar una distorsión perceptiva incómoda.

De acuerdo con realizaciones de la presente invención, una solución para resolver este problema cuando el CELP falla para algunas señales específicas, es que se utilice una codificación en el dominio de la frecuencia en lugar de codificación en el dominio del tiempo.

Por lo general, las señales armónicas de música o señales de voz cantada son más estacionarias que las señales de voz normal. El retardo de paso (o la frecuencia fundamental) de la señal de voz normal cambia constantemente. Sin embargo, el retardo de paso (o la frecuencia fundamental) de la señal de música o la señal de voz cantada, a menudo, mantiene un cambio relativamente lento durante bastante tiempo. El rango de paso muy corto se define de *PIT_MIN0* a *PIT_MIN*. En la frecuencia de muestreo *Fs* = 12,8 kHz, una definición de ejemplo del rango de paso muy corto puede ser de *PIT_MIN0* <= 17 a *PIT_MIN = 34*. Como el candidato de paso es tan corto, la energía de 0 Hz a $F_{MIN} = Fs / PIT\_MIN$ Hz debe ser suficientemente baja. Otras condiciones, como la Detección de Actividad de Voz y la Clasificación de Voz, pueden agregarse durante la detección de la existencia de la señal de paso corto.

Los dos parámetros siguientes pueden ayudar a detectar la posible existencia de la señal de paso muy corto. Uno presenta "*Ausencia de Energía de Muy Baja Frecuencia*" y otro presenta "*Nitidez Espectral*". Como ya se mencionó anteriormente, suponga que la energía máxima en la región [*0,* $F_{MIN}$] (Hz) de frecuencia es *Energía0* (dB), la energía máxima en la región [$F_{MIN}$, 900] (Hz) de frecuencia es *Energía1* (dB), la relación de energía relativa entre *Energía0* y *Energía1* se proporciona en la Ecuación (18) a continuación.

$$Relación = Energía1 - Energía0 \qquad (18)$$

Esta relación de energía puede ponderarse multiplicando un valor *Entonación* de correlación de paso normalizado promedio, que se muestra a continuación en la Ecuación (19).

$$Relación \Leftarrow Relación \cdot \max\{Entonación,\ 0{,}5\} \qquad (19)$$

La razón para hacer la ponderación de la Ecuación (19) utilizando un factor de *Entonación* es que la detección de paso corto es significativa para la voz sonora o música armónica, y no es significativa para voz no sonora o música no armónica. Antes de utilizar el parámetro *Relación* para detectar la ausencia de energía de baja frecuencia, es mejor suavizarlo para reducir la incertidumbre como en la Ecuación (20).

$$\begin{aligned} &si\ \ (VAD = 1)\ \ \{ \\ &\quad LF\_RelaciónEnergía\_sm \Leftarrow (15 \cdot LF\_RelaciónEnergía\_sm + Relación) / 16 \\ &\} \end{aligned} \qquad (20)$$

Si *LF_ausencia_bandera= 1* significa que se detecta la ausencia de energía de baja frecuencia (de lo contrario *LF_ausencia_bandera= 0*), *LF_ausencia_bandera* se puede determinar mediante el siguiente procedimiento.

```
si   (LF_RelaciónEnergía_sm > 30) o (Relación > 48) o
     (LF_RelaciónEnergía_sm > 22 y Relación > 38)   {
       LF_ausencia_bandera = 1;
}
sino si   (LF_RelaciónEnergía_sm < 13)   {
       LF_ausencia_bandera = 0;
}
sino   {
       LF_ausencia_bandera se mantiene sin cambios
}
```

Los parámetros relacionados de *Nitidez Espectral* se determinan de la siguiente manera. Supongamos que *Energía1* (dB) es la energía máxima en la región [$F_{MIN}$, 900] (Hz) de baja frecuencia, *i_pico* es la ubicación del pico armónico de energía máxima en la región [$F_{MIN}$, 900] (Hz) de frecuencia y *Energía2* (dB) es la energía promedio en la región [*i_pico*, *i_pico* + 400] (*Hz*) de frecuencia. Un parámetro de nitidez espectral se define como en la ecuación (21).

$$NitidezEspec = max\{Energía1 - Energía2, 0\} \quad (21)$$

Un parámetro de nitidez espectral suavizada se da como sigue.

```
si   (VAD = 1)   {
       NitidezEspec_sm = (7 · NitidezEspec_sm + NitidezEspec) / 8
}
```

Una bandera nitidez espectral que indica la posible existencia de señal de paso corto se evalúa mediante lo siguiente.

```
si   ( NitidezEspec_sm > 50 o NitidezEspec > 80)   {
       NitidezEspec_bandera = 1;  //posible paso corto o tonos
}
si   (NitidezEspec_sm < 8)   {
       NitidezEspec_bandera = 0;
}
si no se cumple alguna de las condiciones anteriores, NitidezEspec_bandera se mantiene sin cambios.
```

En diversas realizaciones, los parámetros estimados anteriormente pueden utilizarse para mejorar la clasificación o selección de la codificación en el dominio del tiempo y de la codificación en el dominio de la frecuencia. Supongamos que *Sp_Aud_Deci = 1* denota que se selecciona la codificación en el dominio de la frecuencia y *Sp_Aud_Deci = 0* indica que se selecciona la codificación en el dominio del tiempo. El siguiente procedimiento proporciona un ejemplo de algoritmo para mejorar la clasificación de la codificación en el dominio del tiempo y de la codificación en el dominio de la frecuencia para diferentes tasas de bits de codificación.

Las realizaciones de la presente invención se pueden utilizar para mejorar tasas de bits altas, por ejemplo, la tasa de bits de codificación es mayor o igual que 46200 bps. Cuando la tasa de bits de codificación es muy alta y posiblemente exista una señal de paso corto, se selecciona la codificación en el dominio de la frecuencia, porque la codificación en el dominio de la frecuencia puede ofrecer una calidad robusta y confiable, mientras que la codificación en el dominio del tiempo corre el riesgo influirse negativamente por la detección de paso incorrecto. En contraste, cuando la señal de paso corto no existe y la señal es de voz no sonora o de voz normal, se selecciona la codificación en el dominio del tiempo, porque la codificación en el dominio del tiempo puede ofrecer mejor calidad que la codificación en el dominio de la frecuencia para la señal de voz normal.

```
/* para posible señal de paso corto, seleccionar codificación en el dominio de la frecuencia */
si (LF_ausencia_bandera = 1 o NitidezEspec_bandera = 1)  {
      SP_Aud_Deci = 1;  //seleccionar codificación en el dominio de la frecuencia
}

/*para voz no sonora o voz normal, seleccionar codificación en el dominio del tiempo */
si (LF_ausencia_bandera = 0 y NitidezEspec_bandera = 0) {
      si  ( (Inclinación > 40) y (Entonación < 0,5) y (clase_voz = NO SONORA) y
          (VAD = 1) ) {
                Sp_Aud_Deci = 0; //seleccionar codificación en el dominio del tiempo
      }
       si (Voz_bandera = 1) {
            SP_Aud_Deci = 0;  //seleccionar codificación en el dominio del tiempo
      }
}
```

Las realizaciones de la presente invención se pueden utilizar para mejorar la codificación de la tasa de bits intermedia, por ejemplo, cuando la tasa de bits de codificación está entre 24,4 kbps y 46200 bps. Cuando posiblemente existe una señal de paso corto y la periodicidad de entonación es baja, se selecciona la codificación en el dominio de la frecuencia, porque la codificación en el dominio de la frecuencia puede ofrecer una calidad robusta y confiable, mientras que la codificación en el dominio del tiempo corre el riesgo de influirse negativamente por la baja periodicidad de entonación. Cuando no existe una señal de paso corto y la señal es de voz no sonora o de voz normal, se selecciona la codificación en el dominio del tiempo, porque la codificación en el dominio del tiempo puede ofrecer una mejor calidad que la codificación en el dominio de la frecuencia para la señal de voz normal. Cuando la periodicidad de entonación es muy fuerte, se selecciona la codificación en el dominio del tiempo, porque la codificación en el dominio del tiempo puede beneficiarse mucho de la alta ganancia de LTP con periodicidad de entonación fuerte.

Las realizaciones de la presente invención también se pueden utilizar para mejorar tasas de bits altas, por ejemplo, la tasa de bits de codificación es menor que 24,4 kbps. Cuando existe una señal de paso corto y la periodicidad de entonación no es baja con detección de retardo de paso corto correcta, no se selecciona la codificación en el dominio de la frecuencia, porque la codificación en el dominio de la frecuencia no puede ofrecer calidad robusta y confiable en tasa baja, mientras que la codificación en el dominio del tiempo puede beneficiarse de la función de LTP.

El siguiente algoritmo ilustra una realización específica de las realizaciones anteriores, como una ilustración. Todos los parámetros se pueden calcular como se describió anteriormente en una o más realizaciones.

```
/*   preparar parámetros o umbrales  */
      si (trama anterior es de codificación en el dominio del tiempo) {
            DPIT = 0,4;
            TH1 = 0,92;
            TH2 = 0,8;
      }
```

```
    sino {
        DPIT = 0,9;
        TH1 = 0,9;
        TH2 = 0,7;
    }
Stab_Paso_Bandera = (|P0 - P1| < DPIT) y (|P1 - P2| < DPIT) y (|P2 - P3| < DPIT);
Entonación_Alta = (Entonación_sm > TH1) y (Entonación > TH2);


/* para posible señal de paso corto con periodicidad baja (entonación baja) seleccionar
codificación en el dominio de la frecuencia */

si  ( (LF_ausencia_bandera = 1) o (NitidezEspec_bandera = 1) ) {
    si  (((Stab_Paso_Bandera = 0 o Entonación_Alta = 0) y (Inclinación_sm <= -50))
        o (Inclinación_sm <= -60) )
    {
        Sp_Aud_Deci = 1;  //seleccionar codificación en el dominio de la frecuencia
    }
}

/* para señal de voz no sonora o señal de voz normal, seleccionar codificación en el
dominio del tiempo */

si  ( LF_ausencia_bandera = 0 y NitidezEspec_bandera = 0)
{
    si (Inclinación > 40 y Entonación < 0,5 y clase_voz = NO SONORA y Vad = 1)
    {

        Sp_Aud_Deci = 0;  //seleccionar codificación en el dominio del tiempo
    }
    si (Voz_bandera = 1)
    {
        Sp_Aud_Deci = 0;  //seleccionar codificación en el dominio del tiempo
    }

/* para señal de entonación fuerte, seleccionar codificación en el dominio del tiempo */


si ( Inclinación_sm > -60 y (clase_voz no es NO SONORA) )
{
    si ( Entonación_Alta = 1 y
        (Stab_Paso_Bandera = 1 o (LF_ausencia_bandera = 0 y NitidezEspec_bandera = 0) ) )

    {
        Sp_Aud_Deci = 0;  //seleccionar codificación en el dominio del tiempo
    }
}
```

En diversas realizaciones, la clasificación o selección de la codificación en el dominio del tiempo y de la codificación en el dominio de la frecuencia, puede utilizarse para mejorar significativamente la calidad perceptiva de algunas señales de voz específicas o señal de música.

La codificación de audio basada en la tecnología de banco de filtros, se utiliza ampliamente en la codificación en el dominio de la frecuencia. En el procesamiento de señales, un banco de filtros es un conjunto de filtros de paso de banda que separa la señal de entrada en múltiples componentes, cada una de las cuales transporta una subbanda de frecuencia única de la señal de entrada original. El proceso de descomposición realizado por el banco de filtros se denomina análisis y la salida del análisis del banco de filtros se denomina una señal de subbanda que tiene tantas subbandas como filtros en el banco de filtros. El proceso de reconstrucción se denomina síntesis de banco de filtros. En el procesamiento de señales digitales, el término banco de filtros también se aplica comúnmente a un banco de receptores, que también puede convertir descendentemente las subbandas a una frecuencia central baja que se puede volver a muestrear a una tasa reducida. El mismo resultado sintetizado a veces también se puede lograr al submuestrear las subbandas de paso de banda. La salida del análisis del banco de filtros puede estar en una forma de coeficientes complejos. Cada uno de los coeficientes complejos tiene un elemento real y un elemento imaginario que representan, respectivamente, un término de coseno y un término de seno para cada una de las subbandas del banco de filtros.

El Análisis del Banco de Filtros y la Síntesis del Banco de Filtros es un tipo de par de transformación que transforma una señal en el dominio del tiempo en coeficientes en el dominio de la frecuencia y transforma inversamente los coeficientes en el dominio de la frecuencia de nuevo en una señal en el dominio del tiempo. También se pueden utilizar otros pares de transformación populares, tales como (*FFT* e *iFFT*), (*DFT* e *iDFT*) y (*MDCT* e *iMDCT*), en la codificación de voz/audio.

En la aplicación de bancos de filtros para la compresión de señales, algunas frecuencias son perceptivamente más importantes que otras. Después de la descomposición, las frecuencias perceptivamente significativas pueden codificarse con una resolución fina, ya que las pequeñas diferencias en estas frecuencias son perceptivamente perceptibles para justificar utilizar un esquema de codificación que conserve estas diferencias. Por otro lado, las frecuencias perceptivamente menos significativas no se replican con la misma precisión. Por lo tanto, se puede utilizar un esquema de codificación más grueso, aunque algunos de los detalles más finos se perderán en la codificación. Un esquema típico de codificación más grueso puede basarse en el concepto de Extensión de Ancho de Banda (BWE), también conocido como Extensión de Banda Alta (HBE). Un enfoque de BWE o de HBE específico recientemente popular se conoce como Réplica de Subbanda (SBR) o Replicación de Banda Espectral (SBR). Estas técnicas son similares en cuanto a que codifican y decodifican algunas subbandas de frecuencia (generalmente bandas altas) con poco o ningún presupuesto de tasa de bits, produciendo así una tasa de bits significativamente más baja que un enfoque de codificación/decodificación normal. Con la tecnología de SBR, una estructura espectral fina en la banda de frecuencia alta se copia de la banda de baja frecuencia y se puede agregar ruido aleatorio. A continuación, se modela una envolvente espectral de la banda de frecuencia alta utilizando información lateral transmitida desde el codificador al decodificador.

La utilización del principio psicoacústico o efecto de enmascaramiento perceptivo tiene sentido para el diseño de compresión de audio. El equipo de audio/voz o la comunicación están destinados a la interacción con humanos, con todas sus capacidades y limitaciones de percepción. El equipo de audio tradicional trata de reproducir las señales con la mayor fidelidad al original. Un objetivo dirigido de manera más apropiada y, a menudo, más eficiente es lograr la fidelidad perceptible por humanos. Este es el objetivo de los codificadores perceptivos.

Si bien un objetivo principal de los codificadores perceptivos de audio digital es la reducción de datos, la codificación perceptiva también se puede utilizar para mejorar la representación del audio digital a través de la asignación avanzada de bits. Uno de los ejemplos de codificadores perceptivos podría ser los sistemas multibanda, que dividen el espectro de una manera que imita las bandas críticas de la psicoacústica. Al modelar la percepción humana, los codificadores perceptivos pueden procesar señales de manera muy similar a como lo hacen los humanos y aprovechar fenómenos tales como el enmascaramiento. Si bien este es su objetivo, el proceso se basa en un algoritmo preciso. Debido al hecho de que es difícil tener un modelo perceptivo muy preciso que cubra el comportamiento auditivo humano común, la precisión de cualquier expresión matemática del modelo perceptivo es aún limitada. Sin embargo, con una precisión limitada, el concepto de percepción ha ayudado en el diseño de códecs de audio. Numerosos esquemas de codificación de audio de MPEG se han beneficiado de la exploración del efecto de enmascaramiento perceptivo. Varios códecs del estándar de la UIT también utilizan el concepto perceptivo. Por ejemplo, el UIT G.729.1 realiza la llamada asignación dinámica de bits basada en el concepto de enmascaramiento perceptivo. El concepto de asignación dinámica de bits basado en la importancia perceptiva también se utiliza en el códec de EVS del 3GPP reciente.

Las Figuras 9A y 9B ilustran el esquema de un códec perceptivo en el dominio de la frecuencia típico. La Figura 9A ilustra un codificador en el dominio de la frecuencia, mientras que la Figura 9B ilustra un decodificador en el dominio de la frecuencia.

La señal 901 original se transforma primero en el dominio de la frecuencia para obtener coeficientes 902 en el dominio de la frecuencia sin cuantizar. Antes de cuantizar los coeficientes, la función de enmascaramiento (importancia perceptiva) divide el espectro de frecuencias en muchas subbandas (a menudo espaciadas igualmente para la simplicidad). Cada una de las subbandas asigna dinámicamente el número necesario de bits, mientras que mantiene que el número total de bits distribuidos a todas las subbandas no esté más allá del límite superior. A algunas subbandas pueden asignarse 0 bits si se considera que está por debajo del umbral de enmascaramiento. Una vez que se determina como lo qué se puede descartar, al resto se le asigna el número disponible de bits. Debido a que los bits no se desperdician en espectro enmascarado, se pueden distribuir en mayor cantidad al resto de la señal.

De acuerdo con los bits asignados, los coeficientes se cuantizan y el flujo 703 de bits se envía al decodificador. Aunque el concepto de enmascaramiento perceptivo ayudó mucho durante el diseño del códec, todavía no es perfecto debido a diversas razones y limitaciones.

Haciendo referencia a la Figura 9B, el postprocesamiento del lado del decodificador puede mejorar aún más la calidad perceptiva de la señal decodificada producida con tasas de bits limitadas. El decodificador utiliza primero los bits 904 recibidos para reconstruir los coeficientes 905 cuantizados. Luego, se postprocesan por un módulo 906 diseñado correctamente para obtener los coeficientes 907mejorados. Se realiza una transformación inversa en los coeficientes mejorados para tener la salida 908 final en el dominio del tiempo.

La Figura 10 ilustra un esquema de las operaciones en un codificador antes de codificar una señal de voz que comprende datos de audio, de acuerdo con realizaciones de la presente invención.

Haciendo referencia a la Figura 10, el método comprende seleccionar la codificación en el dominio de la frecuencia o la codificación en el dominio del tiempo (recuadro 1000) en base a una tasa de bits de codificación a ser utilizada para codificar la señal digital y un retardo de paso de la señal digital.

Seleccionar la codificación en el dominio de la frecuencia o la codificación en el dominio del tiempo, comprende el paso de determinar si la señal digital comprende una señal de paso corto, para la cual el retardo de paso es más corto que un límite de retardo de paso (recuadro 1010). Además, se determina si la tasa de bits de codificación es mayor que un límite superior de tasa de bits (recuadro 1020). Si la señal digital comprende una señal de paso corto y la tasa de bits de codificación es mayor que un límite superior de tasa de bits, se selecciona la codificación en el dominio de la frecuencia para codificar la señal digital.

De lo contrario, se determina si la tasa de bits de codificación es menor que un límite inferior de tasa de bits (recuadro 1030). Si la señal digital comprende una señal de paso corto y la tasa de bits de codificación es menor que un límite inferior de tasa de bits, se selecciona la codificación en el dominio del tiempo para codificar la señal digital.

De lo contrario, se determina si la tasa de bits de codificación está entre medias de un límite inferior de tasa de bits y un límite superior de tasa de bits (recuadro 1040). A continuación, se determina la periodicidad de entonación (recuadro 1050). Si la señal digital comprende una señal de paso corto y la tasa de bits de codificación está entre medias, y la periodicidad de entonación es baja, se selecciona la codificación en el dominio de la frecuencia para codificar la señal digital. Alternativamente, si la señal digital comprende una señal de paso corto y la tasa de bits de codificación está entre medias y la periodicidad de entonación es muy fuerte, se selecciona la codificación en el dominio del tiempo para codificar la señal digital.

Alternativamente, haciendo referencia al recuadro 1010, la señal digital no comprende una señal de paso corto para la cual el retardo de paso es más corto que un límite de retardo de paso. Se determina si la señal digital se clasifica como voz no sonora o voz normal (recuadro 1070). Si la señal digital no comprende una señal de paso corto y si la señal digital se clasifica como voz no sonora o voz normal, se selecciona la codificación en el dominio del tiempo para codificar la señal digital.

Por consiguiente, en diversas realizaciones, un método para procesar señales de voz antes de codificar una señal digital que comprende datos de audio, incluye seleccionar codificación en el dominio de la frecuencia o codificación en el dominio del tiempo en base a una tasa de bits de codificación a ser utilizada para codificar la señal digital y una detección de retardo de paso corto de la señal digital. La señal digital comprende una señal de paso corto para la cual el retardo de paso es más corto que un límite de retardo de paso. En diversas realizaciones, el método para seleccionar la codificación en el dominio de la frecuencia o la codificación en el dominio del tiempo comprende

seleccionar la codificación en el dominio de la frecuencia para codificar la señal digital, cuando la tasa de bits de codificación es mayor que un límite superior de tasa de bits, y seleccionar la codificación en el dominio del tiempo para codificar la señal digital, cuando la tasa de bits de codificación es menor que un límite inferior de tasa de bits. La tasa de bits de codificación es mayor que el límite superior de tasa de bits cuando la tasa de bits de codificación es mayor o igual que 46200 bps. La tasa de bits de codificación es menor que un límite inferior de tasa de bits cuando la tasa de bits de codificación es menor que 24,4 kbps.

De manera similar, en otra realización, un método para procesar señales de voz antes de la codificación de una señal digital que comprende datos de audio, comprende seleccionar la codificación en el dominio de la frecuencia para codificar la señal digital, cuando una tasa de bits de codificación es mayor que un límite superior de tasa de bits. Alternativamente, el método selecciona la codificación en el dominio del tiempo para codificar la señal digital, cuando la tasa de bits de codificación es menor que un límite inferior de tasa de bits. La señal digital comprende una señal de paso corto para la cual el retardo de paso es más corto que un límite de retardo de paso. La tasa de bits de codificación es mayor que el límite superior de tasa de bits, cuando la tasa de bits de codificación es mayor o igual que 46200 bps. La tasa de bits de codificación es menor que un límite inferior de tasa de bits, cuando la tasa de bits de codificación es menor que 24,4 kbps.

De manera similar, en otra realización, un método para procesar señales de voz antes de codificar comprende seleccionar la codificación en el dominio del tiempo para codificar una señal digital que comprende datos de audio, cuando la señal digital no comprende la señal de paso corto y la señal digital se clasifica como voz no sonora o voz normal. El método comprende, además, seleccionar la codificación en el dominio de la frecuencia para codificar la señal digital, cuando la tasa de bits de codificación está entre medias de un límite inferior de tasa de bits y un límite superior de tasa de bits. La señal digital comprende una señal de paso corto y la periodicidad de entonación es baja. El método incluye, además, seleccionar la codificación en el dominio del tiempo para codificar la señal digital, cuando la tasa de bits de codificación está entre medias y la señal digital comprende la señal de paso corto y una periodicidad de entonación es muy fuerte. El límite inferior de tasa de bits es de 24,4 kbps y el límite superior de tasa de bits es de 46,2 kbps.

La Figura11 ilustra un sistema10 de comunicación, de acuerdo con una realización de la presente invención.

El sistema10 de comunicación tiene dispositivos 7 y 8 de acceso de audio acoplados a una red 36 a través de enlaces 38 y 40 de comunicación. En una realización, el dispositivo 7 y 8 de acceso de audio son dispositivos de voz sobre protocolo de Internet (VOIP) y la red 36 es una red de área amplia (WAN), una red telefónica pública conmutada (PTSN) y/o la Internet. En otra realización, los enlaces 38 y 40 de comunicación son conexiones de banda ancha cableadas y/o inalámbricas. En una realización alternativa, los dispositivos 7 y 8 de acceso de audio son teléfonos celulares o móviles, los enlaces 38 y 40 son canales de telefonía móvil inalámbrica y la red 36 representa una red de telefonía móvil.

El dispositivo 7 de acceso de audio utiliza un micrófono 12 para convertir sonido, tal como música o la voz de una persona, en una señal 28 de entrada de audio analógica. Una interfaz 16 de micrófono convierte la señal 28 de entrada de audio analógica en una señal 33 de audio digital para la entrada en un codificador 22 de un CODEC 20. El codificador 22 produce una señal de TX de audio codificada para la transmisión a una red 26 a través de una interfaz 26 de red, de acuerdo con realizaciones de la presente invención. Un decodificador 24 dentro del CODEC 20 recibe la señal de RX de audio codificada desde la red 36, a través de la interfaz 26 de red, y convierte la señal de RX de audio codificada en una señal 34 de audio digital. La interfaz 18 de altavoz convierte la señal 34 de audio digital en la señal de audio 30 adecuada para accionar el altavoz 14.

En realizaciones de la presente invención, donde el dispositivo 7 de acceso de audio es un dispositivo de VOIP, algunos o todos los componentes dentro del dispositivo 7 de acceso de audio están implementados dentro de un teléfono. Sin embargo, en algunas realizaciones, el micrófono 12 y el altavoz 14 son unidades separadas, y la interfaz 16 de micrófono, la interfaz 18 de altavoz, el CODEC 20 y la interfaz 26 de red se implementan dentro de una computadora personal. El CODEC 20 puede implementarse en un software que se ejecuta en una computadora o un procesador dedicado, o mediante un hardware dedicado, por ejemplo, en un circuito integrado de aplicación específica (ASIC). La interfaz 16 de micrófono se implementa mediante un convertidor analógico a digital (A/D), así como otra circuitería de interfaz ubicada dentro del teléfono y/o dentro de la computadora. Igualmente, la interfaz 18 de altavoz se implementa mediante un convertidor digital a analógico y otra circuitería de interfaz ubicada dentro del teléfono y/o dentro de la computadora. En otras realizaciones, el dispositivo 7 de acceso de audio puede implementarse y particionarse de otras maneras conocidas en la técnica.

En realizaciones de la presente invención, donde el dispositivo 7 de acceso de audio es un teléfono celular o móvil, los elementos dentro del dispositivo 7 de acceso de audio se implementan dentro de un teléfono celular. El CODEC 20 se implementa mediante un software que se ejecuta en un procesador dentro del teléfono o por un hardware

dedicado. En otras realizaciones de la presente invención, el dispositivo de acceso de audio puede implementarse en otros dispositivos, tales como sistemas de comunicación cableados e inalámbricos de pares, tales como intercomunicadores y teléfonos de radio. En aplicaciones, tales como dispositivos de audio de consumo, el dispositivo de acceso de audio puede contener un CODEC con solo el codificador 22 o el decodificador 24, por ejemplo, en un sistema de micrófono digital o un dispositivo de reproducción de música. En otras realizaciones de la presente invención, el CODEC 20 puede utilizarse sin el micrófono 12 ni el altavoz 14, por ejemplo, en estaciones base celulares que acceden a la PTSN.

El procesamiento de voz para mejorar la clasificación no sonora/sonora descrito en diversas realizaciones de la presente invención, puede implementarse en el codificador 22 o en el decodificador 24, por ejemplo. El procesamiento de voz para mejorar la clasificación no sonora/sonora puede implementarse en hardware o en software, en diversas realizaciones. Por ejemplo, el codificador 22 o el decodificador 24 pueden ser parte de un chip de procesamiento de señal digital (DSP).

La Figura12 ilustra un diagrama de bloques de un sistema de procesamiento que puede utilizarse para implementar los dispositivos y métodos dados a conocer en el presente documento. Los dispositivos específicos pueden utilizar todos los componentes mostrados, o solo un subconjunto de los componentes, y los niveles de integración pueden variar de un dispositivo a otro. Además, un dispositivo puede contener múltiples instancias de un componente, tal como múltiples unidades de procesamiento, procesadores, memorias, transmisores, receptores, etc. El sistema de procesamiento puede comprender una unidad de procesamiento equipada con uno o más dispositivos de entrada/salida, tales como un altavoz, un micrófono, un ratón, una pantalla táctil, un teclado numérico, un teclado, una impresora, una pantalla y similares. La unidad de procesamiento puede incluir una unidad central de procesamiento (CPU), una memoria, un dispositivo de almacenamiento masivo, un adaptador de video y una interfaz de E/S conectados a un bus.

El bus puede ser uno o más de cualquier tipo de varias arquitecturas de bus, incluyendo un bus de memoria o controlador de memoria, un bus periférico, bus de vídeo, o similar. La CPU puede comprender cualquier tipo de procesador electrónico de datos. La memoria puede comprender cualquier tipo de memoria del sistema, tal como memoria de acceso aleatorio estático (SRAM), memoria de acceso aleatorio dinámico (DRAM), DRAM síncrona (SDRAM), memoria de solo lectura (ROM), una combinación de las mismas, o similares. En una realización, la memoria puede incluir ROM para utilización en el arranque y DRAM para el almacenamiento de programas y datos para utilización mientras se ejecutan programas.

El dispositivo de almacenamiento masivo puede comprender cualquier tipo de dispositivo de almacenamiento configurado para almacenar datos, programas y otra información, y para hacer los datos, los programas y otra información accesibles a través del bus. El dispositivo de almacenamiento masivo puede comprender, por ejemplo, uno o más de una unidad de estado sólido, una unidad de disco duro, una unidad de disco magnético, una unidad de disco óptico o similares.

El adaptador de video y la interfaz de E/S proporcionan interfaces para acoplar dispositivos externos de entrada y salida a la unidad de procesamiento. Como se ilustra, los ejemplos de dispositivos de entrada y salida incluyen la pantalla acoplada al adaptador de video y el ratón/teclado/impresora acoplados a la interfaz de E/S. Se pueden acoplar otros dispositivos a la unidad de procesamiento y se pueden utilizar tarjetas de interfaz adicionales o menos. Por ejemplo, se puede utilizar una interfaz en serie, tal como el Bus Universal en Serie (USB) (no mostrado), para proporcionar una interfaz para una impresora.

La unidad de procesamiento también incluye una o más interfaces de red, que pueden comprender enlaces cableados, tales como un cable Ethernet o similar, y/o enlaces inalámbricos para acceder a nodos o a diferentes redes. La interfaz de red permite que la unidad de procesamiento se comunique con las unidades remotas a través de las redes. Por ejemplo, la interfaz de red puede proporcionar comunicación inalámbrica a través de uno o más transmisores/antenas de transmisión y uno o más receptores/antenas de recepción. En una realización, la unidad de procesamiento está acoplada a una red de área local o a una red de área amplia para el procesamiento de datos y las comunicaciones con dispositivos remotos, tales como otras unidades de procesamiento, la Internet, instalaciones de almacenamiento remoto, o similares.

Si bien esta invención se ha descrito con referencia a realizaciones ilustrativas, esta descripción no pretende ser interpretada en un sentido limitante. Diversas modificaciones y combinaciones de las realizaciones ilustrativas, así como otras realizaciones de la invención, serán evidentes para los expertos en la técnica en referencia a la descripción. Por ejemplo, se pueden combinar entre sí diversas realizaciones descritas anteriormente.

Con referencia a la Figura13, se describe una realización de un aparato 130 para procesar señales de voz antes de codificar una señal digital. El aparato incluye:

un selector 131 de codificación configurado para seleccionar la codificación en el dominio de la frecuencia o la codificación en el dominio del tiempo en base a una tasa de bits de codificación a ser utilizada para codificar la señal digital y a una detección de retardo de paso corto de la señal digital.

En donde, cuando la señal digital incluye una señal de paso corto para la cual el retardo de paso es más corto que un límite de retardo de paso, el selector de codificación está configurado para

seleccionar la codificación en el dominio de la frecuencia para codificar la señal digital, cuando una tasa de bits de codificación es mayor que un límite superior de tasa de bits y

seleccionar la codificación en el dominio del tiempo para codificar la señal digital, cuando la tasa de bits de codificación es menor que un límite inferior de tasa de bits.

En donde, cuando la señal digital incluye una señal de paso corto para la cual el retardo de paso es más corto que un límite de retardo de paso, el selector de codificación está configurado para seleccionar la codificación en el dominio de la frecuencia para codificar la señal digital cuando la tasa de bits de codificación está entre medias de un límite inferior de tasa de bits y un límite superior de tasa de bits y en donde una periodicidad de entonación es baja.

En donde, cuando la señal digital no incluye una señal de paso corto para la cual el retardo de paso es más corto que un límite de retardo de paso, el selector de codificación está configurado para seleccionar la codificación en el dominio del tiempo para codificar la señal digital, cuando la señal digital se clasifica como voz no sonora o voz normal.

En donde, cuando la señal digital incluye una señal de paso corto para la cual el retardo de paso es más corto que un límite de retardo de paso, el selector de codificación está configurado para seleccionar la codificación en el dominio del tiempo para codificar la señal digital, cuando la tasa de bits de codificación está entre medias de un límite de inferior de tasa de bits y un límite superior de tasa de bits, y una periodicidad de entonación es muy fuerte.

El aparato incluye además una unidad 132 de codificación, la unidad de codificación está configurada para codificar la señal digital utilizando la codificación en el dominio de la frecuencia, seleccionada por el selector 131, o la codificación en el dominio del tiempo, seleccionada por el selector 131.

El selector de codificación y la unidad de codificación pueden implementarse por la CPU o por algunos circuitos de hardware, tales como FPGA, ASIC.

Con referencia a la Figura14, se describe una realización de un aparato 140 para procesar señales de voz antes de codificar una señal digital. El aparato incluye:

una unidad 141 de selección de codificación, la unidad de selección de codificación está configurada para

seleccionar la codificación en el dominio del tiempo para codificar una señal digital que comprende datos de audio, cuando la señal digital no incluye una señal de paso corto y la señal digital se clasifica como voz no sonora o voz normal;

seleccionar la codificación en el dominio de la frecuencia para codificar la señal digital, cuando la tasa de bits de codificación está entre medias de un límite inferior de tasa de bits y un límite superior de tasa de bits, y la señal digital incluye una señal de paso corto y la periodicidad de entonación es baja; y

seleccionar la codificación en el dominio del tiempo para codificar la señal digital, cuando la tasa de bits de codificación está entre medias y la señal digital incluye una señal de paso corto y una periodicidad de entonación es muy fuerte.

El aparato incluye, además, una segunda 142 unidad de codificación, la segunda unidad de codificación está configurada para codificar la señal digital utilizando la codificación en el dominio de la frecuencia, seleccionada por la unidad 141 de selección de codificación, o la codificación en el dominio del tiempo, seleccionada por la unidad 141 de selección de codificación.

La unidad de selección de codificación y la unidad de codificación pueden implementarse por la CPU o por algunos circuitos de hardware, tales como FPGA, ASIC.

Aunque la presente invención y sus ventajas se han descrito en detalle, debe entenderse que pueden realizarse diversos cambios, sustituciones y alteraciones en el presente documento sin apartarse del alcance de la invención,

como se define por las reivindicaciones adjuntas. Por ejemplo, muchas de las características y funciones discutidas anteriormente pueden implementarse en software, hardware o firmware, o una combinación de los mismos. Además, el alcance de la presente solicitud no pretende limitarse a las realizaciones particulares del proceso, la máquina, la fabricación, la composición de materia, los medios, los métodos y los pasos descritos en la memoria descriptiva.

**REIVINDICACIONES**

1. Un método para procesar señales de voz antes de codificar una señal digital que comprende datos de audio, el método que comprende:

seleccionar la codificación en el dominio de la frecuencia o la codificación en el dominio del tiempo en base a

una tasa de bits de codificación a ser utilizada para codificar la señal digital y

una detección de retardo de paso corto de la señal digital; en donde la detección de retardo de paso corto comprende detectar si la señal digital comprende una señal de paso corto para la cual el retardo de paso es más corto que un límite de retardo de paso, en donde el límite de retardo de paso es un paso mínimo permitido para un algoritmo de Predicción Lineal Excitada por Código (CELP) para codificar la señal digital.

2. El método de la reivindicación 1, en donde la señal digital comprende una señal de paso corto para la cual el retardo de paso es más corto que un límite de retardo de paso y, en donde, seleccionar la codificación en el dominio de la frecuencia o la codificación en el dominio del tiempo comprende:

seleccionar la codificación en el dominio de la frecuencia para codificar la señal digital, cuando una tasa de bits de codificación es mayor que un límite superior de tasa de bits, y

seleccionar la codificación en el dominio del tiempo para codificar la señal digital, cuando la tasa de bits de codificación es menor que un límite inferior de tasa de bits.

3. El método de la reivindicación 2, en donde la tasa de bits de codificación es mayor que el límite superior de tasa de bits, cuando la tasa de bits de codificación es mayor o igual que 46200 bps y, en donde, la tasa de bits de codificación es menor que un límite inferior de tasa de bits, cuando la tasa de bits de codificación es menor que 24,4 kbps.

4. El método de la reivindicación 1, en donde la señal digital comprende una señal de paso corto para la cual el retardo de paso es más corto que un límite de retardo de paso y, en donde, seleccionar la codificación en el dominio de la frecuencia o la codificación en el dominio del tiempo comprende:

seleccionar la codificación en el dominio de la frecuencia para codificar la señal digital, cuando la tasa de bits de codificación está entre medias de un límite inferior de tasa de bits y un límite superior de tasa de bits y, en donde, la periodicidad de entonación es baja.

5. El método de la reivindicación 1, en donde la señal digital no comprende una señal de paso corto para la cual el retardo de paso es más corto que un límite de retardo de paso y, en donde, seleccionar la codificación en el dominio de la frecuencia o la codificación en el dominio del tiempo comprende:

seleccionar la codificación en el dominio del tiempo para codificar la señal digital, cuando la señal digital se clasifica como voz no sonora o voz normal.

6. El método de la reivindicación 1, en donde la señal digital comprende una señal de paso corto para la cual el retardo de paso es más corto que un límite de retardo de paso y, en donde, seleccionar la codificación en el dominio de la frecuencia o la codificación en el dominio del tiempo comprende:

seleccionar la codificación en el dominio del tiempo para codificar la señal digital, cuando la tasa de bits de codificación está entre medias de un límite inferior de tasa de bits y un límite superior de tasa de bits, y una periodicidad de entonación es muy fuerte.

7. El método de la reivindicación 1, que comprende, además, codificar la señal digital utilizando la codificación en el dominio de la frecuencia seleccionada o la codificación en el dominio del tiempo seleccionada.

8. El método de la reivindicación 1, en donde seleccionar la codificación en el dominio de la frecuencia o la codificación en el dominio del tiempo en base al retardo de paso de la señal digital comprende detectar una señal de paso corto en base a determinar un parámetro para detectar la ausencia de energía de muy baja frecuencia o un parámetro para la nitidez espectral.

9. Un aparato para procesar señales de voz antes de codificar una señal digital que comprende datos de audio, el aparato que comprende un selector de codificación configurado para realizar el método de una cualquiera de las reivindicaciones 1-8.

10. El aparato de la reivindicación 9, el aparato que comprende, además, una unidad de codificación que está configurada para codificar la señal digital utilizando la codificación en el dominio de la frecuencia, seleccionada por el selector, o la codificación en el dominio del tiempo, seleccionada por el selector.

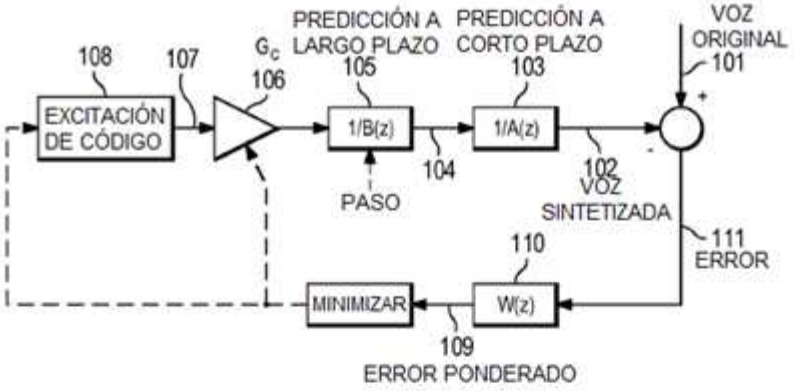


**Figura 1**

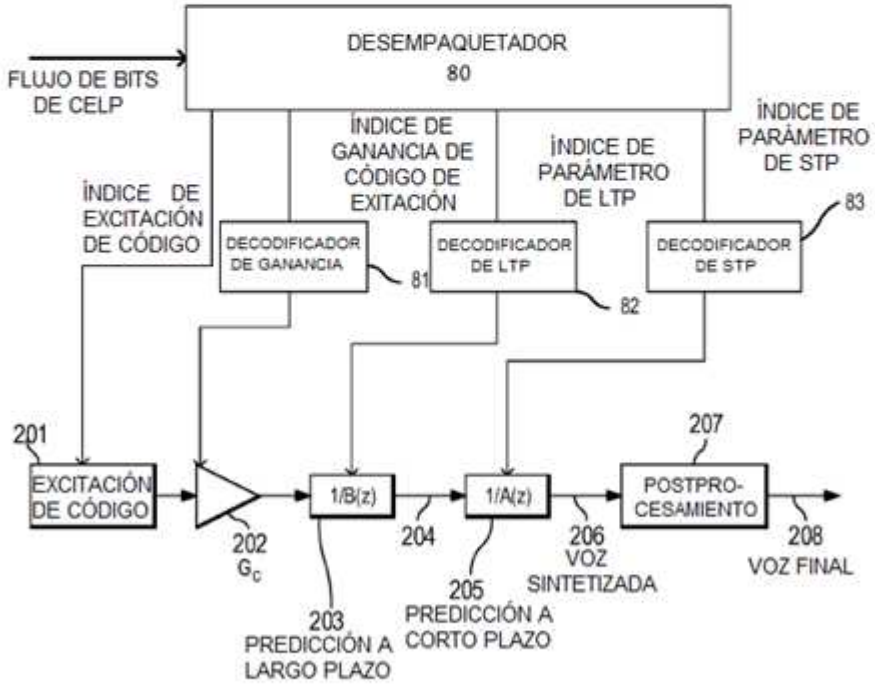


**Figura 2**

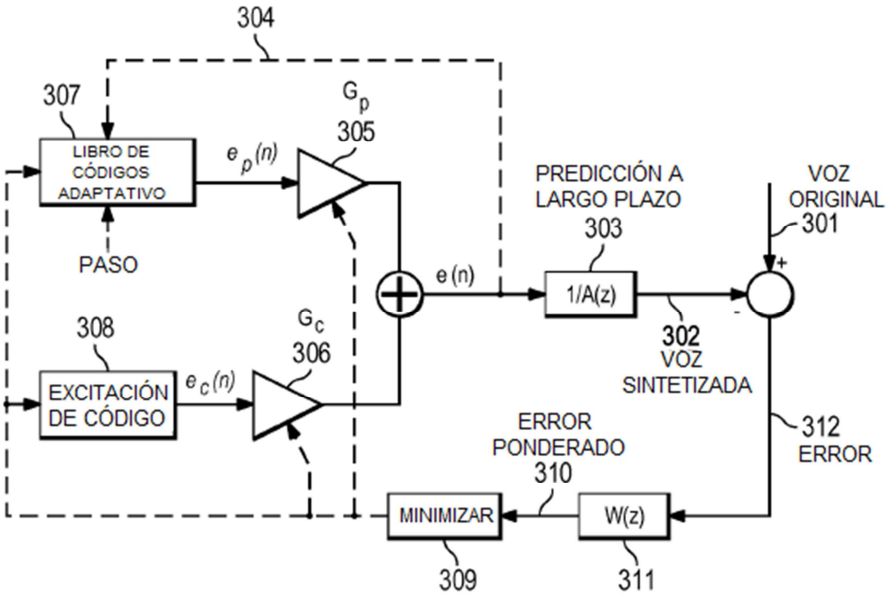
304
307
LIBRO DE CÓDIGOS ADAPTATIVO
PASO
$e_p(n)$
$G_p$
305
e (n)
PREDICCIÓN A LARGO PLAZO
303
1/A(z)
302
VOZ SINTETIZADA
VOZ ORIGINAL
301
+
-
308
EXCITACIÓN DE CÓDIGO
$e_c(n)$
$G_c$
306
ERROR PONDERADO
310
312
ERROR
MINIMIZAR
309
W(z)
311

***Figura 3***

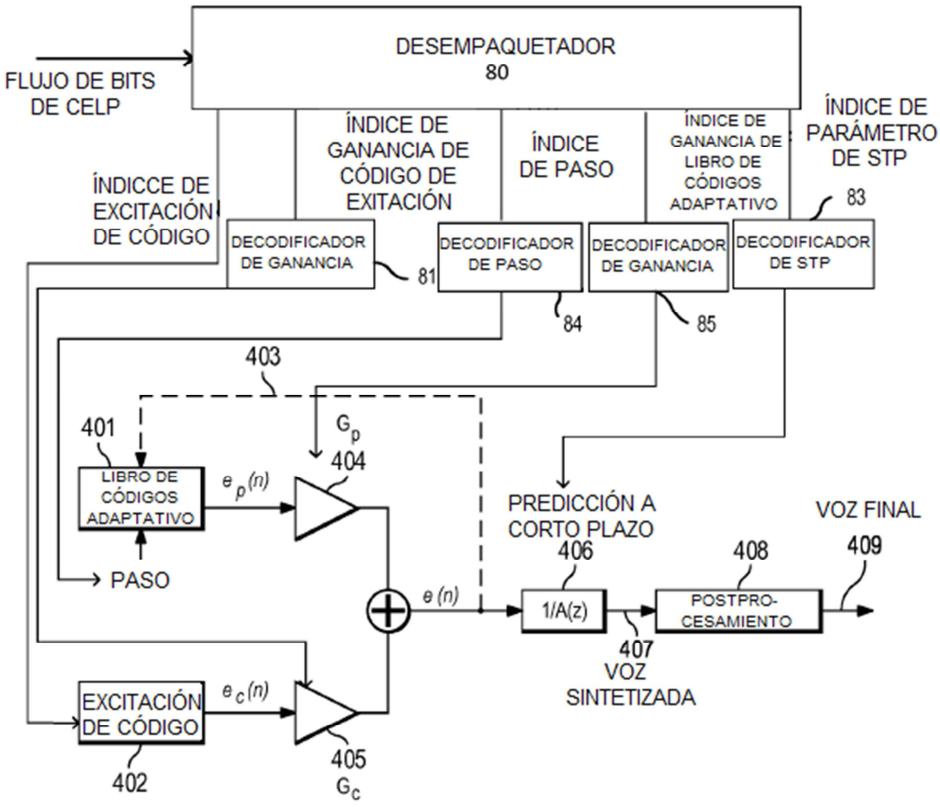
DESEMPAQUETADOR
80
FLUJO DE BITS DE CELP
ÍNDICE DE GANANCIA DE CÓDIGO DE EXITACIÓN
ÍNDICE DE PASO
ÍNDICE DE GANANCIA DE LIBRO DE CÓDIGOS ADAPTATIVO
ÍNDICE DE PARÁMETRO DE STP
ÍNDICCE DE EXCITACIÓN DE CÓDIGO
83
DECODIFICADOR DE GANANCIA
81
DECODIFICADOR DE PASO
DECODIFICADOR DE GANANCIA
DECODIFICADOR DE STP
84
85
403
401
$G_p$
404
LIBRO DE CÓDIGOS ADAPTATIVO
$e_p(n)$
PASO
PREDICCIÓN A CORTO PLAZO
406
VOZ FINAL
408
409
$e(n)$
1/A(z)
POSTPRO-CESAMIENTO
407
VOZ SINTETIZADA
EXCITACIÓN DE CÓDIGO
$e_c(n)$
402
405
$G_c$

***Figura 4***

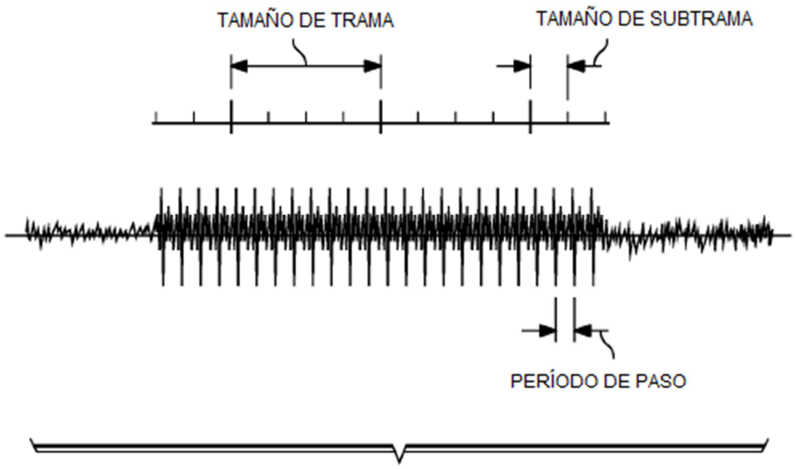


***Figura 5***

(TÉCNICA ANTERIOR)

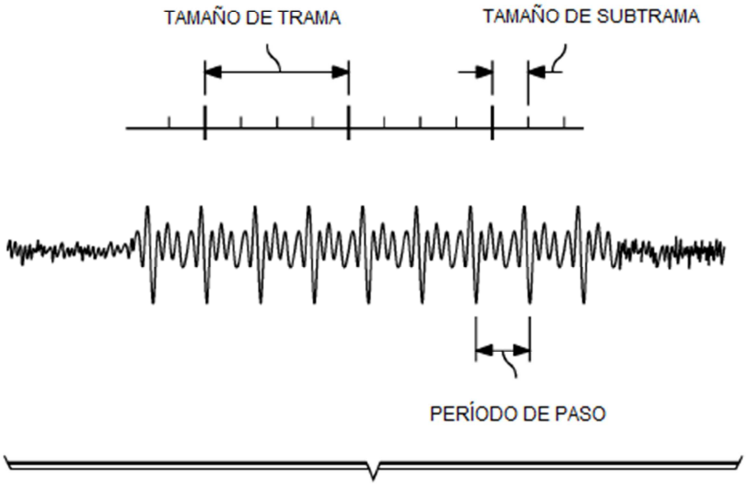


***Figura 6***

(TÉCNICA ANTERIOR)

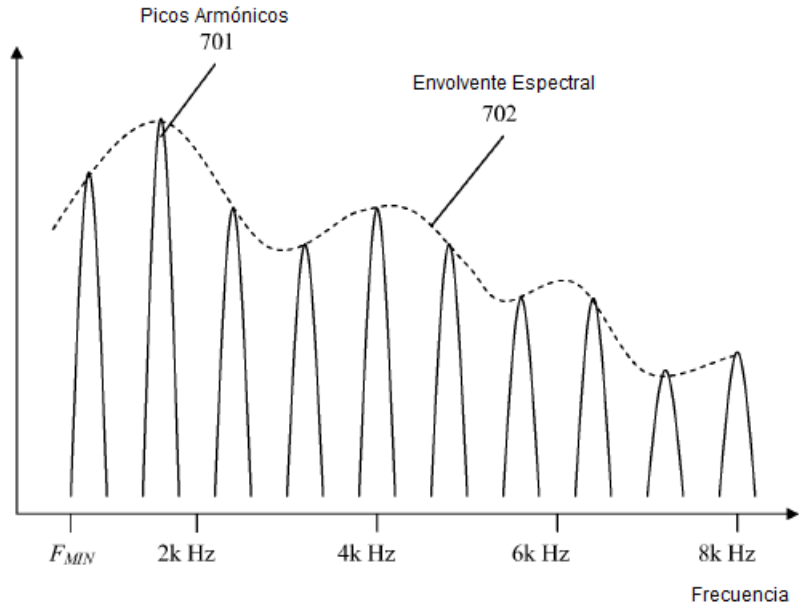
Picos Armónicos
701
Envolvente Espectral
702
$F_{MIN}$
2k Hz
4k Hz
6k Hz
8k Hz
Frecuencia

***Figura 7***

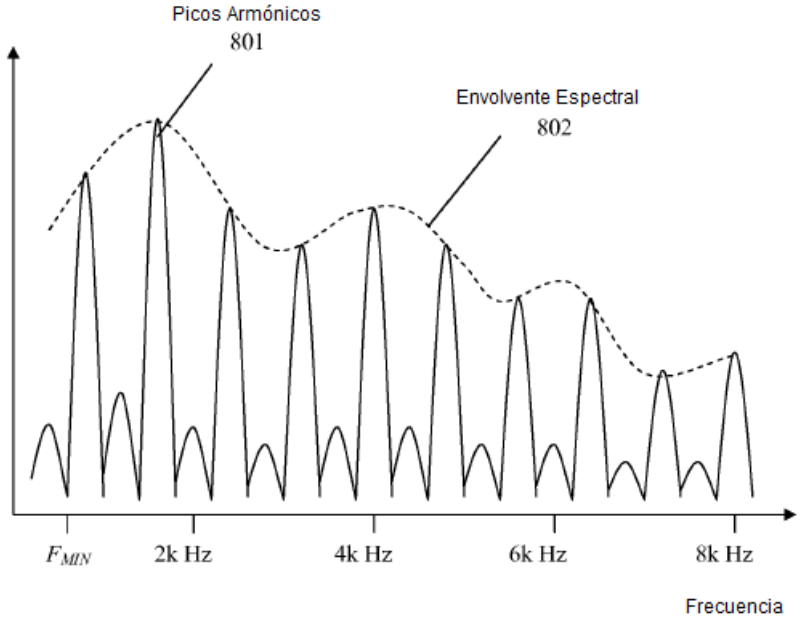
Picos Armónicos
801
Envolvente Espectral
802
$F_{MIN}$
2k Hz
4k Hz
6k Hz
8k Hz
Frecuencia

***Figura 8***

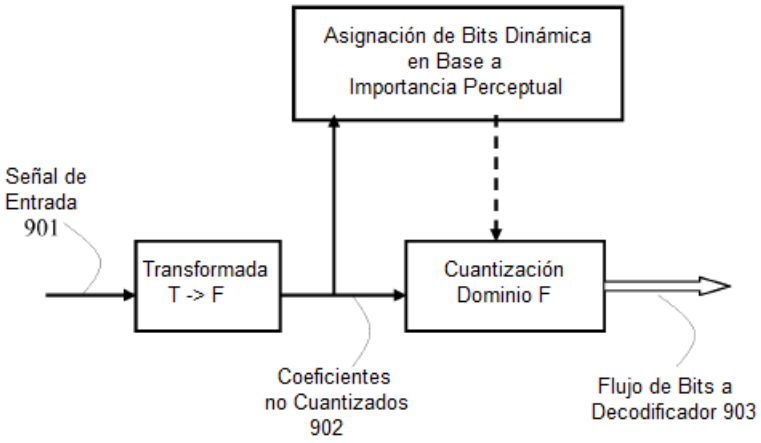


***Figura 9A***

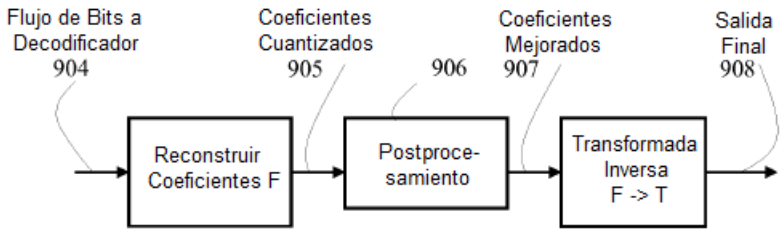


***Figura 9B***

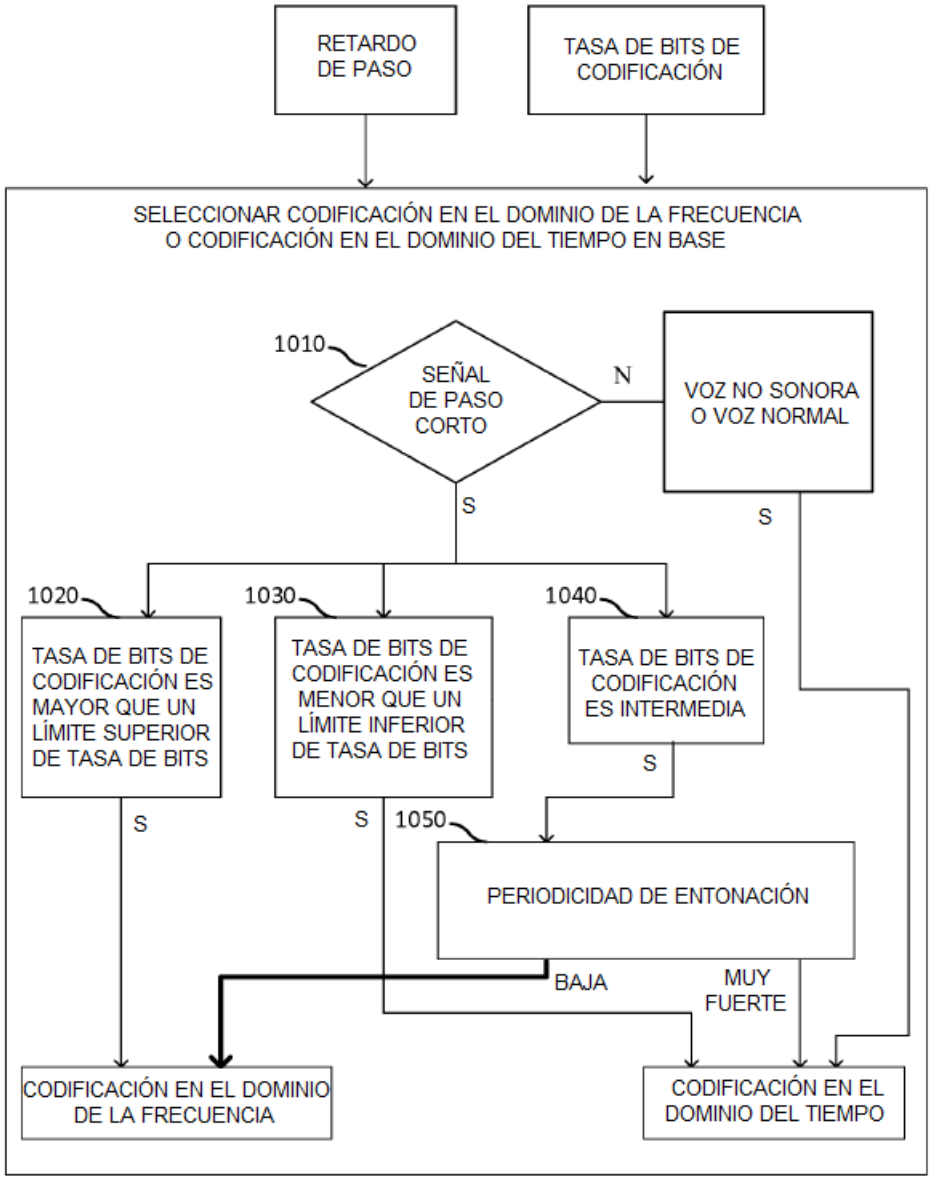
RETARDO DE PASO
TASA DE BITS DE CODIFICACIÓN
SELECCIONAR CODIFICACIÓN EN EL DOMINIO DE LA FRECUENCIA O CODIFICACIÓN EN EL DOMINIO DEL TIEMPO EN BASE
1010
SEÑAL DE PASO CORTO
N
VOZ NO SONORA O VOZ NORMAL
S
S
1020
TASA DE BITS DE CODIFICACIÓN ES MAYOR QUE UN LÍMITE SUPERIOR DE TASA DE BITS
1030
TASA DE BITS DE CODIFICACIÓN ES MENOR QUE UN LÍMITE INFERIOR DE TASA DE BITS
1040
TASA DE BITS DE CODIFICACIÓN ES INTERMEDIA
S
S
S
1050
PERIODICIDAD DE ENTONACIÓN
BAJA
MUY FUERTE
CODIFICACIÓN EN EL DOMINIO DE LA FRECUENCIA
CODIFICACIÓN EN EL DOMINIO DEL TIEMPO

***Figura 10***

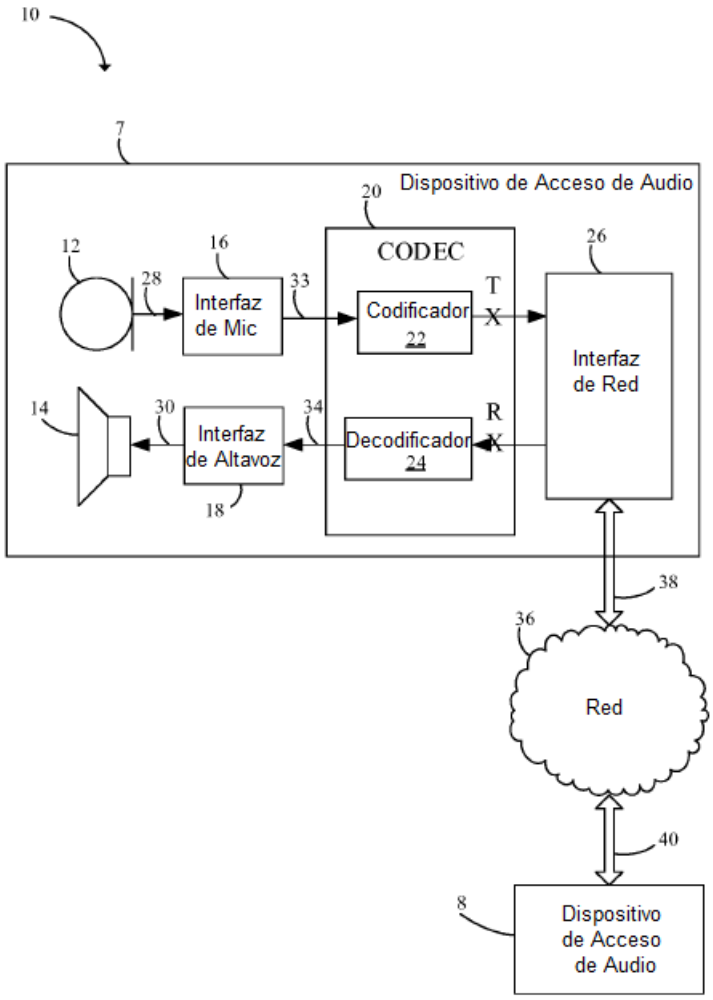
10
7
Dispositivo de Acceso de Audio
20
CODEC
12
16
28
Interfaz de Mic
33
Codificador 22
T X
26
Interfaz de Red
14
30
Interfaz de Altavoz
34
Decodificador 24
R X
18
38
36
Red
40
8
Dispositivo de Acceso de Audio

***Figura 11***

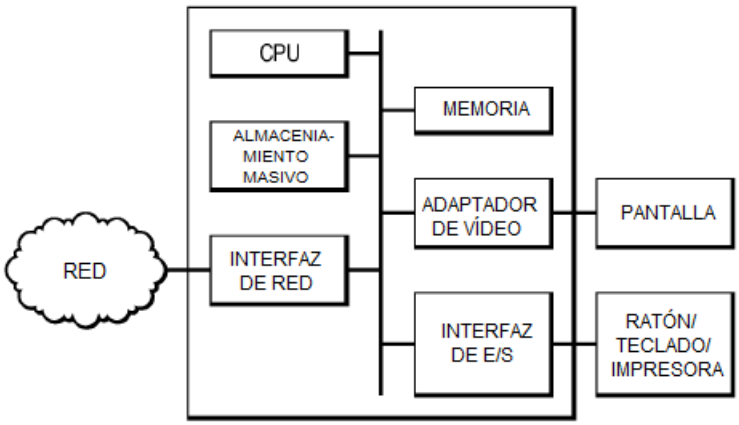


*Figura 12*

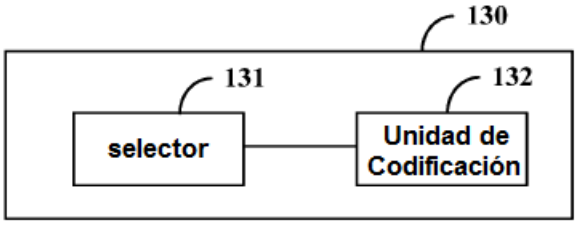


*Figura 13*

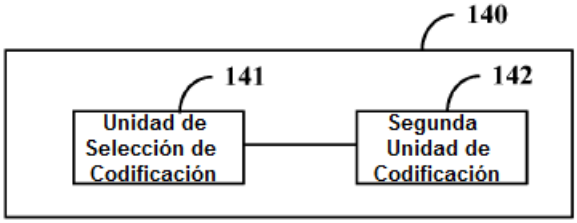


*Figura 14*