

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 644 730**

51 Int. Cl.:

G10L 19/26 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **05.04.2006 PCT/US2006/012641**

87 Fecha y número de publicación internacional: **07.12.2006 WO06130226**

96 Fecha de presentación y número de la solicitud europea: **05.04.2006 E 06740546 (4)**

97 Fecha y número de publicación de la concesión europea: **26.07.2017 EP 1899962**

54 Título: **Postfiltro de código de audio**

30 Prioridad:

31.05.2005 US 142603

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

30.11.2017

73 Titular/es:

MICROSOFT TECHNOLOGY LICENSING, LLC

(100.0%)

One Microsoft Way

Redmond, WA 98052, US

72 Inventor/es:

SUN, XIAOQIN;

WANG, TIAN;

KHALIL, HOSAM A.;

KOISHIDA, KAZUHITO y

CHEN, WEI-GE

74 Agente/Representante:

CARVAJAL Y URQUIJO, Isabel

ES 2 644 730 T3

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

DESCRIPCIÓN

Postfiltro de código de audio

Campo técnico

5 Las herramientas y técnicas descritas se refieren a códecs de audio, y en particular al postprocesamiento de habla decodificada.

Antecedentes

10 Con la aparición de las redes telefónicas digitales inalámbricas, la transferencia continua de audio a través de Internet y la telefonía por Internet, el procesamiento digital y la entrega de voz se han convertido en un lugar común. Los ingenieros usan una variedad de técnicas para procesar el habla de manera eficiente a la vez que se mantiene la calidad. Entender estas técnicas, ayuda a comprender cómo se representa y procesa la información de audio en un ordenador.

I. Representación de información de audio en un ordenador

15 Un ordenador procesa la información de audio como una serie de números que representan el audio. Un número único puede representar una muestra de audio, que es un valor de amplitud en un momento determinado. Varios factores afectan la calidad del audio, incluyendo la profundidad de la muestra y la frecuencia de muestreo.

Profundidad (o precisión) de muestra indica el rango de números utilizados para representar una muestra. Más valores posibles para cada muestra típicamente producen una salida de mayor calidad porque se pueden representar variaciones más sutiles en amplitud. Una muestra de ocho bits tiene 256 valores posibles, mientras que una muestra de dieciséis bits tiene 65.536 valores posibles.

20 La velocidad de muestreo (usualmente medida como el número de muestras por segundo) también afecta a la calidad. Cuanto mayor sea la frecuencia de muestreo, mayor será la calidad, ya que se pueden representar más frecuencias de sonido. Algunas frecuencias comunes de muestreo son 8.000, 11.025, 22.050, 32.000, 44.100, 48.000 y 96.000 muestras/segundo (Hz). La Tabla 1 muestra varios formatos de audio con diferentes niveles de calidad, junto con los correspondientes costes de la velocidad de bits sin procesar.

25 Tabla 1: Velocidades binarias para audio de calidad diferente

Profundidad de la muestra (bits/muestra)	Frecuencia de Muestreo (muestras/segundo)	Modo de Canal	Velocidad de Bits sin Procesar (bits/segundo)
8	8,000	mono	64,000
8	11,025	mono	88,200
16	44,100	estéreo	1,411,200

30 Tal como se muestra en la Tabla 1, el coste de un audio de alta calidad es una alta velocidad de bits. La información de audio de alta calidad consume grandes cantidades de capacidad de almacenamiento y transmisión del ordenador. Muchos ordenadores y redes informáticas carecen de los recursos necesarios para procesar el audio digital bruto. La compresión (también llamada codificar o codificación) disminuye el costo de almacenar y transmitir información de audio convirtiendo la información en una forma de velocidad de bits más baja. La compresión puede ser sin pérdidas (en la cual la calidad no sufre) o con pérdidas (en la cual la calidad sufre, pero la reducción de la velocidad de bits de la compresión sin pérdidas subsiguiente es más dramática). La descompresión (también llamada decodificación) extrae una versión reconstruida de la información original a partir de la forma comprimida.

35 Un códec es un sistema codificador/descodificador.

II. Codificadores y descodificadores de voz

Un objetivo de la compresión de audio es representar digitalmente señales de audio para proporcionar una calidad de señal máxima para una cantidad dada de bits. Dicho de otra manera, este objetivo es representar las señales de

audio con los mínimos bits para un nivel dado de calidad. Otros objetivos tales como la resiliencia a los errores de transmisión y la limitación del retardo general debido a la codificación/transmisión/descodificación se aplican en algunos escenarios.

5 Los diferentes tipos de señales de audio tienen características diferentes. La música se caracteriza por grandes rangos de frecuencias y amplitudes, y a menudo incluye dos o más canales. Por otro lado, el habla se caracteriza por menores rangos de frecuencias y amplitudes, y se representa comúnmente en un solo canal. Ciertos códecs y técnicas de procesamiento están adaptados para la música y el audio en general; otros códecs y técnicas de procesamiento están adaptados para el habla.

10 Un tipo de códec de conversación convencional utiliza predicción lineal ("LP") para lograr la compresión. La codificación del habla incluye varias etapas. El codificador encuentra y cuantifica coeficientes para un filtro de predicción lineal, que se utiliza para predecir los valores de la muestra como combinaciones lineales de los valores de la muestra anterior. Una señal residual (representada como una señal de "excitación") indica partes de la señal original no pronosticadas con precisión por el filtrado. En algunas etapas, el códec de voz utiliza diferentes técnicas de compresión para segmentos sonoros (caracterizados por vibración de cuerdas vocales), segmentos sin voz y segmentos silenciosos, y a que diferentes tipos de habla tienen características diferentes. Los segmentos de voz típicamente exhiben patrones de expresión de voz altamente repetitivos, incluso en el dominio residual. Para segmentos con voz, el codificador consigue una compresión adicional comparando la señal residual actual con los ciclos residuales anteriores y codificando la señal residual actual en términos de información de retraso o aplazamiento con relación a los ciclos previos. El codificador maneja otras discrepancias entre la señal original y la representación codificada predicha (a partir de la información de predicción lineal y de retardo) usando libros de códigos especialmente diseñados.

15 Aunque los códecs de voz como se ha descrito anteriormente tienen un buen rendimiento global para muchas aplicaciones, tienen varios inconvenientes. Por ejemplo, los códecs con pérdidas típicamente reducen la velocidad de bits reduciendo la redundancia en una señal de voz, lo que resulta en ruido u otros artefactos indeseables en el habla descodificada. Por consiguiente, algunos códecs filtran el habla descodificada para mejorar su calidad. Tales postfiltros normalmente vienen en dos tipos: postfiltros de dominio de tiempo y postfiltros de dominio de frecuencia.

25 Dada la importancia de la compresión y la descompresión para representar las señales de habla en los sistemas informáticos, no es sorprendente que el postfiltrado del habla reconstruida haya atraído la investigación. Cualesquiera que sean las ventajas de las técnicas anteriores para el procesamiento de la voz reconstruida u otro audio, no tienen las ventajas de las técnicas y herramientas descritas en la presente memoria.

30 El documento EP 1 308 932 A2 describe un método para procesar una señal de voz descodificada que incluye marcos DS sucesivos, incluyendo cada marco DS muestras DS, en el que el método comprende: filtrar de forma adaptativa la señal DS para producir una señal filtrada; incrementar la escala de la señal filtrada con una ganancia adaptativa actualizada una vez un marco DS, produciendo de este modo una señal escalonada de ganancia; y realizar una operación de suavizado para suavizar posibles discontinuidades de la forma de onda en la señal escalonada en ganancia.

35 El documento US 5 864 798 A describe el ajuste de la forma de un espectro de una señal de voz que incluye las etapas de utilizar un primer filtro con función de transferencia de polos cero $A(z)/B(z)$ para someter una señal de voz a un segundo filtro conectado en cascada con el primer filtro, para compensar una inclinación espectral debida al primer filtro, derivando independientemente dos coeficientes de filtro utilizados en el segundo filtro para compensar la inclinación espectral a partir de la función de transferencia polo-cero, y compensando la inclinación espectral correspondiente a la función de transferencia polo-cero de acuerdo con los coeficientes de filtro derivados.

40 El documento WO 2003/102923 A2 describe un método y un dispositivo para la mejora de tono selectiva en frecuencia de voz sintetizada, en el que una señal de sonido descodificada con vistas a mejorar una calidad percibida de esta señal de sonido descodificada se divide en una pluralidad de sub-bandas de frecuencia señales y postprocesamiento se aplica a al menos una de la señal de sub-banda de frecuencia. Después del postprocesamiento de esta al menos una señal de sub-banda de frecuencia, se pueden añadir las señales de sub-banda de frecuencia para producir una señal de sonido descodificada de salida postprocesada.

45 El documento US 6.064.962 describe un método de énfasis de formantes para enfatizar el formante como un pico espectral de una señal de voz de entrada y atenuar el valle espectral de la señal de voz de entrada. Un filtro de énfasis de espectro realiza el procesamiento para enfatizar el formante de la señal de voz de entrada y atenuar el valle de la señal de voz de entrada. Un filtro de característica variable de primer orden cuya característica cambia adaptativamente de acuerdo con la característica de la señal de voz de entrada y un filtro de característica fija de primer orden compensan una inclinación espectral incluida en una señal de salida del filtro de énfasis espectral.

55 Resumen

En resumen, la descripción detallada está dirigida a diversas técnicas y herramientas para códecs de audio, y específicamente a herramientas y técnicas relacionadas con el filtrado de voz descodificada. Las realizaciones descritas implementan una o más de las técnicas y herramientas descritas incluyendo, pero no limitándose a, lo siguiente:

5 En un aspecto, se calcula un conjunto de coeficientes de filtro para la aplicación a una señal de audio reconstruida. El cálculo incluye realizar uno o más cálculos de dominio de frecuencia. Se produce una señal de audio filtrada filtrando al menos una porción de la señal de audio reconstruida en un dominio de tiempo usando el conjunto de coeficientes de filtro.

10 En otro aspecto, se produce un conjunto de coeficientes de filtro para la aplicación a una señal de audio reconstruida. La producción de los coeficientes incluye procesar un conjunto de valores de coeficientes que representan uno o más picos y uno o más valles. El procesamiento del conjunto de valores de coeficientes incluye el recorte de uno o más de los picos o valles. Al menos una porción de la señal de audio reconstruida se filtra utilizando los coeficientes de filtro.

15 En otro aspecto, se recibe una señal compuesta reconstruida sintetizada a partir de varias señales de sub-banda de frecuencia reconstruida. Las señales de sub-banda incluyen una señal de sub-banda de primera frecuencia reconstruida para una primera banda de frecuencia y una segunda señal de sub-banda de frecuencia reconstruida para una segunda banda de frecuencia. En una región de frecuencia alrededor de una intersección entre la primera banda de frecuencia y la segunda banda de frecuencia, la señal compuesta reconstruida se mejora selectivamente.

Las diversas técnicas y herramientas se pueden usar en combinación o independientemente.

20 Características y ventajas adicionales se pondrán de manifiesto a partir de la siguiente descripción detallada de diferentes realizaciones que procede con referencia a los dibujos adjuntos.

Breve descripción de los dibujos

La figura 1 es un diagrama de bloques de un entorno informático adecuado en el que puede implementarse una o más de las realizaciones descritas.

25 La Figura 2 es un diagrama de bloques de un entorno de red en conjunción con el cual puede implementarse una o más de las realizaciones descritas.

La Figura 3 es un gráfico que representa una posible estructura de sub-banda de frecuencia que puede usarse para la codificación de sub-banda.

30 La figura 4 es un diagrama de bloques de un codificador de banda de voz en tiempo real en conjunción con el cual puede implementarse una o más de las realizaciones descritas.

La figura 5 es un diagrama de flujo que representa la determinación de parámetros de libro de códigos en una implementación.

La Figura 6 es un diagrama de bloques de un descodificador de banda de voz en tiempo real en conjunción con el cual puede implementarse una o más de las realizaciones descritas.

35 La Figura 7 es un diagrama de flujo que representa una técnica para determinar los coeficientes postfiltro que pueden usarse en algunas implementaciones.

Descripción detallada

40 Las realizaciones descritas se dirigen a técnicas y herramientas para procesar información de audio en codificación y/o descodificación. Con estas técnicas se mejora la calidad del habla derivada de un códec de voz, tal como un códec de voz en tiempo real. Tales mejoras pueden resultar del uso de diversas técnicas y herramientas por separado o en combinación.

45 Dichas técnicas y herramientas pueden incluir un postfiltro que se aplica a una señal de audio descodificada en el dominio del tiempo usando coeficientes que se diseñan o procesan en el dominio de la frecuencia. Las técnicas también pueden incluir valores de coeficiente de filtro de recorte o de limitación para uso en dicho filtro, o en algún otro tipo de postfiltro.

Las técnicas también pueden incluir un postfiltro que realce la magnitud de una señal de audio descodificada en regiones de frecuencia en las que la energía puede haberse atenuado debido a la descomposición en bandas de frecuencia. Como ejemplo, el filtro puede mejorar la señal en regiones de frecuencia cerca de intersecciones de bandas adyacentes.

5 Aunque las operaciones para las diversas técnicas se describen en un orden particular y secuencial por razones de presentación, debe entenderse que esta forma de descripción abarca reordenamientos menores en el orden de las operaciones, a menos que se requiera un orden particular. Por ejemplo, las operaciones descritas secuencialmente pueden en algunos casos ser reordenadas o realizadas simultáneamente. Además, en aras de la simplicidad, los diagramas de flujo pueden no mostrar las diversas maneras en que se pueden usar técnicas particulares conjuntamente con otras técnicas.

10 Aunque se describen a continuación características particulares de entorno informático y características de códec de audio, una o más de las herramientas y técnicas pueden utilizarse con diversos tipos de entornos informáticos y/o con diversos tipos diferentes de códecs. Por ejemplo, se pueden usar una o más de las técnicas postfiltro con códecs que no utilizan el modelo de codificación CELP, tales como códecs de modulación de código de pulso diferencial adaptativo, códecs de transformación y/u otros tipos de códecs. Como otro ejemplo, se pueden usar una o más de las técnicas postfiltro con códecs de banda única o códecs de sub-banda. Como otro ejemplo, se pueden aplicar una o más de las técnicas postfiltro a una única banda de un códec de múltiples bandas y/o a una señal sintetizada o no codificada que incluye contribuciones de múltiples bandas de un códec de múltiples bandas.

I. Entorno informático

20 La figura 1 ilustra un ejemplo generalizado de un entorno (100) informático adecuado en el que puede implementarse una o más de las realizaciones descritas. El entorno (100) informático no pretende sugerir ninguna limitación en cuanto al alcance de uso o funcionalidad de la invención, ya que la presente invención puede ser implementada en diversos entornos informáticos de propósito general o de propósito específico.

25 Haciendo referencia a la Figura 1, el entorno (100) informático incluye al menos una unidad (110) de procesamiento y una memoria (120). En la figura 1, esta configuración (130) más básica está incluida dentro de una línea discontinua. La unidad (110) de procesamiento ejecuta instrucciones ejecutables por ordenador y puede ser un procesador real o virtual. En un sistema de procesamiento múltiple, múltiples unidades de procesamiento ejecutan instrucciones ejecutables por ordenador para aumentar la potencia de procesamiento. La memoria (120) puede ser memoria volátil (por ejemplo, registros, memoria caché, RAM), memoria no volátil (por ejemplo, ROM, EEPROM, memoria instantánea, etc.) o alguna combinación de los dos. La memoria (120) almacena software (180) que implementa una o más de las técnicas de postfiltrado descritas aquí para un descodificador de voz.

30 Un entorno (100) informático puede tener características adicionales. En la figura 1, el entorno (100) informático incluye almacenamiento (140), uno o más dispositivos (150) de entrada, uno o más dispositivos (160) de salida y una o más conexiones (170) de comunicación. Un mecanismo de interconexión (no mostrado), tal como un bus, controlador o red, interconecta los componentes del entorno (100) informático. Típicamente, el software del sistema operativo (no mostrado) proporciona un entorno operativo para otro software que se ejecuta en el entorno (100) informático, y coordina las actividades de los componentes del entorno (100) informático.

35 El almacenamiento (140) puede ser extraíble o no extraíble, y puede incluir discos magnéticos, cintas magnéticas o casetes, CD-ROM, CD-RW, DVD o cualquier otro medio que pueda usarse para almacenar información y que se puede acceder dentro del entorno (100) informático. El almacenamiento (140) almacena instrucciones para el software (180).

40 El dispositivo o dispositivos (150) de entrada puede ser un dispositivo de entrada táctil tal como un teclado, ratón, lápiz o bola, un dispositivo de entrada de voz, un dispositivo de exploración, adaptador de red, u otro dispositivo que proporcione entrada al entorno (100) informático. Para el audio, el dispositivo (150) de entrada puede ser una tarjeta de sonido, micrófono u otro dispositivo que acepta entrada de audio en forma analógica o digital, o un lector de CD/DVD que proporciona muestras de audio al entorno (100) informático. El dispositivo (160) de salida puede ser una pantalla, una impresora, un altavoz, un grabador de CD/DVD, un adaptador de red u otro dispositivo que proporcione salida desde el entorno (100) informático.

45 Las conexiones (170) de comunicación permiten la comunicación a través de un medio de comunicación a otra entidad informática. El medio de comunicación transmite información tal como instrucciones ejecutables por ordenador, información de voz comprimida u otros datos en una señal de datos modulada. Una señal de datos modulada es una señal que tiene una o más de sus características establecidas o cambiadas de tal manera que codifican información en la señal. A modo de ejemplo, y no de limitación, los medios de comunicación incluyen técnicas cableadas o inalámbricas implementadas con un portador eléctrico, óptico, RF, infrarrojo, acústico u otro.

La invención se puede describir en el contexto general de medios legibles por ordenador. Los medios legibles por ordenador son cualquier medio disponible al que se pueda acceder dentro de un entorno informático. A modo de ejemplo, y no limitativo, con el entorno (100) informático, los medios legibles por ordenador incluyen memoria (120), almacenamiento (140), medios de comunicación y combinaciones de cualquiera de los anteriores.

5 La invención se puede describir en el contexto general de las instrucciones ejecutables por ordenador, tales como las incluidas en los módulos de programa, que se ejecutan en un entorno informático en un procesador real o virtual de destino. Generalmente, los módulos de programa incluyen rutinas, programas, bibliotecas, objetos, clases, componentes, estructuras de datos, etc. que realizan tareas particulares o implementan tipos de datos abstractos
10 particulares. La funcionalidad de los módulos de programa se puede combinar o dividir entre módulos de programa como se desee en varias realizaciones. Las instrucciones ejecutables por ordenador para los módulos de programa pueden ejecutarse en un entorno informático local o distribuido.

En aras de la presentación, la descripción detallada puede usar términos como "determinar", "generar", "ajustar" y "aplicar" para describir operaciones de ordenador en un entorno informático. Estos términos son abstracciones de alto nivel para operaciones realizadas por un ordenador, y no deben confundirse con actos realizados por un ser humano.
15 Las operaciones informáticas reales correspondientes a estos términos varían dependiendo de la implementación.

II. Entorno de red generalizado y códec de voz en tiempo real

La figura 2 es un diagrama de bloques de un entorno (200) de red generalizada en conjunción con el cual puede implementarse una o más de las realizaciones descritas. Una red (250) separa varios componentes del lado del codificador de diversos componentes del lado del descodificador.
20

Las funciones primarias de los componentes del lado del codificador y del lado del descodificador son codificación y descodificación de voz, respectivamente. En el lado del codificador, una memoria (210) intermedia de entrada acepta y almacena la entrada (202) de voz. El codificador (230) de voz toma la entrada (202) de voz desde la memoria (210) intermedia de entrada y la codifica.

25 Específicamente, un divisor (212) de marcos divide las muestras de la entrada (202) de voz en marcos. En una implementación, los marcos tienen uniformemente veinte ms de largo -160 muestras para la entrada de ocho kHz y 320 muestras para la entrada de dieciséis kHz-. En otras implementaciones, los marcos tienen diferentes duraciones, son no uniformes o se superponen, y/o la velocidad de muestreo de la entrada (202) es diferente. Los marcos pueden estar organizados en supermarco/marco, marco/submarco u otra configuración para diferentes etapas de la codificación y descodificación.
30

Un clasificador (214) de marcos clasifica los marcos de acuerdo con uno o más criterios, tales como energía de la señal, velocidad de cruce nula, ganancia de predicción a largo plazo, diferencial de ganancia y/o otros criterios para submarcos o marcos enteros. Basado en los criterios, el clasificador (214) de marcos clasifica los diferentes marcos en clases tales como silenciosa, sin voz, sonora y de transición (por ejemplo, sin voz a sonora). Además, los marcos pueden clasificarse de acuerdo con el tipo de codificación redundante, si existe, que se utiliza para el cuadro. La clase de marco afecta a los parámetros que se calcularán para codificar el marco. Además, la clase de marco puede afectar a la resiliencia de resolución y pérdida con la que se codifican los parámetros, con el fin de proporcionar más resolución y pérdida de resiliencia a clases y parámetros de marcos más importantes. Por ejemplo, los marcos silenciosos típicamente son codificados a una velocidad muy baja, son muy fáciles de recuperar por ocultación si se pierden, y pueden no necesitar protección contra pérdidas. Los marcos sin voz típicamente están codificados a una velocidad ligeramente superior, son razonablemente sencillos de recuperar por ocultación si se pierden, y no están significativamente protegidos contra la pérdida. Los marcos de voz y de transición suelen codificarse con más bits, dependiendo de la complejidad de el marco, así como de la presencia de transiciones. Los marcos de voz y de transición también son difíciles de recuperar si se pierden, y por lo tanto están más protegidos contra la pérdida.
35 40 45 Alternativamente, el clasificador (214) de marcos utiliza otras clases de marcos y/o clases de marcos adicionales.

La señal de voz de entrada puede dividirse en señales de sub-banda antes de aplicar un modelo de codificación, tal como el modelo de codificación CELP, a la información de sub-bandas para un marco. Esto se puede hacer usando una serie de uno o más bancos (216) de filtros de análisis (como filtros de análisis QMF). Por ejemplo, si se va a utilizar una estructura de tres bandas, entonces la banda de baja frecuencia se puede dividir pasando la señal a través de un filtro de paso bajo. Del mismo modo, la banda alta se puede dividir pasando la señal a través de un filtro de paso alto. La banda media se puede dividir pasando la señal a través de un filtro de paso de banda, que puede incluir un filtro de paso bajo y un filtro de paso alto en serie. Alternativamente, pueden usarse otros tipos de disposiciones de filtro para la descomposición de sub-bandas y/o la temporización de filtrado (por ejemplo, antes de la división del marco). Si sólo se desea descodificar una banda para una parte de la señal, dicha parte puede pasar por alto los bancos (216) de filtros de análisis.
50 55

El número de bandas n se puede determinar por velocidad de muestreo. Por ejemplo, en una implementación, se utiliza una estructura de banda única para una frecuencia de muestreo de ocho kHz. Para tasas de muestreo de 16 kHz y 22.05 kHz, se usa una estructura de tres bandas como se muestra en la Figura 3. En la estructura de tres bandas de la Figura 3, la banda (310) de baja frecuencia se extiende a la mitad del total del ancho de banda F (de 0 a $0,5 F$). La otra mitad del ancho de banda se divide por igual entre la banda (320) media y la banda (330) alta. Cerca de las intersecciones de las bandas, la respuesta de frecuencia para una banda disminuye gradualmente desde el nivel de paso hasta el nivel de parada, que se caracteriza por una atenuación de la señal en ambos lados a medida que se aproxima la intersección. También se pueden usar otras divisiones de la anchura de banda de frecuencia. Por ejemplo, para una frecuencia de muestreo de treinta y dos kHz, se puede usar una estructura de cuatro bandas igualmente espaciada.

La banda de baja frecuencia es típicamente la banda más importante para las señales de voz porque la energía de la señal se descompone típicamente hacia los rangos de frecuencias más altas. En consecuencia, la banda de baja frecuencia se codifica a menudo utilizando más bits que las otras bandas. En comparación con una estructura de codificación de banda única, la estructura de sub-bandas es más flexible y permite un mejor control del ruido de cuantificación a través de la banda de frecuencias. En consecuencia, se cree que la calidad de la voz perceptiva se mejora significativamente mediante el uso de la estructura de sub-banda. Sin embargo, como se discute a continuación, la descomposición de sub-bandas puede causar pérdida de energía de la señal en las regiones de frecuencia cerca de la intersección de bandas adyacentes. Esta pérdida de energía puede degradar la calidad de la señal de voz descodificada resultante.

En la Figura 2, cada sub-banda se codifica por separado, como se ilustra mediante componentes (232, 234) de codificación. Mientras que los componentes (232, 234) de codificación de banda se muestran por separado, la codificación de todas las bandas puede hacerse por un único codificador, o pueden ser codificadas por codificadores separados. Tal codificación de banda se describe con más detalle a continuación con referencia a la Figura 4. Alternativamente, el códec puede funcionar como un códec de banda única. El habla codificada resultante se proporciona al software para una o más capas (240) de red a través de un multiplexor ("MUX") (236). Las capas (240) de red procesan el habla codificado para su transmisión a través de la red (250). Por ejemplo, el software de capa de red empaqueta marcos de información de voz codificada en paquetes que siguen al protocolo RTP, que son retransmitidos a través de Internet usando UDP, IP y varios protocolos de capa física. Alternativamente, se utilizan otras y/o capas adicionales de software o protocolos de red.

La red (250) es una red área amplia, con conmutación de paquetes, tal como Internet. Alternativamente, la red (250) es una red de área local u otro tipo de red.

En el lado del descodificador, el software para una o más capas (260) de interconexión recibe y procesa los datos transmitidos. Los protocolos y software de red, transporte y capa superior y software en la capa o capas (260) de red del lado del descodificador normalmente corresponden a los de las capas (240) de red del lado del codificador. Las capas de red proporcionan la información de voz codificada al descodificador (270) de voz a través de un desmultiplexor ("DEMUX") (276).

El descodificador (270) descodifica cada una de las sub-bandas separadamente, tal como se representa en componentes (272, 274) de descodificación de banda. Todas las sub-bandas pueden ser descodificadas por un solo descodificador, o pueden ser descodificadas por descodificadores de banda independientes.

Las sub-bandas descodificadas son luego sintetizadas en una serie de uno o más bancos (280) de filtros de síntesis (tales como filtros de síntesis QMF), que emiten una voz (292) descodificada. Alternativamente, se utilizan otros tipos de disposiciones de filtro para la síntesis de sub-banda. Si solamente está presente una banda única, entonces la banda descodificada puede evitar los bancos (280) de filtros. Si están presentes múltiples bandas, la emisión (292) de voz descodificada también se puede pasar a través de un postfiltro (284) de mejora de frecuencia media para mejorar la calidad de la salida (294) de voz mejorada resultante. Una implementación del postfiltro de mejora de frecuencia media se discute con más detalle a continuación.

A continuación, se describe un descodificador generalizado de banda de voz en tiempo real con referencia a la figura 6, pero se pueden usar otros descodificadores de voz. Además, algunas o todas las herramientas y técnicas descritas pueden usarse con otros tipos de codificadores y descodificadores de audio, tales como codificadores y descodificadores de música, o codificadores y descodificadores de audio de propósito general.

Aparte de estas funciones de codificación y descodificación primarias, los componentes también pueden compartir información (mostrada en líneas discontinuas en la Figura 2) para controlar la elasticidad de velocidad, calidad y/o pérdida del habla codificada. El controlador (220) de velocidad considera una variedad de factores tales como la complejidad de la entrada de corriente en la memoria (210) intermedia de entrada, la plenitud de la memoria intermedia de las memorias intermedias de salida en el codificador (230) u otro lugar, la velocidad de salida deseada, el ancho de banda de red actual, congestión de la red/condiciones de ruido y/o velocidad de pérdida del descodificador. El descodificador (270) devuelve la información de la velocidad de pérdida del descodificador al

controlador (220) de velocidad. Las capas (240, 260) de interconexión de red recopilan o estiman información sobre el ancho de banda de la red actual y las condiciones de congestión/ruido, que son devueltas al controlador (220) de velocidad. Alternativamente, el controlador (220) de velocidad considera otros factores y/o factores adicionales.

5 El controlador (220) de velocidad dirige el codificador (230) de voz para cambiar la resiliencia de velocidad, calidad y/o pérdida con la que se codifica el habla. El codificador (230) puede cambiar la velocidad y la calidad ajustando factores de cuantificación para parámetros o cambiando la resolución de códigos de entropía que representan los parámetros. Adicionalmente, el codificador puede cambiar la resiliencia de pérdida ajustando la velocidad o el tipo de codificación redundante. De este modo, el codificador (230) puede cambiar la asignación de bits entre las funciones de codificación primaria y las funciones de resiliencia de pérdida en función de las condiciones de la red.

10 La figura 4 es un diagrama de bloques de un codificador (400) de banda de voz generalizada en conjunción con el cual puede implementarse una o más de las realizaciones descritas. El codificador (400) de banda corresponde generalmente a uno cualquiera de los componentes (232, 234) de codificación de banda de la figura 2.

15 El codificador (400) de banda acepta la entrada (402) de banda de los bancos de filtros (u otros filtros) si la señal se divide en múltiples bandas. Si la señal no se divide en múltiples bandas, entonces la entrada (402) de banda incluye muestras que representan toda la anchura de banda. El codificador de banda produce salida (492) de banda codificada.

20 Si una señal se divide en múltiples bandas, entonces un componente (420) de reducción por muestreo puede realizar una reducción por muestreo en cada banda. Por ejemplo, si la velocidad de muestreo se establece en dieciséis kHz y cada marco tiene una duración de veinte ms, cada marco incluye 320 muestras. Si no se realizó una reducción por muestreo y el marco se dividió en la estructura de tres bandas mostrada en la Figura 3, entonces tres veces más muestras (es decir, 320 muestras por banda, o 960 muestras totales) se codificarían y descodificarían para el cuadro. Sin embargo, cada banda puede ser reducida por muestreo. Por ejemplo, la banda (310) de baja frecuencia puede ser reducida por muestreo de 320 muestras a 160 muestras, y cada una de las bandas (320) media y banda (330) alta puede ser reducida por muestreo de 320 muestras a 80 muestras, donde las bandas (310, 320, 330) se extienden más de la mitad, un cuarto y un cuarto del rango de frecuencia, respectivamente. (El grado de reducción (420) por muestreo en esta implementación varía en relación con los rangos de frecuencia de las bandas (310, 320, 330). Sin embargo, son posibles otras implementaciones. En etapas posteriores, se utilizan típicamente menos bits para las bandas más altas debido la energía de la señal típicamente declina hacia los rangos de frecuencia más altos). De acuerdo con esto, esto proporciona un total de 320 muestras para codificar y
 25 30 descodificar para el marco.

35 El componente (430) de análisis LP calcula los coeficientes (432) de predicción lineal. En una implementación, el filtro LP utiliza diez coeficientes para la entrada de ocho kHz y dieciséis coeficientes para la entrada de dieciséis kHz, y el componente (430) de análisis LP calcula un conjunto de coeficientes de predicción lineal por cuadro para cada banda. Alternativamente, el componente (430) de análisis LP calcula dos conjuntos de coeficientes por marco para cada banda, uno para cada una de dos ventanas centradas en lugares diferentes, o calcula un número diferente de coeficientes por banda y/o por marco.

40 El componente (435) de procesamiento LPC recibe y procesa los coeficientes (432) de predicción lineal. Típicamente, el componente (435) de procesamiento LPC convierte los valores LPC en una representación diferente para una cuantificación y codificación más eficientes. Por ejemplo, el componente (435) de procesamiento de LPC convierte los valores de LPC en una representación de par espectral de línea (LSP), y los valores de LSP se cuantifican (por ejemplo, por cuantificación de vector) y se codifican. Los valores de LSP pueden estar codificados o predichos de otros valores de LSP. Varias representaciones, técnicas de cuantificación y técnicas de codificación son posibles para los valores de LPC. Los valores de LPC se proporcionan de alguna forma como parte de la salida (492) de banda codificada para empaquetar y transmitir (junto con cualquier parámetro de cuantificación y otra información necesaria para la reconstrucción). Para su uso posterior en el codificador (400), el componente (435) de procesamiento LPC reconstruye los valores LPC. El componente (435) de procesamiento LPC puede realizar interpolación para valores LPC (tal como equivalentemente en representación LSP u otra representación) para suavizar las transiciones entre diferentes conjuntos de coeficientes LPC o entre los coeficientes LPC utilizados para diferentes submarcos de marcos.

50 El filtro (440) de síntesis (o "predicción a corto plazo") acepta valores (438) LPC reconstruidos y los incorpora en el filtro. El filtro (440) de síntesis recibe una señal de excitación y produce una aproximación de la señal original. Para un marco dado, el filtro (440) de síntesis puede amortiguar un número de muestras reconstruidas (por ejemplo, diez para un filtro de diez etapas) del marco anterior para el inicio de la predicción.

55 Los componentes (450, 455) de ponderación perceptiva aplican la ponderación perceptiva a la señal original y a la salida modelada del filtro (440) de síntesis para desactivar selectivamente la estructura formante de las señales de voz para hacer que los sistemas auditivos sean menos sensibles a la cuantificación de errores. Los componentes (450, 455) de ponderación perceptiva explotan fenómenos psicoacústicos como el enmascaramiento. En una

implementación, los componentes (450, 455) de ponderación perceptiva aplican pesos basados en los valores (432) LPC originales recibidos del componente (430) de análisis LP. Alternativamente, los componentes (450, 455) de ponderación perceptiva aplican otros pesos y/o adicionales.

5 Después de los componentes (450, 455) de ponderación perceptiva, el codificador (400) calcula la diferencia entre la señal original ponderada perceptivamente y la salida ponderada perceptivamente del filtro (440) de síntesis para producir una señal (434) de diferencia. Alternativamente, el codificador (400) utiliza una técnica diferente para calcular los parámetros de voz.

10 El componente (460) de parametrización de excitación busca encontrar la mejor combinación de índices de libro de códigos adaptativos, índices de libro de códigos fijos e índices de libro de códigos de ganancia en términos de minimización de la diferencia entre la señal original perceptivamente ponderada y la señal sintetizada (en términos de cuadrado medio ponderado error u otros criterios). Muchos parámetros se calculan por submarco, pero más generalmente los parámetros pueden ser por supermarco, marco o submarco. Como se ha discutido anteriormente, los parámetros para diferentes bandas de un marco o submarco pueden ser diferentes. La Tabla 2 muestra los tipos de parámetros disponibles para diferentes clases de marco en una implementación.

15 Tabla 2: Parámetros para diferentes clases de marcos

Clase de marco	Parámetros
Silencioso	Información de clase; LSP; ganancia (por marco, para el ruido generado)
Sin voz	Información de clase; LSP; parámetros de libro de códigos de pulso, aleatorio y ganancia
Con voz	Información de clase; LSP; parámetros de libro de códigos adaptativo, pulso, aleatorio y ganancia (por submarco)
Transición	

20 En la figura 4, el componente (460) de parametrización de excitación divide el cuadro en submarcos y calcula los índices del libro de códigos y las ganancias para cada submarco según sea apropiado. Por ejemplo, el número y el tipo de etapas de libro de códigos por utilizar y las resoluciones de índices de libro de códigos pueden determinarse inicialmente mediante un modo de codificación, en el que el modo está dictado por el componente de control de velocidad discutido anteriormente. Un modo particular también puede dictar parámetros de codificación y descodificación distintos del número y tipo de etapas de libro de códigos, por ejemplo, la resolución de los índices de libro de códigos. Los parámetros de cada etapa de libro de códigos se determinan optimizando los parámetros para minimizar el error entre una señal diana y la contribución de esa etapa de libro de códigos a la señal sintetizada.

25 (Como se usa aquí, el término "optimizar" significa encontrar una solución adecuada bajo restricciones aplicables tales como reducción de distorsión, tiempo de búsqueda de parámetros, complejidad de búsqueda de parámetros, velocidad de bits de parámetros, etc., en lugar de realizar una búsqueda completa en el espacio de parámetros. De manera similar, el término "minimizar" debe entenderse en términos de encontrar una solución adecuada bajo restricciones aplicables). Por ejemplo, la optimización se puede hacer usando una técnica de error cuadrático medio modificado. La señal objetivo para cada etapa es la diferencia entre la señal residual y la suma de las contribuciones de las etapas de libro de códigos anteriores, si las hay, a la señal sintetizada. Alternativamente, pueden usarse otras técnicas de optimización.

30

35 La figura 5 muestra una técnica para determinar parámetros de libro de códigos de acuerdo con una implementación. El componente (460) de parametrización de excitación realiza la técnica, potencialmente junto con otros componentes tales como un controlador de velocidad. Alternativamente, otro componente en un codificador realiza la técnica.

40 Haciendo referencia a la figura 5, para cada submarco en un marco de voz o de transición, el componente (460) de parametrización de excitación determina (510) si puede utilizarse un libro de códigos adaptativo para el submarco actual. (Por ejemplo, el control de velocidad puede dictar que no se use un libro de códigos adaptativo para un marco particular). Si no se va a usar el libro de códigos adaptativo, entonces un conmutador de libro de códigos adaptativo indicará que no se usarán libros (535) de códigos adaptativos. Por ejemplo, esto podría hacerse estableciendo un indicador de un bit en el nivel de marco indicando que no se usan libros de códigos adaptativos en el marco, especificando un modo de codificación particular en el nivel de marco o estableciendo un indicador de un bit para cada submarco indicando que no se utiliza ningún libro de códigos adaptativo en el submarco.

Haciendo referencia todavía a la Figura 5, si se puede usar un libro de códigos adaptativo, entonces el componente (460) determina los parámetros de libro de códigos adaptativos. Estos parámetros incluyen un índice, o valor de tono, que indica un segmento deseado del historial de la señal de excitación, así como una ganancia que se aplicará al segmento deseado. En las figuras 4 y 5, el componente (460) realiza una búsqueda (520) de tono de bucle cerrado. Esta búsqueda comienza con el tono determinado por el componente (425) de búsqueda de tono de bucle abierto opcional en la figura 4. Un componente de búsqueda (425) de tono de bucle abierto analiza la señal ponderada producida por el componente (450) de ponderación para estimar su tono. Comenzando con este paso estimado, la búsqueda (520) de tono de bucle cerrado optimiza el valor de tono para disminuir el error entre la señal objetivo y la señal sintetizada ponderada generada a partir de un segmento indicado de la historia de la señal de excitación. También se optimiza el valor de ganancia (525) de libro de códigos adaptativo. El valor de ganancia de libro de códigos adaptativo indica un multiplicador para aplicar a los valores predichos de tono (los valores del segmento indicado del historial de señales de excitación), para ajustar la escala de los valores. La ganancia multiplicada por los valores de tono predeterminados es la contribución de libro de códigos adaptativa a la señal de excitación para el marco o submarco actual. La optimización de ganancia (525) y la búsqueda (520) de paso de bucle cerrado producen un valor de ganancia y un valor de índice, respectivamente, que minimizan el error entre la señal objetivo y la señal sintetizada ponderada a partir de la contribución de libro de códigos adaptativo.

Si el componente (460) determina (530) que se va a utilizar el libro de códigos adaptativo, entonces los parámetros de libro de códigos adaptativos son señalizados (540) en la corriente de bits. Si no es así, se indica que no se utiliza ningún libro de códigos adaptativo para el submarco (535), como por ejemplo estableciendo un indicador de nivel de submarco de un bit, como se ha explicado anteriormente. Esta determinación (530) puede incluir determinar si la contribución de libro de códigos adaptativo para el submarco particular es suficientemente significativa como para valer el número de bits requerido para señalar los parámetros de libro de códigos adaptativos. Alternativamente, se puede usar alguna otra base para la determinación. Además, aunque la Figura 5 muestra la señalización después de la determinación, alternativamente, las señales son procesadas por lotes hasta que la técnica termina para un marco o supermarco.

El componente (460) de parametrización de excitación determina (550) también si se utiliza un libro de códigos de impulsos. El uso o no uso del libro de códigos de impulsos se indica como parte de un modo de codificación global para el marco actual, o puede indicarse o determinarse de otras maneras. Un libro de códigos de impulsos es un tipo de libro de códigos fijos que especifica uno o más impulsos que se deben aportar a la señal de excitación. Los parámetros del libro de códigos de pulsos incluyen pares de índices y signos (las ganancias pueden ser positivas o negativas). Cada par indica un pulso que debe incluirse en la señal de excitación, indicando el índice la posición del impulso y el signo que indica la polaridad del impulso. El número de impulsos incluidos en el libro de códigos de impulsos y utilizados para contribuir a la señal de excitación puede variar dependiendo del modo de codificación. Además, el número de pulsos puede depender de si se está utilizando o no un libro de códigos adaptativo.

Si se utiliza el libro de códigos de impulsos, entonces se optimizan (555) los parámetros de libro de códigos de impulsos para minimizar el error entre la contribución de los impulsos indicados y una señal de destino. Si no se utiliza un libro de códigos adaptativo, entonces la señal objetivo es la señal original ponderada. Si se utiliza un libro de códigos adaptativo, entonces la señal objetivo es la diferencia entre la señal original ponderada y la contribución del libro de códigos adaptativo a la señal sintetizada ponderada. En algún punto (no mostrado), los parámetros del libro de códigos de impulsos son entonces señalizados en la corriente de bits.

El componente (460) de parametrización de excitación determina (565) también si se van a utilizar cualquiera de las etapas de libro de códigos fijos aleatorios. El número (si lo hay) de las etapas de libro de códigos aleatorio se indica como parte de un modo de codificación global para el marco actual, o puede determinarse de otras maneras. Un libro de códigos aleatorio es un tipo de libro de códigos fijos que utiliza un modelo de señal predefinido para los valores que codifica. Los parámetros del libro de códigos pueden incluir el punto de partida para un segmento indicado del modelo de señal y un signo que puede ser positivo o negativo. La longitud o rango del segmento indicado está típicamente fijo y por lo tanto no se señala típicamente, sino que alternativamente se señala una longitud o extensión del segmento indicado. Una ganancia se multiplica por los valores en el segmento indicado para producir la contribución del libro de códigos aleatorio a la señal de excitación.

Si se utiliza al menos una etapa de libro de códigos aleatorio, a continuación, se optimizan (570) los parámetros de la etapa de libro de códigos para el libro de códigos para minimizar el error entre la contribución de la etapa de libro de códigos aleatorio y una señal de destino. La señal objetivo es la diferencia entre la señal original ponderada y la suma de la contribución a la señal sintetizada ponderada del libro de códigos adaptativo (si existe), el libro de códigos de impulsos (si existe) y las etapas de libro de códigos aleatorias previamente determinadas (si las hay). En algún punto (no mostrado), los parámetros de libro de códigos aleatorios son entonces señalizados en la corriente de bits.

El componente (460) determina (580) entonces si se van a utilizar otras etapas de libro de códigos aleatorias. Si es así, entonces los parámetros de la siguiente etapa de libro de códigos aleatorio se optimizan (570) y se señalizan como se ha descrito anteriormente. Esto continúa hasta que se han determinado todos los parámetros para las

etapas de libro de códigos aleatorio. Todas las etapas de libro de códigos aleatorio pueden usar el mismo modelo de señal, aunque probablemente indicarán diferentes segmentos del modelo y tendrán diferentes valores de ganancia. Alternativamente, se pueden usar diferentes modelos de señales para diferentes etapas de libro de códigos aleatorio.

- 5 Cada ganancia de excitación puede cuantificarse independientemente o dos o más ganancias pueden cuantificarse conjuntamente, según lo determinado por el controlador de velocidad y/u otros componentes.

Aunque se ha establecido un orden particular para optimizar los diversos parámetros de libro de códigos, se pueden usar otros órdenes y técnicas de optimización. Por ejemplo, todos los libros de códigos aleatorios podrían ser optimizados simultáneamente. Por lo tanto, aunque la Figura 5 muestra el cálculo secuencial de diferentes parámetros de libro de códigos, alternativamente, dos o más parámetros de libro de códigos diferentes se optimizan conjuntamente (por ejemplo, variando conjuntamente los parámetros y evaluando los resultados según alguna técnica de optimización no lineal). Adicionalmente, podrían usarse otras configuraciones de libros de códigos u otros parámetros de señal de excitación.

15 La señal de excitación en esta implementación es la suma de cualesquiera contribuciones del libro de códigos adaptativo, del libro de códigos de impulsos y de la o las etapas de libro de códigos aleatorio. Alternativamente, el componente (460) de la Figura 4 puede calcular otros parámetros y/o adicionales para la señal de excitación.

Haciendo referencia a la figura 4, los parámetros de libro de códigos para la señal de excitación se señalan o proporcionan de otro modo a un descodificador (465) local (encerrado por líneas discontinuas en la figura 4), así como a la salida (492) de banda. Por lo tanto, para cada banda, la salida (492) de codificador incluye la salida del componente (435) de procesamiento LPC discutido anteriormente, así como la salida del componente (460) de parametrización de excitación.

La velocidad de bits de la salida (492) depende en parte de los parámetros utilizados por los libros de códigos, y el codificador (400) puede controlar la velocidad y/o la calidad de bits conmutando entre diferentes conjuntos de índices de libro de códigos, utilizando códigos embebidos, o utilizando otras técnicas. Diferentes combinaciones de los tipos y etapas de libro de códigos pueden producir diferentes modos de codificación para diferentes marcos, bandas y/o submarcos. Por ejemplo, un marco sin voz puede usar solamente una etapa de libro de códigos aleatoria. Puede utilizarse un libro de códigos adaptativo y un libro de códigos de impulsos para un marco de voz de baja velocidad. Un marco de alta velocidad puede codificarse usando un libro de códigos adaptativo, un libro de códigos de impulsos y una o más etapas de libro de códigos aleatorio. En un marco, la combinación de todos los modos de codificación para todas las sub-bandas juntas puede denominarse un modo establecido. Puede haber varios conjuntos de modos predefinidos para cada velocidad de muestreo, con diferentes modos correspondientes a diferentes velocidades de bits de codificación. El módulo de control de velocidad puede determinar o influir en el modo establecido para cada marco.

Haciendo referencia todavía a la figura 4, la salida del componente (460) de parametrización de excitación es recibida por componentes (470, 472, 474, 476) de reconstrucción de libro de códigos y componentes (480, 482, 484, 486) de aplicación de ganancia correspondientes a los libros de códigos utilizados por el componente (460) de parametrización. Las etapas (470, 472, 474, 476) de libro de códigos y los correspondientes componentes (480, 482, 484, 486) de aplicación de ganancia reconstruyen las contribuciones de los libros de códigos. Dichas contribuciones se suman para producir una señal (490) de excitación, que es recibida por el filtro (440) de síntesis, donde se usa junto con las muestras "predichas" a partir de las cuales se produce una predicción lineal subsiguiente. Las porciones retardadas de la señal de excitación también se utilizan como una señal de historial de excitación por el componente (470) de reconstrucción de libro de códigos adaptativos para reconstruir los siguientes parámetros de libro de códigos adaptativos (por ejemplo, contribución de tono) y por el componente (460) de parametrización (por ejemplo, el índice de tono y los valores de ganancia de tono).

Haciendo referencia de nuevo a la figura 2, la salida de banda para cada banda es aceptada por el MUX (236), junto con otros parámetros. Tales otros parámetros pueden incluir, entre otra información, información (222) de clase de marco del clasificador (214) de marco y modos de codificación de marco. El MUX (236) construye paquetes de la capa de aplicación para pasar a otro software, o el MUX (236) pone los datos en las cargas útiles de los paquetes que siguen un protocolo como RTP. El MUX puede almacenar en memoria intermedia parámetros de modo que permita la repetición selectiva de los parámetros para la corrección de errores hacia adelante en paquetes posteriores. En una implementación, el MUX (236) empaqueta en un único paquete la información de voz codificada primaria para un marco, junto con información de corrección de errores hacia adelante para todo o parte de uno o más marcos anteriores.

El MUX (236) proporciona realimentación tal como la plenitud de la memoria intermedia de corriente para propósitos de control de velocidad. Más generalmente, varios componentes del codificador (230) (incluyendo el clasificador (214) de marcos y MUX (236)) pueden proporcionar información a un controlador (220) de velocidad tal como el mostrado en la figura 2.

La corriente de bits DEMUX (276) de la Figura 2 acepta la información de voz codificada como entrada y la analiza para identificar y procesar parámetros. Los parámetros pueden incluir la clase de marco, alguna representación de los valores LPC y los parámetros del libro de códigos. La clase de marco puede indicar qué otros parámetros están presentes para un marco dada. Más en general, el DEMUX (276) utiliza los protocolos utilizados por el codificador (230) y extrae los parámetros que el codificador (230) empaqueta en paquetes. Para paquetes recibidos a través de una red conmutada por paquetes dinámica, el DEMUX (276) incluye una memoria intermedia de fluctuación para suavizar las fluctuaciones a corto plazo en la velocidad de paquetes durante un periodo de tiempo dado. En algunos casos, el descodificador (270) regula el retardo de la memoria intermedia y gestiona cuando los paquetes se leen de la memoria intermedia para integrar el retardo, el control de calidad, el ocultamiento de los marcos faltantes, etc. en la descodificación. En otros casos, un componente de capa de aplicación gestiona la memoria intermedia de fluctuación de fase y la memoria intermedia de fluctuación de fase se llena a una velocidad variable y se agota mediante el descodificador (270) a una velocidad constante o relativamente constante.

El DEMUX (276) puede recibir múltiples versiones de parámetros para un segmento dado, incluyendo una versión codificada primaria y una o más versiones secundarias de corrección de errores. Cuando falla la corrección de errores, el descodificador (270) utiliza técnicas de ocultación tales como repetición o estimación de parámetros basándose en la información que se recibió correctamente.

La Figura 6 es un diagrama de bloques de un descodificador (600) de banda de voz en tiempo real generalizado en conjunción con el cual se pueden implementar una o más realizaciones descritas. El descodificador (600) de banda corresponde generalmente a uno cualquiera de los componentes (272, 274) de descodificación de banda de la figura 2.

El descodificador (600) de banda acepta información (692) de voz codificada para una banda (que puede ser la banda completa o una de múltiples sub-bandas) como entrada y produce una salida (604) reconstruida filtrada después de descodificar y filtrar. Los componentes del descodificador (600) tienen componentes correspondientes en el codificador (400), pero en general el descodificador (600) es más simple porque carece de componentes para la ponderación perceptiva, el bucle de procesamiento de excitación y el control de velocidad.

El componente (635) de procesamiento LPC recibe información que representa los valores LPC en la forma proporcionada por el codificador (400) de banda (así como cualquier parámetro de cuantificación y otra información necesaria para la reconstrucción). El componente (635) de procesamiento LPC reconstruye los valores (638) LPC utilizando el inverso de la conversión, cuantificación, codificación, etc. aplicados previamente a los valores LPC. El componente (635) de procesamiento LPC también puede realizar interpolación para valores LPC (en representación LPC u otra representación tal como LSP) para suavizar las transiciones entre diferentes conjuntos de coeficientes LPC.

Las etapas (670, 672, 674, 676) de libro de códigos y componentes (680, 682, 684, 686) de aplicación de ganancia descodifican los parámetros de cualquiera de las etapas de libro de códigos correspondientes utilizadas para la señal de excitación y calculan la contribución de cada etapa de libro de códigos que se utiliza. Generalmente, la configuración y las operaciones de las etapas (670, 672, 674, 676) de libro de códigos y componentes de ganancia (680, 682, 684, 686) corresponden a la configuración y operaciones de las etapas (470, 472, 474, 476) de libro de códigos y componentes (480, 482, 484, 486) de ganancia en el codificador (400). Las aportaciones de las etapas de libro de códigos usadas se suman y la señal (690) de excitación resultante se alimenta al filtro (640) de síntesis. Los valores de retardados de la señal (690) de excitación también se usan como un historial de excitación por el libro (670) de códigos adaptativo al calcular la contribución del libro de códigos adaptativo para porciones subsiguientes de la señal de excitación.

El filtro (640) de síntesis acepta valores (638) LPC reconstruidos y los incorpora en el filtro. El filtro (640) de síntesis almacena muestras previamente reconstruidas para su procesamiento. La señal (690) de excitación se hace pasar a través del filtro de síntesis para formar una aproximación de la señal de voz original.

La señal (602) de sub-banda reconstruida también se alimenta a un postfiltro (694) de corto plazo. El postfiltro a corto plazo produce una salida (604) de sub-banda filtrada. A continuación, se describen varias técnicas para calcular coeficientes para el postfiltro (694) de corto plazo. Para el postfiltrado adaptativo, el descodificador (270) puede calcular los coeficientes a partir de parámetros (por ejemplo, valores LPC) para el habla codificada. Alternativamente, los coeficientes se proporcionan a través de alguna otra técnica.

Haciendo referencia a la figura 2, como se ha expuesto anteriormente, si hay múltiples sub-bandas, la salida de sub-banda para cada sub-banda se sintetiza en los bancos (280) de filtros de síntesis para formar la salida (292) de habla.

Las relaciones mostradas en las Figuras 2-6 indican flujos generales de información; otras relaciones no se muestran por motivos de simplicidad. Dependiendo de la implementación y del tipo de compresión deseada, los

componentes pueden ser añadidos, omitidos, divididos en múltiples componentes, combinados con otros componentes y/o reemplazados con componentes similares. Por ejemplo, en el entorno (200) mostrado en la figura 2, el controlador (220) de velocidad se puede combinar con el codificador (230) de voz. Los componentes potencialmente añadidos incluyen una aplicación de codificación multimedia (o reproducción) que gestiona el codificador de voz (o descodificador) así como otros codificadores (o descodificadores) y recoge información de condición de red y descodificador y que realiza funciones de corrección de errores adaptativas. En realizaciones alternativas, diferentes combinaciones y configuraciones de componentes procesan información de voz utilizando las técnicas descritas en la presente memoria.

III. Técnicas de postfiltro

En algunas realizaciones, un descodificador u otra herramienta aplica un postfiltro a corto plazo al audio reconstruido, tal como el habla reconstruida, después de haber sido descodificado. Tal filtro puede mejorar la calidad perceptiva del habla reconstruida.

Los postfiltros son típicamente postfiltros de dominio de tiempo o postfiltros de dominio de frecuencia. Un postfiltro de dominio de tiempo convencional para un códec CELP incluye un filtro de síntesis de coeficiente de predicción lineal de todos los polos escalado por un factor constante y un filtro inverso de coeficiente de predicción lineal totalmente cero escalado por otro factor constante.

Además, un fenómeno conocido como "inclinación espectral" se produce en muchas señales de voz porque las amplitudes de frecuencias más bajas en el habla normal son a menudo más altas que las amplitudes de frecuencias más altas. Por lo tanto, el espectro de amplitud del dominio de frecuencia de una señal de voz a menudo incluye una pendiente, o "inclinación". Por consiguiente, la inclinación espectral del habla original debería estar presente en una señal de voz reconstruida. Sin embargo, si los coeficientes de un postfiltro incorporan también dicha inclinación, entonces el efecto de la inclinación se amplificará en la salida del postfiltro de modo que la señal de habla filtrada se distorsione. Por lo tanto, algunos postfiltros de dominio del tiempo también tienen un filtro de paso alto de primer orden para compensar la inclinación espectral.

Por lo tanto, las características de los postfiltros del dominio del tiempo se controlan típicamente mediante dos o tres parámetros, lo que no proporciona mucha flexibilidad.

Un postfiltro de dominio de frecuencia, por otra parte, tiene una forma más flexible de definir las características postfiltro. En un postfiltro de dominio de frecuencia, los coeficientes de filtro se determinan en el dominio de frecuencia. La señal de voz descodificada se transforma en el dominio de la frecuencia y se filtra en el dominio de la frecuencia. La señal filtrada se transforma a continuación en el dominio del tiempo. Sin embargo, la señal de dominio de tiempo filtrada resultante típicamente tiene un número diferente de muestras que la señal de dominio de tiempo no filtrada original. Por ejemplo, un marco que tiene 160 muestras puede convertirse en el dominio de frecuencia usando una transformada de 256 puntos, tal como una transformada de Fourier rápida de 256 puntos ("FFT"), después de relleno o inclusión de muestras posteriores. Cuando se aplica una FFT inversa de 256 puntos para convertir el marco de nuevo en el dominio de tiempo, se obtendrán 256 muestras de dominio de tiempo. Por lo tanto, produce un extra noventa y seis muestras. Las noventa y seis muestras adicionales pueden ser solapadas y añadidas a muestras respectivas en las primeras noventa y seis muestras del siguiente marco. Esto se denomina a menudo técnica de superposición-adición. La transformación de la señal de voz, así como la implementación de técnicas tales como la técnica de complemento de superposición, puede aumentar significativamente la complejidad del descodificador global, especialmente para códecs que no incluyen ya componentes de transformada de frecuencia. Por consiguiente, los postfiltros del dominio de la frecuencia se usan típicamente solamente para los códecs vocales sinusoidales porque la aplicación de tales filtros a códecs no sinusoidales introduce demasiado retraso y complejidad. Los postfiltros de dominio de frecuencia también tienen típicamente menos flexibilidad para cambiar el tamaño de marco si el tamaño de marco de códec varía durante la codificación porque la complejidad de la técnica de adición de solapamiento discutida anteriormente puede llegar a ser prohibitiva si se encuentra un marco de tamaño diferente (tal como un marco con 80 muestras, en lugar de 160 muestras).

Aunque se describen anteriormente características particulares de entorno informático y características de códec de audio, pueden utilizarse una o más de las herramientas y técnicas con diversos tipos distintos de entornos informáticos y/o con diversos tipos diferentes de códecs. Por ejemplo, se pueden usar una o más de las técnicas postfiltro con códecs que no utilizan el modelo de codificación CELP, tales como códecs de modulación de código de pulso diferencial adaptativo, códecs de transformación y/u otros tipos de códecs. Como otro ejemplo, se pueden usar una o más de las técnicas postfiltro con códecs de banda única o códecs de sub-banda. Como otro ejemplo, se pueden aplicar una o más de las técnicas postfiltro a una única banda de un códec de múltiples bandas y/o a una señal sintetizada o no codificada que incluye contribuciones de múltiples bandas de un códec de múltiples bandas.

A. Ejemplo de postfiltros híbridos a corto plazo

En algunas realizaciones, un descodificador tal como el descodificador (600) mostrado en la figura 6 incorpora un filtro híbrido adaptativo de tiempo-frecuencia para postprocesamiento, o tal filtro se aplica a la salida del descodificador (600). Alternativamente, dicho filtro se incorpora o se aplica a la salida de algún otro tipo de descodificador de audio o herramienta de procesamiento, por ejemplo, un códec de voz descrito en otra parte de la presente solicitud.

Haciendo referencia a la Figura 6, en algunas implementaciones el postfiltro (694) de corto plazo es un filtro "híbrido" basado en una combinación de procesos de dominio de tiempo y de dominio de frecuencia. Los coeficientes del postfiltro (694) pueden ser diseñados flexibles y de manera eficiente principalmente en el dominio de la frecuencia, y los coeficientes pueden aplicarse al postfiltro (694) de corto plazo en el dominio del tiempo. La complejidad de este enfoque es típicamente más baja que los postfiltros estándar de dominio de frecuencia, y puede implementarse de una manera que introduce un retraso insignificante. Además, el filtro puede proporcionar más flexibilidad que los postfiltros de dominio de tiempo tradicionales. Se cree que un filtro híbrido de este tipo puede mejorar significativamente la calidad del habla de salida sin requerir excesiva demora o complejidad del descodificador. Además, debido a que el filtro (694) se aplica en el dominio del tiempo, puede aplicarse a marcos de cualquier tamaño.

En general, el postfiltro (694) puede ser un filtro de respuesta de impulso finito (FIR), cuya respuesta en frecuencia es el resultado de procesos no lineales realizados en el logaritmo de un espectro de magnitud de un filtro de síntesis LPC. El espectro de magnitud del postfiltro puede diseñarse de manera que el filtro (694) sólo se atenúe en valles espectrales y, en algunos casos, por lo menos parte del espectro de magnitud se recorta para que sea plana alrededor de las regiones formantes. Como se discute a continuación, los coeficientes de postfiltro de FIR pueden obtenerse truncando una secuencia normalizada que resulta de la transformada de Fourier inversa del espectro de magnitud procesado.

El filtro (694) se aplica al habla reconstruida en el dominio del tiempo. El filtro puede aplicarse a toda la banda o a una sub-banda. Adicionalmente, el filtro puede usarse solo o en conjunción con otros filtros, tales como postfiltros a largo plazo y/o el filtro de mejora de frecuencia media discutido con más detalle a continuación.

El postfiltro descrito puede ser operado conjuntamente con códecs que utilizan diversas velocidades de bits, diferentes velocidades de muestreo y diferentes algoritmos de codificación. Se cree que el postfiltro (694) es capaz de producir una mejora de calidad significativa con respecto al uso de códecs de voz sin el postfiltro. Específicamente, se cree que el postfiltro (694) reduce el ruido de cuantificación perceptible en regiones de frecuencia en las que la potencia de señal es relativamente baja, es decir, en valles espectrales entre formantes. En estas regiones la relación señal/ruido es típicamente pobre. En otras palabras, debido a la señal débil, el ruido que está presente es relativamente más fuerte. Se cree que el postfiltro mejora la calidad global del habla atenuando el nivel de ruido en estas regiones.

Los coeficientes (638) LPC reconstruidos contienen a menudo información formante porque la respuesta en frecuencia del filtro de síntesis LPC sigue típicamente la envolvente espectral del habla de entrada. En consecuencia, se utilizan coeficientes (638) LPC para deducir los coeficientes del postfiltro a corto plazo. Debido a que los coeficientes (638) LPC cambian de un marco a la siguiente o sobre alguna otra base, los coeficientes postfiltro derivados de ellos también se adaptan de marco a marco o de alguna otra base.

En la figura 7 se ilustra una técnica para calcular los coeficientes de filtro para el postfiltro (694). El descodificador (600) de la figura 6 realiza la técnica. Alternativamente, otro descodificador o una herramienta de postfiltrado realiza la técnica.

El descodificador (600) obtiene un espectro LPC mediante relleno (715) cero de un conjunto de coeficientes (710) LPC $a(i)$, donde $i = 0, 1, 2, \dots, P$ y donde $a(0) = 1$. El conjunto de coeficientes (710) LPC se puede obtener a partir de una corriente de bits si se utiliza un códec de predicción lineal, tal como un códec CELP. Alternativamente, el conjunto de coeficientes (710) LPC se puede obtener analizando una señal de voz reconstruida. Esto puede hacerse incluso si el códec no es un códec de predicción lineal. P es el orden LPC de los coeficientes LPC $a(i)$ que se utilizarán para determinar los coeficientes postfiltro. En general, el relleno cero implica extender una señal (o espectro) con ceros para extender sus límites de tiempo (o banda de frecuencia). En el proceso, el relleno cero asigna una señal de longitud P a una señal de longitud N , donde $N > P$. En una implementación de códec de banda completa, P es diez para una velocidad de muestreo de ocho kHz y dieciséis para tasas de muestreo mayores de ocho kHz. Alternativamente, P es otro valor. Para los códecs de sub-banda, P puede ser un valor diferente para cada sub-banda. Por ejemplo, para una frecuencia de muestreo de dieciséis kHz utilizando la estructura de tres sub-bandas ilustrada en la figura 3, P puede ser diez para la banda (310) de baja frecuencia, seis para la banda (320) media y cuatro para la banda (330) alta. En una implementación, N es 128. Alternativamente, N es algún otro número, tal como 256.

El descodificador (600) realiza entonces una transformada N-punto, tal como una FFT (720), sobre los coeficientes de relleno cero, produciendo un espectro de magnitud $A(k)$. $A(k)$ es el espectro del filtro inverso LPC con relleno

cero, para $k = 0, 1, 2, \dots, N-1$. La inversa del espectro de magnitud (es decir, $1/|A(k)|$) da el espectro de magnitud del filtro de síntesis LPC.

El espectro de magnitud del filtro de síntesis LPC se convierte opcionalmente en el dominio (725) logarítmico para disminuir su rango de magnitud. En una implementación, esta conversión es la siguiente:

$$H(k) = \ln \frac{1}{|A(k)|}$$

donde \ln es el logaritmo natural. Sin embargo, se podrían usar otras operaciones para disminuir el rango. Por ejemplo, podría utilizarse una operación de logaritmo de base diez en lugar de una operación de logaritmo natural.

Tres operaciones no lineales opcionales se basan en los valores de $H(k)$: normalización (730), compresión (735) no lineal y recorte (740).

La normalización (730) tiende a hacer que el intervalo de $H(k)$ sea más consistente de marco a marco y banda a banda. Tanto la normalización (730) como la compresión (735) no lineal reducen el rango del espectro de magnitud no lineal de manera que la señal de voz no se altera demasiado por el postfiltro. Alternativamente, se podrían usar técnicas adicionales y/u otras para reducir el rango del espectro de magnitud.

En una implementación, la normalización (730) inicial se realiza para cada banda de un códec de múltiples bandas de la siguiente manera:

$$\hat{H}(k) = H(k) - H_{\min} + 0.1$$

donde H_{\min} es el valor mínimo de $H(k)$, para $k = 0, 1, 2, \dots, N-1$.

La normalización (730) se puede realizar para un códec de banda completa como sigue:

$$\hat{H}(k) = \frac{H(k) - H_{\min}}{H_{\max} - H_{\min}} + 0.1$$

donde H_{\min} es el valor mínimo de $H(k)$, y H_{\max} es el valor máximo de $H(k)$, para $k = 0, 1, 2, \dots, N-1$. En ambas ecuaciones de normalización anteriores, se añade un valor constante de 0,1 para evitar que los valores máximo y mínimo de $H(k)$ sean 1 y 0, respectivamente, haciendo más efectiva la compresión no lineal. Alternativamente, se pueden utilizar otros valores constantes, u otras técnicas, para evitar valores cero.

La compresión (735) no lineal se realiza para ajustar adicionalmente el intervalo dinámico del espectro no lineal como sigue:

$$H_c(k) = \beta * |\hat{H}(k)|^\gamma$$

donde $k = 0, 1, \dots, N - 1$. En consecuencia, si se utilizó una FFT de 128 puntos para convertir los coeficientes en el dominio de frecuencia, entonces $k = 0, 1, \dots, 127$. Además, $\beta = \eta * (H_{\max} - H_{\min})$, con η y γ tomadas como factores constantes elegidos apropiadamente. Los valores de η y γ pueden elegirse de acuerdo con el tipo de códec de voz y la velocidad de codificación. En una implementación, los parámetros η y γ se eligen experimentalmente. Por ejemplo, η se elige como un valor del intervalo de 0,125 a 0,135, y γ se elige entre el intervalo de 0,5 a 1,0. Los valores constantes se pueden ajustar según las preferencias. Por ejemplo, se obtiene un rango de valores constantes analizando la distorsión espectral prevista (principalmente alrededor de picos y valles) resultante de varios valores constantes. Típicamente, es deseable elegir un intervalo que no supere un nivel predeterminado de distorsión predicha. Los valores finales se eligen entonces entre un conjunto de valores dentro del rango usando los resultados de las pruebas de audición subjetiva. Por ejemplo, en un postfiltro con una frecuencia de muestreo de ocho kHz, η es 0,5 y γ es 0,125, y en un postfiltro con una frecuencia de muestreo de dieciséis kHz, η es 1,0 y γ es 0,135.

El recorte (740) se puede aplicar al espectro comprimido, $H_c(k)$, como sigue:

$$H_{pf}(k) = \begin{cases} \lambda * H_{promedio} & H_c(k) > \lambda * H_{promedio} \\ H_c(k) & \text{para todo lo demás} \end{cases}$$

donde $H_{promedio}$ es el valor medio de $H_c(k)$, y λ es una constante. El valor de λ se puede elegir de manera diferente según el tipo de códec de voz y la velocidad de codificación. En algunas implementaciones, λ se elige experimentalmente (tal como un valor de 0,95 a 1,1), y puede ajustarse según las preferencias. Por ejemplo, los valores finales de λ se pueden elegir utilizando los resultados de las pruebas de audición subjetiva. Por ejemplo, en un postfiltro con una frecuencia de muestreo de ocho kHz, λ es 1,1, y en postfiltro que opera a una frecuencia de muestreo de dieciséis kHz, λ es 0,95.

Esta operación de recorte limita los valores de $H_{pf}(k)$ a un máximo, o techo. En las ecuaciones anteriores, este máximo se representa como $\lambda * H_{promedio}$. Alternativamente, se usan otras operaciones para tapan los valores del espectro de magnitud. Por ejemplo, el techo podría basarse en el valor mediano de $H_c(k)$, en lugar del valor medio. Además, en lugar de recortar todos los valores altos de $H_c(k)$ a un valor máximo específico (tal como $\lambda * H_{promedio}$), los valores podrían recortarse según una operación más compleja.

El recorte tiende a dar como resultado coeficientes de filtro que atenúan la señal de voz en sus valles sin cambiar significativamente el espectro de voz en otras regiones, tales como regiones formantes. Esto puede evitar que el postfiltro distorsione los formantes de voz, produciendo así una salida de voz de mayor calidad. Además, el recorte puede reducir los efectos de la inclinación espectral porque el recorte aplana el espectro postfiltro reduciendo los grandes valores al valor tapado, mientras que los valores alrededor de los valles permanecen sustancialmente sin cambios.

Cuando se realizó la conversión al dominio logarítmico, el espectro de magnitud recortado resultante, $H_{pf}(k)$, se convierte (745) del dominio de registro al dominio lineal, por ejemplo, como sigue:

$$H_{pl}(k) = \exp(H_{pf}(k))$$

donde \exp es la función del logaritmo natural inverso.

Una transformada (750) de Fourier rápida inversa de N puntos se realiza en $H_{pl}(k)$, produciendo una secuencia temporal de $f(n)$, donde $n = 0, 1, \dots, N-1$ y N es el mismo que en la operación FFT (720) discutida anteriormente. Así, $f(n)$ es una secuencia de tiempo de N puntos.

En la Figura 7, los valores de $f(n)$ se truncan (755) ajustando los valores a cero para $n > M-1$, como sigue:

$$h(n) = \begin{cases} f(n) & n = 0, 1, 2, \dots, M-1 \\ 0 & n > M-1 \end{cases}$$

donde M es el orden del postfiltro a corto plazo. En general, un valor más alto de M produce un habla filtrada de mayor calidad. Sin embargo, la complejidad del postfiltro aumenta a medida que M aumenta. El valor de M se puede elegir teniendo en cuenta estas compensaciones. En una implementación, Mis diecisiete.

Los valores de $h(n)$ están opcionalmente normalizados (760) para evitar cambios repentinos entre marcos. Por ejemplo, esto se hace de la siguiente manera:

$$h_{pf}(n) = \begin{cases} 1 & n = 0 \\ h(n) / h(0) & n = 1, 2, 3, \dots, M-1 \end{cases}$$

Alternativamente, se utiliza alguna otra operación de normalización. Por ejemplo, se puede utilizar la siguiente operación:

$$h_n(n) = \frac{h(n)}{\sqrt{\sum_{n=0}^{M-1} h^2(n)}}$$

En una implementación donde la normalización proporciona coeficientes postfiltro $h_{pi}(n)$ (765), se aplica un filtro FIR con coeficientes de $h_{pi}(n)$ (765) al habla sintetizada en el dominio del tiempo. Por lo tanto, en esta implementación, el coeficiente de postfiltro de primer orden ($n = 0$) se establece en un valor de uno para cada marco para evitar desviaciones significativas de los coeficientes de filtro de un marco al siguiente.

- 5 A la vista de las muchas realizaciones posibles a las que pueden aplicarse los principios de nuestra invención, reivindicamos como nuestra invención todas las realizaciones que pueden entrar dentro del alcance de las siguientes reivindicaciones.

REIVINDICACIONES

1. Un procedimiento implementado por ordenador que comprende:

calcular un conjunto de coeficientes de filtro para su aplicación a una señal de audio reconstruida, en el que el cálculo del conjunto de coeficientes de filtro comprende:

5 realizar (720) una transformación de un conjunto de valores de dominio de tiempo inicial de un dominio de tiempo en un dominio de frecuencia, produciendo de este modo un conjunto de valores de dominio de frecuencia inicial;

realizar uno o más cálculos de dominio de frecuencia usando los valores de dominio de frecuencia para producir un conjunto de valores de dominio de frecuencia procesados, en donde los valores de dominio de frecuencia representan un espectro derivado de coeficientes de predicción lineal; y

10 realizar (750) una transformación de los valores de dominio de frecuencia procesados desde el dominio de frecuencia en el dominio de tiempo, produciendo de este modo un conjunto de valores de dominio de tiempo procesados

producir una señal de audio filtrada filtrando al menos una parte de la señal de audio reconstruida en un dominio de tiempo usando el conjunto de coeficientes (694) de filtro, en el que

15 realizar (740) uno o más cálculos de dominio de frecuencia usando los valores de dominio de frecuencia para producir un conjunto de valores de dominio de frecuencia procesados comprende recortar los valores de dominio de frecuencia en el dominio de frecuencia de tal manera que sólo se recortan los valores de dominio de frecuencia que exceden un valor de recorte máximo.

20 2. El método de la reivindicación 1, en el que la señal de audio filtrada representa una sub-banda de frecuencia de la señal de audio reconstruida.

3. El método de la reivindicación 1, en el que el cálculo del conjunto de coeficientes de filtro comprende:

truncar el conjunto de valores de dominio de tiempo en el dominio de tiempo.

4. El método de la reivindicación 1, en el que el procesamiento del conjunto de coeficientes de predicción lineal comprende reducir un intervalo de un espectro derivado del conjunto de coeficientes de predicción lineal.

25 5. El método de la reivindicación 1, en el que uno o más cálculos de dominio de frecuencia comprende uno o más cálculos en un dominio logarítmico.

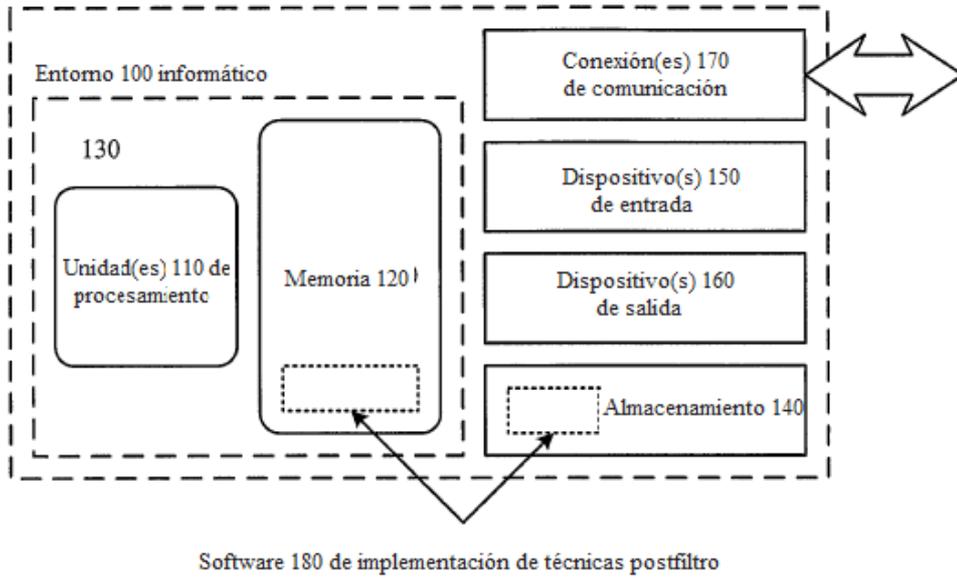


Figura 1

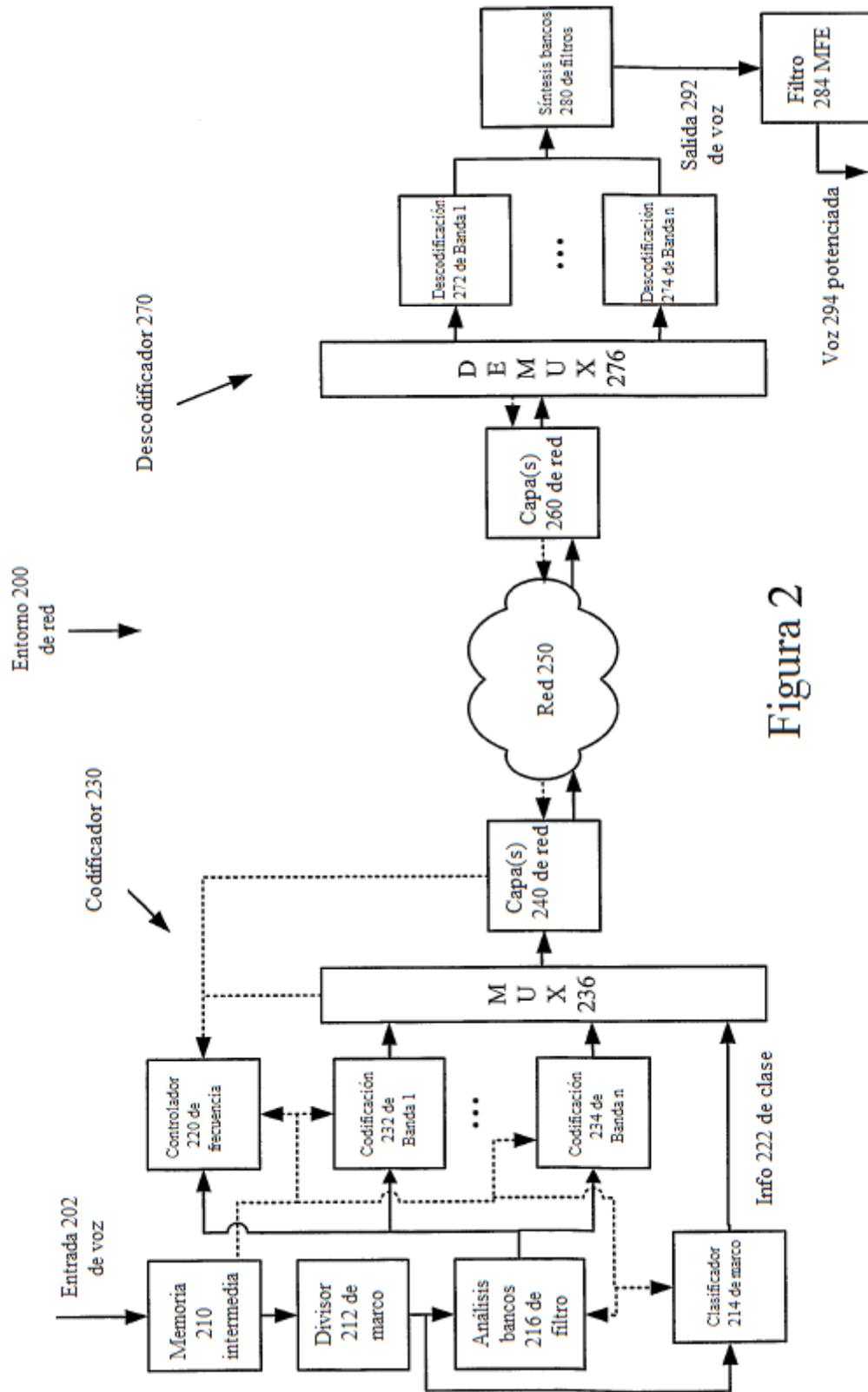


Figura 2

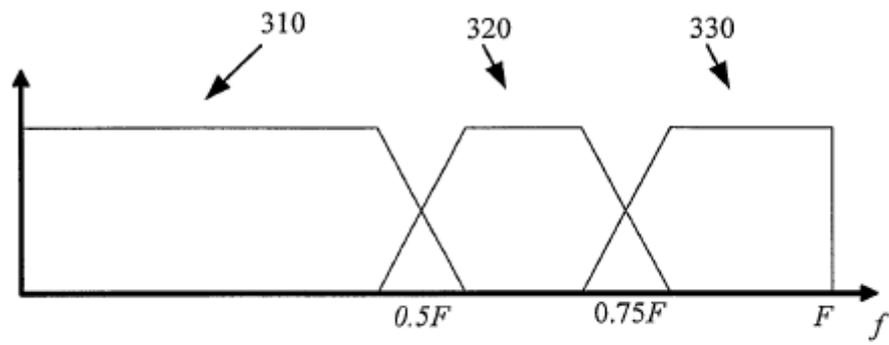


Figura 3

Figura 4

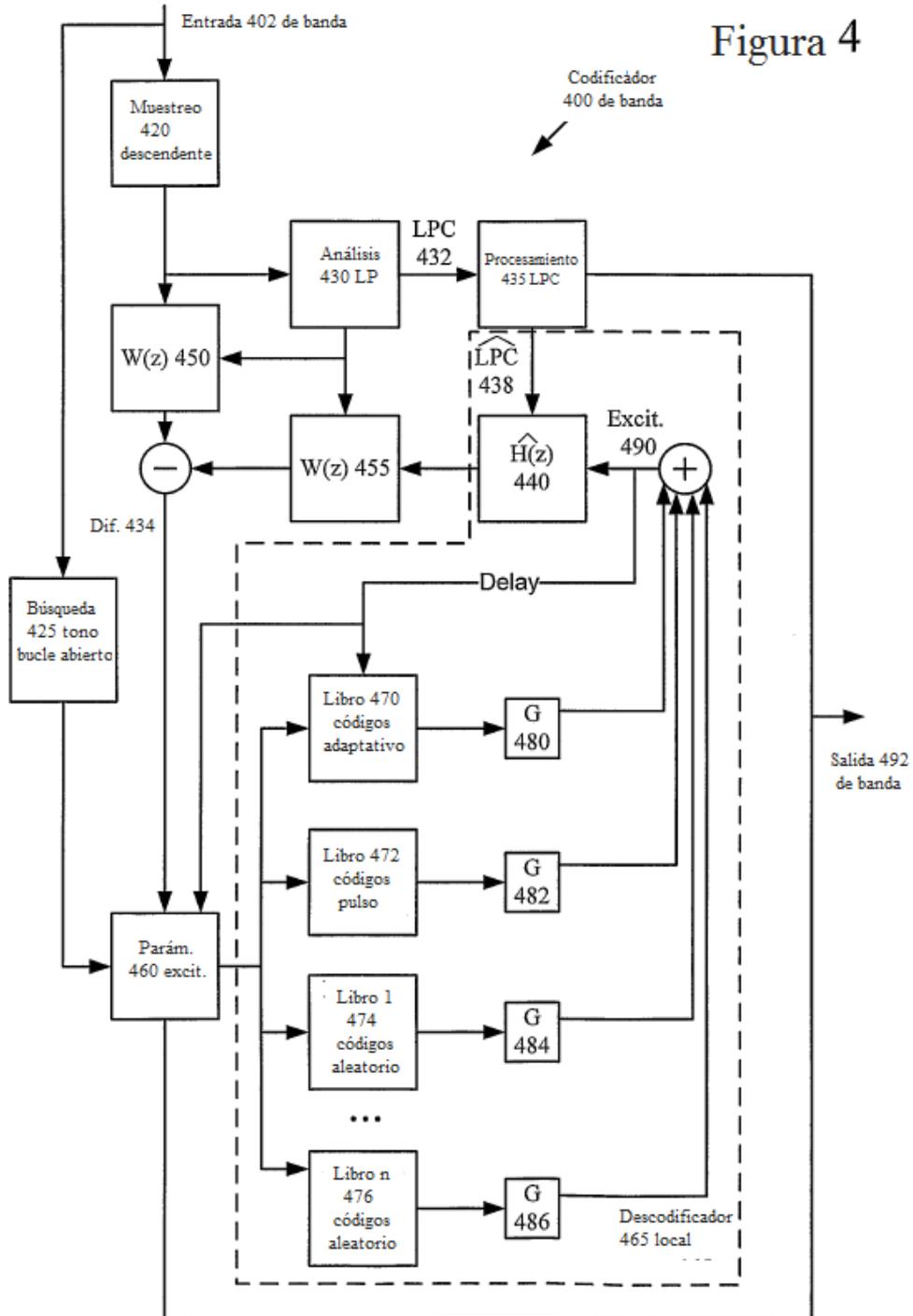
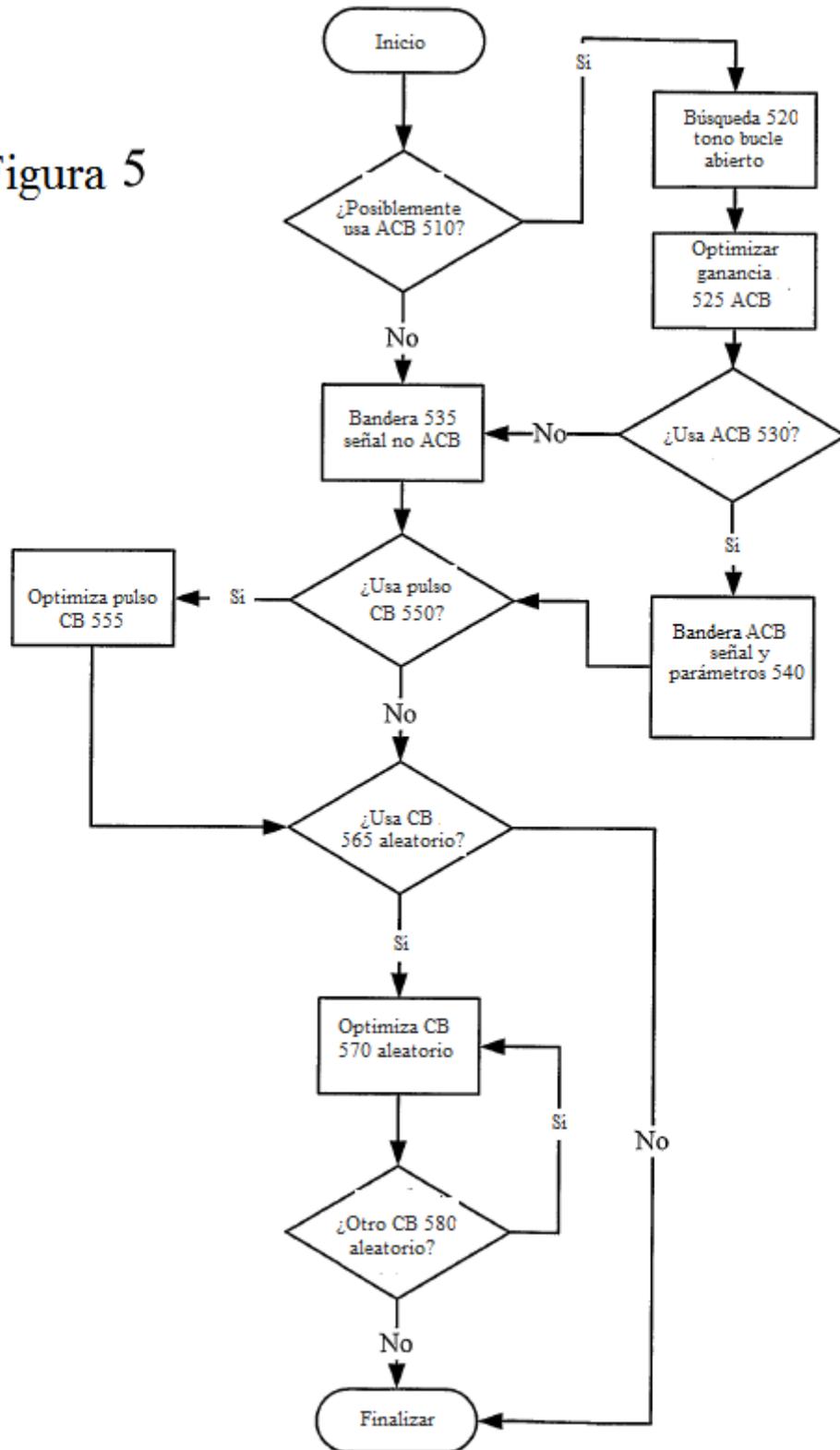


Figura 5



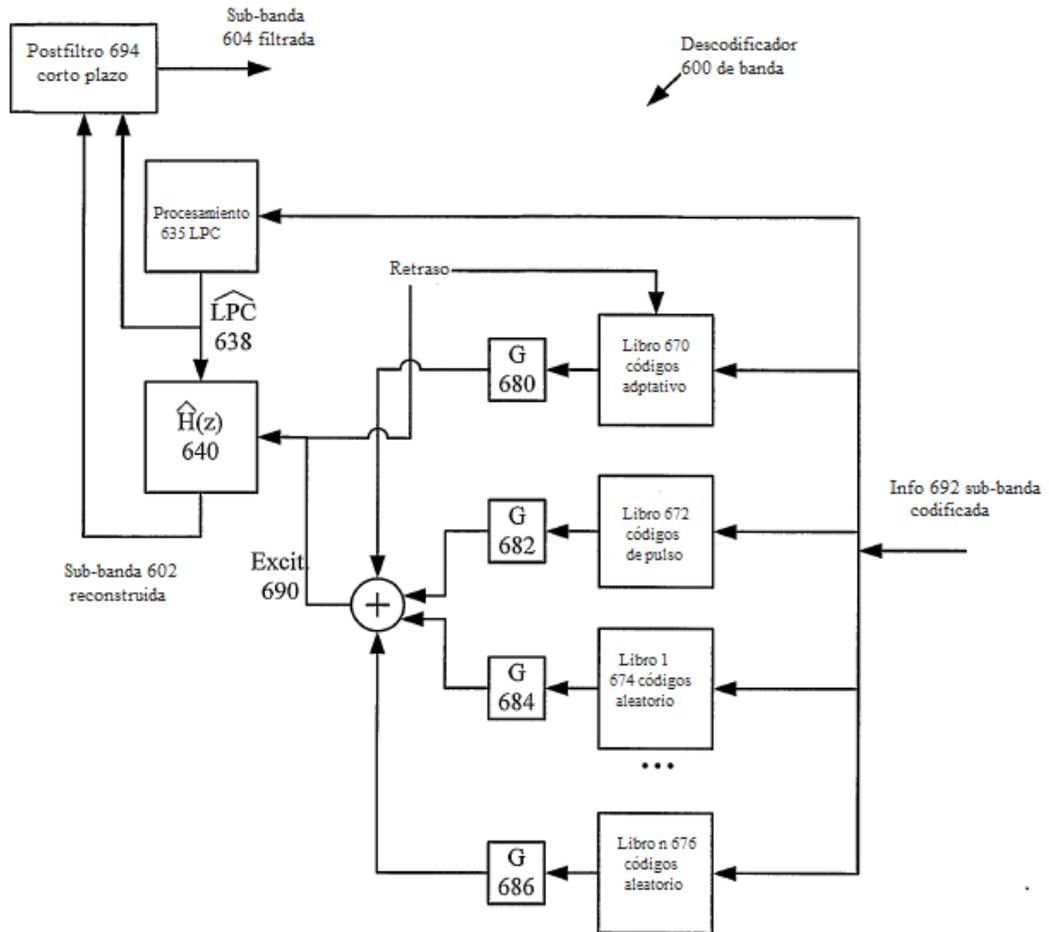


Figura 6

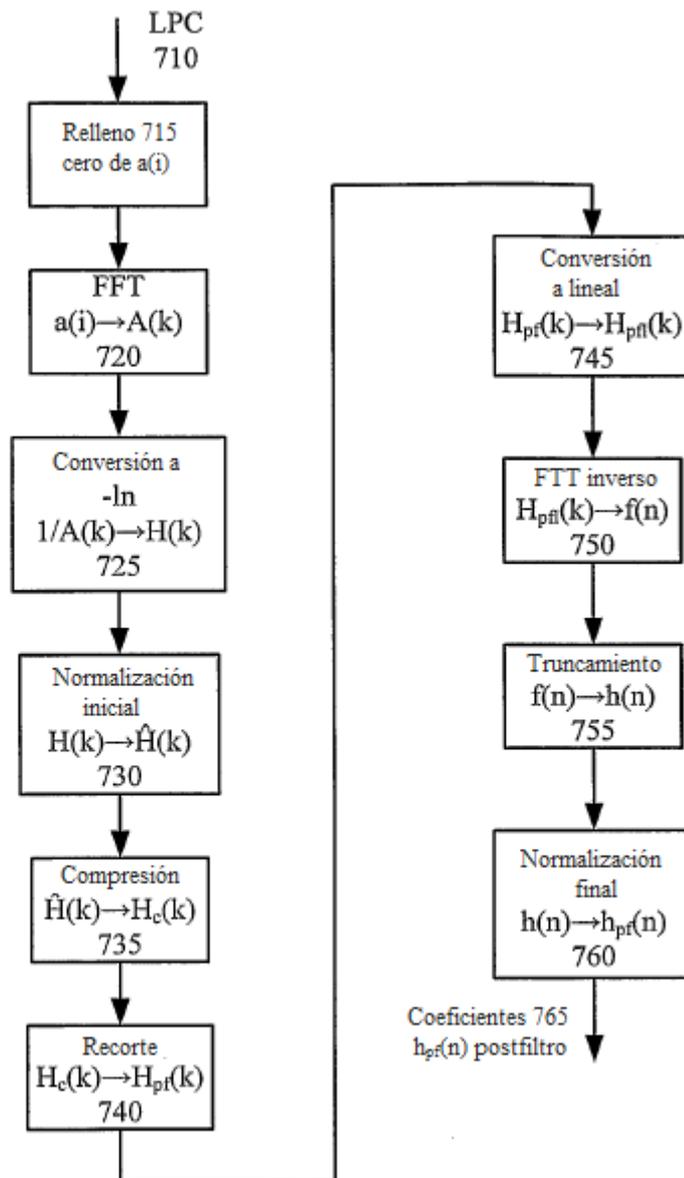


Figura 7