

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **3 020 834**

51 Int. Cl.:

G10L 21/038 (2013.01)

G10L 19/02 (2013.01)

G10L 19/08 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **09.09.2014** **E 23168838 (3)**

97 Fecha y número de publicación de la concesión europea: **22.01.2025** **EP 4258261**

54 Título: **Extensión adaptativa del ancho de banda y aparato para la misma**

30 Prioridad:

10.09.2013 US 201361875690 P
05.09.2014 US 201414478839

45 Fecha de publicación y mención en BOPI de la
traducción de la patente:
23.05.2025

73 Titular/es:

HUAWEI TECHNOLOGIES CO., LTD. (100.00%)
Huawei Administration Building
Shenzhen, Guangdong 518129, CN

72 Inventor/es:

GAO, YANG

74 Agente/Representante:

LEHMANN NOVO, María Isabel

ES 3 020 834 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

Extensión adaptativa del ancho de banda y aparato para la misma

5 Campo técnico

La presente invención pertenece, en general, al campo del procesamiento del habla y, en particular, a la extensión adaptativa del ancho de banda y al aparato para la misma.

10 Antecedentes

En sistemas de comunicación modernos con señales digitales de audio/habla, una señal digital se comprime en un codificador; la información comprimida (corriente de bits) se puede paquetizar y enviar a un decodificador a través de un canal de comunicación, trama a trama. El sistema de codificador y decodificador, juntos, se denomina códec. Se puede usar compresión de habla/audio para reducir el número de bits que representan la señal de habla/audio, reduciendo por ello la tasa de bits necesaria para la transmisión. La tecnología de compresión de habla/audio se puede clasificar, en general, en codificación en el dominio del tiempo y codificación en el dominio de la frecuencia. La codificación en el dominio del tiempo se utiliza usualmente para codificar señales de habla o para codificar señales de audio a tasas de bits bajas. La codificación en el dominio de la frecuencia se utiliza usualmente para codificar señales de audio o para codificar señales de habla a tasas de bits altas. La Extensión del ancho de banda (BWE) puede ser una parte de la codificación en el dominio del tiempo o de la codificación en el dominio de la frecuencia a fin de generar una señal de banda alta a tasa de bits muy baja o a tasa de bits nula.

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

La codificación del habla difiere de otras formas de codificación de audio en que el habla es una señal mucho más sencilla que la mayoría de las otras señales de audio, y está disponible mucha más información estadística sobre las propiedades del habla. Como consecuencia, parte de la información auditiva, que es relevante en codificación de audio, puede ser innecesaria en el contexto de la codificación del habla. En la codificación del habla, el criterio más importante es la conservación de la inteligibilidad y "lo agradable" del habla, con una cantidad restringida de datos transmitidos.

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

La redundancia de las formas de onda del habla se puede considerar con respecto a varios tipos diferentes de señal del habla, tales como las señales del habla con voz y sin voz. Los sonidos con voz, p. ej., 'a', 'b', se deben esencialmente a vibraciones de las cuerdas vocales, y son oscilatorios. Por lo tanto, durante periodos cortos de tiempo, están bien modelados por sumas de señales periódicas, tales como sinusoides. En otras palabras, para el habla con voz, la señal del habla es esencialmente periódica. Sin embargo, esta periodicidad puede ser variable durante la duración de un segmento del habla y la forma de la onda periódica cambia usualmente de manera gradual de segmento a segmento. Una codificación del habla de tasa de bits baja podría beneficiarse mucho de la exploración de tal periodicidad. El periodo del habla con voz se denomina también tono, y la predicción del tono se llama a menudo Predicción a largo plazo (LTP). En contraste a esto, los sonidos sin voz, tales como 's', 'sh', son más semejantes al ruido. Esto se debe a que una señal del habla sin voz es más semejante a un ruido aleatorio y tiene una magnitud más pequeña de predictibilidad.

Tradicionalmente, todos los métodos de codificación del habla paramétricos, tales como la codificación en el dominio del tiempo, hacen uso de la redundancia intrínseca en la señal del habla para reducir la cantidad de información que se debe enviar y para estimar los parámetros de las muestras del habla de una señal a intervalos cortos. Esta redundancia surge principalmente de la repetición de las formas de onda del habla a una tasa cuasiperiódica y de la envolvente espectral de cambio lento de la señal del habla.

La redundancia de las formas de onda del habla se puede considerar con respecto a varios tipos diferentes de señal del habla, tales como con voz y sin voz. Aunque la señal del habla es esencialmente periódica para el habla con voz, esta periodicidad puede ser variable durante la duración de un segmento del habla y la forma de la onda periódica cambia usualmente de manera gradual de segmento a segmento. Una codificación del habla de tasa de bits baja podría beneficiarse mucho de la exploración de tal periodicidad. El periodo del habla con voz se denomina también tono, y la predicción del tono se llama a menudo Predicción a largo plazo (LTP). Como para el habla sin voz, la señal es más semejante a un ruido aleatorio y tiene una magnitud más pequeña de predictibilidad.

En cualquier caso, se puede usar codificación paramétrica para reducir la redundancia de los segmentos del habla separando de la componente de la envolvente espectral la componente de excitación de la señal del habla. La envolvente espectral que cambia lentamente se puede representar por Codificación de predicción lineal (LPC), denominada también Predicción a corto plazo (STP). Una codificación del habla de tasa de bits baja podría beneficiarse también mucho de la exploración de tal predicción a corto plazo. La ventaja de la codificación surge de la tasa lenta a la que cambian los parámetros. No obstante, es raro que los parámetros sean significativamente diferentes de los valores mantenidos en unos pocos milisegundos. Por consiguiente, a la tasa de muestreo de 8 kHz, 12,8 kHz o 16 kHz, el algoritmo de codificación del habla es tal que la duración nominal de la trama está en el intervalo de diez a treinta milisegundos. Una duración de la trama de veinte milisegundos es la elección más común.

Se usa ampliamente la codificación de audio basada en la tecnología de banco de filtros, p. ej., la codificación en el dominio de la frecuencia. En el procesamiento de señales, un banco de filtros es una agrupación de filtros de paso de banda que separa la señal de entrada en múltiples componentes, cada una llevando una única subbanda de frecuencia de la señal original. El proceso de descomposición realizado por el banco de filtros se denomina *análisis*, y a la salida del análisis del banco de filtros se hace referencia como a una señal de subbanda, con tantas subbandas como filtros hay en el banco de filtros. El proceso de reconstrucción se denomina *síntesis* del banco de filtros. En el procesamiento de señales digitales, la expresión banco de filtros se aplica también comúnmente a un banco de receptores. La diferencia es que los receptores también convierten por reducción las subbandas a una frecuencia central baja que puede volverse a muestrear a una tasa reducida. El mismo resultado se puede conseguir a veces submuestreando las subbandas de paso de banda. La salida del *análisis* del banco de filtros podría tener la forma de coeficientes complejos. Cada coeficiente complejo contiene un *elemento real* y un *elemento imaginario* que representa, respectivamente, el *término del coseno* y el *término del seno* para cada subbanda del banco de filtros.

En estándares bien conocidos más recientes, tales como G.723.1, G.729, G.718, se han adoptado la Tasa completa mejorada (EFR), el Codificador de voz de modo seleccionable (SMV), la Tasa múltiple adaptativa (AMR), la Banda ancha multimodo de tasa variable (VMR-WB) o la Banda ancha de tasa múltiple adaptativa (AMR-WB), y la Técnica de predicción lineal excitada de códigos ("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 usa principalmente para codificar señales del habla beneficiándose de las características específicas de la voz humana o del modelo de producción de la voz vocal humana. La codificación del habla CELP es un principio algorítmico muy popular en el área de la compresión del habla, aunque los detalles de la CELP para códecs diferentes podrían ser significativamente diferentes. Debido a su popularidad, el algoritmo CELP se ha usado en diversos estándares como ITU-T, MPEG, 3GPP y 3GPP2. Las variantes de la CELP incluyen la CELP algebraica, la CELP relajada, la CELP de bajo retardo y la predicción lineal excitada de suma de vectores, y otras. CELP es un término genérico para una clase de algoritmos y no para un códec particular.

El algoritmo CELP está basado en cuatro ideas principales. En primer lugar, se usa un modelo de filtro fuente de la producción del habla a través de la predicción lineal (LP). El modelo de filtro fuente de la producción del habla modela el habla como una combinación de una fuente sonora, 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 fuente de producción del habla, la fuente sonora, o señal de excitación, está modelada a menudo como un tren de impulsos periódico, para el habla con voz, o como ruido blanco, para el habla sin voz. En segundo lugar, se usa un libro de códigos adaptativo, y uno fijo, como la entrada (excitación) del modelo LP. En tercer lugar, se realiza una búsqueda de bucle cerrado en un "dominio perceptualmente ponderado". En cuarto lugar, se aplica la cuantificación de vectores (VQ).

El documento US2002128839A1 describe un método para generar una señal del habla de banda ancha a partir de una primera señal del habla de banda estrecha, que extiende la estructura armónica de la señal del habla durante los segmentos del habla con voz y presenta una cantidad estimada linealmente de energía del habla en la banda ancha de frecuencias.

El documento US2001044722A1 describe un método para mejorar las señales del habla que realiza una extensión del ancho de banda copiando partes seleccionadas de una señal de excitación de banda baja a una banda de frecuencia alta, mediante el cual dichas partes se pueden seleccionar basándose en un análisis del espectro de audio descodificado de banda baja y la información de tono disponible.

Ulrich Kornagel: "Spectral widening of the excitation signal for telephone-band speech enhancement" (2001) propone diferentes métodos para generar la señal de excitación de banda ancha a partir de una versión limitada de banda telefónica.

Compendio

Una realización de la presente invención describe un método para descodificar una corriente de bits de audio codificada y generar una extensión del ancho de banda de frecuencia en un descodificador. El método comprende descodificar la corriente de bits de audio para producir una señal de audio descodificada de banda baja y generar un espectro de excitación de banda baja correspondiente a una banda de baja frecuencia. Se selecciona una zona de

subbanda del interior de la banda de baja frecuencia usando un parámetro que indica la información de energía de una envolvente espectral de la señal de audio descodificada de banda baja. Se genera un espectro de excitación de banda alta para una banda de alta frecuencia copiando un espectro de excitación de subbanda desde la zona seleccionada de subbanda hasta una zona de subbanda alta correspondiente a la banda de alta frecuencia. Usando el espectro de excitación generado de banda alta, se genera una señal de audio de banda alta extendida aplicando una envolvente espectral de banda alta. La señal de audio de banda alta extendida se añade a la señal de audio descodificada de banda baja para generar una señal de salida de audio que tiene un ancho de banda de frecuencia extendido.

De acuerdo con una realización alternativa de la presente invención, un descodificador para descodificar una corriente de bits de audio codificada y generar un ancho de banda de frecuencia comprende una unidad de descodificación de banda baja configurada para descodificar la corriente de bits de audio para producir una señal de audio descodificada de banda baja y para generar un espectro de excitación de banda baja correspondiente a una banda de baja frecuencia. El descodificador incluye además una unidad de extensión del ancho de banda acoplada a la unidad de descodificación de banda baja. La unidad de extensión del ancho de banda comprende una unidad de selección de subbandas y una unidad copiadora. La unidad de selección de subbandas está configurada para seleccionar una zona de subbanda del interior de la banda de baja frecuencia usando un parámetro que indica la información de energía de una envolvente espectral de la señal de audio descodificada de banda baja. La unidad copiadora está configurada para generar un espectro de excitación de banda alta para una banda de alta frecuencia copiando un espectro de excitación de subbanda desde la zona seleccionada de subbanda hasta una zona de subbanda alta correspondiente a la banda de alta frecuencia.

De acuerdo con una realización alternativa de la presente invención, un descodificador para el procesamiento del habla comprende un procesador y un soporte de almacenamiento legible por ordenador, que almacena una programación para que la ejecute el procesador. La programación incluye instrucciones para descodificar la corriente de bits de audio a fin de producir una señal de audio descodificada de banda baja y generar un espectro de excitación de banda baja correspondiente a una banda de baja frecuencia. La programación incluye instrucciones para seleccionar una zona de subbanda del interior de la banda de baja frecuencia usando un parámetro que indica la información de energía de una envolvente espectral de la señal de audio descodificada de banda baja, y generar un espectro de excitación de banda alta para una banda de alta frecuencia copiando un espectro de excitación de subbanda desde la zona seleccionada de subbanda hasta una zona de subbanda alta correspondiente a la banda de alta frecuencia. La programación incluye además instrucciones para usar el espectro de excitación generado de banda alta para generar una señal de audio de banda alta extendida aplicando una envolvente espectral de banda alta, y añadir la señal de audio de banda alta extendida a la señal de audio descodificada de banda baja para generar una señal de salida de audio que tiene un ancho de banda de frecuencia extendido.

Una realización alternativa de la presente invención describe un método para descodificar una corriente de bits de audio codificada y generar una extensión del ancho de banda de frecuencia en un descodificador. El método comprende descodificar la corriente de bits de audio para producir una señal de audio descodificada de banda baja y generar un espectro de banda baja correspondiente a una banda de baja frecuencia, y seleccionar una zona de subbanda del interior de la banda de baja frecuencia usando un parámetro que indica la información de energía de una envolvente espectral de la señal de audio descodificada de banda baja. El método incluye además generar un espectro de banda alta copiando un espectro de subbanda desde la zona seleccionada de subbanda hasta una zona de subbanda alta, y usar el espectro generado de banda alta para generar una señal de audio de banda alta extendida aplicando energía de la envolvente espectral de banda alta. El método incluye además añadir la señal de audio de banda alta extendida a la señal de audio descodificada de banda baja para generar una señal de salida de audio que tiene un ancho de banda de frecuencia extendido.

Breve descripción de los dibujos

Para una comprensión más completa de la presente invención, y de sus ventajas, se hace referencia a continuación a las siguientes descripciones, consideradas en unión con los dibujos que se acompañan, en los que:

la figura 1 ilustra las operaciones realizadas durante la codificación de un habla original usando un codificador CELP convencional;

la figura 2 ilustra las operaciones realizadas durante la descodificación de un habla original usando un descodificador CELP en las realizaciones de implementación de la presente invención, como se describirá más adelante;

la figura 3 ilustra las operaciones realizadas durante la codificación de un habla original en un codificador CELP convencional;

la figura 4 ilustra un descodificador CELP básico correspondiente al codificador de la figura 5 en las realizaciones de implementación de la presente invención, como se describirá en lo que sigue;

las figuras 5A y 5B ilustran un ejemplo de codificación/descodificación con Extensión del ancho de banda (BWE), en donde la figura 5A ilustra las operaciones en el codificador con la información conexas de la BWE, mientras que la figura 5B ilustra las operaciones en el descodificador con la BWE;

5 las figuras 6A y 6B ilustran otro ejemplo de codificación/descodificación con una BWE sin transmitir información conexas, en donde la figura 6A ilustra las operaciones en un codificador, mientras que la figura 6B ilustra las operaciones en un descodificador;

10 la figura 7 ilustra un ejemplo de un espectro de excitación ideal para el habla con voz o la música armónica cuando se usa el tipo CELP de códec;

la figura 8 muestra un ejemplo de una extensión convencional del ancho de banda de un espectro de excitación descodificado para el habla con voz o la música armónica cuando se usa el tipo CELP de códec;

15 la figura 9 ilustra un ejemplo de una realización de la presente invención de la extensión del ancho de banda aplicada al espectro de excitación descodificado para el habla con voz o la música armónica cuando se usa el tipo CELP de códec;

20 la figura 10 ilustra las operaciones en un descodificador de acuerdo con las realizaciones de la presente invención para implementar el desplazamiento o la copia de subbandas para la BWE;

la figura 11 ilustra una realización alternativa del descodificador para implementar el desplazamiento o la copia de subbandas para la BWE;

25 la figura 12 ilustra las operaciones realizadas en un descodificador de acuerdo con las realizaciones de la presente invención;

las figuras 13A y 13B ilustran un descodificador que implementa la extensión del ancho de banda de acuerdo con las realizaciones de la presente invención;

30 la figura 14 ilustra un sistema de comunicación según una realización de la presente invención; y

la figura 15 ilustra un diagrama de bloques de un sistema de procesamiento que se puede usar para implementar los dispositivos y los métodos descritos en la presente memoria.

35 Descripción detallada de las realizaciones ilustrativas

En sistemas de comunicación modernos con señales digitales de audio/habla, una señal digital se comprime en un codificador, y la información comprimida o corriente de bits se puede paquetizar y enviar a un descodificador, trama a trama, a través de un canal de comunicación. El descodificador recibe y descodifica la información comprimida para obtener la señal digital de audio/habla.

45 La presente invención se refiere, en general, a la codificación de señales de habla/audio y a la extensión del ancho de banda de señales de habla/audio. En particular, se pueden usar las realizaciones de la presente invención para mejorar el estándar del codificador del habla ITU-T AMR-WB en el campo de la extensión del ancho de banda.

Algunas frecuencias son más importantes que otras. Las frecuencias importantes se pueden codificar con una gran resolución. Son significativas las pequeñas diferencias en estas frecuencias y se necesita un esquema de codificación que conserve estas diferencias. Por otro lado, las frecuencias menos importantes no tienen que ser exactas. Se puede usar un esquema de codificación más basta, incluso aunque algunos de los detalles más finos se perderán en la codificación. Un esquema típico de codificación más basta está basado en el concepto de Extensión del ancho de banda (BWE). Este concepto de la tecnología se denomina también Extensión de banda alta (HBE), Réplica de subbanda (SBR) o Replicación de banda espectral (SBR). Aunque el nombre podría ser diferente, todos tienen el significado similar de codificar/descodificar algunas subbandas de frecuencia (usualmente bandas altas) con poca asignación de tasa de bits (incluso con asignación nula de tasa de bits) o una tasa de bits significativamente menor que el enfoque normal de codificación/descodificación.

En la tecnología SBR, la estructura fina espectral en una banda de alta frecuencia se copia desde una banda de baja frecuencia y se puede añadir algo de ruido aleatorio. Entonces, la envolvente espectral en una banda de alta frecuencia se conforma usando información conexas transmitida de codificador a descodificador. El desplazamiento o la copia de bandas de frecuencia desde banda baja hasta banda alta es normalmente la primera etapa para la tecnología BWE.

65 Se describirán las realizaciones de la presente invención para mejorar la tecnología BWE usando un proceso adaptativo para seleccionar una banda de desplazamiento basada en el nivel de energía de la envolvente espectral.

La figura 1 ilustra las operaciones realizadas durante la codificación de un habla original usando un codificador CELP convencional.

La figura 1 ilustra un codificador CELP inicial convencional en el que se minimiza a menudo un error ponderado entre un habla sintetizada y un habla original usando un enfoque de análisis por síntesis, lo que significa que la codificación (análisis) se realiza optimizando perceptualmente la señal descodificada (síntesis) en un bucle cerrado.

El principio básico que aprovechan todos los codificadores del habla es el hecho de que las señales del habla son formas de onda altamente correlacionadas. Como ilustración, el habla se puede representar usando un modelo autorregresivo (AR) como en la Ecuación (11) siguiente.

$$X_n = \sum_{i=1}^L a_i X_{n-i} + e_n \quad (11)$$

En la Ecuación (11), cada muestra se representa como una combinación lineal de las L muestras previas más un ruido blanco. Los coeficientes de ponderación $a_1, a_2, \dots, a_L$ se denominan Coeficientes de predicción lineal (los LPC). Para cada trama, los coeficientes de ponderación $a_1, a_2, \dots, a_L$ se eligen de manera que el espectro de $\{X_1, X_2, \dots, X_N\}$, generado usando el modelo anterior, coincide mucho con el espectro de la trama del habla de entrada.

Alternativamente, las señales del habla pueden estar representadas también por una combinación de un modelo armónico y un modelo de ruido. La parte armónica del modelo es eficazmente una representación de las series de Fourier de la componente periódica de la señal. En general, para señales con voz, el modelo armónico, más el de ruido, del habla está compuesto por una mezcla tanto de armónicos como de ruido. La proporción de armónicos y ruido en un habla con voz depende de varios factores, incluyendo las características del hablante (p. ej., hasta qué punto la voz del hablante es normal o entrecortada); el carácter de los segmentos del habla (p. ej., hasta qué punto un segmento del habla es periódico) y de la frecuencia. Las frecuencias más altas del habla con voz tienen una proporción más alta 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 señales del habla. El modelo de predicción lineal es particularmente bueno en modelar la envolvente espectral del habla, mientras que el modelo de ruido armónico es bueno en modelar la estructura fina del habla. Los dos métodos se pueden combinar para aprovecharse de sus fortalezas relativas.

Como se ha indicado previamente, antes de la codificación CELP, la señal de entrada al micrófono del aparato se filtra y se muestrea, por ejemplo, a una tasa de 8.000 muestras por segundo. Cada muestra se cuantifica a continuación, por ejemplo, con 13 bits por muestra. El habla muestreada se segmenta en segmentos o tramas de 20 ms (p. ej., en este caso 160 muestras).

La señal del habla se analiza y se extraen su modelo LP, sus señales de excitación y su tono. El modelo LP representa la envolvente espectral del habla. Se convierte en un conjunto de coeficientes de frecuencias espectrales lineales (LSF), que es una representación alternativa de los parámetros de predicción lineal, puesto que los coeficientes LSF tienen buenas propiedades de cuantificación. Los coeficientes LSF se pueden cuantificar de modo escalar o, más eficientemente, se pueden cuantificar en vectores usando libros de códigos de vectores LSF previamente cualificados.

La excitación de códigos incluye un libro de códigos que comprende vectores de códigos, que tienen componentes que se eligen todas independientemente de manera que cada vector de códigos puede tener un espectro aproximadamente 'blanco'. Para cada subtrama del habla de entrada, cada uno de los vectores de códigos se filtra a través del filtro de predicción lineal a corto plazo y del filtro de predicción a largo plazo, y la salida se compara con las muestras del habla. En cada subtrama, se elige el vector de códigos, cuya salida es la que mejor coincide con el habla de entrada (error minimizado), para representar esa subtrama.

La excitación codificada comprende normalmente una señal similar al pulso o una señal similar al ruido, que está construida matemáticamente o guardada en un libro de códigos. El libro de códigos está disponible tanto para el codificador como para el descodificador de recepción. La excitación codificada, que puede ser un libro de códigos estocástico o fijo, puede ser un diccionario de cuantificación de vectores que está codificado de manera fija (implícita o explícitamente) en el códec. Tal libro de códigos fijo puede ser una predicción lineal excitada de códigos algebraicos o estar almacenado explícitamente.

Un vector de códigos del libro de códigos se aumenta a escala por una ganancia apropiada para hacer la energía igual a la energía del habla de entrada. Por consiguiente, la salida de la excitación codificada se aumenta a escala por una ganancia antes de pasar por los filtros lineales.

El filtro de predicción lineal a corto plazo 103 conforma el espectro 'blanco' del vector de códigos para asemejarse al espectro del habla de entrada. Equivalentemente, en el dominio del tiempo, el filtro de predicción lineal a corto plazo 103 incorpora correlaciones a corto plazo (correlación con muestras previas) en la secuencia blanca. El filtro que conforma la excitación tiene un modelo de todo polos de la forma $1/A(z)$ (filtro de predicción lineal a corto plazo 103), donde $A(z)$ se denomina el filtro de predicción y se puede obtener usando predicción lineal (p. ej., el algoritmo Levinson-Durbin). En una o más realizaciones, se puede usar un filtro de todo polos, puesto que es una buena representación del tracto vocal humano y puesto que es fácil de calcular por ordenador.

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

$$A(z) = \sum_{i=1}^P 1 + a_i \cdot z^{-i}, i=1,2,\dots,P \quad (12)$$

Como se ha descrito previamente, unas regiones del habla con voz presentan periodicidad a largo plazo. El filtro de tonos $1/(B(z))$ introduce en el espectro sintetizado este período, conocido como tono. La salida del filtro de predicción a largo plazo 105 depende del tono y de la ganancia del tono. En una o más realizaciones, el tono se puede estimar a partir de la señal original, la señal residual o la señal original ponderada. En una realización, la función de predicción a largo plazo ($B(z)$) se puede expresar usando la Ecuación (13) como sigue.

$$B(z) = 1 - G_p \cdot z^{-Tono} \quad (13)$$

El filtro de ponderación 110 está relacionado con el filtro anterior de predicción a corto plazo. Uno de los filtros de ponderación típico se puede representar como se describe en la Ecuación (14).

$$W(z) = \frac{A(z/\alpha)}{1 - \beta \cdot z^{-1}} \quad (14)$$

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

En otra realización, el filtro de ponderación $W(z)$ se puede obtener del filtro LPC usando la expansión del ancho de banda como se ilustra en una realización en la Ecuación (15) siguiente.

$$W(z) = \frac{A(z/\gamma_1)}{A(z/\gamma_2)} \quad (15),$$

en la Ecuación (15), $\gamma_1 > \gamma_2$, que son los factores con los que los polos son desplazados hacia el origen.

Por consiguiente, para cada trama del habla, se calculan por ordenador los LPC y el tono y se actualizan los filtros. Para cada subtrama del habla, el vector de códigos que produce la 'mejor' salida filtrada se elige para representar la subtrama. El valor cuantificado correspondiente de ganancia se tiene que transmitir al descodificador para una descodificación apropiada. Se tienen también que cuantificar los LPC y los valores del tono y enviar cada trama para reconstruir los filtros en el descodificador. Por consiguiente, se transmiten al descodificador el índice de excitación codificada, el índice de ganancia cuantificado, el índice paramétrico de predicción a largo plazo cuantificado y el índice paramétrico de predicción a corto plazo cuantificado.

La figura 2 ilustra las operaciones realizadas durante la descodificación de un habla original usando un descodificador CELP en las realizaciones de implementación de la presente invención, como se describirá en lo que sigue.

La señal del habla se reconstruye en el descodificador haciendo pasar los vectores de códigos recibidos a través de los filtros correspondientes. Por consiguiente, cada bloque, excepto el de posprocesamiento, tiene la misma definición que se describe en el codificador de la figura 1.

La corriente de bits CELP codificada es recibida y despaquetizada 80 en un dispositivo de recepción. Para cada subtrama recibida, el índice de excitación codificada recibido, el índice de ganancia cuantificado, el índice paramétrico de predicción a largo plazo cuantificado y el índice paramétrico de predicción a corto plazo cuantificado se utilizan para encontrar los parámetros correspondientes usando descodificadores correspondientes, por ejemplo, un descodificador de ganancia 81, un descodificador de predicción a largo plazo 82 y un descodificador de predicción a corto plazo 83. Por ejemplo, las posiciones y los signos de amplitud de los pulsos de excitación y el vector de códigos algebraicos de la excitación de códigos 402 se pueden determinar a partir del índice de excitación codificada recibido.

Haciendo referencia a la figura 2, el descodificador es una combinación de varios bloques que incluyen una excitación codificada 201, una predicción a largo plazo 203 y una predicción a corto plazo 205. El descodificador inicial incluye además un bloque de posprocesamiento 207 después de un habla sintetizada 206. El posprocesamiento puede comprender además posprocesamiento a corto plazo y posprocesamiento a largo plazo.

La figura 3 ilustra un codificador CELP convencional.

La figura 3 ilustra un codificador CELP básico usando un libro de códigos adaptativo adicional para mejorar la predicción lineal a largo plazo. La excitación se produce sumando las contribuciones desde un libro de códigos 307 adaptativo y una excitación de códigos 308, que puede ser un libro de códigos estocástico o fijo como se ha descrito previamente. Las entradas en el libro de códigos adaptativo comprenden versiones retardadas de la excitación. Esto hace posible codificar eficientemente señales periódicas tales como sonidos con voz.

Haciendo referencia a la figura 3, un libro de códigos 307 adaptativo comprende una excitación sintetizada anterior 304 o un ciclo de tonos de excitación anterior repetitiva en un período de tonos. Un retraso del tono se puede codificar en valor entero cuando es grande o largo. El retraso del tono se codifica a menudo en un valor fraccionario más preciso cuando es pequeño o corto. La información periódica del tono se emplea para generar la componente adaptativa de la excitación. Esta componente de excitación se aumenta a escala a continuación por una ganancia G_p 305 (denominada también ganancia de tono).

La predicción a largo plazo desempeña un papel muy importante para la codificación del habla con voz, puesto que el habla con voz tiene una periodicidad acusada. Los ciclos de tono adyacentes del habla con voz son similares entre sí, lo que significa matemáticamente que la ganancia de tono G_p en la siguiente expresión de excitación es alta o próxima a 1. La excitación resultante se puede expresar como en la Ecuación (16), como combinación de las excitaciones individuales.

$$e(n) = G_p \cdot e_p(n) + G_c \cdot e_c(n) \quad (16)$$

donde, $e_p(n)$ es una subtrama de las series de muestras indexadas por n , procedentes del libro de códigos 307 adaptativo que comprende la excitación anterior 304 a través del bucle de realimentación (figura 3). $e_p(n)$ se puede filtrar con paso bajo de modo adaptativo dado que la zona de baja frecuencia es a menudo más periódica o más armónica que la zona de alta frecuencia. $e_c(n)$ procede del libro de códigos de excitación codificada 308 (denominado también libro de códigos fijo) que es una contribución a la excitación actual. Además, $e_c(n)$ se puede mejorar también tal como usando mejora del filtrado de paso alto, mejora del tono, mejora de la dispersión, mejora del formante, y otras.

Para el habla con voz, la contribución de $e_p(n)$ a partir del libro de códigos 307 adaptativo puede ser dominante y la ganancia de tono G_p 305 está alrededor del valor de 1. La excitación está actualizada usualmente para cada subtrama. El tamaño de trama típico es 20 milisegundos y el tamaño de subtrama típico es 5 milisegundos.

Como se describe en la figura 1, la excitación codificada 308 fija se aumenta a escala por una ganancia G_c 306 antes de pasar por los filtros lineales. Las dos componentes de excitación aumentadas a escala desde la excitación codificada 108 fija y el libro de códigos 307 adaptativo se añaden unas con otras antes del filtrado a través del filtro de predicción lineal a corto plazo 303. Las dos ganancias (G_p y G_c) se cuantifican y se transmiten a un descodificador. Por consiguiente, el índice de excitación codificada, el índice de libro de códigos adaptativo, los índices de ganancia cuantificados y el índice paramétrico de predicción a corto plazo cuantificado se transmiten al dispositivo de audio de recepción.

La corriente de bits CELP codificada usando un dispositivo ilustrado en la figura 3 se recibe en un dispositivo de recepción. La figura 4 ilustra el descodificador correspondiente del dispositivo de recepción.

La figura 4 ilustra un descodificador CELP básico correspondiente al codificador en la figura 5. La figura 4 incluye un bloque de posprocesamiento 408 que recibe el habla sintetizada 407 desde el descodificador principal. Este descodificador es similar al de la figura 3, excepto en el libro de códigos 307 adaptativo.

Para cada subtrama recibida, el índice de excitación codificada recibido, el índice de ganancia de excitación codificada cuantificado, el índice de tono cuantificado, el índice de ganancia del libro de códigos adaptativo cuantificado y el índice paramétrico de predicción a corto plazo cuantificado se utilizan para encontrar los parámetros correspondientes usando descodificadores correspondientes, por ejemplo, un descodificador de ganancia 81, un descodificador de tono 84, un descodificador de ganancia 85 del libro de códigos adaptativo y un descodificador de predicción a corto plazo 83.

En diversas realizaciones, el descodificador CELP es una combinación de varios bloques y comprende una excitación codificada 402, un libro de códigos 401 adaptativo, una predicción a corto plazo 406 y un

posprocesamiento 408. Cada bloque, excepto el de posprocesamiento, tiene la misma definición que se describe en el codificador de la figura 3. El posprocesamiento puede incluir además posprocesamiento a corto plazo y posprocesamiento a largo plazo.

Como ya se ha mencionado, la CELP se usa principalmente para codificar la señal del habla beneficiándose de las características específicas de la voz humana o del modelo de producción de la voz vocal humana. A fin de codificar la señal del habla más eficientemente, dicha señal del habla se puede clasificar en clases diferentes y cada clase se codifica de modo diferente. La clasificación sin voz/con voz o la Decisión sin voz puede ser una clasificación importante y básica entre todas las clasificaciones de clases diferentes. Para cada clase, el filtro LPC o STP se usa siempre para representar la envolvente espectral. Pero puede ser diferente la excitación al filtro LPC. Las señales sin voz se pueden codificar con una excitación similar al ruido. Por otro lado, las señales con voz se pueden codificar con una excitación similar al pulso.

El bloque de excitación de códigos (al que se ha hecho referencia con la marca 308 en la figura 3 y 402 en la figura 4) ilustra el lugar del Libro de códigos fijo (FCB) para una codificación CELP general. Un vector de códigos seleccionado a partir del FCB se aumenta a escala por una ganancia señalada a menudo como G_c 306.

Las figuras 5A y 5B ilustran un ejemplo de codificación/descodificación con Extensión del ancho de banda (BWE). La figura 5A ilustra las operaciones en el codificador con la información conexas de la BWE, mientras que la figura 5B ilustra las operaciones en el descodificador con la BWE.

Una señal de banda baja 501 se codifica usando unos parámetros de banda baja 502. Los parámetros de banda baja 502 están cuantificados y el índice de cuantificación generado se puede transmitir a través de un canal de corrientes de bits 503. La señal de banda alta extraída de una señal de audio/habla 504 se codifica con una pequeña cantidad de bits usando los parámetros conexos de banda alta 505. Los parámetros conexos de banda alta cuantificados (índice de información conexas) se transmiten a través del canal de corrientes de bits 506.

Haciendo referencia a la figura 5B, en el descodificador, la corriente de bits de banda baja 507 se usa para producir una señal descodificada de banda baja 508. La corriente de bits conexas de banda alta 510 se usa para descodificar los parámetros conexos de banda alta 511. La señal de banda alta 512 se genera a partir de la señal de banda baja 508 con ayuda de los parámetros conexos de banda alta 511. La señal final de audio/habla 509 se produce combinando la señal de banda baja 508 y la señal de banda alta 512.

Las figuras 6A y 6B ilustran otro ejemplo de codificación/descodificación con una BWE sin transmitir información conexas. La figura 6A ilustra las operaciones en un codificador, mientras que la figura 6B ilustra las operaciones en un descodificador.

Haciendo referencia a la figura 6A, una señal de banda baja 601 se codifica usando unos parámetros de banda baja 602. Los parámetros de banda baja 602 están cuantificados para generar un índice de cuantificación, que se puede transmitir a través del canal de corrientes de bits 603.

Haciendo referencia a la figura 6B, en el descodificador, la corriente de bits de banda baja 604 se usa para producir una señal descodificada de banda baja 605. La señal de banda alta 607 se genera a partir de la señal de banda baja 605 sin ayuda de la transmisión de información conexas. La señal final de audio/habla 606 se produce combinando la señal de banda baja 605 y la señal de banda alta 607.

La figura 7 ilustra un ejemplo de un espectro de excitación ideal para el habla con voz o la música armónica cuando se usa el tipo CELP de códec.

El espectro de excitación 702 ideal es casi plano después de eliminar la envolvente espectral LPC 704. El espectro de excitación de banda baja 701 ideal se puede usar como referencia para la codificación de excitación de banda baja. El espectro de excitación de banda alta 703 ideal no está disponible en el descodificador. Teóricamente, el espectro de excitación de banda alta ideal, o sin cuantificar, podría tener casi el mismo nivel de energía que el espectro de excitación de banda baja.

En la práctica, el espectro de excitación sintetizado o descodificado no parece tan bueno como el espectro de excitación ideal mostrado en la figura 7.

La figura 8 muestra un ejemplo de un espectro de excitación descodificado para el habla con voz o la música armónica cuando se usa el tipo CELP de códec.

El espectro de excitación descodificado 802 es casi plano después de eliminar la envolvente espectral LPC 804. El espectro de excitación de banda baja descodificado 801 está disponible en el descodificador. La calidad del espectro de excitación de banda baja descodificado 801 empeora o se distorsiona más especialmente en la región en la que la energía de la envolvente es baja. Esto se debe a varias razones. Por ejemplo, las dos razones principales son que la codificación CELP de bucle cerrado destaca más en una zona de alta energía que en una zona de baja energía, y

que la coincidencia de las formas de onda para una señal de baja frecuencia es más fácil que para una señal de alta frecuencia, debido al cambio más rápido de la señal de alta frecuencia. Para la codificación CELP de baja tasa de bits, tal como la AMR-WB, la banda alta no se codifica usualmente, sino que se genera en el descodificador con la tecnología BWE. En este caso, el espectro de excitación de banda alta 803 se puede copiar simplemente desde el espectro de excitación de banda baja 801 y la envolvente de la energía espectral de banda alta se puede predecir o estimar a partir de la envolvente de la energía espectral de banda baja. Siguiendo un modo tradicional, el espectro de excitación generado de banda alta 803 después de 6.400 Hz se copia desde la subbanda justamente antes de 6.400 Hz. Esto puede ser bueno si la calidad del espectro es equivalente de 0 Hz a 6.400 Hz. Sin embargo, para un códec CELP de baja tasa de bits, la calidad del espectro puede variar mucho de 0 Hz a 6.400 Hz. La subbanda copiada desde la zona extrema de la banda de baja frecuencia, justamente antes de 6.400 Hz, puede ser de una mala calidad, lo que introduce entonces sonido ruidoso adicional en la zona de banda alta de 6.400 Hz a 8.000 Hz.

El ancho de banda de la banda de alta frecuencia extendida es usualmente mucho menor que el de la banda codificada de baja frecuencia. Por lo tanto, en diversas realizaciones, la mejor subbanda de la banda baja se selecciona y se copia en la zona de banda alta.

La subbanda de alta calidad existe posiblemente en cualquier lugar del interior de toda la banda de baja frecuencia. El lugar más posible de la subbanda de alta calidad está en el interior de la región correspondiente a la zona de alta energía espectral, la zona del formante espectral.

La figura 9 ilustra un ejemplo del espectro de excitación descodificado para el habla con voz o la música armónica cuando se usa el tipo CELP de códec.

El espectro de excitación descodificado 902 es casi plano después de eliminar la envolvente espectral LPC 904. El espectro de excitación de banda baja descodificado 901 está disponible en el descodificador, pero no está disponible en la banda alta 903. La calidad del espectro de excitación de banda baja descodificado 901 empeora o se distorsiona más especialmente en la región en la que la energía de la envolvente espectral 904 es más baja.

En el caso ilustrado de la figura 9, en una realización, la subbanda de alta calidad está situada alrededor de la primera zona del formante del habla (p. ej., alrededor de 2.000 Hz en esta realización a modo de ejemplo). En diversas realizaciones, la subbanda de alta calidad puede estar situada en cualquier lugar entre 0 y 6.400 Hz.

Después de determinar el lugar de la mejor subbanda, se copia desde el interior de la banda baja hasta el interior de la banda alta, como se ilustra además en la figura 9. El espectro de excitación de banda alta 903 se genera así copiando desde la subbanda seleccionada. La calidad perceptual de la banda alta 903 en la figura 9 suena mucho mejor que la banda alta 803 en la figura 8, debido al espectro de excitación mejorado.

En una o más realizaciones, si la envolvente del espectro de banda baja está disponible en el dominio de la frecuencia en el descodificador, la mejor subbanda se puede determinar buscando la energía de subbanda más alta a partir de todas las subbandas candidatas.

Alternativamente, en una o más realizaciones, si la envolvente del espectro en el dominio de la frecuencia no está disponible, el lugar de alta energía se puede determinar también a partir de cualquier parámetro que puede reflejar la envolvente de energía espectral o el pico del formante espectral. El mejor lugar de subbanda para la BWE corresponde al lugar del pico espectral más alto.

El intervalo de búsqueda del mejor punto de partida de la subbanda puede depender de la tasa de bits del códec. Por ejemplo, para un códec de tasa de bits muy baja, el intervalo de búsqueda puede ser de 0 a $6.400 - 1.600 = 4.800$ Hz (2.000 Hz a 4.800 Hz), suponiendo que el ancho de banda de la banda alta es 1.600 Hz. En otro ejemplo, para un códec de tasa de bits mediana, el intervalo de búsqueda puede ser de 2.000 Hz a $6.400 - 1.600 = 4.800$ Hz (2.000 Hz a 4.800 Hz), suponiendo que el ancho de banda de la banda alta es 1.600 Hz.

Dado que la envolvente espectral cambia lentamente de una trama a la siguiente trama, normalmente, se cambia lentamente el mejor punto de partida de la subbanda correspondiente a la energía más alta del formante espectral. A fin de evitar la fluctuación o el cambio frecuente del mejor punto de partida de la subbanda de una trama a otra trama, se puede aplicar algo de suavizado durante la misma región con voz en el dominio del tiempo, a menos que la energía pico espectral se cambie espectacularmente de una trama a la siguiente trama o llegue una nueva región con voz.

La figura 10 ilustra las operaciones en un descodificador de acuerdo con las realizaciones de la presente invención para implementar el desplazamiento o la copia de subbandas para la BWE.

La señal de banda baja en el dominio del tiempo 1002 se descodifica usando la corriente de bits 1001 recibida. La excitación en el dominio del tiempo de banda baja 1003 está disponible usualmente en el descodificador. A veces, está disponible también la excitación en el dominio de la frecuencia de banda baja. Si no está disponible, la

excitación en el dominio del tiempo de banda baja 1003 puede ser transformada en dominio de la frecuencia para conseguir la excitación en el dominio de la frecuencia de banda baja.

La envolvente espectral del habla con voz o la señal de música está representada a menudo por parámetros LPC. A veces, la envolvente espectral directa en el dominio de la frecuencia está disponible en el decodificador. En cualquier caso, la información 1004 de distribución de la energía se puede extraer de los parámetros LPC o de la envolvente espectral directa en el dominio de la frecuencia o de cualquier parámetro tal como el dominio DFT o el dominio FFT. Usando la información 1004 de distribución de la energía de banda baja, la mejor subbanda a partir de la banda baja se selecciona buscando el pico de energía relativamente alto. La subbanda seleccionada se copia a continuación desde la banda baja hasta la zona de banda alta. Una envolvente espectral predicha o estimada de banda alta se aplica a continuación a la zona de banda alta, o una excitación de banda alta en el dominio del tiempo 1005 pasa por un filtro de banda alta predicho o estimado que representa la envolvente espectral de banda alta. La salida del filtro de banda alta es la señal de banda alta 1006. La señal final de salida de habla/audio 1007 se obtiene combinando la señal de banda baja 1002 y la señal de banda alta 1006.

La figura 11 ilustra una realización alternativa del decodificador para implementar el desplazamiento o la copia de subbandas para la BWE.

A diferencia de la figura 10, la figura 11 supone que está disponible el espectro de banda baja en el dominio de la frecuencia. La mejor subbanda en la banda de baja frecuencia se selecciona simplemente buscando el pico de energía relativamente alto en el dominio de la frecuencia. Entonces, la subbanda seleccionada se copia desde la banda baja hasta la banda alta. Después de aplicar una envolvente espectral estimada de banda alta, se forma el espectro de banda alta 1103. El espectro final de habla/audio en el dominio de la frecuencia se obtiene combinando el espectro de banda baja 1102 y el espectro de banda alta 1103. La salida final de señales de habla/audio en el dominio del tiempo se produce transformando el espectro de habla/audio en el dominio de la frecuencia al dominio del tiempo.

Cuando el análisis y la síntesis de bancos de filtros están disponibles en el decodificador cubriendo el intervalo deseado del espectro, un algoritmo SBR puede realizar el desplazamiento de bandas de frecuencia copiando coeficientes de banda de baja frecuencia de la salida correspondientes a la banda baja seleccionada desde el análisis de bancos de filtros hasta la zona de banda de alta frecuencia.

La figura 12 ilustra las operaciones realizadas en un decodificador de acuerdo con las realizaciones de la presente invención.

Haciendo referencia a la figura 12, un método para decodificar en un decodificador una corriente de bits de audio codificada incluye recibir una corriente de bits de audio codificada. En una o más realizaciones, la corriente de bits de audio recibida ha sido codificada con la CELP. En particular, solamente la banda de baja frecuencia está codificada con la CELP. La CELP produce una calidad de espectro en una zona de energía espectral más alta relativamente mayor que en una zona de energía espectral más baja. Por consiguiente, las realizaciones de la presente invención incluyen la decodificación de la corriente de bits de audio para generar una señal de audio decodificada de banda baja y un espectro de excitación de banda baja correspondiente a una banda de baja frecuencia (recuadro 1210). Se selecciona una zona de subbanda del interior de la banda de baja frecuencia usando información de energía de una envolvente espectral de la señal de audio decodificada de banda baja (recuadro 1220). Se genera un espectro de excitación de banda alta para una banda de alta frecuencia copiando un espectro de excitación de subbanda desde la zona seleccionada de subbanda hasta una zona de subbanda alta correspondiente a la banda de alta frecuencia (recuadro 1230). Se genera una señal de salida de audio usando el espectro de excitación de banda alta (recuadro 1240). En particular, usando el espectro de excitación generado de banda alta, se genera una señal de audio de banda alta extendida aplicando una envolvente espectral de banda alta. La señal de audio de banda alta extendida se añade a la señal de audio decodificada de banda baja para generar la señal de salida de audio que tiene un ancho de banda de frecuencia extendido.

Como se ha descrito previamente usando las figuras 10 y 11, las realizaciones de la presente invención se pueden aplicar de modo distinto dependiendo de si está disponible la envolvente del espectro en el dominio de la frecuencia. Por ejemplo, si está disponible la envolvente del espectro en el dominio de la frecuencia, se puede seleccionar la subbanda con la energía de subbanda más alta. Por otro lado, si no está disponible la envolvente del espectro en el dominio de la frecuencia, la distribución de energía de la envolvente espectral se puede identificar a partir de los parámetros de codificación predictiva lineal (LPC), los parámetros del dominio de la Transformada de Fourier discreta (DFT) o del dominio de la Transformada de Fourier rápida (FFT). De modo similar, la información del pico del formante espectral, si está disponible (o es calculable por ordenador), se puede usar en alguna realización. Si solamente está disponible la excitación en el dominio del tiempo de banda baja, la excitación en el dominio de la frecuencia de banda baja se puede calcular por ordenador transformando la excitación en el dominio del tiempo de banda baja al dominio de la frecuencia.

En diversas realizaciones, la envolvente espectral se puede calcular por ordenador usando cualquier método conocido, como sabría un experto en la técnica. Por ejemplo, en el dominio de la frecuencia, la envolvente espectral

puede ser simplemente un conjunto de energías que representan energías de un conjunto de subbandas. De modo similar, en otro ejemplo, en el dominio del tiempo, la envolvente espectral puede estar representada por parámetros LPC. Los parámetros LPC pueden tener muchas formas tales como Coeficientes de reflexión, Coeficientes LPC, Coeficientes LSP y Coeficientes LSF en diversas realizaciones.

Las figuras 13A y 13B ilustran un descodificador que implementa la extensión del ancho de banda de acuerdo con las realizaciones de la presente invención.

Haciendo referencia a la figura 13A, un descodificador para descodificar una corriente de bits de audio codificada comprende una unidad de descodificación de banda baja 1310 configurada para descodificar la corriente de bits de audio a fin de generar un espectro de excitación de banda baja correspondiente a una banda de baja frecuencia.

El descodificador incluye además una unidad de extensión del ancho de banda 1320 acoplada a la unidad de descodificación de banda baja 1310 y que comprende una unidad de selección de subbandas 1330 y una unidad copiadora 1340. La unidad de selección de subbandas 1330 está configurada para seleccionar una zona de subbanda del interior de la banda de baja frecuencia usando información de energía de una envolvente espectral de la corriente de bits de audio descodificada. La unidad copiadora 1340 está configurada para generar un espectro de excitación de banda alta para una banda de alta frecuencia copiando un espectro de excitación de subbanda desde la zona seleccionada de subbanda hasta una zona de subbanda alta correspondiente a la banda de alta frecuencia.

Un generador de señales de banda alta 1350 está acoplado a la unidad copiadora 1340. El generador de señales de banda alta 1350 está configurado para aplicar una envolvente espectral predicha de banda alta a fin de generar una señal en el dominio del tiempo de banda alta. Un generador de salida está acoplado al generador de señales de banda alta 1350 y a la unidad de descodificación de banda baja 1310. El generador de salida 1360 está configurado para generar una señal de salida de audio combinando una señal en el dominio del tiempo de banda baja obtenida descodificando la corriente de bits de audio con la señal en el dominio del tiempo de banda alta.

La figura 13B ilustra una realización alternativa de un descodificador que implementa la extensión del ancho de banda.

Similar al de la figura 13A, el descodificador de la figura 13B incluye también una unidad de descodificación de banda baja 1310 y una unidad de extensión del ancho de banda 1320, que está acoplada a la unidad de descodificación de banda baja 1310, y que comprende una unidad de selección de subbandas 1330 y una unidad copiadora 1340.

Haciendo referencia a la figura 13B, el descodificador incluye además un generador de espectros de banda alta 1355, que está acoplado a la unidad copiadora 1340. El generador de señales de banda alta 1355 está configurado para aplicar energía de la envolvente espectral de banda alta a fin de generar un espectro de banda alta para la banda de alta frecuencia usando el espectro de excitación de banda alta.

Un generador de espectros de salida 1365 está acoplado al generador de espectros de banda alta 1355 y a la unidad de descodificación de banda baja 1310. El generador de espectros de salida está configurado para generar un espectro de audio en el dominio de la frecuencia combinando un espectro de banda baja obtenido descodificando la corriente de bits de audio desde la unidad de descodificación de banda baja 1310 con el espectro de banda alta desde el generador de espectros de banda alta 1355.

Un generador de señales de transformada inversa 1370 está configurado para generar una señal de audio en el dominio del tiempo por transformación inversa del espectro de audio en el dominio de la frecuencia al dominio del tiempo.

Los diversos componentes descritos en las figuras 13A y 13B se pueden implementar en hardware en una o más realizaciones. En algunas realizaciones, se pueden implementar en software y diseñar para funcionar en un procesador de señales.

Por consiguiente, las realizaciones de la presente invención se pueden usar para mejorar la extensión del ancho de banda en un descodificador que descodifica una corriente de bits de audio codificada CELP.

La figura 14 ilustra un sistema de comunicación 10 según una realización de la presente invención.

El sistema de comunicación 10 tiene unos dispositivos de acceso a audio 7 y 8 acoplados a una red 36 a través de unos enlaces de comunicación 38 y 40. En una realización, los dispositivos de acceso a audio 7 y 8 son dispositivos de voz sobre protocolo de internet (VOIP) y la red 36 es una red de área extensa (WAN), una red telefónica conmutada pública (PTSN) y/o la de internet. En otra realización, unos enlaces de comunicación 38 y 40 son conexiones de banda ancha inalámbricas y/o alámbricas. En una realización alternativa, los dispositivos de acceso a audio 7 y 8 son teléfonos celulares o móviles, los enlaces 38 y 40 son canales de telefonía móvil inalámbricos y la red 36 representa una red de telefonía móvil.

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

En las realizaciones de la presente invención, en las que el dispositivo de acceso a audio 7 es un dispositivo VOIP, algunos o todos los componentes en el interior del dispositivo de acceso a audio 7 están implementados en el interior de un aparato. En algunas realizaciones, sin embargo, el micrófono 12 y el altavoz 14 son unidades independientes, y la interfaz de micrófono 16, la interfaz de altavoz 18, el CÓDEC 20 y la interfaz de red 26 están implementados en el interior de un ordenador personal. El CÓDEC 20 puede estar implementado en software que se ejecuta en un ordenador o en un procesador específico, o mediante hardware específico, por ejemplo, sobre un circuito integrado de aplicación específica (ASIC). La interfaz de micrófono 16 está implementada por un convertidor analógico a digital (A/D), así como otra circuitería de interfaz situada en el interior del aparato y/o del ordenador. Igualmente, la interfaz de altavoz 18 está implementada por un convertidor digital a analógico y otra circuitería de interfaz situada en el interior del aparato y/o del ordenador. En las realizaciones adicionales, el dispositivo de acceso a audio 7 puede estar implementado y dividido de otros modos conocidos en la técnica.

En las realizaciones de la presente invención en las que el dispositivo de acceso a audio 7 es un teléfono celular o móvil, los elementos en el interior del dispositivo de acceso a audio 7 están implementados en el interior de un aparato celular. El CÓDEC 20 está implementado por software que se ejecuta en un procesador en el interior del aparato o por hardware específico. En las realizaciones adicionales de la presente invención, el dispositivo de acceso a audio puede estar implementado en otros dispositivos tales como sistemas de comunicación digitales inalámbricos y alámbricos de par a par, tales como intercomunicadores, y aparatos de radio. En aplicaciones tales como en dispositivos de audio para consumidores, el dispositivo de acceso a audio puede contener un CÓDEC solamente con un codificador 22 o un descodificador 24, por ejemplo, en un sistema de micrófono digital o un dispositivo de reproducción de música. En otras realizaciones de la presente invención, el CÓDEC 20 se puede usar sin micrófono 12 y altavoz 14, por ejemplo, en estaciones base celulares que acceden a la PTSN.

El procesamiento del habla para mejorar la clasificación sin voz/con voz descrita en diversas realizaciones de la presente invención se puede implementar en el codificador 22 o el descodificador 24, por ejemplo. El procesamiento del habla para mejorar la clasificación sin voz/con voz se puede implementar en hardware o software en diversas realizaciones. Por ejemplo, el codificador 22 o el descodificador 24 puede ser parte de un chip de procesamiento de señales digitales (DSP).

La figura 15 ilustra un diagrama de bloques de un sistema de procesamiento que se puede usar para implementar los dispositivos y los métodos descritos en la presente memoria. Unos dispositivos específicos pueden utilizar todos los componentes mostrados, o solamente un subconjunto de los componentes, y los niveles de integración pueden variar de dispositivo a dispositivo. Además, un dispositivo puede contener múltiples ejemplos de un componente, tales 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 similar. La unidad de procesamiento puede incluir una unidad central de procesamiento (CPU), una memoria, un dispositivo de almacenamiento másico, un adaptador de vídeo y una interfaz I/O conectada a un bus.

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

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

El adaptador de vídeo y la interfaz I/O proporcionan interfaces para acoplar dispositivos de entrada y salida externos a la unidad de procesamiento. Como se ilustra, los ejemplos de dispositivos de entrada y salida incluyen la pantalla

acoplada al adaptador de vídeo y el ratón/teclado/impresora acoplados a la interfaz I/O. Otros dispositivos pueden estar acoplados a la unidad de procesamiento, y se pueden utilizar tarjetas de interfaz adicionales o en menor cantidad. Por ejemplo, se puede usar una interfaz en serie, tal como un Bus en serie universal (USB) (no mostrado), a fin de proporcionar una interfaz para una impresora.

- 5 La unidad de procesamiento incluye también una o más interfaces de red, que pueden comprender enlaces alámbricos, tales como un cable de Ethernet o similar, y/o enlaces inalámbricos para acceder a nodos o a diferentes redes. La interfaz de red permite a la unidad de procesamiento comunicarse con unidades remotas a través de las redes. Por ejemplo, la interfaz de red puede proporcionar la comunicación inalámbrica a través de uno o más
- 10 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 extensa para el procesamiento de datos y las comunicaciones con dispositivos remotos, tales como otras unidades de procesamiento, Internet, instalaciones de almacenamiento remotas, o similares.
- 15 Aunque esta invención se ha descrito con referencia a las realizaciones ilustrativas, esta descripción no está destinada a interpretarse en un sentido limitativo. Diversas modificaciones y combinaciones de las realizaciones ilustrativas, así como otras realizaciones de la invención, serán evidentes para los expertos en la técnica tras hacer referencia a la descripción. Por ejemplo, se pueden combinar entre sí diversas realizaciones descritas anteriormente.
- 20 Aunque la presente invención y sus ventajas se han descrito con detalle, se debería entender que se pueden realizar en la presente memoria diversos cambios, sustituciones y modificaciones sin alejarse del alcance de la invención tal como se define mediante las reivindicaciones adjuntas. Por ejemplo, muchas de las características y funciones descritas anteriormente se pueden implementar en software, hardware o firmware, o una combinación de los mismos. Además, no se pretende que el alcance de la presente solicitud esté limitado a las realizaciones
- 25 particulares del proceso, máquina, fabricaciones, composiciones de materia, medios, métodos y etapas descritos en la memoria descriptiva. Como el experto en la técnica apreciará fácilmente a partir de la descripción de la presente invención, se pueden utilizar, según la presente invención, procesos, máquinas, fabricaciones, composiciones de materia, medios, métodos o etapas, existentes en la actualidad o a desarrollar más adelante, que realizan sustancialmente la misma función o consiguen sustancialmente el mismo resultado que las realizaciones
- 30 correspondientes descritas en la presente memoria. En consecuencia, se pretende que las reivindicaciones adjuntas incluyan dentro de su alcance dichos procesos, máquinas, fabricaciones, composiciones de materia, medios, métodos o etapas.

REIVINDICACIONES

1. Un método para descodificar una corriente de bits de audio codificada y generar una extensión del ancho de banda de frecuencia en un descodificador, comprendiendo el método:
5 descodificar la corriente de bits de audio para producir una señal de audio descodificada de banda baja y generar un espectro de excitación de banda baja correspondiente a una banda de baja frecuencia;
determinar una zona de subbanda del interior de la banda de baja frecuencia usando un parámetro que indica la información de energía de una envolvente espectral de la señal de audio descodificada de banda baja, en donde el
10 lugar de la zona de subbanda se corresponde con el lugar del pico espectral más alto, en donde la banda de baja frecuencia está entre 0 y 6400 Hz;
generar un espectro de excitación de banda alta para una banda de alta frecuencia copiando un espectro de excitación de subbanda desde la zona determinada de subbanda hasta una zona de subbanda alta correspondiente a la banda de alta frecuencia;
15 usar el espectro de excitación generado de banda alta para generar una señal de audio de banda alta extendida aplicando una envolvente espectral de banda alta; y
añadir la señal de audio de banda alta extendida a la señal de audio descodificada de banda baja para generar una señal de salida de audio que tiene un ancho de banda de frecuencia extendido.
- 20 2. El método de la reivindicación 1, en donde un intervalo de búsqueda del punto de partida de la zona de subbanda depende de una tasa de bits del códec, y el intervalo de búsqueda es una región de frecuencias en el interior de la banda de baja frecuencia, en donde el punto de partida se corresponde con la energía más alta del formante espectral.
- 25 3. El método de las reivindicaciones 1 o 2, en donde aplicar la envolvente espectral de banda alta comprende filtrar el espectro de excitación de banda alta usando un filtro de banda alta que representa una envolvente espectral de banda alta para obtener la señal de audio de banda alta extendida.
- 30 4. Un descodificador para procesamiento del habla que comprende:
un procesador; y
un soporte de almacenamiento legible por ordenador que almacena una programación para que la ejecute el procesador, incluyendo la programación instrucciones para:
descodificar la corriente de bits de audio para producir una señal de audio descodificada de banda baja y generar un
35 espectro de excitación de banda baja correspondiente a una banda de baja frecuencia,
determinar una zona de subbanda del interior de la banda de baja frecuencia usando un parámetro que indica la información de energía de una envolvente espectral de la señal de audio descodificada de banda baja, en donde el
lugar de la zona de subbanda se corresponde con el lugar del pico espectral más alto;
generar un espectro de excitación de banda alta para una banda de alta frecuencia copiando un espectro de
40 excitación de subbanda desde la zona de subbanda determinada hasta una zona de subbanda alta correspondiente a la banda de alta frecuencia,
usar el espectro de excitación de banda alta generado para generar una señal de audio de banda alta extendida aplicando una envolvente espectral de banda alta; y
añadir la señal de audio de banda alta extendida a la señal de audio descodificada de banda baja para generar una
45 señal de salida de audio que tenga un ancho de banda de frecuencia extendido.
- 50 5. El descodificador de la reivindicación 4, en donde un intervalo de búsqueda del punto de partida de la zona de subbanda depende de una tasa de bits del códec, y el intervalo de búsqueda es una región de frecuencias en el interior de la banda de baja frecuencia, en donde el punto de partida se corresponde con la energía más alta del formante espectral.
- 55 6. El descodificador de las reivindicaciones 4 o 5, en donde aplicar la envolvente espectral de banda alta comprende filtrar el espectro de excitación de banda alta usando un filtro de banda alta que representa una envolvente espectral de banda alta para obtener la señal de audio de banda alta extendida.
- 60 7. Un soporte de almacenamiento legible por ordenador que almacena instrucciones que, cuando son ejecutadas por un procesador, hacen que el procesador lleve a cabo las operaciones de una cualquiera de las reivindicaciones 1 a 3.
8. Un dispositivo de acceso a audio que comprende un CÓDEC con un descodificador, en donde el descodificador está configurado para implementar el método de una cualquiera de las reivindicaciones 1 a 3.
9. El dispositivo de acceso a audio de la reivindicación 8, en donde el codificador o el descodificador forma parte de un chip de procesamiento de señales digitales, DSP.
- 65 10. El dispositivo de acceso a audio de la reivindicación 8, en donde el CÓDEC se implementa mediante software que se ejecuta en un procesador, o mediante hardware específico.

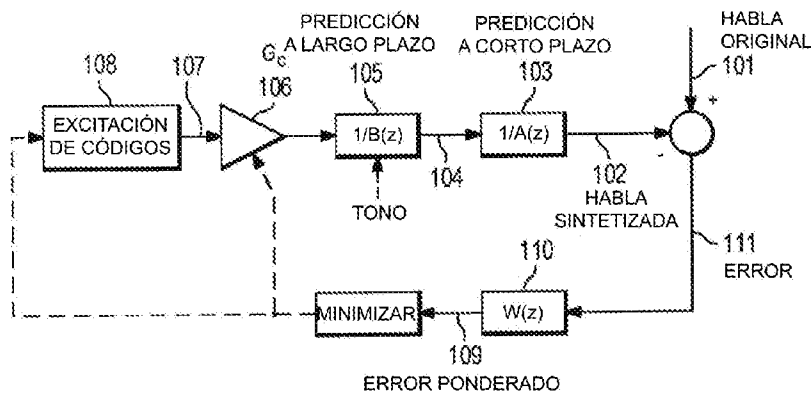


Figura 1

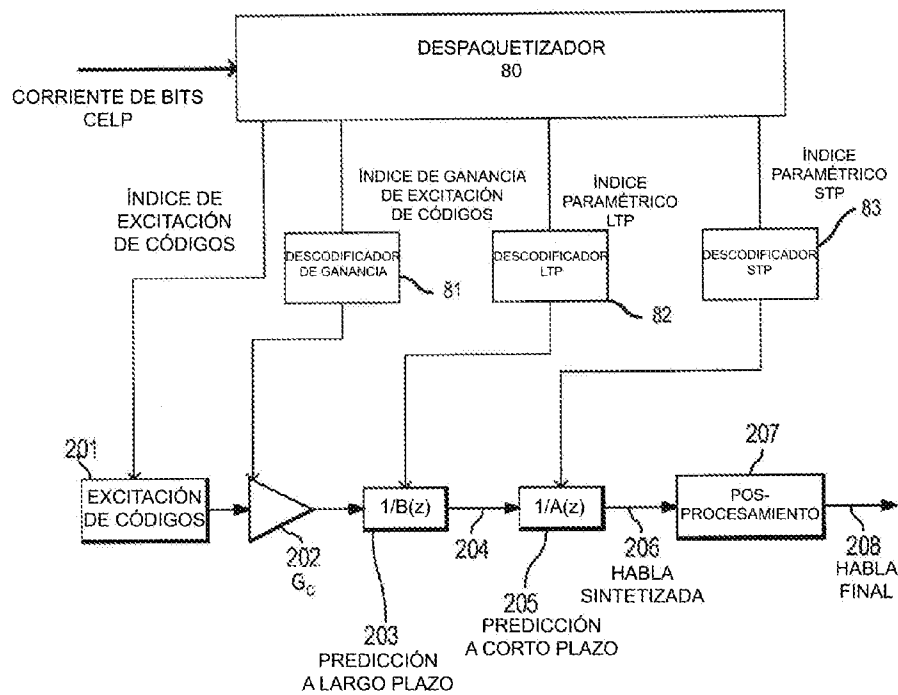


Figura 2

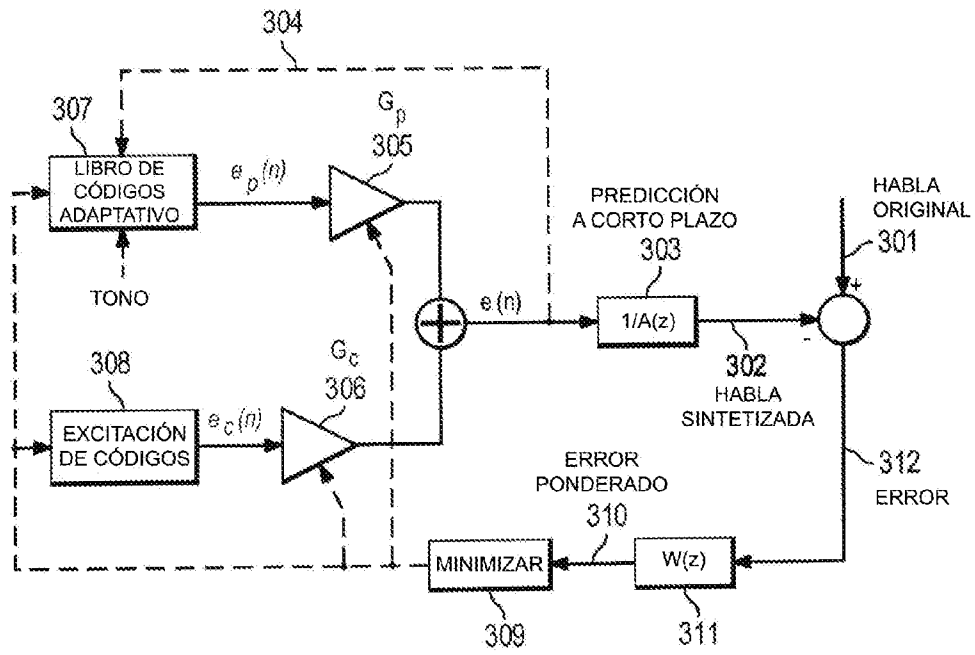


Figura 3

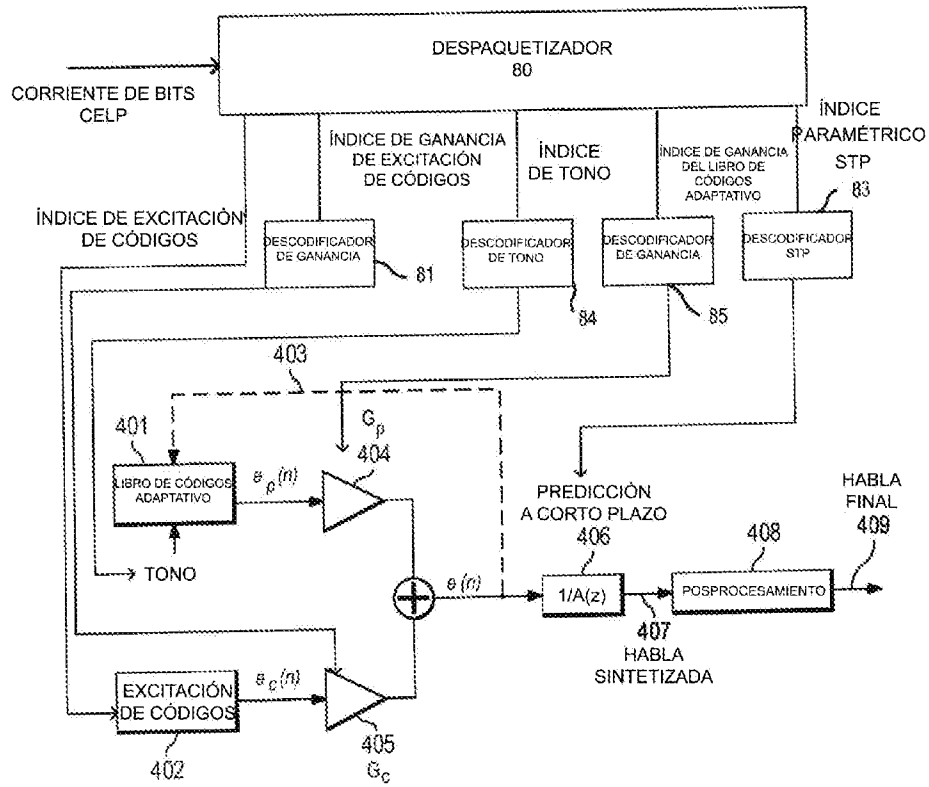


Figura 4

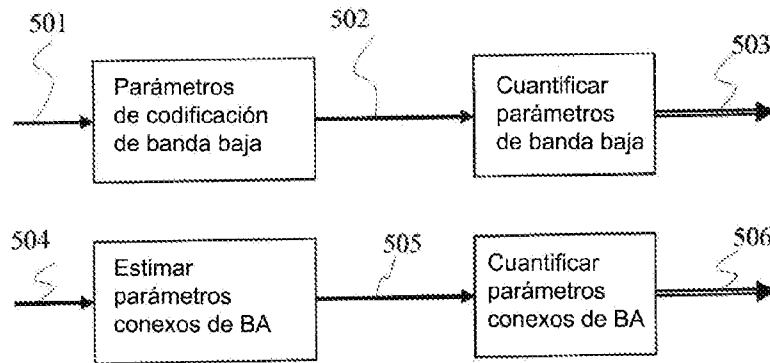


Figura 5A

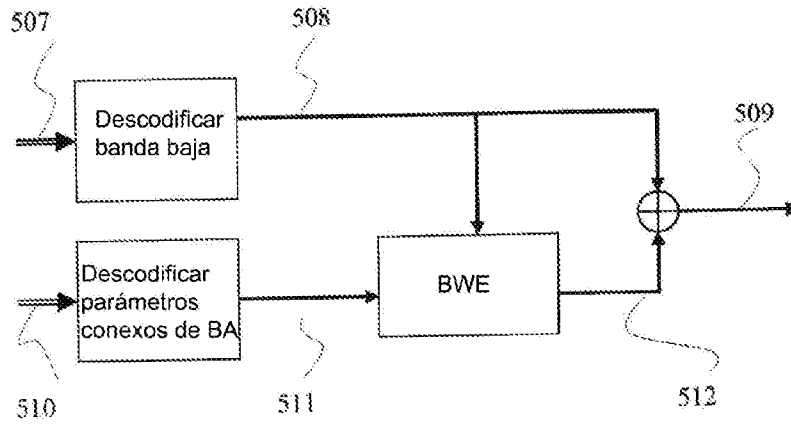


Figura 5B

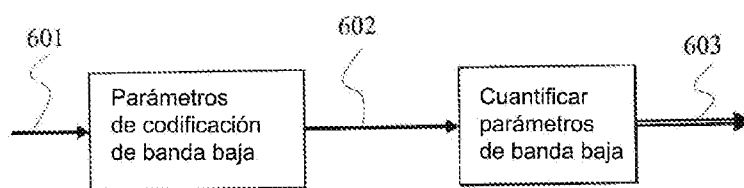


Figura 6A

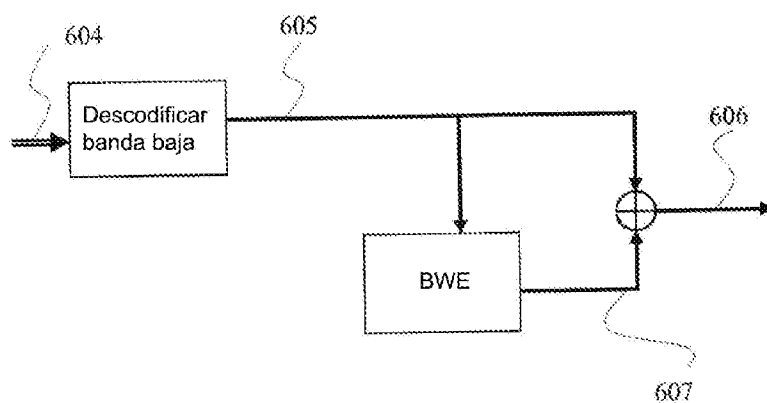


Figura 6B

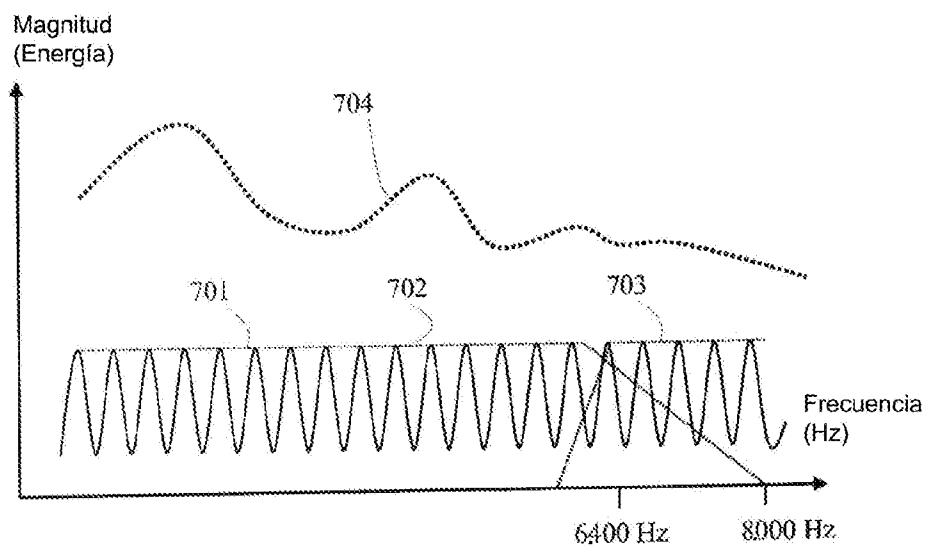


Figura 7

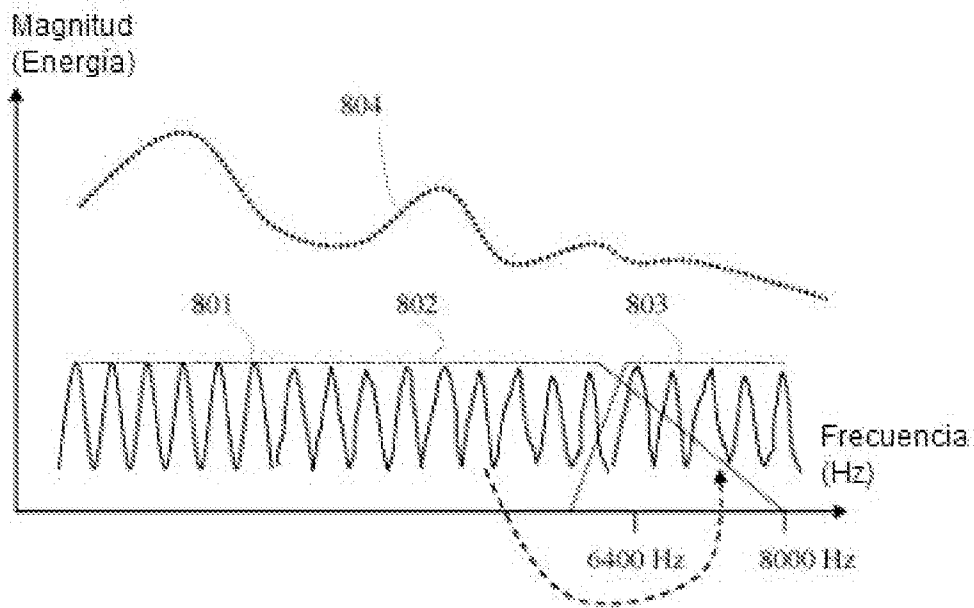


Figura 8

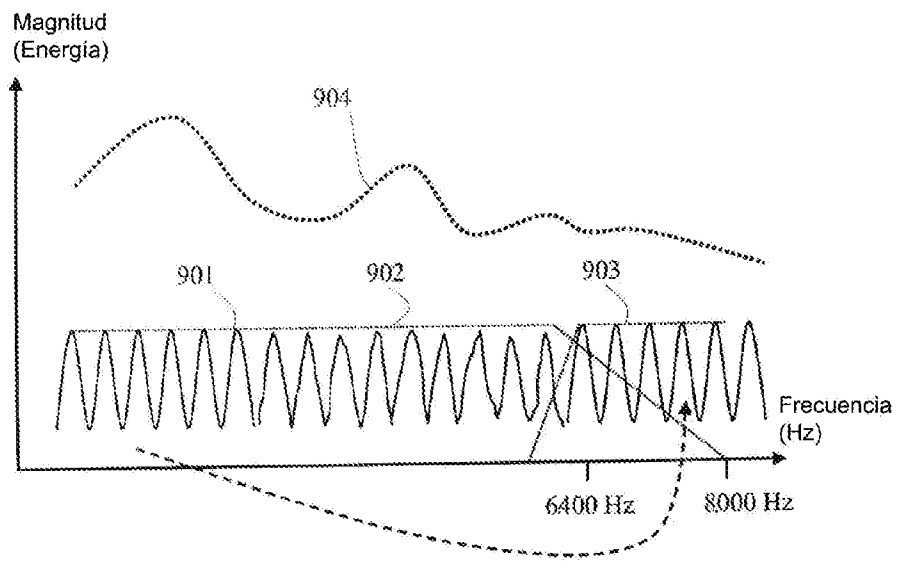


Figura 9

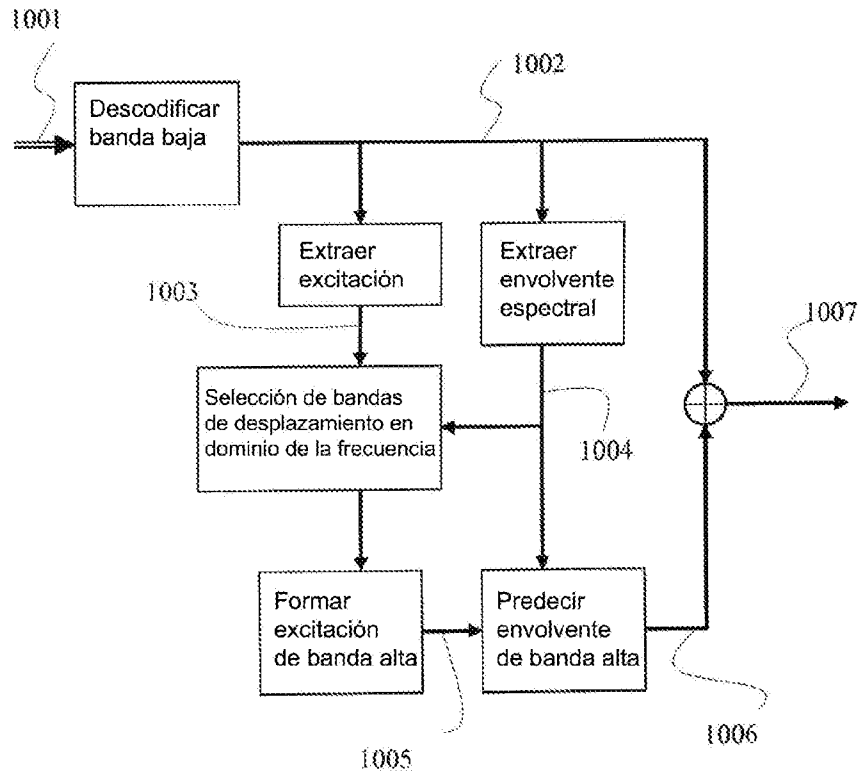
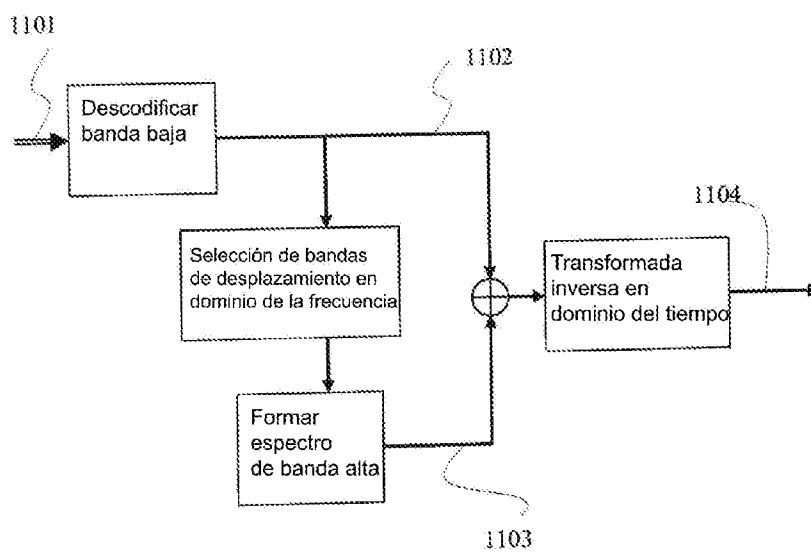


Figura 10

**Figura 11**

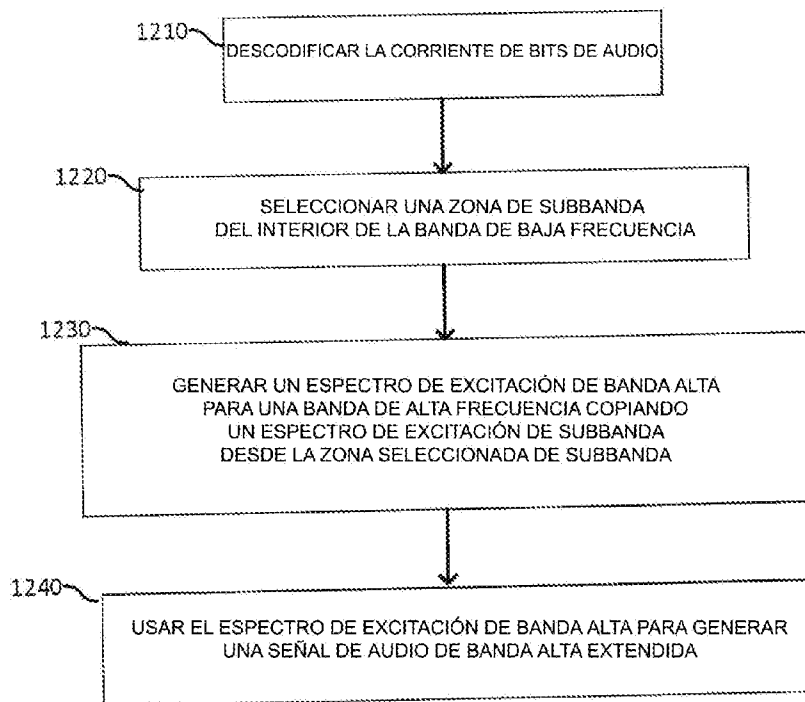


Figura 12

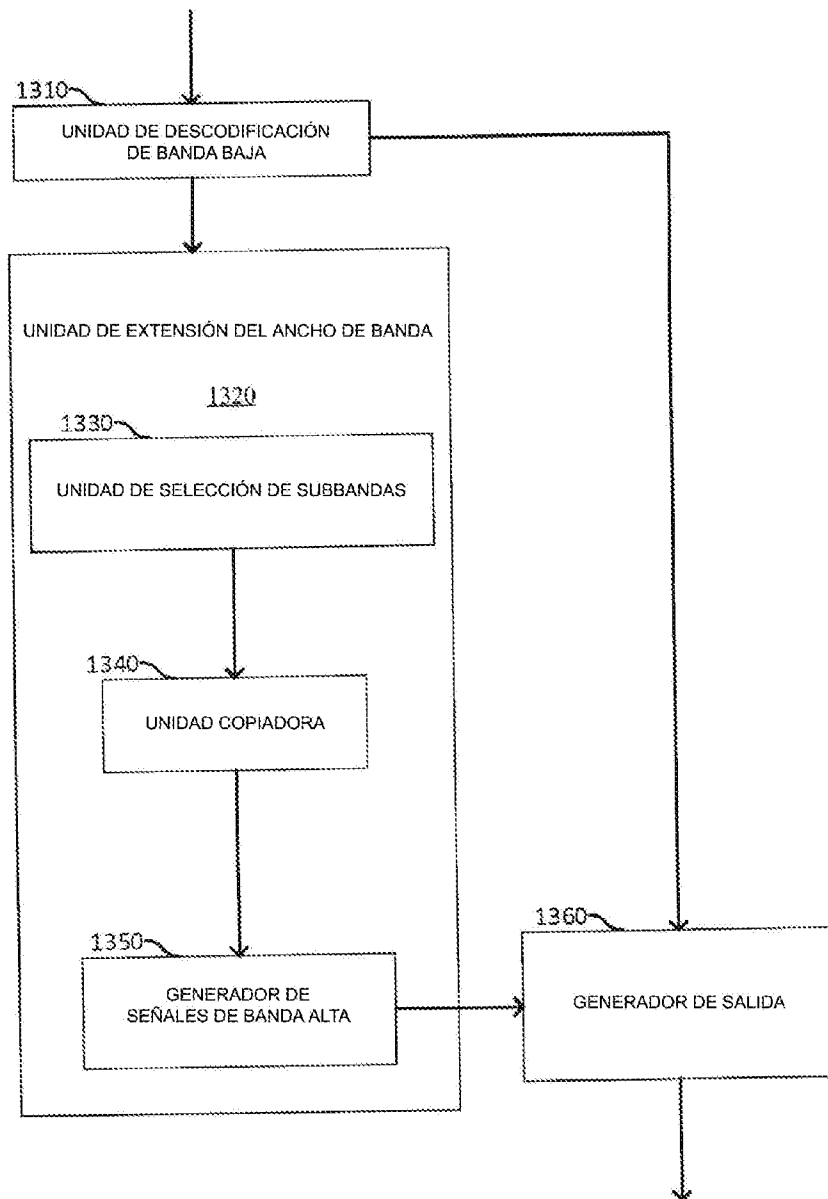


Figura 13A

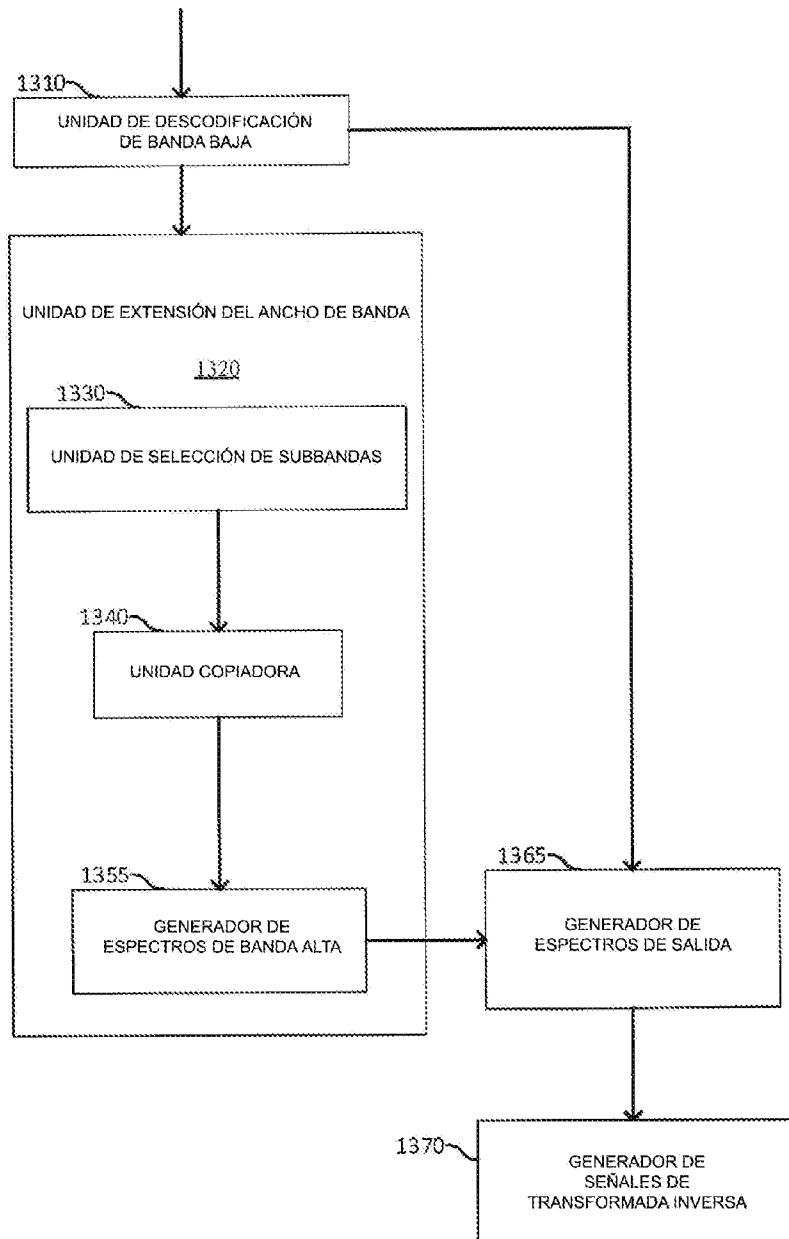


Figura 13B

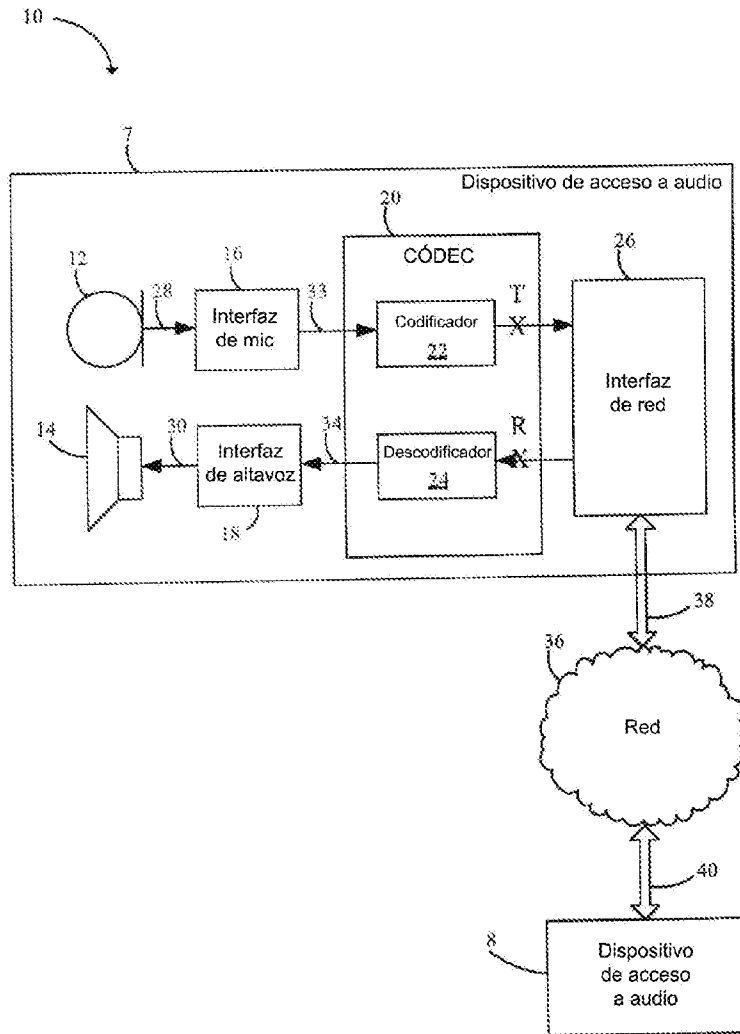


Figura 14

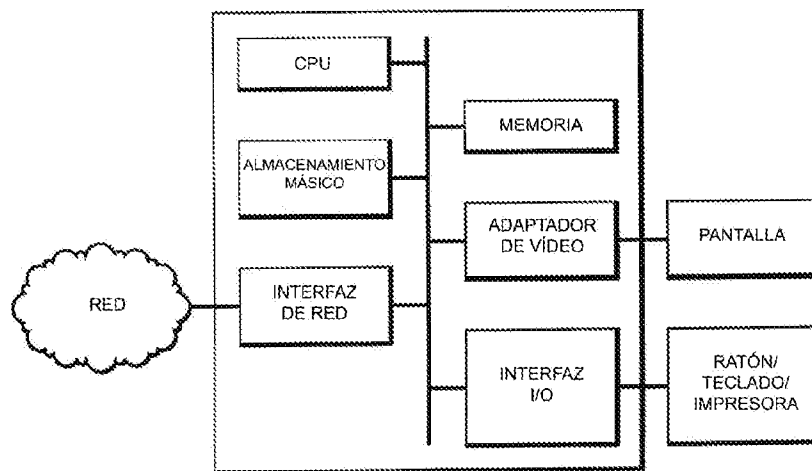


Figura 15