

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 978 967**

51 Int. Cl.:

**G10L 21/038** (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **04.02.2015** **E 17206569 (0)**

97 Fecha y número de publicación de la concesión europea: **10.04.2024** **EP 3327722**

54 Título: **Extensión de banda de frecuencia mejorada en un decodificador de señales de audiofrecuencia**

30 Prioridad:

**07.02.2014 FR 1450969**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**23.09.2024**

73 Titular/es:

**KONINKLIJKE PHILIPS N.V. (100.0%)**  
**High Tech Campus 52**  
**5656 AG Eindhoven, NL**

72 Inventor/es:

**KANIEWSKA, MAGDALENA y**  
**RAGOT, STÉPHANE**

74 Agente/Representante:

**ISERN JARA, Jorge**

ES 2 978 967 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 de banda de frecuencia mejorada en un decodificador de señales de audiofrecuencia

5 La presente invención se refiere al campo de la codificación/decodificación y del procesamiento de señales de audiofrecuencia (tales como señales de voz, de música u otras) para su transmisión o almacenamiento.

De manera más particular, la invención se refiere a un procedimiento y a un dispositivo de extensión de banda de frecuencia en un decodificador o un procesador realizando una mejora de la señal de audiofrecuencia.

10

Existen muchas técnicas para comprimir (con pérdida) una señal de audiofrecuencia, como la de voz o música.

Los métodos clásicos de codificación para aplicaciones conversacionales se clasifican generalmente como *codificación de forma de onda* (MIC, siglas de "Modulación por impulsos codificados", MICDA, siglas de "Modulación por impulsos codificados diferencial adaptativa", codificación por transformada...), *codificación paramétrica* (LPC, siglas en inglés de "Linear Predictive Coding" o Codificación predictiva lineal, codificación sinusoidal...) y codificación paramétrica híbrida con cuantificación de parámetros por "análisis por síntesis" de los que la codificación CELP (siglas en inglés de "Code Excited Linear Prediction" o predicción lineal excitada por código) es el ejemplo más conocido.

15

Para aplicaciones no conversacionales, el estado de la técnica sobre codificación de señales de audio (mono) consiste en la codificación perceptual por transformada o en subbandas, con una codificación paramétrica de altas frecuencias por replicación de banda (SBR, siglas en inglés de "Spectral Band Replication" o replicación de banda espectral). Se puede encontrar una revisión de los métodos clásicos de codificación de voz y audio en los trabajos de W.B. Kleijn y K.K. Paliwal (Eds.), *Speech Coding and Synthesis*, Elsevier, 1995; M. Bosi, R.E. Goldberg, *Introduction to Digital Audio Coding and Standards*, Springer 2002; J. Benesty, M.M. Sondhi, Y. Huang (Eds.), *Handbook of Speech Processing*, Springer 2008.

20

25

En el presente documento se considerará particularmente con más interés el códec (codificador y decodificador) estandarizado 3GPP AMR-WB (siglas en inglés de "Adaptive Multi-Rate Wideband" o banda ancha adaptativa de múltiples velocidades) que funciona a una frecuencia de entrada/salida de 16 kHz y en el que la señal se divide en dos subbandas, la banda baja (0-6,4 kHz) que se muestrea a 12,8 kHz y se codifica según el modelo CELP y la banda alta (6,4-7 kHz) que se reconstruye paramétricamente por "*extensión de banda*" (o BWE, siglas en inglés de "Bandwidth Extension" o extensión de ancho de banda) con o sin información adicional según el modo de la trama actual. En este caso, se puede observar que la limitación de la banda codificada del códec AMR-WB a 7 kHz está esencialmente relacionada con el hecho de que la respuesta en frecuencia de emisión de los terminales de banda ampliada se aproximó en el momento de la estandarización (ETSI/3GPP y luego UIT-T) según la máscara de frecuencia definida en la norma UIT-T P.341 y más específicamente utilizando un filtro denominado "P341" definido en la norma UIT-T G.191 que corta frecuencias por encima de 7 kHz (este filtro respeta la máscara definida en P.341). Sin embargo, en teoría, es bien sabido que una señal muestreada a 16 kHz puede tener una banda de audio definida de 0 a 8000 Hz; por tanto, el códec AMR-WB introduce una limitación de la banda alta en comparación con el ancho de banda teórico de 8 kHz.

30

35

40

El códec de voz 3GPP AMR-WB se estandarizó en 2001 principalmente para aplicaciones de telefonía en modo circuito (CS) sobre GSM (2G) y UMTS (3G). Este mismo códec también se estandarizó en 2003 al UIT-T como recomendación G.722.2 "Wideband coding speech at around 16kbit/s using Adaptive Multi-Rate Wideband (AMR-WB)".

45

Comprende nueve velocidades, denominadas modos, de 6,6 a 23,85 kbit/s, y comprende mecanismos de transmisión continua (DTX siglas en inglés de "Discontinuous Transmission" o transmisión discontinua) con detección de actividad de voz (VAD, siglas en inglés de "Voice Activity Detection" o detección de actividad de voz) y generación de ruido de confort (CNG, siglas en inglés de "Confort Noise Generation") a partir de tramas de descripción del silencio (SID, siglas en inglés de "Silence Insertion Descriptor" o descriptor de inserción de silencio), así como mecanismos de corrección de tramas perdidas (FEC, siglas en inglés de "Frame Erasure Concealment" u ocultamiento del borrado de tramas", a veces llamado PLC por sus siglas en inglés de "Packet Loss Concealment" u ocultamiento de pérdida de paquetes).

50

En el presente documento no se repetirán los detalles del algoritmo de codificación y decodificación AMR-WB, ya que se puede encontrar una descripción detallada de este códec en las especificaciones 3GPP (TS 26.190, 26.191, 26.192, 26.193, 26.194, 26.204) y UIT-TG.722.2 (y los Anexos y Apéndices correspondientes), así como en el artículo de B. Bessette *et al.* titulado "The adaptive multirate wideband speech codec (AMR-WB)", *IEEE Transactions on Speech and Audio Processing*, vol. 10, n.º 8, 2002, páginas. 620-636 y los códigos fuente de los estándares 3GPP y UIT-T asociados.

55

60

El principio de la extensión de banda en el códec AMR-WB es bastante rudimentario. En efecto, la banda alta (6,4-7 kHz) se genera conformando un ruido blanco a través de una envolvente temporal (aplicada en forma de ganancias por subtrama) y de frecuencia (mediante la aplicación de un filtro de síntesis de predicción lineal o LPC, de sus siglas en inglés de "Linear Predictive Coding" o codificación lineal predictiva). Esta técnica de extensión de banda se ilustra en la figura 1.

65

Un ruido blanco,  $u_{HB1}(n)$ ,  $n = 0, \dots, 79$ , se genera a 16 kHz por subtrama de 5 ms mediante un generador congruencial lineal (bloque 100). Este ruido  $u_{HB1}(n)$  se conforma a lo largo del tiempo aplicando ganancias por subtrama; esta operación se descompone en dos etapas de procesamiento (bloques 102, 106 o 109):

- Se calcula un primer factor (bloque 101) para poner el ruido blanco  $u_{HB1}(n)$  (bloque 102) a un nivel similar al de la excitación,  $u(n)$ ,  $n = 0, \dots, 63$ , decodificado a 12,8 kHz en la banda baja:

$$u_{HB2}(n) = u_{HB1}(n) \sqrt{\frac{\sum_{l=0}^{63} u(l)^2}{\sum_{l=0}^{79} u_{HB1}(l)^2}}$$

En este caso, se puede observar que la normalización de energías se realiza comparando bloques de diferente tamaño (64 para  $u(n)$  y 80 para  $u_{HB1}(n)$ ), sin compensación de las diferencias en las frecuencias de muestreo (12,8 o 16 kHz).

- A continuación, se obtiene la excitación en la banda alta (bloque 106 o 109) en forma de:

$$u_{HB}(n) = \hat{g}_{HB} u_{HB2}(n)$$

donde la ganancia  $\hat{g}_{HB}$  se obtiene de manera diferente dependiendo de la velocidad. Si la velocidad de la trama actual es  $< 23,85$  kbit/s, la ganancia  $\hat{g}_{HB}$  se estima "a ciegas" (es decir, sin información adicional); en ese caso, el bloque 103 filtra la señal de banda baja decodificada mediante un filtro paso alto que tiene una frecuencia de corte de 400 Hz para obtener una señal  $\hat{s}_{hp}(n)$ ,  $n = 0, \dots, 63$  - este filtro paso alto elimina la influencia de frecuencias muy bajas que pueden sesgar la estimación realizada en el bloque 104 - luego se calcula el "tilt" (indicador de pendiente espectral) denotado  $e_{tilt}$  de la señal  $\hat{s}_{hp}(n)$  por autocorrelación normalizada (bloque 104):

$$e_{tilt} = \frac{\sum_{n=1}^{63} \hat{s}_{hp}(n) \hat{s}_{hp}(n-1)}{\sum_{n=0}^{63} \hat{s}_{hp}(n)^2}$$

y por último se calcula  $\hat{g}_{HB}$  en forma de:

$$\hat{g}_{HB} = w_{SP} g_{SP} + (1 - w_{SP}) g_{BG}$$

donde  $g_{SP} = 1 - e_{tilt}$  es la ganancia aplicada en las tramas de voz activas (SP, siglas en inglés de "speech" o voz),  $g_{BG} = 1,25 g_{SP}$  es la ganancia aplicada en las tramas de voz inactivas asociadas con un ruido de fondo (BG, por sus siglas en inglés de "Background" o fondo) y  $w_{SP}$  es una función de ponderación que depende de la detección de actividad de voz (VAD). Se debe entender que la estimación de inclinación del tilt ( $e_{tilt}$ ) permite adaptar el nivel de la banda alta en función de la naturaleza espectral de la señal; esta estimación es particularmente importante cuando la pendiente espectral de la señal decodificada CELP es tal que la energía media disminuye cuando aumenta la frecuencia (caso de una señal sonorizada donde  $e_{tilt}$  se aproxima a 1, por tanto,  $g_{SP} = 1 - e_{tilt}$  se reduce así). También se observa que el factor  $\hat{g}_{HB}$  en la decodificación AMR-WB está limitado para tomar valores en el intervalo  $[0,1, 1,0]$ . De hecho, para las señales cuyo espectro tiene más energía en altas frecuencias ( $e_{tilt}$  cerca de -1,  $g_{SP}$  cerca de 2), la ganancia  $\hat{g}_{HB}$  suele subestimarse.

A 23,85 kbit/s, una información de corrección es transmitida por el codificador AMR-WB y decodificada (bloques 107, 108) para afinar la ganancia estimada por subtrama (4 bits cada 5 ms, es decir, 0,8 kbit/s).

La excitación artificial  $u_{HB}(n)$  se filtra a continuación, (bloque 111) mediante un filtro de síntesis LPC de función de transferencia  $1/A_{HB}(z)$  y que funciona a la frecuencia de muestreo de 16 kHz. La realización de este filtro depende de la velocidad de la trama actual:

- A 6,6 kbit/s, el filtro  $1/A_{HB}(z)$  se obtiene ponderando por un factor  $\gamma = 0,9$  un filtro LPC de orden 20,  $1/\hat{A}^{ext}(z)$  que "extrapola" el filtro LPC de orden 16,  $1/\hat{A}(z)$ , decodificado en la banda baja (a 12,8 kHz) - los detalles de la extrapolación en el dominio de los parámetros ISF (siglas en inglés de "Imittance Spectral Frequency" o frecuencia espectral de inmitancia) se describen en la norma G.722.2 en la sección 6.3.2.1; en ese caso,

$$1/A_{HB}(z) = 1/\hat{A}^{ext}(z/\gamma)$$

- A velocidades > 6,6 kbit/s, el filtro  $1/A_{HB}(z)$  es de orden 16 y simplemente corresponde a:

$$1/A_{HB}(z) = 1/\hat{A}(z/\gamma)$$

donde  $\gamma = 0,6$ . Se observa que en este caso el filtro  $1/\hat{A}(z/\gamma)$  se utiliza a 16 kHz, lo que resulta en un ensanchamiento (por homotecia) de la respuesta en frecuencia de este filtro de [0, 6,4 kHz] a [0, 8 kHz].

El resultado,  $S_{HB}(n)$ , se procesa por último mediante un filtro paso banda (bloque 112) de tipo FIR ("Finite Impulse Response" o respuesta finita al impulso), para conservar solo la banda de 6 - 7 kHz; a 23,85 kbit/s, un filtro paso bajo también de tipo FIR (bloque 113) se añade al procesamiento para atenuar aún más las frecuencias superiores a 7 kHz. La síntesis de altas frecuencias (AF) se suma por último (bloque 130) a la síntesis de baja frecuencia (BF) obtenida con los bloques 120 a 123 y remuestreada a 16 kHz (bloque 123). De este modo, incluso si la banda alta se extiende en teoría de 6,4 a 7 kHz en el códec AMR-WB, la síntesis de AF está más bien comprendida en la banda de 6-7 kHz antes de la suma con la síntesis de BF.

Se pueden identificar varios inconvenientes en la técnica de extensión de banda del códec AMR-WB:

- La señal en la banda alta es un ruido blanco conformado (por ganancias temporales por subtrama, mediante un filtrado por  $1/A_{HB}(z)$  y un filtrado paso banda), lo que no es un buen modelo general de la señal en la banda de 6,4-7 kHz. Por ejemplo, hay señales de música muy armónicas para las que la banda de 6,4-7 kHz contiene componentes sinusoidales (o tonos) y ningún ruido (o poco ruido), para estas señales, la extensión de banda del códec AMR-WB degrada en gran medida la calidad.
- El filtro paso bajo de 7 kHz (bloque 113) introduce un desfase de casi 1 ms entre las bandas bajas y altas, lo que puede degradar potencialmente la calidad de algunas señales desincronizando ligeramente las dos bandas a 23,85 kbit/s; esta desincronización también puede suponer un problema cuando se produce una conmutación de velocidad de 23,85 kbit/s a otros modos.
- La estimación de ganancias por subtrama (bloque 101, 103 a 105) no es óptima. En parte, se basa en una ecualización de la energía "absoluta" por subtrama (bloque 101) entre señales a diferentes frecuencias: la excitación artificial a 16 kHz (ruido blanco) y una señal a 12,8 kHz (excitación ACELP decodificada). Cabe destacar, en particular, que este enfoque induce implícitamente una atenuación de la excitación de banda alta (en una relación  $12,8/16 = 0,8$ ); de hecho, también se observará que no se realiza ninguna desacentuación (o desénfasis) en la banda alta en el códec AMR-WB, lo que induce implícitamente una amplificación relativa próxima a 0,6 (que corresponde al valor de la respuesta en frecuencia de  $1/(1-0,68z^{-1})$  a 6400 Hz). De hecho, los factores de 1/0,8 y 0,6 se compensan aproximadamente.
- En la voz, las pruebas de caracterización del códec 3GPP AMR-WB documentadas en el informe 3GPP TR 26.976 han demostrado que el modo a 23,85 kbit/s tiene peor calidad que a 23,05 kbit/s, de hecho, su calidad es similar a la del modo a 15,85 kbit/s. Esto demuestra en particular que el nivel de la señal de AF artificial debe controlarse con mucha prudencia, ya que la calidad se degrada a 23,85 kbit/s, mientras que se supone que los 4 bits por trama permiten aproximarse más a la energía de las altas frecuencias originales.
- La limitación de la banda codificada a 7 kHz es el resultado de la aplicación de un modelo estricto de la respuesta de emisión de los terminales acústicos (filtro P.341 en la norma UIT-T G.191). Ahora bien, para una frecuencia de muestreo de 16 kHz, las frecuencias en la banda de 7-8 kHz siguen siendo significativas, en particular, para las señales musicales, para garantizar un buen nivel de calidad.

El algoritmo de decodificación AMR-WB ha mejorado en parte con el desarrollo del códec escalable UIT-T G.718 que se estandarizó en 2008.

La norma UIT-T G.718 comprende un modo denominado interoperable, para el que la codificación de núcleo es compatible con la codificación G.722.2 (AMR-WB) a 12,65 kbit/s; además, el decodificador G.718 tiene la particularidad de poder decodificar un tren binario AMR-WB/G.722.2 a todas las velocidades posibles del códec AMR-WB (de 6,6 a 23,85 kbit/s).

El decodificador interoperable G.718 en modo de retardo bajo "low delay" en inglés (G.718-LD) se ilustra en la figura 2. A continuación, se enumeran las mejoras realizadas en la funcionalidad de decodificación de tren binario AMR-WB en el decodificador G.718, con referencias a la figura 1 cuando sea necesario:

La extensión de banda (descrita, por ejemplo, en la cláusula 7.13.1 de la recomendación G.718, bloque 206) es idéntica a la del decodificador AMR-WB, salvo porque el filtro paso banda de 6-7 kHz y el filtro de síntesis  $1/A_{HB}(z)$  (bloques 111 y 112) están en orden inverso. Además, a 23,85 kbit/s, los 4 bits transmitidos por subtramas por el codificador AMR-WB no se utilizan en el decodificador interoperable G.718; la síntesis de altas frecuencias (AF) a 23,85 kbit/s es, por tanto, idéntica que a 23,05 kbit/s, lo que evita el problema conocido de la calidad de la decodificación AMR-WB a 23,85 kbit/s. Con mayor motivo, no se utiliza el filtro paso bajo a 7 kHz (bloque 113) y se omite la decodificación específica del modo a 23,85 kbit/s (bloques 107-109).

En G.718 se implementa un postprocesamiento de síntesis a 16 kHz (véase la cláusula 7.14 de G.718) mediante una "noise gate" o puerta de ruido en el bloque 208 (para "mejorar" la calidad de los silencios reduciendo el nivel),

de filtrado paso alto (bloque 209), post-filtro de baja frecuencia (denominado "*bass postfilter*" o *postfiltro de bajos*) en el bloque 210 atenuando el ruido entre armónicos a bajas frecuencias y una conversión a enteros de 16 bits con control de saturación (con control de ganancia o AGC) en el bloque 211.

5 Sin embargo, la extensión de banda en los códecs AMR-WB y/o G.718 (modo interoperable) todavía está limitada en varios aspectos.

En particular, la síntesis de altas frecuencias mediante ruido blanco conformado (mediante un enfoque temporal de tipo fuente-filtro LPC) es un modelo muy limitado de la señal en la banda de frecuencias superiores a 6,4 kHz.

10 Solo la banda de 6,4-7 kHz se resintetiza artificialmente, mientras que, en la práctica, una banda más ancha (hasta 8 kHz) es teóricamente posible a la frecuencia de muestreo de 16 kHz, lo que potencialmente puede mejorar la calidad de las señales, si no están preprocesadas por un filtro de tipo P.341 (50-7000 Hz) como se define en la Software Tool Library (norma G.191) de la UIT-T.

15 Por lo tanto, existe la necesidad de mejorar la extensión de banda en un códec de tipo AMR-WB o una versión interoperable de este códec o, de manera más general, de mejorar la extensión de banda de una señal de audio, para mejorar, en concreto, el contenido de frecuencia de la extensión de banda.

20 La presente invención pretende mejorar la situación.

La invención propone para tal efecto, un procedimiento de extensión de banda de frecuencia de una señal de audiofrecuencia durante un proceso de decodificación o mejora que incluye una etapa de obtención de la señal decodificada en una primera banda de frecuencia denominada banda baja, tal como se reivindica por la reivindicación 1. Se observará que en lo sucesivo la "extensión de banda" se tomará en un sentido amplio e incluirá no solo el caso de la extensión de una subbanda a altas frecuencias, sino también el caso de una sustitución de subbandas puestas a cero (del tipo "noise filling" o relleno de ruido en la codificación por transformada).

30 De este modo, tanto la toma en consideración de componentes tonales como de una señal ambiental extraída de la señal resultante de la decodificación de la banda baja permite realizar la extensión de banda con un modelo de señal adaptado a la verdadera naturaleza de la señal, a diferencia de la utilización de ruido artificial. De este modo se mejora la calidad de la extensión de banda y, en concreto, para ciertos tipos de señales, como las señales musicales.

35 En efecto, la señal decodificada en la banda baja incluye una parte correspondiente al entorno sonoro que se puede transponer a alta frecuencia de modo que una mezcla de las componentes armónicas y el entorno existente permite asegurar una banda alta reconstruida coherente.

40 Se observará que, aunque la invención está motivada por la mejora de la calidad de la extensión de banda en el contexto de la codificación interoperable AMR-WB, los diferentes modos de realización se aplican al caso más general de la extensión de banda de una señal de audio, en particular, en un dispositivo de mejora que realiza un análisis de la señal de audio para extraer los parámetros necesarios para la extensión de banda.

45 Los diferentes modos particulares de realización mencionados a continuación pueden añadirse independientemente o combinados entre sí, a las etapas del procedimiento de extensión definido anteriormente.

En un modo de realización, la extensión de banda se realiza en el dominio de la excitación y la señal de banda baja decodificada es una señal de excitación de banda baja decodificada.

50 La ventaja de este modo de realización es que es posible una transformación sin ventana (o de manera equivalente con una ventana rectangular implícita de la longitud de la trama) en el dominio de la excitación. En este caso, ningún artefacto (efectos de bloque) es entonces audible.

55 En un primer modo de realización, la extracción de las componentes tonales y de la señal ambiental se lleva a cabo según las siguientes etapas:

- la detección de las componentes tonales dominantes de la señal de banda baja decodificada o decodificada y extendida en el dominio de la frecuencia;
- el cálculo de una señal residual mediante la extracción de las componentes tonales dominantes para obtener la señal ambiental.

60 Este modo de realización permite una detección precisa de las componentes tonales.

En una segunda realización, de baja complejidad, la extracción de las componentes tonales y de la señal ambiental se realiza de acuerdo con las siguientes etapas:

- obtención de la señal ambiental calculando un valor medio del espectro de la señal de banda baja decodificada o

- decodificada y extendida;
- obtención de las componentes tonales sustrayendo la señal ambiental calculada de la señal de banda baja decodificada o decodificada y extendida.

5 En un modo de realización de la etapa de combinación, se calcula un factor de control del nivel de energía utilizado para la mezcla adaptativa en función de la energía total de la señal de banda baja decodificada o decodificada y extendida y de las componentes tonales.

10 La aplicación de este factor de control permite a la etapa de combinación adaptarse a las características de la señal para optimizar la proporción relativa de señal ambiental en la mezcla. El nivel de energía se controla así para evitar artefactos audibles.

15 En un modo de realización preferido, la señal de banda baja decodificada se somete a una etapa de descomposición en subbandas mediante una transformada o un banco de filtros, realizándose entonces las etapas de extracción y combinación en el dominio de la frecuencia o en subbandas.

20 La implementación de la extensión de banda en el dominio de la frecuencia permite obtener una finura en el análisis de frecuencia que no está disponible con un enfoque temporal, y también permite tener una resolución de frecuencia suficiente para detectar las componentes tonales.

En un modo de realización detallado, la señal de banda baja decodificada y extendida se obtiene de acuerdo con la siguiente ecuación:

$$U_{HB1}(k) = \begin{cases} 0 & k = 0, \dots, 199 \\ U(k) & k = 200, \dots, 239 \\ U(k + start\_band - 240) & k = 240, \dots, 319 \end{cases}$$

25 siendo  $k$  el índice de la muestra,  $U(k)$  el espectro de la señal obtenida después de una etapa de transformada  $U_{HB1}(k)$  el espectro de la señal extendida y  $start\_band$  una variable predefinida. De este modo, esta función comprende un remuestreo de la señal añadiendo muestras al espectro de esa señal. Sin embargo, son posibles otras formas de extender la señal, por ejemplo, por traslación en un procesamiento de subbandas.

30 La presente invención también se refiere a un dispositivo de extensión de banda de frecuencia de una señal de audiofrecuencia, habiendo sido la señal decodificada en una primera banda de frecuencia denominada banda baja, tal como se reivindica por la reivindicación 9.

35 Este dispositivo presenta las mismas ventajas que el procedimiento descrito anteriormente, que este implementa.

La invención se refiere a un decodificador que incluye un dispositivo como el descrito.

40 Se refiere a un programa informático que incluye instrucciones de código para la implementación de las etapas del procedimiento de extensión de banda tal como se ha descrito, cuando estas instrucciones son ejecutadas por un procesador.

45 Por último, la invención hace referencia a un medio de almacenamiento, legible por un procesador, integrado o no en el dispositivo de extensión de banda, eventualmente amovible, que tiene en memoria un programa informático que implementa un procedimiento de extensión de banda tal como el que se ha descrito anteriormente.

Otras características y ventajas de la invención se pondrán de manifiesto de manera más clara tras la lectura de la siguiente descripción, aportada únicamente a modo de ejemplo no limitativo y hecha con referencia a los dibujos adjuntos, en los que:

- la figura 1 ilustra parte de un decodificador de tipo AMR-WB que implementa las etapas de extensión de banda de frecuencia del estado de la técnica y tal como se ha descrito anteriormente;
- la figura 2 ilustra un decodificador del tipo interoperable G.718-LD a 16 kHz según el estado de la técnica y tal como se ha descrito anteriormente;
- la figura 3 ilustra un decodificador interoperable con la codificación AMR-WB e integra un dispositivo de extensión de banda según un modo de realización de la invención;
- la figura 4 ilustra en forma de organigrama, las etapas principales de un procedimiento de extensión de banda según un modo de realización de la invención;
- la figura 5 ilustra un modo de realización en el dominio de la frecuencia de un dispositivo de extensión de banda según la invención integrado en un decodificador; y
- la figura 6 ilustra una realización material de un dispositivo de extensión de banda según la invención.

La figura 3 ilustra un ejemplo de decodificador, compatible con la norma AMR-WB/G.722.2 en la que existe un postprocesamiento similar al introducido en G.718 y descrito con referencia a la figura 2 y una extensión de banda mejorada según el procedimiento de extensión de la invención, implementado por el dispositivo de extensión de banda ilustrado por el bloque 309.

A diferencia de la decodificación AMR-WB que funciona con una frecuencia de muestreo de salida de 16 kHz y de la decodificación G.718 que funciona a 8 o 16 kHz, en el presente documento, se considera un decodificador que puede funcionar con una señal de salida (síntesis) a la frecuencia  $f_s = 8, 16, 32$  o  $48$  kHz. Se observa que en el presente documento se asume que la codificación se ha realizado según el algoritmo AMR-WB con una frecuencia interna de  $12,8$  kHz para la codificación CELP en banda baja y a  $23,85$  kbit/s una codificación de ganancia por subtrama a la frecuencia de  $16$  kHz, pero también son posibles variantes interoperables del codificador AMR-WB; incluso si la invención se describe en el presente documento a nivel de decodificación, en el presente documento se supone que la codificación también puede funcionar con una señal de entrada a una frecuencia  $f_s = 8, 16, 32$  o  $48$  kHz y las operaciones de remuestreo adecuadas, que van más allá del alcance de la invención, se implementan en la codificación en función del valor de  $f_s$ . Se puede observar que cuando  $f_s = 8$  kHz en el decodificador, en el caso de una decodificación compatible con AMR-WB, no es necesario extender la banda baja  $0-6,4$  kHz, ya que la banda de audio reconstruida a la frecuencia  $f_s$  está limitada a  $0-4000$  Hz.

En la figura 3, la decodificación CELP (BF siglas de bajas frecuencias) siempre funciona a una frecuencia interna de  $12,8$  kHz, como en AMR-WB y G.718, y la extensión de banda (AF, siglas de altas frecuencias) que constituye el objeto de la invención funciona a una frecuencia de  $16$  kHz, las síntesis de BF y AF se combinan (bloque 312) con la frecuencia  $f_s$  después de un remuestreo adecuado (bloques 307 y 311). En unas variantes de la invención, la combinación de las bandas baja y alta se puede hacer a  $16$  kHz, después de remuestrear la banda baja de  $12,8$  a  $16$  kHz, antes de volver a muestrear la señal combinada a la frecuencia  $f_s$ .

La decodificación según la figura 3 depende del modo AMR-WB (o velocidad) asociado con la trama actual recibida. A modo indicativo y sin que esto afecte al bloque 309, la decodificación de la parte CELP en banda baja incluye las siguientes etapas:

- Demultiplexación de parámetros codificados (bloque 300) en caso de una trama correctamente recibida ( $bfi=0$  donde  $bfi$  es el "bad frame indicator" o indicador de trama incorrecta que vale 0 para una trama recibida y 1 para una trama perdida).
- Decodificación de parámetros ISF con interpolación y conversión a coeficientes LPC (bloque 301) como se describe en la cláusula 6.1 de la norma G.722.2.
- Decodificación de la excitación CELP (bloque 302), con una parte adaptativa y fija para reconstruir la excitación ( $exc$  o  $u'(n)$ ) en cada subtrama de longitud de  $64$  a  $12,8$  kHz:

$$u'(n) = \hat{g}_p v(n) + \hat{g}_c c(n), \quad n = 0, \dots, 63$$

según las notaciones de la cláusula 7.1.2.1 de G.718 con respecto a la decodificación CELP, donde  $v(n)$  y  $c(n)$  son, respectivamente, las palabras de código de los diccionarios adaptativo y fijo, y  $\hat{g}_p$  y  $\hat{g}_c$  son las ganancias decodificadas asociadas. Esta excitación  $u'(n)$  se utiliza en el diccionario adaptativo de la siguiente subtrama; luego se postprocesa y se distingue, al igual que en G.718, la excitación  $u'(n)$  (también denotada  $exc$ ) de su versión modificada postprocesada  $u''(n)$  (también denotada  $exc2$ ) que sirve de entrada al filtro de síntesis,  $1/\hat{A}(z)$ , en el bloque 303. En unas variantes que se pueden implementar para la invención, se pueden modificar los postprocesamientos aplicados a la excitación (por ejemplo, se puede mejorar la dispersión de fase) o se pueden extender estos postprocesamientos (por ejemplo, se puede implementar una reducción de ruido entre armónicos), sin afectar a la naturaleza del procedimiento de extensión de banda según la invención.

- Filtrado de síntesis por  $1/\hat{A}(z)$  (bloque 303) donde el filtro LPC decodificado  $\hat{A}(z)$  es de orden 16
- Postprocesamiento de banda estrecha (bloque 304) según la cláusula 7.3 de G.718 si  $f_s = 8$  kHz.
- Desacentuación (bloque 305) por filtro  $1/(1-0,68z^{-1})$
- Postprocesamiento de bajas frecuencias (bloque 306) tal como se describe en la cláusula 7.14.1.1 de G.718. Este procesamiento introduce un retardo que se tiene en cuenta en la decodificación de la banda alta ( $> 6,4$  kHz).
- Remuestreo de la frecuencia interna de  $12,8$  kHz a la frecuencia de salida  $f_s$  (bloque 307). Son posibles varias realizaciones. Sin pérdida de generalidad, en este caso se considera a modo de ejemplo que si  $f_s = 8$  o  $16$  kHz, el remuestreo descrito en la cláusula 7.6 de G.718 se reproduce en este caso, y si  $f_s = 32$  o  $48$  kHz, se utilizan filtros de respuesta finita al impulso (FIR) adicionales.
- Cálculo de los parámetros del "noise gate" (bloque 308) que preferentemente se realiza como se describe en la cláusula 7.14.3 de G.718.

En unas variantes que se pueden implementar para la invención, se pueden modificar los postprocesamientos aplicados a la excitación (por ejemplo, se puede mejorar la dispersión de fase) o se pueden extender estos postprocesamientos (por ejemplo, se puede implementar una reducción de ruido entre armónicos), sin afectar a la naturaleza de la extensión de banda. En el presente documento no se describe el caso de decodificación de banda

baja cuando se pierde la trama actual ( $bfi=1$ ) y que es informativo en la norma 3GPP AMR-WB; en general, ya se trate del decodificador AMR-WB o de un decodificador general basado en el modelo de fuente-filtro, normalmente, consiste en estimar con el menor margen de error la excitación LPC y los coeficientes del filtro LPC de síntesis para reconstituir la señal perdida mientras se mantiene el modelo fuente-filtro. Cuando  $bfi=1$  se considera en el presente documento que la extensión de banda (bloque 309) puede operar como en el caso  $bfi=0$  y a una velocidad  $<23,85$  kbit/s; de este modo, la descripción de la invención asumirá a continuación y sin pérdida de generalidad que,  $bfi=0$ .

Se puede observar que la utilización de los bloques 306, 308, 314 es opcional.

También se observará que la decodificación de la banda baja descrita anteriormente asume una trama actual denominada "activa" con una velocidad de entre 6,6 y 23,85 kbit/s. De hecho, cuando el modo DTX (transmisión continua en español) está activado, algunas tramas pueden codificarse como "inactivas" y en ese caso es posible transmitir un descriptor de silencio (en 35 bits) o no transmitir nada. En particular, se recuerda que la trama SID del codificador AMR-WB describe varios parámetros: parámetros ISF promediados sobre 8 tramas, energía media en 8 tramas, "flag dithering" o difuminado de bandera para la reconstrucción de ruido no estacionario. En cualquier caso, en el decodificador, se encuentra el mismo modelo de decodificación que para una trama activa, con una reconstrucción de la excitación y de un filtro LPC para la trama actual, lo que permite aplicar la invención incluso en tramas inactivas. La misma constatación se aplica a la decodificación de "tramas perdidas" (o FEC, PLC) en la que se aplica el modelo LPC.

Este ejemplo de decodificador funciona en el dominio de la excitación y por lo tanto incluye una etapa de decodificación de la señal de excitación de banda baja. El dispositivo de extensión de banda y el procedimiento de extensión de banda en el sentido de la invención funciona, asimismo, en un dominio diferente al dominio de la excitación y en concreto con una señal directa decodificada en banda baja o una señal ponderada mediante un filtro perceptual.

A diferencia de la decodificación AMR-WB o G.718, el decodificador descrito permite extender la banda baja decodificada (50-6400 Hz teniendo en cuenta el filtrado paso alto a 50 Hz en el decodificador, 0-6400 Hz en el caso general) a una banda extendida cuyo ancho varía, pasando aproximadamente de 50-6900 Hz a 50-7700 Hz dependiendo del modo implementado en la trama actual. Por tanto, se puede hablar de una primera banda de frecuencia de 0 a 6400 Hz y de una segunda banda de frecuencia de 6400 a 8000 Hz. En realidad, en el modo de realización preferido, la excitación para las altas frecuencias se genera en el dominio de la frecuencia en una banda de 5000 a 8000 Hz, para permitir un filtrado paso banda de un ancho de 6000 a 6900 o 7700 Hz, cuya pendiente no sea demasiado pronunciada en la banda superior rechazada.

La parte de síntesis de banda alta se realiza en el bloque 309 que representa el dispositivo de extensión de banda según la invención y que se detalla en la figura 5 en un modo de realización.

Para alinear las bandas bajas y altas decodificadas, se introduce un retardo (bloque 310) para sincronizar las salidas de los bloques 306 y 309 y la banda alta sintetizada a 16 kHz se remuestrea de 16 kHz a la frecuencia  $f_s$  (salida del bloque 311). El valor del retardo  $T$  deberá estar adaptado para los otros casos ( $f_s=32, 48$  kHz) dependiendo de los procesamientos implementados. Se recuerda que cuando  $f_s=8$  kHz, no es necesario aplicar los bloques 309 a 311 porque la banda de la señal a la salida del decodificador está limitada a 0-4000 Hz.

Se observará que el procedimiento de extensión de la invención implementado en el bloque 309 según el primer modo de realización preferentemente no introduce ningún retardo adicional con respecto a la banda baja reconstruida a 12,8 kHz; sin embargo, en unas variantes de la invención (por ejemplo, utilizando una transformación de tiempo/frecuencia con superposición), se podrá introducir un retardo. De este modo, de manera general el valor de  $T$  en el bloque 310 deberá ajustarse en función de la implementación específica. Por ejemplo, en el caso de que no se utilice el postprocesamiento de bajas frecuencias (bloque 306), el retardo a introducir para  $f_s=16$  kHz se podrá fijar en  $T=15$ .

Las bandas baja y alta se combinan a continuación (se suman) en el bloque 312 y la síntesis obtenida se postprocesa por filtrado paso alto a 50 Hz (de tipo IIR) de orden 2, cuyos coeficientes dependen de la frecuencia  $f_s$  (bloque 313) y del postprocesamiento de salida con aplicación opcional del "noise gate" de manera similar a G.718 (bloque 314).

El dispositivo de extensión de banda según la invención, ilustrado por el bloque 309 según el modo de realización del decodificador de la figura 5, implementa un procedimiento de extensión de banda (en el sentido amplio) descrito ahora con referencia a la figura 4.

Este dispositivo de extensión también puede ser independiente del decodificador y puede implementar el procedimiento descrito en la figura 4 para realizar una extensión de banda de una señal de audio existente almacenada o transmitida al dispositivo, con un análisis de la señal de audio para extraer, por ejemplo, una excitación y un filtro LPC.

Este dispositivo recibe como entrada una señal decodificada en una primera banda de frecuencia denominada banda baja  $u(n)$  que puede estar en el dominio de la excitación o en el de la señal. En el modo de realización descrito en el

presente documento, se aplica una etapa de descomposición en subbandas (E401b) por transformada de tiempo-frecuencia o banco de filtros a la señal de banda baja decodificada para obtener el espectro de la señal de banda baja decodificada  $u(n)$  para una implementación en el dominio de la frecuencia.

Una etapa E401 de extensión de la señal de banda baja decodificada a una segunda banda de frecuencia superior a la primera banda de frecuencia, para obtener una señal de banda baja decodificada extendida  $U_{HB1}(k)$ , se realiza en esta señal de banda baja decodificada antes o después de la etapa de análisis (descomposición en subbandas). Esta etapa de extensión puede incluir tanto una etapa de remuestreo como una etapa de extensión o simplemente una etapa de traslación o transposición de frecuencia en función de la señal obtenida en la entrada. Se observará que, en unas variantes, no cubiertas por las reivindicaciones, la etapa E401a se podría realizar al final del procesamiento descrito en la figura 4, es decir, en la señal combinada, realizándose este procesamiento principalmente en la señal de banda baja antes de la extensión, siendo el resultado equivalente.

Esta etapa se detalla más adelante en el modo de realización descrito con referencia a la figura 5.

Una etapa E402 de extracción de una señal ambiental ( $U_{HBA}(k)$ ) y de las componentes tonales ( $y(k)$ ) se realiza a partir de la señal de banda baja decodificada ( $U(k)$ ) o decodificada y extendida ( $U_{HB1}(k)$ ). En el presente documento se define como señal ambiental a la señal residual que se obtiene eliminando los armónicos (o componentes tonales) principales (o dominantes) de la señal existente.

En la mayoría de las señales de banda ampliada (muestreadas a 16 kHz), la banda alta (>6 kHz) contiene información ambiental que generalmente es similar a la presente en la banda baja.

La etapa de extracción de las componentes tonales y de la señal ambiental comprende por ejemplo las siguientes etapas:

- la detección de las componentes tonales dominantes de la señal de banda baja decodificada (o decodificada y extendida), en el dominio de la frecuencia; y
- el cálculo de una señal residual mediante la extracción de las componentes tonales dominantes para obtener la señal ambiental.

Esta etapa también se puede obtener mediante:

- la obtención de la señal ambiental calculando un valor medio de la señal de banda baja decodificada (o decodificada y extendida); y
- la obtención de las componentes tonales sustrayendo la señal ambiental calculada de la señal de banda baja decodificada (o decodificada y extendida).

A continuación, las componentes tonales y la señal ambiental se combinan de manera adaptativa con la ayuda de factores de control del nivel de energía en la etapa E403 para obtener la denominada señal combinada ( $U_{HB2}(k)$ ). La etapa de extensión E401a se puede implementar entonces, si aún no se ha realizado, en la señal de banda baja decodificada.

De este modo, la combinación de estos dos tipos de señales permite obtener una señal combinada con características mejor adaptadas para ciertos tipos de señales como las señales musicales y más rica en contenido de frecuencia y en la banda de frecuencia extendida correspondiente a cualquier banda de frecuencia, incluyendo la primera y la segunda banda de frecuencia.

La extensión de banda según el procedimiento mejora la calidad para este tipo de señales en comparación con la extensión descrita en la norma AMR-WB.

El hecho de utilizar una combinación de señal ambiental y componentes tonales permite enriquecer esta señal de extensión para aproximarla a las características de la señal real y no de una señal artificial.

Esta etapa de combinación se detallará más adelante con referencia a la figura 5.

Una etapa de síntesis, que corresponde al análisis de 401b, se realiza en E404b para llevar la señal al dominio del tiempo.

Opcionalmente, se puede realizar una etapa de ajuste del nivel de energía de la señal de banda alta en E404a, antes y/o después de la etapa de síntesis, mediante la aplicación de una ganancia y/o mediante un filtrado adecuado. Esta etapa se explicará con más detalle en el modo de realización descrito en la figura 5 para los bloques 501 a 507.

En un ejemplo de realización, el dispositivo de extensión de banda 500 se describe ahora con referencia a la figura 5, que ilustra tanto este dispositivo como también los módulos de procesamiento adecuados para la implementación en un decodificador de tipo interoperable con una codificación AMR-WB. Este dispositivo 500 implementa el

procedimiento de extensión de banda descrito anteriormente con referencia a la figura 4.

De este modo, el bloque de procesamiento 510 recibe una señal de banda baja decodificada ( $u(n)$ ). En un modo de realización particular, la extensión de banda utiliza la excitación decodificada a 12,8 kHz (exc2 o ( $u(n)$ )) a la salida del bloque 302 de la figura 3.

Esta señal es descompuesta en subbandas de frecuencia por el módulo 510 de descomposición en subbandas (que implementa la etapa E401b de la figura 4) que generalmente realiza una transformada o aplica un banco de filtros, para obtener una descomposición en subbandas  $U(k)$  de la señal  $u(n)$ .

En un modo de realización particular, una transformada de tipo DCT-IV (por sus siglas en inglés de "*Discrete Cosine Transform- Type IV*" o transformada de coseno discreta de tipo IV) (bloque 510) se aplica a la trama actual de 20 ms (256 muestras), sin ventana, lo que equivale a transformar directamente  $u(n)$  siendo  $n = 0, \dots, 255$  de acuerdo con la siguiente fórmula:

$$U(k) = \sum_{n=0}^{N-1} u(n) \cos\left(\frac{\pi}{N} \left(n + \frac{1}{2}\right) \left(k + \frac{1}{2}\right)\right)$$

donde  $N = 256$  y  $k = 0, \dots, 255$ .

Es posible una transformación sin ventana (o de manera equivalente con una ventana rectangular implícita de la longitud de la trama) cuando el procesamiento se realiza en el dominio de la excitación y no en el dominio de la señal. En este caso, ningún artefacto (efectos de bloque) es audible, lo que constituye una ventaja importante de este modo de realización de la invención.

En este modo de realización, la transformación DCT-IV es implementada por FFT de acuerdo con el algoritmo denominado "*Evolved DCT* (EDCT) o DCT evolucionada" que se describe en el artículo de D.M. Zhang, H.T. Li, A Low Complexity Transform - Evolved DCT, IEEE 14th International Conference on Computational Science and Engineering (CSE), ago. 2011, páginas 144-149, e implementado en las normas UIT-T G.718 Anexo B y G.729.1 Anexo E.

En unas variantes de la invención y sin pérdida de generalidad, la transformación DCT-IV podría reemplazarse por otras transformaciones de tiempo-frecuencia a corto plazo de la misma longitud y en el dominio de la excitación o en el dominio de la señal, como una FFT (siglas en inglés de "*Fast Fourier Transform*" o Transformada rápida de Fourier) o una DCT-II (*Transformada de coseno discreta* - Tipo II). Como alternativa, se podría reemplazar la DCT-IV en la trama por una transformación con superposición-adición y ventana de longitud superior a la longitud de la trama actual, por ejemplo, utilizando una MDCT (siglas en inglés de "*Modified Discrete Cosine Transform*" o Transformada de coseno discreta modificada). En este caso el retardo  $T$  en el bloque 310 de la figura 3, deberá ajustarse (reducirse) adecuadamente en función del retardo adicional debido al análisis/síntesis mediante esta transformada.

En otro modo de realización, la descomposición en subbandas se realiza mediante la aplicación de un banco de filtros, por ejemplo, de tipo PQMF (Pseudo-QMF) reales o complejos. Para algunos bancos de filtros, se obtiene, para cada subbanda en una trama dada, no un valor espectral, sino una serie de valores temporales asociados a la subbanda; en ese caso, el modo de realización preferido de la invención se puede aplicar realizando, por ejemplo, una transformada de cada subbanda y calculando la señal ambiental en el dominio de los valores absolutos, obteniéndose siempre las componentes tonales por la diferencia entre la señal (en valor absoluto) y la señal ambiental. En el caso de un banco de filtro complejo, el módulo complejo de las muestras reemplazará el valor absoluto.

En otros modos de realización, la invención se aplicará en un sistema que utiliza dos subbandas, siendo analizada la banda baja por transformada o por banco de filtros.

En el caso de una DCT, el espectro DCT,  $U(k)$ , de 256 muestras que cubre la banda de 0-6400 Hz (a 12,8 kHz), se extiende a continuación (bloque 511) en un espectro de 320 muestras que cubre la banda 0-8000 Hz (a 16 kHz) de la siguiente forma:

$$U_{HBI}(k) = \begin{cases} 0 & k = 0, \dots, 199 \\ U(k) & k = 200, \dots, 239 \\ U(k + start\_band - 240) & k = 240, \dots, 319 \end{cases}$$

donde se toma preferentemente una  $start\_band = 160$ .

El bloque 511 implementa la etapa E401a de la figura 4, es decir, la extensión de la señal de banda baja decodificada. Esta etapa también puede incluir un remuestreo de 12,8 a 16 kHz en el dominio de la frecuencia, añadiendo 1/4 de las muestras ( $k = 240, \dots, 319$ ) al espectro, la relación entre 16 y 12,8 es 5/4.

5 En la banda de frecuencia correspondiente a las muestras que van desde los índices 200 a 239, se conserva el espectro original, para poder aplicarle una respuesta de atenuación progresiva del filtro paso alto en esta banda de frecuencia y también para no introducir defectos audibles durante la etapa de adición de la síntesis de baja frecuencia a la síntesis de alta frecuencia.

10 Se observará que, en este modo de realización, la generación del espectro extendido sobremuestreado se realiza en una banda de frecuencia que va de 5 a 8 kHz, por lo que incluye una segunda banda de frecuencia (6,4-8 kHz) superior a la primera banda de frecuencia (0-6,4 kHz).

15 De este modo, la extensión de la señal de banda baja decodificada se realiza al menos sobre la segunda banda de frecuencia, pero también sobre una parte de la primera banda de frecuencia.

Aunque evidentemente, los valores que definen estas bandas de frecuencia pueden ser diferentes dependiendo del decodificador o del dispositivo de procesamiento en el que se aplique la invención.

20 Además, el bloque 511 realiza un filtrado paso alto implícito en la banda de 0-5000 Hz ya que las primeras 200 muestras de  $U_{HB1}(k)$  se ponen a cero; como se explica más adelante, este filtrado paso alto también puede complementarse con una parte de atenuación progresiva de los valores espectrales de los índices  $k = 200, \dots, 255$  en la banda de 5000-6400 Hz, esta atenuación progresiva se implementa en el bloque 501, pero podría llevarse a cabo por separado fuera del bloque 501. De manera equivalente y en unas variantes de la invención, la implementación del  
25 filtrado paso alto separado en bloques de coeficientes de índice  $k = 0, \dots, 199$  puestos a cero, de coeficientes  $k = 200, \dots, 255$  atenuados, en el dominio transformado, podrá realizarse, por lo tanto, en una sola etapa.

30 En este ejemplo de realización y según la definición de  $U_{HB1}(k)$ , se observa que la banda de 5000-6000 Hz de  $U_{HB1}(k)$  (que corresponde a los índices  $k = 200, \dots, 239$ ) se copia a partir de la banda 5000-6000 Hz de  $U(k)$ . Este enfoque permite conservar el espectro original en esta banda y evita introducir distorsiones en la banda de 5000-6000 Hz durante la suma de la síntesis de AF con la síntesis de BF, en particular, se preserva la fase de la señal (representada implícitamente en el dominio DCT-IV) en esta banda.

35 La banda de 6000-8000 Hz  $U_{HB1}(k)$  se define en este caso copiando la banda 4000-6000 Hz de  $U(k)$  ya que el valor de *start\_band* se fija preferentemente en 160.

40 En una variante del modo de realización, el valor de *start\_band* se puede volver adaptativo en torno al valor de 160, sin modificar la naturaleza de la invención. Los detalles de la adaptación del valor *start\_band* no se describen en el presente documento porque van más allá del ámbito de la invención sin cambiar su alcance.

45 En la mayoría de las señales de banda ampliada (muestreadas a 16 kHz), la banda alta (>6 kHz) contiene información ambiental que naturalmente es similar a la presente en la banda baja. En el presente documento se define como señal ambiental a la señal residual que se obtiene eliminando los armónicos principales (o dominantes) de la señal existente. El nivel de armonización en la banda de 6000-8000 Hz generalmente está correlacionado con el de las bandas de frecuencias inferiores.

50 Esta señal de banda baja decodificada y extendida se suministra a la entrada del dispositivo de extensión 500 y en concreto a la entrada del módulo 512. De este modo, el bloque 512 de extracción de componentes tonales y de una señal ambiental, implementa la etapa E402 de la figura 4 en el dominio de la frecuencia. La señal ambiental,  $U_{HBA}(k)$  para  $k = 240, \dots, 319$  (80 muestras) se obtiene de este modo para una segunda banda de frecuencia denominada alta frecuencia para luego combinarla adaptativamente con las componentes tonales extraídas  $y(k)$ , en el bloque de combinación 513.

55 En un modo de realización particular, la extracción de las componentes tonales y de la señal ambiental (en la banda 6000-8000 Hz) se realiza de acuerdo con las siguientes operaciones:

- Cálculo de la energía total de la señal de banda baja decodificada extendida  $ener_{HB}$ :

$$ener_{HB} = \sum_{k=240}^{319} U_{HB1}(k)^2 + \varepsilon$$

60 donde  $\varepsilon = 0,1$  (este valor puede ser diferente, en el presente documento se ha fijado a modo de ejemplo).

- Cálculo de la parte ambiental (en valor absoluto) que corresponde en este caso al nivel medio del espectro  $lev(i)$

(línea a línea) y cálculo de la energía  $ener_{tonal}$  de las partes tonales dominantes (en el espectro de altas frecuencias)

Para  $i = 0 \dots L-1$ , se obtiene este nivel medio mediante la siguiente ecuación:

$$lev(i) = \frac{1}{fn(i) - fb(i) + 1} \sum_{j=fb(i)}^{fn(i)} |U_{HB1}(j + 240)|$$

Esto corresponde al nivel medio (en valor absoluto) y, por tanto, representa una especie de envolvente del espectro. En este modo de realización,  $L = 80$  y representa la longitud del espectro y el índice  $i$  de  $0$  a  $L-1$  corresponde a los índices  $j+240$  de  $240$  a  $319$ , es decir, el espectro de  $6$  a  $8$  kHz.

En general  $fb(i) = i-7$  y  $fn(i) = i + 7$ , sin embargo, los  $7$  primeros y los últimos índices ( $i = 0, \dots, 6$  e  $i=L-7, \dots, L-1$ ) requieren un procesamiento especial y sin pérdida de generalidad se define entonces:

$$fb(i) = 0 \text{ y } fn(i) = i + 7 \text{ para } i = 0, \dots, 6$$

$$fb(i) = i-7 \text{ y } fn(i) = L-1 \text{ para } i = L-7, \dots, L-1$$

En unas variantes de la invención, la media de  $|U_{HB1}(j + 240)|$ ,  $j = fb(i), \dots, fn(i)$ , podrá reemplazarse por un valor mediano en el mismo conjunto de valores, es decir,  $lev(i) = \text{mediana}_{j=fb(i), \dots, fn(i)} (|U_{HB1}(j + 240)|)$ . Esta variante tiene el defecto de ser más compleja (en términos de número de cálculos) que una media móvil. En otras variantes, se podría aplicar una ponderación no uniforme a los términos promediados, o se podría reemplazar el filtrado medio, por ejemplo, por otros filtros no lineales de tipo "stack filters" o filtros apilados.

También se calcula la señal residual:

$$y(i) = |U_{HB1}(i + 240)| - lev(i), \quad i = 0, \dots, L-1$$

que corresponde (aproximadamente) a las componentes tonales si el valor  $y(i)$  en una línea  $i$  dada es positivo ( $y(i) > 0$ ).

Por tanto, este cálculo implica una detección implícita de las componentes tonales. Por tanto, las partes tonales se detectan implícitamente con la ayuda del término intermedio  $y(i)$  que representa un umbral adaptativo. Siendo la condición de detección  $y(i) > 0$ . En unas variantes de la invención, esta condición podría cambiarse, por ejemplo, definiendo un umbral adaptativo como función de la envolvente local de la señal o en forma de  $y(i) > lev(i) + x \text{ dB}$  donde  $x$  tiene un valor predefinido (por ejemplo,  $x = 10 \text{ dB}$ ).

La energía de las partes tonales dominantes viene definida por la siguiente ecuación:

$$ener_{tonal} = \sum_{i=0 \dots L-1 |y(i)| > 0} y(i)^2$$

Por supuesto, pueden contemplarse otros métodos de extracción de la señal ambiental. Por ejemplo, esta señal ambiental se puede extraer de una señal de baja frecuencia o eventualmente de otra banda de frecuencia (o varias bandas de frecuencia).

La detección de picos o componentes tonales se podría hacer de una manera diferente.

La extracción de esta señal ambiental también podría realizarse en la excitación decodificada pero no extendida, es decir, antes de la etapa de extensión o de traslación espectral, es decir, por ejemplo, en una porción de la señal de baja frecuencia en lugar de directamente en la señal de alta frecuencia.

En una variante de realización, la extracción de las componentes tonales y de la señal ambiental se lleva a cabo en un orden diferente y según las siguientes etapas:

- la detección de las componentes tonales dominantes de la señal de banda baja decodificada (o decodificada y extendida), en el dominio de la frecuencia;
- el cálculo de una señal residual mediante la extracción de las componentes tonales dominantes para obtener la señal ambiental.

Esta variante se puede lograr de la siguiente manera: Se detecta un pico (o componente tonal) en una línea de índice

$i$  en el espectro de amplitud  $|U_{HB1}(i + 240)|$  si se cumple el siguiente criterio:

$$|U_{HB1}(i + 240)| > |U_{HB1}(i + 240 - 1)| \text{ y } |U_{HB1}(i + 240)| > |U_{HB1}(i + 240 + 1)|,$$

5 para  $i = 0, \dots, L-1$ . Tan pronto como se detecta un pico en la línea índice  $i$ , se aplica un modelo sinusoidal para estimar los parámetros de amplitud, de frecuencia y posiblemente de fase de un componente tonal asociado con este pico. Los detalles de esta estimación no se presentan en el presente documento, pero la estimación de frecuencia puede utilizar normalmente una interpolación parabólica de 3 puntos para localizar el máximo de la parábola que se aproxima a los 3 puntos de amplitud  $|U_{HB1}(i+240)|$  (se reduce a dB), obteniéndose la estimación de la amplitud por medio de esta misma interpolación. Dado que el dominio de la transformada utilizado en este caso (DCT-IV) no permite obtener la fase directamente, es posible descartar este término en una realización, pero en las variantes es posible aplicar una transformada en cuadratura de tipo DST para estimar un término de fase. El valor inicial de  $y(i)$  se establece en cero para  $i = 0, \dots, L-1$ . Estimando los parámetros sinusoidales (frecuencia, amplitud y posiblemente fase) de cada componente tonal, se calcula entonces el término  $y(i)$  como la suma de prototipos (espectros) predefinidos de sinusoides puras transformadas en el dominio DCT-IV (u otro si se utiliza otra descomposición en subbandas) según los parámetros sinusoidales estimados. Finalmente, se aplica un valor absoluto a los términos  $y(i)$  para llevar al dominio del espectro de amplitud en valores absolutos.

20 Son posibles otros métodos para determinar las componentes tonales, por ejemplo, también sería posible calcular una envolvente de señal  $env(i)$  mediante interpolación spline de los valores máximos locales (picos detectados) de  $|U_{HB1}(i + 240)|$ , para reducir esta envolvente en un cierto nivel en dB para detectar componentes tonales tales como picos que exceden esta envolvente y definir  $y(i)$  como

$$y(i) = \max(|U_{HB1}(i + 240)| - env(i), 0)$$

25 En esta variante, la señal ambiental se obtiene por tanto por la ecuación:

$$lev(i) = |U_{HB1}(i + 240)| - y(i), \quad i = 0, \dots, L-1$$

30 En otras variantes de la invención, el valor absoluto de los valores espectrales será reemplazado, por ejemplo, por el cuadrado de los valores espectrales, sin cambiar el principio de la invención; en este caso será necesaria una raíz cuadrada para volver al dominio de la señal, lo que es más complejo de realizar.

35 El módulo de combinación 513 realiza una etapa de combinación mediante una mezcla adaptativa de la señal ambiental y las componentes tonales. Para ello, se define un factor  $\Gamma$  de control del nivel ambiental mediante la siguiente ecuación:

$$\Gamma = \beta \frac{ener_{HB} - ener_{tonal}}{ener_{HB} - \beta ener_{tonal}}$$

40 siendo  $\beta$  un factor, del que se aporta un ejemplo más adelante.

Para obtener la señal extendida, se obtiene primero la señal combinada en valores absolutos para  $i = 0 \dots L-1$ :

$$y'(i) = \begin{cases} \Gamma y(i) + \frac{1}{\Gamma} lev(i) & y(i) > 0 \\ y(i) + \frac{1}{\Gamma} lev(i) & y(i) \leq 0 \end{cases}$$

45 a la que se aplican los signos de  $U_{HB1}(k)$ :

$$y''(i) = \text{sgn}(U_{HB1}(i+240)) \cdot y'(i)$$

50 donde la función  $\text{sgn}(\cdot)$  da el signo:

$$\text{sgn}(x) = \begin{cases} 1 & x \geq 0 \\ -1 & x < 0 \end{cases}$$

55 Por definición, el factor  $\Gamma$  es  $>1$ . Las componentes tonales, detectadas línea a línea por la condición  $y(i) > 0$ , son reducidas por el factor  $\Gamma$ ; el nivel medio está amplificado por el factor  $1/\Gamma$ .

En el bloque 513 de mezcla adaptativa, se calcula un factor de control del nivel de energía en función de la energía total de la señal de banda baja decodificada (o decodificada y extendida) y de las componentes tonales.

60 En un modo de realización preferido de la mezcla adaptativa, el ajuste de energía se realiza de la siguiente manera:

$$U_{HB2}(k) = fac.y''(k - 240), k = 240, \dots, 319$$

siendo  $U_{HB2}(k)$  la señal combinada de extensión de banda.

5 El factor de ajuste viene definido por la siguiente ecuación:

$$fac = \gamma \sqrt{\frac{ener_{HB}}{\sum_{i=0}^{L-1} y''(i)}}$$

Donde  $\gamma$  permite evitar una sobreestimación de la energía. En un ejemplo de realización, se calcula  $\beta$  de manera que se mantenga el mismo nivel de señal ambiental con respecto a la energía de las componentes tonales en las bandas consecutivas de la señal. Se calcula la energía de las componentes tonales en tres bandas: 2000-4000 Hz, 4000-6000 Hz y 6000-8000 Hz, con

$$E_{N2-4} = \sum_{k \in N(80,159)} U'^2(k)$$

$$E_{N4-6} = \sum_{k \in N(160,239)} U'^2(k)$$

$$E_{N4-6} = \sum_{k \in N(240,319)} U'^2(k)$$

donde

$$U'(k) = \begin{cases} \sqrt{\frac{\sum_{k=160}^{239} U^2(k)}{\sum_{k=80}^{159} U^2(k)}} U(k) & k = 80, \dots, 159 \\ U(k) & k = 160, \dots, 239 \\ \sqrt{\frac{\sum_{k=160}^{239} U^2(k)}{\sum_{k=240}^{319} U_{HB1}^2(k)}} U_{HB1}(k) & k = 240, \dots, 319 \end{cases}$$

Y donde  $N(k_1, k_2)$  es el conjunto de índices  $k$  para los que el coeficiente de índice  $k$  está clasificado como asociado a las componentes tonales. Este conjunto se puede obtener, por ejemplo, detectando los picos locales en  $U'(k)$  verificando  $|U'(k)| > lev(k)$  donde  $lev(k)$  se calcula como el nivel medio del espectro línea a línea.

Se puede observar que son posibles otros métodos para calcular la energía de las componentes tonales, por ejemplo, tomando el valor mediano del espectro en la banda considerada.

Se fija  $\beta$  de modo que la relación entre la energía de las componentes tonales en las bandas de 4-6 kHz y 6-8 kHz sea la misma que entre las bandas de 2-4 kHz y 4-6 kHz:

$$\beta = \frac{\rho - E_{N6-8}}{\sum_{k=160}^{239} U^2(k) - E_{N6-8}}$$

donde

$$E_{N4-6} = \max(E_{N4-6}, E_{N2-4}), \rho = \frac{E_{N4-6}^2}{E_{N2-4}}, \rho = \max(\rho, E_{N6-8})$$

y  $\max(\dots)$  es la función que da el máximo de los dos argumentos.

En unas variantes de la invención, el cálculo de  $\beta$  se podría reemplazar por otros métodos. Por ejemplo, en una variante, se podría extraer (calcular) diferentes parámetros (o "features" en inglés) que caracterizan la señal en banda baja, incluyendo un parámetro de "tilt" (o pendiente) similar al calculado en el códec AMR-WB, y se estimará el factor  $\beta$  en función de una regresión lineal a partir de estos diferentes parámetros limitando su valor entre 0 y 1. La regresión lineal podría estimarse, por ejemplo, de manera supervisada estimando el factor  $\beta$  aportándose la banda alta original en una base de aprendizaje. Se observará que el modo de cálculo de  $\beta$  no limita la naturaleza de la invención.

A continuación, el parámetro  $\beta$  se puede utilizar para calcular  $\gamma$  teniendo en cuenta el hecho de que una señal con una señal ambiental añadida en una banda dada se percibe generalmente como más fuerte que una señal armónica con la misma energía en la misma banda. Si se define  $\alpha$  como la cantidad de señal ambiental añadida a la señal armónica:

$\alpha = \sqrt{1 - \beta}$  se podría calcular  $\gamma$  como una función decreciente de  $\alpha$ , por ejemplo,  $\gamma = b - a\sqrt{\alpha}$ ,  $b = 1,1$ ,  $a = 1,2$  y estando  $\gamma$  limitado entre 0,3 y 1. De nuevo, otras definiciones de  $\alpha$  y  $\gamma$  son posibles en el contexto de la invención.

A la salida del dispositivo de extensión de banda 500, el bloque 501, en un modo de realización particular realizado, una doble operación de aplicación de la respuesta en frecuencia del filtro paso banda y de filtrado de desacentuación (o desénfasis) en el dominio de la frecuencia.

En una variante de la invención no cubierta por las reivindicaciones, el filtrado de desacentuación se podrá realizar en el dominio del tiempo, después del bloque 502 o incluso antes del bloque 510; sin embargo, en ese caso, el filtrado paso banda realizado en el bloque 501 puede dejar ciertas componentes de baja frecuencia de niveles muy bajos que se amplifican por desacentuación, lo que puede modificar ligeramente de forma ligeramente perceptible la banda baja decodificada. Por esta razón, en el presente documento se prefiere realizar la desacentuación en el dominio de la frecuencia. En el modo de realización preferido, los coeficientes del índice  $k = 0, \dots, 199$  se ponen a cero, de este modo, la desacentuación está limitada a los coeficientes superiores.

La excitación primero se desacentúa de acuerdo con la siguiente ecuación:

$$U_{HB2}'(k) = \begin{cases} 0 & k = 0, \dots, 199 \\ G_{deemph}(k - 200)U_{HB2}(k) & k = 200, \dots, 255 \\ G_{deemph}(55)U_{HB2}(k) & k = 256, \dots, 319 \end{cases}$$

donde  $G_{deemph}(k)$  es la respuesta en frecuencia del filtro  $1/(1-0,68z^{-1})$  sobre una banda de frecuencia discreta restringida. Teniendo en cuenta las frecuencias discretas (impares) de la DCT-IV, en el presente documento se define  $G_{deemph}(k)$  como:

$$G_{deemph}(k) = \frac{1}{|e^{j\theta_k} - 0,68|}, k = 0, \dots, 255$$

donde

$$\theta_k = \frac{256 - 80 + k + \frac{1}{2}}{256}.$$

En el caso de que se utilice una transformación distinta a la DCT-IV, la definición de  $\theta_k$  se podría ajustar (por ejemplo, para frecuencias pares).

Se observará que la desacentuación se aplica en dos fases para  $k = 200, \dots, 255$  correspondiente a la banda de frecuencia 5000-6400 Hz, donde la respuesta  $1/(1-0,68z^{-1})$  se aplica como a 12,8 kHz, y para  $k = 256, \dots, 319$  correspondiente a la banda de frecuencia 6400-8000 Hz, donde la respuesta se extiende de 16 kHz, que en este caso tiene un valor constante, en la banda de 6,4-8 kHz.

Se puede observar que en el códec AMR-WB, la síntesis de AF no está desacentuada.

En el modo de realización que se presenta en el presente documento, la señal de alta frecuencia, por el contrario, se desacentúa para llevarla a un dominio consistente con la señal de baja frecuencia (0-6,4 kHz) que sale del bloque 305

de la figura 3. Esto es importante para la posterior estimación y ajuste de la energía de la síntesis de AF.

En una variante del modo de realización, para reducir la complejidad, se podría fijar  $G_{deemph}(k)$  a un valor constante independiente de  $k$ , tomando, por ejemplo,  $G_{deemph}(k) = 0,6$  lo que corresponde aproximadamente al valor medio de  $G_{deemph}(k)$  para  $k = 200, \dots, 319$  en las condiciones del modo de realización descrito anteriormente.

En otra variante del modo de realización del decodificador, la desacentuación podría realizarse de manera equivalente en el dominio del tiempo después de la DCT inversa.

Además de la desacentuación, se aplica un filtrado paso banda con dos partes separadas: una fija de paso alto, la otra de paso bajo adaptativo (en función de la velocidad).

Este filtrado se realiza en el dominio de la frecuencia.

En el modo de realización más preferido, la respuesta parcial del filtro paso bajo en el dominio de la frecuencia se calcula de la siguiente manera:

$$G_{lp}(k) = 1 - 0,999 \frac{k}{N_{lp} - 1}$$

donde  $N_p = 60$  a 6,6 kbit/s, 40 a 8,85 kbit/s, 20 a velocidades  $>8,85$  bit/s.

A continuación, se aplica un filtro paso banda en forma de:

$$U_{HB3}(k) = \begin{cases} 0 & k = 0, \dots, 199 \\ G_{hp}(k - 200)U_{HB2}'(k) & k = 200, \dots, 255 \\ U_{HB2}'(k) & k = 256, \dots, 319 - N_{lp} \\ G_{lp}(k - 320 - N_{lp})U_{HB2}'(k) & k = 320 - N_{lp}, \dots, 319 \end{cases}$$

La definición de  $G_{hp}(k)$ ,  $k = 0, \dots, 55$ , se da, por ejemplo, en la siguiente tabla 1.

Tabla 1

| $K$ | $g_{hp}(k)$ | $K$ | $g_{hp}(k)$ | $K$ | $g_{hp}(k)$ | $k$ | $g_{hp}(k)$ |
|-----|-------------|-----|-------------|-----|-------------|-----|-------------|
| 0   | 0,001622428 | 14  | 0,114057967 | 28  | 0,403990611 | 42  | 0,776551214 |
| 1   | 0,004717458 | 15  | 0,128865425 | 29  | 0,430149896 | 43  | 0,800503267 |
| 2   | 0,008410494 | 16  | 0,144662643 | 30  | 0,456722014 | 44  | 0,823611104 |
| 3   | 0,012747280 | 17  | 0,161445005 | 31  | 0,483628433 | 45  | 0,845788355 |
| 4   | 0,017772424 | 18  | 0,179202219 | 32  | 0,510787115 | 46  | 0,866951597 |
| 5   | 0,023528982 | 19  | 0,197918220 | 33  | 0,538112915 | 47  | 0,887020781 |
| 6   | 0,030058032 | 20  | 0,217571104 | 34  | 0,565518011 | 48  | 0,905919644 |
| 7   | 0,037398264 | 21  | 0,238133114 | 35  | 0,592912340 | 49  | 0,923576092 |
| 8   | 0,045585564 | 22  | 0,259570657 | 36  | 0,620204057 | 50  | 0,939922577 |
| 9   | 0,054652620 | 23  | 0,281844373 | 37  | 0,647300005 | 51  | 0,954896429 |
| 10  | 0,064628539 | 24  | 0,304909235 | 38  | 0,674106188 | 52  | 0,968440179 |
| 11  | 0,075538482 | 25  | 0,328714699 | 39  | 0,700528260 | 53  | 0,980501849 |
| 12  | 0,087403328 | 26  | 0,353204886 | 40  | 0,726472003 | 54  | 0,991035206 |
| 13  | 0,100239356 | 27  | 0,378318805 | 41  | 0,751843820 | 55  | 1,000000000 |

Se observará que en las variantes de la invención los valores de  $G_{hp}(k)$  se podrían modificar manteniendo una atenuación progresiva. Asimismo, el filtrado paso bajo con ancho de banda variable,  $G_{lp}(k)$ , se podría ajustar con diferentes valores o un soporte de frecuencia diferente, sin cambiar el principio de esta etapa de filtrado.

También se observará que el filtrado paso banda se puede adaptar definiendo una sola etapa de filtrado que combine el filtrado paso alto y el paso bajo.

En otro modo de realización, el filtrado paso banda se puede realizar de manera equivalente en el dominio del tiempo (como en el bloque 112 de la figura 1) con diferentes coeficientes de filtro según la velocidad, después de una etapa DCT inversa. Sin embargo, se observará que es ventajoso realizar esta etapa directamente en el dominio de la frecuencia porque el filtrado se realiza en el dominio de la excitación LPC y por lo tanto los problemas de convolución

circular y de efectos de borde son muy limitados en este dominio.

El bloque 502 de transformada inversa realiza una DCT inversa sobre 320 muestras para encontrar la señal de alta frecuencia muestreada a 16 kHz. Su implementación es idéntica a la del bloque 510, porque la DCT-IV es ortonormal, salvo porque la longitud de la transformada es de 320 en lugar de 256, y se obtiene:

$$u_{HB}(n) = \sum_{k=0}^{N_{16k}-1} U_{HB3}(k) \cos\left(\frac{\pi}{N_{16k}}\left(k + \frac{1}{2}\right)\left(n + \frac{1}{2}\right)\right)$$

donde

$$N_{16k} = 320 \text{ y } k = 0, \dots, 319.$$

En caso de que el bloque 510 no sea una DCT, sino otra transformación o descomposición en subbandas, el bloque 502 realiza la síntesis correspondiente al análisis realizado en el bloque 510.

La señal muestreada a 16 kHz se pone entonces opcionalmente a escala mediante las ganancias definidas por subtrama de 80 muestras (bloque 504).

En un modo de realización preferido, se calcula primero (bloque 503) una ganancia  $g_{HB1}(m)$  por subtrama mediante las relaciones de energía de las subtramas, tal que en cada subtrama de índice  $m=0, 1, 2$  o  $3$  de la trama actual:

$$g_{HB1}(m) = \sqrt{\frac{e_3(m)}{e_2(m)}}$$

donde

$$e_1(m) = \sum_{n=0}^{63} u(n+64m)^2 + \varepsilon$$

$$e_2(m) = \sum_{n=0}^{79} u_{HB}(n+80m)^2 + \varepsilon$$

$$e_3(m) = e_1(m) \frac{\sum_{n=0}^{319} u_{HB}(n)^2 + \varepsilon}{\sum_{n=0}^{255} u(n)^2 + \varepsilon}$$

siendo  $\varepsilon = 0,01$ . Se puede expresar la ganancia por subtrama  $g_{HB1}(m)$  en forma de:

$$g_{HB1}(m) = \sqrt{\frac{\sum_{n=0}^{63} u(n+64m)^2 + \varepsilon}{\sum_{n=0}^{255} u(n)^2 + \varepsilon} \frac{\sum_{n=0}^{79} u_{HB}(n+80m)^2 + \varepsilon}{\sum_{n=0}^{319} u_{HB}(n)^2 + \varepsilon}}$$

lo que demuestra que en la señal  $u_{HB}$  se garantiza la misma relación entre la energía por subtrama y la energía por

trama que en la señal  $u(n)$ .

El bloque 504 realiza la puesta a escala de la señal combinada (comprendida en la etapa E404a de la figura 4) según la siguiente ecuación:

$$u_{HB}'(n) = g_{HB1}(m)u_{HB}(n), n = 80m, \dots, 80(m+1)-1$$

Se observará que la realización del bloque 503 difiere de la del bloque 101 de la figura 1, porque se tiene en cuenta la energía al nivel de la trama actual además de la de la subtrama. Esto permite tener la relación de la energía de cada subtrama con respecto a la energía de la trama. Por lo tanto, se comparan las relaciones de energía (o energías relativas) en lugar de las energías absolutas entre la banda baja y la banda alta.

De este modo, esta etapa de puesta a escala permite conservar en la banda alta la relación de energía entre la subtrama y la trama de la misma forma que en la banda baja.

Opcionalmente, el bloque 506 realiza a continuación la puesta a escala de la señal (comprendida en la etapa E404a de la figura 4) de acuerdo con la siguiente ecuación:

$$u_{HB}''(n) = g_{HB2}(m)u_{HB}'(n), n = 80m, \dots, 80(m+1)-1$$

donde la ganancia  $g_{HB2}(m)$  se obtiene a partir del bloque 505 ejecutando los bloques 103, 104 y 105 del códec AMR-WB (siendo la entrada del bloque 103 la excitación de banda baja decodificada,  $u(n)$ ). Los bloques 505 y 506 son útiles para ajustar el nivel del filtro de síntesis LPC (bloque 507), en este caso en función del *tilt* de la señal. Otros métodos para calcular la ganancia  $g_{HB2}(m)$  son posibles sin cambiar la naturaleza de la invención.

Por último, la señal,  $u_{HB}'(n)$  o  $u_{HB}''(n)$ , es filtrada por el módulo de filtrado 507 que en este caso se puede realizar tomando como función de transferencia  $1/\hat{A}(z/\gamma)$ , donde  $\gamma = 0,9$  a  $6,6$  kbit/s y  $\gamma = 0,6$  a otras velocidades, lo que limita el orden del filtro al orden 16.

En una variante, este filtrado se podría realizar de la misma forma que se describe para el bloque 111 de la figura 1 del decodificador AMR-WB, sin embargo, el orden del filtro pasa a 20 a una velocidad de  $6,6$ , lo que no cambia significativamente la calidad de la señal sintetizada. En otra variante, el filtrado de síntesis LPC se podría realizar en el dominio de la frecuencia, después de haber calculado la respuesta en frecuencia del filtro implementado en el bloque 507.

En unas variantes de realización de la invención, la codificación de la banda baja (0-6,4 kHz) podrá reemplazarse por un codificador CELP distinto al utilizado en AMR-WB, como, por ejemplo, el codificador CELP de G.718 a 8 kbit/s. Sin pérdida de generalidad de otros codificadores en banda ampliada o funcionando a frecuencias superiores a 16 kHz, en los que se podría utilizar la codificación de la banda baja a una frecuencia interna de 12,8 kHz. Asimismo, obviamente, la invención se puede adaptar a otras frecuencias de muestreo distintas de 12,8 kHz, cuando un codificador de bajas frecuencias funciona a una frecuencia de muestreo inferior a la de la señal original o reconstruida. Cuando la decodificación de banda baja no utiliza una predicción lineal, no se dispone de una señal de excitación para extenderla, en ese caso, se podrá realizar un análisis LPC de la señal reconstruida en la trama actual y se calculará una excitación LPC para poder aplicar la invención.

Por último, en otra variante de la invención, la excitación o la señal de banda baja ( $u(n)$ ) se vuelve a muestrear, por ejemplo, por interpolación lineal o "spline" cúbica, de 12,8 a 16 kHz antes de la transformación (por ejemplo, DCT-IV) de longitud 320. Esta variante tiene el inconveniente de ser más compleja, porque la transformada (DCT-IV) de la excitación o de la señal se calcula entonces sobre una longitud mayor y no se realiza el remuestreo en el dominio de la transformada.

Además, en unas variantes de la invención, todos los cálculos necesarios para estimar las ganancias ( $G_{HBN}$ ,  $g_{HB1}(m)$ ,  $g_{HB2}(m)$ ,  $g_{HBN}, \dots$ ) se podrían realizar en un dominio logarítmico.

La figura 6 representa un ejemplo de realización material de un dispositivo de extensión de banda 600 según la invención. Este puede formar parte integrante de un decodificador de señal de audiofrecuencia o de un equipo que recibe señales de audiofrecuencias, decodificadas o no.

Este tipo de dispositivo incluye un procesador PROC que coopera con un bloque de memoria BM que incluye una memoria de almacenamiento y/o de trabajo MEM.

Dicho dispositivo incluye un módulo de entrada E capaz de recibir una señal de audio decodificada o extraída en una primera banda de frecuencia denominada banda baja llevada al dominio de la frecuencia. ( $U(k)$ ). Incluye un módulo de salida S capaz de transmitir la señal de extensión en una segunda banda de frecuencia ( $U_{HB2}(k)$ ), por ejemplo, a un módulo de filtrado 501 de la figura 5.

El bloque de memoria puede incluir, ventajosamente, un programa informático que incluya instrucciones de código para la implementación de las etapas del procedimiento de extensión de banda en el sentido de la invención, cuando estas instrucciones son ejecutadas por el procesador PROC y, en concreto, las etapas de extracción (E402) de componentes tonales y de una señal ambiental a partir de una señal que se origina en la señal de banda baja decodificada ( $U(k)$ ), de combinación (E403) de las componentes tonales ( $y(k)$ ) y de la señal ambiental ( $U_{HBA}(k)$ ) mediante una mezcla adaptativa utilizando factores de control del nivel de energía para obtener una señal de audio, denominada señal combinada  $U_{HB2}(k)$ , de extensión (E401a) sobre al menos una segunda banda de frecuencia superior a la primera banda de frecuencia de la señal de banda baja decodificada antes de la etapa de extracción o de la señal combinada después de la etapa de combinación.

Normalmente, la descripción de la figura 4 retoma las etapas de un algoritmo de un programa informático de este tipo. El programa informático también puede estar almacenado en un medio de memoria legible por un lector del dispositivo o puede descargarse en el espacio de memoria del mismo.

La memoria MEM graba de manera general, todos los datos necesarios para la implementación del procedimiento.

En un posible modo de realización, el dispositivo así descrito también puede incluir las funciones de decodificación de banda baja y otras funciones de procesamiento descritas, por ejemplo, en las figuras 5 y 3, además de las funciones de extensión de banda según la invención.

## REIVINDICACIONES

1. Procedimiento de extensión de banda de frecuencia de una señal de audiofrecuencia durante un proceso de decodificación o de mejora que incluye una etapa de obtención de la señal decodificada en una primera banda de frecuencia denominada banda baja, incluyendo el procedimiento las siguientes etapas realizadas en el dominio de la frecuencia:

- extensión (E401a) en al menos una segunda banda de frecuencia por encima de la primera banda de frecuencia de la señal decodificada de banda baja para formar una señal de banda baja decodificada extendida,
- extracción (E402) de componentes tonales y de una señal ambiental a partir de la señal que se origina en la señal de banda baja decodificada extendida;
- combinación (E403) de las componentes tonales y de la señal ambiental mediante una mezcla adaptativa utilizando factores de control del nivel de energía para obtener una señal combinada;

estando el método caracterizado por que comprende las siguientes etapas realizadas en el dominio de la frecuencia:

- aplicación de filtrado de desacentuación y aplicación de respuesta de frecuencia de filtro paso banda en la señal combinada para obtener una señal combinada desacentuada; y
- síntesis (E404b) de una señal de audio para llevar al dominio temporal una señal que se origina en la señal combinada desacentuada.

2. Procedimiento según la reivindicación 1, en el que el filtrado de desacentuación se limita a los coeficientes superiores de la señal combinada.

3. Procedimiento según la reivindicación 2, en el que la señal combinada se desacentúa según la ecuación:

$$U'_{HB2}(k) = \begin{cases} 0 & k = 0, \dots, 199 \\ G_{deemph}(k-200)U_{HB2}(k) & k = 200, \dots, 255 \\ G_{deemph}(55)U_{HB2}(k) & k = 256, \dots, 319 \end{cases}$$

donde  $G_{deemph}(k)$  es la respuesta de frecuencia del filtro  $1/(1-0,68z^{-1})$  en una banda de frecuencia discreta restringida.

4. Procedimiento según la reivindicación 3, en el que la respuesta de frecuencia  $G_{deemph}(k)$  se define por

$$G_{deemph}(k) = \frac{1}{|e^{j\theta_k - 0,68}|}, k = 0, \dots, 255.$$

$$\text{donde } \theta_k = \frac{256 - 80 + k + \frac{1}{2}}{256}.$$

5. Procedimiento según una de las reivindicaciones anteriores, en el que el filtro paso banda se aplica con una parte de paso alto fija y una parte de paso bajo adaptativa.

6. Procedimiento según la reivindicación 5, en el que la respuesta parcial del filtro paso bajo en el dominio de la frecuencia es de la forma:

$$G_{lp}(k) = 1 - 0,999 \frac{k}{N_{lp}-1}$$

donde  $N_{lp}$  = 60 a 6,6 kbit/s, 40 a 8,85 kbit/s, 20 a velocidades superiores a 8,85 kbit/s.

7. Procedimiento según la reivindicación 6, en el que el filtro paso banda se aplica en la forma:

$$U_{HB3}(k) = \begin{cases} 0 & k = 0, \dots, 199 \\ G_{hp}(k-200)U'_{HB2}(k) & k = 200, \dots, 255 \\ U'_{HB2}(k) & k = 256, \dots, 319 - N_{lp} \\ G_{lp}(k-320-N_{lp})U'_{HB2}(k) & k = 320 - N_{lp}, \dots, 319 \end{cases}$$

donde  $U'_{HB2}(k)$  es la señal combinada desacentuada, y  $G_{hp}(k)$  son los valores del filtro paso alto.

8. Procedimiento según la reivindicación 7, en el que los valores del filtro paso alto  $G_{hp}(k)$  son los siguientes:

| k  | $G_{hp}(k)$ | k  | $G_{hp}(k)$ | k  | $G_{hp}(k)$ | k  | $G_{hp}(k)$ |
|----|-------------|----|-------------|----|-------------|----|-------------|
| 0  | 0,001622428 | 14 | 0,114047967 | 28 | 0,403990611 | 42 | 0,776551214 |
| 1  | 0,004717458 | 15 | 0,128865425 | 29 | 0,430149896 | 43 | 0,800503267 |
| 2  | 0,008410494 | 16 | 0,144662643 | 30 | 0,456722014 | 44 | 0,823611104 |
| 3  | 0,012747280 | 17 | 0,161445005 | 31 | 0,483628433 | 45 | 0,845788355 |
| 4  | 0,017772424 | 18 | 0,179202219 | 32 | 0,510787115 | 46 | 0,866951597 |
| 5  | 0,023528982 | 19 | 0,197918220 | 33 | 0,538112915 | 47 | 0,887020781 |
| 6  | 0,030058032 | 20 | 0,217571104 | 34 | 0,565518011 | 48 | 0,905919644 |
| 7  | 0,037398264 | 21 | 0,238133114 | 35 | 0,592912340 | 49 | 0,923576092 |
| 8  | 0,045585564 | 22 | 0,259570657 | 36 | 0,620204057 | 50 | 0,939922577 |
| 9  | 0,054652620 | 23 | 0,281844373 | 37 | 0,647300005 | 51 | 0,954896429 |
| 10 | 0,064628539 | 24 | 0,304909235 | 38 | 0,674106188 | 52 | 0,968440179 |
| 11 | 0,075538482 | 25 | 0,328714699 | 39 | 0,700528260 | 53 | 0,980501849 |
| 12 | 0,087403328 | 26 | 0,353204886 | 40 | 0,726472003 | 54 | 0,991035206 |
| 13 | 0,100239356 | 27 | 0,378318805 | 41 | 0,751843820 | 55 | 1,000000000 |

9. Dispositivo de extensión de banda de frecuencia de una señal de audiofrecuencia, habiéndose decodificado la señal en una primera banda de frecuencia denominada banda baja, incluyendo el dispositivo:

- un módulo de extensión (512), en el dominio de la frecuencia sobre al menos una segunda banda de frecuencia superior a la primera banda de frecuencia implementada en la señal de banda baja decodificada para formar la señal de banda baja extendida  $U_{HB1}(k)$ ,
- un módulo de extracción (512) en el dominio de la frecuencia de componentes tonales y de una señal ambiental a partir de una señal procedente de la señal de banda baja decodificada extendida;
- un módulo de combinación (513) de las componentes tonales y de la señal ambiental en el dominio de la frecuencia mediante mezcla adaptativa utilizando factores de control del nivel de energía para obtener una señal combinada;

estando el dispositivo caracterizado por que incluye:

- un bloque (501) de filtrado de desacentuación y de filtrado de frecuencia de paso banda, en el dominio de la frecuencia para obtener una señal combinada desacentuada; y
- un módulo de síntesis (502) de una señal de audio para llevar la señal combinada desacentuada al dominio del tiempo.

10. Decodificador de señales de audiofrecuencia caracterizado por que incluye un dispositivo de extensión de banda de frecuencia conforme a la reivindicación 9.

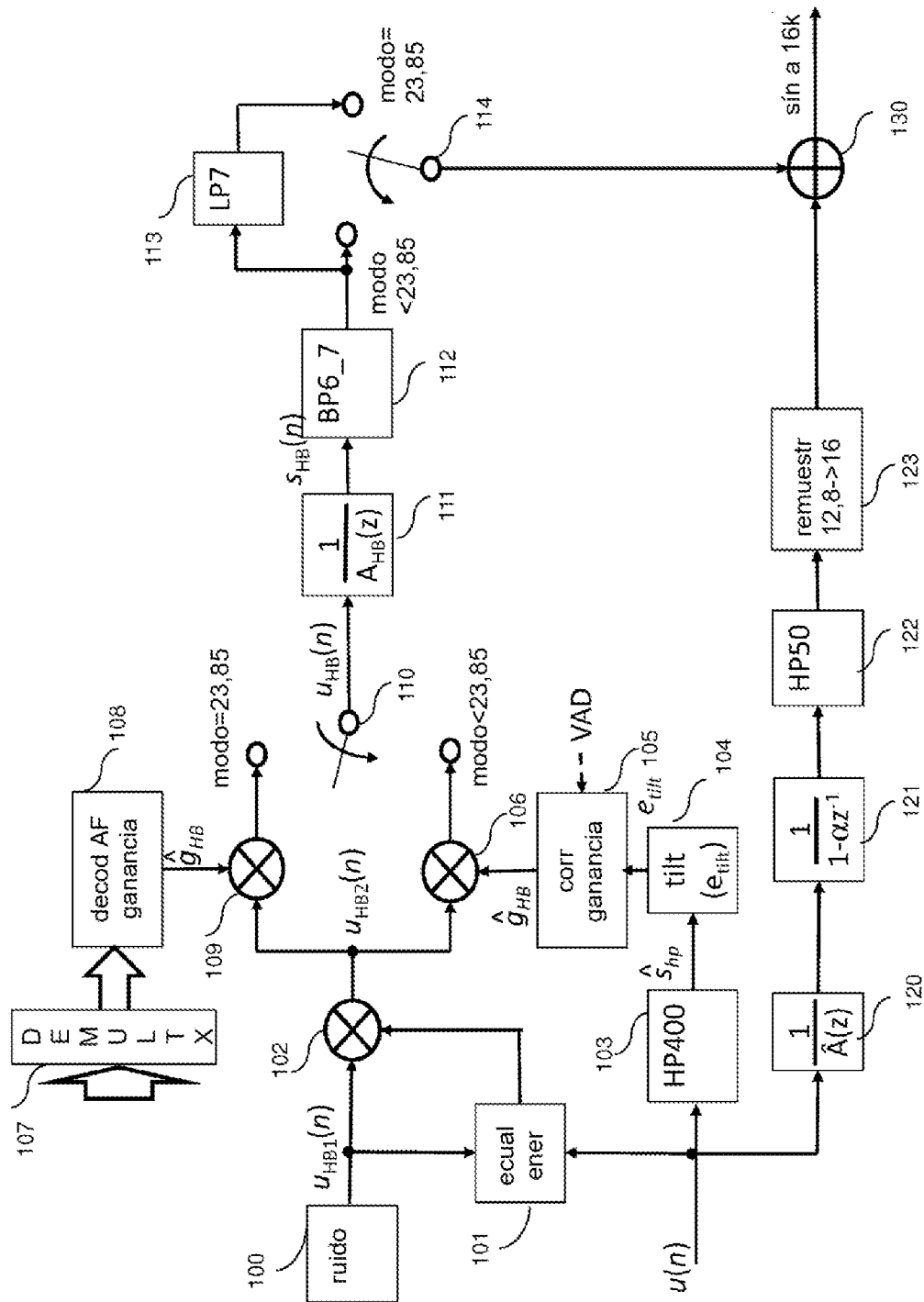


Fig.1

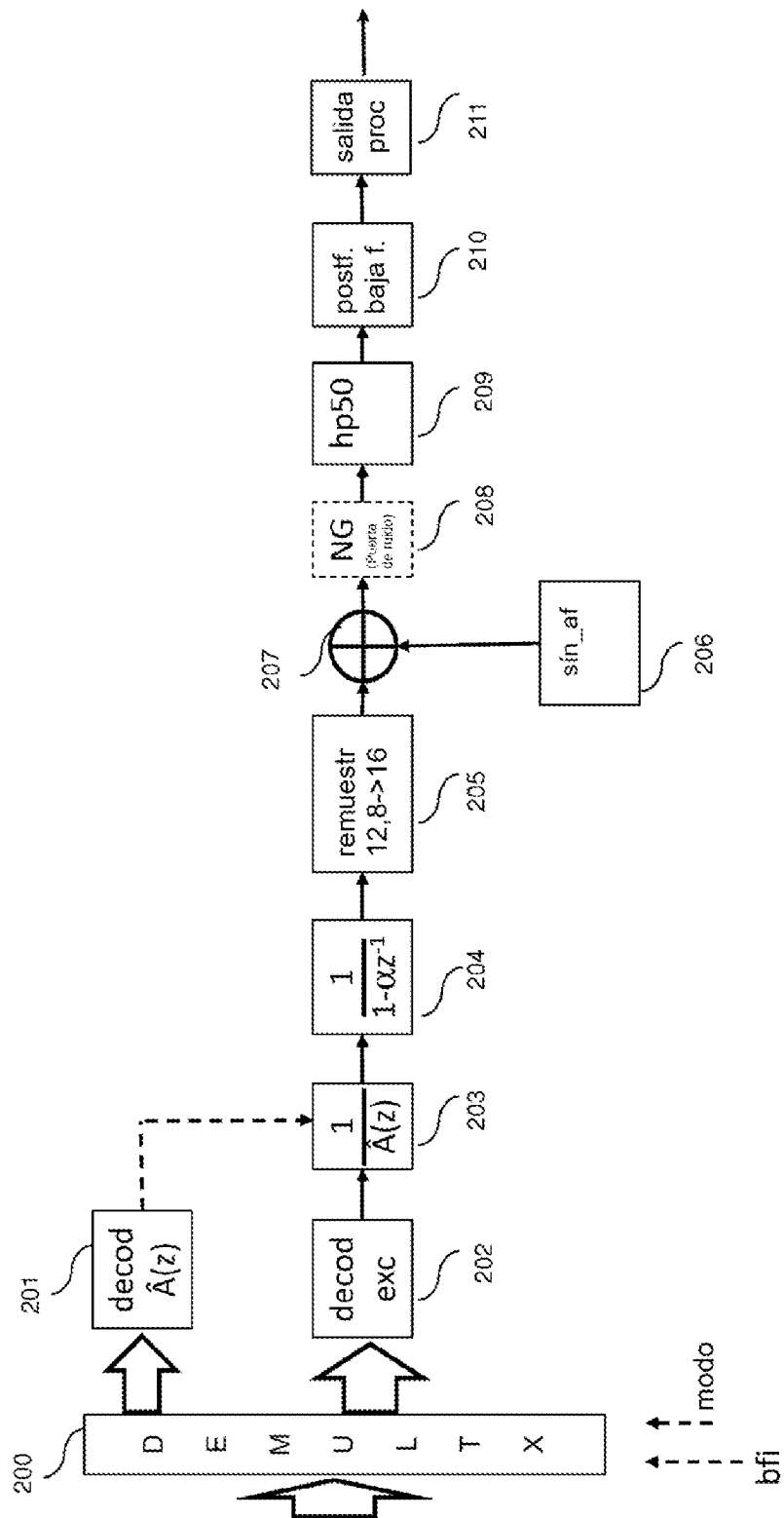
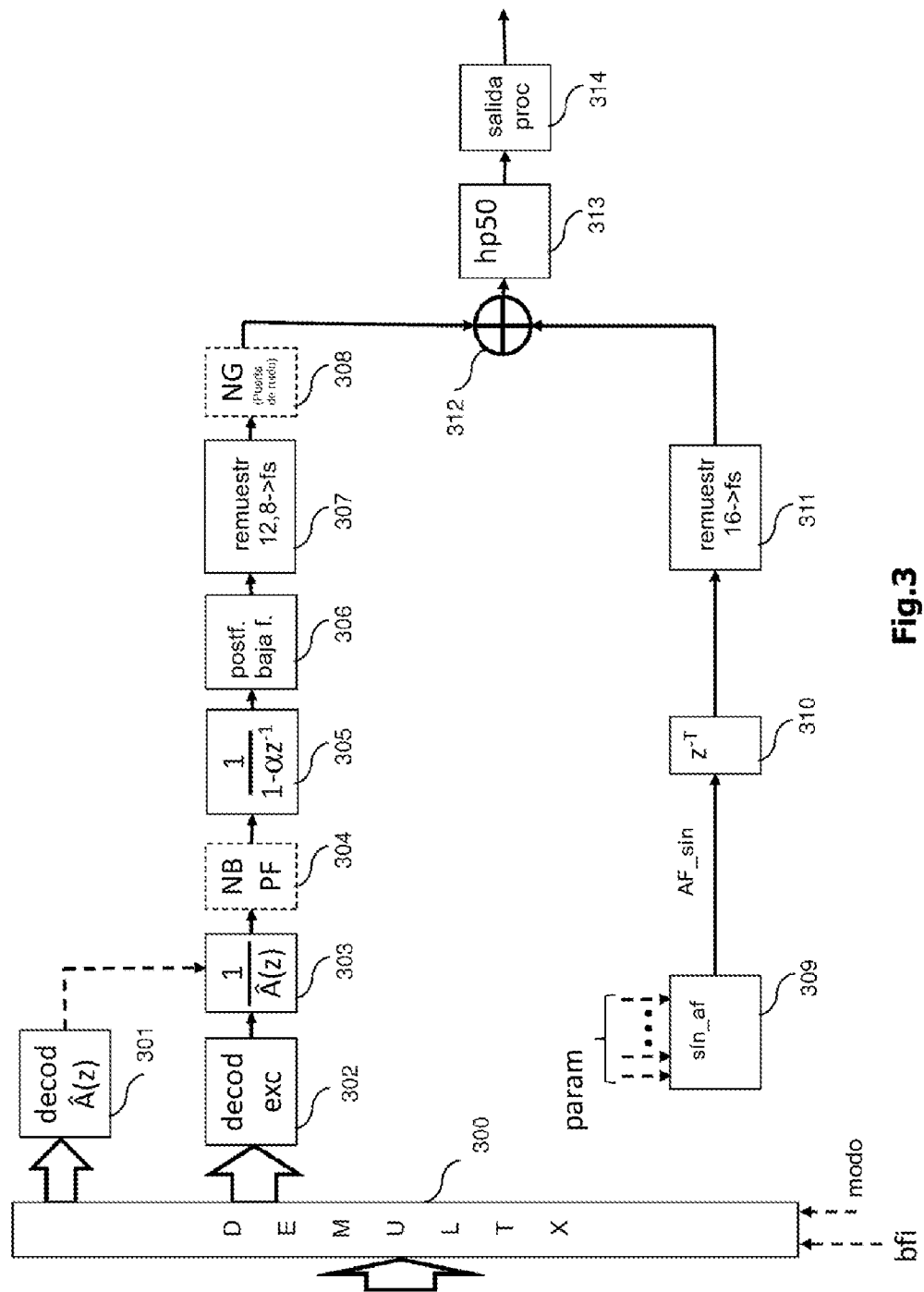
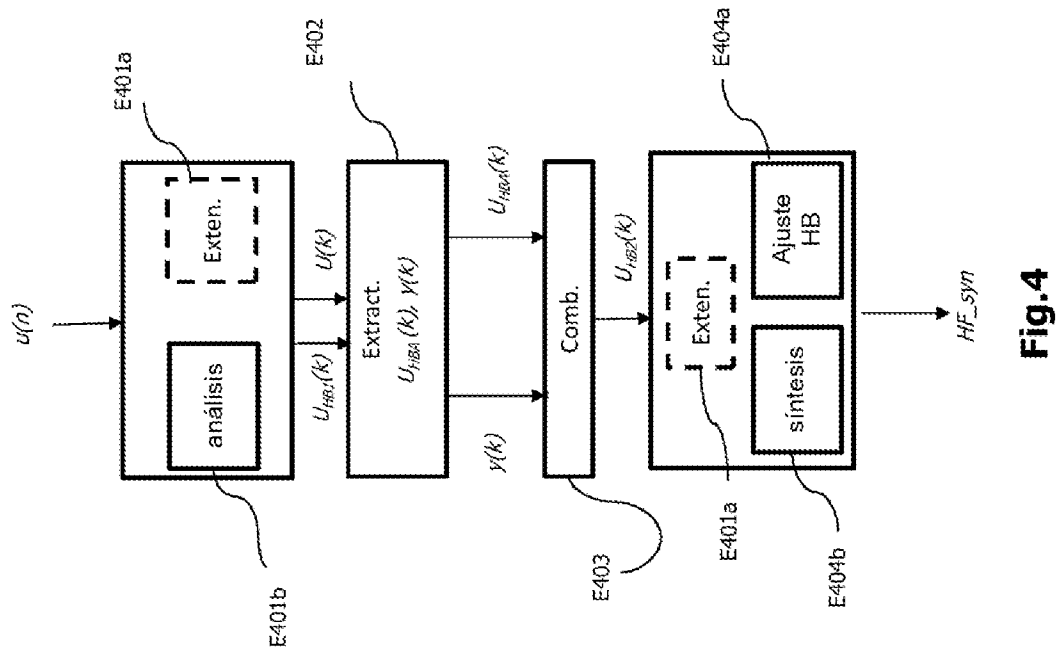
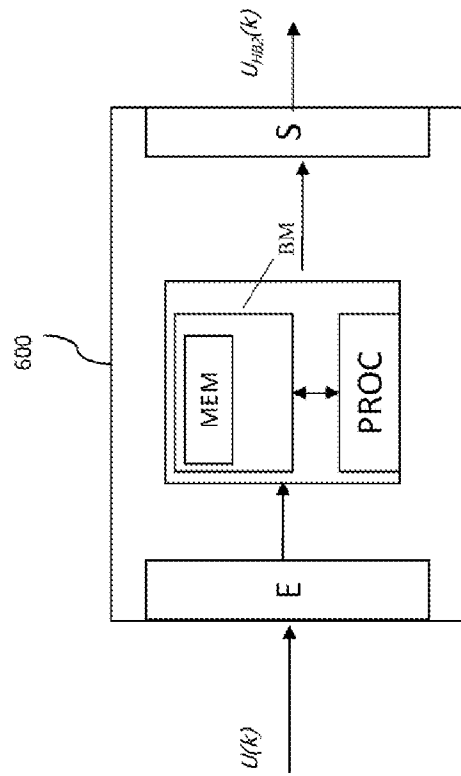


Fig.2





**Fig.4**



**Fig.6**

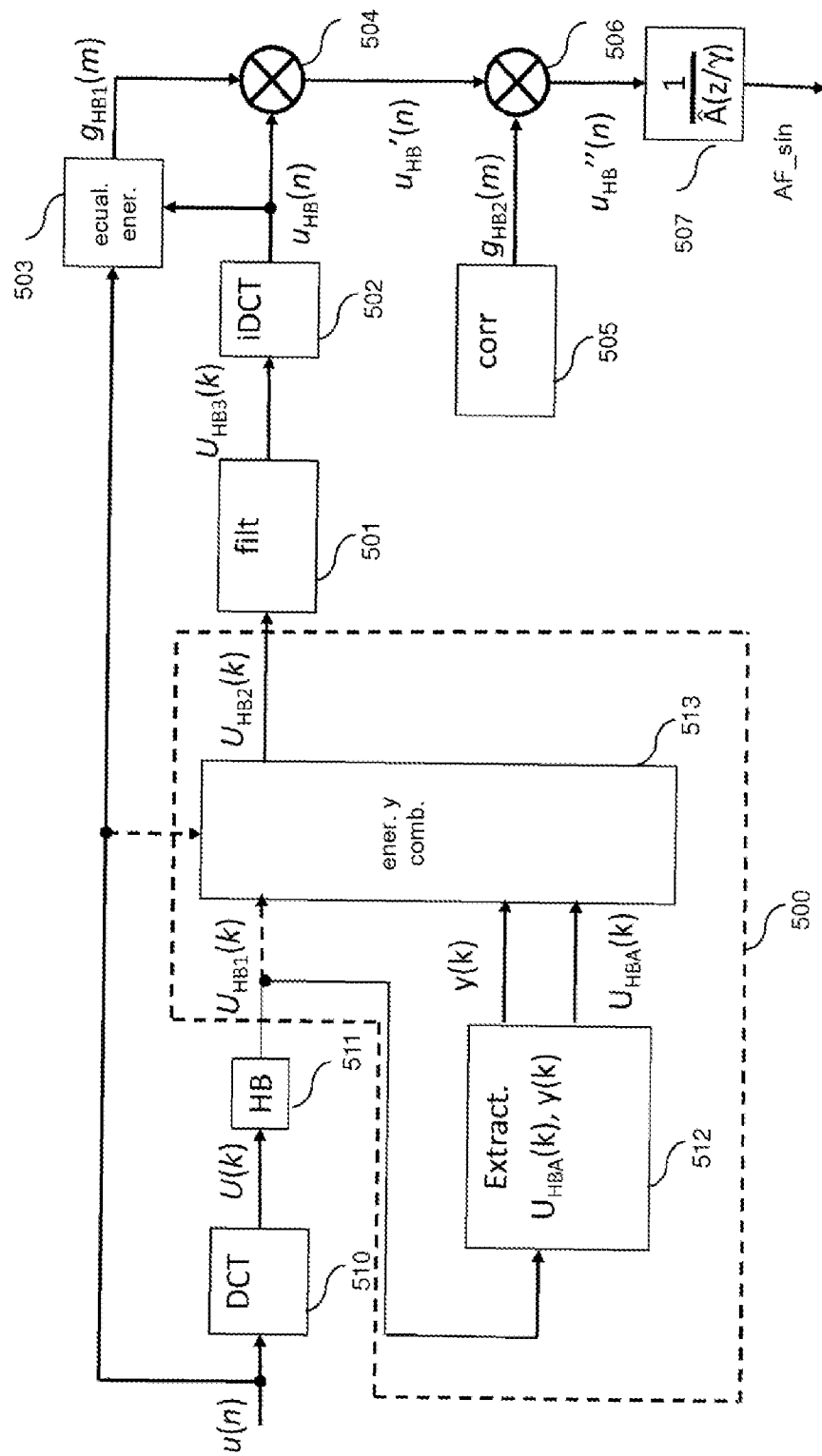


Fig.5