

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 377 551**

51 Int. Cl.:
G10L 21/02 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 96 Número de solicitud europea: **10712439 .8**
96 Fecha de presentación: **01.04.2010**
97 Número de publicación de la solicitud: **2269189**
97 Fecha de publicación de la solicitud: **05.01.2011**

54 Título: **Aparato, método y programa informático para generar una representación de una señal de ancho de banda ampliado basándose en una representación de señal de entrada usando una combinación de una ampliación de ancho de banda armónica y una ampliación de ancho de banda no armónica**

30 Prioridad:
02.04.2009 US 166125 P
09.04.2009 US 168068 P
30.12.2009 EP 09181008

45 Fecha de publicación de la mención BOPI:
28.03.2012

45 Fecha de la publicación del folleto de la patente:
28.03.2012

73 Titular/es:
**Fraunhofer-Gesellschaft zur Förderung der
angewandten Forschung e.V.
Hansastraße 27c
80686 München, DE**

72 Inventor/es:
**NAGEL, Frederik;
NEUENDORF, Max;
RETTELBACH, Nikolaus;
LECOMTE, Jérémie;
MULTRUS, Markus;
GRILL, Bernhard y
DISCH, Sascha**

74 Agente/Representante:
Arizti Acha, Monica

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).

ES 2 377 551 T3

DESCRIPCIÓN

5 Aparato, método y programa informático para generar una representación de una señal de ancho de banda ampliado basándose en una representación de señal de entrada usando una combinación de una ampliación de ancho de banda armónica y una ampliación de ancho de banda no armónica.

Campo técnico

10 Realizaciones según la invención se refieren a un aparato para generar una representación de una señal de ancho de banda ampliado basándose en una representación de señal de entrada. Otras realizaciones según la invención se refieren a un método para generar una representación de una señal de ancho de banda ampliado basándose en una representación de señal de entrada. Realizaciones adicionales según la invención se refieren a un programa informático para realizar tal método.

Algunas realizaciones según la invención se refieren a métodos de parcheado novedosos dentro de la replicación de banda espectral.

Antecedentes de la invención

15 El almacenamiento o transmisión de señales de audio está sujeto a menudo a estrictas restricciones de tasa de transmisión de bits. Estas restricciones se superan habitualmente mediante una codificación de la señal. En el pasado, se forzaba a los codificadores a reducir drásticamente el ancho de banda del audio transmitido cuando sólo había disponible una tasa de transmisión de bits muy baja. Los códecs de audio modernos pueden preservar hoy en día el ancho de banda audible usando métodos de ampliación de ancho de banda (BWE). Tales métodos se describen, por ejemplo, en las referencias [1] a [12]. Estos algoritmos se basan en una representación paramétrica del contenido de alta frecuencia (HF), que se genera a partir de la parte de baja frecuencia (LF) con forma de onda codificada de la señal decodificada por medio de transposición a la región espectral de HF ("parcheado") y la aplicación de un procesamiento posterior dirigido por parámetros.

20 En la técnica, métodos de ampliación de ancho de banda, tales como replicación de banda espectral (SBR), se usan como un método eficaz para generar señales de alta frecuencia en códecs basados en HFR (reconstrucción de alta frecuencia).

30 La replicación de banda espectral descrita en la referencia [1], que también se designa de manera abreviada como "SBR", usa un banco de filtros de espejo en cuadratura (QMF) para generar la información de HF. Con la ayuda del denominado proceso de "parcheado", se copian bandas QMF más bajas a una posición (frecuencia) más alta dando lugar a una replicación de la información de la parte de LF en la parte de HF. La parte de HF generada se adapta después a la parte de HF original con la ayuda de parámetros que adoptan (o ajustan) la envolvente espectral y la tonalidad (por ejemplo usando un formateado de envolvente).

35 En SBR estándar, el parcheado siempre se lleva a cabo mediante una operación de copia dentro del dominio de QMF. Se ha descubierto que esto a veces puede llevar a artefactos auditivos, particularmente si se copian sinusoides en la proximidad unos de otros en el borde de la parte de LF y la de HF generada. Por tanto, puede afirmarse que la SBR estándar tiene el problema de artefactos auditivos. Además, algunas implementaciones convencionales del concepto de ampliación de ancho de banda conllevan una complejidad comparativamente alta. Además, en algunas implementaciones de los conceptos de ampliación de ancho de banda de la invención, el espectro se vuelve muy disperso para parches altos (factores de alargamientos altos), lo que puede dar como resultado artefactos de audio (audibles) no deseados.

40 La referencia [14] da a conocer un método de ampliación de ancho de banda usando transposición de frecuencia del espectro de banda más bajo.

La referencia [15] da a conocer una ampliación de ancho de banda iterativa, en la que se obtienen parches espectrales y se usan para una posterior generación de parches.

45 En vista de los comentarios anteriores, es un objetivo de la presente invención crear un concepto para generar una representación de una señal de ancho de banda ampliado basándose en una representación de señal de entrada, que conlleve un equilibrio mejorado entre complejidad y calidad de audio.

Sumario de la invención

50 El objetivo se consigue proporcionando un aparato según la reivindicación 1, un método según la reivindicación 14, un aparato según la reivindicación 15, un método según la reivindicación 16 y un programa informático según la reivindicación 17.

Realizaciones según la invención crean un aparato para generar una representación de una señal de ancho de banda ampliado basándose en una representación de señal de entrada. El aparato comprende un vocoder de fase configurado para obtener valores de una representación de dominio espectral de un primer parche de la señal de ancho de banda

ampliado basándose en la representación de señal de entrada. El aparato también comprende un copiador de valores configurado para copiar un conjunto de valores de la representación de dominio espectral del primer parche, valores que se proporcionan por el vocoder de fase, para obtener un conjunto de valores de una representación de dominio espectral de un segundo parche. El segundo parche está asociado con frecuencias más altas que el primer parche. El aparato está configurado para obtener la representación de la señal de ancho de banda ampliado usando los valores de la representación de dominio espectral del primer parche y los valores de la representación de dominio espectral del segundo parche.

La idea clave de la presente invención es que se obtiene un equilibrio particularmente bueno entre complejidad computacional y calidad de audio de una señal de ancho de banda ampliado combinando un vocoder de fase con un copiador de valores, de manera que el primer parche de la señal de ancho de banda ampliado se obtiene por el vocoder de fase, y de manera que el segundo parche de la señal de ancho de banda ampliado se obtiene basándose en el primer parche usando el copiador de valores.

Por consiguiente, el contenido del primer parche es una versión armónicamente transpuesta del contenido de la parte de baja frecuencia (LF) de la señal de entrada (representada por la representación de señal de entrada), y el segundo parche es (o representa) una versión desplazada en frecuencia (no armónicamente) del contenido de la señal del primer parche. Por consiguiente, el segundo parche puede obtenerse con una complejidad computacional relativamente baja porque el copiado de los valores es más sencillo desde un punto de vista computacional que una operación de vocodificación de fase. Además, se evita que haya grandes huecos espectrales en el segundo parche, porque los valores espectrales del primer parche normalmente están suficientemente poblados (es decir comprenden valores distintos de cero), de manera que se reducen o evitan artefactos audibles, que se provocarían, en algunos casos, si el segundo parche estuviese sólo escasamente poblado.

En resumen, el concepto de la invención conlleva ventajas significativas frente a métodos de parcheado convencionales, porque la ampliación de ancho de banda armónica, usando el vocoder de fase, se aplica sólo para obtener valores de la representación de dominio espectral del primer parche, es decir para la parte más baja del espectro, mientras que se usa una ampliación de ancho de banda no armónica, que se basa en un copiado de valores de la representación de dominio espectral del primer parche para obtener valores de la representación de dominio espectral del primer parche, para frecuencias más altas. Por consiguiente, el rango más bajo (que también se designa como "primer parche") de la parte de frecuencia de ampliación (que es una parte de frecuencia por encima de la frecuencia de cruce) se proporciona como una ampliación armónica del rango de frecuencia fundamental (es decir en el rango de frecuencia de la señal de entrada, que abarca frecuencias más bajas que las frecuencias de la parte de frecuencia de ampliación, por ejemplo frecuencias por debajo de la frecuencia de cruce), lo que conlleva una buena impresión de la señal de ancho de banda ampliado cuando se escucha. Además, se ha descubierto que la sencilla generación de los valores de la representación de dominio espectral del rango más alto de la parte de frecuencia de ampliación (que también se designa como "segundo parche"), que se realiza usando el copiador, no conlleva artefactos auditivos significativos porque la audición humana no es particularmente sensible a detalles espectrales del rango más alto de la parte de frecuencia de ampliación (segundo parche).

En resumen, el concepto de la invención conlleva una buena impresión de escucha con una complejidad computacional comparativamente pequeña.

En una realización preferida, el vocoder de fase está configurado para copiar un conjunto de valores de magnitud asociados con una pluralidad de subrangos de frecuencia dados de la representación espectral de entrada, para obtener un conjunto de valores de magnitud asociados con correspondientes subrangos de frecuencia del primer parche, en el que un par de un subrango de frecuencia dado de la representación espectral de entrada y un correspondiente subrango de frecuencia del primer parche abarca (o comprende) un par de una frecuencia fundamental y un armónico de la frecuencia fundamental (por ejemplo un primer armónico de la frecuencia fundamental). El vocoder de fase también está preferiblemente configurado para multiplicar valores de fase asociados con la pluralidad de subrangos de frecuencia dados de la representación espectral de entrada por un factor predeterminado (por ejemplo 2), para obtener valores de fase asociados con correspondientes subrangos de frecuencia del primer parche. Preferiblemente, el copiador de valores está configurado para copiar un conjunto de valores asociados con una pluralidad de subrangos de frecuencia dados del primer parche, para obtener un conjunto de valores asociados con correspondientes subrangos de frecuencia del segundo parche. El copiador de valores está preferiblemente configurado para dejar valores de fase sin cambiar en el copiado. Por consiguiente, el vocoder de fase realiza, al menos aproximadamente, una transposición armónica, mientras que el copiador de valores realiza un desplazamiento de frecuencia no armónico. Los subrangos de frecuencia pueden ser, por ejemplo, rangos de frecuencia asociados con coeficientes de una transformada rápida de Fourier (o cualquier transformada comparable). Alternativamente, los subrangos de frecuencia pueden ser rangos de frecuencia asociados con señales individuales de un banco de filtros de QMF. Normalmente, un ancho de los subrangos de frecuencia es comparativamente pequeño en comparación con la frecuencia central, de manera que los subrangos de frecuencia abarcan un intervalo de frecuencia que tiene una razón de frecuencia entre una frecuencia final y una frecuencia inicial, que es significativamente menor que 2:1. En otras palabras, incluso aunque los subrangos de frecuencia de la representación espectral de entrada (que pueden, por ejemplo, adoptar la forma de coeficientes de FFT, o la forma de señales de banco de filtros de QMF) y los subrangos de frecuencia del primer parche no necesiten ser exactamente armónicos unos con respecto a otros, normalmente es posible identificar una asociación entre un subrango de frecuencia (por ejemplo, que tiene un índice de frecuencia k) de la representación espectral de entrada y

un correspondiente subrango de frecuencia (por ejemplo, que tiene un índice de frecuencia $2k$) del primer parche, de manera que el subrango de frecuencia ($2k$) del primer parche representa, al menos aproximadamente, una frecuencia armónica del correspondiente subrango de frecuencia (k) de la representación espectral de entrada.

5 Por consiguiente, una transposición armónica se realiza por el vocoder de fase, teniendo en cuenta los valores de fase, que se procesan usando un ajuste a escala de fase. En cambio, el copiador de valores simplemente realiza (al menos aproximadamente), una operación de desplazamiento de frecuencia no armónica.

En una realización preferida, el copiador de valores está configurado para copiar los valores de manera que se obtiene un desplazamiento espectral común (o desplazamiento de frecuencia) de valores del primer parche a valores del segundo parche.

10 En una realización preferida, el vocoder de fase está configurado para obtener los valores de la representación de dominio espectral del primer parche de manera que los valores de la representación de dominio espectral del primer parche representan una versión convertida en sentido ascendente de manera armónica de un rango de frecuencia fundamental de la representación de señal de entrada (por ejemplo, un rango de frecuencia fundamental por debajo de una denominada frecuencia de cruce). El copiador de valores está preferiblemente configurado para obtener los valores de la representación de dominio espectral del segundo parche representan una versión desplazada en frecuencia del primer parche. Por consiguiente se obtienen las ventajas anteriormente descritas. En particular, la implementación es sencilla al tiempo que se obtiene una buena impresión auditiva.

20 En una realización preferida, el aparato está configurado para recibir datos de audio de entrada modulados por impulsos codificados (PCM), para muestrear en sentido descendente los datos de audio de entrada modulados por impulsos codificados con el fin de obtener datos de audio modulados por impulsos codificados muestreados en sentido descendente. Además, el aparato está configurado para dividir en ventanas los datos de audio modulados por impulsos codificados muestreados en sentido descendente, con el fin de obtener datos de entrada divididos en ventanas, y para convertir o transformar los datos de entrada divididos en ventanas en un dominio de frecuencia, con el fin de obtener la representación de señal de entrada. El aparato también está preferiblemente configurado para calcular valores de magnitud a_k (también designados con α_k) y valores de fase φ_k , que representan un intervalo de frecuencia k (en el que k es un índice de intervalo de frecuencia) de la representación de señal de entrada, y para copiar los valores de magnitud valores de magnitud a_k , para obtener valores de magnitud copiados a_{sk} (también designados con α_{sk}) que representan un intervalo de frecuencia que tiene un índice de intervalo de frecuencia sk del primer parche, en el que s es un factor de alargamiento siendo $s=2$. Además, el aparato está preferiblemente configurado para copiar y ajustar a escala valores de fase φ_k asociados con un intervalo de frecuencia que tiene un índice de intervalo de frecuencia k de la representación de señal de entrada, para obtener valores de fase copiados y ajustados a escala φ_{sk} asociados con un intervalo de frecuencia que tiene un índice de frecuencia sk del primer parche. Además, el aparato está preferiblemente configurado para copiar valores $\beta_{k-i\zeta}$ asociados con un intervalo de frecuencia $k-i\zeta$ de la representación de dominio espectral del primer parche, para obtener valores β_k de la representación de dominio espectral del segundo parche. Además, el aparato está preferiblemente configurado para convertir la representación de la señal de ancho de banda ampliado (que comprende la representación de dominio espectral del primer parche y la representación de dominio espectral del segundo parche) en el dominio de tiempo, para obtener una representación de dominio de tiempo, y para aplicar una ventana de síntesis a la representación de dominio de tiempo. Usando el concepto anteriormente descrito, es posible obtener una señal de ancho de banda ampliado con una complejidad computacional moderada. La ampliación de ancho de banda se realiza en el dominio de frecuencia, pudiendo realizarse una transformada a un dominio espectral, por ejemplo, a un dominio FFT o un dominio de QMF.

45 En una realización preferida, el aparato comprende un convertidor del dominio de tiempo al dominio espectral (por ejemplo, un medio de transformada rápida de Fourier o un banco de filtros de QMF) configurado para proporcionar, como la representación de señal de entrada, valores de una representación de dominio espectral (por ejemplo, coeficientes de transformada rápida de Fourier o señales de subbanda de QMF) de una señal de audio de entrada, o de una versión previamente procesada (por ejemplo muestreada en sentido descendente y/o dividida en ventanas) de la señal de audio de entrada (por ejemplo una señal modulada por impulsos codificados proporcionada por un núcleo de decodificador de audio). El aparato preferiblemente comprende un convertidor del dominio espectral al dominio de tiempo (por ejemplo, un medio de transformada rápida de Fourier inversa o un medio de síntesis de QMF) configurado para proporcionar una representación de dominio de tiempo de la señal de ancho de banda ampliado usando valores de la representación de dominio espectral (por ejemplo coeficientes de FFT, o señales de subbanda de QMF) del primer parche y valores de la representación de dominio espectral (por ejemplo coeficientes de FFT, o señales de subbanda de QMF) del segundo parche. El convertidor del dominio espectral al dominio de tiempo está preferiblemente configurado de manera que un número de diferentes valores espectrales (por ejemplo intervalos de FFT o bandas de QMF) recibidos por el convertidor del dominio espectral al dominio de tiempo es mayor que un número de diferentes valores espectrales (por ejemplo un número de intervalos de frecuencia de FFT, o un número de bandas de QMF) proporcionado por el convertidor del dominio de tiempo al dominio espectral (por ejemplo medios de transformada rápida de Fourier o banco de filtros de QMF), de manera que el convertidor del dominio espectral al dominio de tiempo está configurado para procesar un mayor número de intervalos de frecuencia (por ejemplo intervalos de frecuencia de transformada rápida de Fourier o bandas de frecuencia de QMF) que el convertidor del dominio de tiempo al dominio de frecuencia. Por

consiguiente, una ampliación de ancho de banda se consigue por el hecho de que el convertidor del dominio espectral al dominio de tiempo comprende un mayor número de intervalos de frecuencia que el convertidor del dominio de tiempo al dominio de frecuencia.

5 En una realización preferida, el aparato comprende un divisor en ventanas de análisis configurado para dividir en ventanas una señal de audio de entrada de dominio de tiempo, para obtener una versión dividida en ventanas de la señal de audio de entrada de dominio de tiempo, que constituye la base para obtener la representación de señal de entrada. Además, el aparato comprende un divisor en ventanas de síntesis configurado para dividir en ventanas una parte de una representación de dominio de tiempo de la señal de ancho de banda ampliado, para obtener una parte dividida en ventanas de la representación de dominio de tiempo de la señal de ancho de banda ampliado. Por
10 consiguiente, se reducen o incluso se evitan artefactos en la señal de ancho de banda ampliado.

En una realización preferida, el aparato está configurado para procesar una pluralidad de partes desplazadas en el tiempo que se solapan temporalmente de la señal de audio de entrada de dominio de tiempo, para obtener una pluralidad de partes divididas en ventanas desplazadas en el tiempo que se solapan temporalmente de la representación de dominio de tiempo de la señal de ancho de banda ampliado. Un desfase de tiempo entre partes
15 desplazadas en el tiempo temporalmente adyacentes de la señal de audio de entrada de dominio de tiempo es menor que o igual a una cuarta parte de una longitud de ventana de la ventana de análisis. Se ha descubierto que un solapamiento temporal comparativamente grande entre partes desplazadas en el tiempo adyacentes de la señal de audio de entrada de dominio de tiempo (y/o un solapamiento temporal comparativamente grande entre partes desplazadas en el tiempo temporalmente adyacentes de la representación de dominio de tiempo de la señal de ancho de banda ampliado) da como resultado una ampliación de ancho de banda que conlleva una buena impresión de
20 escucha, porque no se tienen en cuenta estacionaridades de la señal debido al solapamiento temporal comparativamente grande.

En una realización preferida, el aparato comprende un proveedor de información de transitorios configurado para proporcionar una información que indica la presencia de un transitorio en la señal de entrada (representado por la representación de señal de entrada). El aparato también comprende una primera rama de procesamiento para proporcionar una representación de una parte de señal de ancho de banda ampliado basándose en una parte sin transitorio de la representación de señal de entrada y una segunda rama de procesamiento para proporcionar una representación de una parte de señal de ancho de banda ampliado basándose en una parte con transitorio de la representación de señal de entrada. La segunda rama de procesamiento está configurada para procesar una representación de dominio espectral de la señal de entrada que tiene una resolución espectral más alta que una representación de dominio espectral de la señal de entrada procesada por la primera rama de procesamiento. Por consiguiente, pueden tratarse partes de señal que comprenden un transitorio con resolución espectral más alta, lo que evita artefactos audibles en presencia de transitorios. Por otro lado, puede usarse una resolución espectral reducida para partes de señal sin transitorio (es decir, para partes de señal en las que el proveedor de información de transitorios no identifica un transitorio). Por tanto, se mantiene alta la eficiencia computacional, y la resolución espectral aumentada se usa sólo cuando conlleva ventajas (por ejemplo, porque da como resultado una mejor impresión de escucha en la proximidad de transitorios).
25
30
35

En una realización preferida, el aparato comprende un rellenador de ceros de dominio de tiempo configurado para rellenar con ceros una parte con transitorio de la señal de entrada, con el fin de obtener una parte con transitorio temporalmente ampliada de la señal de entrada. En este caso, la primera rama de procesamiento comprende un (primer) convertidor del dominio de tiempo al dominio de frecuencia configurado para proporcionar un primer número de valores de dominio espectral asociados con una parte sin transitorio de la señal de entrada, y la segunda rama de procesamiento comprende un (segundo) convertidor del dominio de tiempo al dominio de frecuencia configurado para proporcionar un segundo número de valores de dominio espectral asociados con la parte con transitorio temporalmente ampliada de la señal de entrada. El segundo número de valores de dominio espectral es mayor, al menos en un factor de 1,5, que el primer número de valores de dominio espectral. Por consiguiente, se obtiene un buen manejo de transitorios.
40
45

En una realización preferida, la segunda rama de procesamiento comprende un eliminador de ceros configurado para eliminar una pluralidad de valores cero de una parte de señal de ancho de banda ampliado obtenida basándose en la parte con transitorio temporalmente ampliada de la señal de entrada. Por consiguiente, se invierte la ampliación temporal de la señal de entrada, que se obtiene mediante el relleno con ceros.
50

En una realización preferida, el aparato comprende un muestreador en sentido descendente configurado para muestrear en sentido descendente una representación de dominio de tiempo de la señal de entrada. Mediante el muestreo en sentido descendente de la señal de entrada, puede mejorarse una eficiencia computacional si la señal de entrada no abarca el ancho de banda Nyquist completo de un flujo de entrada de muestras moduladas por impulsos codificados.
55

Otra realización según la invención crea un aparato en el que se invierte el orden de procesamiento del procesamiento por el coprador de valores y el vocoder de fase. 15. Un aparato de este tipo para generar una representación de una señal de ancho de banda ampliado basándose en una representación (110; 383) de señal de entrada comprende un coprador de valores configurado para copiar un conjunto de valores de la representación de señal de entrada, para obtener un conjunto de valores de una representación de dominio espectral de un primer parche, en el que el primer
60

5 parche está asociado con frecuencias más altas que la representación de señal de entrada. El aparato también comprende un vocoder (130; 406) de fase configurado para obtener valores ($\beta_{2\zeta} \dots \beta_{3\zeta}$) de una representación de dominio espectral de un segundo parche de la señal de ancho de banda ampliado basándose en los valores ($\beta_{4/3\zeta} \dots \beta_{2\zeta}$) de la representación de dominio espectral del primer parche, en el que el segundo parche está asociado con frecuencias más altas que el primer parche. El aparato está configurado para obtener la representación (120; 426) de la señal de ancho de banda ampliado usando los valores de la representación de dominio espectral del primer parche y los valores de la representación de dominio espectral del segundo parche.

10 Este aparato puede obtener una señal de ancho de banda ampliado con complejidad computacional comparativamente baja al tiempo que todavía se consigue una buena impresión de escucha de la señal de ancho de banda ampliado. Realizando la vocodificación de fase tras la operación de copiado, el vocoder de fase puede hacerse funcionar con una razón de frecuencia comparativamente pequeña (razón entre frecuencia de salida del vocoder y frecuencia de entrada del vocoder), lo que da como resultado un buen llenado espectral y evita la presencia de grandes huecos espectrales. Además, se ha descubierto que la impresión de escucha usando este concepto es aún mejor que para un concepto que se basa meramente en operaciones de copiado, sin una acción de vocoder de fase, a pesar de que el primer parche (parche de frecuencia más baja) se obtiene usando la operación de copiado, y sólo el segundo parche (parche de frecuencia más alta) se obtiene usando la operación de vocodificación de fase. Además, la complejidad computacional es menor que en sistemas en los que todos los parches se generan usando vocoders de fase, y los huecos espectrales se reducen en comparación con tales conceptos.

20 Naturalmente, esta realización puede complementarse mediante cualquiera de las funcionalidades comentadas en el presente documento.

Otras realizaciones según la invención crean métodos para generar una representación de una señal de ancho de banda ampliado basándose en una representación de señal de entrada. Dicho método se basa en las mismas ideas que el aparato anteriormente comentado.

Otra realización según la invención crea un programa informático para implementar el método.

25 Breve descripción de las figuras

La figura 1 muestra un diagrama de bloques esquemático de un aparato para generar una representación de una señal de ancho de banda ampliado basándose en una representación de señal de entrada, según una realización de la invención;

30 la figura 2 muestra una representación esquemática del concepto de ampliación de ancho de banda, según la presente invención;

la figura 3 muestra un diagrama de bloques esquemático detallado de un decodificador de audio que comprende un aparato para generar una representación de una señal de ancho de banda ampliado basándose en una representación de señal de entrada, según una realización de la invención;

35 la figura 4 muestra un diagrama de flujo de un método para generar una representación de una señal de ancho de banda ampliado basándose en una representación de señal de entrada, según una realización de la invención;

la figura 5 muestra un diagrama de bloques esquemático de un decodificador de audio, según un primer ejemplo de comparación; y

la figura 6 muestra un diagrama de bloques esquemático de un decodificador de audio, según un segundo ejemplo de comparación.

40 Descripción detallada de las realizaciones

1. Aparato según la figura 1

45 La figura 1 muestra un diagrama de bloques esquemático de un aparato 100 para generar una representación de una señal de ancho de banda ampliado basándose en una representación de señal de entrada. El aparato 100 está configurado para recibir una representación 110 de señal de entrada y proporcionar, basándose en la misma, una señal 120 de ancho de banda ampliado. El aparato 100 comprende un vocoder de fase configurado para obtener valores de una representación 130 de dominio espectral de un primer parche de la señal 120 de ancho de banda ampliado basándose en la representación 110 de señal de entrada. Los valores de la representación de dominio espectral del primer parche se designan, por ejemplo, con β_{ζ} a $\beta_{2\zeta}$. El aparato 100 también comprende un coprador 140 de valores configurado para copiar un conjunto de valores de la representación 132 de dominio espectral del primer parche, que se proporcionan por el vocoder 130 de fase, para obtener un conjunto de valores de una representación 142 de dominio espectral de un segundo parche, en el que el segundo parche está asociado con frecuencias más altas que el primer parche. Los valores de la representación 142 de dominio espectral del segundo parche se designan, por ejemplo, con $\beta_{2\zeta}$ a $\beta_{3\zeta}$. El aparato 100 está configurado para obtener la representación 120 de la señal de ancho de banda ampliado usando los valores β_{ζ} a $\beta_{2\zeta}$ de la representación 132 de dominio espectral del primer parche y los valores $\beta_{2\zeta}$ a $\beta_{3\zeta}$ de la

representación 142 de dominio espectral del segundo parche. Por ejemplo, la representación 120 de la señal de ancho de banda ampliado puede comprender los valores tanto de la representación 132 de dominio espectral del primer parche como de la representación 142 de dominio espectral del segundo parche. Además, la representación 120 de la señal de ancho de banda ampliado puede comprender, por ejemplo, valores de una representación de dominio espectral de la señal de entrada (representados, por ejemplo, por la representación 110 de señal de entrada). Sin embargo, la representación 120 de la señal de ancho de banda ampliado puede ser una representación de dominio de tiempo, que puede basarse en los valores de la representación 132 de dominio espectral del primer parche y los valores de la representación 142 de dominio espectral del segundo parche (y, opcionalmente, valores adicionales, por ejemplo valores de la representación 116 de dominio espectral de la señal de entrada, y/o valores de una representación de dominio espectral de parches adicionales).

A continuación se describirá en detalle la funcionalidad y el funcionamiento del aparato 100 haciendo referencia a la figura 2, que muestra una representación esquemática del concepto de la invención para generar una representación de una señal de ancho de banda ampliado basándose en una representación de señal de entrada.

Una primera representación 200 gráfica muestra una transposición armónica de la señal de entrada (representada por la representación 110 de señal de entrada), que se realiza por el vocoder 130 de fase. Como puede verse, la señal de entrada está representada, por ejemplo, mediante un conjunto de valores de magnitud α_k . El índice k designa un intervalo espectral (por ejemplo, un intervalo que tiene el índice k de una transformada rápida de Fourier, o una banda de frecuencia que tiene el índice k de una conversión de QMF). La representación 110 de señal de entrada puede comprender, por ejemplo, valores de magnitud α_k para $k = 1$ a $k = \zeta$, donde ζ puede designar un denominado intervalo de frecuencia de cruce y describe un inicio de frecuencia de la ampliación de ancho de banda. Un rango de frecuencia fundamental se describe adicionalmente, por ejemplo, mediante valores de fase φ_k , donde k es un índice de intervalo de frecuencia, tal como se comentó anteriormente.

De manera similar, el primer parche se describe mediante un conjunto de valores de una representación de dominio espectral, por ejemplo, valores β_k estando k entre ζ y 2ζ . Alternativamente, el primer parche puede estar representado por valores de magnitud α_k y valores de fase φ_k , estando el índice de intervalo de frecuencia k entre ζ y 2ζ .

Tal como se ha mencionado, el vocoder 130 de fase está configurado para realizar una transposición armónica basándose en la representación 110 de señal de entrada para obtener valores de la representación 132 de dominio espectral del primer parche. Con este fin, el vocoder 130 de fase puede ajustar un valor de magnitud α_{2k} de un intervalo de frecuencia que tiene el índice (de intervalo de frecuencia) $2k$ para que sea igual al valor de magnitud α_k de un intervalo de frecuencia que tiene el índice (de intervalo de frecuencia) k. Además, el vocoder 130 de fase puede estar configurado para ajustar el valor de fase φ_{2k} de un intervalo de frecuencia que tiene el índice $2k$ a un valor que es igual a 2 veces el valor de fase φ_k asociado con el intervalo de frecuencia que tiene el índice k. En este caso, el intervalo de frecuencia que tiene el índice k puede ser un intervalo de frecuencia de la representación 110 de señal de entrada, y el intervalo de frecuencia que tiene el índice $2k$ puede ser un intervalo de frecuencia de la representación 132 de dominio espectral del primer parche. Además, un intervalo de frecuencia que tiene el índice $2k$ puede comprender una frecuencia, que es un primer armónico de una frecuencia incluida en el intervalo de frecuencia que tiene el índice k. Por consiguiente, pueden obtenerse valores de magnitud α_{2k} y valores de fase φ_{2k} que son valores de la representación 132 de dominio espectral del primer parche, para $2k$ oscilando desde ζ a 2ζ , de manera que $\alpha_{2k} = \alpha_k$ y $\varphi_{2k} = 2\varphi_k$. Alternativamente, y de manera equivalente, los valores β_{2k} , que son valores de la representación 132 de dominio espectral del primer parche, pueden obtenerse para $2k$ entre ζ y 2ζ , de manera que $\beta_{2k} = \beta_{\alpha_k} e^{i2\varphi_k}$.

En resumen, suponiendo que los intervalos de frecuencia que tienen índices k (o de manera equivalente, $2k$, etc.), que son, por ejemplo, intervalos de frecuencia de una representación de transformada rápida de Fourier o bandas de frecuencia de una representación de dominio de QMF, están separados linealmente en frecuencia (de manera que el índice de intervalo de frecuencia, por ejemplo k o $2k$, es al menos aproximadamente proporcional a una frecuencia comprendida en el respectivo intervalo de frecuencia, por ejemplo, una frecuencia central de un k-ésimo intervalo de frecuencia de transformada rápida de Fourier o una frecuencia central de una k-ésima banda de QMF), se obtiene una transposición armónica por el vocoder 130 de fase.

Sin embargo, los valores de la representación 142 de dominio espectral del segundo parche se obtienen por el copiador 140 de valores, que realiza un copiado ascendente no armónico de valores de la representación 132 de dominio espectral del primer parche.

Haciendo referencia ahora a la representación 250 gráfica se comentará brevemente el copiado ascendente no armónico. Como puede verse, el primer parche está representado por valores β_ζ a $\beta_{2\zeta}$ (o, de manera equivalente, por valores de magnitud α_ζ a $\alpha_{2\zeta}$ y valores de fase φ_ζ a $\varphi_{2\zeta}$). Por consiguiente, los valores $\beta_{2\zeta}$ a $\beta_{3\zeta}$ (o, de manera equivalente, valores de magnitud $\alpha_{2\zeta}$ a $\alpha_{3\zeta}$ y valores de fase $\varphi_{2\zeta}$ a $\varphi_{3\zeta}$) de la representación 142 de dominio espectral del segundo parche se obtienen mediante un copiado no armónico, que se realiza por el copiador 140 de valores. Por ejemplo, valores espectrales de valor complejo $\beta_{2\zeta}$ a $\beta_{3\zeta}$ de la representación 142 de dominio espectral del segundo parche pueden obtenerse basándose en correspondientes valores β_ζ a $\beta_{2\zeta}$ de la representación 132 de dominio espectral del primer parche según $\beta_k = \beta_{k-\zeta}$ para k entre 2ζ y 3ζ . De manera equivalente, valores de magnitud $\alpha_{2\zeta}$ a $\alpha_{3\zeta}$ de la representación 142 de dominio espectral del segundo parche pueden obtenerse basándose en valores de magnitud

de la representación 132 de dominio espectral del primer parche según $\alpha_k = \alpha_{k-\zeta}$ para k entre 2_ζ y 3_ζ . En este caso, pueden obtenerse valores de fase φ_{2_ζ} a φ_{3_ζ} de la representación 142 de dominio espectral del segundo parche basándose en valores de fase φ_ζ a φ_{2_ζ} de la representación 132 de dominio espectral del primer parche según $\varphi_k = \varphi_{k-\zeta}$ para k entre 2_ζ y 3_ζ .

5 Por consiguiente, los valores de la representación 142 de dominio espectral del segundo parche representan una señal, que está desplazada en frecuencia de manera no armónica (es decir, linealmente) con respecto a una señal representada por los valores de la representación 132 de dominio espectral del primer parche.

10 Los valores β_ζ a β_{2_ζ} de la representación 132 de dominio espectral del primer parche y los valores β_{2_ζ} a β_{3_ζ} de la representación 142 de dominio espectral del segundo parche pueden usarse para obtener la representación 120 de la señal de ancho de banda ampliado. Dependiendo de los requisitos, la representación 120 de la señal de ancho de banda ampliado puede ser una representación de dominio espectral o una representación de dominio de tiempo. Si se desea obtener una representación de dominio de tiempo, puede usarse un convertidor del dominio de frecuencia al dominio de tiempo para derivar la representación de dominio de tiempo basándose en los valores β_ζ a β_{2_ζ} de la representación 132 de dominio espectral del primer parche y los valores β_{2_ζ} a β_{3_ζ} de la representación 142 de dominio espectral del segundo parche. Alternativamente (y de manera equivalente) los valores α_ζ a α_{2_ζ} , φ_ζ a φ_{2_ζ} , α_{2_ζ} a α_{3_ζ} y φ_{2_ζ} a φ_{3_ζ} pueden usarse con el fin de derivar la representación 120 de la señal de ancho de banda ampliado (ya sea en el dominio espectral o en el dominio de tiempo).

20 Tal como se comentó anteriormente, el concepto descrito con respecto a las figuras 1 y 2 conlleva una buena impresión de escucha y una complejidad computacional comparativamente baja. La vocodificación de fase sólo se requiere una vez, incluso aunque se use una pluralidad de parches (por ejemplo, el primer parche y el segundo parche). Además, se evita que haya grandes huecos espectrales en el segundo parche, lo que ocurriría si se usara otro vocoder de fase para obtener el segundo parche. Por tanto, el concepto de la invención conlleva un equilibrio muy bueno entre complejidad computacional y una impresión de escucha alcanzable.

25 Además, debe observarse que pueden obtenerse parches adicionales basándose en los valores de la representación 132 de dominio espectral del primer parche en algunas realizaciones. Por ejemplo, en una ampliación opcional del concepto de la invención, pueden obtenerse valores de una representación de dominio espectral de un tercer parche basándose en los valores de la representación 132 de dominio espectral del primer parche usando otro coprador de valores, tal como se describirá en más detalle haciendo referencia a la figura 3.

30 Las realizaciones según las figuras 1 y 2 (y también las demás realizaciones) pueden modificarse de una amplia variedad de formas. Por ejemplo, puede obtenerse un primer parche usando un vocoder de fase, y pueden obtenerse parches segundo, tercero y cuarto mediante una operación de copiado ascendente de valores espectrales. Alternativamente, un primer y un segundo parche pueden obtenerse usando vocoders de fase, y un tercer y un cuarto parche pueden obtenerse usando un copiado ascendente de valores espectrales. Naturalmente, pueden aplicarse diferentes combinaciones de la operación de vocodificación de fase y la operación de copiado ascendente.

35 Alternativamente, sin embargo, puede obtenerse un primer parche usando una operación de copiado ascendente (coprador de valores) de valores espectrales a partir de la representación de señal de entrada, y puede obtenerse un segundo parche usando un vocoder de fase (basándose en los valores copiados del primer parche, obtenidos usando el coprador de valores).

2. Realización según la figura 3

40 A continuación, se describirá un decodificador 300 de audio haciendo referencia a la figura 3, mostrando la figura 3 un diagrama de bloques esquemático detallado de un decodificador 300 de audio de este tipo que comprende un aparato para generar una representación de una señal de ancho de banda ampliado basándose en una representación de señal de entrada.

2.1. Visión general del decodificador de audio

45 El decodificador 300 de audio está configurado para recibir un flujo 310 de datos y para proporcionar, basándose en el mismo, una forma 312 de onda de audio. El decodificador 300 de audio comprende un decodificador 320 núcleo, que está configurado para proporcionar, por ejemplo, datos 322 modulados por impulsos codificados ("datos PCM") basándose en el flujo 310 de datos. El decodificador 320 núcleo puede ser, por ejemplo, un decodificador de audio tal como se describe en la norma internacional ISO/IEC 14496-3: 2005(e), parte 3: audio, subparte 4: codificación general de audio (GA)-AAC, Twin VQ, BSAC. Por ejemplo, el decodificador 320 núcleo puede ser un denominado decodificador núcleo de codificación de audio avanzada (AAC), que se describe en dicha norma, y que los expertos en la técnica conocen ampliamente. Por tanto, los datos 322 de audio modulados por impulsos codificados pueden proporcionarse por el decodificador 220 núcleo basándose en el flujo 310 de datos. Por ejemplo, los datos 322 de audio modulados por impulsos codificados pueden comprender la longitud de trama de 1024 muestras.

55 El decodificador 300 de audio también comprende una ampliación 330 de ancho de banda (o amplificador de ancho de banda), que está configurada para recibir los datos 322 de audio modulados por impulsos codificados (por ejemplo, una

longitud de trama de 1024 muestras) y para proporcionar, basándose en los mismos, la forma 312 de onda. La ampliación 330 de ancho de banda (o ampliador de ancho de banda) también recibe algunos datos 332 de control procedentes del flujo 310 de datos. La ampliación 330 de ancho de banda comprende una provisión 340 de datos de QMF parcheados (o proveedor de datos de QMF parcheados), que recibe los datos 322 de audio modulados por impulsos codificados y que proporciona, basándose en los mismos, datos 342 de QMF parcheados. La ampliación 330 de ancho de banda también comprende un formateo 344 de envolvente (o formateador de envolvente), que recibe los datos 342 de QMF parcheados y datos 346 de control de formateo de envolvente y proporciona, basándose en los mismos, datos 348 de QMF parcheados y con envolvente formateada. La ampliación 330 de ancho de banda también comprende una síntesis 350 de QMF (o sintetizador de QMF), que recibe los datos 348 de QMF parcheados y con envolvente formateada y proporciona, basándose en los mismos, la forma 312 de onda realizando una síntesis de QMF.

2.2. Provisión 340 de datos de QMF parcheados

2.2.1. Provisión 340 de datos de QMF parcheados – Visión general

La provisión 340 de datos de QMF parcheados (que puede realizarse mediante un proveedor 340 de datos de QMF parcheados en una implementación de hardware) puede conmutar entre dos modos, concretamente un primer modo, en el que se realiza un parcheado de replicación de banda espectral (SBR), y un segundo modo en el que se realiza un parcheado de ampliación de ancho de banda armónica (HBE). Por ejemplo, los datos 322 de audio modulados por impulsos codificados pueden retardarse mediante un retardador 360, para obtener datos 362 de audio modulados por impulsos codificados retardados, y los datos 362 de audio modulados por impulsos codificados retardados pueden convertirse a un dominio de QMF usando un analizador 364 de QMF de 32 bandas. El resultado del analizador 364 de QMF de 32 bandas, por ejemplo, una representación 365 de dominio de QMF de 32 bandas (es decir dominio espectral) de los datos 362 de audio modulados por impulsos codificados retardados, puede proporcionarse a un parcheador 366 de SBR y a un parcheador 368 de ampliación de ancho de banda armónica.

El parcheador 366 de replicación de banda espectral puede realizar, por ejemplo, un parcheado de replicación de banda espectral, que se describe, por ejemplo, en la sección 4.6.18 “SBR tool” de la norma internacional ISO/IEC 14496-3:2005(e), parte 3, subparte 4. Por consiguiente, puede proporcionarse una representación 370 de dominio de QMF de 64 bandas mediante el parcheador 366 de replicación de banda espectral.

Alternativamente, o además, el parcheador 368 de ampliación de ancho de banda armónica puede proporcionar una representación 372 de dominio de QMF de 64 bandas, que es una representación de ancho de banda ampliado de los datos 322 de audio PCM. Un conmutador 374, que se controla en dependencia de los datos 332 de control de ampliación de ancho de banda extraídos del flujo 310 de datos, puede usarse para decidir si se aplica el parcheado 366 de replicación de banda espectral o parcheado 368 de ampliación de ancho de banda armónica con el fin de obtener los datos 342 de QMF parcheados (que puede ser igual a la representación 370 de dominio de QMF 64 bandas o igual a la representación 372 de dominio de QMF de 64 bandas dependiendo del estado del conmutador 374).

2.2.2. Provisión de datos de QMF parcheados - Ampliación 368 de ancho de banda armónica

A continuación se describirá en más detalle el parcheado 368 de ampliación de ancho de banda (al menos parcialmente) armónica. El parcheado 368 de ampliación de ancho de banda armónica comprende una trayectoria de señal, en la que se convierten datos 322 de audio modulados por impulsos codificados, o una versión previamente procesada de los mismos, a un dominio espectral (por ejemplo, a un dominio de coeficiente de transformada rápida de Fourier o un dominio de QMF), en el que se realiza una ampliación de ancho de banda armónica en el dominio espectral, y en el que la representación de dominio espectral obtenida de la señal de ancho de banda ampliado, o una representación derivada de la misma, se usa para el parcheado de ampliación de ancho de banda armónica.

En la realización de la figura 3, los datos 322 de audio modulados por impulsos codificados se muestrean en sentido descendente en un muestreador 380 en sentido descendente, por ejemplo, en un factor de 2, para obtener datos 381 de audio modulados por impulsos codificados muestreados en sentido descendente. Los datos 381 de audio modulados por impulsos codificados muestreados en sentido descendente se dividen a continuación en ventanas mediante un divisor 382 en ventanas, que puede comprender, por ejemplo, una longitud de ventana de 512 muestras. Debe observarse que la ventana está desplazada, por ejemplo, 64 muestras de los datos 381 de audio modulados por impulsos codificados muestreados en sentido descendente en etapas de procesamiento posteriores, de manera que se obtiene un solapamiento comparativamente grande de las partes 383 divididas en ventanas de los datos de audio modulados por impulsos codificados muestreados en sentido descendente.

El decodificador 300 de audio también comprende un detector 384 de transitorios, que está configurado para detectar un transitorio dentro de los datos 322 de audio modulados por impulsos codificados. El detector 384 de transitorios puede detectar la presencia de un transitorio ya sea basándose en los propios datos 322 de audio PCM, o basándose en una información secundaria, que está incluida en el flujo 310 de datos.

Las partes 383 divididas en ventanas de los datos 381 de audio PCM muestreados en sentido descendente pueden procesarse selectivamente usando una primera rama 386 de procesamiento o una segunda rama 388 de procesamiento. La primera rama 386 puede usarse para procesar una parte 383 sin transitorio dividida en ventanas de los datos de audio PCM muestreados en sentido descendente (para los que el detector 384 de transitorio niega la

presencia de un transitorio), y puede usarse una segunda rama 388 para un procesamiento de una parte 383 con transitorio dividida en ventanas de los datos de audio PCM muestreados en sentido descendente (para los que el detector 384 de transitorios indica la presencia de un transitorio).

5 La primera rama 386 recibe una parte 383 sin transitorio dividida en ventanas y proporciona, basándose en la misma, una representación 387, 434 de ancho de banda ampliado de la parte 383 dividida en ventanas. De forma similar, la segunda rama 388 recibe una parte 383 con transitorio dividida en ventanas de los datos 381 de audio PCM muestreados en sentido descendente y proporciona, basándose en la misma, una representación 389 de ancho de banda ampliado de la parte 383 (con transitorio) dividida en ventanas. Tal como se comentó anteriormente, el detector 384 de transitorio decide si la parte 383 dividida en ventanas actual es una parte sin transitorio dividida en ventanas o
10 una parte con transitorio dividida en ventanas, de manera que el procesamiento de la parte 383 dividida en ventanas actual se realiza usando o bien la primera rama 386 o bien la segunda rama 388. Por tanto, diferentes partes 383 divididas en ventanas pueden procesarse por diferentes ramas 386, habiendo un solapamiento temporal significativo entre las subsiguientes representaciones 387, 389 de ancho de banda ampliado de las subsiguientes partes 383 divididas en ventanas (porque hay un solapamiento temporal significativo de partes 383 divididas en ventanas
15 temporalmente subsiguientes).

La ampliación 368 de ancho de banda armónica comprende además un solapador-y-sumador 390, que está configurado para solapar-y-sumar las diferentes representaciones 387, 389 de ancho de banda ampliado asociadas con diferentes partes 383 divididas en ventanas (temporalmente subsiguientes). Un incremento de solapamiento-y-suma puede ajustarse, por ejemplo, a 256 muestras. Por consiguiente, se obtiene una señal 392 solapada-y-sumada.

20 La ampliación 368 de ancho de banda armónica también comprende un analizador 394 de QMF de 64 bandas, que está configurado para recibir la señal 392 solapada-y-sumada y para proporcionar, basándose en el mismo, una señal 396 de dominio de QMF de 64 bandas. La señal 396 de dominio de QMF de 64 bandas puede representar, por ejemplo, un rango de frecuencia más amplio que la señal 365 de dominio de QMF de 32 bandas proporcionada por el analizador 364 de QMF de 32 bandas.

25 La ampliación 368 de ancho de banda armónica también comprende un combinador 398, que está configurado para recibir tanto la señal de dominio de QMF de 32 bandas proporcionada por el analizador 364 de QMF de 32 bandas como la señal 396 de dominio de QMF de 64 bandas y para combinar esas señales. Por ejemplo, las componentes de rango de baja frecuencia (o rango de frecuencia fundamental) de la señal 396 de dominio de QMF de 64 bandas pueden sustituirse por, o combinarse con, la señal 365 de dominio de QMF de 32 bandas proporcionada por el analizador 364 de QMF de 32 bandas, de manera que, por ejemplo, las 32 componentes de rango de frecuencia más baja (o rango de frecuencia fundamental) de la señal 372 de dominio de QMF de 64 bandas se determinan mediante la salida del analizador 364 de QMF de 32 bandas, y de manera que las 32 componentes de frecuencia más alta de la señal 372 de dominio de QMF de 64 bandas se determinan por las 32 componentes de rango de frecuencia más alta de la señal 396 de dominio de QMF de 64 bandas.

35 Naturalmente, el número de componentes de las señales de dominio de QMF puede variar, dependiendo de los requisitos específicos. Naturalmente, una posición de frecuencia de una transición entre un rango de frecuencia fundamental (también designado como rango de frecuencia más baja) y un rango de frecuencia de ancho de banda ampliado (también designado como rango de frecuencia más alta) puede depender de la frecuencia de cruce, o, de manera equivalente, el ancho de banda de la señal de audio representada por los datos 322 de audio modulados por impulsos codificados.
40

A continuación se describirán detalles relativos a la primera rama 386 de procesamiento. La primera rama 386 comprende un convertidor 400 del dominio de tiempo al dominio de frecuencia, que está implementado, por ejemplo, en forma de un medio de transformada rápida de Fourier configurado para proporcionar 512 coeficientes de transformada rápida de Fourier basándose en una parte 383 dividida en ventanas de 512 muestras de dominio de tiempo de los datos 381 de audio modulados por impulsos codificados muestreados en sentido descendente. Por consiguiente, los intervalos de frecuencia de transformada rápida de Fourier se designan con índices de intervalo de frecuencia enteros subsiguientes k en un rango entre 1 y $N=512$.
45

La primera rama 386 también comprende un proveedor 402 de valores de magnitud, que está configurado para proporcionar valores de magnitud α_k de los coeficientes de transformada rápida de Fourier. Además, la primera rama 386 comprende un proveedor 404 de valores de fase configurado para proporcionar valores de fase φ_k de los coeficientes de transformada rápida de Fourier.
50

La primera rama 386 también comprende un vocoder 406 de fase, que puede recibir los valores de magnitud α_k y los valores de fase φ_k como una representación de señal de entrada, y que puede comprender la funcionalidad del vocoder 130 de fase comentado anteriormente. Por consiguiente, el vocoder 406 de fase puede emitir valores β_{2k} , en un rango entre $\beta_{2\xi}$ y $\beta_{2\zeta}$, de una representación de dominio espectral de un primer parche. Los valores β_{2k} se designan con 408, y pueden ser equivalentes a los valores de la representación 132 de dominio espectral de un primer parche. La primera rama 386 también comprende un copiador 410 de valores, que puede asumir la funcionalidad del copiador 140 de valores, y que puede recibir, como información de entrada, los valores β_{2k} (por ejemplo, en un rango entre $\beta_{2\xi}$ y $\beta_{2\zeta}$). Por consiguiente, el primer copiador 410 de valores puede proporcionar valores β_k en un rango entre $\beta_{2\xi}$ y $\beta_{3\zeta}$, que se
55

designan con 412 y que pueden ser equivalentes a los valores $\beta_{2\xi}$ a $\beta_{3\xi}$ de la representación 142 de dominio espectral del segundo parche. Además, la primera rama 386 puede comprender (opcionalmente) un segundo copiador 414 de valores, que está configurado para recibir los valores β_{ξ} y $\beta_{2\xi}$ (también designados con 408) proporcionados por el vocoder 406 de fase y para proporcionar, basándose en los mismos, valores espectrales $\beta_{3\xi}$ a $\beta_{4\xi}$ usando una operación de copia (que da como resultado, efectivamente, un desplazamiento de frecuencia no armónico del espectro descrito por los valores β_{ξ} a $\beta_{2\xi}$ (408)). Por consiguiente, el segundo copiador 414 de valores proporciona valores espectrales $\beta_{3\xi}$ a $\beta_{4\xi}$ de una representación de dominio espectral de un tercer parche, que también se designan como 416.

La primera rama 386 puede comprender un interpolador 420 opcional, que puede estar configurado para recibir los valores 412, 416 de las representaciones de dominio espectral del segundo parche y del tercer parche (y, opcionalmente, también los valores 408 de la representación de dominio espectral del primer parche) y para proporcionar valores 422 interpolados de la representación de dominio espectral del segundo y tercer parche (y, opcionalmente, también del primer parche).

La primera rama 386 puede comprender adicionalmente un rellenador 424 de ceros, que está configurado para recibir los valores 422 interpolados (o, alternativamente, los valores 412, 416 originales) de las representaciones de dominio espectral del segundo y tercer parche (y, opcionalmente también del primer parche) y para obtener, basándose en las mismas, una versión rellenada con ceros de valores de una representación de dominio espectral, que está rellenada con ceros con el fin de adaptarse a una dimensión de un convertidor 428 del dominio espectral al dominio de tiempo.

El convertidor 428 del dominio espectral al dominio de tiempo puede implementarse, por ejemplo, como un transformador rápido de Fourier inverso. Por ejemplo, el transformador 428 rápido de Fourier inverso puede estar configurado para recibir un conjunto de 2048 valores espectrales (opcionalmente interpolados y rellenados con cero), y para proporcionar, basándose en los mismos, una representación 430 de dominio de tiempo de la parte de señal de ancho de banda ampliado. La primera trayectoria 386 también comprende un divisor 432 en ventanas de síntesis, que está configurado para recibir la representación 430 de dominio de tiempo de la parte de señal de ancho de banda ampliado y para aplicar una división en ventanas de síntesis, con el fin de obtener una representación de dominio de tiempo dividida en ventanas de síntesis de la parte 430 de señal de ancho de banda ampliado.

El decodificador 300 de audio también comprende una segunda trayectoria 388 de procesamiento, que realiza un procesamiento muy similar en comparación con la primera trayectoria 386. Sin embargo, la segunda trayectoria 388 comprende un rellenador 438 de ceros de dominio de tiempo, que está configurado para recibir la parte 383 con transitorio dividida en ventanas de los datos 381 de audio modulados por impulsos codificados muestreados en sentido descendente y para derivar una versión 439 rellenada con ceros a partir de la parte 383 dividida en ventanas, de manera que un inicio de la parte 439 rellenada con ceros y un final de la parte 439 rellenada con ceros estén rellenados con ceros, y de manera que el transitorio esté dispuesto en una región central (entre las muestras de inicio rellenadas con ceros y las muestras de final rellenadas con ceros) de la parte 439 rellenada con ceros.

La segunda trayectoria 388 también comprende un transformador 440 del dominio de tiempo al dominio espectral, por ejemplo, un transformador rápido de Fourier o un QMF (banco de filtros de espejo en cuadratura). El transformador 440 del dominio de tiempo al dominio espectral normalmente comprende un mayor número de intervalos de frecuencia (por ejemplo, intervalos de frecuencia de transformada rápida de Fourier, o bandas de QMF) que el transformador 400 del dominio de tiempo al dominio espectral de la primera rama. Por ejemplo, el transformador 440 rápido de Fourier puede estar configurado para derivar 1024 coeficientes de transformada rápida de Fourier a partir de la parte 439 rellenada con ceros de 1024 muestras de dominio de tiempo.

La segunda rama 388 también comprende un determinador 442 de valores de magnitud y un determinador 444 de valores de fase, que pueden comprender la misma funcionalidad que los medios 402, 404 correspondientes de la primera rama 386, aunque con una dimensión $N=1024$ aumentada. De forma similar, la segunda rama 388 también comprende un vocoder 446 de fase, un primer copiador 450 de valores, un segundo copiador 454 de valores, un interpolador 460 opcional, y un rellenador 464 de ceros opcional, que pueden comprender las mismas funcionalidades que los medios correspondientes de la primera rama 386, aunque con dimensiones aumentadas. En particular, el índice ξ de la banda de cruce puede ser superior en la segunda rama 388 que en la primera rama 386, por ejemplo, en un factor de 2.

Por consiguiente, una representación de dominio espectral que comprende, por ejemplo, 4096 coeficientes de transformada rápida de Fourier puede proporcionarse a un transformador 468 rápido de Fourier inverso, que a su vez proporciona una señal 470 de dominio de tiempo que tiene 4096 muestras.

La segunda rama 388 también comprende un divisor 472 en ventanas de síntesis, que está configurado para proporcionar una versión dividida en ventanas de la representación 470 de dominio de tiempo de la parte de señal de ancho de banda ampliado.

La segunda rama 388 también comprende un eliminador de ceros configurado para proporcionar una representación 478 de dominio de tiempo acortada, dividida en ventanas, de la parte de señal de ancho de banda ampliado, representación 478 de dominio de tiempo acortada, dividida en ventanas que puede comprender, por ejemplo, 2048 muestras.

5 Por consiguiente, la representación 387 de dominio de tiempo se usa para partes sin transitorio (por ejemplo, tramas de audio) de los datos 322 de audio modulados por impulsos codificados, y la representación 478 de dominio de tiempo se usa para partes con transitorio de los datos 322 de audio modulados por impulsos codificados. Por consiguiente, las partes con transitorio se procesan con una resolución de dominio espectral más alta en la segunda rama 388 de procesamiento, mientras que las partes sin transitorio se procesan con una resolución espectral más baja en la primera rama 386 de procesamiento.

2.3. Formateo 344 de envolvente

A continuación se resumirá brevemente el formateo 344 de envolvente. Además, se hace referencia a las respectivas observaciones realizadas en la sección de introducción, que también se aplican al concepto de la invención.

10 Los datos 342 de QMF parcheados, que se obtienen basándose en la señal 396 de dominio de QMF de 64 bandas, se procesan por el formateo 344 de envolvente, para obtener la representación 348 de señal, que se introduce en el sintetizador 350 de QMF. El formateo de envolvente puede adaptar, por ejemplo, las señales de banda de dominio de QMF de los datos 342 de QMF parcheados con el fin de realizar un llenado de ruido, con el fin de reconstruir armónicos que ausentes, y/o con el fin de obtener un filtrado inverso. Variaciones de llenado de ruido, inserción de armónicos ausentes y filtrado inverso pueden controlarse, por ejemplo, mediante una información 346 secundaria, que puede extraerse del flujo 310 de datos. Para más detalles se hace referencia, por ejemplo, a los comentarios de la herramienta de SBR en la sección 4.6.18 de la norma internacional ISC/IEC 14496-3:2005(e), parte 3, subparte 4. Sin embargo, también pueden aplicarse conceptos de formateo de envolvente diferentes según los requisitos.

3. Comentario y comparación de diferentes soluciones

20 A continuación se proporcionará un breve comentario y sumario de la solución de la invención.

Realizaciones según la presente invención, por ejemplo el aparato 100 según la figura 1 y el decodificador 300 de audio según la figura 3, son (o comprenden) nuevos algoritmos de parcheado dentro de la replicación de banda espectral (SBR). Puede usarse un parcheado de dominio espectral de diferentes maneras con el fin de tener en cuenta diferentes características de señal o restricciones dictadas por requisitos de software o hardware.

25 En SBR estándar, el parcheado siempre se lleva a cabo mediante una operación de copia dentro del dominio de QMF. Esto puede llevar a veces a artefactos auditivos, particularmente si se copian sinusoides en la proximidad unos de otros en el borde de la parte de LF y de HF generada. Por tanto, se ha introducido un nuevo algoritmo de parcheado que evita algunos problemas usando un vocoder de fase (véase, por ejemplo, la referencia [13]). Este algoritmo se ilustra en la figura 5 como un ejemplo de comparación.

30 La SBR estándar tiene el problema de los artefactos auditivos. El enfoque del vocoder de fase presentado en la referencia [13] tiene una complejidad, particularmente debido al alto número de transformadas rápidas de Fourier que tienen que calcularse. Adicionalmente, el espectro se vuelve muy disperso para parches altos (factores de alargamiento altos), lo que puede dar como resultado artefactos de audio no deseados.

35 Dos realizaciones evitan el elevado número de transformadas rápidas de Fourier moviendo la generación de diferentes parches del dominio de tiempo al dominio de frecuencia. En la figura 6 se da un ejemplo en el que la transformación al dominio de frecuencia se consigue con la ayuda de una transformada rápida de Fourier. En lugar de la transformación de Fourier pueden usarse, sin embargo, otras transformaciones de tiempo-frecuencia.

40 La figura 3 muestra una solución híbrida del algoritmo de la figura 6 para parcheado de SBR. Sólo se genera el primer parche mediante el algoritmo del vocoder de fase (por ejemplo, bloque 406 de la primera rama 386, y bloque 446 de la segunda rama 388) mientras que parches superiores (por ejemplo, el segundo parche y el tercer parche) se crean simplemente copiando el primer parche (por ejemplo, usando los copiadores 410, 414 de valores de la primera rama 386, y/o los copiadores 450, 454 de valores de la segunda rama 388). Esto da lugar a un espectro menos disperso.

45 A continuación se explicarán brevemente el algoritmo de comparación, que se implementa en el decodificador de audio mostrado en la figura 6, y el algoritmo de la invención, que se implementa en el decodificador de audio mostrado en la figura 3:

El algoritmo de comparación o de referencia, que se implementa en el decodificador de audio mostrado en la figura 6, comprende las siguientes etapas:

1. Muestreo en sentido descendente de la señal (si no se daña el criterio Nyquist)

50 2. La señal se divide en ventanas (se proponen ventanas "Hann" aunque pueden usarse otras formas de ventana) y se toman denominados granos (por ejemplo, partes 383 de señal divididas en ventanas) de longitudes N a partir de la señal. Las ventanas se desplazan sobre la señal con un tamaño de salto H. Se propone un solapamiento de $N/H=8$ veces.

3. Si el grano (por ejemplo, una parte 383 de señal dividida en ventanas) contiene un evento transitorio en los bordes, se rellena (por ejemplo, mediante el rellenedor 438 de ceros) con ceros lo que lleva a un sobremuestreo en dominio de frecuencia.
- 5 4. Los granos se transforman al dominio de frecuencia (por ejemplo, usando los transformadores 400, 440 del dominio de tiempo al dominio espectral).
5. Los granos de dominio de frecuencia se rellenan (opcionalmente) hasta una longitud de salida deseada del algoritmo de parcheado.
6. Se calculan magnitud y fase (por ejemplo, usando los medios 402, 404, 442, 444).
- 10 7. El contenido n de intervalo de frecuencia se copia a una posición sn para el factor de alargamiento s . La fase se multiplica por el factor de alargamiento s . Esto se hace para todos los factores de alargamiento s (sólo para las regiones en el espectro que abarcan los parches deseados). (a) $\zeta \cdot (s-1)/s \leq n \leq \zeta$ o (b) $\zeta/s \leq n \leq \zeta$; (b) da lugar a un espectro más denso que (a) ya que los parches se solapan. ζ indica la frecuencia más alta de la parte de LF, la denominada frecuencia de cruce. En términos generales, la fase se corrige para una nueva posición de muestra (por ejemplo, posición de frecuencia), lo que puede conseguirse usando el algoritmo comentado en el presente documento o cualquier algoritmo alternativo apropiado.
- 15 8. Los intervalos de dominio de frecuencia que no consiguen datos mediante el copiado pueden llenarse aplicando una función de interpolación (por ejemplo, usando los interpoladores 420, 460).
9. Los granos se transforman de vuelta al dominio de tiempo (por ejemplo, usando los transformadores 428, 468 rápidos de Fourier inversos).
- 20 10. Los granos de dominio de tiempo se multiplican por una ventana de síntesis (de nuevo se proponen ventanas Hann) (por ejemplo, usando los divisores 432, 472 en ventanas de síntesis).
11. Si se llevó a cabo un relleno con ceros en la etapa 3, los ceros se eliminan de nuevo (por ejemplo, usando el eliminador 476 de ceros).
- 25 12. Se crea una señal o trama de ancho de banda ampliado (por ejemplo, señal 392), respectivamente, usando solapamiento y suma (OLA) (por ejemplo, usando el solapamiento-y-suma 390).
- Sin embargo, el orden de las etapas individuales también puede intercambiarse en algunas realizaciones alternativas, y algunas de las etapas pueden fusionarse en una única etapa en algunas realizaciones alternativas.
- El algoritmo de la invención, que se implementa en el decodificador de audio mostrado en la figura 3, comprende las siguientes etapas:
- 30 1. Muestreo descendente de la señal (si no se daña el criterio Nyquist)
2. La señal se divide en ventanas (se proponen ventanas "Hann" aunque pueden usarse otras formas de ventana) y se toman denominados granos (por ejemplo, partes 383 de señal divididas en ventanas) de longitudes N de la señal. Las ventanas se desplazan sobre la señal con un tamaño de salto H . Se propone un solapamiento de $N/H=8$ veces.
- 35 3. Si el grano (por ejemplo, una parte 383 de señal dividida en ventanas) contiene un evento transitorio en los bordes, se rellena (por ejemplo, por el rellenedor 438 de ceros) con ceros lo que lleva a un sobremuestreo en el dominio de frecuencia.
4. Los granos se transforman al dominio de frecuencia (por ejemplo, usando los transformadores 400, 440 del dominio de tiempo al dominio espectral).
- 40 5. Los granos de dominio de frecuencia se rellenan (opcionalmente) hasta una longitud de salida deseada del algoritmo de parcheado.
6. Se calculan magnitud y fase (por ejemplo, usando los medios 402, 404, 442, 444).
7. a) El contenido n de intervalo de frecuencia se copia a la posición $2n$. La fase se multiplica por 2.
(a) $\zeta \cdot (s-1)/s \leq n \leq \zeta$ o (b) $\zeta/s \leq n \leq \zeta$ (véase más arriba).
- 45 7. b) El contenido $2n$ de intervalo de frecuencia se copia a la posición sn para todos los factores de alargamiento $s > 2$ en los rangos $1 \leq n \leq \zeta$.
8. Los intervalos de dominio de frecuencia que no consiguen datos mediante el copiado pueden llenarse aplicando una función de interpolación (por ejemplo, usando los interpoladores 420, 460).

9. Los granos se transforman de vuelta al dominio de tiempo (por ejemplo, usando los transformadores 428, 468 rápidos de Fourier inversos).

10. Los granos de dominio de tiempo se multiplican por una ventana de síntesis (de nuevo se proponen ventanas Hann) (por ejemplo usando los divisores 432, 472 en ventanas de síntesis).

5 11. Si se llevó a cabo un relleno con ceros en la etapa 3, los ceros se eliminan de nuevo (por ejemplo, usando el eliminador 476 de ceros).

12. La señal o trama de ancho de banda ampliado (por ejemplo, señal 392), respectivamente, se crea usando solapamiento y suma (OLA) (por ejemplo, usando el solapamiento-y-suma 390).

10 Sin embargo, el orden de las etapas individuales también puede intercambiarse en algunas realizaciones alternativas, y algunas de las etapas pueden fusionarse en una única etapa en algunas realizaciones alternativas.

Por tanto, todas las etapas son idénticas en el algoritmo de referencia (que se implementa en el decodificador de audio mostrado en la figura 6) y el algoritmo de la invención (que se implementa en el decodificador de audio mostrado en la figura 3), excepto por la etapa 7, que se ha sustituido por las siguientes etapas:

7. a) El contenido n de intervalo de frecuencia se copia a la posición $2n$. La fase se multiplica por 2.

15 (a) $\zeta \cdot (s-1)/s \leq n \leq \zeta$ o (b) $\zeta/s \leq n \leq \zeta$ (véase más arriba).

7. b) El contenido $2n$ de intervalo de frecuencia se copia a la posición sn para todos los factores de alargamiento $s > 2$ en los rangos $1 \leq n \leq \zeta$.

En resumen, las realizaciones según las figuras 1, 2, 3 y 4 (y también el decodificador de audio mostrado en la figura 6) reducen en primer lugar la complejidad drásticamente en comparación con las soluciones convencionales mencionadas.

20 En segundo lugar, permiten diferentes modificaciones de espectro diferentes respecto o bien el plano SBR o bien según se presenta en la figura 5 (véase, por ejemplo, la referencia [13]).

Por ejemplo, las señales de voz podrían beneficiarse del algoritmo, que se realiza en el aparato, decodificador de audio y método según las figuras 1, 2, 3 y 4, ya que la estructura de tren de impulsos, que es típica de las señales de voz, se mantiene mejor que con el enfoque presentado en la referencia [13].

25 Las aplicaciones más destacadas de realizaciones según la invención son decodificadores de audio, que suelen implementarse en dispositivos portátiles y por tanto funcionan con un suministro eléctrico mediante batería.

4. Método según la figura 4.

30 A continuación se describirá un método 400 para generar una representación de una señal de ancho de banda ampliado basándose en una representación de señal de entrada haciendo referencia a la figura 4, que muestra un diagrama de flujo de tal método. El método 400 comprende una etapa 410 de obtener valores de una representación de dominio espectral de un primer parche de la señal de ancho de banda ampliado basándose en la representación de señal de entrada usando una vocodificación de fase. El método 400 también comprende una etapa 420 de copiar un conjunto de valores de la representación de dominio espectral del primer parche, valores que se obtienen usando la vocodificación de fase, para obtener un conjunto de valores de una representación de dominio espectral de un segundo parche, en el que el segundo parche está asociado con frecuencias más altas que el primer parche. El método 400 también comprende una etapa 430 de obtener una representación de la señal de ancho de banda ampliado usando los valores de la representación de dominio espectral del primer parche y los valores de la representación de dominio espectral del segundo parche.

40 El método 400 puede complementarse mediante cualquiera de los medios y funcionalidades comentados en el presente documento con respecto al aparato de la invención.

5. Alternativas de implementación

45 Aunque algunos aspectos se han descrito en el contexto de un aparato, está claro que estos aspectos también representan una descripción del correspondiente método, correspondiendo un bloque o dispositivo a una etapa de método o una característica de una etapa de método. De manera análoga, aspectos descritos en el contexto de una etapa de método también representan una descripción de un correspondiente bloque o elemento o característica de un correspondiente aparato. Algunas o todas las etapas de método pueden ejecutarse mediante (o usando) un aparato de hardware, como por ejemplo, un microprocesador, un ordenador programable o un circuito electrónico. En algunas realizaciones, alguna o más de las etapas de método más importantes pueden ejecutarse mediante tal aparato.

50 Dependiendo de determinados requisitos de implementación, las realizaciones de la invención pueden implementarse en hardware o en software. La implementación puede realizarse usando un medio de almacenamiento digital, por ejemplo, un disco flexible, un DVD, un Blue-Ray, un CD, una ROM, una PROM, una EPROM, una EEPROM o una memoria FLASH, con señales de control legibles electrónicamente almacenadas en el mismo, que actúan

conjuntamente (o pueden actuar conjuntamente) con un sistema informático programable de manera que se lleve a cabo el respectivo método. Por tanto, el medio de almacenamiento digital puede ser legible por ordenador.

5 Algunas realizaciones según la invención comprenden un soporte de datos que tiene señales de control legibles electrónicamente, que pueden actuar conjuntamente con un sistema informático programable, de manera que se lleve a cabo uno de los métodos descritos en el presente documento.

Generalmente, realizaciones de la presente invención pueden implementarse como un producto de programa informático con un código de programa, siendo el código de programa operativo para realizar uno de los métodos cuando el producto de programa informático se ejecuta en un ordenador. El código de programa puede almacenarse, por ejemplo, en un soporte legible por máquina.

10 Otras realizaciones comprenden el programa informático para llevar a cabo uno de los métodos descritos en el presente documento, almacenado en un soporte legible por máquina.

En otras palabras, una realización del método de la invención es, por tanto, un programa informático que tiene un código de programa para llevar a cabo uno de los métodos descritos en el presente documento, cuando el programa informático se ejecuta en un ordenador.

15 Una realización adicional de los métodos de la invención es, por tanto, un soporte de datos (o un medio de almacenamiento digital, o un medio legible por ordenador) que comprende, registrado en el mismo, el programa informático para llevar a cabo uno de los métodos descritos en el presente documento.

20 Una realización adicional del método de la invención es, por tanto, un flujo de datos o una secuencia de señales que representan el programa informático para llevar a cabo uno de los métodos descritos en el presente documento. El flujo de datos o la secuencia de señales pueden configurarse por ejemplo para transferirse a través de una conexión de comunicación de datos, por ejemplo, a través de Internet.

Una realización adicional comprende un medio de procesamiento, por ejemplo un ordenador, o un dispositivo lógico programable, configurado o adaptado para llevar a cabo uno de los métodos descritos en el presente documento.

25 Una realización adicional comprende un ordenador que tiene instalado en el mismo el programa informático para llevar a cabo uno de los métodos descritos en el presente documento.

En algunas realizaciones, un dispositivo lógico programable (por ejemplo una disposición de puertas programables en campo) puede usarse para llevar a cabo algunas o todas las funcionalidades de los métodos descritos en el presente documento. En algunas realizaciones, una disposición de puertas programables en campo puede actuar conjuntamente con un microprocesador con el fin de llevar a cabo uno de los métodos descritos en el presente documento. Generalmente, los métodos se llevan a cabo preferiblemente mediante cualquier aparato de hardware.

30 Las realizaciones descritas anteriormente son meramente ilustrativas de los principios de la presente invención. Se entiende que modificaciones y variaciones de las disposiciones y los detalles descritos en el presente documento resultarán evidentes para los expertos en la técnica. Se pretende, por tanto, limitarse únicamente al alcance de las siguientes reivindicaciones de patente y no a los detalles específicos presentados a modo de descripción y explicación de las realizaciones en el presente documento.

6. Ejemplo de comparación según la figura 5

40 A continuación se comentará brevemente un ejemplo de comparación haciendo referencia a la figura 5. La funcionalidad del ejemplo de comparación según la figura 5 es similar a la función del decodificador de audio según la figura 3, de manera que los medios y funcionalidades no se explicarán de nuevo. Sin embargo, el ejemplo de comparación según la figura 5 se basa en el uso de tres vocoders 590, 592, 594, ó 596, 597, 598 de fase por cada rama. Transformadores rápidos de Fourier inversos, divisores en ventanas de síntesis, solapadores-y-sumadores, individuales, están asociados a los vocoders de fase individuales, como puede verse en la figura 5. Además, en algunas de las subramas, se usa muestreo en sentido descendente individual (\downarrow factor) y retardo individual ($z^{-\text{muestras}}$). Por consiguiente, el aparato 500 según la figura 5 no es tan eficiente desde el punto de vista computacional como el aparato 300 según la figura 3. No obstante, el aparato 500 conlleva significativas mejoras frente a algunos decodificadores de audio convencionales.

7. Ejemplo de comparación según la figura 6

50 La figura 6 muestra otro decodificador 600 de audio, según un ejemplo de comparación. El decodificador 600 de audio según la figura 6 es similar a los decodificadores 300, 500 de audio según las figuras 3 y 5. Sin embargo, el decodificador 600 de audio también se basa en el uso de una pluralidad de vocoders 690, 692, 694 ó 696, 697, 698 de fase individuales por cada rama, lo que hace que el aparato 600 sea más exigente desde el punto de vista computacional que el aparato 300, y lo que conlleva artefactos audibles en algunos casos. No obstante, el aparato 500 conlleva significativas mejoras frente a algunos decodificadores de audio convencionales.

8. Conclusión

En vista de los comentarios anteriores, puede verse que el aparato 100 según la figura 1, el decodificador 300 de audio según la figura 3 y el método 400 según la figura 4 conllevan diversas ventajas frente a los ejemplos de comparación, que se han comentado brevemente con referencia a las figuras 5 y 6.

- 5 El concepto de la invención es aplicable en una amplia variedad de aplicaciones y puede modificarse de un gran número de formas. En particular, los transformadores rápidos de Fourier pueden sustituirse por bancos de filtros de QMF, y los transformadores rápidos de Fourier inversos pueden sustituirse por sintetizadores de QMF.

- 10 Además, en algunas realizaciones algunas o todas las etapas de procesamiento pueden resumirse en una única etapa. Por ejemplo, una secuencia de procesamiento que comprende una síntesis de QMF y un posterior análisis de QMF puede simplificarse omitiendo las transformadas repetidas.

Bibliografía:

- [1] M. Dietz, L. Liljeryd, K. Kjörling y O. Kunz, "Spectral Band Replication, a novel approach in audio coding", en 112th AES Convention, Múnich, mayo de 2002.
- 15 [2] S. Meltzer, R. Böhm y F. Henn, "SBR enhanced audio codecs for digital broadcasting such as "Digital Radio Mondiale" (DRM)", en 112th AES Convention, Múnich, mayo de 2002.
- [3] T. Ziegler, A. Ehret, P. Ekstrand y M. Lutzky, "Enhancing mp3 with SBR: Features and Capabilities of the new mp3PRO Algorithm", en 112th AES Convention, Múnich, mayo de 2002.
- [4] International Standard ISO/IEC 14496-3: 2001/FPDAM 1, "Bandwidth Extension", ISO/IEC, 2002. Speech bandwidth extension method and apparatus Vasu lyengar et al.
- 20 [5] E. Larsen, R. M. Aarts y M. Danessis. Efficient high-frequency bandwidth extension of music and speech. En AES 112th Convention, Múnich, Alemania, mayo de 2002.
- [6] R. M. Aarts, E. Larsen y O. Ouweltjes. A unified approach to low- and high frequency bandwidth extension. En AES 115th Convention, Nueva York, EE.UU., octubre de 2003.
- 25 [7] K. Käyhkö. A Robust Wideband Enhancement for Narrowband Speech Signal. Research Report, Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, 2001.
- [8] E. Larsen y R. M. Aarts. Audio Bandwidth Extension - Application to psychoacoustics, Signal Processing and Loudspeaker Design. John Wiley & Sons, Ltd, 2004.
- [9] E. Larsen, R. M. Aarts y M. Danessis. Efficient high-frequency bandwidth extension of music and speech. En AES 112th Convention, Múnich, Alemania, mayo de 2002.
- 30 [10] J. Makhoul. Spectral Analysis of Speech by Linear Prediction. IEEE Transactions on Audio and Electroacoustics, AU-21 (3), junio de 1973.
- [11] Solicitud de patente estadounidense 08/951.029, Ohmori, et al. Audio band width extending system and method.
- [12] Patente estadounidense 6895375, Malah, D & Cox, R. V.: System for bandwidth extension of Narrowband speech.
- 35 [13] Frederik Nagel, Sascha Disch, "A harmonic bandwidth extension method for audio codecs", ICASSP International Conference on Acoustics, Speech and Signal Processing, IEEE CNF, Taipei, Taiwán, abril de 2009.
- [14] Solicitud de patente internacional WO 98/57436.
- [15] Solicitud de patente estadounidense US 20071238415.

REIVINDICACIONES

1. Aparato (100; 386) para generar una representación (120; 426) de una señal de ancho de banda ampliado basándose en una representación (110; 383) de señal de entrada, comprendiendo el aparato:
- 5 un vocoder (130; 406) de fase configurado para obtener valores ($\beta_{\zeta} \dots \beta_{2\zeta}$, 408) de una representación de dominio espectral de un primer parche de la señal de ancho de banda ampliado basándose en la representación de señal de entrada; y
- un copiador (140; 410, 416) de valores configurado para copiar un conjunto de valores ($\beta_{\zeta} \dots \beta_{2\zeta}$, 408) de la representación de dominio espectral del primer parche, valores que se proporcionan por el vocoder de fase, para obtener un conjunto de valores ($\beta_{\zeta} \dots \beta_{3\zeta}$, 408) de una representación de dominio espectral de un segundo parche, en el
- 10 que el segundo parche está asociado con frecuencias más altas que el primer parche;
- en el que el aparato está configurado para obtener la representación (120; 426) de la señal de ancho de banda ampliado usando los valores de la representación de dominio espectral del primer parche y los valores de la representación de dominio espectral del segundo parche.
2. Aparato (100; 386) según la reivindicación 1, en el que el vocoder (130; 406) de fase está configurado para
- 15 copiar un conjunto de valores de magnitud ($\alpha_{\zeta/2} \dots \alpha_{\zeta}$) asociados con una pluralidad de subrangos de frecuencia dados de la representación (110; 383) de señal de entrada, para obtener un conjunto de valores de magnitud ($\alpha_{\zeta} \dots \alpha_{2\zeta}$) asociados con correspondientes subrangos de frecuencia del primer parche,
- en el que un par de un subrango de frecuencia dado de la representación de señal de entrada y de un correspondiente subrango de frecuencia del primer parche abarcan un par de una frecuencia fundamental y un armónico de la frecuencia
- 20 fundamental,
- en el que el vocoder (130; 406) de fase está configurado para multiplicar valores de fase ($\varphi_{\zeta/2} \dots \varphi_{\zeta}$) asociados con la pluralidad de subrangos de frecuencia dados de la representación de señal de entrada por un factor predeterminado, para obtener un conjunto de valores de fase ($\varphi_{\zeta} \dots \varphi_{2\zeta}$) asociado con los correspondientes subrangos de frecuencia del primer parche, y
- 25 en el que el copiador (140; 410) de valores está configurado para copiar un conjunto de valores ($\beta_{\zeta} \dots \beta_{2\zeta}$) asociados con una pluralidad de subrangos de frecuencia dados del primer parche, para obtener un conjunto de valores ($\beta_{2\zeta} \dots \beta_{3\zeta}$) asociados con correspondientes subrangos de frecuencia del segundo parche, en el que el copiador de valores está configurado para dejar valores de fase sin cambiar en el copiado.
3. Aparato (100; 386) según la reivindicación 2, en el que el copiador (140; 410) de valores está configurado para
- 30 copiar los valores de manera que se obtiene un desplazamiento espectral común entre valores ($\beta_{\zeta} \dots \beta_{2\zeta}$) del primer parche y correspondientes valores ($\beta_{2\zeta} \dots \beta_{3\zeta}$) del segundo parche.
4. Aparato (100; 386) según una de las reivindicaciones 1 a 3, en el que el vocoder (130; 410) de fase está configurado para obtener los valores ($\beta_{\zeta} \dots \beta_{2\zeta}$) de la representación (132; 408) de dominio espectral del primer parche de manera que los valores de la representación de dominio espectral del primer parche representan una versión
- 35 convertida en sentido ascendente de manera armónica de un rango de frecuencia fundamental de la representación (110; 383) de señal de entrada; y
- en el que el copiador (140; 410) de valores está configurado para obtener los valores ($\beta_{2\zeta} \dots \beta_{3\zeta}$) de la representación (142; 412) de dominio espectral del segundo parche de manera que los valores de la representación de dominio espectral del segundo parche representan una versión desplazada en frecuencia del contenido de audio del primer
- 40 parche.
5. Aparato (100; 380, 382, 386) según una de las reivindicaciones 1 a 4, en el que el aparato está configurado para recibir datos (322) de audio de entrada,
- para muestrear en sentido descendente (380) los datos (322) de audio de entrada, con el fin de obtener datos (381) de audio muestreados en sentido descendente,
- 45 para dividir en ventanas (382) los datos (381) de audio muestreados en sentido descendente, con el fin de obtener datos (383) de entrada divididos en ventanas,
- para convertir (400) o transformar los datos (383) de entrada divididos en ventanas en un dominio espectral, con el fin de obtener la representación (383) de señal de entrada en forma de una representación (401) de dominio espectral,
- 50 para calcular (402, 404) valores de magnitud α_k y valores de fase φ_k que representan un intervalo de frecuencia que tiene un índice k de la representación (383) de señal de entrada,

para usar (130; 406) una pluralidad de valores de magnitud φ_k que representan intervalos de frecuencia que tienen índices de intervalo de frecuencia k de la representación (383) de señal de entrada, para obtener valores de magnitud α_{2k} que representan intervalos de frecuencia que tienen índices de intervalo de frecuencia sk del primer parche, cuando s es un factor de alargamiento estando s entre 1,5 y 2,5, y

5 para copiar y ajustar a escala (130; 406) valores de fase φ_k asociados a intervalos de frecuencia que tienen índices de intervalo de frecuencia k de la representación (383) de señal de entrada, para obtener valores de fase copiados y ajustados a escala $\varphi_{2k}=S\varphi_k$ asociados con intervalos de frecuencia que tienen índices de intervalo de frecuencia $2k$ del primer parche,

10 para copiar (140; 410) valores $\beta_{k-i\beta}$ asociados con intervalos de frecuencia que tienen índices de intervalo de frecuencia $k-i\zeta$ de la representación (132; 408) de dominio espectral del primer parche, para obtener valores β_k de la representación (142; 412) de dominio espectral del segundo parche,

para convertir (428) la representación (426) de la señal de ancho de banda ampliado en el dominio de tiempo, para obtener una representación (430) de dominio de tiempo, y

para dividir (432) en ventanas de síntesis la representación de dominio de tiempo.

15 6. Aparato (100; 386) según una de las reivindicaciones 1 a 5, en el que el aparato comprende un convertidor (400) del dominio de tiempo al dominio espectral configurado para proporcionar, como la representación (401) de señal de entrada, valores de una representación de dominio espectral de una señal (322) de audio de entrada, o de una versión (383) previamente procesada de la señal (322) de audio de entrada; y

20 en el que el aparato comprende un convertidor (428) del dominio espectral al dominio de tiempo configurado para proporcionar una representación (430) de dominio de tiempo de la señal de ancho de banda ampliado usando valores (β_{ζ} , $\beta_{2\zeta}$, 408) de la representación de dominio espectral del primer parche y valores $\beta_{2\zeta} \dots \beta_{3\zeta}$, 412) de la representación de dominio espectral del segundo parche;

25 en el que el convertidor (428) del dominio espectral al dominio de tiempo está configurado de manera que un número ($N=2048$) de diferentes valores (426) espectrales recibidos por el convertidor (428) del dominio espectral al dominio de tiempo es mayor que un número ($N=512$) de diferentes valores (401) espectrales proporcionados por el convertidor (400) del dominio de tiempo al dominio espectral, de manera que el convertidor (428) del dominio espectral al dominio de tiempo está configurado para procesar un mayor número de intervalos de frecuencia que el convertidor (400) del dominio de tiempo al dominio espectral.

30 7. Aparato (100; 382, 386) según una de las reivindicaciones 1 a 6, en el que el aparato comprende un divisor (382) en ventanas de análisis configurado para dividir en ventanas una señal (322) de audio de entrada de dominio de tiempo, para obtener una versión (383) dividida en ventanas de la señal de audio de entrada de dominio de tiempo, que constituye la base para obtener la representación de señal de entrada en forma de una representación (401) de dominio espectral; y

35 en el que el aparato comprende un divisor (432) en ventanas de síntesis configurado para dividir en ventanas una parte de una representación (430) de dominio de tiempo de la señal de ancho de banda ampliado, para obtener una parte (434) dividida en ventanas de la representación de dominio de tiempo de la señal de ancho de banda ampliado.

40 8. Aparato (100; 382, 386) según la reivindicación 7, en el que el aparato está configurado para procesar una pluralidad de partes desplazadas en el tiempo que se solapan temporalmente de la señal (322) de audio de entrada de dominio de tiempo, para obtener una pluralidad de partes (434) divididas en ventanas desplazadas en el tiempo que se solapan temporalmente de la representación de dominio de tiempo de la señal de ancho de banda ampliado,

en el que un desfase de tiempo ($\text{Inc}=64$) entre partes desplazadas en el tiempo temporalmente adyacentes de la señal (322) de audio de entrada de dominio de tiempo es menor que o igual a una cuarta parte de una longitud (512) de ventana del divisor (382) en ventanas de análisis.

45 9. Aparato (100; 382,386) según una de las reivindicaciones 1 a 8, en el que el aparato comprende un proveedor (384) de información de transitorios configurado para proporcionar una información que indica la presencia de un transitorio en la señal (322) de entrada; y

50 en el que el aparato comprende una primera rama (386) de procesamiento para proporcionar una representación (434) de una parte de señal de ancho de banda ampliado basándose en una parte sin transitorio de la representación (383) de señal de entrada y una segunda rama (388) de procesamiento para proporcionar una representación (478) de una parte de señal de ancho de banda ampliado basándose en una parte con transitorio de la representación (383) de señal de entrada;

en el que la segunda rama (388) de procesamiento está configurada para procesar una representación (441) de dominio espectral de la señal de entrada con una resolución espectral más alta ($N=1024$) que una representación (401) de dominio espectral de la señal de entrada procesada por la primera rama (386) de procesamiento.

10. Aparato (100; 382, 386) según la reivindicación 9, en el que la segunda rama (388) de procesamiento comprende un rellenedor (438) de ceros de dominio de tiempo configurado para rellenar con ceros una parte (383) que contiene transitorios de la señal de entrada, con el fin de obtener una parte (439) que contiene transitorios temporalmente ampliada de la señal de entrada; y
- 5 en el que la primera rama (386) de procesamiento comprende un convertidor (400) del dominio de tiempo al dominio de frecuencia configurado para proporcionar un primer número (N=512) de valores (401) de dominio espectral asociados con la parte (383) sin transitorio de la señal de entrada; y
- en el que la segunda rama (388) de procesamiento comprende un convertidor (440) del dominio de tiempo al dominio de frecuencia configurado para proporcionar un segundo número (N=1024) de valores (441) de dominio espectral asociados con la parte (439) que contiene transitorios temporalmente ampliada de la señal de entrada,
- 10 en el que el segundo número (N=1024) de valores de dominio espectral es mayor, al menos en un factor de 1,5, que el primer número(N=512) de valores de dominio espectral.
11. Aparato (100; 382, 386) según la reivindicación 10, en el que la segunda rama de procesamiento comprende un eliminador (476) de ceros configurado para eliminar una pluralidad de valores cero de una parte (474) de señal de ancho de banda ampliado obtenida basándose en la parte (439) que contiene transitorios temporalmente ampliada de la señal de entrada.
- 15 12. Aparato (100; 380, 382, 386) según una de las reivindicaciones 1 a 11, en el que el aparato comprende un muestreador (380) en sentido descendente configurado para muestrear en sentido descendente una representación (322) de dominio de tiempo de la señal de entrada.
- 20 13. Decodificador (300) de audio que comprende un aparato (100; 386) según una de las reivindicaciones 1 a 12.
14. Método (400) para generar una representación de una señal de ancho de banda ampliado basándose en una representación de señal de entrada, comprendiendo el método:
- obtener (410), usando una vocodificación de fase, valores de una representación de dominio espectral de un primer parche de la señal de ancho de banda ampliado basándose en la representación de señal de entrada; y
- 25 copiar (420) un conjunto de valores de la representación de dominio espectral del primer parche, valores que se proporcionan por la vocodificación de fase, para obtener un conjunto de valores de una representación de dominio espectral de un segundo parche, en el que el segundo parche está asociado con frecuencias más altas que el primer parche; y
- 30 obtener (430) la representación de la señal de ancho de banda ampliado usando los valores de la representación de dominio espectral del primer parche y los valores de la representación de dominio espectral del segundo parche.
15. Aparato (100; 386) para generar una representación (120; 426) de una señal de ancho de banda ampliado basándose en una representación (110; 383) de señal de entrada, comprendiendo el aparato:
- un coprador de valores configurado para copiar un conjunto de valores ($\beta_1 \dots \beta_c$) de la representación de señal de entrada, para obtener un conjunto de valores ($\beta_c \dots \beta_{2c}$) de una representación de dominio espectral de un primer parche, en el que el primer parche está asociado con frecuencias más altas que la representación de señal de entrada;
- 35 y
- un vocoder (130; 406) de fase configurado para obtener valores ($\beta_{2c} \dots \beta_{3c}$) de una representación de dominio espectral de un segundo parche de la señal de ancho de banda ampliado basándose en los valores ($\beta_{4/3c} \dots \beta_{2c}$) de la representación de dominio espectral del primer parche, en el que el segundo parche está asociado con frecuencias más altas que el primer parche; y
- 40 en el que el aparato está configurado para obtener la representación (120;426) de la señal de ancho de banda ampliado usando los valores de la representación de dominio espectral del primer parche y los valores de la representación de dominio espectral del segundo parche.
- 45 16. Método (400) para generar una representación de una señal de ancho de banda ampliado basándose en una representación de señal de entrada, comprendiendo el método:
- copiar valores de la representación de señal de entrada, para obtener valores de una representación de dominio espectral de un primer parche de la señal de ancho de banda ampliado basándose en la representación de señal de entrada, en el que el primer parche está asociado con frecuencias más altas que la representación de señal de entrada;
- y
- 50 obtener, usando una vocodificación de fase, un conjunto de valores de la representación de dominio espectral del segundo parche basándose en un conjunto de valores de la representación de dominio espectral del primer parche,

valores de la representación de dominio espectral del primer parche que se obtienen mediante el copiado, en el que el segundo parche está asociado con frecuencias más altas que el primer parche; y

obtener (430) la representación de la señal de ancho de banda ampliado usando los valores de la representación de dominio espectral del primer parche y los valores de la representación de dominio espectral del segundo parche.

- 5 17. Programa informático para realizar el método según la reivindicación 14 o la reivindicación 16, cuando el programa informático se ejecuta en un ordenador.

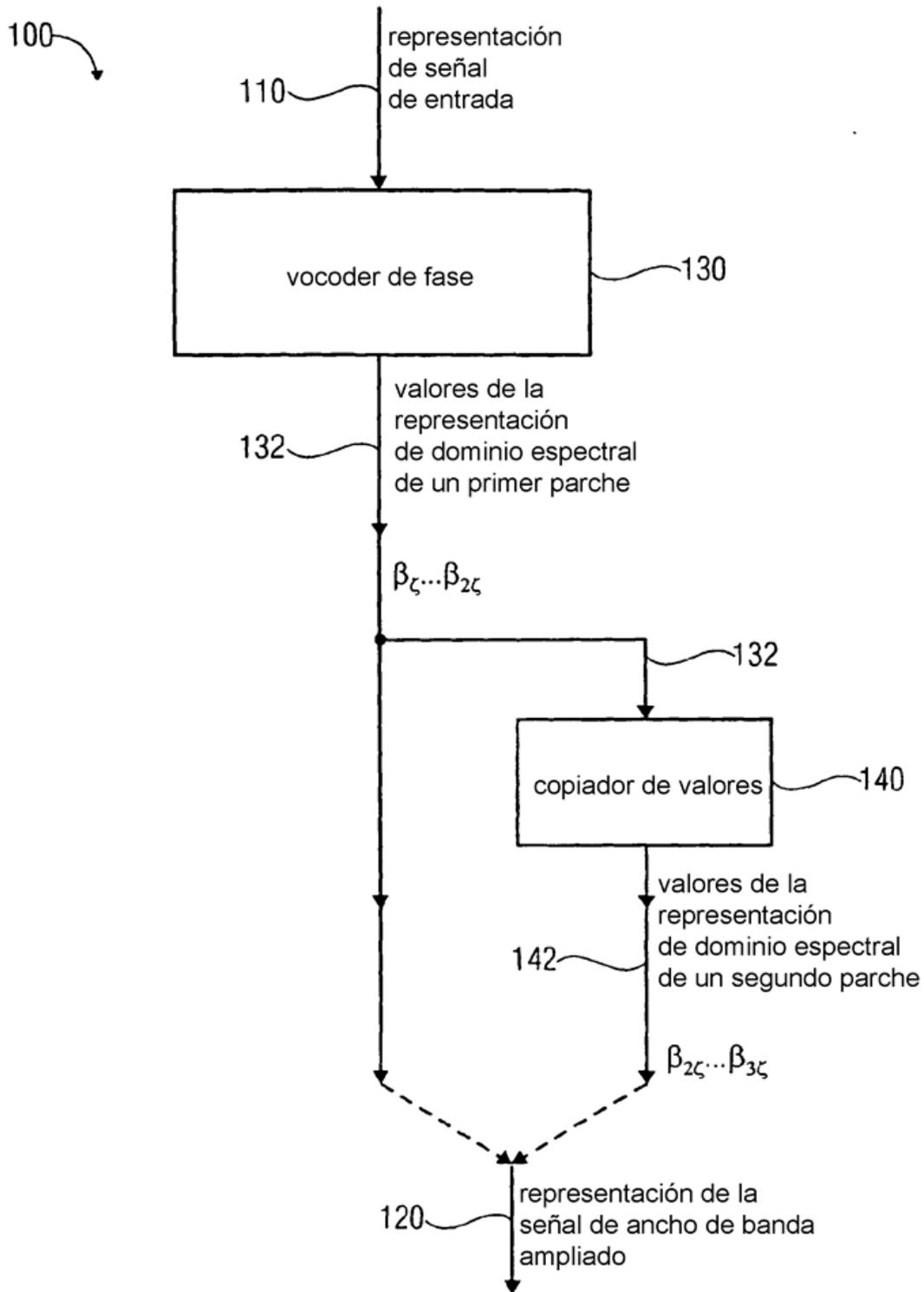


FIG 1

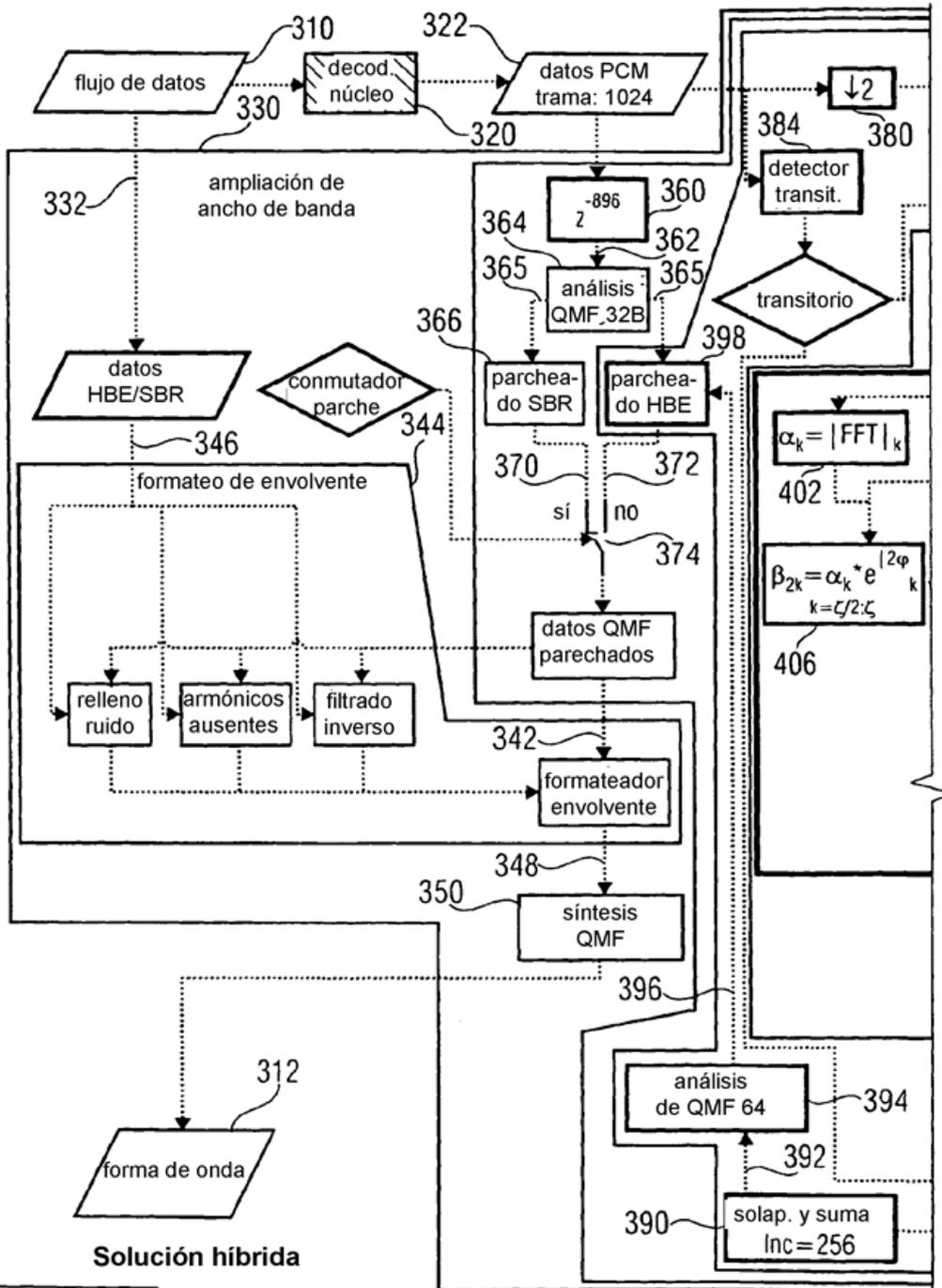


FIG 3A FIG 3B

FIG 3A

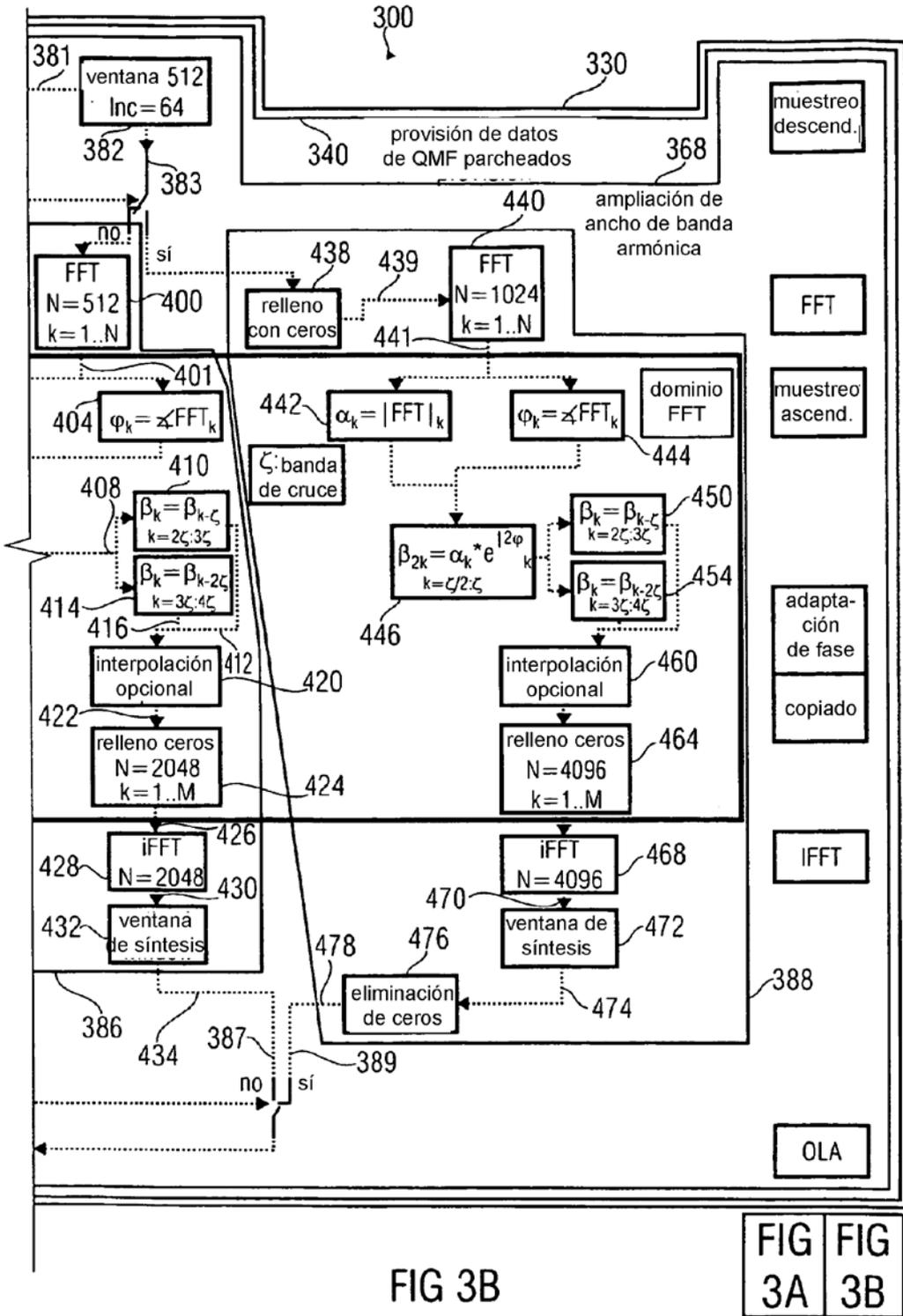


FIG 3B

FIG 3A
FIG 3B

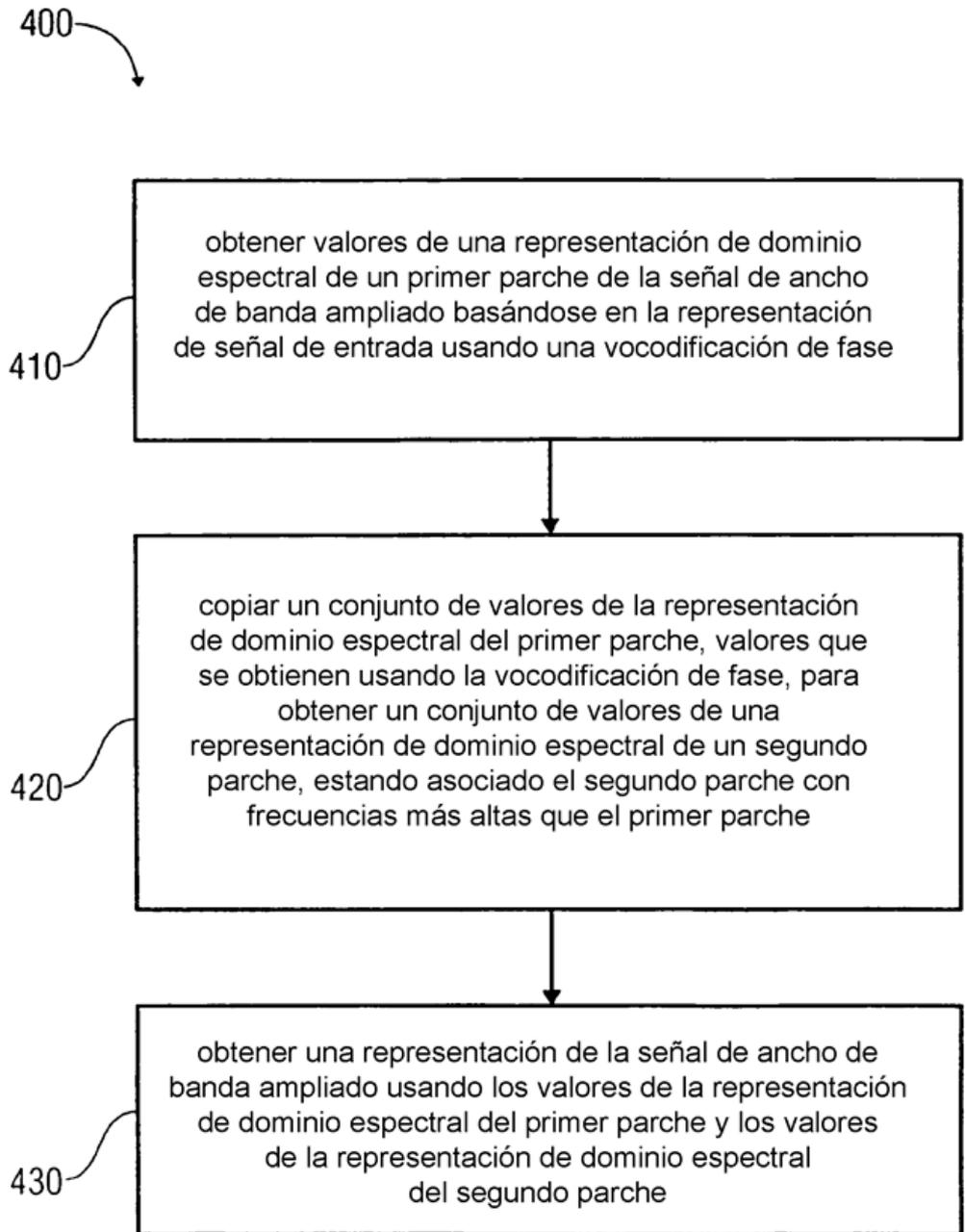


FIG 4

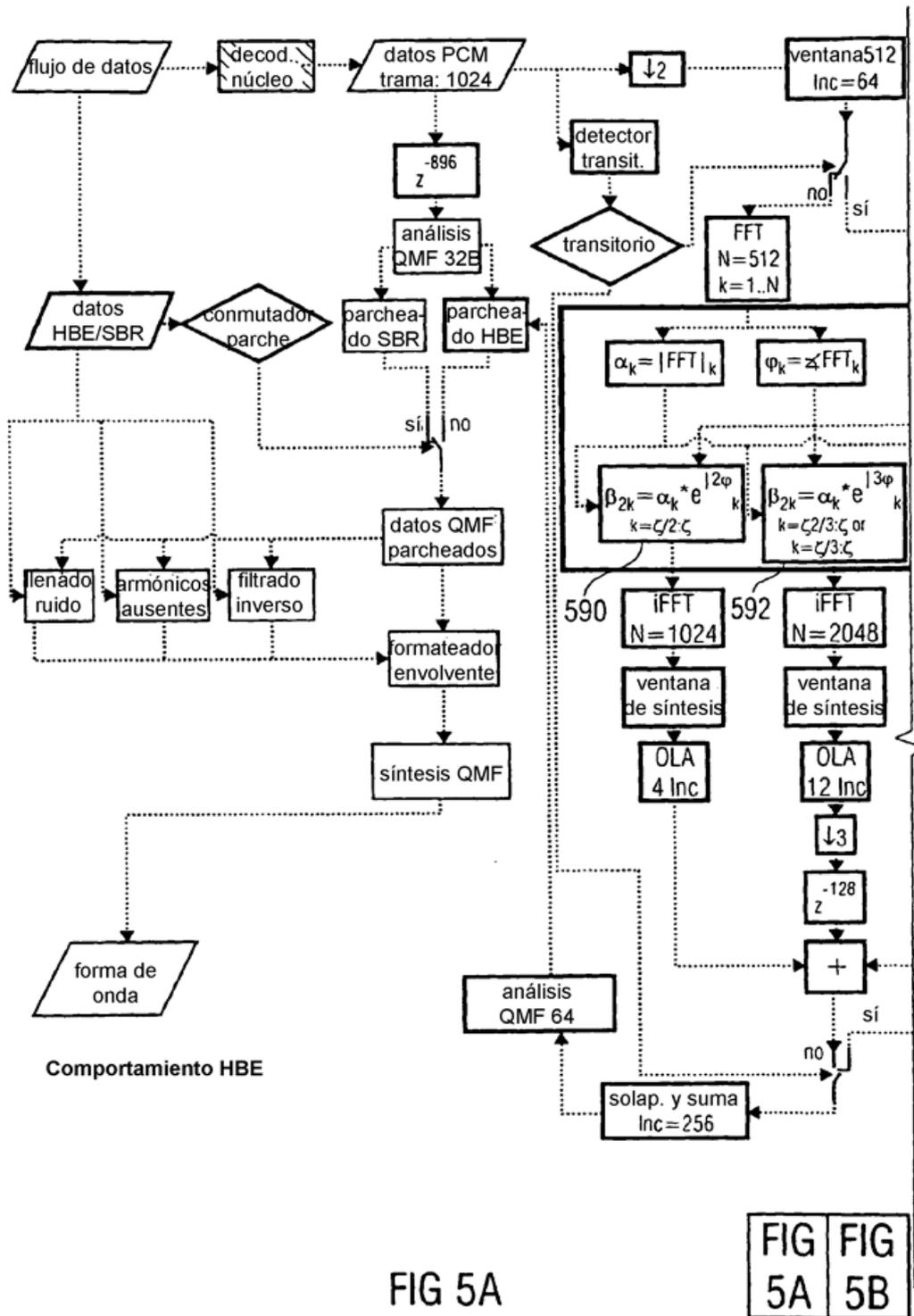


FIG 5A

FIG 5A	FIG 5B
-----------	-----------

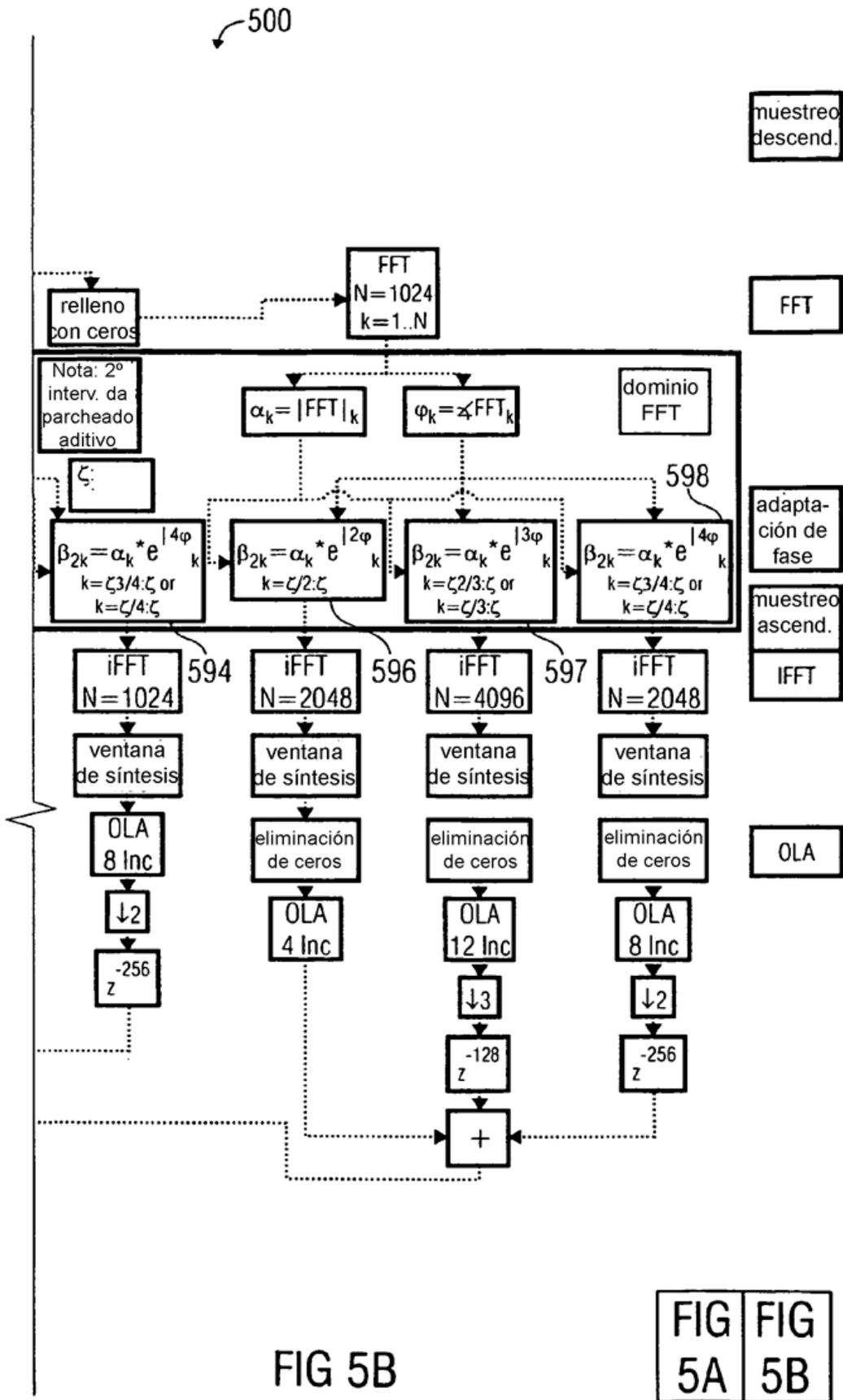
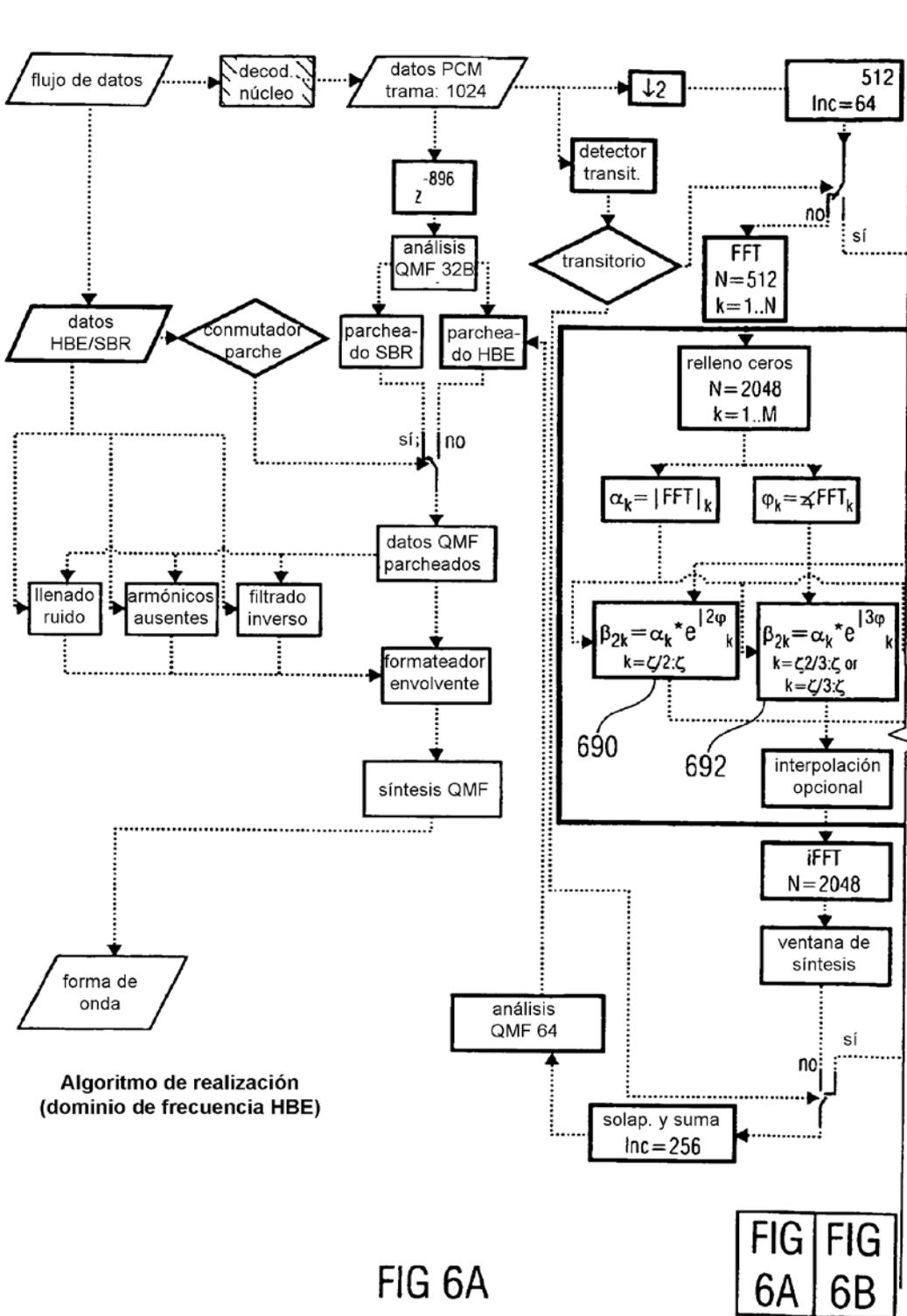


FIG 5B

FIG 5A FIG 5B



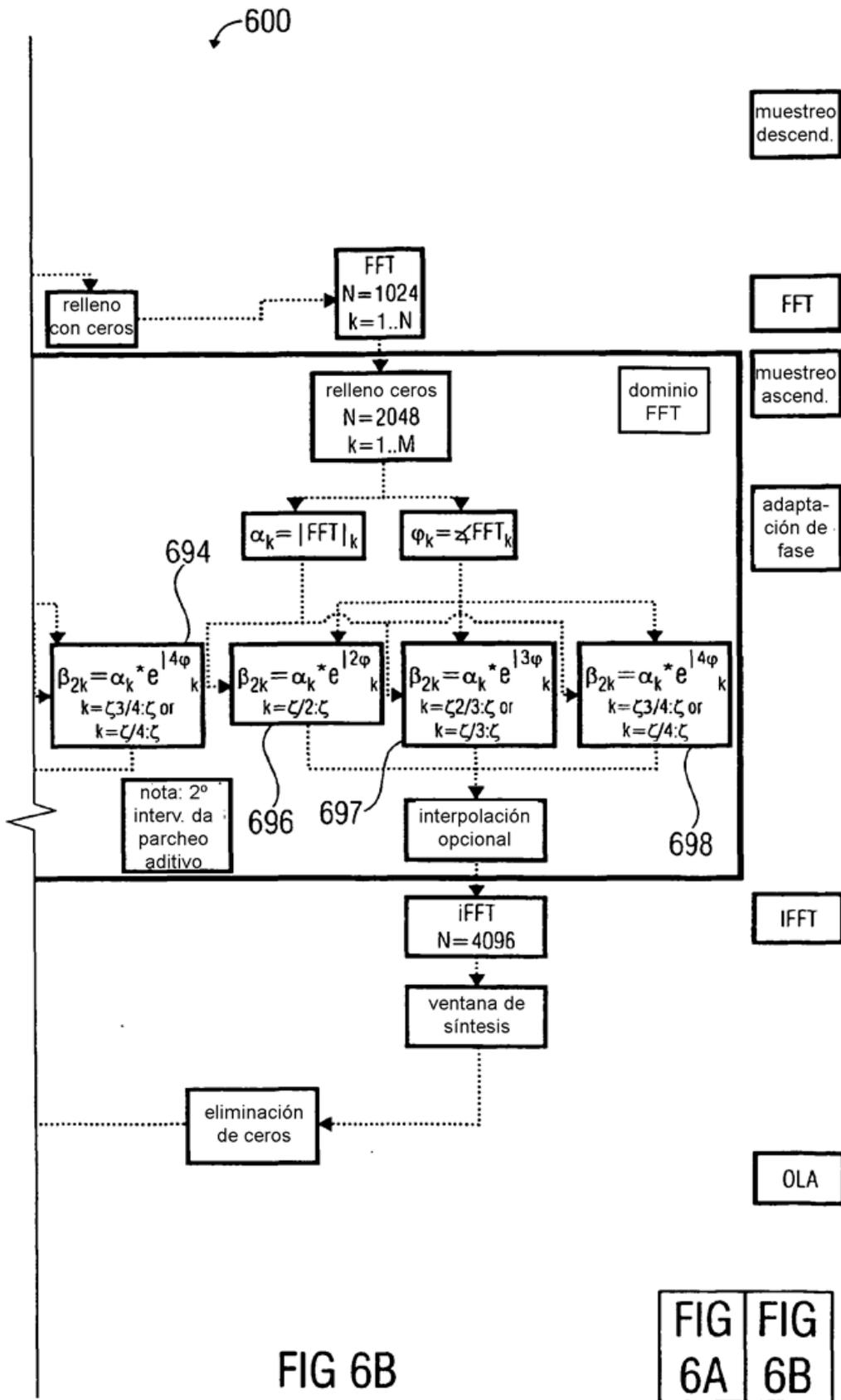


FIG 6B

FIG 6A
FIG 6B