



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS
ESPAÑA



⑪ Número de publicación: **2 391 292**

⑯ Int. Cl.:
G01L 21/02
(2006.01)

⑫

TRADUCCIÓN DE PATENTE EUROPEA

T3

- ⑯ Número de solicitud europea: **06784345 .8**
⑯ Fecha de presentación: **03.04.2006**
⑯ Número de publicación de la solicitud: **1864101**
⑯ Fecha de publicación de la solicitud: **12.12.2007**

⑭ Título: **Sistemas, procedimientos y aparato para la generación de una señal de excitación de banda alta**

⑯ Prioridad:
01.04.2005 US 667901 P
22.04.2005 US 673965 P

⑬ Titular/es:
QUALCOMM INCORPORATED (100.0%)
5775 Morehouse Drive
San Diego, CA 92121, US

⑯ Fecha de publicación de la mención BOPI:
23.11.2012

⑬ Inventor/es:
VOS, KOEN BERNARD y
KANDHADAI, ANANTHAPADMANABHAN A.

⑯ Fecha de la publicación del folleto de la patente:
23.11.2012

⑬ Agente/Representante:
CARPINTERO LÓPEZ, Mario

ES 2 391 292 T3

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

DESCRIPCIÓN

Sistemas, procedimientos y aparato para la generación de una señal de excitación de banda alta

Solicitudes relacionadas

- 5 La presente solicitud reivindica el beneficio de la solicitud provisional de patente estadounidense nº 60/667.901, titulada "CODING THE HIGH-FREQUENCY BAND OF WIDEBAND SPEECH", presentada el 1 de abril de 2005. La presente solicitud también reivindica el beneficio de la solicitud provisional de patente estadounidense nº 60/673.965, titulada "PARAMETER CODING IN A HIGH-BAND SPEECH CODER", presentada el 22 de abril de 2005.

Campo de la invención

- 10 La presente invención se refiere a procesamiento de señales.

Antecedentes

Las comunicaciones de voz a través de la red telefónica pública conmutada (RTPC) se han limitado tradicionalmente en ancho de banda al intervalo de frecuencia de 300-3400 kHz. Las nuevas redes para comunicaciones de voz, tales como telefonía celular y voz sobre IP (protocolo de Internet, VoIP) pueden no tener los mismos límites de ancho de banda, y puede ser deseable transmitir y recibir comunicaciones de voz que incluyan un intervalo de frecuencias de banda ancha a través de tales redes. Por ejemplo, puede ser deseable soportar un intervalo de frecuencias de audio que se extienda desde 50 Hz y/o hasta 7 u 8 kHz. También puede ser deseable soportar otras aplicaciones, tales como conferencia de audio o audio/vídeo de alta calidad, que pueden tener contenido de voz de audio en intervalos fuera de los límites de la RTPC tradicional.

20 El ensanchamiento del intervalo soportado por un codificador de voz a frecuencias superiores puede mejorar la inteligibilidad. Por ejemplo, la información que diferencia fricativas tales como 's' y 'f' está en gran parte en las frecuencias altas. El ensanchamiento de la banda alta también puede mejorar otras cualidades del habla, tal como la presencia. Por ejemplo, incluso una vocal vocalizada puede tener energía espectral muy por encima del límite de la RTPC.

25 Un enfoque con respecto a la codificación de voz de banda ancha implica ajustar a escala una técnica de codificación de voz de banda estrecha (por ejemplo, una configurada para codificar el intervalo de 0-4 kHz) para cubrir el espectro de banda ancha. Por ejemplo, una señal de voz puede muestrearse a una tasa superior para incluir componentes a frecuencias altas, y una técnica de codificación de banda estrecha puede volver a configurarse para usar más coeficientes de filtro para representar esta señal de banda ancha. Sin embargo, las técnicas de codificación de banda estrecha tales como CELP (predicción lineal excitada por código) son muy exigentes en lo que al cálculo se refiere, y un codificador CELP de banda ancha puede consumir demasiados ciclos de procesamiento como para ser práctico para muchas aplicaciones móviles y otras aplicaciones integradas. La codificación de todo el espectro de una señal de banda ancha a una calidad deseada usando tal técnica puede conducir a un aumento inaceptablemente grande en el ancho de banda. Además, se requeriría la transcodificación de una señal codificada de este tipo antes incluso de que su parte de banda estrecha pudiera transmitirse hacia un sistema que solo soporte codificación de banda estrecha y/o ser decodificada por él.

Otro enfoque con respecto a la codificación de voz de banda ancha implica extraer la envolvente espectral de banda alta a partir de la envolvente espectral de banda estrecha codificada. Aunque puede implementarse un enfoque de este tipo sin ningún aumento en el ancho de banda y sin necesidad de transcodificación, la envolvente espectral aproximada o estructura formante de la porción de banda alta de una señal de voz generalmente no puede predecirse con precisión a partir de la envolvente espectral de la parte de banda estrecha.

50 Puede ser deseable implementar la codificación de voz de banda ancha de tal manera que al menos la porción de banda estrecha de la señal codificada puede enviarse a través de un canal de banda estrecha (tal como un canal de RTPC) sin transcodificación u otra modificación significativa. También puede ser deseable la eficacia de la extensión de codificación de banda ancha, por ejemplo, para evitar una reducción significativa en el número de usuarios a los que puede darse servicio en aplicaciones tales como telefonía celular inalámbrica y difusión a través de canales inalámbricos y por cable. Se llama la atención, además, sobre el documento WO 03/044777, que versa acerca de un sistema de transmisión que comprende un transmisor para transmitir a un receptor una señal de audio de banda estrecha por medio de un canal de transmisión. El receptor comprende un ensanchador del ancho de banda para generar una señal de audio de banda ancha a partir de la señal de audio de banda estrecha. El ensanchador del ancho de banda comprende medios de plegado espectral para generar una señal de audio plegada espectralmente plegando espectralmente al menos parte de la señal de audio de banda estrecha. El ensanchador de ancho de banda del sistema de transmisión comprende un conformador de ruido para generar una señal conformada de ruido conformando una señal de ruido según al menos parte de la señal de audio plegada espectralmente, comprendiendo el ensanchador de ancho de banda un combinador para combinar la señal conformada de ruido y la señal de audio plegada espectralmente en la señal de audio de banda ancha.

Resumen

En la presente invención se proporcionan un procedimiento de generación de una señal de excitación de banda alta, según se expone en la reivindicación 1, un medio de almacenamiento de datos, según se expone en la reivindicación 17, y un aparato, según se expone en la reivindicación 18. En las reivindicaciones dependientes se reivindican realizaciones adicionales.

En una realización, un procedimiento de generación de una señal de excitación de banda alta incluye ensanchar armónicamente el espectro de una señal que se basa en una señal de excitación de banda baja; calcular una envolvente de dominio temporal de una señal que se basa en la señal de excitación de banda baja; y modular una señal de ruido según la envolvente de dominio temporal. El procedimiento también incluye combinar (A) una señal ensanchada armónicamente con base en un resultado del ensanchamiento armónico y (B) una señal modulada de ruido con base en un resultado de la modulación. En este procedimiento, la señal de excitación de banda alta está basada en un resultado de la combinación.

En otra realización, un aparato incluye un ensanchador de espectro configurado para llevar a cabo un ensanchamiento armónico del espectro de una señal que se basa en una señal de excitación de banda baja; un calculador de envolvente configurado para calcular una envolvente de dominio temporal de una señal que se basa en la señal de excitación de banda baja; un primer combinador configurado para llevar a cabo una modulación de una señal de ruido según la envolvente de dominio temporal; y un segundo combinador configurado para calcular una suma de (A) una señal ensanchada armónicamente con base en un resultado del ensanchamiento armónico y (B) una señal modulada de ruido con base en un resultado de la modulación. La señal de excitación de banda alta está basada en un resultado de la suma.

En otra realización, un aparato incluye medios para ensanchar armónicamente el espectro de una señal que se basa en una señal de excitación de banda baja; medios para calcular una envolvente de dominio temporal de una señal que se basa en la señal de excitación de banda baja; medios para modular una señal de ruido según la envolvente de dominio temporal; y medios para combinar (A) una señal ensanchada armónicamente con base en un resultado de dicho ensanchamiento armónico y (B) una señal modulada de ruido con base en un resultado de dicha modulación. En este aparato, la señal de excitación de banda alta está basada en un resultado de dicha combinación.

En otra realización, un procedimiento de generación de una señal de excitación de banda alta incluye calcular una señal ensanchada armónicamente aplicando una función no lineal a una señal de excitación de banda baja derivada de una porción de baja frecuencia de una señal de voz; y mezclar la señal ensanchada armónicamente con una señal modulada de ruido para generar una señal de excitación de banda alta.

Breve descripción de los dibujos

La FIGURA 1a muestra un diagrama de bloques de un codificador A100 de voz de banda ancha según una realización.

35 La FIGURA 1b muestra un diagrama de bloques de una implementación A102 del codificador A100 de voz de banda ancha.

La FIGURA 2a muestra un diagrama de bloques de un decodificador B100 de voz de banda ancha según una realización.

40 La FIGURA 2b muestra un diagrama de bloques de una implementación B102 del decodificador B100 de voz de banda ancha.

La FIGURA 3a muestra un diagrama de bloques de una implementación A112 del banco A110 de filtros.

La FIGURA 3b muestra un diagrama de bloques de una implementación B122 del banco B120 de filtros.

La FIGURA 4a muestra la cobertura de ancho de banda de las bandas alta y baja para un ejemplo del banco A110 de filtros.

45 La FIGURA 4b muestra la cobertura de ancho de banda de las bandas alta y baja para otro ejemplo del banco A110 de filtros.

La FIGURA 4c muestra un diagrama de bloques de una implementación A114 del banco A112 de filtros.

La FIGURA 4d muestra un diagrama de bloques de una implementación B124 del banco B122 de filtros.

50 La FIGURA 5a muestra un ejemplo de una representación gráfica del logaritmo de amplitud en función de la frecuencia para una señal de voz.

- La FIGURA 5b muestra un diagrama de bloques de un sistema básico de codificación de predicción lineal.
- La FIGURA 6 muestra un diagrama de bloques de una implementación A122 del codificador A120 de banda estrecha.
- 5 La FIGURA 7 muestra un diagrama de bloques de una implementación B112 de codificador B110 de banda estrecha.
- La FIGURA 8a muestra un ejemplo de una representación gráfica del logaritmo de amplitud en función de la frecuencia para una señal residual para habla vocalizada.
- La FIGURA 8b muestra un ejemplo de una representación gráfica del logaritmo de amplitud en función del tiempo para una señal residual para habla vocalizada.
- 10 La FIGURA 9 muestra un diagrama de bloques de un sistema básico de codificación de predicción lineal que también lleva a cabo predicción a largo plazo.
- La FIGURA 10 muestra un diagrama de bloques de una implementación A202 del codificador A200 de banda alta.
- La FIGURA 11 muestra un diagrama de bloques de una implementación A302 del generador A300 de excitación de banda alta.
- 15 La FIGURA 12 muestra un diagrama de bloques de una implementación A402 del ensanchador A400 de espectro.
- La FIGURA 12a muestra representaciones gráficas de espectros de señal en diversos puntos en un ejemplo de una operación de ensanchamiento espectral.
- La FIGURA 12b muestra representaciones gráficas de espectros de señal en diversos puntos de otro ejemplo de una operación de ensanchamiento espectral.
- 20 La FIGURA 13 muestra un diagrama de bloques de una implementación A304 del generador A302 de excitación de banda alta.
- La FIGURA 14 muestra un diagrama de bloques de una implementación A306 del generador A302 de excitación de banda alta.
- La FIGURA 15 muestra un diagrama de flujo para una tarea T100 de cálculo de envolvente.
- 25 La FIGURA 16 muestra un diagrama de bloques de una implementación 492 del combinador 490.
- La FIGURA 17 ilustra un enfoque para calcular una medida de periodicidad de la señal S30 de banda alta.
- La FIGURA 18 muestra un diagrama de bloques de una implementación A312 del generador A302 de excitación de banda alta.
- 30 La FIGURA 19 muestra un diagrama de bloques de una implementación A314 del generador A302 de excitación de banda alta.
- La FIGURA 20 muestra un diagrama de bloques de una implementación A316 del generador A302 de excitación de banda alta.
- La FIGURA 21 muestra un diagrama de flujo para una tarea T200 de cálculo de ganancia.
- 35 La FIGURA 22 muestra un diagrama de flujo para una implementación T210 de la tarea T200 de cálculo de ganancia.
- La FIGURA 23a muestra un diagrama de una función ventana.
- La FIGURA 23b muestra una aplicación de una función ventana, tal como se muestra en la FIGURA 23a, a subtramas de una señal de voz.
- 40 La FIGURA 24 muestra un diagrama de bloques para una implementación B202 del decodificador B200 de banda alta.
- La FIGURA 25 muestra un diagrama de bloques de una implementación AD10 del codificador A100 de voz de banda ancha.
- La FIGURA 26a muestra un diagrama esquemático de una implementación D122 de la línea D120 de retardo.
- La FIGURA 26b muestra un diagrama esquemático de una implementación D124 de la línea D120 de retardo.

La FIGURA 27 muestra un diagrama esquemático de una implementación D130 de línea D120 de retardo.

La FIGURA 28 muestra un diagrama de bloques de una implementación AD12 del codificador AD10 de voz de banda ancha.

5 La FIGURA 29 muestra un diagrama de flujo de un procedimiento de procesamiento MD100 de señales según una realización.

La FIGURA 30 muestra un diagrama de flujo para un procedimiento M100 según una realización.

La FIGURA 31a muestra un diagrama de flujo para un procedimiento M200 según una realización.

La FIGURA 31b muestra un diagrama de flujo para una implementación M210 del procedimiento M200.

La FIGURA 32 muestra un diagrama de flujo para un procedimiento M300 según una realización.

10 En las Figuras y la descripción adjunta, las mismas etiquetas de referencia se refieren a elementos o señales idénticos o análogos.

Descripción detallada

Tal como se describen en el presente documento, las realizaciones incluyen sistemas, procedimientos y aparatos que pueden configurarse para proporcionar una extensión a un codificador de voz de banda estrecha para soportar la transmisión y/o el almacenamiento de señales de voz de banda ancha en un aumento del ancho de banda de solo

15 aproximadamente 800 a 1000 bps (bits por segundo). Las ventajas potenciales de tales implementaciones incluyen la codificación integrada para soportar la compatibilidad con sistemas de banda estrecha, la asignación y la reasignación relativamente fáciles de bits entre los canales de codificación banda estrecha y de banda alta, evitando una operación de síntesis de banda ancha exigente desde el punto de vista del cálculo, y manteniendo una tasa de muestreo baja para señales que van a procesarse mediante rutinas de codificación de forma de onda exigentes desde el punto de vista del cálculo.

A no ser que se limite expresamente por su contexto, el término "calcular" se usa en el presente documento para indicar cualquiera de sus significados ordinarios, tales como computar, generar y seleccionar de una lista de valores. Cuando la expresión "que comprende" se usa en la presente descripción y las reivindicaciones, no excluye otros 25 elementos u operaciones. La expresión "A se basa en B" se usa para indicar cualquiera de sus significados habituales, incluyendo los casos (i) "A es igual a B" y (ii) "A se basa en al menos B". La expresión "protocolo de Internet" incluye la versión 4, tal como se describe en RFC (Petición de comentarios) 791de IETF (Grupo de Trabajo en Ingeniería de Internet), y versiones posteriores tal como la versión 6.

30 La FIGURA 1a muestra un diagrama de bloques de un codificador A100 de voz de banda ancha según una realización. El banco A110 de filtros está configurado para filtrar una señal S10 de voz de banda ancha para producir una señal S20 de banda estrecha y una señal S30 de banda alta. El codificador A120 de banda estrecha está configurado para codificar la señal S20 de banda estrecha para producir parámetros S40 de filtro de banda estrecha (BE) y una señal S50 residual de banda estrecha. Tal como se describe con mayor detalle en el presente documento, el codificador A120 de banda estrecha está configurado típicamente para producir parámetros S40 de 35 filtro de banda estrecha y la señal codificada S50 de excitación de banda estrecha como índices de código o en otra forma cuantificada. El codificador A200 de banda alta está configurado para codificar la señal S30 de banda alta según información en la señal codificada S50 de excitación de banda estrecha para producir parámetros S60 de codificación de banda alta. Tal como se describe con mayor detalle en el presente documento, el codificador A200 de banda alta está configurado típicamente para producir parámetros S60 de codificación de banda alta como 40 índices de código o en otra forma cuantificada. Un ejemplo particular del codificador A100 de voz de banda ancha está configurado para codificar la señal S10 de voz de banda ancha a una tasa de aproximadamente 8,55 kbps (kilobits por segundo), usándose aproximadamente 7,55 kbps para parámetros S40 de filtro de banda estrecha y la señal codificada S50 de excitación de banda estrecha, y usándose aproximadamente 1 kbps para parámetros S60 de codificación de banda alta.

45 Puede desecharse combinar las señales de banda estrecha y banda alta codificadas en una corriente de bits única. Por ejemplo, puede desecharse multiplexar las señales codificadas entre sí para su transmisión (por ejemplo, a través de un canal de transmisión por cable, óptico o inalámbrico), o para su almacenamiento, como una señal de voz de banda ancha codificada. La FIGURA 1b muestra un diagrama de bloques de una implementación A102 del codificador A100 de voz de banda ancha que incluye un multiplexor A130 configurado para combinar parámetros S40 de filtro de banda estrecha, la señal codificada S50 de excitación de banda estrecha y parámetros S60 de filtro de banda alta en una señal multiplexada S70.

55 Un aparato que incluye el codificador A102 también puede incluir un conjunto de circuitos configurado para transmitir la señal multiplexada S70 en un canal de transmisión tal como un canal por cable, óptico o inalámbrico. Un aparato tal también puede estar configurado para llevar a cabo una o más operaciones de codificación de canal en la señal, tal como codificación de corrección de errores (por ejemplo, codificación convolucional compatible con la tasa) y/o

codificación de detección de errores (por ejemplo, codificación de redundancia cíclica) y/o una o más capas de codificación de protocolo de red (por ejemplo, Ethernet, TCP/IP, cdma 2000).

Puede ser deseable que el multiplexor A130 esté configurado para integrar la señal codificada de banda estrecha (incluyendo parámetros S40 de filtro de banda estrecha y la señal codificada S50 de excitación de banda estrecha)

5 como una subcorriente separable de la señal multiplexada S70, de tal manera que pueda recuperarse la señal codificada de banda estrecha y decodificarse independientemente de otra porción de la señal multiplexada S70, tal como una señal de banda alta y/o de banda baja. Por ejemplo, puede disponerse la señal multiplexada S70 de tal manera que pueda recuperarse la señal codificada de banda estrecha eliminando los parámetros S60 de filtro de banda alta. Una ventaja potencial de tal característica es evitar la necesidad de transcodificar la señal de banda ancha codificada antes de pasarla a un sistema que soporte la decodificación de la señal de banda estrecha pero no soporte la decodificación de la porción de banda alta.

La FIGURA 2a es un diagrama de bloques de un decodificador B100 de voz de banda ancha según una realización. El decodificador B110 de banda estrecha está configurado para decodificar parámetros S40 de filtro de banda estrecha y la señal codificada S50 de excitación de banda estrecha para producir una señal S90 de banda estrecha.

15 El decodificador B200 de banda alta está configurado para decodificar parámetros S60 de codificación de banda alta según una señal S80 de excitación de banda estrecha, basándose en la señal codificada S50 de excitación de banda estrecha, para producir una señal S100 de banda alta. En este ejemplo, el decodificador B110 de banda estrecha está configurado para proporcionar la señal S80 de excitación de banda estrecha al decodificador B200 de banda alta. El banco B120 de filtros está configurado para combinar la señal S90 de banda estrecha y la señal S100 de banda alta para producir una señal S110 de voz de banda ancha.

20 La FIGURA 2b es un diagrama de bloques de una implementación B102 del decodificador B100 de voz de banda ancha que incluye un demultiplexor B130 configurado para producir señales S40, S50 y S60 codificadas a partir de la señal multiplexada S70. Un aparato que incluye un decodificador B102 puede incluir un conjunto de circuitos configurado para recibir la señal multiplexada S70 de un canal de transmisión tal como un canal por cable, óptico o

25 inalámbrico. Un aparato de este tipo también puede estar configurado para llevar a cabo una o más operaciones de decodificación de canal en la señal, tal como decodificación de corrección de errores (por ejemplo, decodificación convolucional compatible con la tasa) y/o decodificación de detección de errores (por ejemplo, decodificación de redundancia cíclica) y/o una o más capas de decodificación de protocolo de red (por ejemplo, Ethernet, TCP/IP, cdma 2000).

30 El banco A110 de filtros está configurado para filtrar una señal de entrada según un esquema de división en bandas para producir una subbanda de baja frecuencia y una subbanda de alta frecuencia. Dependiendo de los criterios de diseño para la aplicación particular, las subbandas de salida pueden tener anchos de banda iguales o desiguales y pueden superponerse o no superponerse. También es posible una configuración del banco A110 de filtros que produzca más de dos subbandas. Por ejemplo, un banco de filtros de este tipo puede estar configurado para producir una o más señales de banda baja que incluyen componentes en un intervalo de frecuencias por debajo del de la señal S20 de banda estrecha (tal como el intervalo de 50-300 Hz). También es posible que un banco de filtros de este tipo esté configurado para producir una o más señales de banda alta adicionales que incluyen componentes en un intervalo de frecuencias por encima del de la señal S30 de banda alta (tal como un intervalo de 14-20 kHz, 16-20 kHz o 16-32 kHz). En tal caso, puede implementarse el codificador A100 de voz de banda ancha para codificar esta señal o estas señales por separado, y el multiplexor A130 puede estar configurado para incluir la señal o las señales codificadas adicionales en la señal multiplexada S70 (por ejemplo, como una porción separable).

35 La FIGURA 3a muestra un diagrama de bloques de una implementación A112 del banco A110 de filtros que está configurado para producir dos señales de subbanda que tienen tasas de muestreo reducidas. El banco A110 de filtros está dispuesto para recibir una señal S10 de voz de banda ancha que tiene una porción de alta frecuencia (o

45 banda alta) y una porción de baja frecuencia (o banda baja). El banco A112 de filtros incluye un trayecto de procesamiento de banda baja configurado para recibir la señal S10 de voz de banda ancha y para producir la señal S20 de voz de banda estrecha, y un trayecto de procesamiento de banda alta configurado para recibir la señal S10 de voz de banda ancha y para producir la señal S30 de voz de banda alta. El filtro 110 de paso bajo filtra la señal S10 de voz de banda ancha para pasar una subbanda de baja frecuencia seleccionada, y el filtro 130 de paso alto filtra la señal S10 de voz de banda ancha para pasar una subbanda de alta frecuencia seleccionada. Debido a que ambas señales de subbanda tienen más anchos de banda estrechos que la señal S10 de voz de banda ancha, pueden reducirse sus tasas de muestreo hasta cierto punto sin pérdida de información. El submuestreador 120 reduce la tasa de muestreo de la señal de paso bajo según un factor de diezmado deseado (por ejemplo, eliminando muestras de la señal y/o sustituyendo muestras con valores promedio), y asimismo el submuestreador 140 reduce la tasa de muestreo de la señal de paso alto según otro factor de diezmado deseado.

50 La FIGURA 3b muestra un diagrama de bloques de una implementación B122 correspondiente del banco B120 de filtros. El sobremuestreador 150 aumenta la tasa de muestreo de la señal S90 de banda estrecha (por ejemplo, rellenando con ceros y/o duplicando muestras), y el filtro 160 de paso bajo filtra la señal sobremuestreada para que pase solo una porción de banda baja (por ejemplo, para prevenir el solapamiento). Asimismo, el sobremuestreador 170 aumenta la tasa de muestreo de la señal S100 de banda alta y el filtro 180 de paso alto filtra la señal

sobremuestreada para que pase solo una parte de banda alta. Entonces las dos señales de banda de paso se suman para formar la señal S110 de voz de banda ancha. En algunas implementaciones del decodificador B100, el banco B120 de filtros está configurado para producir una suma ponderada de las dos señales de banda de paso según uno o más pesos recibidos y/o calculados por el decodificador B200 de banda alta. También se contempla una configuración del banco B120 de filtros que combine más de dos señales de banda de paso.

5 Puede implementarse cada uno de los filtros 110, 130, 160, 180 como un filtro de respuesta de impulsos finita (FIR) o como un filtro de respuesta de impulsos infinita (IIR). Las respuestas de frecuencia de los filtros 110 y 130 de codificador pueden tener regiones de transición conformadas de manera simétrica o no similar entre la banda de parada y la banda de paso. Asimismo, las respuestas de frecuencia de los filtros 160 y 180 de decodificador pueden

10 tener regiones de transición conformadas de manera simétrica o no similar entre la banda de parada y la banda de paso. Puede ser deseable aunque no estrictamente necesario que el filtro 110 de paso bajo tenga la misma respuesta que el filtro 160 de paso bajo, y que el filtro 130 de paso alto tenga la misma respuesta que el filtro 180 de paso alto. En un ejemplo, los dos pares 110, 130 y 160, 180 de filtros son bancos de filtros de espejo en cuadratura (QMF), teniendo el par 110, 130 de filtros los mismos coeficientes que el par 160, 180 de filtros.

15 En un ejemplo típico, el filtro 110 de paso bajo tiene una banda de paso que incluye el intervalo de RTPC limitado de 300-3400 Hz (por ejemplo, la banda desde 0 kHz hasta 4 kHz). Las FIGURAS 4a y 4b muestran anchos de banda relativos de la señal S10 de voz de banda ancha, la señal S20 de banda estrecha y la señal S30 de banda alta en dos ejemplos de implementación diferentes. En ambos de estos ejemplos particulares, la señal S10 de voz de banda ancha tiene una tasa de muestreo de 16 kHz (que representa componentes de frecuencia dentro del intervalo de 0 a 8 kHz), y la señal S20 de banda estrecha tiene una tasa de muestreo de 8 kHz (que representa componentes de frecuencia dentro del intervalo de 0 a 4 kHz).

20 En el ejemplo de la FIGURA 4a, no hay superposición significativa entre las dos subbandas. Puede obtenerse una señal S30 de banda alta, tal como se muestra en este ejemplo usando un filtro 130 de paso alto con una banda de paso de 4-8 kHz. En tal caso, puede ser deseable reducir la tasa de muestreo a 8 kHz submuestreando la señal filtrada en un factor de dos. Una operación de este tipo, que puede esperarse que reduzca significativamente la complejidad de cálculo de operaciones de procesamiento adicionales en la señal, moverá la energía de banda de paso hasta el intervalo de 0 a 4 kHz sin pérdida de información.

25 En el ejemplo alternativo de la FIGURA 4b, las subbandas superior e inferior tienen una superposición apreciable, de tal manera que la región de 3,5 a 4 kHz se describe mediante ambas señales de subbanda. Puede obtenerse una señal S30 de banda alta como en este ejemplo usando un filtro 130 de paso alto con una banda de paso de 3,5-7 kHz. En tal caso, puede ser deseable reducir la tasa de muestreo a 7 kHz submuestreando la señal filtrada en un factor de 16/7. Una operación de este tipo, que puede esperarse que reduzca significativamente la complejidad de cálculo de operaciones de procesamiento adicionales en la señal, moverá la energía de banda de paso hasta el intervalo de 0 a 3,5 kHz sin pérdida de información.

30 35 En un microteléfono típico para comunicación telefónica, uno o más de los transductores (es decir, el microteléfono y el auricular o altavoz) carece de una respuesta apreciable sobre el intervalo de frecuencia de 7-8 kHz. En el ejemplo de la FIGURA 4b, la porción de la señal S10 de voz de banda ancha entre 7 y 8 kHz no se incluye en la señal codificada. Otros ejemplos particulares del filtro 130 de paso alto tienen bandas de paso de 3,5-7,5 kHz y 3,5-8 kHz.

40 45 En algunas implementaciones, proporcionar una superposición entre subbandas como en el ejemplo de la FIGURA 4b permite el uso de un filtro de paso bajo y/o de paso alto que tiene unas atenuaciones progresivas suaves sobre la región superpuesta. Normalmente, tales filtros son más fáciles de diseñar, menos complejos de calcular y/o introducen menos retardo que los filtros con respuestas más agudas o de "pared". Los filtros que tienen regiones de transición agudas tienden a tener lóbulos laterales más altos (que pueden provocar solapamiento) que los filtros de orden similar que tienen atenuaciones progresivas suaves. Los filtros que tienen regiones de transición agudas también pueden tener respuestas de impulso largas que pueden provocar artefactos de repiqueteo. Para las implementaciones del banco de filtros que tienen uno o más filtros de IIR, permitir una atenuación progresiva suave sobre la región superpuesta puede permitir el uso de un filtro o filtros cuyos polos están más alejados del círculo unitario, que pueden ser importantes para garantizar una implementación de punto fijo estable.

50 55 La superposición de subbandas permite una mezcla suave de banda baja y de banda alta que puede conducir a menos artefactos audibles, solapamiento reducido y/o una transición menos perceptible desde una banda a la otra. Además, la eficacia de codificación del codificador A120 de banda estrecha (por ejemplo, un codificador de forma de onda) puede caer con una frecuencia creciente. Por ejemplo, puede reducirse la calidad de codificación del codificador de banda estrecha a bajas tasas de transmisión de bits, especialmente en presencia de ruido de fondo. En tales casos, proporcionar una superposición de las subbandas puede aumentar la calidad de los componentes de frecuencia reproducidos en la región superpuesta.

Además, la superposición de subbandas permite una mezcla suave de banda baja y de banda alta que puede conducir a menos artefactos audibles, solapamiento reducido y/o una transición menos perceptible de una banda a la otra. Tal característica puede ser especialmente deseable para una implementación en la que el codificador A120 de banda estrecha y el codificador A200 de banda alta operan según metodologías de codificación diferentes. Por

ejemplo, técnicas de codificación diferentes pueden producir señales que suenan bastante diferentes. Un codificador que codifica una envolvente espectral en forma de índices de código puede producir una señal que tiene un sonido diferente que un codificador que codifica, en cambio, el espectro de amplitud. Un codificador en el dominio temporal (por ejemplo, una modulación por impulsos codificados o codificador PCM) puede producir una señal que tiene un sonido diferente que un codificador en el dominio frecuencial. Un codificador que codifica una señal con una representación de la envolvente espectral y la señal residual correspondiente puede producir una señal que tiene un sonido diferente que un codificador que codifica una señal con solo una representación de la envolvente espectral. Un codificador que codifica una señal como una representación de su forma de onda puede producir una salida que tiene un sonido diferente que el de un codificador sinusoidal. En tales casos, el uso de filtros que tienen regiones de transición agudas para definir subbandas no superpuestas puede llevar a una transición abrupta y perceptible entre las subbandas en la señal de banda ancha sintetizada.

Aunque a menudo se usan bancos de filtros QMF que tienen respuestas de frecuencia de superposición complementarias en técnicas de subbanda, tales filtros son inadecuados para al menos algunas de las implementaciones de codificación de banda ancha descritas en el presente documento. Un banco de filtros QMF en el codificador está configurado para crear un grado significativo de solapamiento que se cancela en el banco de filtros QMF correspondiente en el decodificador. Una disposición de este tipo puede no ser apropiada para una aplicación en la que la señal experimenta una cantidad significativa de distorsión entre los bancos de filtros, ya que la distorsión puede reducir la efectividad de la propiedad de cancelación de solapamiento. Por ejemplo, las aplicaciones descritas en el presente documento incluyen implementaciones de codificación configuradas para operar a tasas de transmisión de bits muy bajas. Como consecuencia de la tasa de transmisión de bits muy baja, es probable que la señal decodificada aparezca significativamente distorsionada en comparación con la señal original, de tal manera que el uso de bancos de filtros QMF puede llevar a un solapamiento no cancelado.

Además, un codificador puede estar configurado para producir una señal sintetizada que es perceptiblemente similar a la señal original pero que en realidad difiere significativamente de la señal original. Por ejemplo, un codificador que deriva la excitación de banda alta a partir del resto de banda estrecha, tal como se describe en el presente documento, puede producir una señal de este tipo, ya que el resto de banda alta real puede estar completamente ausente de la señal decodificada. El uso de bancos de filtros QMF en tales aplicaciones puede conducir a un grado significativo de distorsión provocado por solapamiento no cancelado. Las aplicaciones que usan bancos de filtros QMF tienen normalmente tasas de transmisión de bits superiores (por ejemplo, por encima de 12 kbps para AMR, y de 64 kbps para G.722).

Puede reducirse la cantidad de distorsión provocada por solapamiento de QMF si la subbanda afectada es estrecha, ya que el efecto del solapamiento se limita a un ancho de banda igual al ancho de la subbanda. Sin embargo, para ejemplos tal como se describen en el presente documento en los que cada subbanda incluye aproximadamente la mitad del ancho de banda de banda ancha, la distorsión provocada por solapamiento no cancelado podría afectar a una parte significativa de la señal. También puede verse afectada la calidad de la señal por la ubicación de la banda de frecuencia sobre la que se produce el solapamiento no cancelado. Por ejemplo, la distorsión creada cerca del centro de una señal de voz de banda ancha (por ejemplo, entre 3 y 4 kHz) puede ser mucho más inaceptable que la distorsión que se produce cerca de un borde de la señal (por ejemplo, por encima de 6 kHz).

Aunque las respuestas de los filtros de un banco de filtros QMF están relacionadas estrictamente entre sí, pueden configurarse los trayectos de banda baja y banda alta de los bancos A110 y B120 de filtros para que tengan espectros que no tienen relación ninguna aparte de la superposición de las dos subbandas. Se define la superposición de las dos subbandas como la distancia desde el punto en el que la respuesta de frecuencia del filtro de banda alta cae a -20 dB hasta el punto en el que la respuesta de frecuencia del filtro de banda baja cae a -20 dB. En diversos ejemplos del banco A110 y/o B120 de filtros, esta superposición oscila desde aproximadamente 200 Hz hasta aproximadamente 1 kHz. El intervalo de aproximadamente 400 hasta aproximadamente 600 Hz puede representar un equilibrio deseado entre la eficacia de codificación y la suavidad de percepción. En un ejemplo particular, tal como se ha mencionado anteriormente, la superposición es aproximadamente 500 Hz.

Puede ser deseable implementar el banco A112 y/o B122 de filtros para llevar a cabo operaciones tal como se ilustra en las FIGURAS 4a y 4b en varias etapas. Por ejemplo, la FIGURA 4c muestra un diagrama de bloques de una implementación A114 del banco A112 de filtros que lleva a cabo un equivalente funcional de operaciones de submuestreo y filtrado de paso alto usando una serie de operaciones de interpolación, realización de nuevo muestreo, diezmado y otras. Puede ser más fácil de diseñar una implementación de este tipo y/o puede permitir la reutilización de bloques funcionales de lógica y/o código. Por ejemplo, puede usarse el mismo bloque funcional para llevar a cabo las operaciones de diezmado a 14 kHz y diezmado a 7 kHz, tal como se muestra en la FIGURA 4c. Puede implementarse la operación de inversión espectral multiplicando la señal por la función $e^{-j\pi n}$ o la secuencia $(-1)^n$, cuyos valores alternan entre +1 y -1. Puede implementarse la operación de conformación espectral como un filtro de paso bajo configurado para conformar la señal para obtener una respuesta de filtro global deseada.

Se hace notar que como consecuencia de la operación de inversión espectral, se invierte el espectro de la señal S30 de banda alta. En consecuencia, pueden configurarse operaciones posteriores en el codificador y el decodificador correspondiente. Por ejemplo, el generador A300 de excitación de banda alta, tal como se describe en el presente

documento, puede estar configurado para producir una señal S120 de excitación de banda alta que también tiene una forma invertida espectralmente.

La FIGURA 4d muestra un diagrama de bloques de una implementación B124 del banco B122 de filtros que lleva a cabo un equivalente funcional de operaciones de sobremuestreo y filtrado de paso alto usando una serie de operaciones de interpolación, realización de nuevo muestreo y otras. El banco B124 de filtros incluye una operación de inversión espectral en la banda alta que invierte una operación similar tal como se lleva a cabo, por ejemplo, en un banco de filtros del codificador tal como el banco A114 de filtros. En este ejemplo particular, el banco B124 de filtros también incluye filtros de muesca en la banda baja y la banda alta que atenúan un componente de la señal a 7100 Hz, aunque tales filtros son opcionales y no es preciso que estén incluidos. La solicitud de patente "SYSTEMS, METHODS, AND APPARATUS FOR SPEECH SIGNAL FILTERING", presentada con el presente documento, expediente de agente 050551, incluye una descripción y figuras adicionales relativas a respuestas de elementos de implementaciones particulares de los bancos A110 y B120 de filtros, y este material se incorpora al presente documento por referencia.

Se implementa el codificador A120 de banda estrecha según un modelo de filtro de fuente que codifica la señal de voz de entrada como (A) un conjunto de parámetros que describen un filtro y (B) una señal de excitación que hace que el filtro descrito produzca una reproducción sintetizada de la señal de voz de entrada. La FIGURA 5a muestra un ejemplo de una envolvente espectral de una señal de voz. Los picos que caracterizan esta envolvente espectral representan resonancias del tracto vocal y se denominan formantes. La mayoría de codificadores de voz codifican al menos esta estructura espectral tosca como un conjunto de parámetros tales como coeficientes de filtro.

La FIGURA 5b muestra un ejemplo de una disposición básica de filtro de fuente tal como se aplica para la codificación de la envolvente espectral de la señal S20 de banda estrecha. Un módulo de análisis calcula un conjunto de parámetros que caracterizan un filtro que corresponde a un sonido de voz durante un periodo de tiempo (normalmente 20 ms). Un filtro blanqueador (también denominado un filtro de análisis o de error de predicción) configurado según esos parámetros de filtro elimina la envolvente espectral para aplanar espectralmente la señal. La señal blanqueada resultante (también denominada un resto) tiene menos energía y, por lo tanto, menos varianza y es más fácil de codificar que la señal de voz original. También pueden propagarse los errores resultantes de la codificación de la señal residual más uniformemente en todo el espectro. Los parámetros de filtro y el resto se cuantifican normalmente para su transmisión eficaz a través del canal. En el decodificador, un filtro de síntesis configurado según los parámetros de filtro se excita por una señal basada en el resto para producir una versión sintetizada del sonido de voz original. Típicamente, el filtro de síntesis está configurado para tener una función de transferencia que es la inversa de la función de transferencia del filtro blanqueador.

La FIGURA 6 muestra un diagrama de bloques de una implementación A122 básica del codificador A120 de banda estrecha. En este ejemplo, un módulo 210 de análisis de codificación de predicción lineal (LPC) codifica la envolvente espectral de la señal S20 de banda estrecha como un conjunto de coeficientes de predicción lineal (LP) (por ejemplo, coeficientes de un filtro omnipolar 1/A(z)). Normalmente, el módulo de análisis procesa la señal de entrada como una serie de tramas no superpuestas, calculándose un nuevo conjunto de coeficientes para cada trama. Generalmente, el periodo de trama es un periodo durante el que puede esperarse que la señal sea estacionaria localmente; un ejemplo común es 20 milisegundos (equivalente a 160 muestras a una tasa de muestreo de 8 kHz). En un ejemplo, el módulo 210 de análisis de LPC está configurado para calcular un conjunto de diez coeficientes de filtro de LP para caracterizar la estructura formante de cada trama de 20 milisegundos. También es posible implementar el módulo de análisis para procesar la señal de entrada como una serie de tramas superpuestas.

El módulo de análisis puede estar configurado para analizar las muestras de cada trama directamente, o pueden ponderarse las muestras primero según una función ventana (por ejemplo, una ventana de Hamming). También puede llevarse a cabo el análisis a través de una ventana que sea más grande que la trama, tal como una ventana de 30 ms. Esta ventana puede ser simétrica (por ejemplo 5-20-5, de tal manera que incluye los 5 milisegundos inmediatamente anteriores y posteriores a la trama de 20 milisegundos) o asimétrica (por ejemplo 10-20, de tal manera que incluya los últimos 10 milisegundos de la trama anterior). Normalmente, un módulo de análisis de LPC está configurado para calcular los coeficientes de filtro de LP usando una recursión de Levinson-Durbin o el algoritmo de Leroux-Gueguen. En otra implementación, el módulo de análisis puede estar configurado para calcular un conjunto de coeficientes cepstrales para cada trama en lugar de un conjunto de coeficientes de filtro de LP.

La velocidad de salida del codificador A120 puede reducirse significativamente, con relativamente poco efecto sobre la calidad de reproducción, cuantificando los parámetros de filtro. Los coeficientes de filtro de predicción lineal son difíciles de cuantificar eficazmente y habitualmente se correlacionan con otra representación, tal como pares espirales de línea (LSP) o frecuencias espirales de línea (LSF), para cuantificación y/o codificación de entropía. En el ejemplo de la FIGURA 6, la transformada 220 de coeficiente de filtro de LP a LSF transforma el conjunto de coeficientes de filtro de LP en un conjunto correspondiente de las LSF. Otras representaciones uno a uno de coeficientes de filtro de LP incluyen coeficientes de correlación parcial; valores de la razón de área logarítmica; pares espirales de inmitancia (ISP); y frecuencias espirales de inmitancia (ISF), que se usan en el códec de AMR-WB (ancho de banda de tasa múltiple adaptativa) de GSM (Sistema general para las comunicaciones móviles).

Normalmente una transformada entre un conjunto de coeficientes de filtro de LP y un conjunto correspondiente de las LSF es reversible, pero las realizaciones también incluyen implementaciones del codificador A120 en las que la transformada no es reversible sin error.

5 El cuantificador 230 está configurado para cuantificar el conjunto de las LSF de banda estrecha (u otra representación de coeficiente), y el codificador A122 de banda estrecha está configurado para emitir el resultado de esta cuantificación como los parámetros S40 de filtro de banda estrecha. Normalmente, un cuantificador de este tipo incluye un cuantificador de vector que codifica el vector de entrada como un índice con respecto a una entrada de vector correspondiente en una tabla o código.

10 Tal como se observa en la FIGURA 6, el codificador A122 de banda estrecha también genera una señal residual dejando pasar la señal S20 de banda estrecha a través de un filtro blanqueador 260 (también denominado filtro de análisis o de error de predicción) que está configurado según el conjunto de coeficientes de filtro. En este ejemplo particular, se implementa el filtro blanqueador 260 como un filtro de FIR, aunque también pueden usarse implementaciones de IIR. Normalmente, esta señal residual contendrá información importante en cuanto a la percepción de la trama de voz, tal como la estructura a largo plazo relativa a la altura tonal, que no se representa en los parámetros S40 de filtro de banda estrecha. El cuantificador 270 está configurado para calcular una representación cuantificada de esta señal residual para su emisión como la señal codificada S50 de excitación de banda estrecha. Normalmente, un cuantificador de este tipo incluye un cuantificador de vector que codifica el vector de entrada como un índice con respecto a una entrada de vector correspondiente en una tabla o código. Como alternativa, un cuantificador de este tipo puede estar configurado para enviar uno o más parámetros a partir de los cuales puede generarse el vector dinámicamente en el decodificador, en lugar de recuperarse del almacenamiento, como en un procedimiento de código disperso. Se usa un procedimiento de este tipo en esquemas de codificación tales como CELP algebraica (predicción lineal de excitación por código) y códecs tales como el EVRC (Códec de tasa variable mejorada) de 3GPP2 (Asociación 2 de Tercera Generación).

25 Es deseable que el codificador A120 de banda estrecha genere la señal codificada de excitación de banda estrecha según los mismos valores de parámetros de filtro que estarán disponibles para el correspondiente decodificador de banda estrecha. De esta manera, la señal codificada de excitación de banda estrecha resultante ya puede representar en cierta medida las no idealidades en esos valores de parámetros, tales como el error de cuantificación. Por consiguiente, es deseable configurar el filtro blanqueador usando los mismos valores de coeficiente que estarán disponibles en el decodificador. En el ejemplo básico del codificador A122 tal como se muestra en la FIGURA 6, el cuantificador inverso 240 descuantifica los parámetros S40 de codificación de banda estrecha, la transformada 250 de coeficiente de filtro de LSF a LP vuelve a correlacionar los valores resultantes con un conjunto correspondiente de coeficientes de filtro de LP, y este conjunto de coeficientes se usa para configurar el filtro blanqueador 260 para generar la señal residual que es cuantificada por el cuantificador 270.

35 Algunas implementaciones del codificador A120 de banda estrecha están configuradas para calcular la señal codificada S50 de excitación de banda estrecha identificando uno entre un conjunto de vectores de código que mejor se ajusta a la señal residual. Sin embargo, se hace notar que también puede implementarse el codificador A120 de banda estrecha para calcular una representación cuantificada de la señal residual sin generar en realidad la señal residual. Por ejemplo, el codificador A120 de banda estrecha puede estar configurado para usar un número de vectores de código para generar señales sintetizadas correspondientes (por ejemplo, según un conjunto actual de parámetros de filtro), y para seleccionar el vector de código asociado con la señal generada que mejor se ajuste a la señal S20 de banda estrecha original en un dominio ponderado según la percepción.

45 La FIGURA 7 muestra un diagrama de bloques de una implementación B112 del decodificador B110 de banda estrecha. El cuantificador inverso 310 descuantifica los parámetros S40 de filtro de banda estrecha (en este caso, a un conjunto de LSF), y la transformada 320 de coeficiente de filtro de LP a LSF transforma las LSF en un conjunto de coeficientes de filtro (por ejemplo, tal como se describió anteriormente con referencia al cuantificador inverso 240 y la transformada 250 del codificador A122 de banda estrecha). El cuantificador inverso 340 descuantifica la señal codificada S50 de excitación de banda estrecha para producir una señal S80 de excitación de banda estrecha. Basándose en los coeficientes de filtro y la señal S80 de excitación de banda estrecha, el filtro 330 de síntesis de banda estrecha sintetiza la señal S90 de banda estrecha. En otras palabras, el filtro 330 de síntesis de banda estrecha está configurado para conformar espectralmente la señal S80 de excitación de banda estrecha según los coeficientes de filtro descuantificados para producir la señal S90 de banda estrecha. El decodificador B112 de banda estrecha también proporciona la señal S80 de excitación de banda estrecha al codificador A200 de banda alta, que la usa para derivar la señal S120 de excitación de banda alta, tal como se describe en el presente documento. En algunas implementaciones, tal como se describe en lo que sigue, el decodificador B110 de banda estrecha puede estar configurado para proporcionar información adicional al decodificador B200 de banda alta que se relaciona con la señal de banda estrecha, tal como inclinación espectral, ganancia y retraso de altura tonal, y modo de voz.

55 El sistema del codificador A122 de banda estrecha y el decodificador B112 de banda estrecha es un ejemplo básico de un códec de voz de análisis por síntesis. La codificación de predicción lineal de excitación de código (CELP) es una familia popular de codificación de análisis por síntesis, y las implementaciones de tales codificadores pueden 60 llevar a cabo la codificación de forma de onda del resto, incluyendo operaciones tales como la selección de entradas

a partir de códigos fijos y adaptativos, operaciones de minimización de error y/u operaciones de ponderación perceptual. Otras implementaciones de codificación de análisis por síntesis incluyen la codificación de predicción lineal de excitación mixta (MELP), CELP algebraica (ACELP), CELP de relajación (RCELP), excitación de impulsos regulares (RPE), CELP de impulsos múltiples (MPE), y predicción lineal excitada por suma de vectores (VSELP).

- 5 Los procedimientos de codificación relacionados incluyen la codificación de excitación de bandas múltiples (MBE) y la interpolación de forma de onda de prototipo (PWI). Los ejemplos de códecs estandarizados de voz de análisis por síntesis incluyen el códec de tasa completa de GSM (GSM 06.10) del ETSI (Instituto Europeo de Normas de Telecommunicaciones), que usa predicción lineal excitada residual (RELP); el códec de tasa completa mejorada de GSM (ETSI-GSM 06.60); el codificador del anexo E G.729 de 11,8 kb/s de la norma de ITU (Unión Internacional de Telecommunicaciones); los códecs de IS (Norma provisional)-641 para IS-136 (un esquema de acceso múltiple por división de tiempo); los códecs de velocidades múltiples adaptativas de GSM (GSM-AMR); y el códec de 4GV™ (Vocoder™ de cuarta generación) (QUALCOMM Incorporated, San Diego, California). Pueden implementarse el codificador A120 de banda estrecha y el correspondiente decodificador B110 según cualquiera de estas tecnologías o cualquier otra tecnología de codificación de voz (ya sea conocida o que vaya a desarrollarse) que represente una señal de voz como (A) un conjunto de parámetros que describen un filtro y (B) una señal de excitación usada para hacer que el filtro descrito reproduzca la señal de voz.

Incluso después de que el filtro blanqueador haya eliminado la envolvente espectral tosca de la señal S20 de banda estrecha, puede permanecer una cantidad considerable de estructura armónica fina, especialmente para habla vocalizada. La FIGURA 8a muestra una representación gráfica espectral de un ejemplo de una señal residual, tal

- 20 como puede ser producida mediante un filtro blanqueador, para una señal sonora tal como una vocal. La estructura periódica visible en este ejemplo se refiere a la altura tonal, y diferentes sonidos vocalizados hablados por el mismo orador pueden tener diferentes estructuras formantes, pero estructuras de altura tonal similar. La FIGURA 8b muestra una representación gráfica en el dominio temporal de un ejemplo de una señal residual de este tipo que muestra una secuencia de impulsos de altura tonal en el tiempo.

- 25 Puede aumentarse la eficacia de codificación y/o la calidad de la voz usando uno o más valores de parámetros para codificar características de la estructura de altura tonal. Una característica importante de la estructura de altura tonal es la frecuencia del primer armónico (también denominada frecuencia fundamental), que normalmente está en el intervalo de 60 a 400 Hz. Normalmente, esta característica se codifica como la inversa de la frecuencia fundamental, también denominada retraso de altura tonal. El retraso de altura tonal indica el número de muestras en un periodo de altura tonal y puede codificarse como uno o más índices de código. Las señales de voz de los oradores tienden a tener retrasos de altura tonal más grandes que las señales de voz de las oradoras.

Otra característica de la señal en relación con la estructura de altura tonal es la periodicidad, que indica la intensidad de la estructura armónica o, en otras palabras, el grado en el que la señal es armónica o no armónica. Dos indicadores típicos de periodicidad son los cruces por cero y las funciones de autocorrelación normalizadas (NACF).

- 35 También puede indicarse la periodicidad mediante la ganancia de altura tonal, que se codifica comúnmente como una ganancia de código (por ejemplo, una ganancia de código adaptativa cuantificada).

- 40 El codificador A120 de banda estrecha puede incluir uno o más módulos configurados para codificar la estructura armónica a largo plazo de la señal S20 de banda estrecha. Tal como se muestra en la FIGURA 9, un paradigma de CELP típico que puede usarse incluye un módulo de análisis de LPC de bucle abierto, que codifica las características a corto plazo o la envolvente espectral tosca, seguido por una etapa de análisis de predicción a largo plazo de bucle cerrado, que codifica la altura tonal fina o la estructura armónica. Las características a corto plazo se codifican como coeficientes de filtro, y las características a largo plazo se codifican como valores para parámetros tales como retraso de altura tonal y ganancia de altura tonal. Por ejemplo, el codificador A120 de banda estrecha puede estar configurado para emitir la señal codificada S50 de excitación de banda estrecha en una forma que incluye uno o más índices de código (por ejemplo, un índice de código fijo y un índice de código adaptativo) y valores de ganancia correspondientes. El cálculo de esta representación cuantificada de la señal residual de banda estrecha (por ejemplo, por el cuantificador 270) puede incluir seleccionar tales índices y calcular tales valores. La codificación de la estructura de altura tonal también puede incluir la interpolación de una forma de onda de prototipo de altura tonal, operación que puede incluir calcular una diferencia entre impulsos de altura tonal sucesivos. El modelado de la estructura a largo plazo puede inhabilitarse para tramas correspondientes a habla no vocalizada, que normalmente es de tipo ruido y no estructurada.

- 50 Una implementación del decodificador B110 de banda estrecha según un paradigma como el mostrado en la FIGURA 9 puede estar configurada para emitir la señal S80 de excitación de banda estrecha al decodificador B200 de banda alta tras haberse restaurado la estructura a largo plazo (estructura armónica o de altura tonal). Por ejemplo, un decodificador de este tipo puede estar configurado para emitir la señal S80 de excitación de banda estrecha como una versión descuantificada de la señal codificada S50 de excitación de banda estrecha. Naturalmente, también es posible implementar el decodificador B110 de banda estrecha de tal manera que el decodificador B200 de banda alta lleve a cabo la descuantificación de la señal codificada S50 de excitación de banda estrecha para obtener la señal S80 de excitación de banda estrecha.

En una implementación del codificador A100 de voz de banda ancha según un paradigma tal como el mostrado en la FIGURA 9, el codificador A200 de banda alta puede estar configurado para recibir la señal de excitación de banda estrecha tal como se produjo mediante el análisis a corto plazo o el filtro blanqueador. En otras palabras, el codificador A120 de banda estrecha puede estar configurado para emitir la señal de excitación de banda estrecha al codificador A200 de banda alta antes de la codificación de la estructura a largo plazo. Sin embargo, es deseable que el codificador A200 de banda alta reciba del canal de banda estrecha la misma información de codificación que recibirá el decodificador B200 de banda alta, de tal manera que los parámetros de codificación producidos por el codificador A200 de banda alta puedan ya representar en cierta medida las no idealidades en esa información. Por tanto, puede resultar preferible que el codificador A200 de banda alta reconstruya la señal S80 de excitación de banda estrecha a partir de la misma señal codificada S50 de excitación de banda estrecha parametrizada y/o cuantificada que el codificador A100 de voz de banda ancha va emitir. Una ventaja potencial de este enfoque es el cálculo más preciso de los factores S60b de ganancia de banda alta que se describen a continuación.

Además de los parámetros que caracterizan la estructura a corto plazo y/o largo plazo de la señal S20 de banda estrecha, el codificador A120 de banda estrecha puede producir valores de parámetros que se relacionan con otras 15 características de la señal S20 de banda estrecha. Estos valores, que pueden cuantificarse adecuadamente para ser emitidos mediante el codificador A100 de voz de banda ancha, pueden ser incluidos entre los parámetros S40 de filtro de banda estrecha o ser emitidos por separado. El codificador A200 de banda alta también puede estar configurado para calcular parámetros S60 de codificación de banda alta según uno o más de estos parámetros adicionales (por ejemplo, tras la descuantificación). En el decodificador B100 de voz de banda ancha, un 20 decodificador B200 de banda alta puede estar configurado para recibir los valores de parámetros por medio del decodificador B110 de banda estrecha (por ejemplo, tras la descuantificación). Como alternativa, el decodificador B200 de banda alta puede estar configurado para recibir (y posiblemente para descuantificar) los valores de parámetros directamente.

En un ejemplo de parámetros de codificación de banda estrecha adicionales, el codificador A120 de banda estrecha 25 produce valores para inclinación espectral y parámetros de modo de voz para cada trama. La inclinación espectral se refiere a la forma de la envolvente espectral sobre la banda de paso y se representa normalmente por el primer coeficiente de reflexión cuantificado. Para la mayoría de los sonidos vocalizados, la energía espectral disminuye con la aumento de la frecuencia, de tal manera que el primer coeficiente de reflexión es negativo y puede aproximarse a -1. La mayoría de los sonidos no vocalizados tienen un espectro que es plano, de tal manera que el primer 30 coeficiente de reflexión se aproxima a cero, o bien tiene más energía en frecuencias altas, de tal manera que el primer coeficiente de reflexión es positivo y puede aproximarse a +1.

El modo de voz (también denominado modo de vocalización) indica si la trama actual representa habla vocalizada o no vocalizada. Este parámetro puede tener un valor binario basado en una o más medidas de periodicidad (por ejemplo, cruces por cero, NACF, ganancia de altura tonal) y/o actividad de voz para la trama, tal como una relación 35 entre una medida de este tipo y un valor umbral. En otras implementaciones, el parámetro del modo de voz tiene uno o más estados adicionales para indicar modos tales como silencio o ruido de fondo, o una transición entre silencio y habla vocalizada.

El codificador A200 de banda alta está configurado para codificar la señal S30 de banda alta según un modelo de filtro de fuente, basándose la excitación para este filtro en la señal codificada de excitación de banda estrecha. La 40 FIGURA 10 muestra un diagrama de bloques de una implementación A202 del codificador A200 de banda alta que está configurado para producir una corriente de parámetros S60 de codificación de banda alta que incluye parámetros S60a de filtro de banda alta y factores S60b de ganancia de banda alta. El generador A300 de excitación de banda alta deriva una señal S120 de excitación de banda alta de la señal codificada S50 de excitación de banda estrecha. El módulo A210 de análisis produce un conjunto de valores de parámetros que caracterizan la envolvente 45 espectral de la señal S30 de banda alta. En este ejemplo particular, el módulo A210 de análisis está configurado para llevar a cabo el análisis de LPC para producir un conjunto de coeficientes de filtro de LP para cada trama de la señal S30 de banda alta. La transformada 410 de coeficiente de filtro de predicción lineal a LSF transforma el conjunto de coeficientes de filtro de LP en un conjunto correspondiente de LSF. Tal como se observó anteriormente 50 con referencia al módulo 210 de análisis y la transformada 220, el módulo A210 de análisis y/o la transformada 410 pueden estar configurados para usar otros conjuntos de coeficientes (por ejemplo, coeficientes cepstrales) y/o representaciones de coeficientes (por ejemplo, ISP).

El cuantificador 420 está configurado para cuantificar el conjunto de las LSF de banda alta (u otra representación de 55 coeficientes, tal como las ISP), y el codificador A202 de banda alta está configurado para emitir el resultado de esta cuantificación como los parámetros S60a de filtro de banda alta. Normalmente, un cuantificador de este tipo incluye un cuantificador de vector que codifica el vector de entrada como un índice con respecto a una entrada de vector correspondiente en una tabla o código.

El codificador A202 de banda alta también incluye un filtro A220 de síntesis configurado para producir una señal sintetizada S130 de banda alta según la señal S120 de excitación de banda alta y la envolvente espectral codificada (por ejemplo, el conjunto de coeficientes de filtro de LP) producidas por el módulo A210 de análisis. Normalmente, el filtro A220 de síntesis se implementa como un filtro de IIR, aunque también pueden usarse implementaciones de 60

FIR. En un ejemplo particular, se implementa el filtro A220 de síntesis como un filtro autorregresivo lineal de sexto orden.

El calculador A230 del factor de ganancia de banda alta calcula una o más diferencias entre los niveles de la señal S30 de banda alta original y la señal sintetizada S130 de banda alta para especificar una envolvente de ganancia para la trama. El cuantificador 430, que puede implementarse como un cuantificador de vector que codifica el vector de entrada como un índice con respecto a una entrada de vector correspondiente en una tabla o código, cuantifica el valor o los valores que especifican la envolvente de ganancia, y el codificador A202 de banda alta está configurado para emitir el resultado de esta cuantificación como factores S60b de ganancia de banda alta.

En una implementación como la mostrada en la FIGURA 10, se dispone el filtro A220 de síntesis para recibir los coeficientes de filtro procedentes del módulo A210 de análisis. Una implementación alternativa del codificador A202 de banda alta incluye un cuantificador inverso y una transformada inversa configurados para decodificar los coeficientes de filtro de parámetros S60a de filtro de banda alta, y en este caso se dispone el filtro A220 de síntesis para recibir, en cambio, los coeficientes de filtro decodificados. Tal disposición alternativa puede soportar un cálculo más preciso de la envolvente de ganancia por parte del calculador A230 de ganancia de banda alta.

En un ejemplo particular, el módulo A210 de análisis y el calculador A230 de ganancia de banda alta emiten un conjunto de seis LSF y un conjunto de cinco valores de ganancia por trama, respectivamente, de tal manera que puede lograrse un ensanchamiento a banda ancha de la señal S20 de banda estrecha con solo once valores adicionales por trama. El oído tiende a ser menos sensible a errores de frecuencia en frecuencias altas, de tal manera que la codificación de banda alta en un orden de LPC bajo puede producir una señal que tiene una calidad de percepción comparable con respecto a la codificación de banda estrecha en un orden de LPC más alto. Una implementación típica del codificador A200 de banda alta puede estar configurada para emitir de 8 a 12 bits por trama para la reconstrucción de alta calidad de la envolvente espectral y otros 8 a 12 bits por trama para la reconstrucción de alta calidad de la envolvente temporal. En otro ejemplo particular, el módulo A210 de análisis emite un conjunto de ocho LSF por trama.

Algunas implementaciones del codificador A200 de banda alta están configuradas para producir la señal S120 de excitación de banda alta generando una señal de ruido aleatoria que tiene componentes de frecuencia de banda alta y modulando en amplitud la señal de ruido según la envolvente de dominio temporal de la señal S20 de banda estrecha, la señal S80 de excitación de banda estrecha o la señal S30 de banda alta. Sin embargo, aunque un procedimiento a base de ruido de este tipo puede producir resultados adecuados para sonidos no vocalizados, puede ser no deseable para sonidos vocalizados, cuyos restos son habitualmente armónicos y, por consiguiente, tienen cierta estructura periódica.

El generador A300 de excitación de banda alta está configurado para generar la señal S120 de excitación de banda alta ensanchando el espectro de la señal S80 de excitación de banda alta en el intervalo de frecuencia de banda alta. La FIGURA 11 muestra un diagrama de bloques de una implementación A302 del generador A300 de excitación de banda alta. El cuantificador inverso 450 está configurado para descuantificar la señal codificada S50 de excitación de banda estrecha para producir la señal S80 de excitación de banda estrecha. El ensanchador A400 de espectro está configurado para producir una señal S160 ensanchada armónicamente basándose en la señal S80 de excitación de banda estrecha. El combinador 470 está configurado para combinar una señal de ruido aleatorio generada por el generador 480 de ruido y una envolvente de dominio temporal calculada por el calculador 460 de la envolvente para producir una señal modulada S170 de ruido. El combinador 490 está configurado para mezclar la señal S160 ensanchada armónicamente y la señal modulada S170 de ruido para producir la señal S120 de excitación de banda alta.

En un ejemplo, el ensanchador A400 de espectro está configurado para llevar a cabo una operación de plegado espectral (también denominada reflejo) en la señal S80 de excitación de banda estrecha para producir la señal S160 ensanchada armónicamente. Puede llevarse a cabo el plegado espectral rellenando con ceros la señal S80 de excitación y después aplicando un filtro de paso alto para conservar el solapamiento. En otro ejemplo, el ensanchador A400 de espectro está configurado para producir la señal S160 ensanchada armónicamente trasladando espectralmente la señal S80 de excitación de banda estrecha hacia la banda alta (por ejemplo, por medio de sobremuestreo seguido por multiplicación por una señal de coseno de frecuencia constante).

Los procedimientos de plegado y traslación espectral pueden producir señales ensanchadas espectralmente cuya estructura armónica sea discontinua con la estructura armónica original de la señal S80 de excitación de banda estrecha en fase y/o frecuencia. Por ejemplo, tales procedimientos pueden producir señales que tienen picos que no están ubicados generalmente en múltiplos de la frecuencia fundamental, lo que puede provocar artefactos con sonido metálico en la señal de voz reconstruida. Estos procedimientos también tienden a producir armónicos de alta frecuencia que tienen características tonales fuertes de manera no natural. Además, debido a que una señal de RTPC puede muestrearse a 8 kHz pero limitarse en banda a no más de 3400 Hz, el espectro superior de la señal S80 de excitación de banda estrecha puede contener poco o nada de energía, de tal manera que una señal ensanchada generada según una operación de plegado espectral o traslación espectral puede tener un agujero espectral por encima de 3400 Hz.

Otros procedimientos de generación de la señal S160 ensanchada armónicamente incluyen identificar una o más frecuencias fundamentales de la señal S80 de excitación de banda estrecha y generar tonos armónicos según esa información. Por ejemplo, la estructura armónica de una señal de excitación puede caracterizarse por la frecuencia fundamental junto con información de amplitud y de fase. Otra implementación del generador A300 de excitación de banda alta genera una señal S160 ensanchada armónicamente basándose en la frecuencia y la amplitud fundamentales (tal como se indica, por ejemplo, por el retraso de altura tonal y la ganancia de altura tonal). Sin embargo, a no ser que la señal ensanchada armónicamente sea coherente en fase con la señal S80 de excitación de banda estrecha, la calidad del habla decodificada resultante puede no ser aceptable.

5 Puede usarse una función no lineal para crear una señal de excitación de banda alta que sea coherente en fase con la excitación de banda estrecha y preserve la estructura armónica sin discontinuidad de fase. Una función no lineal también puede proporcionar un nivel de ruido aumentado entre armónicos de alta frecuencia, lo que tiende a sonar más natural que los armónicos de alta frecuencia tonal producidos mediante procedimientos tales como el plegado espectral y la traslación espectral. Las funciones no lineales sin memoria típicas que pueden ser aplicadas por 10 diversas implementaciones del ensanchador A400 de espectro incluyen la función de valor absoluto (también denominada rectificación de onda completa), la rectificación de onda media, la elevación al cuadrado, la elevación al cubo y el recorte. Otras implementaciones del ensanchador A400 de espectro pueden estar configuradas para 15 aplicar una función no lineal que tenga memoria.

10 La FIGURA 12 es un diagrama de bloques de una implementación A402 del ensanchador A400 de espectro que 20 está configurado para aplicar una función no lineal para ensanchar el espectro de la señal S80 de excitación de banda estrecha. El sobremuestreador 510 está configurado para sobremuestrear la señal S80 de excitación de banda estrecha. Puede resultar deseable sobremuestrear la señal lo suficiente para minimizar el solapamiento tras la aplicación de la función no lineal. En un ejemplo particular, el sobremuestreador 510 sobremuestrea la señal en un factor de ocho. El sobremuestreador 510 puede estar configurado para llevar a cabo la operación de sobremuestreo 25 llenando con ceros la señal de entrada y filtrando paso bajo el resultado. El calculador 520 de función no lineal está configurado para aplicar una función no lineal a la señal sobremuestreada. Una ventaja potencial de la función de valor absoluto con respecto a otras funciones no lineales para ensanchamiento espectral, tal como la elevación al cuadrado, es que no se necesita normalización de energía. En algunas implementaciones, puede aplicarse 30 eficazmente la función de valor absoluto eliminando o quitando el bit de signo de cada muestra. El calculador 520 de función no lineal también puede estar configurado para llevar a cabo una distorsión de la amplitud de la señal ensanchada espectralmente o sobremuestreada.

35 El submuestreador 530 está configurado para submuestrear el resultado ensanchado espectralmente de la aplicación de la función no lineal. Puede resultar deseable que el submuestreador 530 lleve a cabo una operación de filtrado de paso de banda para seleccionar una banda de frecuencia deseada de la señal ensanchada espectralmente antes de reducir la tasa de muestreo (por ejemplo, para reducir o evitar el solapamiento o la corrupción por una imagen no deseada). También puede ser deseable que el submuestreador 530 reduzca la tasa de muestreo en más de una etapa.

40 La FIGURA 12a es un diagrama que muestra los espectros de señal en diversos puntos en un ejemplo de una operación de ensanchamiento espectral, siendo la escala de frecuencia la misma en las diversas representaciones gráficas. La representación gráfica (a) muestra el espectro de un ejemplo de la señal S80 de excitación de banda estrecha. La representación gráfica (b) muestra el espectro tras haberse sobremuestreado la señal S80 en un factor 45 de ocho. La representación gráfica (c) muestra un ejemplo del espectro ensanchado tras la aplicación de una función no lineal. La representación gráfica (d) muestra el espectro tras el filtrado de paso bajo. En este ejemplo, la banda de paso se ensancha hasta el límite de frecuencia superior de la señal S30 de banda alta (por ejemplo, 7 kHz u 8 kHz).

45 La representación gráfica (e) muestra el espectro tras una primera etapa de submuestreo, en el que la tasa de muestreo se reduce en un factor de cuatro para obtener una señal de banda ancha. La representación gráfica (f) muestra el espectro tras una operación de filtrado de paso alto para seleccionar la parte de banda alta de la señal ensanchada, y la representación gráfica (g) muestra el espectro tras una segunda etapa de submuestreo, en la que la tasa de muestreo se reduce en un factor de dos. En un ejemplo particular, el submuestreador 530 lleva a cabo el 50 filtrado de paso alto y la segunda etapa de submuestreo dejando pasar la señal de banda ancha a través del filtro 130 de paso alto y el submuestreador 140 del banco A112 de filtros (u otras estructuras o rutinas que tienen la misma respuesta) para producir una señal ensanchada espectralmente que tiene el intervalo de frecuencia y la tasa de muestreo de la señal S30 de banda alta.

55 Tal como puede observarse en la representación gráfica (g), el submuestreo de la señal de paso alto mostrada en la representación gráfica (f) provoca una inversión de su espectro. En este ejemplo, el submuestreador 530 también está configurado para llevar a cabo una operación de cambio espectral en la señal. La representación gráfica (h) muestra un resultado de la aplicación de la operación de inversión espectral, que puede llevarse a cabo multiplicando la señal por la función $e^{j\pi t}$ o la secuencia $(-1)^n$, cuyos valores alternan entre +1 y -1. Una operación de este tipo es equivalente a desplazar el espectro digital de la señal en el dominio frecuencial en una distancia de π . Se hace notar que también puede obtenerse el mismo resultado aplicando las operaciones de submuestreo y de 60 inversión espectral en un orden diferente. También pueden configurarse las operaciones de sobremuestreo y/o

submuestreo para que incluyan un nuevo muestreo para obtener una señal ensanchada espectralmente que tenga la tasa de muestreo de la señal S30 de banda alta (por ejemplo, 7 kHz).

- 5 Tal como se ha hecho notar anteriormente, pueden implementarse bancos A110 y B120 de filtros de tal manera que una o ambas de las señales S20, S30 de banda estrecha y banda alta tengan una forma invertida espectralmente a la salida del banco A110 de filtros, se codifiquen y decodifiquen en la forma invertida espectralmente, y se inviertan espectralmente de nuevo en el banco B120 de filtros antes de que se emitan en la señal S110 de voz de banda ancha. Por supuesto, en tal caso, una operación de inversión espectral como la se muestra en la FIGURA 12a no sería necesaria, puesto que también sería deseable que la señal S120 de excitación de banda alta tuviera una forma invertida espectralmente.
- 10 Pueden configurarse y disponerse de muchas maneras diferentes las diversas tareas de sobremuestreo y submuestreo de una operación de ensanchamiento espectral tal como lleva a cabo el ensanchador A402 de espectro. Por ejemplo, la FIGURA 12b es un diagrama que muestra los espectros de señal en diversos puntos en otro ejemplo de una operación de ensanchamiento espectral, siendo la escala de frecuencia la misma en las diversas representaciones gráficas. La representación gráfica (a) muestra el espectro de un ejemplo de la señal S80 de excitación de banda estrecha. La representación gráfica (b) muestra el espectro tras haberse sobremuestreado la señal S80 en un factor de dos. La representación gráfica (c) muestra un ejemplo del espectro ensanchado tras la aplicación de una función no lineal. En este caso, se acepta el solapamiento que puede producirse en las frecuencias superiores.
- 15 La representación gráfica (d) muestra el espectro tras una operación de inversión espectral. La representación gráfica (e) muestra el espectro tras una única etapa de submuestreo, en la que la tasa de muestreo se reduce en un factor de dos para obtener la señal ensanchada espectralmente deseada. En este ejemplo, la señal está en forma invertida espectralmente y puede usarse en una implementación del codificador A200 de banda alta que procesó la señal S30 de banda alta en una forma de este tipo.
- 20 La señal ensanchada espectralmente producida por el calculador 520 de función no lineal es probable que tenga una caída pronunciada en amplitud a medida que aumenta la frecuencia. El ensanchador espectral A402 incluye un aplanador espectral 540 configurado para llevar a cabo una operación de blanqueo en la señal submuestreada. El aplanador espectral 540 puede estar configurado para llevar a cabo una operación de blanqueo fijo o para llevar a cabo una operación de blanqueo adaptativo. En un ejemplo particular de blanqueo adaptativo, el aplanador espectral 540 incluye un módulo de análisis de LPC configurado para calcular un conjunto de cuatro coeficientes de filtro de la señal submuestreada y un filtro de análisis de cuarto orden configurado para blanquear la señal según esos coeficientes. Otras implementaciones del ensanchador A400 de espectro incluyen configuraciones en las que el aplanador espectral 540 opera en la señal ensanchada espectralmente antes del submuestreador 530.
- 25 Puede implementarse el generador A300 de excitación de banda alta para emitir la señal S160 ensanchada armónicamente como la señal S120 de excitación de banda alta. Sin embargo, en algunos casos, el uso de solo una señal ensanchada armónicamente como la excitación de banda alta puede dar como resultado artefactos auditivos. La estructura armónica de la voz generalmente es menos pronunciada en la banda alta que en la banda baja, y el uso de demasiada estructura armónica en la señal de excitación de banda alta puede dar como resultado un sonido de zumbido. Este artefacto puede ser especialmente perceptible en señales de voz de oradoras.
- 30 Las realizaciones incluyen implementaciones del generador A300 de excitación de banda alta que están configuradas para mezclar la señal S160 ensanchada armónicamente con una señal de ruido. Tal como se muestra en la FIGURA 11, el generador A302 de excitación de banda alta incluye un generador 480 de ruido que está configurado para producir una señal de ruido aleatorio. En un ejemplo, el generador 480 de ruido está configurado para producir una señal blanca de ruido pseudoaleatorio de varianza unitaria, aunque en otras implementaciones la señal de ruido no necesita ser blanca y puede tener una densidad de potencia que varíe con la frecuencia. Puede ser deseable que el generador 480 de ruido esté configurado para emitir la señal de ruido como una función determinista, de tal manera que pueda duplicarse su estado en el decodificador. Por ejemplo, el generador 480 de ruido puede estar configurado para emitir la señal de ruido como una función determinista de información codificada previamente dentro de la misma trama, tal como los parámetros S40 de filtro de banda estrecha y/o la señal codificada S50 de excitación de banda estrecha.
- 35 Antes de mezclarse con la señal S160 ensanchada armónicamente, la señal de ruido aleatorio producida por el generador 480 de ruido puede modularse en amplitud para que tenga una envolvente en el dominio temporal que se aproxime a la distribución de energía en el tiempo de la señal S20 de banda estrecha, la señal S30 de banda alta, la señal S80 de excitación de banda estrecha o la señal S160 ensanchada armónicamente. Tal como se muestra en la FIGURA 11, el generador A302 de excitación de banda alta incluye un combinador 470 configurado para modular en amplitud la señal de ruido producida por el generador 480 de ruido según una envolvente en el dominio temporal calculada por el calculador 460 de envolvente. Por ejemplo, puede implementarse el combinador 470 como un multiplicador dispuesto para ampliar a escala la salida del generador 480 de ruido según la envolvente en el dominio temporal calculada por el calculador 460 de envolvente para producir la señal modulada S170 de ruido.
- 40
- 45
- 50
- 55

En una implementación A304 del generador A302 de excitación de banda alta, tal como se muestra en el diagrama de bloques de la FIGURA 13, se dispone el calculador 460 de envolvente para calcular la envolvente de la señal S160 ensanchada armónicamente. En una implementación A306 del generador A302 de excitación de banda alta, tal como se muestra en el diagrama de bloques de la FIGURA 14, se dispone el calculador 460 de envolvente para

- 5 calcular la envolvente de la señal S80 de excitación de banda estrecha. Pueden configurarse de otro modo otras implementaciones del generador A302 de excitación de banda alta para añadir ruido a la señal S160 ensanchada armónicamente según las ubicaciones de los impulsos de altura tonal de banda estrecha en el tiempo.

El calculador 460 de envolvente puede estar configurado para llevar a cabo un cálculo de envolvente como una tarea que incluye una serie de subtareas. La FIGURA 15 muestra un diagrama de flujo de un ejemplo T100 de una tarea de este tipo. La subtarea T110 calcula el cuadrado de cada muestra de la trama de la señal cuya envolvente debe modelarse (por ejemplo, la señal S80 de excitación de banda estrecha o la señal S160 ensanchada armónicamente) para producir una secuencia de valores al cuadrado. La subtarea T120 lleva a cabo una operación de suavizado en la secuencia de valores al cuadrado. En un ejemplo, la subtarea T120 aplica un filtro de paso bajo IIR de primer orden a la secuencia según la expresión

$$y(n) = ax(n) + (1 - a)y(n-1), \quad (1)$$

- 15 en la que x es la entrada de filtro, y es la salida de filtro, n es un índice en el dominio temporal, y a es un coeficiente de aplanamiento que tiene un valor entre 0,5 y 1. El valor del coeficiente de aplanamiento a puede ser fijo o, en una implementación alternativa, puede ser adaptativo según una indicación de ruido en la señal de entrada, de tal manera que a se aproxima más a 1 en ausencia de ruido y más a 0,5 en presencia de ruido. La subtarea T130 aplica una función de raíz cuadrada a cada muestra de la secuencia aplanada para producir la envolvente en el dominio temporal.

Una implementación de este tipo del calculador 460 de envolvente puede ser configurada para llevar a cabo las diversas subtareas de la tarea T100 en serie y/o en paralelo. En implementaciones adicionales de la tarea T100, la subtarea T110 puede estar precedida por una operación de paso de banda configurada para seleccionar una porción de frecuencia deseada de la señal cuya envolvente debe modelarse, tal como el intervalo de 3-4 kHz.

- 25 El combinador 490 está configurado para mezclar la señal S160 ensanchada armónicamente y la señal modulada S170 de ruido para producir la señal S120 de excitación de banda alta. Pueden configurarse implementaciones del combinador 490, por ejemplo, para calcular la señal S120 de excitación de banda alta como una suma de la señal S160 ensanchada armónicamente y la señal modulada S170 de ruido. Puede configurarse una implementación de este tipo de combinador 490 para calcular la señal S120 de excitación de banda alta como una suma ponderada 30 aplicando un factor de ponderación a la señal S160 ensanchada armónicamente y/o a la señal modulada S170 de ruido antes de la suma. Cada factor de ponderación de este tipo puede calcularse según uno o más criterios y puede ser un valor fijo o, como alternativa, un valor adaptativo que se calcula trama por trama o subtrama por subtrama.

35 La FIGURA 16 muestra un diagrama de bloques de una implementación 492 del combinador 490 que está configurada para calcular la señal S120 de excitación de banda alta como una suma ponderada de la señal S160 ensanchada armónicamente y la señal modulada S170 de ruido. El combinador 492 está configurado para ponderar la señal S160 ensanchada armónicamente según el factor S180 de ponderación de armónicos, para ponderar la señal modulada S170 de ruido según el factor S190 de ponderación de ruidos, y para emitir la señal S120 de excitación de banda alta como una suma de las señales ponderadas. En este ejemplo, el combinador 492 incluye un calculador 550 del factor de ponderación que está configurado para calcular el factor S180 de ponderación de armónicos y el factor S190 de ponderación de ruido.

40 El calculador 550 del factor de ponderación puede estar configurado para calcular los factores S180 y S190 de ponderación según una relación deseada de contenido de armónicos a contenido de ruido en la señal S120 de excitación de banda alta. Por ejemplo, puede ser deseable que el combinador 492 produzca la señal S120 de excitación de banda alta para tener una relación entre energía de armónicos y energía de ruido similar a la de la señal S30 de banda alta. En algunas implementaciones del calculador 550 del factor de ponderación, se calculan los factores S180, S190 de ponderación según uno o más parámetros relativos a una periodicidad de la señal S20 de banda estrecha o de la señal residual de banda estrecha, tal como ganancia de altura tonal y/o modo de voz. Puede configurarse una implementación de este tipo del calculador 550 del factor de ponderación para asignar un valor al factor S180 de ponderación de armónicos que sea proporcional a la ganancia de altura tonal, por ejemplo, y/o para asignar un valor superior al factor S190 de ponderación de ruidos para señales de voz no vocales que para señales de habla vocales.

45 En otras implementaciones, el calculador 550 del factor de ponderación está configurado para calcular valores para el factor S180 de ponderación de armónicos y/o el factor S190 de ponderación de ruidos según una medida de periodicidad de la señal S30 de banda alta. En un ejemplo de este tipo, el calculador 550 del factor de ponderación calcula el factor S180 de ponderación de armónicos como el valor máximo del coeficiente de autocorrelación de la señal S30 de banda alta para la trama o la subtrama actual, llevándose a cabo la autocorrelación en un intervalo de

50 búsqueda que incluye un retardo de un retraso de altura tonal y no incluye un retardo de muestras cero. La FIGURA 55 calcula el factor S180 de ponderación de armónicos como el valor máximo del coeficiente de autocorrelación de la señal S30 de banda alta para la trama o la subtrama actual, llevándose a cabo la autocorrelación en un intervalo de búsqueda que incluye un retardo de un retraso de altura tonal y no incluye un retardo de muestras cero. La FIGURA

17 muestra un ejemplo de un intervalo de búsqueda de este tipo de muestras de longitud n que se centra sobre un retardo de un retraso de altura tonal y tiene un ancho no mayor que un retraso de altura tonal.

La FIGURA 17 también muestra un ejemplo de otro enfoque en el que el calculador 550 del factor de ponderación calcula una medida de periodicidad de la señal S30 de banda alta en varias etapas. En una primera etapa, se divide

5 la trama actual en una serie de subtramas, y se identifica por separado para cada subtrama el retardo para el que el coeficiente de autocorrelación es máximo. Tal como se ha mencionado anteriormente, la autocorrelación se lleva a cabo en un intervalo de búsqueda que incluye un retraso de un retraso de altura tonal y no incluye retardo de cero muestras.

10 En una segunda etapa, se construye una trama retardada aplicando el retraso identificado correspondiente a cada subtrama, concatenando las subtramas resultantes para construir una trama retardada óptimamente, y calculando el factor S180 de ponderación de armónicos como el coeficiente de correlación entre la trama original y la trama retardada óptimamente. En otra alternativa, el calculador 550 del factor de ponderación calcula el factor S180 de ponderación de armónicos como un promedio de los coeficientes de autocorrelación máximos obtenidos en la 15 primera etapa para cada subtrama. También pueden configurarse implementaciones del calculador 550 del factor de ponderación para ajustar a escala el coeficiente de correlación, y/o combinarlo con otro valor, para calcular el valor para el factor S180 de ponderación de armónicos.

20 Puede resultar deseable que el calculador 550 del factor de ponderación calcule una medida de periodicidad de la señal S30 de banda alta solo en los casos en los que se indique de otro modo una presencia de periodicidad en la trama. Por ejemplo, el calculador 550 del factor de ponderación puede estar configurado para calcular una medida de periodicidad de la señal S30 de banda alta según una relación entre otro indicador de periodicidad de la trama actual, tal como la ganancia de altura tonal, y un valor umbral. En un ejemplo, el calculador 550 del factor de ponderación está configurado para llevar a cabo una operación de autocorrelación en la señal S30 de banda alta solo si la ganancia de altura tonal de la trama (por ejemplo, la ganancia del código adaptativo del resto de banda estrecha) tiene un valor de más de 0,5 (como alternativa, al menos de 0,5). En otro ejemplo, el calculador 550 del factor de ponderación está configurado para llevar a cabo una operación de autocorrelación en la señal S30 de banda alta solo para las tramas que tengan estados particulares del modo de voz (por ejemplo, solo para señales vocales). En tales casos, el calculador 550 del factor de ponderación puede estar configurado para asignar un factor de ponderación por defecto para tramas que tengan otros estados del modo de voz y/o valores inferiores de ganancia de altura tonal.

25 30 Las realizaciones incluyen otras implementaciones del calculador 550 del factor de ponderación que están configuradas para calcular factores de ponderación según características diferentes de la periodicidad o además de ella. Por ejemplo, una implementación de este tipo puede estar configurada para asignar un valor superior al factor S190 de ganancia de ruido para señales de voz que tengan un retraso de altura tonal grande que para señales de voz que tengan un retraso de altura tonal pequeño. Otra implementación de este tipo del calculador 550 del factor de ponderación está configurada para determinar una medida de armonicidad de la señal S10 de voz de banda ancha, o de la señal S30 de banda alta, según una medida de la energía de la señal en múltiplos de la frecuencia fundamental en relación con la energía de la señal en otros componentes de frecuencia.

35 40 Algunas implementaciones del codificador A100 de voz de banda ancha están configuradas para emitir una indicación de periodicidad o armonicidad (por ejemplo una bandera de un bit que indique si la trama es armónica o no armónica) basándose en la ganancia de altura tonal y/u otra medida de periodicidad o armonicidad, tal como se describe en el presente documento. En un ejemplo, un decodificador B100 de voz de banda ancha correspondiente usa esta indicación para configurar una operación tal como el cálculo del factor de ponderación. En otro ejemplo, se usa una indicación de este tipo en el codificador y/o el decodificador en el cálculo de un valor para un parámetro de modo de voz.

45 50 Puede resultar deseable que el generador A302 de excitación de banda alta genere la señal S120 de excitación de banda alta de tal manera que la energía de la señal de excitación no se vea afectada sustancialmente por los valores particulares de los factores S180 y S190 de ponderación. En tal caso, el calculador 550 del factor de ponderación puede configurarse para que calcule un valor para el factor S180 de ponderación de armónicos o para el factor S190 de ponderación de ruido (o para recibir un valor de este tipo del almacenamiento u otro elemento del codificador A200 de banda alta) y para que derive un valor para el otro factor de ponderación según una expresión tal como

$$(W_{\text{armónicos}})^2 + (W_{\text{ruido}})^2 = 1, \quad (2)$$

55 en la que $W_{\text{armónicos}}$ indica el factor S180 de ponderación de armónicos y W_{ruido} indica el factor S190 de ponderación de ruido. Como alternativa, el calculador 550 del factor de ponderación puede configurarse para seleccionar, según un valor de una medida de periodicidad para la trama o la subtrama actuales, un par correspondiente entre una pluralidad de pares de factores S180, S190 de ponderación, calculándose los pares previamente para satisfacer una relación de energía constante tal como la expresión (2). Para una implementación del calculador 550 del factor de ponderación en la que se observa la expresión (2), los valores típicos para el factor

5 S180 de ponderación de armónicos oscilan entre aproximadamente 0,7 y aproximadamente 1,0, y los valores típicos para el factor S190 de ponderación de ruidos oscilan entre aproximadamente 0,1 y aproximadamente 0,7. Otras implementaciones del calculador 550 del factor de ponderación pueden configurarse para operar según una versión de la expresión (2) que se modifica según una ponderación de referencia deseada entre la señal S160 ensanchada armónicamente y la señal modulada S170 de ruido.

10 Pueden producirse artefactos en una señal de voz sintetizada cuando se ha usado un código disperso (uno cuyas entradas son en su mayoría valores cero) para calcular la representación cuantificada del resto. La dispersión del código se produce especialmente cuando se codifica la señal de banda estrecha a una tasa de transmisión de bits baja. Los artefactos provocados por la dispersión del código son normalmente casi periódicos en el tiempo y se producen en su mayoría por encima de 3 kHz. Debido a que el oído humano tiene una mejor resolución temporal a 15 frecuencias superiores, estos artefactos pueden ser más perceptibles en la banda alta.

20 Las realizaciones incluyen implementaciones del generador A300 de excitación de banda alta que están configuradas para llevar a cabo un filtrado de antidispersión. La FIGURA 18 muestra un diagrama de bloques de una implementación A312 de un generador A302 de excitación de banda alta que incluye un filtro antidispersión 600 dispuesto para filtrar la señal de excitación de banda estrecha descuantificada producida por el cuantificador 450 inverso. La FIGURA 19 muestra un diagrama de bloques de una implementación A314 del generador A302 de excitación de banda alta que incluye un filtro antidispersión 600 dispuesto para filtrar la señal ensanchada espectralmente producida por el ensanchador A400 de espectro. La FIGURA 20 muestra un diagrama de bloques de una implementación A316 del generador A302 de excitación de banda alta que incluye un filtro antidispersión 600 dispuesto para filtrar la salida del combinador 490 para producir la señal S120 de excitación de banda alta. Por supuesto, se contemplan implementaciones del generador A300 de excitación de banda alta que combinan las características de cualquiera de las implementaciones A304 y A306 con las características de cualquiera de las implementaciones A312, A314 y A316 y son dadas a conocer expresamente por el presente documento. El filtro antidispersión 600 también puede disponerse dentro del ensanchador A400 de espectro; por ejemplo, tras 25 cualquiera de los elementos 510, 520, 530 y 540 en el ensanchador A402 de espectro. Se hace notar expresamente que también puede usarse el filtro antidispersión 600 con implementaciones del ensanchador A400 de espectro que lleven a cabo plegado espectral, traslación espectral o ensanchamiento armónico.

30 El filtro antidispersión 600 puede configurarse para alterar la fase de su señal de entrada. Por ejemplo, puede resultar deseable que el filtro antidispersión 600 esté configurado y dispuesto de tal manera que la fase de la señal S120 de excitación de banda alta sea aleatoria, o, en todo caso, para que se distribuya más uniformemente en el tiempo. También puede ser deseable que la respuesta del filtro antidispersión 600 sea espectralmente plana, de tal manera que no se cambie de manera apreciable el espectro de magnitud de la señal filtrada. En un ejemplo, el filtro antidispersión 600 se implementa como un filtro de paso total que tiene una función de transferencia según la siguiente expresión:

$$H(z) = \frac{-0,7 + z^4}{1 - 0,7z^4} \cdot \frac{0,6 + z^6}{1 + 0,6z^6}. \quad (3)$$

35 Un efecto de un filtro de este tipo puede ser esparcir la energía de la señal de entrada de tal manera que ya no se concentre en solo algunas muestras.

40 Los artefactos provocados por la dispersión del código son habitualmente más perceptibles para señales de tipo ruido, en las que el resto incluye menos información de altura tonal, y también para el habla en ruido de fondo. La dispersión provoca normalmente menos artefactos en casos en los que la excitación tiene una estructura a largo plazo y, de hecho, la modificación de fase puede provocar ruido en señales vocales. Por lo tanto, puede resultar deseable configurar el filtro antidispersión 600 para filtrar señales no vocales y para dejar pasar al menos algunas señales vocales sin alteración. Las señales no vocales se caracterizan por una baja ganancia de altura tonal (por ejemplo, ganancia de código adaptativa de banda estrecha cuantificada) y una inclinación espectral (por ejemplo primer coeficiente de reflexión cuantificado) que es cercana a cero o positiva, lo que indica una envolvente espectral plana o inclinada hacia arriba con una frecuencia creciente. Las implementaciones típicas del filtro antidispersión 600 están configuradas para filtrar sonidos no vocalizados (por ejemplo, tal como se indica por medio del valor de la inclinación espectral), para filtrar sonidos vocalizados cuando la ganancia de altura tonal está por debajo de un valor umbral (como alternativa, no mayor que el valor umbral) y, si no, para dejar pasar la señal sin alteración.

50 Otras implementaciones del filtro antidispersión 600 incluyen dos o más filtros que están configurados para tener diferentes ángulos de modificación de fase máximos (por ejemplo, hasta 180 grados). En tal caso, el filtro antidispersión 600 puede configurarse para seleccionar entre estos filtros de componentes según un valor de la ganancia de altura tonal (por ejemplo, la ganancia de código adaptativo cuantificada o de LTP), de tal manera que se use un ángulo de modificación de fase máximo mayor para tramas que tengan valores de ganancia de altura tonal inferiores. Una implementación del filtro antidispersión 600 también puede incluir diferentes filtros de componentes 55 que estén configurados para modificar la fase en una mayor o menor medida del espectro de frecuencia, de tal

manera que se use un filtro configurado para modificar la fase por un intervalo de frecuencia más amplio de la señal de entrada para tramas que tengan valores de ganancia de altura tonal inferiores.

Para una reproducción precisa de la señal de voz codificada, puede ser deseable que la relación entre los niveles de las porciones de banda alta y banda estrecha de la señal sintetizada S100 de voz de banda ancha sea similar a la de la señal original S10 de voz de banda ancha. Además de una envolvente espectral, tal como se representa por los parámetros S60a de codificación de banda alta, el codificador A200 de banda alta puede configurarse para

5 caracterizar la señal S30 de banda alta especificando una envolvente temporal o de ganancia. Tal como se muestra en la FIGURA 10, el codificador A202 de banda alta incluye un calculador A230 del factor de ganancia de banda alta que está configurado y dispuesto para calcular uno o más factores de ganancia según una relación entre la señal 10 S30 de banda alta y la señal sintetizada S130 de banda alta, tal como una diferencia o relación entre las energías de las dos señales en una trama o alguna porción de la misma. En otras implementaciones del codificador A202 de banda alta, el calculador A230 de ganancia de banda alta puede estar configurado de la misma manera, pero estar 15 dispuesto en cambio para calcular la envolvente de ganancia según una relación de este tipo que varía en el tiempo entre la señal S30 de banda alta y la señal S80 de excitación de banda estrecha o la señal S120 de excitación de banda alta.

Es probable que las envolventes temporales de la señal S80 de excitación de banda estrecha y la señal S30 de banda alta sean similares. Por tanto, la codificación de una envolvente de ganancia que se base en una relación 20 entre la señal S30 de banda alta y la señal S80 de excitación de banda estrecha (o una señal derivada de la misma, tal como la señal S120 de excitación de banda alta o la señal sintetizada S130 de banda alta) será en general más eficaz que la codificación de una envolvente de ganancia basada solo en la señal S30 de banda alta. En una 25 implementación típica, el codificador A202 de banda alta está configurado para emitir un índice cuantificado de ocho a doce bits que especifica cinco factores de ganancia para cada trama.

El calculador A230 del factor de ganancia de banda alta puede configurarse para llevar a cabo el cálculo del factor 30 de ganancia como una tarea que incluya una o más series de subtareas. La FIGURA 21 muestra un diagrama de flujo de un ejemplo T200 de una tarea de este tipo que calcula un valor de ganancia para una subtrama correspondiente según las energías relativas de la señal S30 de banda alta y la señal sintetizada S130 de banda alta. Las tareas 220a y 220b calculan las energías de las subtramas correspondientes de las señales respectivas. Por ejemplo, las tareas 220a y 220b pueden configurarse para calcular la energía como una suma de los cuadrados 35 de las muestras de la subtrama respectiva. La tarea T230 calcula un factor de ganancia para la subtrama como la raíz cuadrada de la relación de esas energías. En este ejemplo, la tarea T230 calcula el factor de ganancia como la raíz cuadrada de la relación de la energía de la señal S30 de banda alta a la energía de la señal sintetizada S130 de banda alta en la subtrama.

Puede ser deseable que el calculador A230 del factor de ganancia de banda alta se configure para calcular las 40 energías de subtrama según una función ventana. La FIGURA 22 muestra un diagrama de flujo de una implementación T210 de este tipo de una tarea T200 de cálculo del factor de ganancia. La tarea T215a aplica una función ventana a la señal S30 de banda alta, y la tarea T215b aplica la misma función ventana a la señal sintetizada S130 de banda alta. Las implementaciones 222a y 222b de las tareas 220a y 220b calculan las energías de las ventanas respectivas, y la tarea T230 calcula un factor de ganancia para la subtrama como la raíz cuadrada de la relación de las energías.

45 Puede ser deseable aplicar una función ventana que superponga subtramas adyacentes. Por ejemplo, una función ventana que produzca factores de ganancia que puedan ser aplicados en un modo de adición de superposición puede contribuir a reducir o evitar una discontinuidad entre subtramas. En un ejemplo, el calculador A230 del factor de ganancia de banda alta está configurado para aplicar una función ventana trapezoidal, tal como se muestra en la FIGURA 23a, en la que la ventana superpone cada una de las dos subtramas adyacentes en un milisegundo. La FIGURA 23b muestra una aplicación de esta función ventana a cada una de las cinco subtramas de una trama de 20 milisegundos. Otras implementaciones del calculador A230 del factor de ganancia de banda alta pueden configurarse para aplicar funciones ventana que tengan diferentes períodos de superposición y/o diferentes formas 50 de ventana (por ejemplo, rectangular, de Hamming) que pueden ser simétricas o asimétricas. También es posible que una implementación del calculador A230 del factor de ganancia de banda alta se configure para aplicar diferentes funciones ventana a diferentes subtramas dentro de una trama y/o que una trama incluya subtramas de diferentes longitudes.

55 Sin limitación, se presentan los siguientes valores como ejemplos para implementaciones particulares. Se supone una trama de 20 ms para estos casos, aunque puede usarse cualquier otra duración. Para una señal de banda alta muestreada a 7 kHz, cada trama tiene 140 muestras. Si se divide una trama de este tipo en cinco subtramas de igual longitud, cada subtrama tendrá 28 muestras, y la ventana, tal como se muestra en la FIGURA 23a, tendrá un ancho de 42 muestras. Para una señal de banda alta muestreada a 8 kHz, cada trama tiene 160 muestras. Si se divide una 60 trama de este tipo en cinco subtramas de igual longitud, cada subtrama tendrá 32 muestras, y la ventana, tal como se muestra en la FIGURA 23a, tendrá un ancho de 48 muestras. En otras implementaciones, pueden usarse subtramas de cualquier ancho, e incluso es posible que una implementación del calculador A230 de ganancia de banda alta se configure para producir un factor de ganancia diferente para cada muestra de una trama.

La FIGURA 24 muestra un diagrama de bloques de una implementación B202 del decodificador B200 de banda alta. El decodificador B202 de banda alta incluye un generador B300 de excitación de banda alta que está configurado para producir la señal S120 de excitación de banda alta basándose en la señal S80 de excitación de banda estrecha. Dependiendo de las opciones particulares de diseño del sistema, el generador B300 de excitación de banda alta puede implementarse según cualquiera de las implementaciones del generador A300 de excitación de banda alta, tal como se describe en el presente documento.

Normalmente es deseable implementar el generador B300 de excitación de banda alta para que tenga la misma respuesta que el generador de excitación de banda alta del codificador de banda alta del sistema de codificación particular. Sin embargo, dado que el decodificador B110 de banda estrecha llevará a cabo normalmente la descuantificación de la señal codificada S50 de excitación de banda estrecha, en la mayoría de los casos el generador B300 de excitación de banda alta puede ser implementado para recibir la señal S80 de excitación de banda estrecha procedente del decodificador B110 de banda estrecha y no es necesario que incluya un cuantificador inverso configurado para descuantificar la señal codificada S50 de excitación de banda estrecha. También es posible que el decodificador B110 de banda estrecha se implemente para incluir una instancia de filtro antidispersión 600 dispuesto para filtrar la señal de excitación de banda estrecha descuantificada antes de su entrada en un filtro de síntesis de banda estrecha tal como el filtro 330.

El cuantificador 560 inverso está configurado para descuantificar los parámetros S60a de filtro de banda alta (en este ejemplo, a un conjunto de LSF), y la transformada 570 del coeficiente de filtro de LSF a LP está configurada para transformar las LSF en un conjunto de coeficientes de filtro (por ejemplo, tal como se ha descrito anteriormente con referencia al cuantificador inverso 240 y la transformada 250 del codificador A122 de banda estrecha). En otras

implementaciones, tal como se mencionó en lo que antecede, pueden usarse diferentes conjuntos de coeficientes (por ejemplo, coeficientes cepstrales) y/o representaciones de coeficiente (por ejemplo, ISP). El filtro B200 de síntesis de banda alta está configurado para producir una señal sintetizada de banda alta según la señal S120 de excitación de banda alta y el conjunto de coeficientes de filtro. Para un sistema en el que el codificador de banda alta incluye un filtro de síntesis (por ejemplo, como en el ejemplo del codificador A202 descrito anteriormente), puede ser deseable implementar el filtro B200 de síntesis de banda alta para que tenga la misma respuesta (por ejemplo, la misma función de transferencia) que ese filtro de síntesis.

El decodificador B202 de banda alta también incluye un cuantificador 580 inverso configurado para descuantificar los factores S60b de ganancia de banda alta, y un elemento 590 de control de ganancia (por ejemplo, un multiplicador o un amplificador) configurado y dispuesto para aplicar los factores de ganancia descuantificados a la señal sintetizada de banda alta para producir la señal S100 de banda alta. Para un caso en el que se especifica la envolvente de ganancia de una trama en más de un factor de ganancia, el elemento 590 de control de ganancia puede incluir una lógica configurada para aplicar los factores de ganancia a las subtramas respectivas, posiblemente según una función ventana que puede ser la misma o una función ventana diferente tal como la aplica un calculador de ganancia (por ejemplo, el calculador A230 de ganancia de banda alta) del correspondiente codificador de banda alta. En otras implementaciones del decodificador B202 de banda alta, el elemento 590 de control de ganancia está configurado de manera similar pero está dispuesto en cambio para aplicar los factores de ganancia descuantificados a la señal S80 de excitación de banda estrecha o a la señal S120 de excitación de banda alta.

Tal como se mencionó en lo que antecede, puede ser deseable obtener el mismo estado en el codificador de banda alta y el decodificador de banda alta (por ejemplo, usando valores descuantificados durante la codificación). Por ello, puede ser deseable en un sistema de codificación según una implementación de este tipo garantizar el mismo estado para correspondientes los generadores de ruido en los generadores A300 y B300 de excitación de banda alta. Por ejemplo, los generadores A300 y B300 de excitación de banda alta de una implementación de este tipo pueden configurarse de tal manera que el estado del generador de ruido sea una función determinista de información ya codificada dentro de la misma trama (por ejemplo, los parámetros S40 de filtro de banda estrecha o una porción de los mismos y/o la señal codificada S50 de excitación de banda estrecha o una porción de la misma).

Uno o más de los cuantificadores de los elementos descritos en el presente documento (por ejemplo, los cuantificadores 230, 420 o 430) pueden configurarse para llevar a cabo una cuantificación clasificada de vectores. Por ejemplo, puede configurarse un cuantificador de este tipo para seleccionar uno de un conjunto de códigos basándose en información que ya ha sido codificada dentro de la misma trama en el canal de banda estrecha y/o en el canal de banda alta. Una técnica de este tipo proporciona normalmente un aumento de la eficacia de codificación a expensas de un almacenamiento de códigos adicional.

Tal como se trató anteriormente con referencia, por ejemplo, a las FIGURAS 8 y 9, una cantidad considerable de la estructura periódica puede permanecer en la señal residual tras la eliminación de la envolvente espectral tosca de la señal S20 de voz de banda estrecha. Por ejemplo, la señal residual puede contener una secuencia de impulsos o picos aproximadamente periódicos en el tiempo. Es especialmente probable que tal estructura, que se relaciona normalmente con la altura tonal, se produzca en señales de habla vocal. El cálculo de una representación cuantificada de la señal residual de banda estrecha puede incluir la codificación de esta estructura de altura tonal según un modelo de periodicidad a largo plazo, tal como la representada, por ejemplo, por uno o más códigos.

La estructura de altura tonal de una señal residual real puede no coincidir exactamente con el modelo de periodicidad. Por ejemplo, la señal residual puede incluir pequeñas fluctuaciones en la regularidad de las

ubicaciones de los impulsos de altura tonal, de tal manera que las distancias entre impulsos de altura tonal sucesivos en una trama no sean exactamente iguales y la estructura no sea muy regular. Estas irregularidades tienden a reducir la eficacia de la codificación.

5 Algunas implementaciones del codificador A120 de banda estrecha están configuradas para llevar a cabo una regularización de la estructura de altura tonal aplicando una distorsión temporal adaptativa al resto antes o durante la cuantificación, o incluyendo de otro modo una distorsión temporal adaptativa en la señal codificada de excitación. Por ejemplo, un codificador de este tipo puede configurarse para seleccionar o calcular de otro modo un grado de distorsión temporal (por ejemplo, según uno o más criterios de ponderación perceptual y/o minimización de errores), de tal manera que la señal de excitación resultante se ajuste óptimamente al modelo de periodicidad a largo plazo.

10 La regularización de la estructura de altura tonal se lleva a cabo mediante un subconjunto de codificadores CELP denominados codificadores de predicción lineal excitada por códigos de relajación (RCELP).

Un codificador RCELP está configurado normalmente para efectuar la distorsión temporal como un desplazamiento temporal adaptativo. Este desplazamiento temporal puede ser un retardo que oscile entre unos pocos milisegundos negativos y unos pocos milisegundos positivos, y habitualmente se varía suavemente para evitar discontinuidades audibles. En algunas implementaciones, un codificador de este tipo está configurado para aplicar la regularización por partes, en el que cada trama se distorsiona en un desplazamiento temporal fijo correspondiente. En otras implementaciones, el codificador está configurado para aplicar la regularización como una función de distorsión continua, de tal manera que una trama o una subtrama sean distorsionadas según un contorno de altura tonal (también denominado trayectoria de altura tonal). En algunos casos (por ejemplo, tal como se describe en la publicación de solicitud de patente estadounidense 2004/0098255), el codificador está configurado para incluir una distorsión temporal en la señal codificada de excitación aplicando el desplazamiento a una señal de entrada ponderada perceptualmente que es usada para calcular la señal codificada de excitación.

25 El codificador calcula una señal codificada de excitación que se regulariza y cuantifica, y el decodificador descuantifica la señal codificada de excitación para obtener una señal de excitación que es usada para sintetizar la señal de voz decodificada. Por ello, la señal de salida decodificada presenta el mismo retardo variable que se incluyó en la señal codificada de excitación mediante la regularización. Normalmente, no se transmite al decodificador ninguna información que especifique las cantidades de regularización.

30 La regularización tiende a hacer que la señal residual sea más fácil de codificar, lo que mejora la ganancia de codificación del predictor a largo plazo y, por ello, potencia la eficacia de codificación general, generalmente sin generar artefactos. Puede ser deseable llevar a cabo la regularización solo en tramas que sean vocalizadas. Por ejemplo, el codificador A124 de banda estrecha puede configurarse para desplazar solo las tramas o las subtramas que tengan una estructura a largo plazo, tales como las señales vocales. Incluso puede ser deseable llevar a cabo la regularización solo en subtramas que incluyan energía de impulso de altura tonal. Se describen diversas implementaciones de codificación por RCELP en las patentes estadounidenses n°s 5.704.003 (Kleijn et al.) y 35 6.879.955 (Rao) y en la publicación de solicitud de patente estadounidense 2004/0098255 (Kovesi et al.). Las implementaciones existentes de los codificadores RCELP incluyen el códec de tasa variable mejorado (EVRC), tal como se describe en la norma IS-127 de la Asociación de la Industria de las Telecomunicaciones (TIA), y el vocodificador en modo seleccionable (SMV) del Proyecto 2 de Asociación de Tercera Generación (3GPP2).

40 Desgraciadamente, la regularización puede provocar problemas para un codificador de voz de banda ancha en el que la excitación de banda alta se derive de la señal codificada de excitación de banda estrecha (tal como un sistema que incluya el codificador A100 de voz de banda ancha y el decodificador B100 de voz de banda ancha). Debido a su derivación de una señal con distorsión temporal, la señal de excitación de banda alta tendrá generalmente un perfil temporal que es diferente del de la señal original de voz de banda alta. En otras palabras, la señal de excitación de banda alta ya no será síncrona con la señal original de voz de banda alta.

45 Un desalineamiento en el tiempo entre la señal de excitación de banda alta alineada y la señal original de voz de banda alta puede provocar varios problemas. Por ejemplo, puede ser que la señal distorsionada de excitación de banda alta ya no proporcione una excitación fuente adecuada para un filtro de síntesis que esté configurado según los parámetros de filtro extraídos de la señal original de voz de banda alta. En consecuencia, la señal sintetizada de banda alta puede contener artefactos audibles que reduzcan la calidad percibida de la señal decodificada de voz de banda ancha.

50 El desalineamiento en el tiempo también puede provocar ineficiencias en la codificación de la envolvente de ganancia. Tal como se mencionó anteriormente, es probable que exista una correlación entre las envolventes temporales de la señal S80 de excitación de banda estrecha y la señal S30 de banda alta. Codificando la envolvente de ganancia de la señal de banda alta según una relación entre estas dos envolventes temporales, puede observarse un aumento en la eficacia de codificación en comparación con la codificación de la envolvente de ganancia directamente. Sin embargo, cuando se regulariza la señal codificada de excitación de banda estrecha, puede debilitarse esta correlación. El desalineamiento en el tiempo entre la señal S80 de excitación de banda estrecha y la señal S30 de banda alta puede provocar que aparezcan fluctuaciones en los factores S60b de ganancia de banda alta y puede caer la eficacia de codificación.

Las realizaciones incluyen procedimientos de codificación de voz de banda ancha que llevan a cabo una distorsión temporal de una señal de voz de banda alta según una distorsión temporal incluido en una señal codificada de excitación de banda estrecha correspondiente. Las ventajas potenciales de tales procedimientos incluyen mejorar la calidad de una señal decodificada de voz de banda ancha y/o mejorar la eficacia de codificación de una envolvente de ganancia de banda alta.

5 La FIGURA 25 muestra un diagrama de bloques de una implementación AD10 del codificador A100 de voz de banda ancha. El codificador AD10 incluye una implementación A124 del codificador A120 de banda estrecha que está configurada para llevar a cabo una regularización durante el cálculo de la señal codificada S50 de excitación de banda estrecha. Por ejemplo, el codificador A124 de banda estrecha puede configurarse según una o más de las 10 implementaciones de RCELP expuestas en lo que antecede.

10 El codificador A124 de banda estrecha también está configurado para emitir una señal SD10 de datos de regularización que especifica el grado de distorsión temporal aplicada. Para diversos casos en los que el codificador A124 de banda estrecha está configurado para aplicar un desplazamiento temporal fijo a cada trama o subtrama, la señal SD10 de datos de regularización puede incluir una serie de valores que indiquen cada cantidad de 15 desplazamiento temporal como un valor entero o no entero en términos de muestras, milisegundos o cualquier otro incremento de tiempo. Para un caso en el que el codificador A124 de banda estrecha está configurado para modificar de otro modo la escala de tiempo de una trama u otra secuencia de muestras (por ejemplo, comprimiendo una porción y expandiendo otra porción), la señal SD10 de información de regularización puede incluir una 20 correspondiente descripción de la modificación, tal como un conjunto de parámetros de función. En un ejemplo particular, el codificador A124 de banda estrecha está configurado para dividir una trama en tres subtramas y para calcular un desplazamiento temporal fijo para cada subtrama, de tal manera que la señal SD10 de datos de regularización indique tres cantidades de desplazamiento temporal para cada trama regularizada de la señal codificada de banda estrecha.

25 El codificador AD10 de voz de banda ancha incluye una línea D120 de retardo configurada para avanzar o retrasar partes de la señal S30 de voz de banda alta, según cantidades de retardo indicadas por una señal de entrada, para producir la señal S30a de voz de banda alta con distorsión temporal. En el ejemplo mostrado en la FIGURA 25, la línea D120 de retardo está configurada para distorsionar temporalmente la señal S30 de voz de banda alta según la distorsión indicada por la señal SD10 de datos de regularización. De tal manera, la misma cantidad de distorsión temporal que se incluyó en la señal codificada S50 de excitación de banda estrecha también se aplica a la porción 30 correspondiente de la señal S30 de voz de banda alta antes del análisis. Aunque este ejemplo muestra la línea D120 de retardo como un elemento separado del codificador A200 de banda alta, en otras implementaciones la línea D120 de retardo se dispone como parte del codificador de banda alta.

35 Otras implementaciones del codificador A200 de banda alta pueden configurarse para llevar a cabo un análisis espectral (por ejemplo, un análisis de LPC) de la señal S30 de voz de banda alta no distorsionada y para llevar a cabo la distorsión temporal de la señal S30 de voz de banda alta antes del cálculo de los parámetros S60b de ganancia de banda alta. Un codificador de este tipo puede incluir, por ejemplo, una implementación de la línea D120 de retardo dispuesta para llevar a cabo la distorsión temporal. Sin embargo, en tales casos, los parámetros S60a de filtro de banda alta basados en el análisis de la señal S30 no distorsionada pueden describir una envolvente espectral desalineada en el tiempo con la señal S120 de excitación de banda alta.

40 La línea D120 de retardo puede configurarse según cualquier combinación de elementos lógicos y elementos de almacenamiento adecuados para aplicar las operaciones de distorsión temporal deseadas a la señal S30 de voz de banda alta. Por ejemplo, la línea D120 de retardo puede configurarse para leer la señal S30 de voz de banda alta desde una memoria intermedia según los desplazamientos temporales deseados. La FIGURA 26a muestra un diagrama esquemático de una implementación D122 de este tipo de la línea D120 de retardo que incluye un registro SR1 de desplazamiento. El registro SR1 de desplazamiento es una memoria intermedia de cierta longitud m que 45 está configurada para recibir y almacenar las m muestras más recientes de la señal S30 de voz de banda alta. El valor m es igual al menos a la suma de los desplazamientos temporales máximos positivos (o de "adelanto") y negativos (o de "retraso") que han de soportarse. Puede ser conveniente que el valor m sea igual a la longitud de una trama o una subtrama de la señal S30 de banda alta.

50 La línea D122 de retardo está configurada para emitir la señal S30a de banda alta con distorsión temporal desde una ubicación OL de desfase del registro SR1 de desplazamiento. La posición de la ubicación OL del desfase varía con respecto a una posición de referencia (desplazamiento temporal cero) según el desplazamiento temporal actual tal como se indica, por ejemplo, por medio de la señal SD10 de datos de regularización. La línea D122 de retardo puede configurarse para soportar límites de adelanto y retardo iguales o, como alternativa, un límite más grande que el otro, de tal manera que pueda llevarse a cabo un mayor desplazamiento en una dirección que en la otra. La FIGURA 26a muestra un ejemplo particular que soporta un mayor desplazamiento temporal positivo que negativo. La línea D122 de retardo puede ser configurada para emitir una o más muestras cada vez (dependiendo, por ejemplo, del ancho de bus de salida).

Un desplazamiento temporal de regularización que tenga una magnitud de más de algunos milisegundos puede provocar artefactos audibles en la señal decodificada. Normalmente, la magnitud de un desplazamiento temporal de regularización, tal como el que lleva a cabo un codificador A124 de banda estrecha, no superará algunos milisegundos, de tal manera que se limitarán los desplazamientos temporales indicados por la señal SD10 de datos de regularización. Sin embargo, puede desearse en tales casos que la línea D122 de retardo esté configurada para imponer un límite máximo a los desplazamientos temporales en la dirección positiva y/o negativa (por ejemplo, para observar un límite más ajustado que el impuesto por el codificador de banda estrecha).

La FIGURA 26b muestra un diagrama esquemático de una implementación D124 de la línea D122 de retardo que incluye una ventana SW de desplazamiento. En este ejemplo, la posición de la ubicación OL del desfase está limitada por la ventana SW de desplazamiento. Aunque la FIGURA 26b muestra un caso en el que la longitud m de memoria intermedia es mayor que el ancho de la ventana SW de desplazamiento, la línea D124 de retardo también puede ser implementada de tal manera que el ancho de la ventana SW de desplazamiento sea igual a m .

En otras implementaciones, la línea D120 de retardo está configurada para escribir la señal S30 de voz de banda alta en una memoria intermedia según los desplazamientos temporales deseados. La FIGURA 27 muestra un diagrama esquemático de tal implementación D130 de la línea D120 de retardo que incluye dos registros SR2 y SR3 de desplazamiento configurados para recibir y almacenar la señal S30 de voz de banda alta. La línea D130 de retardo está configurada para escribir una trama o una subtrama del registro SR2 de desplazamiento al registro SR3 de desplazamiento según un desplazamiento temporal, tal como se indica, por ejemplo, por medio de la señal SD10 de datos de regularización. El registro SR3 de desplazamiento está configurado como una memoria intermedia FIFO dispuesta para emitir la señal S30 de banda alta con distorsión temporal.

En el ejemplo particular mostrado en la FIGURA 27, el registro SR2 de desplazamiento incluye una porción FB1 de memoria intermedia de trama y una porción DB de memoria intermedia de retraso, y el registro SR3 de desplazamiento incluye una porción FB2 de memoria intermedia de trama, una porción AB de memoria intermedia de adelanto, y una parte RB de memoria intermedia de retraso. Las longitudes de la memoria intermedia AB de adelanto y la memoria intermedia RB de retraso pueden ser iguales, o una puede ser más grande que la otra, de tal manera que se soporte un mayor desplazamiento en una dirección que en la otra. La memoria DB intermedia de retardo y la parte RB de memoria intermedia de retraso pueden configurarse para que tengan la misma longitud. Como alternativa, la memoria DB intermedia de retardo puede ser más pequeña que la memoria intermedia RB de retraso para representar un intervalo de tiempo requerido para transferir muestras desde la memoria intermedia FB1 de trama al registro SR3 de desplazamiento, lo que puede incluir otras operaciones de procesamiento tales como la distorsión de las muestras antes de su almacenamiento en el registro SR3 de desplazamiento.

En el ejemplo de la FIGURA 27, la memoria intermedia FB1 de trama está configurada para que tenga una longitud igual a la de una trama de la señal S30 de banda alta. En otro ejemplo, la memoria intermedia FB1 de trama está configurada para que tenga una longitud igual a la de una subtrama de la señal S30 de banda alta. En tal caso, la línea D130 de retardo puede ser configurada para incluir una lógica para aplicar el mismo retardo (por ejemplo, un promedio) a todas las subtramas de una trama que vaya a desplazarse. La línea D130 de retardo también puede incluir una lógica para promediar valores de la memoria intermedia FB1 de trama con valores que van a sobreescibirse en la memoria intermedia RB de retraso o la memoria intermedia AB de adelanto. En un ejemplo adicional, el registro SR3 de desplazamiento puede estar configurado para recibir valores de la señal S30 de banda alta solo por medio de la memoria intermedia FB1 de trama, y en tal caso la línea D130 de retardo puede incluir una lógica para interpolar pausas entre tramas o subtramas sucesivas escritas en el registro SR3 de desplazamiento. En otras implementaciones, la línea D130 de retardo puede ser configurada para llevar a cabo una operación de distorsión en muestras de la memoria intermedia FB1 de trama antes de escribirlas en el registro SR3 de desplazamiento (por ejemplo, según una función descrita por la señal SD10 de datos de regularización).

Puede que resulte deseable que la línea D120 de retardo aplique una distorsión temporal que se base, sin ser idéntica a ella, en la distorsión especificada por la señal SD10 de datos de regularización. La FIGURA 28 muestra un diagrama de bloques de una implementación AD12 del codificador AD10 de voz de banda ancha que incluye un correlacionador D110 del valor de retardo. El correlacionador D110 del valor de retardo está configurado para correlacionar la distorsión indicada por la señal SD10 de datos de regularización con valores SD10a de retardo correlacionados. La línea D120 de retardo se dispone para producir la señal S30a de voz de banda alta con distorsión temporal según la distorsión indicada por los valores SD10a de retardo correlacionados.

Cabe esperar que el desplazamiento temporal aplicado por el codificador de banda estrecha evolucione suavemente en el tiempo. Por lo tanto, normalmente es suficiente calcular el desplazamiento temporal promedio de banda estrecha aplicado a las subtramas durante una trama de voz, y desplazar una trama correspondiente de la señal S30 de voz de banda alta según este promedio. En un ejemplo de este tipo, el correlacionador D110 del valor de retardo está configurado para calcular un promedio de los valores de retardo de subtrama para cada trama, y la línea D120 de retardo está configurada para aplicar el promedio calculado a una trama correspondiente de la señal S30 de banda alta. En otros ejemplos, puede calcularse y aplicarse un promedio en un periodo más corto (tal como dos subtramas, o la mitad de una trama) o un periodo más largo (tal como dos tramas). En un caso en el que el

promedio sea un valor no entero de muestras, el correlacionador D110 del valor de retardo puede estar configurado para redondear el valor a un número entero de muestras antes de emitirlo a la línea D120 de retardo.

El codificador A124 de banda estrecha puede ser configurado para incluir un desplazamiento temporal de regularización de un número no entero de muestras en la señal codificada de excitación de banda estrecha. En un caso de este tipo, puede ser deseable que el correlacionador D110 del valor de retardo se configure para redondear el desplazamiento temporal de banda estrecha a un número entero de muestras y que la línea D120 de retardo aplique el desplazamiento temporal redondeado a la señal S30 de voz de banda alta.

En algunas implementaciones del codificador AD10 de voz de banda ancha, pueden diferir las tasas de muestreo de la señal S20 de voz de banda estrecha y la señal S30 de voz de banda alta. En tales casos, el correlacionador D110 del valor de retardo puede configurarse para ajustar las cantidades de desplazamiento temporal indicadas en la señal SD10 de datos de regularización para justificar una diferencia entre las tasas de muestreo de la señal S20 de voz de banda estrecha (o la señal S80 de excitación de banda estrecha) y la señal S30 de voz de banda alta. Por ejemplo, el correlacionador D110 del valor de retardo puede configurarse para ajustar a escala las cantidades de desplazamiento temporal según una relación de las tasas de muestreo. En un ejemplo particular, tal como se mencionó anteriormente, se muestrea la señal S20 de voz de banda estrecha a 8 kHz y se muestrea la señal S30 de voz de banda alta a 7 kHz. En este caso, el correlacionador D110 del valor de retardo está configurado para multiplicar cada cantidad de desplazamiento por 7/8. También pueden configurarse implementaciones del correlacionador D110 del valor de retardo para llevar a cabo tal operación de ajuste a escala junto con una operación de redondeo a número entero y/o promediación de desplazamiento temporal, tal como se describe en el presente documento.

En otras implementaciones, la línea D120 de retardo está configurada para modificar de otro modo la escala temporal de una trama u otra secuencia de muestras (por ejemplo, comprimiendo una porción y expandiendo otra porción). Por ejemplo, el codificador A124 de banda estrecha puede configurarse para llevar a cabo la regularización según una función tal como una trayectoria o contorno de altura tonal. En tal caso, la señal SD10 de datos de regularización puede incluir una correspondiente descripción de la función, tal como un conjunto de parámetros, y la línea D120 de retardo puede incluir una lógica configurada para distorsionar tramas o subtramas de la señal S30 de voz de banda alta según la función. En otras implementaciones, el correlacionador D110 del valor de retardo está configurado para promediar, ajustar a escala y/o redondear la función antes de que sea aplicada a la señal S30 de voz de banda alta por la línea D120 de retardo. Por ejemplo, el correlacionador D110 del valor de retardo puede configurarse para calcular uno o más valores de retardo según la función, indicando cada valor de retardo una serie de muestras, que son entonces aplicadas por la línea D120 de retardo para distorsionar temporalmente una o más tramas o subtramas correspondientes de la señal S30 de voz de banda alta.

La FIGURA 29 muestra un diagrama de flujo para un procedimiento MD100 de distorsión temporal de una señal de voz de banda alta según una distorsión temporal incluida en una correspondiente señal codificada de excitación de banda estrecha. La tarea TD100 procesa una señal de voz de banda ancha para obtener una señal de voz de banda estrecha y una señal de voz de banda alta. Por ejemplo, la tarea TD100 puede configurarse para filtrar la señal de voz de banda ancha usando un banco de filtros que tiene filtros de paso bajo y paso alto, tales como una implementación del banco A110 de filtros. La tarea TD200 codifica la señal de voz de banda estrecha en al menos una señal codificada de excitación de banda estrecha y una pluralidad de parámetros de filtro de banda estrecha. Pueden cuantificarse la señal codificada de excitación de banda estrecha y/o los parámetros de filtro, y la señal de voz de banda estrecha codificada también puede incluir otros parámetros tales como un parámetro del modo de voz. La tarea TD200 también incluye una distorsión temporal en la señal codificada de excitación de banda estrecha.

La tarea TD300 genera una señal de excitación de banda alta basada en una señal de excitación de banda estrecha. En este caso, la señal de excitación de banda estrecha se basa en la señal codificada de excitación de banda estrecha. Según al menos la señal de excitación de banda alta, la tarea TD400 codifica la señal de voz de banda alta en al menos una pluralidad de parámetros de filtro de banda alta. Por ejemplo, la tarea TD400 puede ser configurada para codificar la señal de voz de banda alta en una pluralidad de LSF cuantificadas. La tarea TD500 aplica un desplazamiento temporal a la señal de voz de banda alta que se basa en información relativa a una distorsión temporal incluida en la señal codificada de excitación de banda estrecha.

La tarea TD400 puede ser configurada para llevar a cabo un análisis espectral (tal como un análisis de LPC) en la señal de voz de banda alta, y/o para calcular una envolvente de ganancia de la señal de voz de banda alta. En tales casos, la tarea TD500 puede estar configurada para aplicar el desplazamiento temporal a la señal de voz de banda alta antes del análisis y/o el cálculo de la envolvente de ganancia.

Otras implementaciones del codificador A100 de voz de banda ancha están configuradas para invertir una distorsión temporal de la señal S120 de excitación de banda alta provocado por una distorsión temporal incluida en la señal codificada de excitación de banda estrecha. Por ejemplo, el generador A300 de excitación de banda alta puede implementarse para que incluya una implementación de la línea D120 de retardo que esté configurada para recibir la señal SD10 de datos de regularización o los valores SD10a de retardo correlacionados, y para que aplique un desplazamiento temporal inverso correspondiente a la señal S80 de excitación de banda estrecha, y/o a una señal

posterior basándose ella, tal como la señal S160 ensanchada armónicamente o la señal S120 de excitación de banda alta.

Otras implementaciones del codificador de voz de banda ancha pueden configurarse para codificar la señal S20 de voz de banda estrecha y la señal S30 de voz de banda alta independientemente entre sí, de tal manera que se codifique la señal S30 de voz de banda alta como una representación de una envolvente espectral de banda alta y una señal de excitación de banda alta. Puede configurarse una implementación de este tipo para llevar a cabo una distorsión temporal de la señal residual de banda alta, o incluir, si no, una distorsión temporal en una señal codificada de excitación de banda alta, según información relativa a una distorsión temporal incluida en la señal codificada de excitación de banda estrecha. Por ejemplo, el codificador de banda alta puede incluir una implementación de la línea D120 de retardo y/o el correlacionador D110 del valor de retardo, tal como se describe en el presente documento, que se configuran para aplicar una distorsión temporal a la señal residual de banda alta. Las ventajas potenciales de una operación de este tipo incluyen una codificación más eficaz de la señal residual de banda alta y una mayor coincidencia entre las señales sintetizadas de voz de banda alta y de banda estrecha.

Tal como se ha mencionado en lo que antecede, las realizaciones, tal como se describen en el presente documento, incluyen implementaciones que pueden usarse para llevar a cabo una codificación integrada, soportando compatibilidad con sistemas de banda estrecha y evitando la necesidad de transcodificación. El soporte para la codificación de banda alta también puede servir para diferenciar, en cuanto al coste, entre chips, conjuntos de chips, dispositivos y/o redes que tienen soporte de banda ancha con compatibilidad hacia atrás, y aquellos que solo tienen soporte de banda estrecha. También puede usarse soporte para codificación de banda alta, tal como se describe en el presente documento, conjuntamente con una técnica para soportar la codificación de banda baja, y un sistema, un procedimiento o un aparato según una realización de este tipo pueden soportar la codificación de componentes de frecuencia, por ejemplo, desde aproximadamente 50 o 100 Hz hasta aproximadamente 7 u 8 kHz.

Tal como se ha mencionado en lo que antecede, la adición de soporte de banda alta a un codificador de voz puede mejorar la inteligibilidad, especialmente en relación con la diferenciación de fricativas. Aunque tal diferenciación puede ser deducida habitualmente por un oyente humano del contexto particular, el soporte de banda alta puede servir como una característica de habilitación en el reconocimiento del habla y otras aplicaciones de interpretación por máquina, tales como sistemas para la navegación automatizada de menú por voz y/o el procesamiento automático de llamadas.

Un aparato según una realización puede integrarse en un dispositivo portátil para las comunicaciones inalámbricas tal como un teléfono celular o una agenda electrónica (PDA). Como alternativa, puede incluirse un aparato de este tipo en otro dispositivo de comunicaciones, tal como un micrófono de VoIP, un ordenador personal configurado para soportar comunicaciones de VoIP o un dispositivo de red configurado para encaminar comunicaciones telefónicas o de VoIP. Por ejemplo, puede implementarse un aparato según una realización en un chip o un conjunto de chips para un dispositivo de comunicaciones. Dependiendo de la aplicación particular, un dispositivo de este tipo también puede incluir características tales como la conversión analógica-digital y/o la digital-analógica de una señal de voz, circuitería para llevar a cabo operaciones de amplificación y/u otras operaciones de procesamiento de señales en una señal de voz, y/o circuitería de radiofrecuencia para la transmisión y/o la recepción de la señal codificada de voz.

Se contempla y se da a conocer explícitamente que las realizaciones pueden incluir y/o usarse con una cualquiera o más de otras características. Tales características incluyen la eliminación de ráfagas de alta energía de corta duración que se producen en la banda alta y están sustancialmente ausentes en la banda estrecha. Tales características incluyen el aplanamiento fijo o adaptativo de representaciones de coeficiente tales como las LSF de banda alta. Tales características incluyen conformación fija o adaptativa del ruido asociado con la cuantificación de representaciones de coeficiente tales como las LSF. Tales características también incluyen el aplanamiento fijo o adaptativo de una envolvente de ganancia, y atenuación adaptativa de una envolvente de ganancia.

Se proporciona la presentación anterior de las realizaciones descritas para permitir que cualquier experto en la técnica realice o use la presente invención. Son posibles diversas modificaciones de estas realizaciones, y los principios genéricos presentados en el presente documento también pueden aplicarse a otras realizaciones. Por ejemplo, una realización puede implementarse en parte o en su totalidad como un circuito cableado, como una configuración de circuito fabricada en un circuito integrado para aplicaciones específicas, o como un programa de soporte lógico inalterable cargado en un almacenamiento no volátil o un programa de soporte lógico cargado desde un medio de almacenamiento de datos, o en el mismo, como código legible por máquina, siendo dicho código instrucciones ejecutables por un conjunto de elementos lógicos tales como un microprocesador u otra unidad de procesamiento de señales digitales. El medio de almacenamiento de datos puede ser un conjunto de elementos de almacenamiento tales como una memoria de semiconductores (que puede incluir, sin limitación, RAM (memoria de acceso aleatorio) estática o dinámica, ROM (memoria de solo lectura) y/o flash RAM), o memoria ferroeléctrica, magnetoresistiva, ovónica, polimérica o de cambio de fase; o un medio de disco tal como un disco magnético y óptico. Debe entenderse que la expresión "soporte lógico" incluye código fuente, código de lenguaje ensamblador, código máquina, código binario, soporte lógico inalterable, macrócódigo, microcódigo, uno cualquiera o más de

conjuntos o secuencias de instrucciones ejecutables por un conjunto de elementos lógicos, y cualquier combinación de tales ejemplos.

Los diversos elementos de implementaciones de los generadores A300 y B300 de excitación de banda alta, el codificador A100 de banda alta, el decodificador B200 de banda alta, el codificador A100 de voz de banda ancha y el

5 decodificador B100 de habla de banda ancha pueden implementarse como dispositivos electrónicos y/u ópticos que residan, por ejemplo, en el mismo chip o entre dos o más chips en un conjunto de chips, aunque también se contemplan otras disposiciones sin tal limitación. Uno o más elementos de un aparato de este tipo pueden implementarse en su totalidad o en parte como uno o más conjuntos de instrucciones dispuestos para ejecutarse en uno o más conjuntos fijos o programables de elementos lógicos (por ejemplo, transistores, puertas), tales como
10 microprocesadores, procesadores integrados, núcleos IP, procesadores de señales digitales, FPGA (matrices de puertas programables in situ), ASSP (productos estándar para aplicaciones específicas) y ASIC (circuitos integrados para aplicaciones específicas). También es posible que uno o más de tales elementos tengan una estructura en común (por ejemplo, un procesador usado para ejecutar porciones de código correspondientes a elementos diferentes en momentos diferentes, un conjunto de instrucciones ejecutadas para llevar a cabo tareas
15 correspondientes a elementos diferentes en momentos diferentes, o una disposición de dispositivos electrónicos y/u ópticos que lleve a cabo operaciones para elementos diferentes en momentos diferentes). Además, es posible que uno o más de tales elementos sea usado para llevar a cabo tareas o ejecutar otros conjuntos de instrucciones que no se relacionen directamente con una operación del aparato, tales como una tarea que se relacione con otra operación de un dispositivo o un sistema en el que se integre el aparato.

20 La FIGURA 30 muestra un diagrama de flujo de un procedimiento M100, según una realización, de codificación de una porción de banda alta de una señal de voz que tiene una porción de banda estrecha y la porción de banda alta. La tarea X100 calcula un conjunto de parámetros de filtro que caracterizan una envolvente espectral de la porción de banda alta. La tarea X200 calcula una señal ensanchada espectralmente aplicando una función no lineal a una señal derivada de la porción de banda estrecha. La tarea X300 genera una señal sintetizada de banda alta según (A) el
25 conjunto de parámetros de filtro y (B) una señal de excitación de banda alta basada en la señal ensanchada espectralmente. La tarea X400 calcula una envolvente de ganancia basada en una relación entre (C) la energía de la porción de banda alta y (D) la energía de una señal derivada de la porción de banda estrecha.

30 La FIGURA 31a muestra un diagrama de flujo de un procedimiento M200 de generación de una señal de excitación de banda alta según una realización. La tarea Y100 calcula una señal ensanchada armónicamente aplicando una función no lineal a una señal de excitación de banda estrecha derivada de una porción de banda estrecha de una señal de voz. La tarea Y200 mezcla la señal ensanchada armónicamente con una señal modulada de ruido para
35 generar una señal de excitación de banda alta. La FIGURA 31b muestra un diagrama de flujo de un procedimiento M210 de generación de una señal de excitación de banda alta según otra realización que incluye las tareas Y300 e Y400. La tarea Y300 calcula una envolvente en el dominio temporal según la energía en el tiempo de una entre la señal de excitación de banda estrecha y la señal ensanchada armónicamente. La tarea Y400 modula una señal de ruido según la envolvente en el dominio temporal para producir la señal modulada de ruido.

40 La FIGURA 32 muestra un diagrama de flujo de un procedimiento M300 según una realización, de decodificación de una porción de banda alta de una señal de voz que tiene una porción de banda estrecha y la porción de banda alta. La tarea Z100 recibe un conjunto de parámetros de filtro que caracterizan una envolvente espectral de la porción de banda alta y un conjunto de factores de ganancia que caracterizan una envolvente temporal de la porción de banda alta. La tarea Z200 calcula una señal ensanchada espectralmente aplicando una función no lineal a una señal derivada de la porción de banda estrecha. La tarea Z300 genera una señal sintetizada de banda alta según (A) el
45 conjunto de parámetros de filtro y (B) una señal de excitación de banda alta basada en la señal ensanchada espectralmente. La tarea Z400 modula una envolvente de ganancia de la señal sintetizada de banda alta basada en el conjunto de factores de ganancia. Por ejemplo, la tarea Z400 puede configurarse para modular la envolvente de ganancia de la señal sintetizada de banda alta aplicando el conjunto de factores de ganancia a una señal de excitación derivada de la porción de banda estrecha, a la señal ensanchada espectralmente, a la señal de excitación de banda alta o a la señal sintetizada de banda alta.

50 Las realizaciones también incluyen procedimientos adicionales de codificación de voz, cifrado y decodificación, tal como se da a conocer expresamente en el presente documento, por ejemplo, mediante descripciones de realizaciones estructurales configuradas para llevar a cabo tales procedimientos. Cada uno de estos procedimientos también puede ser implementado de forma tangible (por ejemplo, en uno o más medios de almacenamiento de datos según se ha enumerado en lo que antecede) como uno o más conjuntos de instrucciones legibles y/o ejecutables por una máquina, incluyendo un conjunto de elementos lógicos (por ejemplo, un procesador, un microprocesador, un
55 microcontrolador u otra máquina de estado finito). Por ello, no se pretende que la presente invención esté limitada a las realizaciones mostradas en lo que antecede, sino que, más bien, debe otorgársele el alcance más amplio coherente con las reivindicaciones adjuntas.

REIVINDICACIONES

1. Un procedimiento de generación de una señal (S120) de excitación de banda alta, comprendiendo dicho procedimiento:

5 ensanchar armónicamente el espectro de una señal que se basa en una señal (S80) de excitación de banda baja;
 calcular una envolvente de dominio temporal de una señal que se basa en la señal (S80) de excitación de banda baja;
 modular una señal de ruido según la envolvente de dominio temporal; y
 10 combinar (A) una señal (S160) ensanchada armónicamente con base en un resultado de dicho ensanchamiento armónico y (B) una señal modulada (S170) de ruido con base en un resultado de dicha modulación, incluyendo dicha combinación el cálculo de una suma ponderada de la señal (S160) ensanchada armónicamente y la señal modulada (S170) de ruido, incluyendo dicho cálculo de una suma ponderada la ponderación de la señal (S160) ensanchada armónicamente según un primer factor de ponderación y la ponderación de la señal modulada (S170) de ruido según un segundo factor de ponderación, comprendiendo dicho procedimiento el cálculo de al menos uno entre los factores de ponderación primero y segundo según al menos uno entre (A) una medida de periodicidad de una señal de voz y (B) un grado vocal de una señal de voz,
 15 en el que la señal de excitación de banda alta se basa en la suma ponderada.

- 20 2. El procedimiento según la reivindicación 1 en el que dicho ensanchamiento armónico comprende aplicar una función no lineal a una señal que se basa en la señal (S80) de excitación de banda baja.

3. El procedimiento según la reivindicación 2 en el que dicha aplicación de una función no lineal comprende aplicar la función no lineal en el dominio temporal.

4. El procedimiento según la reivindicación 2 en el que la función no lineal es una función no lineal sin memoria.

5. El procedimiento según la reivindicación 2 en el que la función no lineal es invariante en el tiempo.

- 25 6. El procedimiento según la reivindicación 2 en el que la función no lineal comprende al menos una de la función de valor absoluto, la función de elevación al cuadrado y una función de recorte.

7. El procedimiento según la reivindicación 2 en el que la función no lineal es la función de valor absoluto.

8. El procedimiento según la reivindicación 1 en el que dicho cálculo de una envolvente de dominio temporal de una señal que se basa en la señal (S80) de excitación de banda baja incluye calcular una envolvente de dominio temporal entre la señal (S80) de excitación de banda baja y la señal (S160) ensanchada armónicamente.

- 30 9. El procedimiento según la reivindicación 1 en el que dicho ensanchamiento armónico incluye ensanchar armónicamente el espectro de una señal sobremuestreada que se basa en la señal (S80) de excitación de banda baja.

- 35 10. El procedimiento según la reivindicación 1, comprendiendo dicho procedimiento aplanar espectralmente la señal ensanchada armónicamente antes de dicha combinación.

11. El procedimiento según la reivindicación 10 en el que dicho aplanamiento espectral comprende:

40 calcular una pluralidad de coeficientes de filtro con base en una señal que ha de ser aplanada espectralmente; y
 filtrar la señal que ha de ser aplanada espectralmente con un filtro blanqueador configurado según la pluralidad de coeficientes de filtro.

12. El procedimiento según la reivindicación 1, comprendiendo dicho procedimiento la generación de la señal de ruido según una función determinista de información dentro de una señal codificada de voz.

- 45 13. El procedimiento según la reivindicación 1, comprendiendo dicho procedimiento la obtención de la señal (S80) de excitación de banda baja y un valor de ganancia de altura tonal a partir de una representación cuantificada de un resto (S50) de banda baja, y comprendiendo dicho procedimiento calcular uno entre los factores de ponderación primero y segundo según al menos el valor de ganancia de altura tonal.

- 50 14. El procedimiento según la reivindicación 1, comprendiendo dicho procedimiento al menos uno de (i) codificar una señal de voz de banda alta según la señal (S120) de excitación de banda alta y (ii) decodificar una señal de voz de banda alta según la señal (S120) de excitación de banda alta.

15. Un medio de almacenamiento de datos que contiene instrucciones ejecutables por máquina para llevar a cabo el procedimiento de procesamiento de señales según la reivindicación 1.

16. Un aparato (A302) que comprende:

5 un medio para ensanchar armónicamente el espectro de una señal que se basa en una señal (S80) de excitación de banda baja;

un medio para calcular una envolvente de dominio temporal de una señal que se basa en la señal (S80) de excitación de banda baja;

un medio para modular una señal de ruido según la envolvente de dominio temporal; y

10 un medio para combinar (A) una señal (S160) ensanchada armónicamente con base en un resultado de dicho ensanchamiento armónico y (B) una señal modulada (S170) de ruido con base en un resultado de dicha modulación, incluyendo dicho medio de combinación un medio de cálculo de una suma ponderada de la señal (S160) ensanchada armónicamente y la señal modulada (S170) de ruido, estando configurado dicho medio de combinación para ponderar la señal (S160) ensanchada armónicamente según un primer factor de ponderación y para ponderar la señal modulada (S170) de ruido según un segundo factor de ponderación, estando configurado dicho medio de combinación para calcular al menos uno entre los factores de ponderación primero y segundo según al menos uno entre (A) una medida de periodicidad de una señal de voz y (B) un grado vocal de una señal de voz,

15 en el que la señal (S120) de excitación de banda alta se basa en la suma ponderada.

17. El aparato (A302) de la reivindicación 16 en el que:

20 el medio para ensanchar armónicamente el espectro de una señal es un ensanchador (A400) de espectro; el medio para calcular una envolvente de dominio temporal de una señal es un calculador (460) de envolvente;

el medio para modular una señal de ruido es un primer combinador (470); y

el medio para combinar (A) y (B) es un segundo combinador (490).

25 18. El aparato (A302) según la reivindicación 17 en el que dicho ensanchador (A400) de espectro está configurado para aplicar una función no lineal para llevar a cabo el ensanchamiento armónico del espectro de una señal que se basa en la señal (S80) de excitación de banda baja.

19. El aparato (A302) según la reivindicación 18 en el que la función no lineal comprende al menos una de la función de valor absoluto, la función de elevación al cuadrado y una función de recorte.

30 20. El aparato (A302) según la reivindicación 18 en el que la función no lineal es la función de valor absoluto.

21. El aparato (A302) según la reivindicación 17 en el que dicho calculador (460) de envolvente está configurado para calcular la envolvente de dominio temporal con base en una entre la señal (S80) de excitación de banda baja y la señal (S160) ensanchada armónicamente.

35 22. El aparato (A302) según la reivindicación 17 en el que dicho ensanchador (A400) de espectro está configurado para llevar a cabo un ensanchamiento armónico del espectro de una señal sobremuestreada que se basa en la señal (S80) de excitación de banda baja.

23. El aparato (A302) según la reivindicación 17, comprendiendo dicho aparato un aplanador espectral configurado para aplanar espectralmente la señal ensanchada armónicamente.

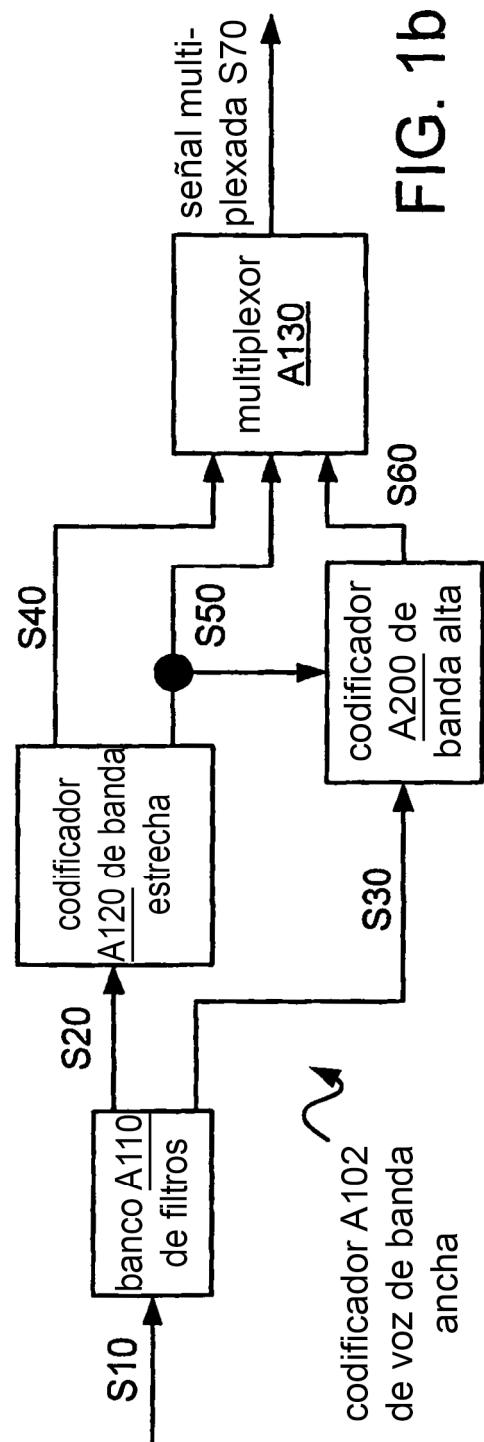
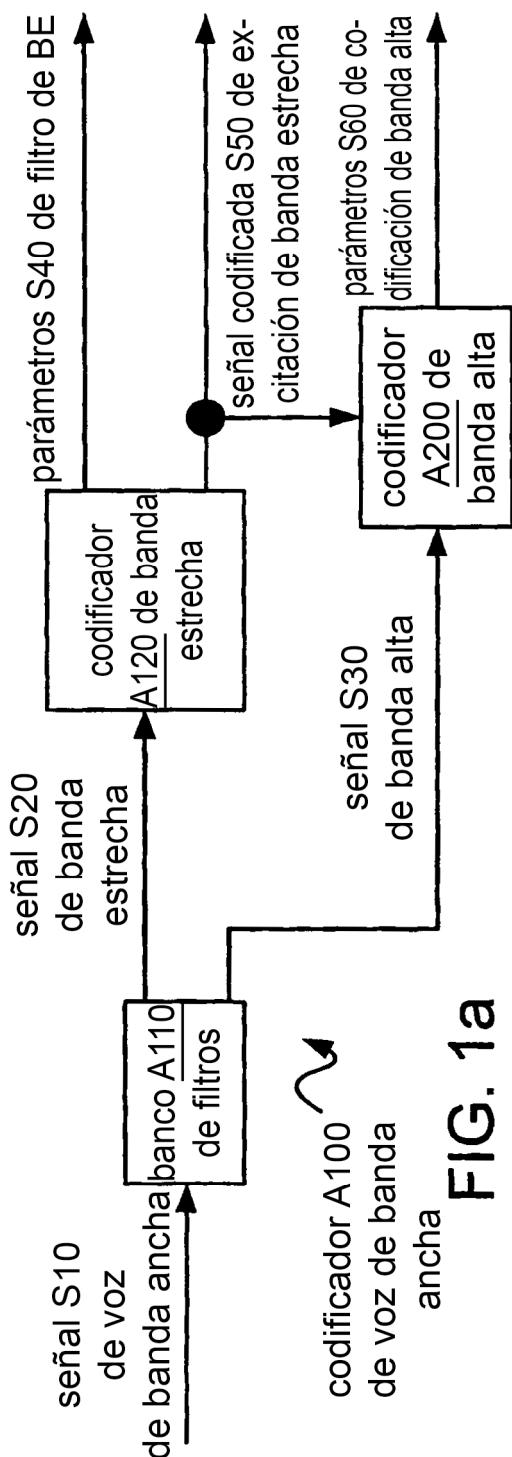
40 24. El aparato (A302) según la reivindicación 23 en el que dicho aplanador espectral está configurado para calcular una pluralidad de coeficientes de filtro con base en una señal que ha de ser aplanada espectralmente y filtrar la señal que ha de ser aplanada espectralmente con un filtro blanqueador configurado según la pluralidad de coeficientes de filtro.

25. El aparato (A302) según la reivindicación 17, comprendiendo dicho aparato un generador de ruido configurado para generar la señal de ruido según una función determinista de información dentro de una señal codificada de voz.

45 26. El aparato (A302) según la reivindicación 16, incluyendo dicho aparato un descuantificador configurado para obtener la señal (S80) de excitación de banda baja y un valor de ganancia de altura tonal a partir de una representación cuantificada de un resto (S50) de banda baja, y estando configurado dicho segundo combinador (490) para calcular al menos uno entre los factores de ponderación primero y segundo según al menos el valor de ganancia de altura tonal.

50 27. El aparato (A302) según la reivindicación 17, incluyendo dicho aparato al menos uno de (i) un codificador de voz de banda alta configurado para codificar una señal de voz de banda alta según la señal de excitación de banda alta y (ii) un decodificador de voz de banda alta configurado para decodificar una señal de voz de banda alta según la señal de excitación de banda alta.

28. El aparato (A302) según la reivindicación 17, comprendiendo dicho aparato un teléfono celular.



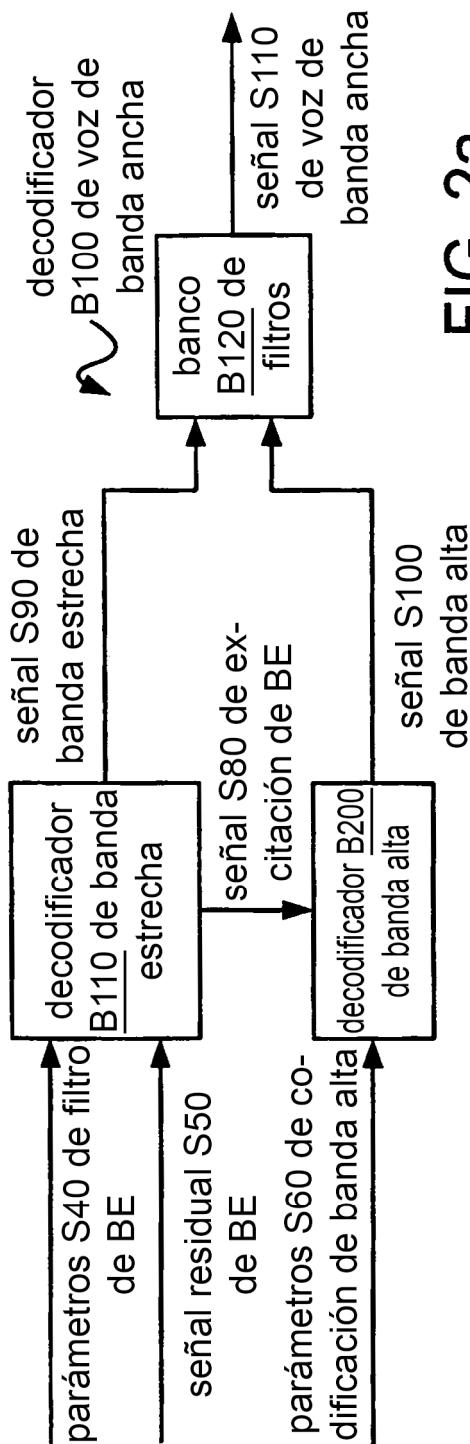


FIG. 2a

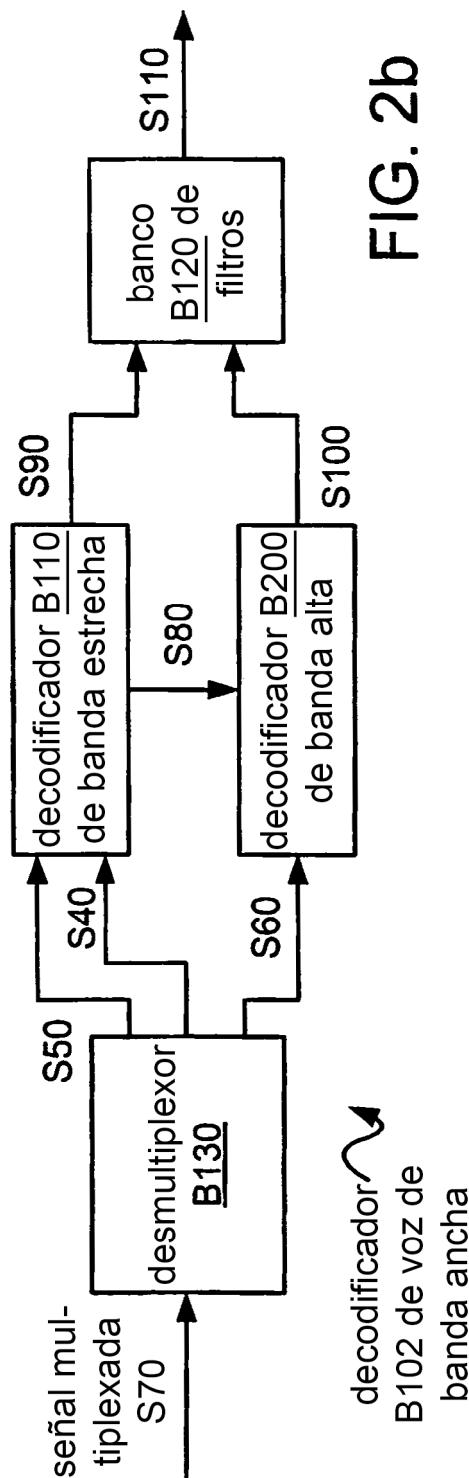


FIG. 2b

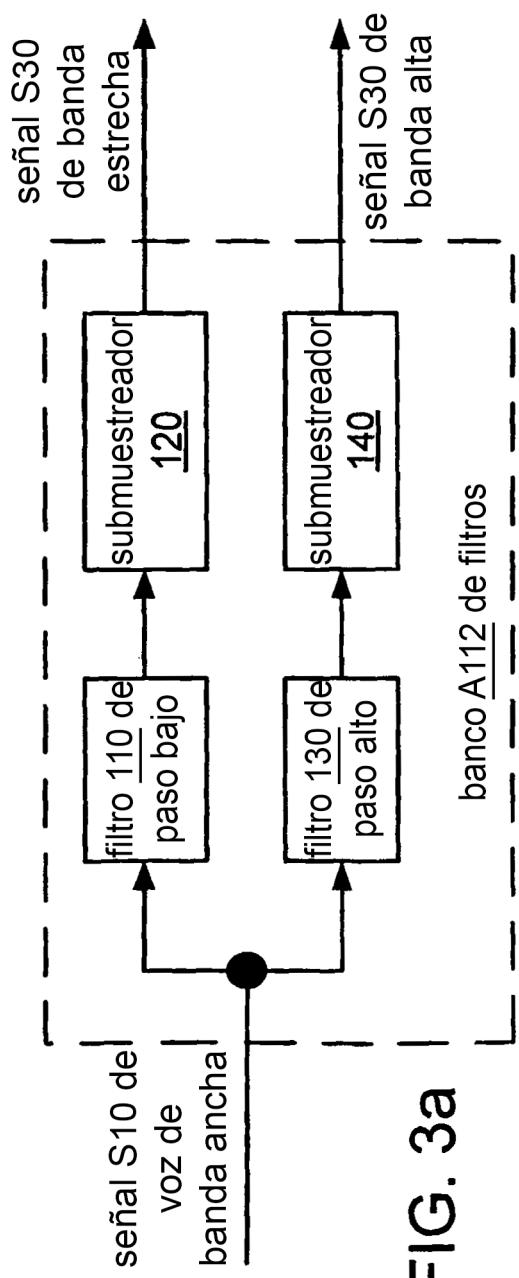


FIG. 3a

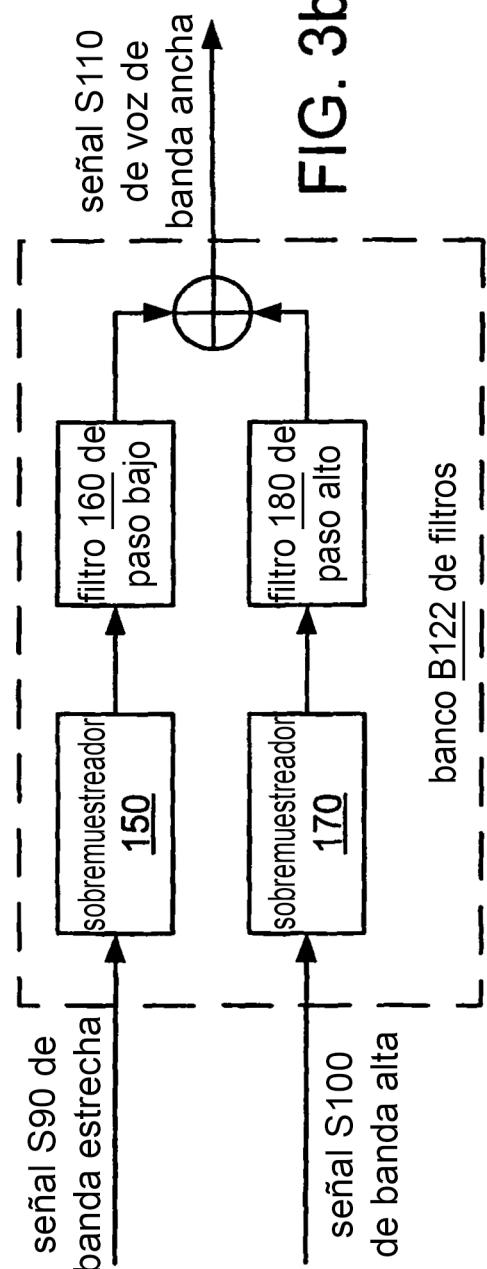


FIG. 3b

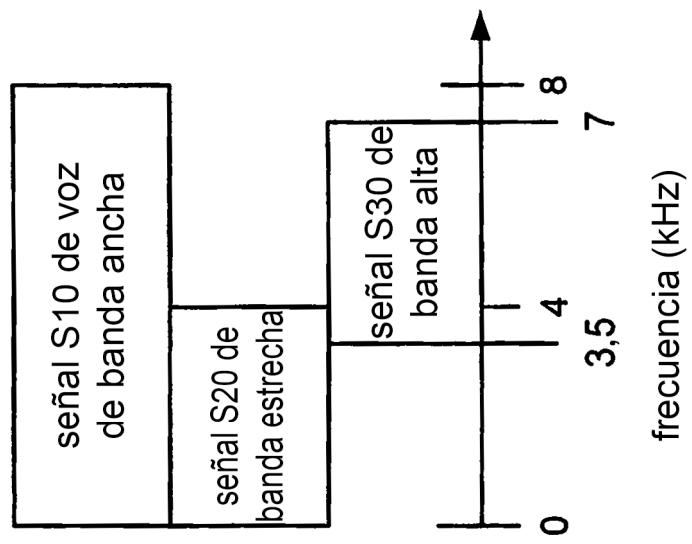


FIG. 4b

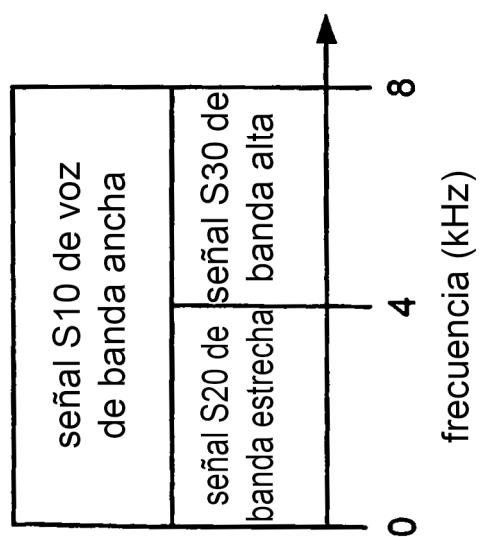


FIG. 4a

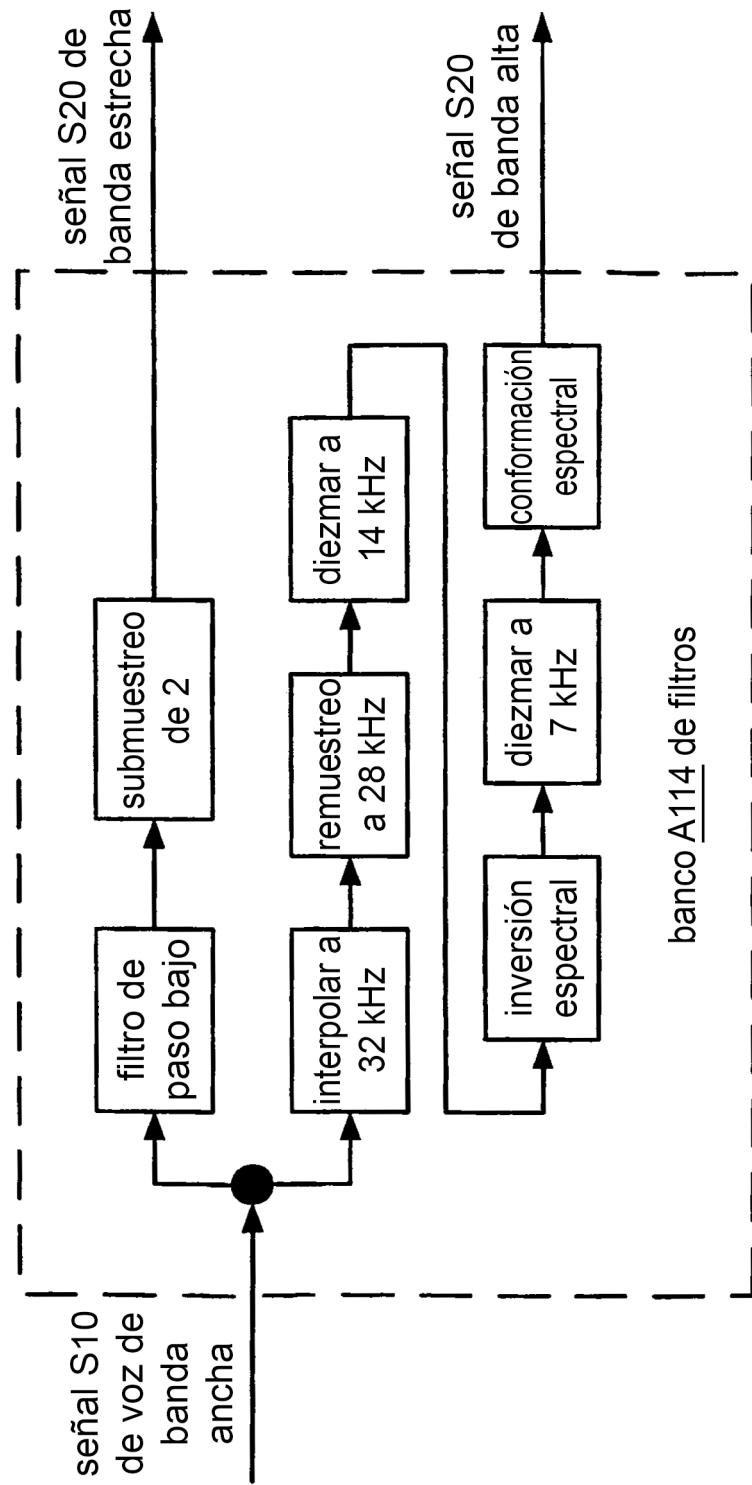


FIG. 4c

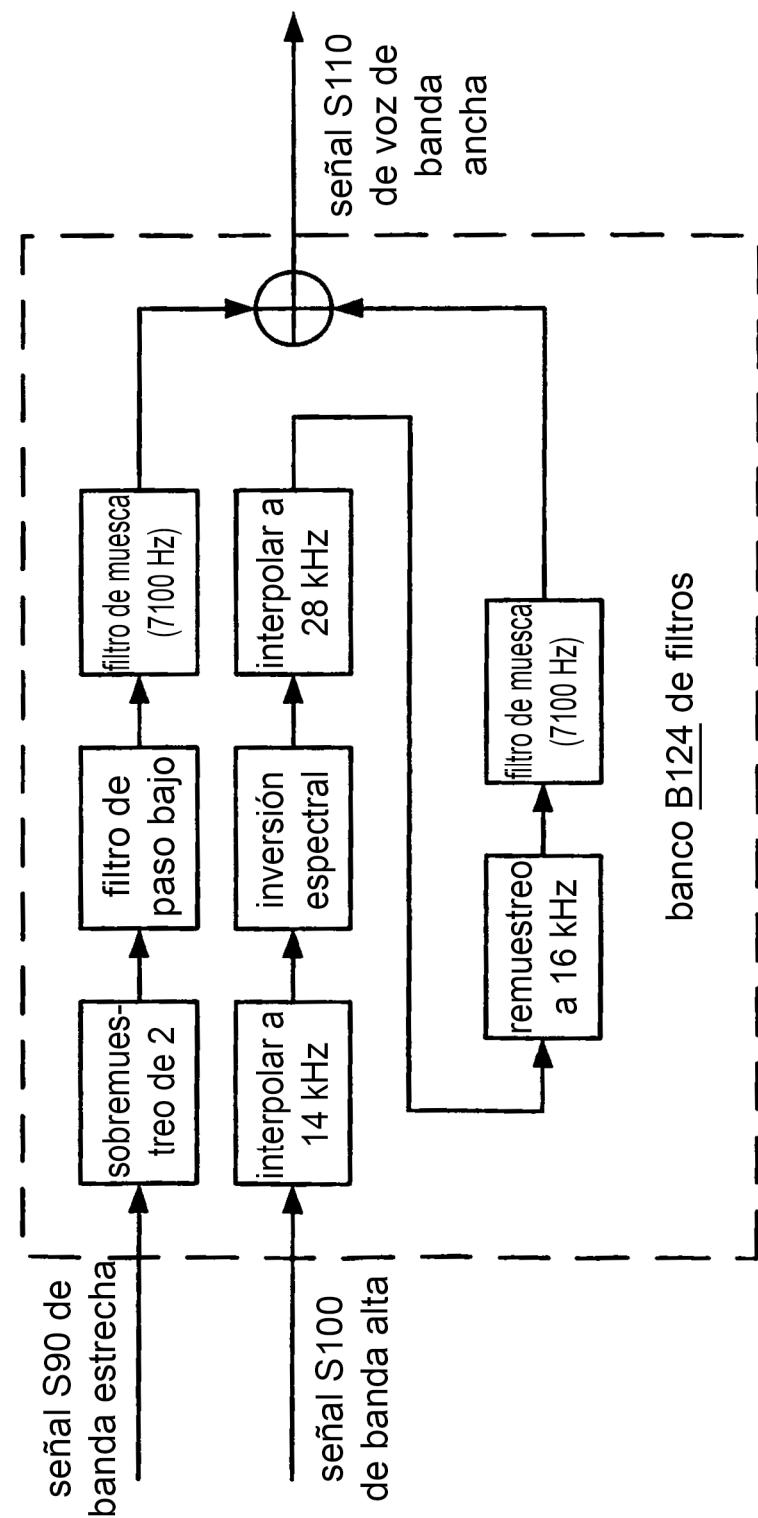


FIG. 4d

FIG. 5a

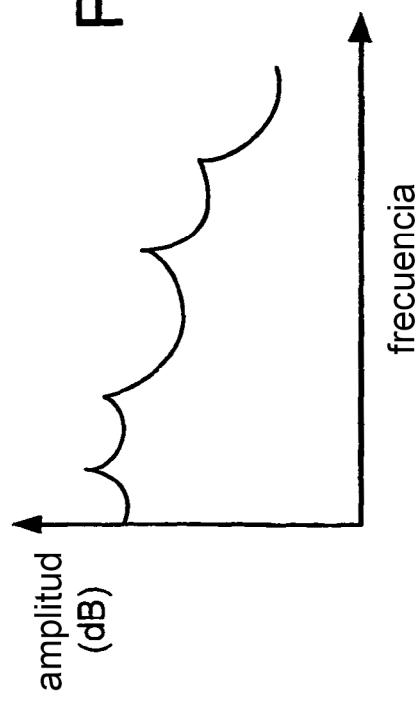
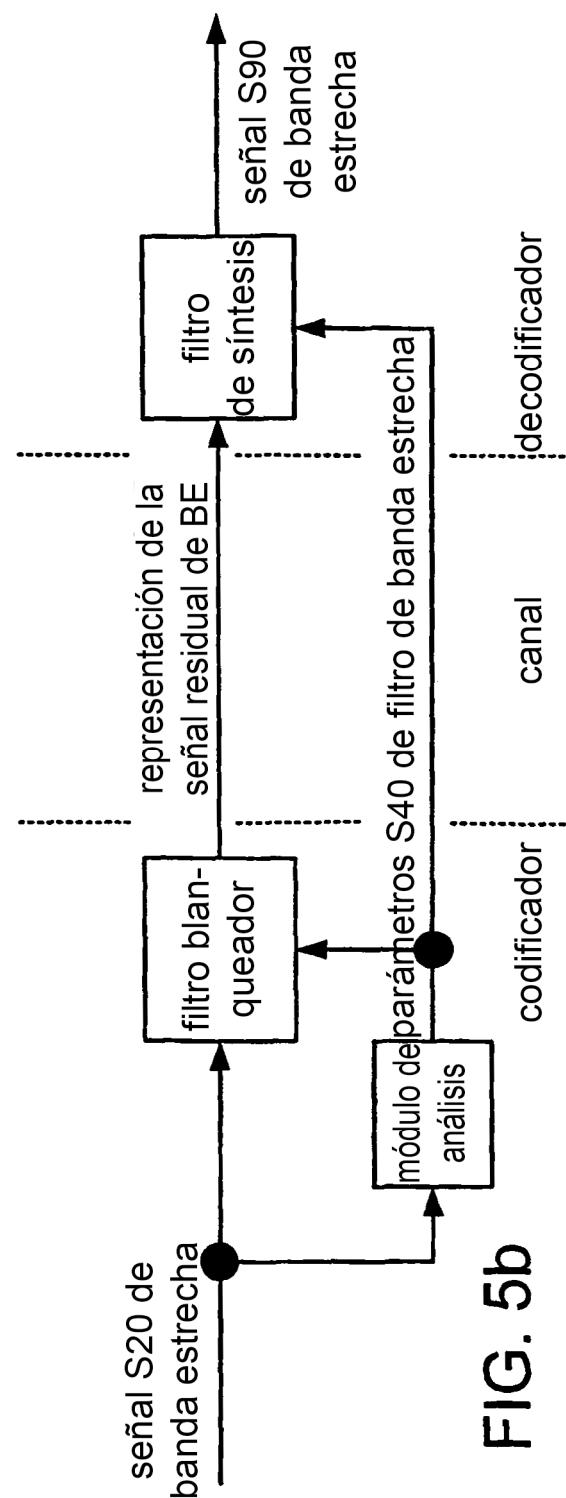


FIG. 5b



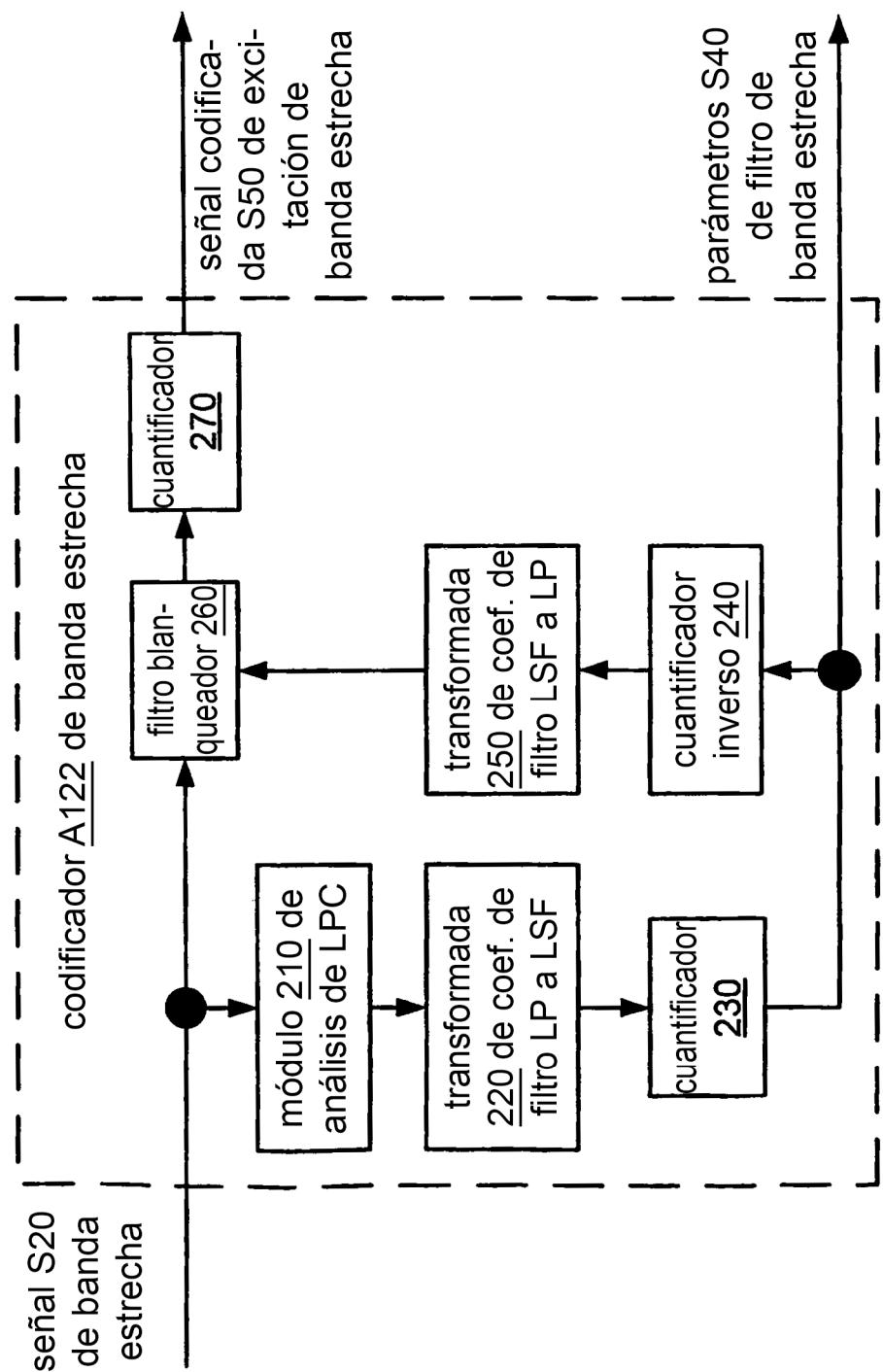


FIG. 6

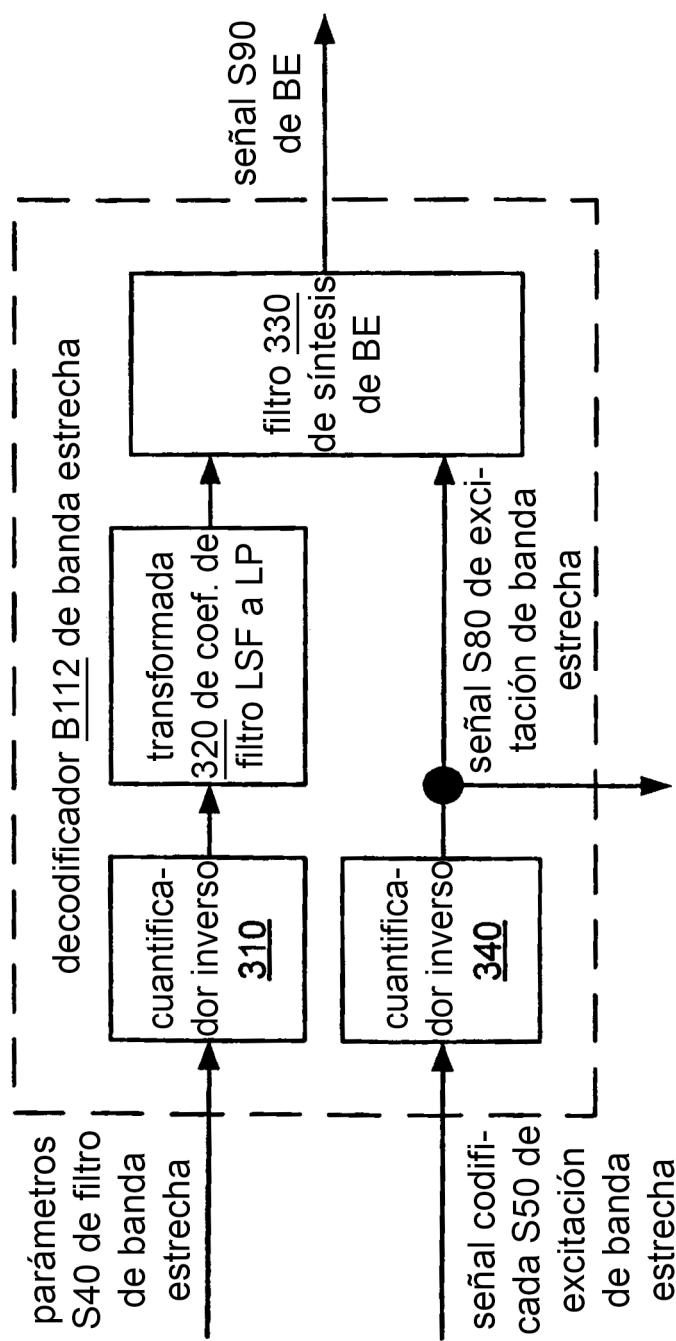
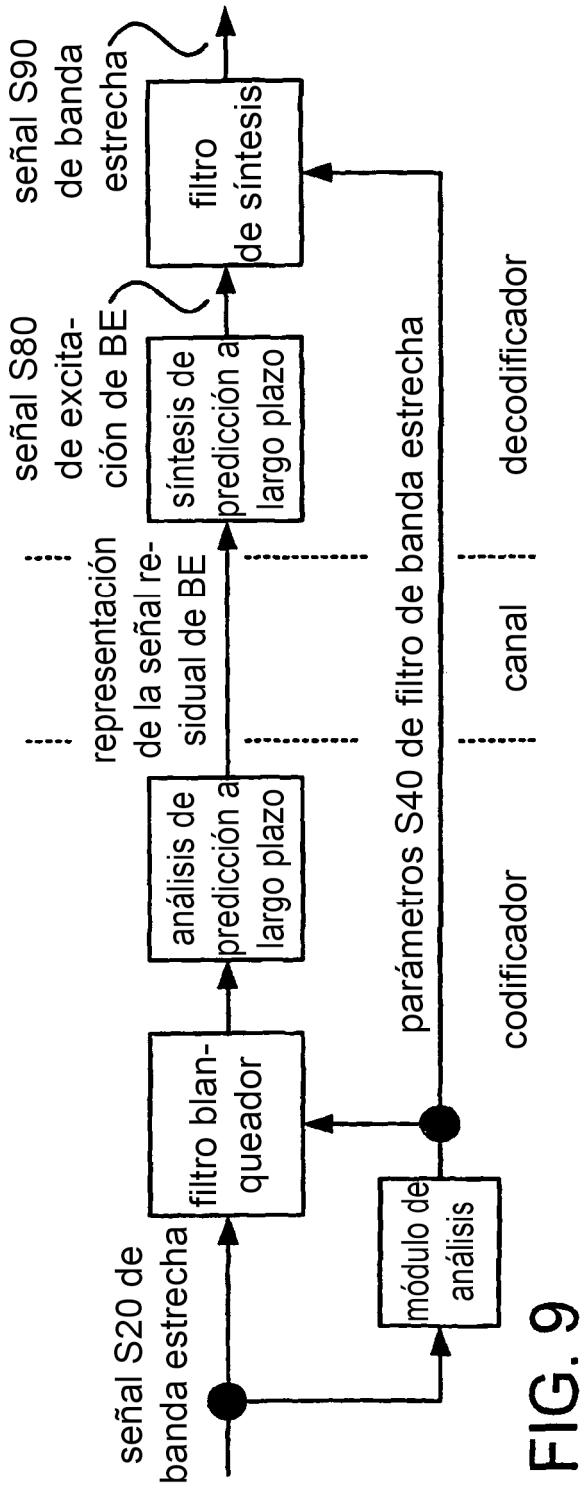
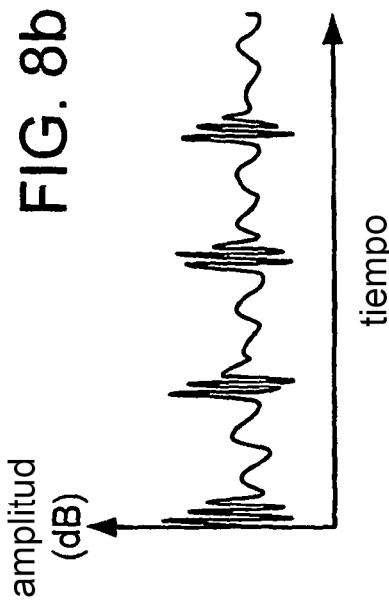
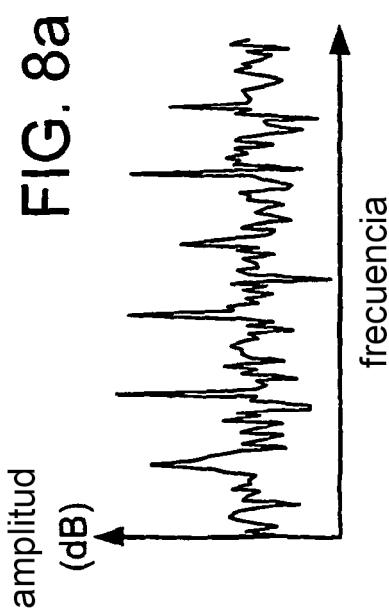


FIG. 7



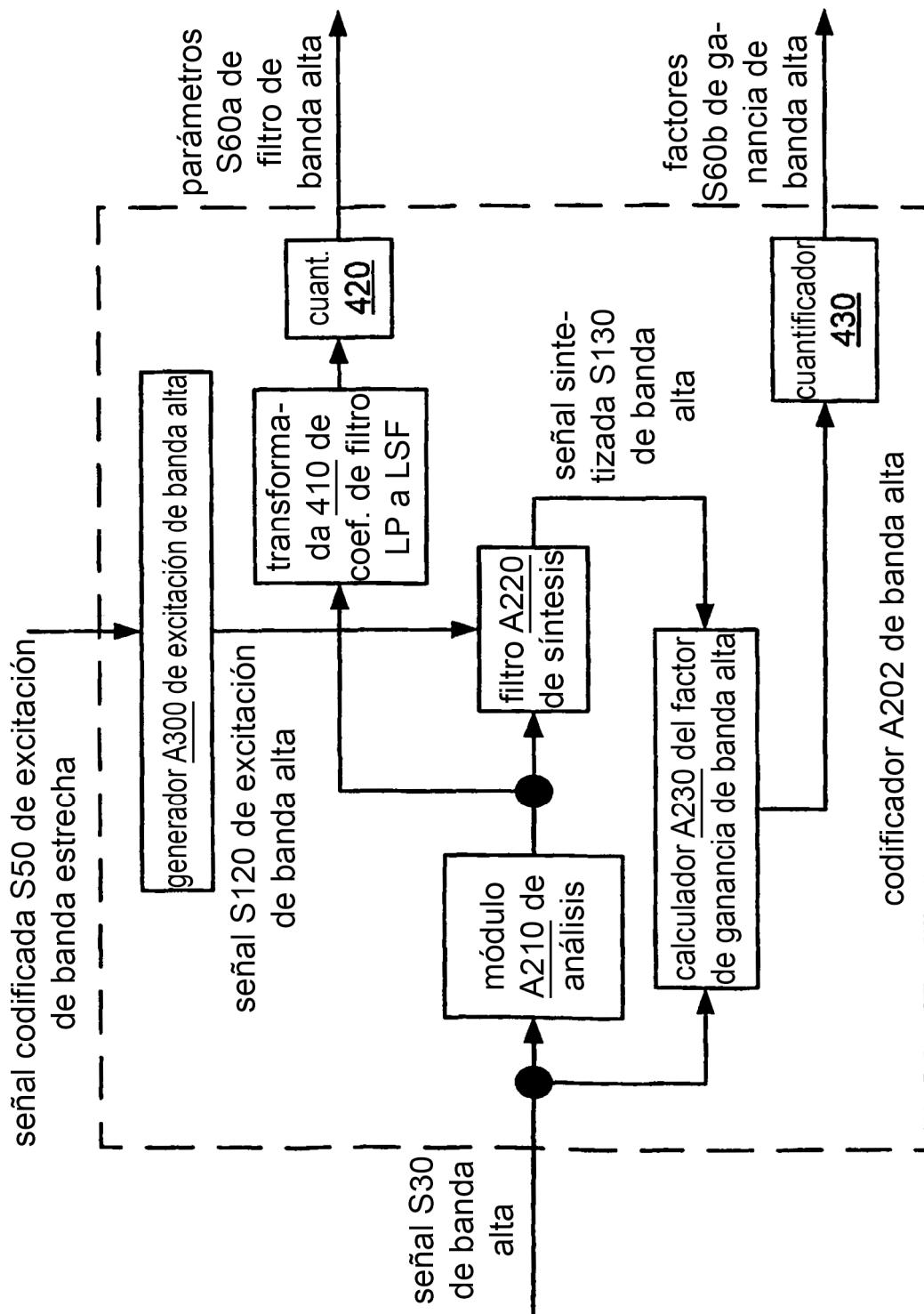


FIG. 10

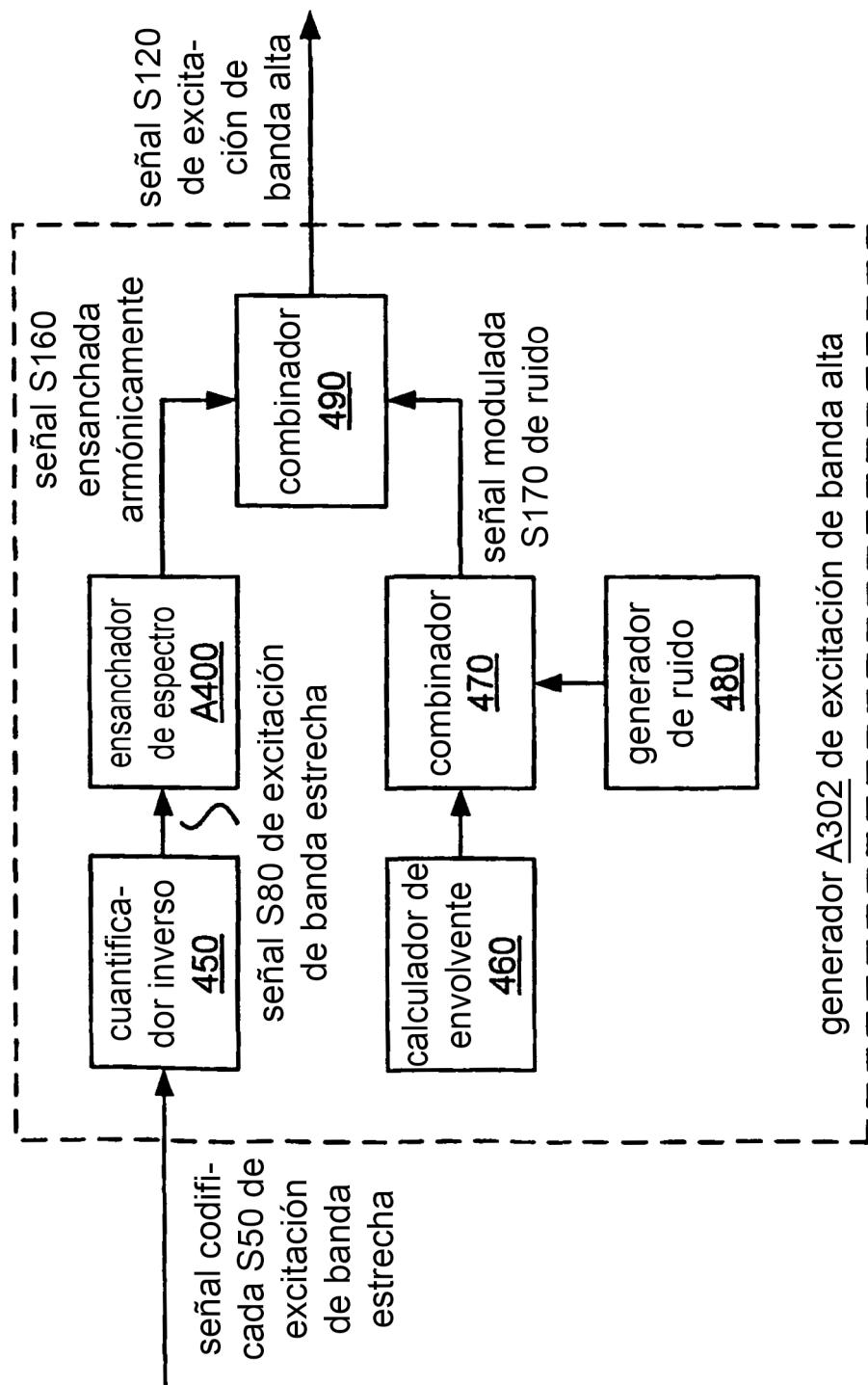


FIG. 11

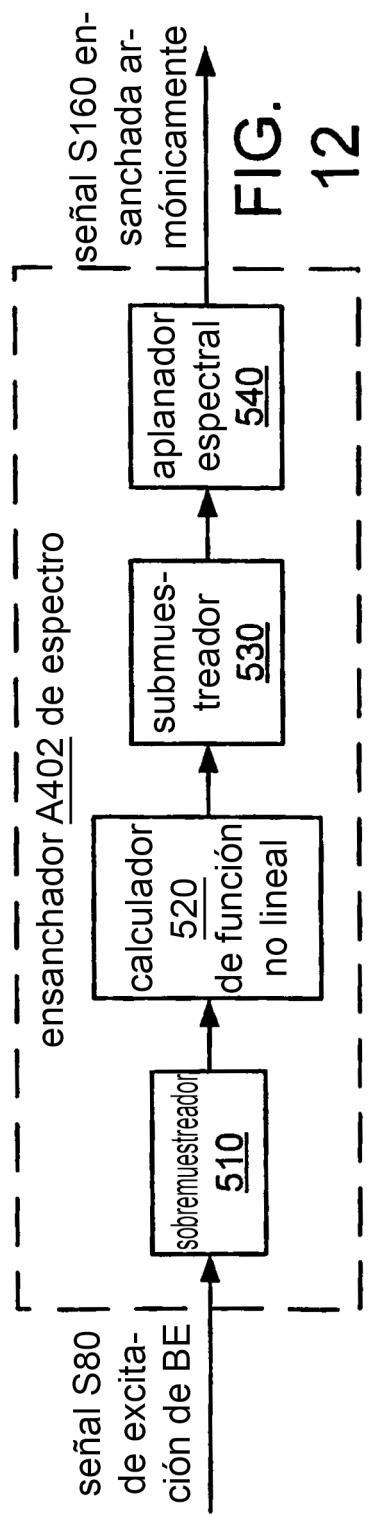


FIG. 12

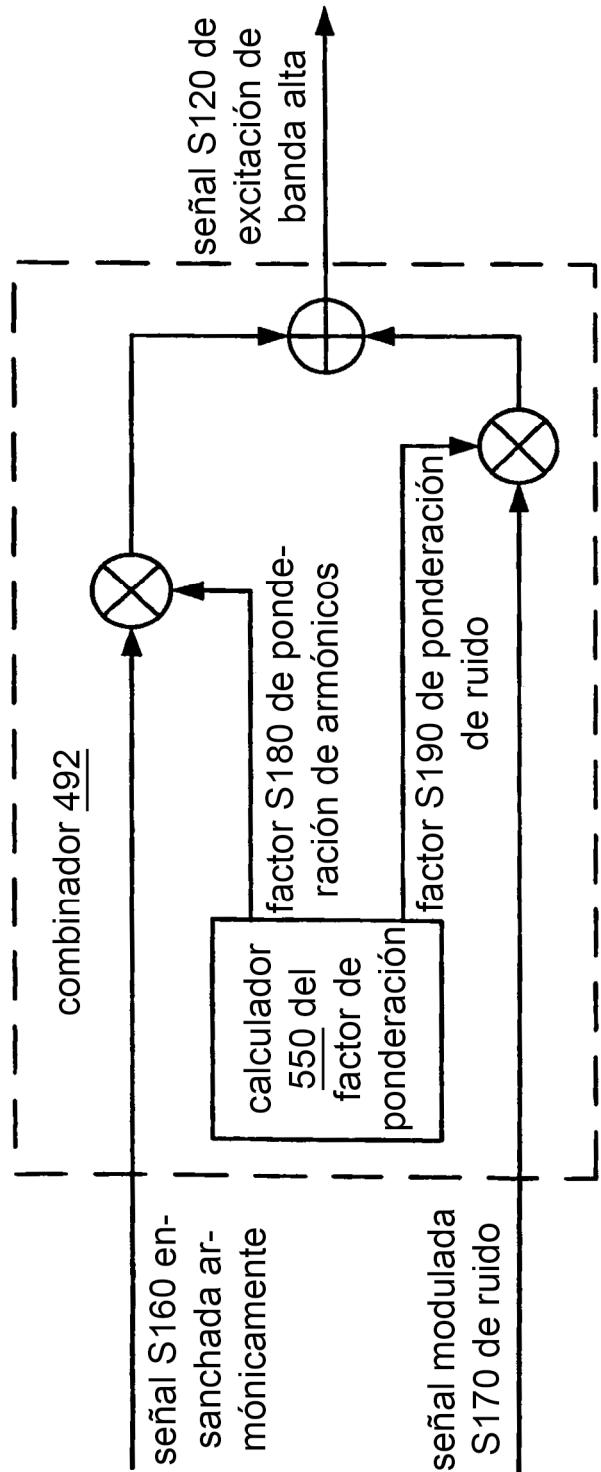


FIG. 16

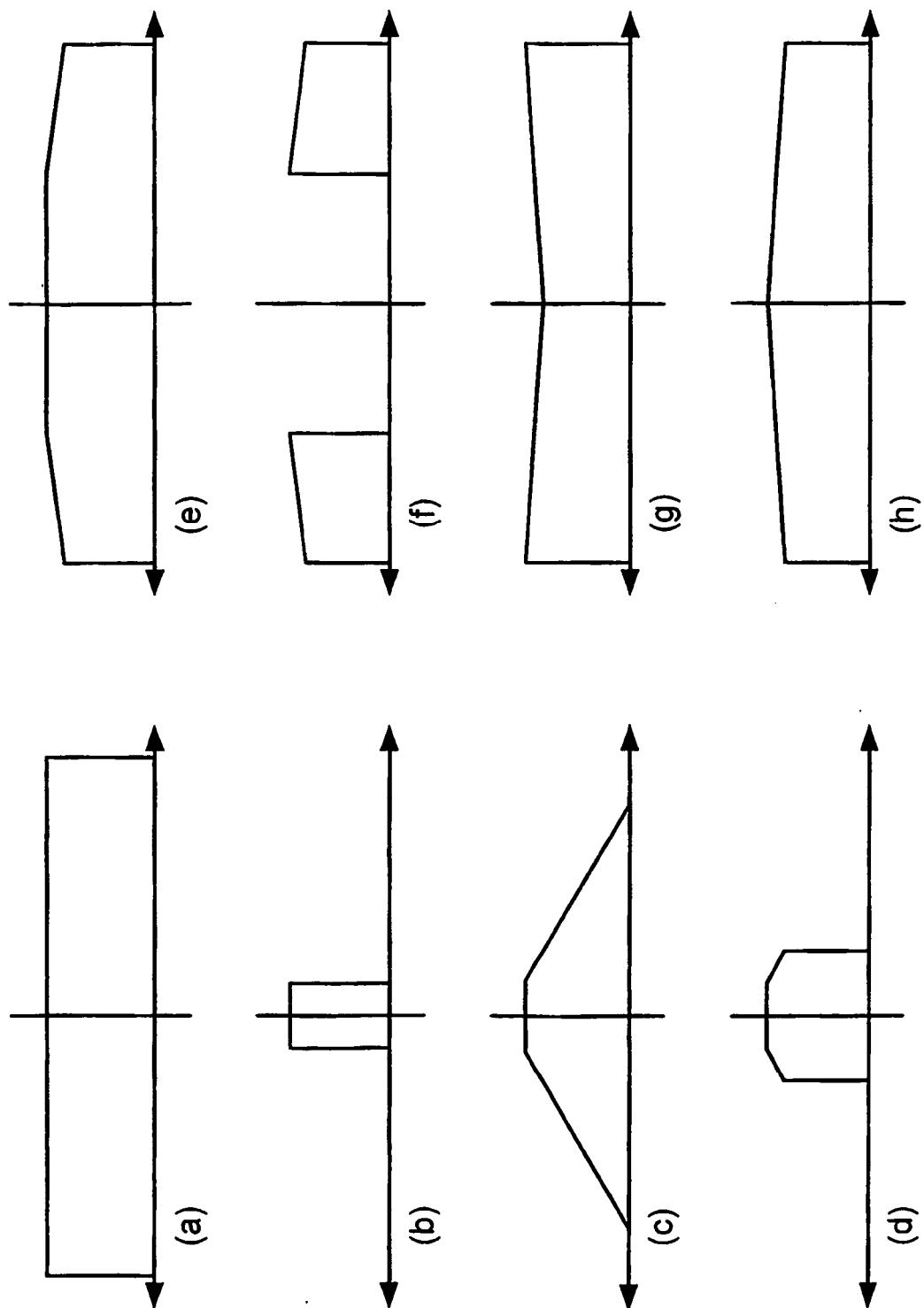
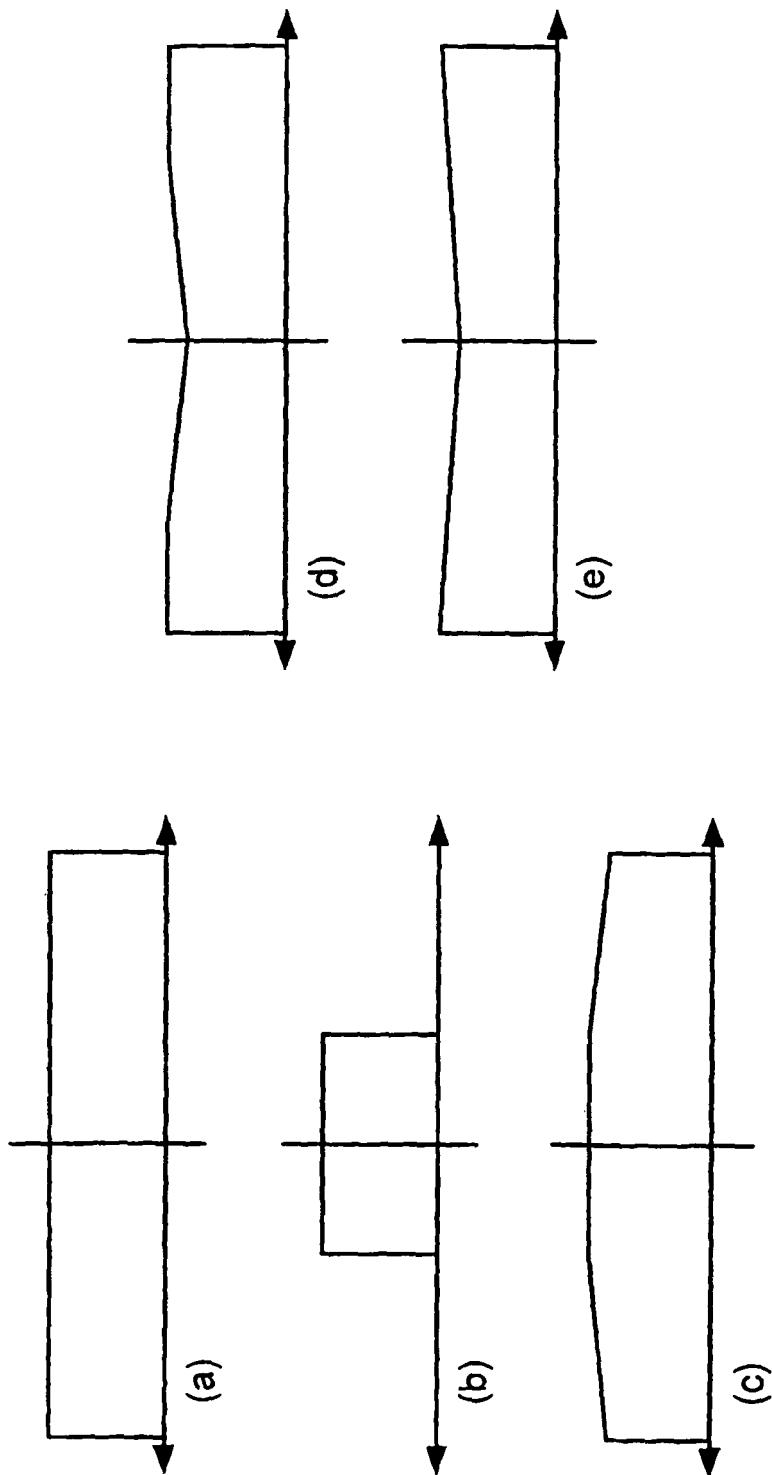


FIG. 12a

FIG. 12b



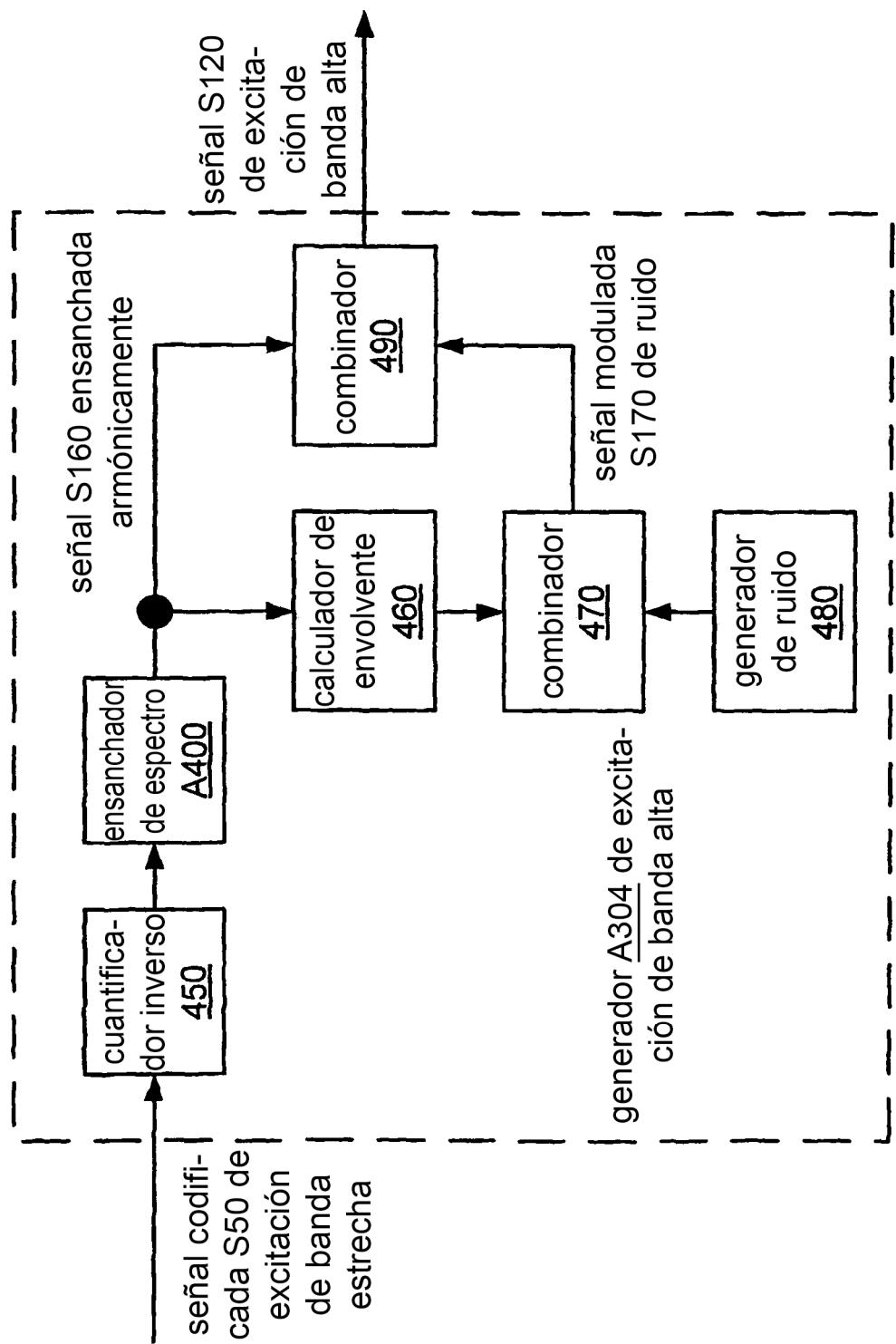


FIG. 13

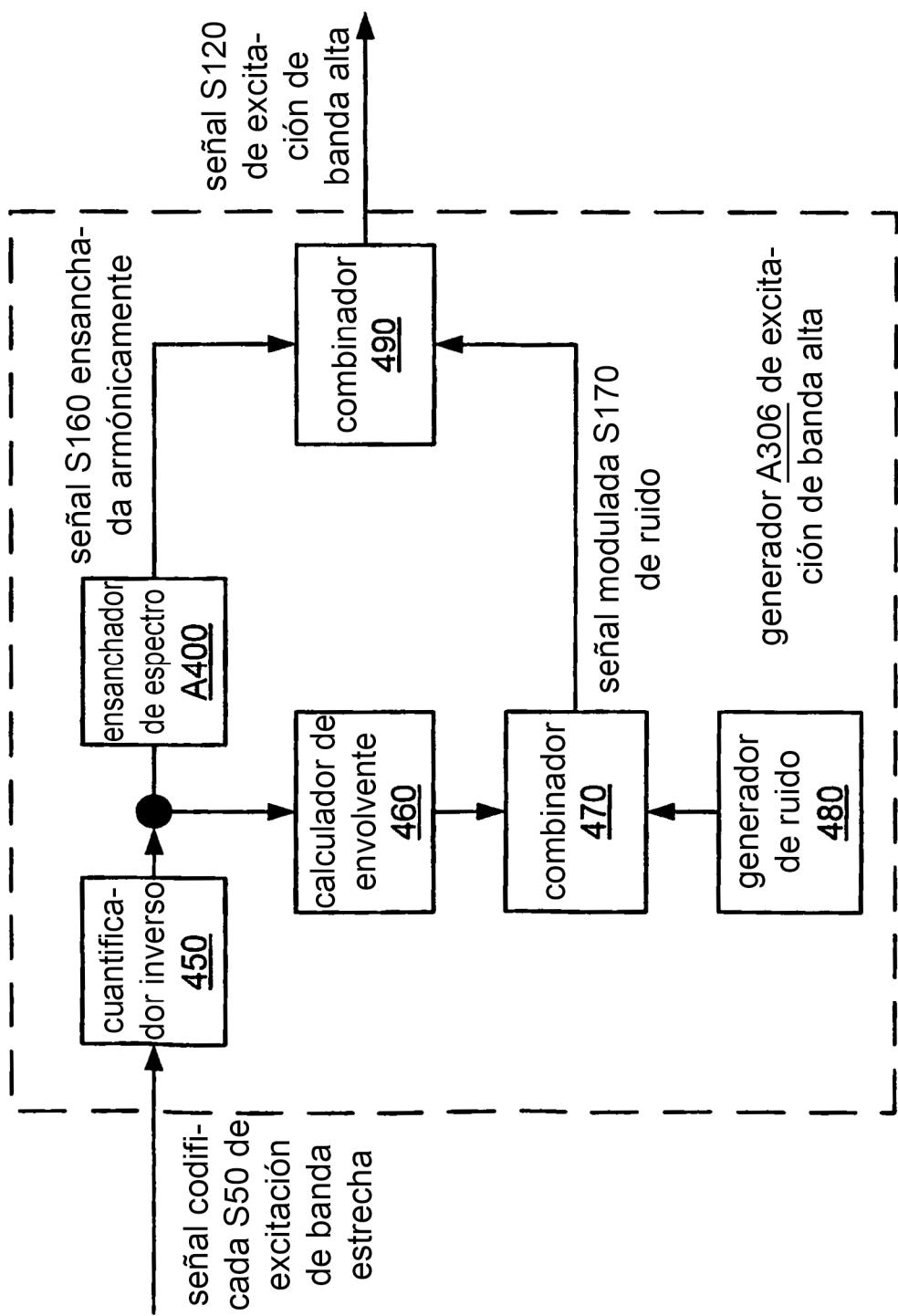


FIG. 14

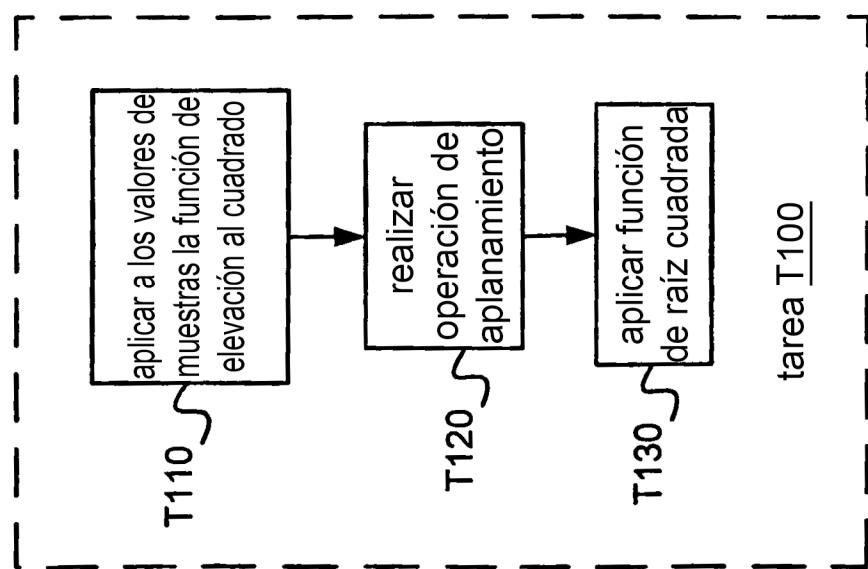


FIG. 15

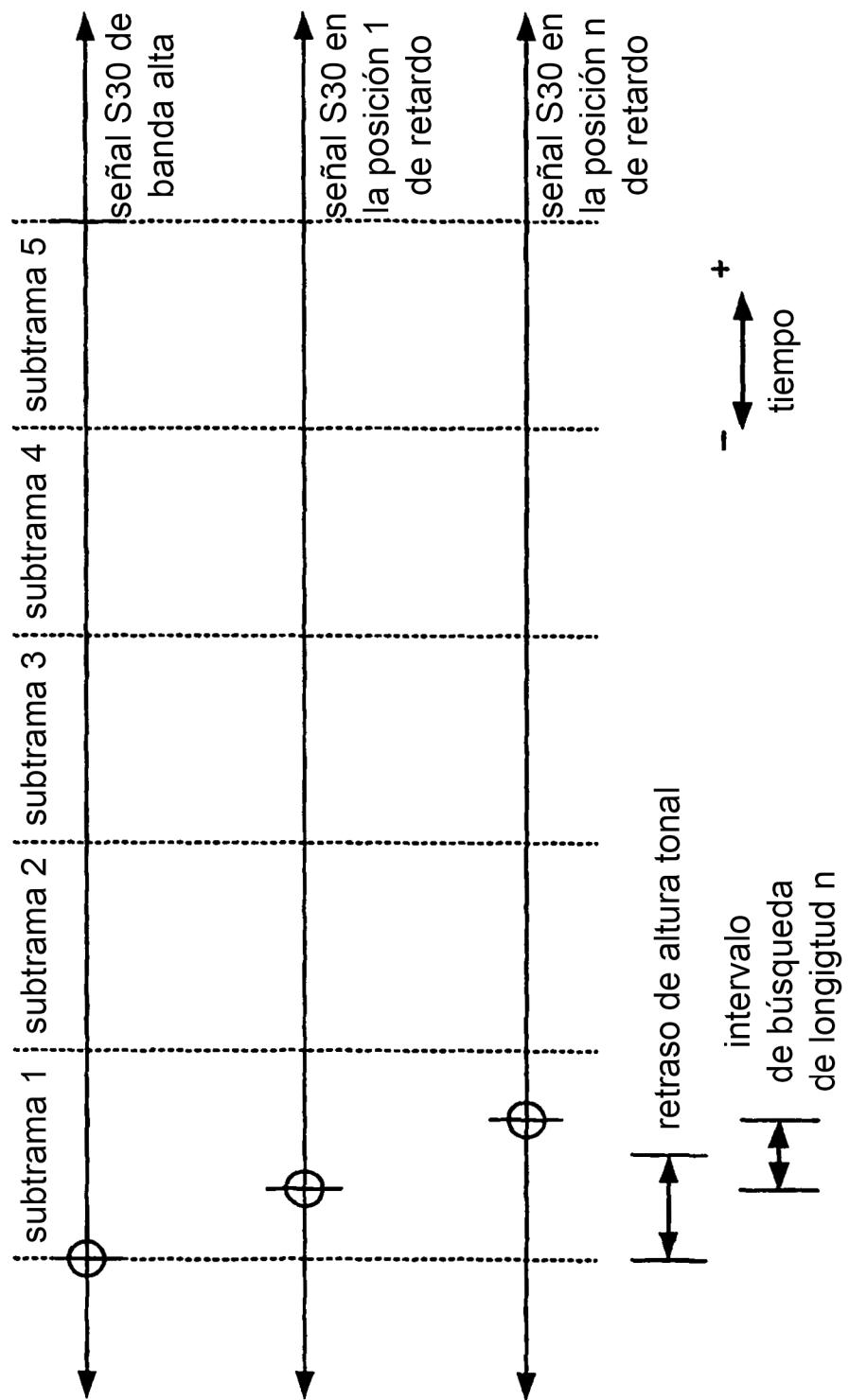


FIG. 17

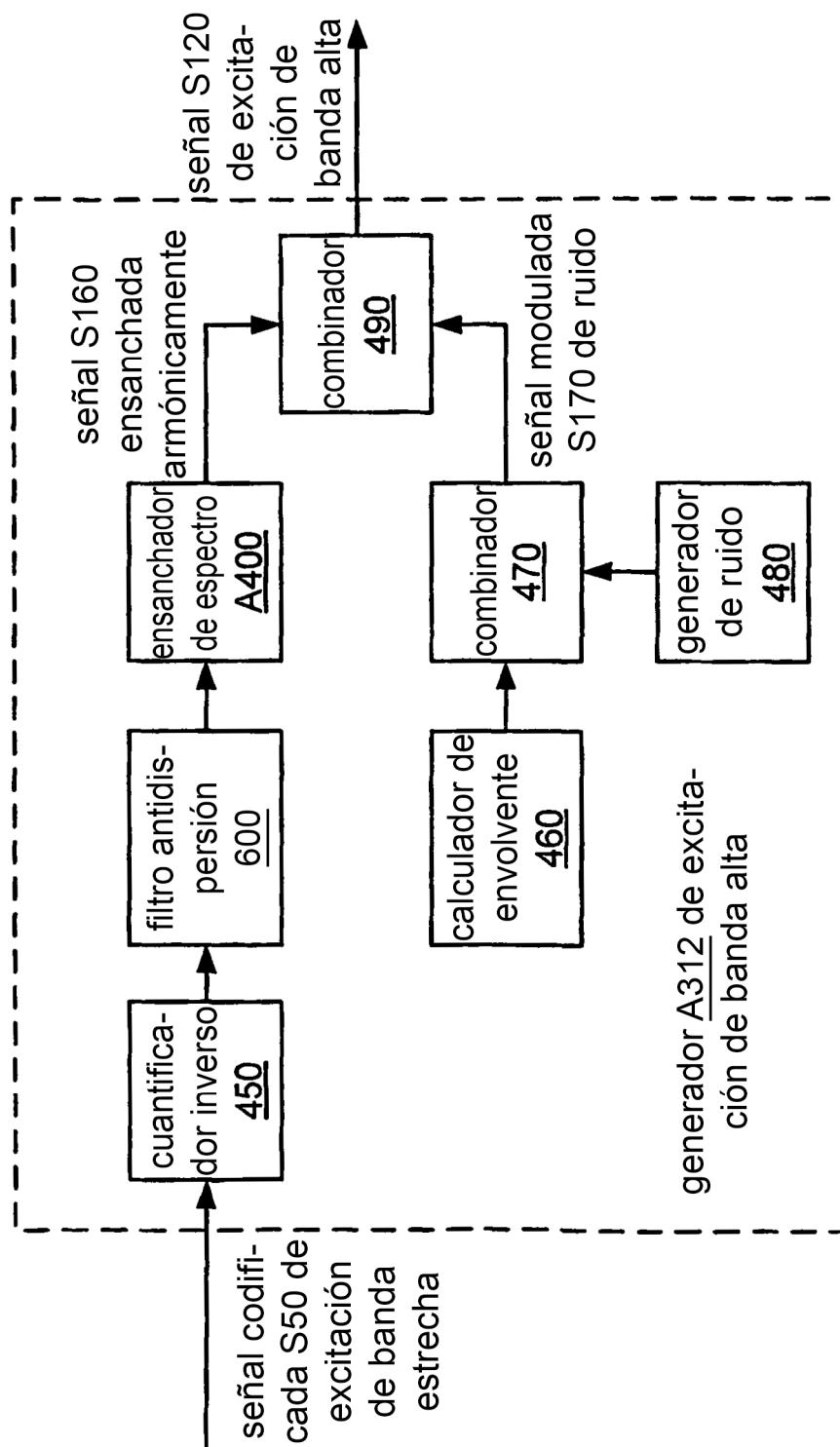


FIG. 18

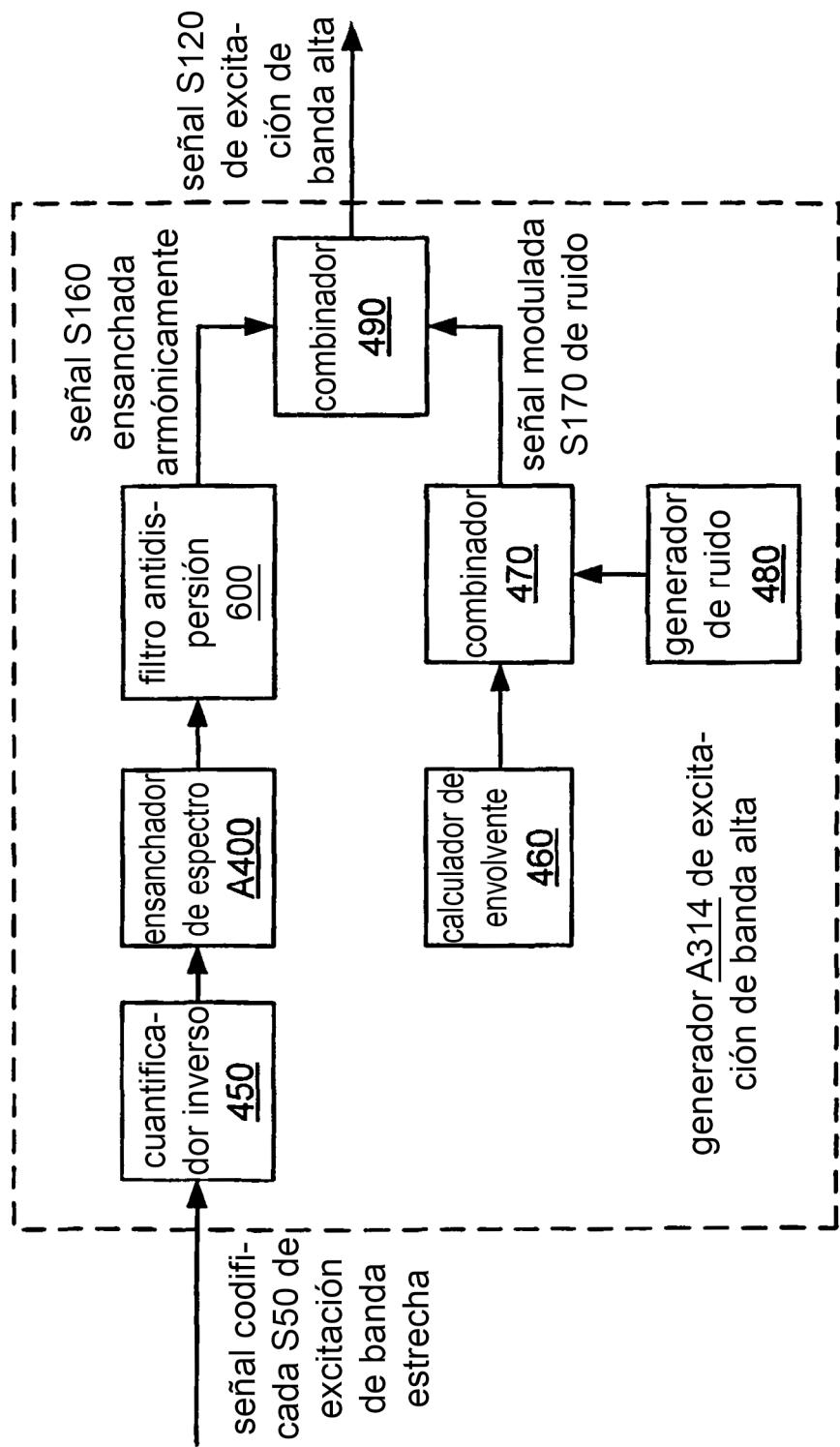


FIG. 19

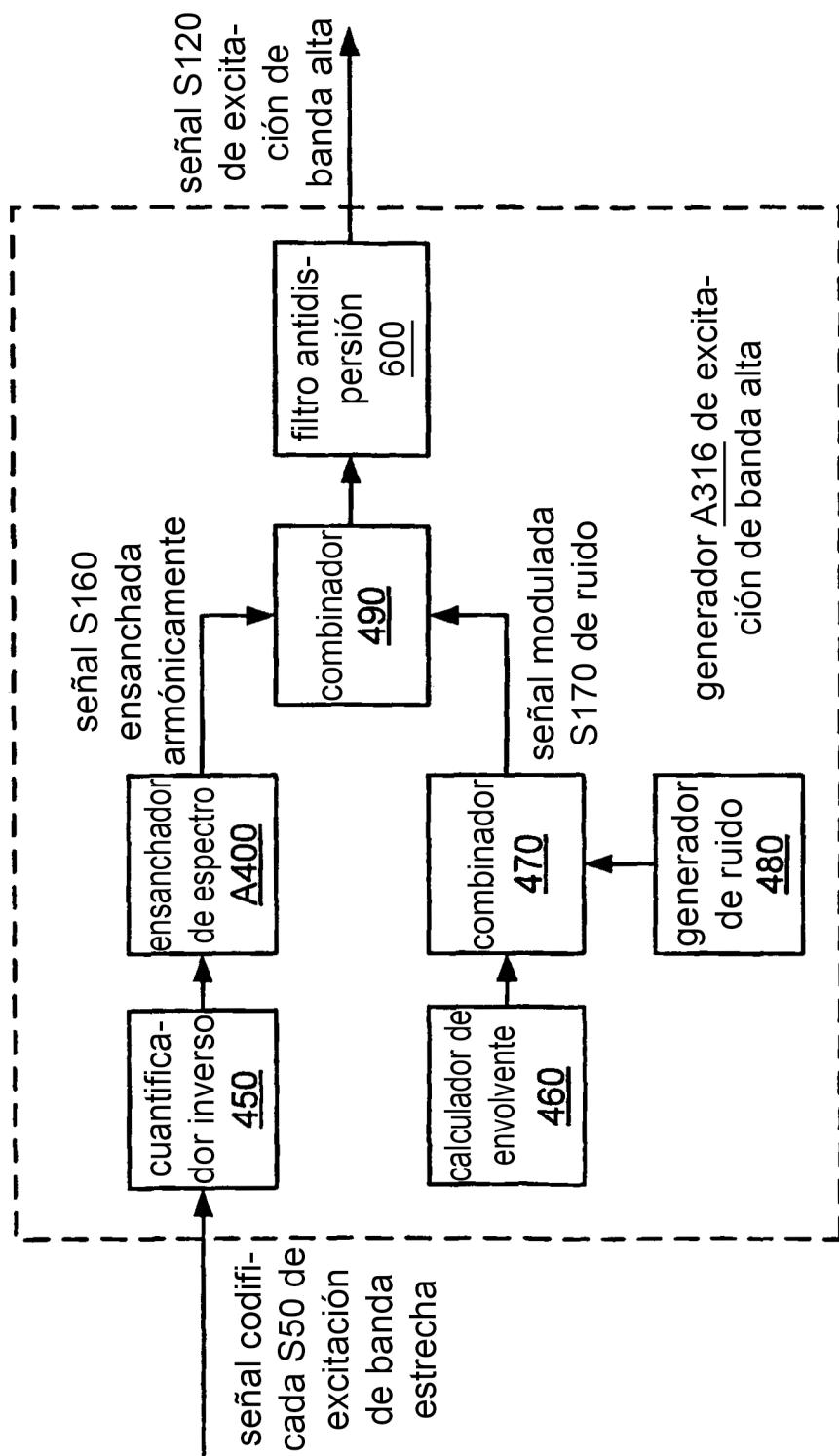


FIG. 20

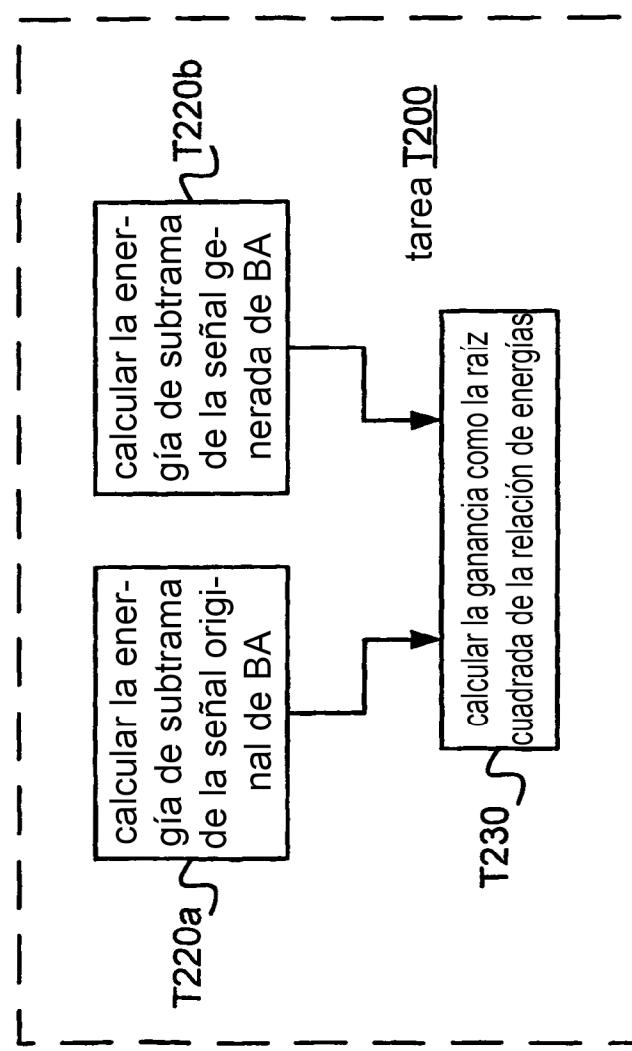


FIG. 21

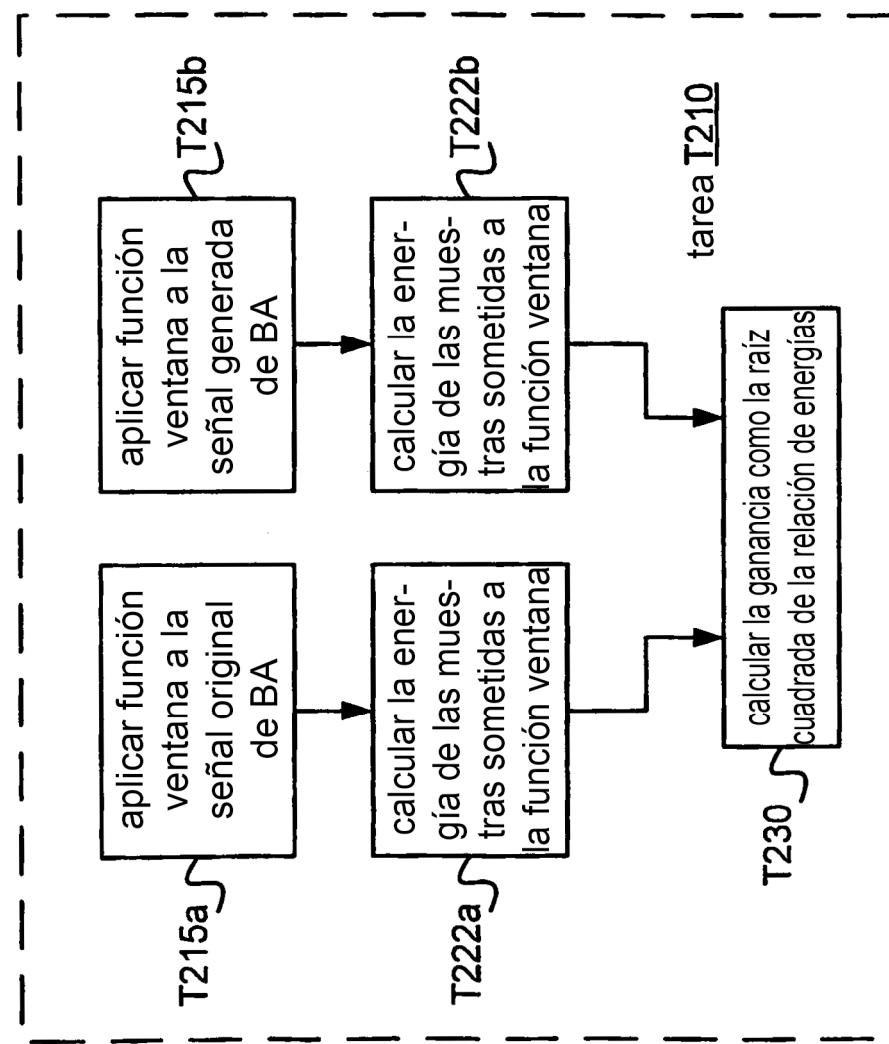


FIG. 22

FIG. 23b

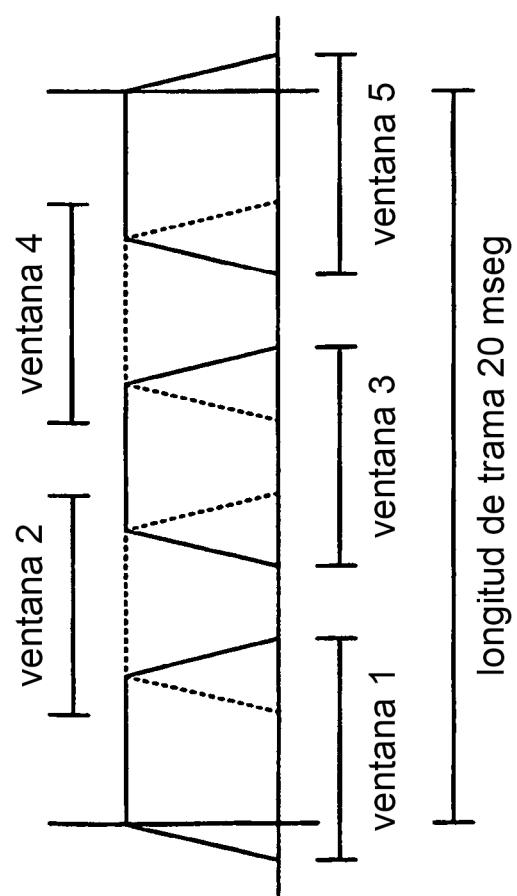


FIG. 23a

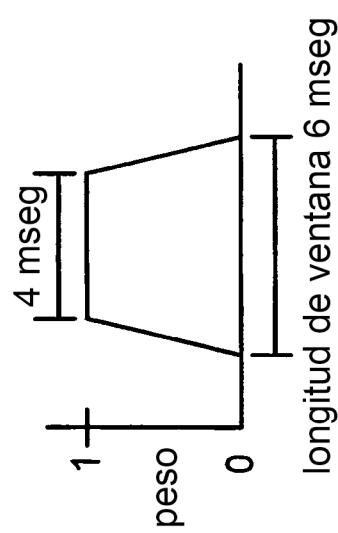
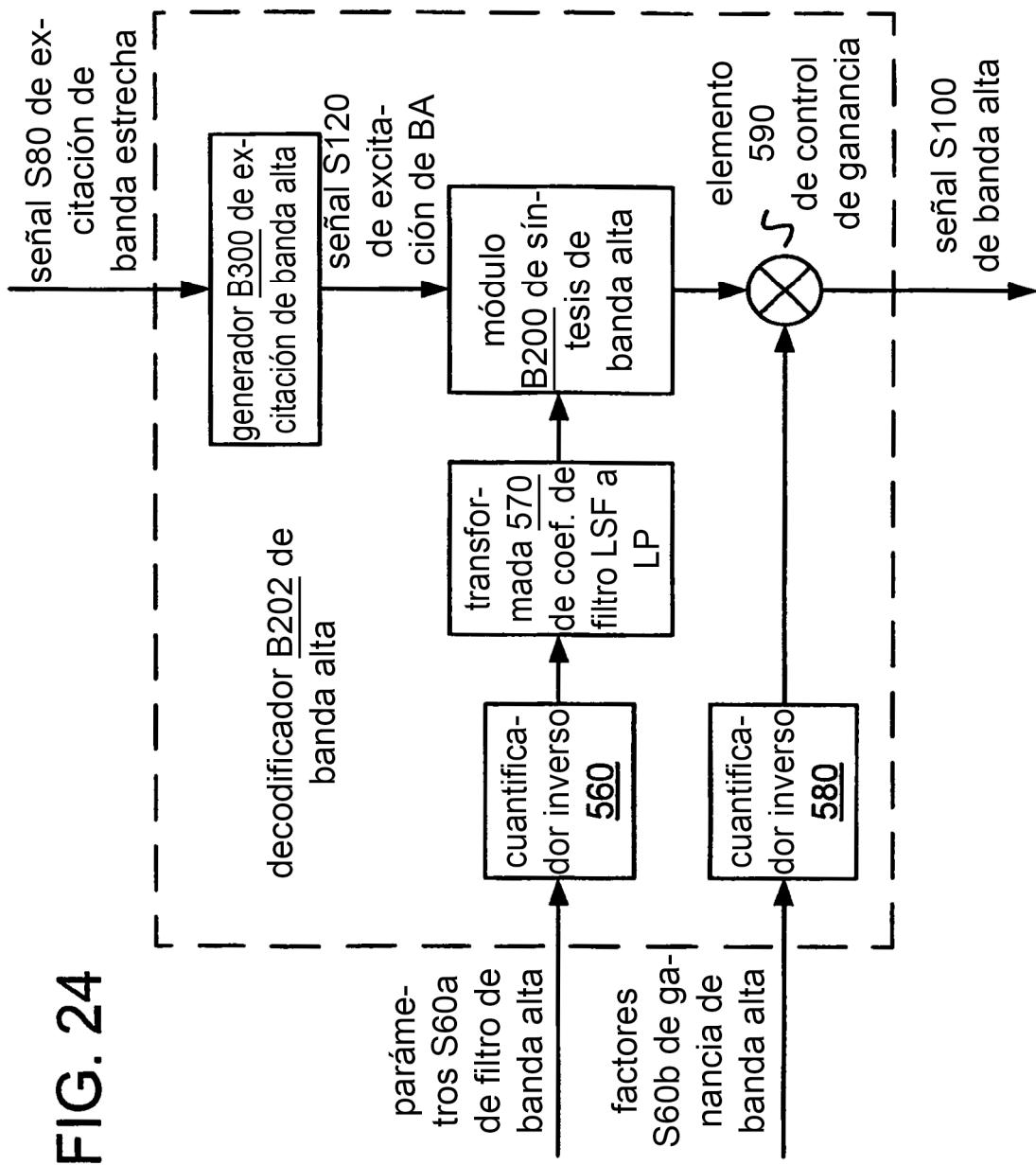


FIG. 24



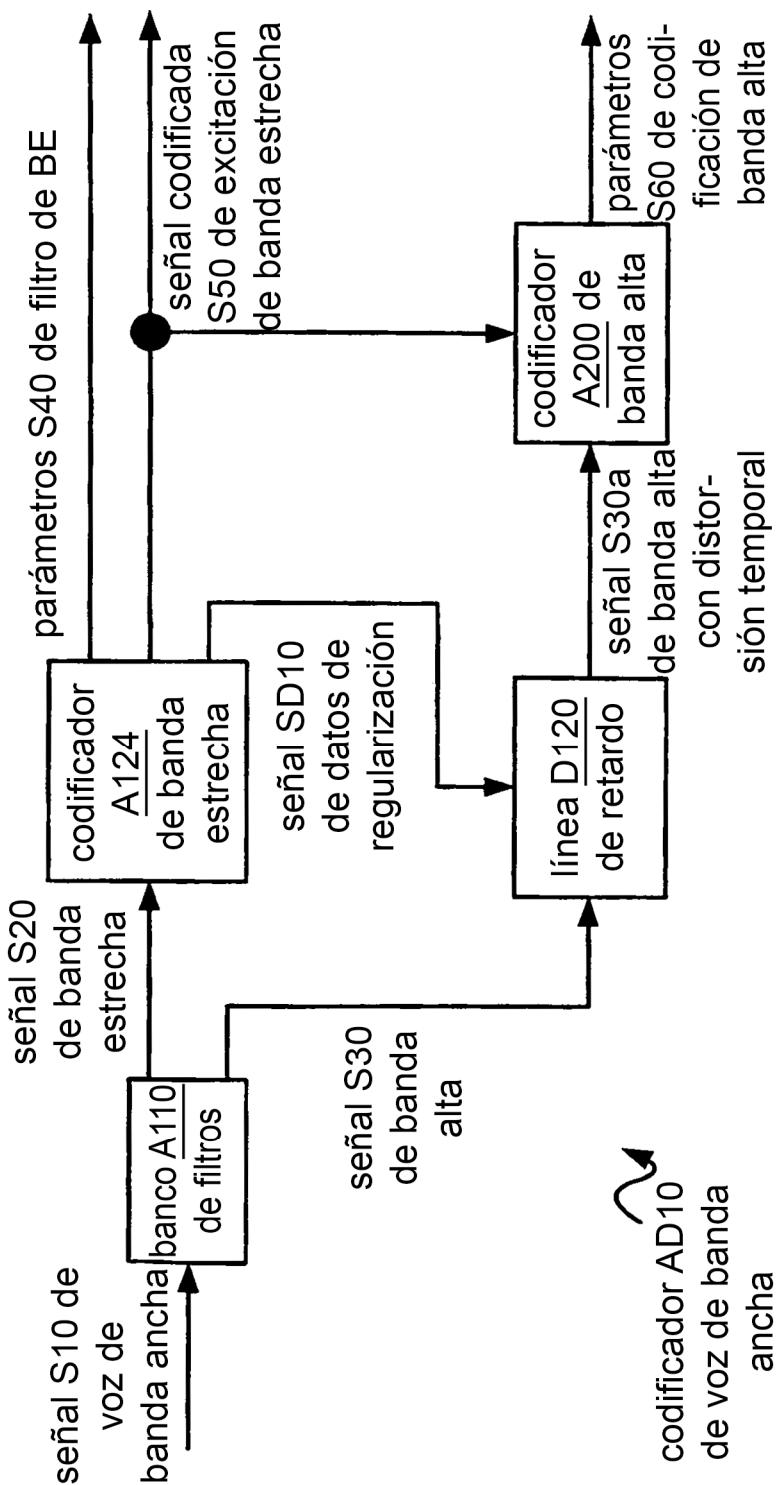


FIG. 25

FIG. 26a

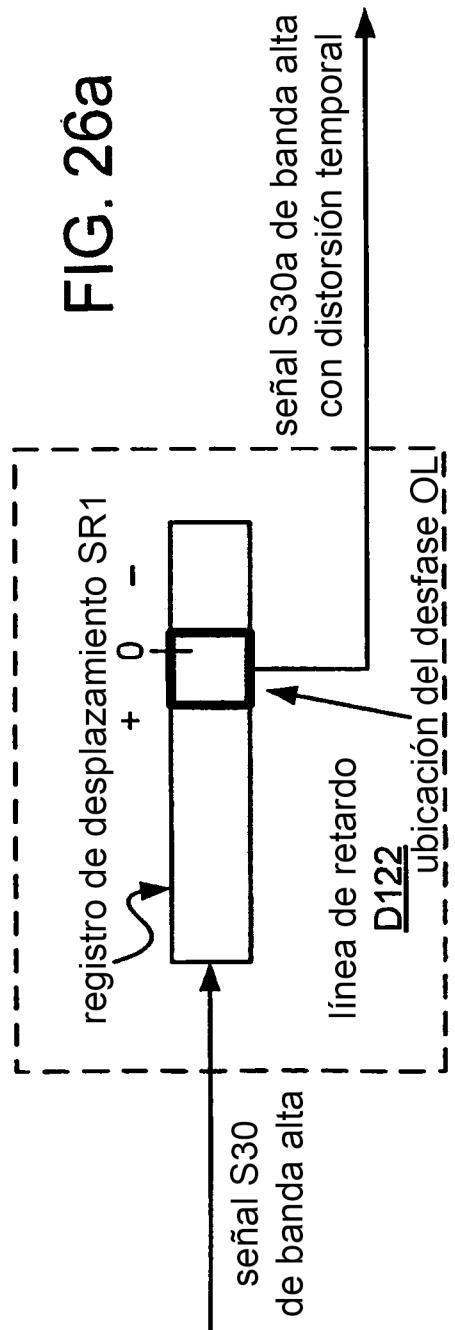
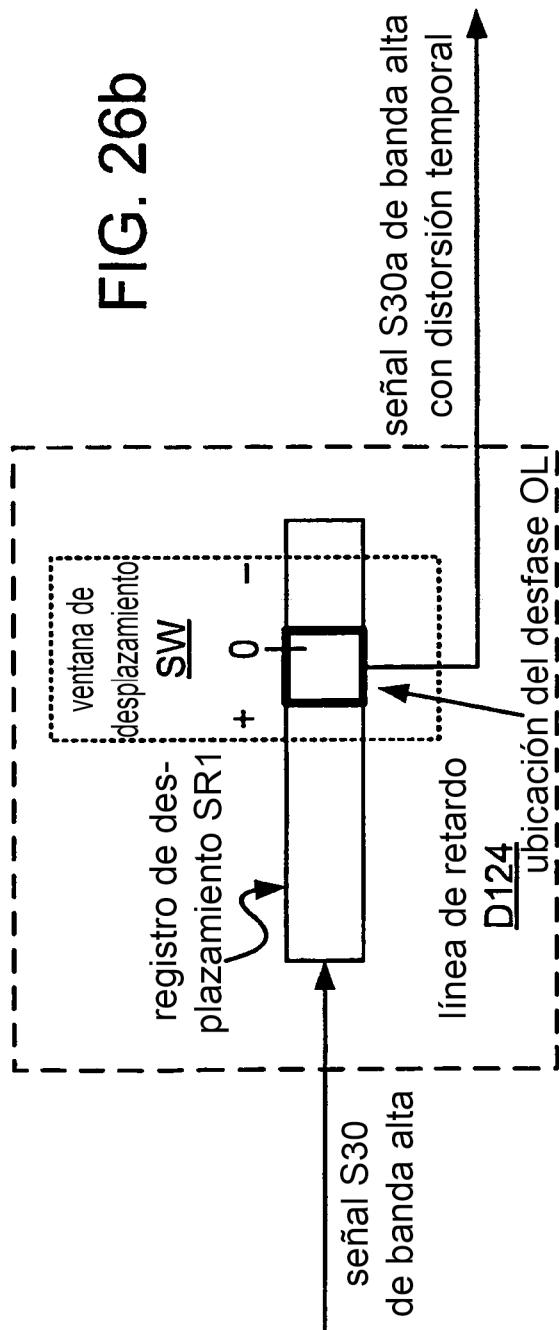


FIG. 26b



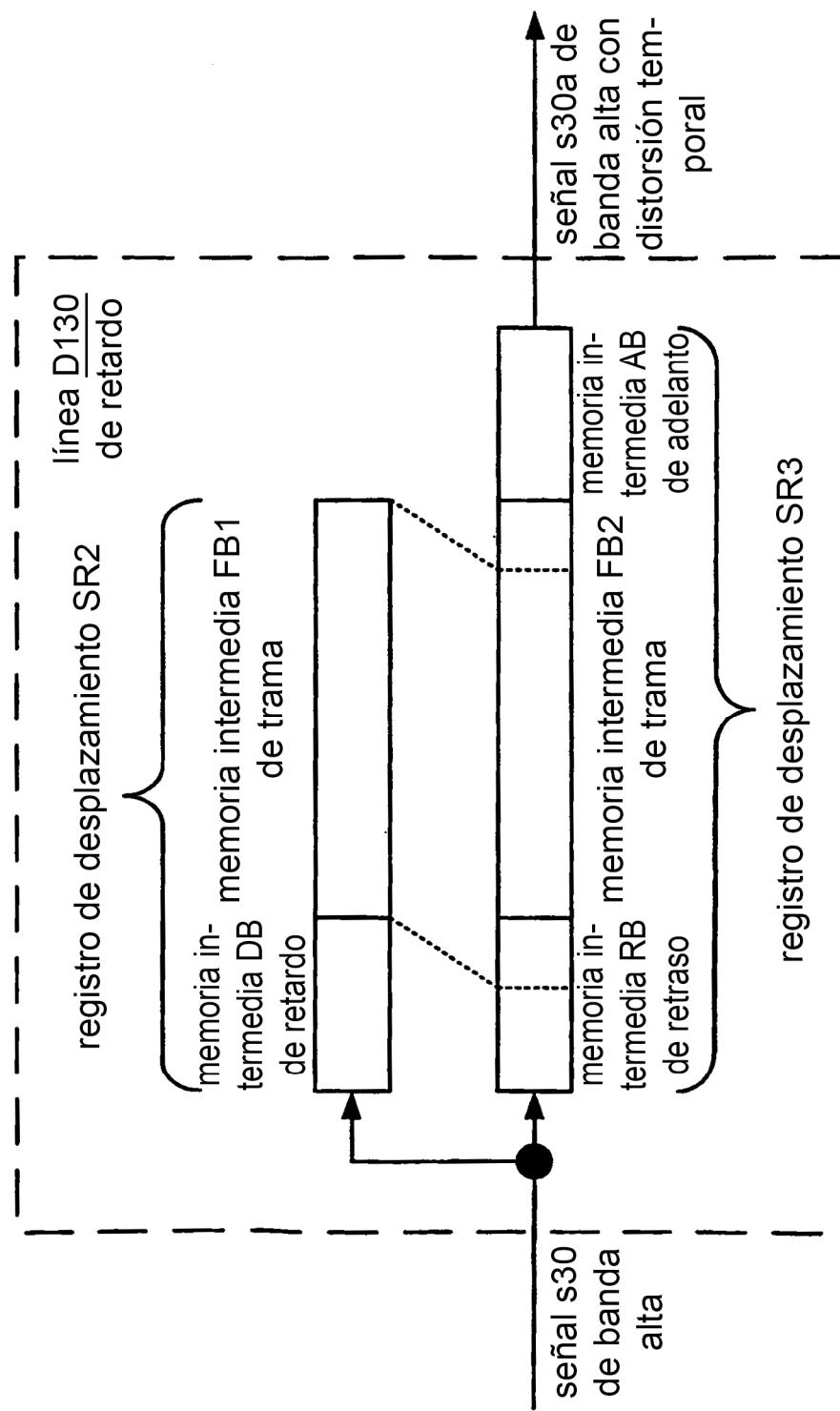


FIG. 27

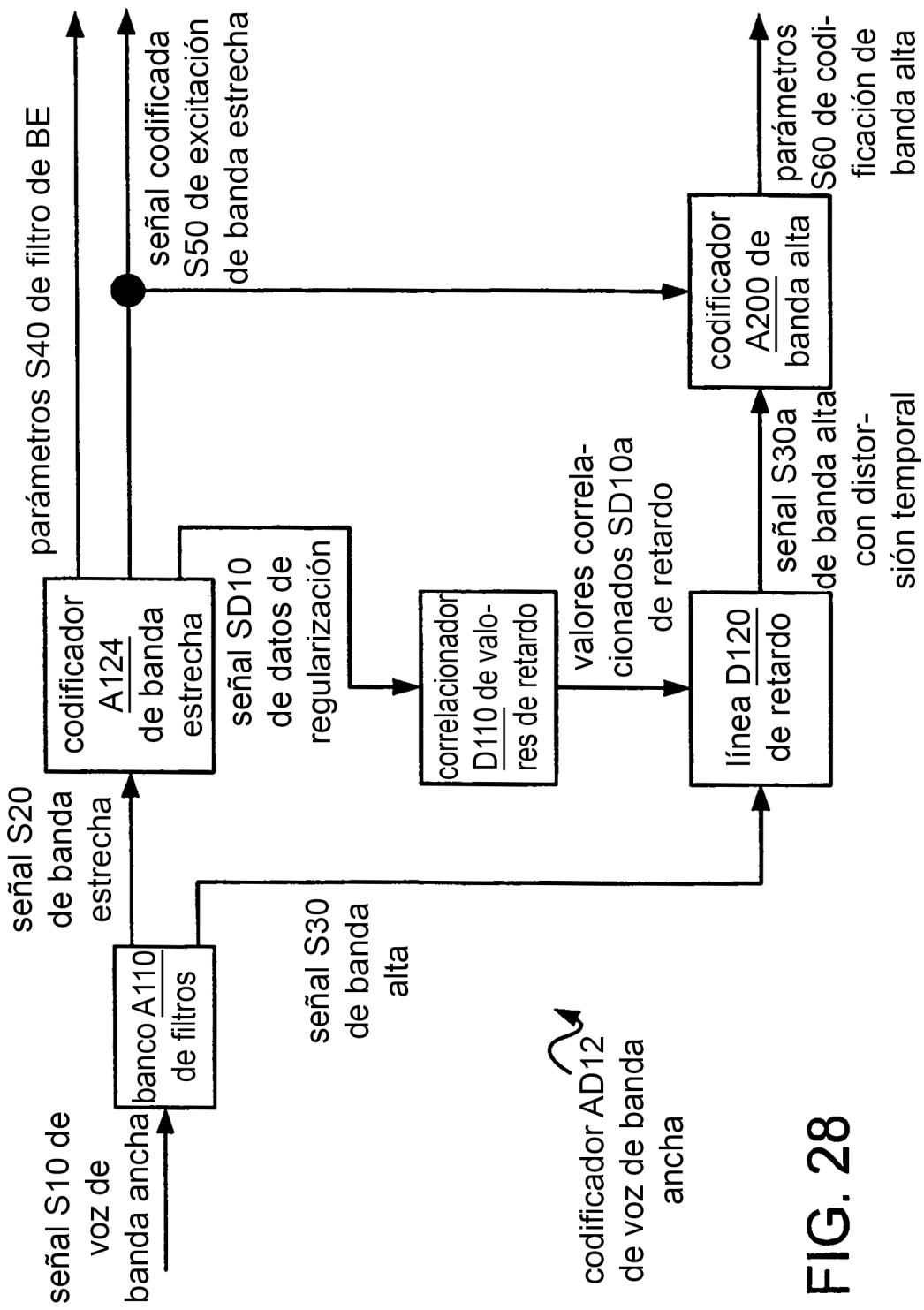


FIG. 28

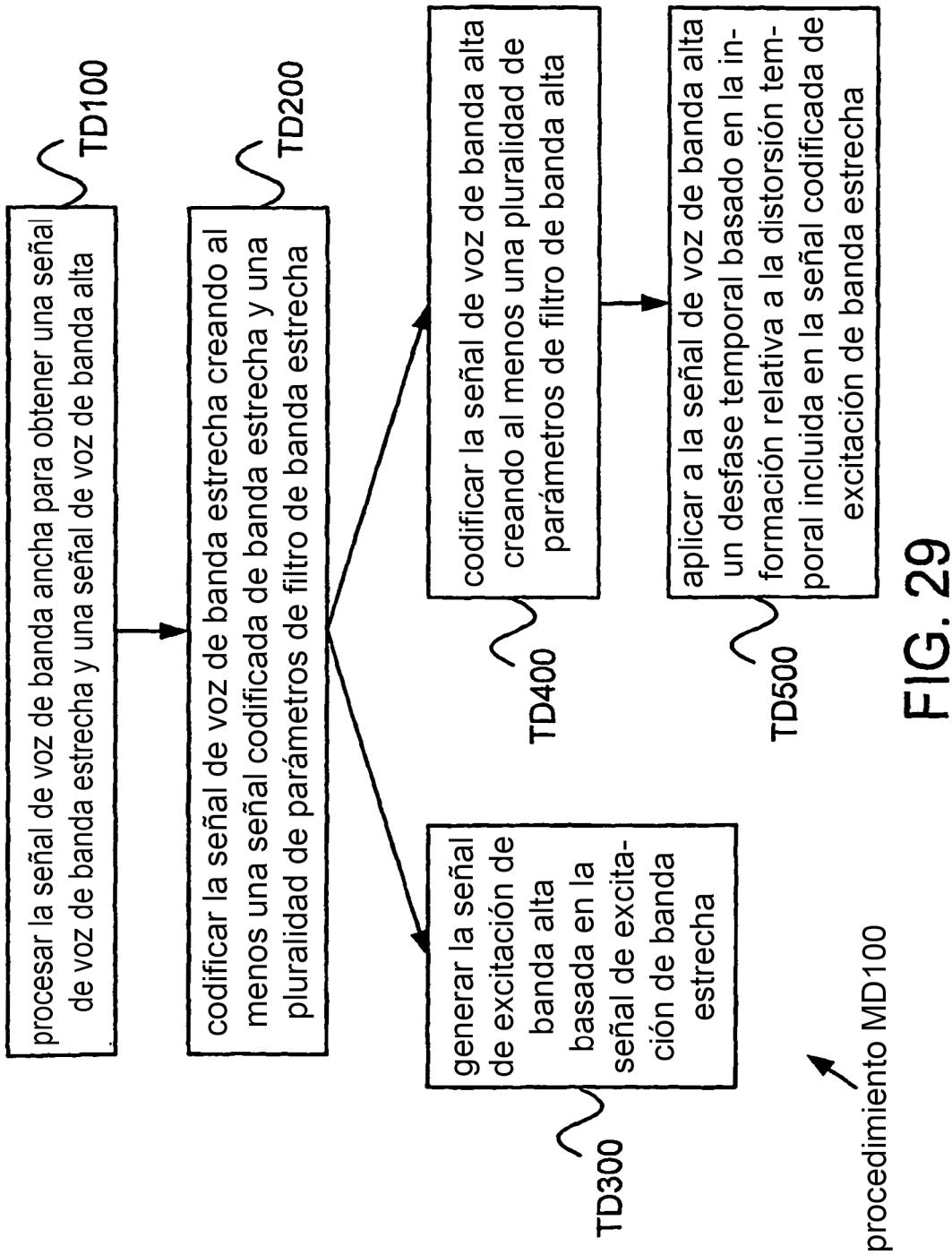


FIG. 29

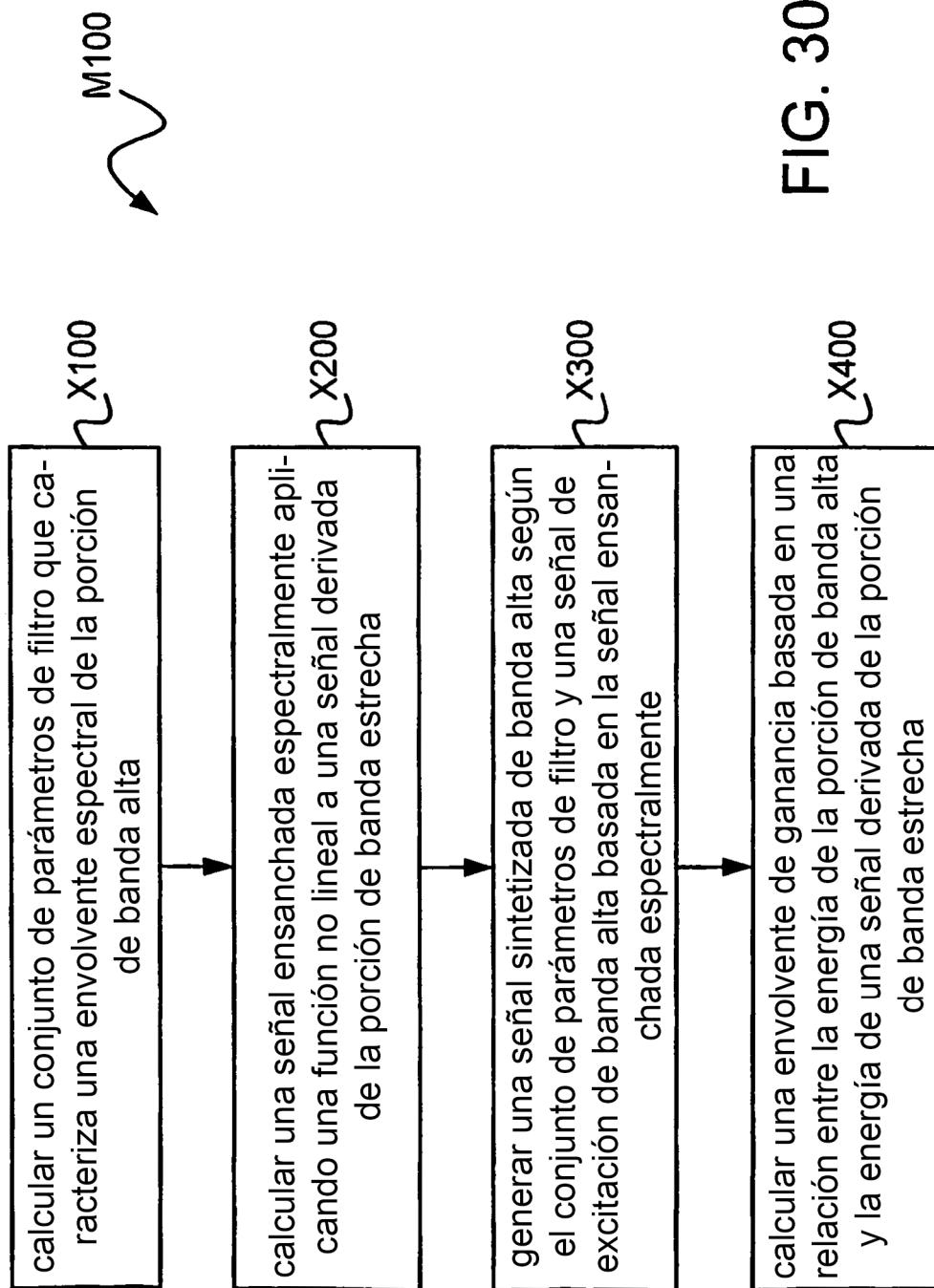


FIG. 30

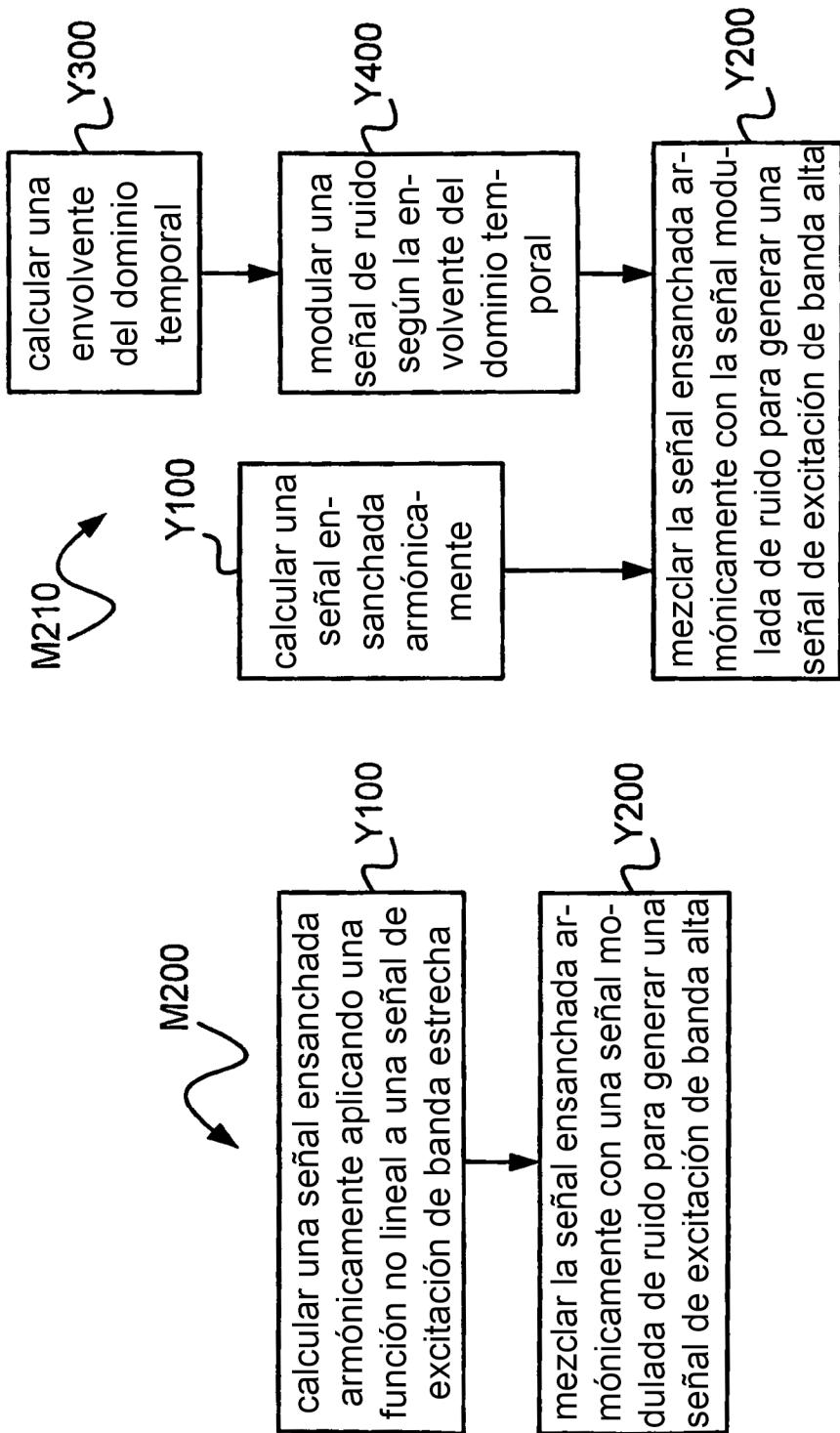


FIG. 31a
FIG. 31b

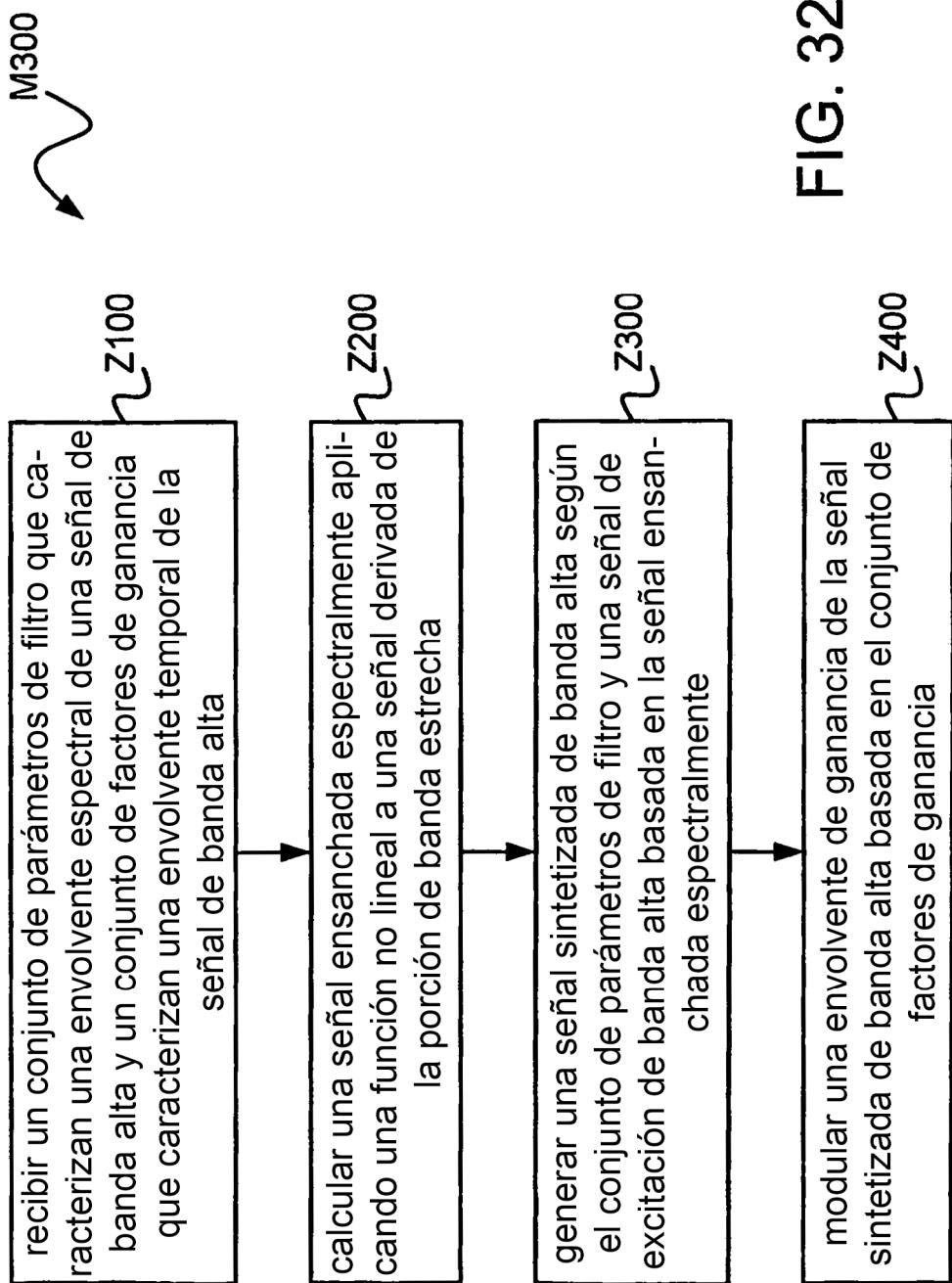


FIG. 32