

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 934 964**

51 Int. Cl.:

G10L 21/038 (2013.01)

G10L 21/0316 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **12.12.2019** **PCT/EP2019/084974**

87 Fecha y número de publicación internacional: **25.06.2020** **WO20126857**

96 Fecha de presentación y número de la solicitud europea: **12.12.2019** **E 19816780 (1)**

97 Fecha y número de publicación de la concesión europea: **02.11.2022** **EP 3899937**

54 Título: **Procesador de audio y procedimiento para generar una señal de audio de frecuencia mejorada usando procesamiento de impulsos**

30 Prioridad:

21.12.2018 EP 18215691

01.04.2019 EP 19166643

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
28.02.2023

73 Titular/es:

**FRAUNHOFER-GESELLSCHAFT ZUR
FÖRDERUNG DER ANGEWANDTEN
FORSCHUNG E.V. (100.0%)**

**Hansastr. 27c
80686 München, DE**

72 Inventor/es:

**DISCH, SASCHA y
STURM, MICHAEL**

74 Agente/Representante:

PONTI & PARTNERS, S.L.P.

ES 2 934 964 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

Procesador de audio y procedimiento para generar una señal de audio de frecuencia mejorada usando procesamiento de impulsos

5

[0001] La presente invención se refiere al procesamiento de señales de audio y en particular, con conceptos para generar una señal de audio de frecuencia mejorada a partir de una señal de una fuente de audio.

[0002] El almacenamiento o transmisión de señales de audio a menudo está sujeto a restricciones de velocidad de bits estrictas. En el pasado, los codificadores eran forzados a reducir drásticamente el ancho de banda de audio transmitido cuando únicamente estaba disponible una velocidad de bits muy baja. Hoy en día los códecs de audio modernos son capaces de codificar señales de banda ancha usando procedimientos de extensión de ancho de banda (BWE) [1, 2, 5].

[0003] Estos algoritmos dependen de una representación paramétrica del contenido de alta frecuencia (HF), el cual se genera a partir de la parte de baja frecuencia (LF) codificada de la forma de onda de la señal decodificada por medio de transposición a la región ("parche") espectral de HF. Al hacerlo, en primer lugar, se genera un parche "sin procesar" y en segundo lugar, se aplica el parámetro controlado después del procesamiento sobre el parche "sin procesar".

20

[0004] Típicamente, el post-procesamiento se aplica para ajustar las propiedades perceptuales importantes que no se han considerado durante la generación de alta frecuencia a través de la transposición y, de este modo, necesitan ajustarse sobre el parche "sin procesar" resultante a posteriori.

[0005] Sin embargo, si, por ejemplo, la estructura fina espectral en un parche copiado a alguna región objetivo es bastante diferente de la estructura fina espectral del contenido original, pueden resultar artefactos no deseados y degradar la calidad perceptual de la señal de audio decodificada. Con frecuencia, en estos casos, el posprocesamiento aplicado no ha sido capaz de corregir completamente las propiedades erróneas del parche "sin procesar".

30

[0006] El objeto de la invención es mejorar la calidad perceptual a través de una nueva generación adaptada de señales novedosas del relleno de espacios o contenido del parche "sin procesar" de señal de alta frecuencia estimada la cual se adapta perceptualmente a la señal de LF.

[0007] Al obtener ya una señal "sin procesar" adaptada perceptualmente, de otro modo las medidas de corrección necesarias a posteriori se minimizan. Además, una señal sin procesar adaptada perceptualmente puede permitir la elección de una frecuencia de cruce más baja entre LF y HF que las estrategias tradicionales [3].

[0008] En los esquemas BWE, la reconstrucción de la región espectral de HF por encima de una llamada frecuencia de cruce dada con frecuencia se basa en el parche espectral. Típicamente, la región de HF está compuesta de múltiples parches apilados y cada uno de esos parches se origina de regiones de paso de banda (BP) del espectro de LF por debajo de la frecuencia de cruce dada.

[0009] Los sistemas del estado de la técnica efectúan eficientemente el parcheo dentro de un banco de filtros o representación de transformación tiempo-frecuencia copiando un conjunto de coeficientes de subbanda adyacentes de una fuente espectral a la región espectral objetivo.

[0010] En una etapa siguiente, se ajusta la tonalidad, el ruido y la envolvente espectral de modo que se asemeje mucho a las propiedades perceptuales y la envolvente de la señal de HF original que ha sido medida en el codificador y transmitida en el flujo de bits como información del lado de BWE.

50

[0011] La Replicación de Banda Espectral (SBR) es una BWE bien conocida en códecs de audio contemporáneos como la codificación de audio avanzada de alta eficiencia (HE-AAC) y usan las técnicas citadas anteriormente [1].

55

[0012] El Llenado Inteligente de Espacios (IGF) denota una técnica de codificación semiparamétrica dentro de los códecs modernos como el códec de audio MPEG-H 3D o 3gpp EVS [2]. El IGF puede aplicarse para rellenar huecos espectrales introducidos por el proceso de cuantización en el codificador debido a restricciones de baja velocidad de bits.

60

[0013] Típicamente, si el presupuesto de bits limitado no permite una codificación transparente, emergen huecos espectrales en la región de alta frecuencia (HF) de la señal primero y afectan de manera creciente a todo el intervalo espectral superior para velocidades de bits más bajas.

[0014] En el lado del decodificador, esos huecos espectrales son sustituidos vía IGF usando contenido de HF

65

sintético generado en una forma semiparamétrica del contenido de baja frecuencia (LF) y post-procesamiento controlado por información del lado paramétrico adicional como el ajuste de envolvente espectral y un “nivel de blanqueo espectral”.

5 **[0015]** Sin embargo, después del post-procesamiento, puede existir un desajuste remanente que puede conducir a la percepción de artefactos. Ese desajuste puede consistir típicamente en

- Desajuste de armónico: artefactos latentes debido a componentes armónicos mal colocados
- Desajuste de fase: dispersión de señales de excitación similares a impulso que conduce a una pérdida percibida en
- 10 el zumbido en la voz sonora o señales de bronce.
- Desajuste de tonalidad: tonalidad exagerada o demasiado poca.

[0016] Por lo tanto, se han propuesto procedimientos de corrección de frecuencia y fase para corregir esos tipos de desajustes [3] a través de post-procesamiento adicional. En la presente invención, proponemos ya evitar

15 introducir esos artefactos en la señal “sin procesar”, en lugar de fijarlas en una etapa de post-procesamiento como se encuentra en los procedimientos del estado de la técnica.

[0017] Otras implementaciones de BWE se basan en técnicas de dominio del tiempo para estimar la señal de HF [4], típicamente a través de la aplicación de una función no lineal en la forma de onda del dominio del tiempo

20 como la rectificación, encuadre o una función de energía. De esta manera, distorsionando la LF, se genera una mezcla rica de sobretonos consonantes y disonantes que pueden usarse como una señal “sin procesar” para restablecer el contenido de HF.

[0018] Aquí, especialmente el desajuste armónico es un problema, puesto que, sobre el contenido polifónico,

25 esas técnicas producen una mezcla densa de sobretonos armónicos deseados inevitablemente mezclados con componentes no armónicos no deseados.

[0019] Aunque el post-procesamiento puede incrementar fácilmente el ruido, falla absolutamente en la eliminación de componentes tonales no inarmónicos no deseados una vez que se introducen en la HF estimada “sin

30 procesar”.

[0020] Un objeto de la presente invención es proporcionar un concepto mejorado para generar una señal de audio de frecuencia mejorada a partir de una señal de una fuente de audio.

35 **[0021]** El objeto se alcanza por medio de un procesador de audio de la reivindicación 1, un procedimiento de generación de una señal de audio mejorada de frecuencia de una señal de audio fuente de la reivindicación 19 o un programa informático de la reivindicación 20.

[0022] La presente invención se basa en el descubrimiento de que se obtiene una calidad de perceptual

40 mejorada de una extensión de ancho de banda de audio o relleno de espacios o, generalmente, mejora en la frecuencia por medio de una generación adaptada de señal novedosa del relleno de espacios o contenido de parche “sin procesar” de señal de alta frecuencia (HF) estimada. Mediante la obtención de una señal “sin procesar” adaptada perceptualmente, las medidas de corrección a posteriori en otras circunstancias necesarias pueden minimizarse o incluso eliminarse. Las realizaciones de la presente invención indicadas como Excitación de Impulso

45 Sincronizado de Envolvente de Forma de Onda (WESPE) se basan en la generación de una señal similar a un tren de impulsos en un dominio del tiempo, donde la colocación del impulso real se sincroniza a una envolvente de dominio del tiempo. La última se deriva de la señal de baja frecuencia (LF) que está disponible en la salida de, por ejemplo, un codificador central o que está disponible en cualquier otra fuente de una señal de una de audio fuente. De este modo, se obtiene una señal “sin procesar” adaptada perceptualmente.

50 **[0023]** Un procesador de audio de acuerdo con un aspecto de la invención se configura para generar el audio de frecuencia mejorada a partir de la señal de una señal de audio fuente y comprende un determinador de envolvente para determinar una envolvente temporal de al menos una porción de la señal de audio fuente. Un analizador se configura para analizar la envolvente temporal para determinar valores de ciertas características de la

55 envolvente temporal. Esos valores son valores temporales. Se coloca un sintetizador de señales para generar una señal de síntesis, donde la generación de la señal de síntesis comprende colocar impulsos con relación a los valores temporales determinados, donde los impulsos se ponderan mediante el uso de pesos derivados de amplitudes de la envolvente temporal, donde las amplitudes se relacionan con los valores temporales, donde se colocan los impulsos. Existe un combinador para combinar al menos una banda de la señal de síntesis que no está incluida en la señal de

60 audio fuente y la señal de audio fuente para obtener la señal de audio de frecuencia mejorada.

[0024] La presente invención es ventajosa dado que, en contraste con la generación un tanto “ciega” de frecuencias más altas a partir de la señal de audio fuente, por ejemplo mediante el uso de un procesamiento no lineal o similar, la presente invención proporciona un procedimiento fácilmente controlado determinando la

65 envolvente temporal de la fuente de señales y colocando impulsos en ciertas características de la envolvente

temporal, como los máximos locales de la envolvente temporal o los mínimos locales sobre la envolvente temporal o colocando impulsos siempre entre dos mínimos locales de la envolvente temporal o en cualquier otra relación con respecto a ciertas características de la envolvente temporal. Un impulso tiene un contenido de frecuencia que es, en general, plano sobre todo el intervalo de frecuencia bajo consideración. De este modo, incluso cuando se usen impulsos que teóricamente no sean ideales si no, por ejemplo, cercanos a eso, el contenido de frecuencia de esos impulsos no ideales, por ejemplo, impulsos que no están en línea con la forma Dirac ideal, no obstante, es relativamente plana en el intervalo de frecuencia interesante, por ejemplo, entre 0 y 20 kHz en el contexto del IGF (relleno inteligente de espacios) o en el intervalo de frecuencia entre, por ejemplo, 8 kHz a 16 kHz o 20 kHz en el contexto de la extensión de ancho de banda de audio, donde la señal fuente se limita en ancho de banda.

[0025] De este modo, la señal de síntesis que consiste en esos impulsos proporciona un contenido de alta frecuencia denso y fácilmente controlado. Colocando varios impulsos por envolvente temporal que, por ejemplo, se extraen de una trama de la señal de audio fuente, se obtiene la formación en el dominio espectral, puesto que los contenidos de frecuencia de los diferentes impulsos colocados con respecto a ciertas características se superponen entre sí en el dominio espectral para hacer coincidir en al menos las características dominantes o, generalmente, ciertas características de la envolvente temporal de la señal de audio fuente. Debido al hecho de que las fases de los valores espectrales representados por impulsos se bloquean entre sí y debido al hecho de que, preferiblemente, cualquier impulso positivo o impulso negativo son colocados por el sintetizador de señales, las fases de los valores espectrales representados por un impulso individual entre los diferentes impulsos se bloquean entre sí. Por tanto, se obtiene una señal de síntesis controlada que tiene características de dominio de frecuencia muy útiles. Típicamente, la señal de síntesis es una señal de banda ancha que se extiende sobre todo el intervalo de frecuencia de audio existente, es decir, que se extiende también hacia el intervalo de LF. Para generar realmente la señal final que eventualmente se combina con la señal de una fuente de audio para el propósito de mejorar la frecuencia, al menos una banda de la señal de síntesis, como la banda alta o una señal determinada por un paso de banda se extrae y suma a la señal de audio fuente.

[0026] El concepto de la invención tiene el potencial para efectuarse completamente en el dominio del tiempo, es decir, sin ninguna transformación específica. El dominio del tiempo es el dominio del tiempo típico o el dominio del tiempo filtrado por codificación por predicción lineal (LPC), es decir, una señal del dominio del tiempo que se ha blanqueado espectralmente y necesita finalmente ser procesada usando un filtro de síntesis de LPC para reintroducir la forma espectral original para que sea útil para renderizar las señales de audio. De este modo, la determinación de la envolvente, el análisis, la síntesis de señal, la extracción de la banda de señal de síntesis y la combinación final pueden efectuarse todas en el dominio del tiempo, de modo que pueden evitarse transformaciones espectrales de tiempo o transformaciones de tiempo espectral que típicamente incurren en retraso. Sin embargo, el contexto de la invención también es flexible dado que también pueden efectuarse varios procedimientos tales como determinación de la envolvente, síntesis de señal y la combinación parcial o totalmente en el dominio espectral. De este modo, la implementación de la presente invención, es decir, si ciertos procedimientos requeridos por la invención se implementan en el tiempo o dominio espectral, siempre pueden adaptarse completamente al marco correspondiente de un diseño de decodificador típico requerido para una cierta aplicación. El contexto de la invención es aún flexible en el contexto de un codificador de voz LPC donde, por ejemplo, se efectúa una mejora de frecuencia de una señal de excitación LPC (por ejemplo, una señal TCX). La combinación de la señal de síntesis y la señal de audio fuente se efectúa en el dominio del tiempo de LPC y la conversión final del dominio del tiempo de LPC al dominio del tiempo normal se efectúa con un filtro de síntesis de LPC donde, específicamente, se efectúa un ajuste de envolvente típicamente preferido de la señal de síntesis dentro de la etapa de filtro de síntesis de LPC para la porción espectral correspondiente representada por al menos una banda de la señal de síntesis. De este modo, las operaciones de post-procesamiento típicamente necesarias se combinan con el ajuste de envolvente dentro de una sola etapa de filtro. Esas operaciones de post-procesamiento pueden involucrar filtración de síntesis de LPC, filtración de énfasis conocido de los decodificadores de voz, otras operaciones de filtración posterior tales como operaciones de filtración posterior de bajos u otros procedimientos de filtración posterior de sonido mejorado basados en LTP (Predicción a Largo Plazo) como se encuentra en los decodificadores TCX u otros decodificadores.

[0027]

Las realizaciones preferidas de la presente invención se discuten posteriormente con respecto a las figuras adjuntas, en las cuales:

La Figura 1 es un diagrama de bloques de una realización de un procesador de audio de acuerdo con la presente invención;

La Figura 2 es una descripción detallada de una realización preferida del determinador de envolvente de la Figura 1;

La Figura 3a es una realización para calcular la envolvente temporal de una señal de audio de subbanda o banda completa;

La Figura 3b es una implementación alternativa de la generación de una envolvente temporal;

La Figura 3c ilustra un diagrama de flujo para una implementación de la determinación de la señal analítica de la Figura 3a usando una transformación de Hilbert;

La Figura 4 ilustra una implementación preferida del analizador de la Figura 1;

La Figura 5 ilustra una implementación preferida del sintetizador de señales de la Figura 1;

La Figura 6 ilustra una realización preferida del procesador de audio como un dispositivo o procedimiento usado en el contexto de un decodificador central;

La Figura 7 ilustra una implementación preferida, donde la combinación de la señal de síntesis y la señal de una fuente de audio se efectúa en el dominio de LPC;

5 La Figura 8 ilustra una realización adicional de la presente invención, donde el filtro de paso alto o de paso de banda, el ajuste de envolvente y la combinación de la señal de audio fuente y la señal de síntesis se efectúan en el dominio espectral;

La Figura 9a ilustra varias señales en el proceso de mejora de frecuencia con respecto a un elemento de sonido "voz masculina alemana";

10 La Figura 9b ilustra un espectrograma para el elemento de sonido "voz masculina alemana";

La Figura 10a ilustra varias señales en el proceso de mejora de frecuencia con respecto a un elemento de sonido "diapasón";

La Figura 10b ilustra un espectrograma para el elemento de sonido "diapasón";

La Figura 11a ilustra varias señales en el proceso de mejora de frecuencia con respecto a un elemento de sonido "Madonna Vogue"; y

15 La Figura 11b ilustra varias señales en el proceso de mejora de frecuencia con respecto a un elemento de sonido "Madonna Vogue".

[0028] La Figura 1 ilustra un procesador de audio para generar una señal de audio de frecuencia mejorada
20 420 en la salida de un combinador 400 de una señal de audio fuente introducida en un determinador de envolvente 100 por un lado y la entrada al combinador 400 por el otro.

[0029] El determinador de envolvente 100 se configura para determinar una envolvente temporal en al menos una porción de la señal de audio fuente. El determinador de envolvente puede usar una señal de audio fuente de
25 banda completa o, por ejemplo únicamente una banda o porción de la señal de una fuente de audio que tiene una cierta frecuencia límite inferior como una frecuencia de, por ejemplo, 100, 200 o 500 Hz. La envolvente temporal se envía desde el determinador de envolvente 100 a un analizador 200 para analizar la envolvente temporal para determinar valores de ciertas características de la envolvente temporal. Esos valores son valores temporales. Ciertas características pueden, por ejemplo, ser máximos locales de la envolvente temporal, mínimos locales de la
30 envolvente temporal, cruces en cero de la envolvente temporal o puntos entre dos mínimos locales o puntos entre dos máximos locales donde, por ejemplo, los puntos entre esas características son valores que tienen la misma distancia temporal hacia las características vecinas. De este modo, esas ciertas características pueden ser también puntos que estén a medio camino entre dos mínimos locales o dos máximos locales. Sin embargo, en realizaciones preferidas, se prefiere la determinación de máximos locales de la envolvente temporal usando, por ejemplo, un
35 procesamiento de cálculo de curva. Los valores temporales de ciertas características de la envolvente temporal se envían a un sintetizador de señales 300 para generar una señal de síntesis. La generación de la señal de síntesis comprende colocar impulsos en relación con las envolventes temporales determinadas, donde los impulsos se ponderan, ya sea antes de la colocación o después de colocarse usando pesos derivados de amplitudes de la envolvente temporal, estando las amplitudes relacionadas con los valores temporales recibidos del analizador o
40 relacionadas con los valores temporales, donde se colocan los impulsos. Al menos una banda de la señal de síntesis o la banda superior completa de la señal de síntesis o varias bandas individuales y distintas de la señal de síntesis o incluso la señal de síntesis completa se envían al combinador 400 para combinar al menos una banda de la señal de síntesis que no se incluyó en la señal de audio fuente y la señal de audio fuente para obtener una señal de audio de frecuencia mejorada.

[0030] En una realización preferida, el determinador de envolvente se configura como se ilustra en la Figura 2. En esta realización, la señal de audio fuente o al menos la porción de la señal de audio fuente se descompone en una pluralidad de señales de subbanda como se ilustra en 105. Una o más o incluso todas las subbandas se seleccionan o usan como se ilustra en 110 para la determinación de envolventes temporales individuales por cada
50 subbanda (seleccionada) como se ilustra en 120. Como se ilustra en 125, las envolventes temporales se normalizan o filtran y, las envolventes temporales individuales se combinan entre sí como se indica en 130 para obtener la envolvente temporal final en la salida del determinador de envolvente. Esta envolvente temporal final puede ser una envolvente combinada como se determina por el procedimiento ilustrado en la Figura 2. Dependiendo de la implementación, puede proporcionarse una etapa de filtración adicional 115 para normalizar o filtrar las subbandas seleccionadas individuales. Si se usan todas las subbandas, todas esas subbandas se normalizan o filtran como se indica en el bloque 115. El procedimiento de normalización indicado en 125 puede evitarse y este procedimiento que no efectúa la normalización o filtración de las envolventes temporales determinadas es útil, cuando las subbandas, de las cuales se determinan las envolventes temporales en el bloque 120 ya se han normalizado o filtrado de manera correspondiente. Naturalmente, ambos procedimientos 115, 125 pueden efectuarse también o, incluso de
60 manera alternativa, únicamente los procedimientos de determinación de la envolvente temporal por cada subbanda (seleccionada) 120 y la combinación subsecuente de las envolventes temporales 130 puede efectuarse sin ningún procedimiento ilustrado por el bloque 115 o 125.

[0031] En una implementación adicional, la descomposición en el bloque 105 no puede efectuarse en su
65 totalidad, pero puede reemplazarse por una filtración de paso alto con una frecuencia de cruce baja tal como una

frecuencia de cruce de 20, 50, 100 o, por ejemplo una frecuencia inferior a 500 Hz y únicamente se determina una sola envolvente temporal del resultado de esta filtración de paso alto. Naturalmente, la filtración de paso alto también puede evitarse y, únicamente se deriva una sola envolvente temporal de la señal de audio fuente y, típicamente, una trama de la señal de audio fuente, donde la señal de audio fuente se procesa preferiblemente en tramas típicamente superpuestas, pero pueden usarse incluso también tramas no superpuestas. Una selección indicada en el bloque 110 se implementa en cierto escenario, cuando, por ejemplo, se determina que una cierta señal de subbanda no satisface criterios específicos con respecto a las características de señal de subbanda o se excluye a partir de la determinación de la envolvente temporal final por cualquier razón.

- 10 **[0032]** La Figura 5 ilustra una implementación preferida del sintetizador de señales 300. El sintetizador de señales 300 recibe, como una entrada del analizador 200, los valores temporales de características y, adicionalmente, información adicional sobre la envolvente. En el elemento 310, el sintetizador de señales 300 ilustrado en la Figura 5 deriva factores de escalamiento de la envolvente temporal que se relacionan con los valores temporales. Por lo tanto, el bloque 310 recibe información de envolvente tal como la amplitud de envolvente por un lado y valores temporales por el otro. La derivación de los factores de escalamiento se efectúa, por ejemplo, usando una función de compresión tal como una función de raíz cuadrada, una función de energía con una energía menor que 1.0 o, por ejemplo, una función de registro.

- [0033]** El sintetizador de señales 300 comprende el procedimiento de colocación 305 de impulsos en los valores temporales donde, preferiblemente, solo se colocan los impulsos negativos o positivos para tener fases sincronizadas de los valores espectrales relacionados que están asociados con un impulso. Sin embargo, en otras realizaciones y dependiendo de, por ejemplo, otros criterios derivados del relleno de espacios típicamente disponible o información del lado de la extensión de ancho de banda, se efectúa una colocación aleatoria de los impulsos, típicamente, cuando la tonalidad de la señal de banda base no es alta. La colocación de impulsos negativos o positivos puede ser controlada por la polaridad de la forma de onda original. La polaridad de los impulsos puede elegirse de modo que sea igual a la polaridad de la forma de onda original que tenga el factor de cresta más alto. En otras palabras, esto significa que los picos positivos son modelados por impulsos positivos y viceversa.

- [0034]** En la etapa 315, los impulsos obtenidos por el bloque 305 se escalan usando el resultado del bloque 310 y, se efectúa un post-procesamiento opcional 320 para los impulsos. La señal de impulso está disponible y la señal de impulso se filtra por paso alto o se filtra por paso de banda como se ilustra en el bloque 325 para obtener una banda de frecuencia de la señal de impulso, es decir, para obtener al menos la banda de la señal de síntesis que se envió al combinador. Sin embargo, se aplica un ajuste de envolvente espectral opcional 330 a la señal emitida por la etapa de filtración 325, donde este ajuste de envolvente espectral se efectúa usando una cierta función envolvente o una cierta selección de parámetros de envolvente como los derivados de la información lateral o alternativamente derivados de la señal de audio fuente en el contexto de, por ejemplo, aplicaciones de extensiones de ancho de banda ciegas.

- [0035]** La Figura 6 ilustra una realización preferida de un procesador de audio o un procedimiento de procesamiento de audio para generar una señal de audio de frecuencia mejorada. La estrategia de la invención llamada Excitación de Impulso Sincronizado de Envolvente de Forma de Onda (WESPE) se basa en la generación de una señal similar a un tren de impulsos, donde la colocación real del impulso se sincroniza con una envolvente de dominio del tiempo dedicada. Esta denominada envolvente común se deriva de la señal de LF que se obtiene en la salida del decodificador central 20, vía un conjunto de señales de paso de banda, cuyas envolventes individuales se combinan en una envolvente común.

- [0036]** La Figura 6 muestra una integración típica del procesamiento WESPE en un decodificador de audio que presenta funcionalidad de extensión de ancho de banda (BWE), el cual es también una realización preferida de la nueva técnica. Esta implementación opera sobre tramas de tiempo con una duración de por ejemplo 20 ms, opcionalmente con superposiciones temporales entre las tramas, por ejemplo 50 %.

[0037] Ventajas de la BWE de la WESPE propuesta

- Mitigación de artefactos ásperos o latentes
- Continuación armónica de señales
- Preserva impulsos
- Adecuada como BWE de voz
- Puede gestionar música también
- El cruce de BWE puede comenzar a 2 kHz o menos ya
- BWE autoajutable con respecto a la tonalidad, alineación de separación, armonicidad, fase

[0038] El procesamiento WESPE comprende las siguientes etapas:

1. Estimación de la envolvente temporal 100: La señal de LF obtenida del decodificador central 20 se divide (105) en una colección de señales de paso de banda. A continuación, se determina la envolvente temporal 120 por

cada una de las señales de paso de banda. Opcionalmente, puede aplicarse la normalización o filtración de las envolventes individuales. A continuación, todas las envolventes temporales se combinan 130 en una envolvente común. Preferiblemente, la operación de combinación es un proceso de promediación.

2. Colocación de impulsos sincronizados: La envolvente común derivada en la etapa 1 se analiza 205 mediante la aplicación del cálculo de curva, preferiblemente para la ubicación de sus máximos locales. Los candidatos máximos obtenidos pueden opcionalmente seleccionarse posteriormente o estabilizarse con respecto a su 210 distancia temporal. Se coloca un impulso de Dirac 305 en la señal "sin procesar" estimada para la generación de HF en cada ubicación de un máximo. Opcionalmente, este proceso puede ser soportado por la información lateral.

3. Escalamiento de la magnitud de impulso individual de la envolvente: el tren de impulso montado en el paso previo 2 se pondera 315 por medio de pesos temporales derivados de la envolvente común.

4. Post-procesamiento, selección de extracción de HF o relleno de espacios: La señal "sin procesar" generada en la etapa 3 se post-procesa opcionalmente 320, por ejemplo por adición de ruido y se filtra 325 para usarse como HF en la BWE o como una señal mosaico objetivo para rellenar espacios.

5. Ajuste de energía: La distribución de energía espectral de la señal filtrada de la etapa 4 se ajusta 330 para usarse como HF en la BWE o como una señal de mosaico objetivo para rellenar espacios. Aquí, se usa la información lateral 40 del flujo de bits sobre la distribución de energía deseada.

6. Mezclado de HF o señal de relleno de espacios con LF: Finalmente, la señal de ajuste de la etapa 5 se mezcla 400 con la salida del codificador central 30 de acuerdo con la BWE usual o principios de relleno de espacios, es decir pasando un filtro de HP y complementando la LF o rellenando huecos espectrales en las regiones espectrales de relleno de espacios.

[0039] En lo sucesivo, se explicará mejor la función de cada una de las etapas comprendidas en el procesamiento WESPE dando señales ejemplares y su efecto sobre los resultados del procesamiento.

[0040] La estimación de envolvente común temporal apropiada es una parte clave de la WESPE. La envolvente común permite la estimación de propiedades perceptuales promediadas y de este modo representativas de cada trama de tiempo individual.

[0041] Si la señal LF es muy tonal con una entonación f_0 y un espectro de líneas de sobretono separadas Δf_0 fuerte, aparecerán varias líneas en cada una de las señales de paso de banda individuales, si su ancho de paso de banda puede acomodarlas, produciendo modulaciones de envolvente coherentes fuertes a través de pulsaciones dentro de todas las bandas de paso de banda. El promedio de las envolventes temporales preservará esa estructura de modulación de envolvente coherente encontrada a través de envolventes de paso de banda y resultarán picos fuertes en lugares aproximadamente equidistantes espaciados $\Delta T_0 = 1/(\Delta f_0)$. Finalmente, a través de la aplicación del cálculo de curva, se colocarán impulsos fuertes en las ubicaciones de picos, formando un tren de impulsos que tiene un espectro consistente de líneas equidistantes discretas en lugares $n \cdot \Delta f_0$, $n=1, \dots, N$.

[0042] En el caso de que una señal tonal no tenga sobretonos en absoluto o el ancho de banda de los filtros de paso de banda no pueda acomodar más de uno de esos sobretonos en cada una de las bandas individuales, la estructura de modulación no se mostrará en todas las señales de paso de banda y en consecuencia, no dominará la envoltura común promediada. La colocación de impulsos resultantes se basará principalmente en los máximos espaciados irregularmente y de este modo, será ruidosa.

[0043] Lo mismo es cierto para señales de LF ruidosas que exhiben una colocación de máximos locales aleatoria en la señal envolvente común: esos conducen a una colocación de impulsos pseudo-aleatoria.

[0044] Los eventos transitorios se preservan puesto que en este caso todas las señales de paso de banda comparten un máximo común alineado temporalmente que, de este modo, también aparecerá en la envolvente común.

[0045] Los pasos de banda se dimensionarán de modo que abarquen una banda perceptual y que puedan acomodar al menos 2 sobretonos para la frecuencia más alta que se requiera resolver. Para una promediación mejor, los pasos de banda pueden tener alguna superposición de sus bandas de transición. De esta manera, la tonalidad de las señales estimadas se adapta intrínsecamente a la señal de LF. Los pasos de banda pueden excluir frecuencias muy bajas por debajo de, por ejemplo, 20 Hz.

[0046] La colocación y el escalamiento de impulsos temporales sincronizada es el otro contribuyente clave de WESPE. La colocación sincronizada de impulsos hereda las propiedades principales representativas condensadas en las modulaciones temporales de la envolvente común y las imprime en una señal de banda completa sin procesar adaptada perceptualmente.

[0047] Cabe señalar que se sabe que la percepción humana del contenido de alta frecuencia funciona a través de la evaluación de modulaciones en envolventes de bandas críticas. Como se ha detallado anteriormente, la colocación sincronizada de impulsos temporales a la envolvente de LF común impone similitud y alineación de la

estructura temporal y espectral perceptualmente relevante entre LF y HF.

[0048] En el caso de señales muy tonales con sobretonos fuertes y limpios, como por ejemplo, diapasón, la WESPE puede asegurar a través de estabilización óptima adicional que la colocación de impulso sea exactamente equidistante, conduciendo a un espectro de sobretono de HF muy tonal de la señal “sin procesar”.

[0049] La ponderación de los impulsos con la envolvente común asegura que las modulaciones dominantes se preserven en impulsos fuertes, mientras que las modulaciones menos importantes resultan en impulsos débiles, contribuyendo además a la propiedad de WESPE de adaptación intrínseca de la señal “sin procesar” a la señal LF.

[0050] En el caso de una señal ruidosa, si la colocación y ponderación de impulsos se vuelve cada vez más aleatoria, esto conduce a una señal “sin procesar” gradualmente más ruidosa, la cual es una propiedad muy deseada.

[0051] Las etapas de procesamiento restantes, extracción de HF, ajuste de energía y mezclado son etapas adicionales que son necesarias para integrar el procesamiento WESPE novedoso en un códec para adaptarse a la funcionalidad completa de BWE o relleno de espacios.

[0052] La Figura 3a ilustra una implementación preferida para la determinación de la envolvente temporal. Como se ilustra en 135, la señal analítica se determina usando una transformación de Hilbert. La salida del bloque 135, es decir, la señal de la transformación de Hilbert se usa para el cálculo de la envolvente ENV(t) ilustrada en 140. Hasta este punto, la envolvente se calcula como el cuadrado de los valores de la señal de una fuente de audio original en ciertos instantes de tiempo y el cuadrado de los valores de la transformación de Hilbert correspondiente en ciertos instantes de tiempo y sumando los valores cuadrados y calculando la raíz cuadrada del resultado de la suma de cada instante de tiempo individual. Mediante este procedimiento, se determina la envolvente temporal con la misma resolución de muestra que la señal de una fuente de audio original a(t). Naturalmente, se efectúa el mismo procedimiento cuando la entrada en el bloque 135 y 140 es una señal de subbanda como la obtenida por el bloque 105 o como la seleccionada por el bloque 110 o como la normalizada y filtrada por el bloque 115 de la Figura 2.

[0053] Otro procedimiento para el cálculo de una envolvente temporal se ilustra en el bloque 145 y 150 de la Figura 3b. Hasta este punto, la forma de onda de la señal de una fuente de audio o una subbanda del audio de la fuente se rectifica (145) y la señal rectificada se filtra por paso bajo (150) y el resultado de esta filtración de paso bajo es la envolvente de la señal de una fuente de audio o es la envolvente de señales de subbanda individuales que se combinan con otras de esas envolventes de otras subbandas, preferiblemente por promediación como se ilustra en 130 en la Fig. 2.

[0054] La publicación “Simple empirical algorithm to obtain signal envelope in the three steps”, de C. Jarne, 20 de marzo de 2017, ilustra otros procedimientos para calcular envolventes temporales como el cálculo de un valor media cuadrática instantánea (RMS) de la forma de onda a través de una ventana deslizante con soporte finito. Otros procedimientos consisten en calcular una aproximación lineal por piezas de la forma de onda donde la envolvente de amplitud se crea encontrando y conectando los picos de la forma de onda en una ventana que se mueve a través de los datos. Los procedimientos adicionales dependen de la determinación de picos permanentes en la señal de la fuente de audio o la señal de subbanda y la derivación de la envolvente por interpolación.

[0055] Otros procedimientos para calcular la envolvente temporal comprenden la interpretación de información lateral que representa la envolvente o efectuar una predicción en el dominio espectral sobre un conjunto de valores espectrales derivados de una trama de un dominio del tiempo como se sabe a partir de TNS (Conformación de Ruido Temporal), donde los coeficientes de predicción correspondientes representan una envolvente temporal de la trama.

[0056] La Figura 3c ilustra una implementación preferida de la determinación de la señal analítica usando la transformación de Hilbert indicada en 135 en la Figura 3a. Los procedimientos para calcular esas transformaciones de Hilbert se ilustran, por ejemplo, en “A Praat-Based Algorithm to Extract the Amplitude Envelope and Temporal Fine Structure Using the Hilbert Transform”, He, Lei, y col., INTERSPEECH 2016-1447, páginas 530-534. En una etapa 136, se calcula un espectro complejo de la señal a(t), por ejemplo, la señal de la fuente de audio o una señal de subbanda. En la etapa 137, se selecciona una parte positiva del espectro complejo o se deselecta la parte negativa. En la etapa 138, la parte positiva del espectro complejo se multiplica por “-j” y en la etapa 139, el resultado de esta multiplicación se convierte al dominio del tiempo y tomando la parte imaginaria, se obtiene la señal analítica $\hat{a}(t)$.

[0057] Naturalmente, están disponibles muchos otros procedimientos para determinar la envolvente temporal y cabe señalar que la envolvente temporal no necesariamente tiene que “envolver” realmente la señal en el dominio del tiempo, pero puede, por supuesto, ser algún máximo o mínimo de la señal en el dominio del tiempo mayor o menor que el valor de la envolvente correspondiente en este punto de tiempo.

[0058] La Figura 4 ilustra una realización preferida del procedimiento de determinación de valores temporales de ciertas características de la envolvente temporal. Con este fin, se introduce una envolvente temporal promedio en el bloque 205 para la determinación de valores temporales iniciales para las características. Esos valores temporales iniciales pueden, por ejemplo, ser los valores temporales de máximos realmente encontrados dentro de la envolvente temporal. Los valores temporales finales de las características, en los cuales se colocan los impulsos reales, se derivan de los valores temporales sin procesar o de valores temporales "iniciales" por medio de una función de optimización, por medio de una información lateral o por medio de la selección y manipulación de las características sin procesar como se indica en el bloque 210. Preferiblemente, el bloque 210 es para que los valores iniciales se manipulen de acuerdo con una regla de procesamiento o de acuerdo con una función de optimización. Particularmente, la función de optimización o regla de procesamiento se implementa de modo que los valores temporales sean colocados en una trama con un espaciado de trama T. Particularmente, el espaciado de trama T y/o la posición de la trama dentro de la envolvente temporal es para que un valor de desviación entre los valores temporales y los valores temporales iniciales tenga una característica predeterminada donde, en realizaciones, el valor de desviación sea una suma sobre las diferencias cuadradas y/o la característica predeterminada sea mínima. De este modo, después de la determinación de los valores temporales iniciales, se coloca una trama de valores temporales equidistantes la cual coincide con la trama no constante de los valores iniciales temporales lo máximo posible, pero ahora muestra un comportamiento tonal claro e ideal. La trama puede determinarse en el dominio muestreado en el sentido ascendente que tiene una granularidad temporal más fina en comparación con el dominio no muestreado en el sentido ascendente o se puede utilizar alternativamente un retraso fraccional para la colocación de impulsos con precisión de submuestra.

[0059] La Figura 7 ilustra una realización adicional de la presente invención en el contexto del procesamiento LPC. Como se ilustra, por ejemplo, en la Figura 1 o la Figura 6, el procesador de audio de la Figura 7 comprende el determinador de envolvente 100, el analizador 200 (ambos no mostrados en la Figura 7) y el sintetizador de señales 300. En contraste con la Figura 6, sin embargo, los datos de salida del decodificador central, es decir, la salida de LF 30 no es una señal de audio de dominio del tiempo, sino una señal de audio que se encuentra en el dominio del tiempo de LPC. Esos datos pueden encontrarse típicamente dentro de un codificador TCX (Excitación Codificada por Transformación) como una representación de señal interna.

[0060] Los datos de TCX generados por el decodificador de audio 20 en la Figura 7 se envían al mezclador que se indica en la Figura 7 como el sumador de dominio de LPC 405. El sintetizador de señales genera datos de frecuencia mejorada TCX. De este modo, la señal de síntesis generada por el sintetizador de señales se deriva de la señal de la fuente de audio siendo una señal de datos TCX en esta realización. De este modo, en la salida del bloque 405, está disponible una señal de audio de frecuencia mejorada que, sin embargo, aún está en el dominio del tiempo de LPC. Un filtro de síntesis de LPC conectado posteriormente 410 efectúa la transformación de la señal de dominio del tiempo de LPC al dominio del tiempo.

[0061] El filtro de síntesis de LPC se configura para efectuar adicionalmente un tipo de deseo de énfasis si se requiere y adicionalmente, este filtro de dominio del tiempo se configura para efectuar también el ajuste de envolvente espectral para la banda de señal de síntesis. De este modo, el filtro de síntesis de LPC 410 en la Figura 7 no solo efectúa la filtración de síntesis de la salida del intervalo de frecuencia de datos de TCX por el decodificador de audio 20, sino que también efectúa el ajuste de la envolvente espectral para los datos en la banda espectral que no se incluyó en la salida de datos de TCX por el decodificador de audio 20. Típicamente, estos datos también se obtienen de la señal de audio codificada 10 por medio del decodificador de audio 20 que extrae los datos de LPC 40a para el intervalo de frecuencia central y adicionalmente extrae el ajuste de la envolvente espectral para la banda alta o para IGF (Relleno Inteligente de Espacio), una o más bandas indicadas en 40b de la Figura 7. De este modo, el combinador o mezclador en la Figura 1 se implementa por medio del sumador de dominio LPC 405 y el filtro de síntesis de LPC conectado posteriormente 410 de la Figura 7 de modo que la salida del filtro de síntesis de LPC 410 indicada en 420 es la señal de audio de dominio del tiempo de frecuencia mejorada. En contraste con el procedimiento en la Figura 6, donde el ajuste de la envolvente espectral 330 se efectúa antes de efectuar la operación del mezclador por el combinador 400, la Figura 7 efectúa el ajuste de la envolvente de la banda alta o banda de relleno después de mezclar o combinar ambas señales.

[0062] La Figura 8 ilustra una implementación adicional del procedimiento ilustrado en la Figura 6. Básicamente, la implementación de la Figura 6 se efectúa en el dominio del tiempo de modo que los bloques 320, 325, 330, 400 se efectúan completamente en el dominio del tiempo. De manera alternativa, la implementación de la Figura 8 depende de una conversión de espectro 105 para la banda baja que, sin embargo, es una medida opcional, pero la operación de conversión de espectro 105 en la Figura 8 para la banda baja se utiliza de manera ventajosa para la implementación del banco de filtro de paso de banda 105 en la Figura 6. Adicionalmente, la implementación de la Figura 8 comprende un convertidor de espectro 345 para convertir la salida del procesador de impulsos 340 que comprende típicamente la colocación de impulsos 305 y el escalamiento de impulsos 315 de la Figura 6. El procesador de impulsos 340 en la Figura 8 puede comprender adicionalmente el bloque estabilizador 210 como una característica opcional y el bloque de búsqueda extremo 205 como una característica opcional.

[0063] Sin embargo, los procedimientos de la filtración de paso alto 325, ajuste de envolvente 330 y la

combinación de la banda baja y la banda alta se efectúa por medio de un banco de filtro de síntesis, es decir, que se efectúa en el dominio espectral y la salida del banco de filtro de síntesis 400 en la Figura 8 es la señal de audio de frecuencia mejorada del dominio del tiempo 420. Sin embargo, si el elemento 400 se implementa como un combinador simple para combinar diferentes bandas, la salida del bloque 400 también puede ser una señal de dominio espectral completa que consiste típicamente en bloques subsecuentes de valores espectrales que se procesan adicionalmente en cualquier forma requerida.

[0064] En lo sucesivo, se proporcionan tres ejemplos de señales características que se han extendido en el ancho de banda usando la BWE de WESPE. La velocidad de muestreo fue de 32 kHz, una DFT (105 en la Figura 8) con un espectro de un solo lado que tiene 513 líneas se usó para extraer 8 señales de paso de banda superpuestas. Para implementar el paso alto de 4 kHz (325 en la Figura 8), se efectuó el ajuste de envolvente espectral (330 en la Figura 8) y el mezclado (400 en la Figura 8) de LF y HF, se empleó una DFT/IDFT similar (345 en la Figura 8) con superposición del 50 % organizada en 16 bandas de factor de escala uniformes. La señal resultante mostrada en los espectrogramas es PCM no codificada de DC a 4 kHz y generada por WESPE de 4 kHz a 16 kHz.

[0065] La Figura 9a muestra un extracto corto (un bloque de 1024 muestras) de la forma de onda, la envolvente común y colocación de impulso sincronizado y escalada resultante por WESPE. Los impulsos grandes con dispersión algo ligera se colocan aproximadamente equidistantes en una estructura periódica amplia.

[0066] La Figura 9b describe el espectrograma de todo el elemento de prueba. La estructura de impulso vertical de voz sonora se mantiene alineada coherentemente entre LF y HF, mientras que las fricativas exhiben un ruido similar a la estructura de HF.

[0067] En consecuencia, la Figura 9a muestra cómo la WESPE modela impulsos de voz, muestra una forma de onda, envolvente común y una generación de impulsos, donde el elemento es "voz masculina Alemana". La Figura 9b muestra cómo WESPE modela impulsos de voz y muestra un espectrograma. El Elemento es "voz masculina Alemana".

[0068] La Figura 10a muestra un extracto corto (un bloque de 1024 muestras) de la forma de onda, de la envolvente común y la colocación de impulsos sincronizada y escalada resultante por WESPE. Se colocan distintos impulsos agudos equidistantemente en una estructura periódica estrecha. La Figura 10b describe el espectrograma de todo el elemento de prueba. La estructura de líneas horizontales del diapason se mantiene alineada entre LF y HF, sin embargo, la HF también es algo ruidosa y podría beneficiarse de la estabilización adicional.

[0069] La Figura 10a muestra cómo WESPE modela sobretonos armónicos y muestra una forma de onda, una envolvente común y generación de impulsos. El Elemento es "Diapasón". La Figura 10b muestra cómo WESPE modela sobretonos armónicos y muestra un espectrograma. El Elemento es "Diapasón".

[0070] La Figura 11a muestra un extracto corto (un bloque de 1024 muestras) de una forma de onda del elemento de muestra Madonna Vogue, la envolvente común y la colocación de impulsos sincronizada y escalada resultante por WESPE. La colocación y escalamiento de impulsos tiene una estructura casi aleatoria. La Figura 11b describe el espectrograma de todo el elemento de prueba. Las estructuras transitorias verticales de música de pop se mantienen alineadas coherentemente entre LF y HF, mientras que la tonalidad de HF es principalmente baja.

[0071] La Figura 11a muestra cómo WESPE modela una mezcla ruidosa y muestra una forma de onda, una envolvente común y una generación de impulsos. El Elemento es "Vogue". La Figura 11b muestra cómo WESPE modela una mezcla ruidosa y muestra un espectrograma. El Elemento es "Vogue".

Fin del inserto C

[0072] La primera imagen en las Figuras 9a, 10a, 11a ilustra una forma de onda de un bloque de 1024 muestras de una señal de la fuente de banda baja. Adicionalmente, muestra la influencia del filtro de análisis para extraer el bloque de muestras dado que la forma de onda es igual a 0 al inicio del bloque, es decir, en la muestra 0 y también es 0 al final del bloque, es decir, en la muestra 1023. Esa forma de onda está, por ejemplo, disponible en la entrada al bloque 100 de la Figura 1 y en 30 en la Figura 6. El eje vertical en las Figuras 9a, 9b, siempre indica una amplitud de dominio del tiempo y el eje horizontal en esas figuras siempre indica la variable de tiempo y, particularmente, el número de muestra que se extiende típicamente de 0 hasta 1023 para un bloque.

[0073] La segunda imagen en las Figuras 9a, 10a, 10b ilustra la envolvente de banda baja promediada y, particularmente, solo la parte positiva de la envolvente de banda baja. Naturalmente, la envolvente de banda baja típicamente es simétrica y también se extiende en el intervalo negativo. Sin embargo, solo se requiere la parte positiva de la envolvente de banda baja. Es visible en las Figuras 9a, 10a y 11a que en esta realización, la envolvente se ha calculado únicamente excluyendo a la vez el primer par de muestras del bloque y el último par de muestras del bloque lo cual, sin embargo, no es totalmente un problema, puesto que los bloques se calculan preferiblemente en una forma superpuesta. De este modo, la segunda imagen de las Figuras 9a, 10a, 11a,

típicamente ilustra la salida del bloque 100 de la Figura 1 o la salida del bloque 130 de la Figura 2, por ejemplo.

[0074] La tercera imagen de las Figuras 9a, 10a, 11a, ilustra la señal de síntesis posterior al escalamiento de impulso, es decir, posterior al procesamiento, en el cual los impulsos se colocan en los valores temporales de las características de la envolvente y han sido ponderados por las amplitudes correspondientes de la envolvente. Las Figuras 9a, 10a, 11a ilustran que los impulsos colocados únicamente se extienden desde la muestra 256 hasta la muestra 768. De este modo, la señal que consiste en los impulsos ponderados únicamente se extiende sobre 512 muestras y no tiene ninguna porción antes de esas muestras y después de esas muestras, es decir, que cubre la parte media de una trama. Esto refleja la situación de que la trama anterior tiene una superposición y la trama subsecuente tiene una superposición también. Para el propósito de generar una señal de impulso con los bloques subsecuentes, la señal de impulso del siguiente bloque también se procesaría dado que el primer cuarto y el último cuarto está ausente y, por lo tanto, la señal de impulso del siguiente bloque se colocaría inmediatamente después de la señal de impulso ilustrada del bloque actual en las Figuras 9a, 10a, 11a. Este procedimiento es muy eficiente, puesto que no son necesarias las operaciones de superposición/suma de la señal de impulso. Sin embargo, cualquier procedimiento de superposición/suma o cualquier procedimiento de atenuación de una trama a la siguiente con respecto a la señal de impulso también puede efectuarse si se requiere.

[0075] Las Figuras 9b, 10b, 11b ilustran espectrogramas. El eje horizontal representa el tiempo, pero no el tiempo con respecto a las muestras como en las Figuras 9a, 10a, 11a, sino el tiempo con respecto al número de bloque de DFT. El eje vertical ilustra el espectro de frecuencia de frecuencias bajas en el fondo de la Figura correspondiente hasta las frecuencias altas en la parte superior de las figuras correspondientes. El intervalo horizontal se extiende desde 0 hasta 16 kHz de modo que el cuarto inferior representa la señal original y los tres cuartos superiores representan la señal de síntesis. De este modo, las Figuras 9b, 10b, 11b ilustran la señal de audio de frecuencia mejorada de modo que el cuarto inferior de esas Figuras ilustra la señal de una fuente de audio.

[0076] Las figuras indican que la estructura de banda baja se refleja muy bien en la banda alta. Esto es particularmente visible con respecto a la Figura 10b que ilustra un diapasón donde se reproducen tres tonos diferentes del diapasón después del otro de izquierda a derecha en la Figura 10b. Particularmente, la primera porción del lado izquierdo de la Figura 10b es el tono más bajo del diapasón, la porción media es el tono medio del diapasón y la porción derecha de la Figura 10b es el tono más alto del diapasón. El diapasón se caracteriza específicamente por un espectro muy tonal y parece que en la presente invención es particularmente útil para replicar muy bien la estructura armónica en los 12 kHz más altos.

[0077] Con respecto al tercer elemento de prueba, se vuelve evidente que la estructura de banda baja para un elemento de música pop se transforma muy bien el intervalo de frecuencia alta por medio del procedimiento de la invención.

[0078] La Figura 12 ilustra una realización más que es un tanto similar a la realización de la Figura 6. Por lo tanto, los números de referencia similares en la Figura 6 indican elementos similares como en la Figura 12. Además de las características en la Figura 6, la realización de la Figura 12 comprende adicionalmente el separador de LF/HF 160, un generador de ruido aleatorio o ruido pseudo aleatorio 170 como una tabla de ruido o similar y un ajustador de energía 180.

[0079] El separador de LF/HF 160 efectúa una descomposición de la envolvente temporal en una envolvente de LF y una envolvente de HF. Preferiblemente, la envolvente de LF se determina por filtración de paso bajo y la envolvente de HF se determina sustrayendo la envolvente de HF de la envolvente de LF.

[0080] El generador de ruido aleatorio o ruido pseudo aleatorio 170 genera una señal de ruido y el ajustador de energía 180 ajusta la energía de ruido a la energía de la envolvente de HF que se estima en el bloque 180 también. Este ruido que tiene una energía ajustada a la energía de la envolvente de HF (sin ninguna contribución de la envolvente de LF) es sumado por el sumador 335 al tren de impulsos ponderados como salida por el bloque 315. Sin embargo, el orden de los bloques o etapas de procesamiento 315, 335, por ejemplo también puede modificarse.

[0081] Por otro lado, los procedimientos con respecto a los elementos 205 a 315 se aplican únicamente a la envolvente de LF según lo determinado por el bloque 160.

[0082] Una realización preferida que depende de la descomposición de la envolvente de banda completa en al menos dos partes comprende los siguientes bloques o etapas siguientes o en cualquier otro orden técnicamente factible:

60 Estimación de la envolvente temporal 100: Rectificación; compresión usando por ejemplo una función $x^{0.75}$; división subsecuente 160 de la envolvente en una envolvente de LF y una envolvente de HF.

[0083] La envolvente de LF se obtiene a través de la filtración de paso bajo, donde una frecuencia de cruce es por ejemplo de 2-6 kHz. En una realización, la envolvente de HF es la diferencia entre la envolvente original y una envolvente de LF ajustada preferiblemente retrasada.

[0084] Colocación de impulso sincronizada 300. La envolvente de LF derivada en la etapa descrita anteriormente se analiza por ejemplo por cálculo de curva y la colocación de impulso se efectúa en lugares máximos de la envolvente de LF.

[0085] Escalamiento de magnitud de impulsos individuales 315 derivada de la envolvente: El tren de impulsos montados en la etapa descrita anteriormente se pondera por pesos temporales derivados de la envolvente de LF.

[0086] Se estima la energía de la envolvente de HF y se suma el ruido aleatorio de la misma energía 335 al tren de impulsos ponderados.

[0087] Post-procesamiento, extracción de HF o selección de relleno de espacios: La señal "sin procesar" generada en la etapa descrita anteriormente en la salida del bloque 335 se post-procesa opcionalmente 320, por ejemplo por adición de ruido y se filtra 325 para usarse como HF en la BWE o como una señal mosaico objetivo de relleno de espacios.

[0088] Ajuste de energía 330: La distribución de energía espectral de la señal filtrada de la estimación de energía como se expone en una etapa anterior se ajusta para usarse como HF en la BWE como una señal mosaico objetivo de relleno de espacios. Aquí, preferiblemente se usa la información lateral del flujo de bits sobre la distribución de energía deseada.

[0089] Mezclado 400 de HF o señal de relleno de espacios con LF: Finalmente, la señal ajustada de la etapa 5 se mezcla con la salida del codificador central de acuerdo con la BWE usual o principios de relleno de espacios, es decir pasando un filtro de HP y completando la LF o rellenando huecos espectrales en las regiones espectrales de relleno de espacios.

[0090] Deberá mencionarse aquí que todas las alternativas o aspectos como se discutió anteriormente y todos los aspectos como se define por las reivindicaciones independientes en las siguientes reivindicaciones pueden usarse individualmente, es decir, sin ninguna otra alternativa u objeto que la alternativa, objeto o reivindicación independiente contemplada. Sin embargo, en otras realizaciones, pueden combinarse dos o más de las alternativas o aspectos de las reivindicaciones independientes entre sí y, en otras realizaciones, todos los aspectos o alternativas y todas las reivindicaciones independientes pueden combinarse entre sí.

[0091] Una señal de audio codificada de la invención puede almacenarse en un medio de almacenamiento digital o un medio de almacenamiento no transitorio o puede transmitirse sobre un medio de transmisión como un medio de transmisión inalámbrica o un medio de transmisión alámbrica tal como Internet. Aunque algunos aspectos se han descrito en el contexto de un aparato, está claro que esos aspectos también representan una descripción del procedimiento correspondiente, donde un bloque o dispositivo corresponde a una etapa de procedimiento o una característica de una etapa de procedimiento. De manera análoga, los aspectos descritos en el contexto de una etapa de procedimiento también representan una descripción de un bloque o elemento o característica correspondiente de un aparato correspondiente.

[0092] Dependiendo de ciertos requisitos de implementación, las realizaciones de la invención pueden implementarse en hardware o en software. La implementación puede efectuarse usando un medio de almacenamiento digital, por ejemplo un disco flexible, un DVD, un CD, una ROM, una PROM, una EPROM, una EEPROM o una memoria FLASH, que tenga señales de control legibles electrónicamente almacenadas en él, las cuales cooperan (o son capaces de cooperar) con un sistema informático programable de modo que se efectúe el procedimiento respectivo.

[0093] Algunas realizaciones según la invención comprenden un soporte de datos que tiene señales de control legibles electrónicamente, que son capaces de cooperar con un sistema informático programable, de modo que se efectúe uno de los procedimientos descritos en esta invención.

[0094] Generalmente, las realizaciones de la presente invención pueden implementarse como un producto de programa informático con un código de programa, operando el código de programa para efectuar uno de los procedimientos cuando el producto de programa informático se ejecuta en un ordenador. El código de programa puede almacenarse por ejemplo en un soporte legible por una máquina.

[0095] Otras realizaciones comprenden el programa informático para efectuar uno de los procedimientos descritos en esta invención, almacenados en un soporte legible por una máquina o un medio de almacenamiento no transitorio.

[0096] En otras palabras, una realización del procedimiento de la invención es, por lo tanto, un programa informático que tiene un código de programa para efectuar uno de los procedimientos descritos en esta invención, cuando el programa informático se ejecuta en un ordenador.

[0097] Una realización adicional del procedimiento de la invención es, por lo tanto, un soporte de datos (o un medio de almacenamiento digital o un medio legible por ordenador) que comprende, registrado en él, el programa informático para efectuar uno de los procedimientos descritos en esta invención.

5

[0098] Una realización adicional del procedimiento de la invención es, por lo tanto, un flujo de datos o una secuencia de señales que representan el programa informático para efectuar uno de los procedimientos descritos en esta invención. El flujo de datos o la secuencia de señales pueden configurarse por ejemplo para transferirse vía una conexión de comunicación de datos, por ejemplo a través de Internet.

10

[0099] Una realización adicional comprende un medio de procesamiento, por ejemplo un ordenador o un dispositivo lógico programable, configurado o adaptado para efectuar uno de los procedimientos descritos en esta invención.

15 **[0100]** Una realización adicional comprende un equipo que tiene instalado en él un programa informático para efectuar uno de los procedimientos descritos en esta invención.

[0101] En algunas realizaciones, puede usarse un dispositivo lógico programable (por ejemplo una matriz de puertas programables en el campo) para efectuar algunas o todas las funcionalidades de los procedimientos descritos en esta invención. En algunas realizaciones, una matriz de puertas programables en el campo puede cooperar con un microprocesador para efectuar uno de los procedimientos descritos en esta invención. Generalmente, los procedimientos se efectúan preferiblemente por medio de cualquier aparato de hardware.

20 **[0102]** Las realizaciones descritas anteriormente son meramente ilustrativas de los principios de la presente invención. Debe comprenderse que las modificaciones y variaciones de las disposiciones y detalles descritos en esta invención serán evidentes a aquellos expertos en el estado de la técnica. Se pretende, por lo tanto, que estén limitadas únicamente por el alcance de las reivindicaciones de patente anexas y no por los detalles específicos presentados a modo de descripción y explicación de las realizaciones en esta invención.

Referencias

[103]

[1] Dietz, M., Liljeryd, L., Kjörling, K., and Kunz, O., "Spectral Band Replication, a Novel Approach in Audio Coding," in Audio Engineering Society Convention 112, 2002.

[2] Disch, S., Niedermeier, A., Helmrich, C. R., Neukam, C., Schmidt, K., Geiger, R., Lecomte, J., Ghido, F., Nagel, F., and Edler, B., "Intelligent Gap Filling in Per-ceptual Transform Coding of Audio," in Audio Engineering Society Convention 141, 2016.

[3] Laitinen M-V., Disch S., Oates C., Pulkki V. "Phase derivative correction of bandwidth extended signals for perceptual audio codecs." In 140th Audio Engineer-ing Society International Convention 2016, AES 2016. Audio Engineering Society. 2016.

[4] Atti, Venkatraman, Venkatesh Krishnan, Duminda A. Dewasurendra, Venkata Chebiyyam, Shaminda Subasingha, Daniel J. Sinder, Vivek Rajendran, Imre Varga, Jon Gibbs, Lei Miao, Volodya Grancharov and Harald Pobloth. "Super-wideband bandwidth extension for speech in the 3GPP EVS codec." 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2015): 5927-5931.

[5] US 2008/027711 A1.

REIVINDICACIONES

1. Procesador de audio para generar una señal de audio de frecuencia mejorada (420) a partir de una señal de audio fuente (30), que comprende:
5 un determinador de envolvente (100) para determinar una envolvente temporal de al menos una porción de la señal de una fuente de audio (30);
 un analizador (200) para analizar la envolvente temporal para determinar valores de ciertas características de la envolvente temporal;
10 un sintetizador de señales (300) para generar una señal de síntesis, comprendiendo la generación colocar impulsos en relación a los valores temporales determinados, donde los impulsos se ponderan usando pesos derivados de las amplitudes de la envolvente temporal relacionada con los valores temporales, donde se colocan los impulsos; y
 un combinador (400) para combinar al menos una banda de la señal de síntesis que no se incluye en la señal de audio fuente (30) y la señal de audio fuente (30) para obtener la señal de audio de frecuencia mejorada (420).
15 2. Procesador de audio de la reivindicación 1, donde el analizador (200) se configura para determinar los valores temporales de máximos locales o mínimos locales como características ciertas.
20 3. Procesador de audio de la reivindicación 1 o 2, donde el determinador de envolvente (100) se configura
 para descomponer (105) una señal de audio fuente (30) en una pluralidad de señales de subbanda,
 para calcular (120) una envolvente temporal seleccionada de una señal de subbanda seleccionada de la pluralidad de señales de subbanda, siendo la envolvente temporal seleccionada la envolvente temporal, o
25 para calcular (120) al menos dos envolventes temporales de al menos dos señales de subbanda de la pluralidad de señales de subbanda y combinar (130) al menos las dos señales de subbanda para obtener una envolvente combinada temporal como la envolvente temporal.
30 4. Procesador de audio de la reivindicación 3,
 donde el determinador de la envolvente (100) se configura para normalizar (115, 125) o filtrar las señales de subbanda seleccionadas o las envolventes temporales antes de la combinación (130), o donde la combinación comprende una operación de promediación, o donde el determinador de envolvente (100) se configura para
35 calcular envolventes temporales de todas las señales de subbanda de la pluralidad de señales de subbanda, o donde el determinador de envolvente (100) se configura para determinar una sola envolvente temporal de banda ancha de la señal de audio fuente (30) como la envolvente temporal.
40 5. Procesador de audio de una de las reivindicaciones anteriores, donde el determinador de envolvente (100) se configura para determinar la envolvente temporal
 usando un seguidor de envolvente configurado para rectificar (145) una forma de onda y filtración de paso bajo (150) de la forma de onda rectificadora, o
 usando el cálculo de valores o potencias absolutas de valores absolutos de una forma de onda digital y
45 posteriormente filtrar por paso bajo como resultado, o
 usando el cálculo de un valor de la media cuadrática instantánea de la raíz de la forma de onda a través de una ventana deslizante con un ancho de ventana definido, o
 usando una aproximación lineal por piezas de la forma de onda, donde la envolvente temporal se determina encontrando y conectando picos de la forma de onda en una ventana deslizable que se mueve a través de un
50 resultado de la aproximación lineal por piezas, o
 usando una transformación de Hilbert para generar una señal analítica para la forma de onda y calculando la envolvente de la señal de audio fuente (30) y usando la señal analítica operaciones de cuadratura, operaciones de suma y operaciones de raíz cuadrada.
55 6. Procesador de audio de una de las reivindicaciones anteriores, donde el analizador (200) se configura
 para determinar (205) los valores temporales iniciales de ciertas características, y
 para derivar (210), de los valores temporales iniciales, los valores temporales usando la función de optimización o, usando información lateral asociada con la señal de audio fuente (30) o, seleccionando o manipulando los
60 valores adicionales de acuerdo con una regla de procesamiento.
7. Procesador de audio de la reivindicación 6, donde la regla de procesamiento de la función de optimización se implementa de modo que los valores temporales se colocan en una trama con un espaciado de trama (T), donde el espaciado de trama (T) y una posición de la trama dentro de la envolvente temporal es para que
65 un valor de desviación entre los valores temporales y los valores temporales iniciales tenga una característica

predeterminada.

8. Procesador de audio de la reivindicación 7, donde el valor de desviación es una suma sobre las diferencias cuadradas y donde la característica predeterminada es una característica mínima.

9. Procesador de audio de una de las reivindicaciones anteriores, donde el sintetizador de señales (300) se configura

para colocar (305) únicamente impulsos positivos o únicamente negativos para obtener un tren de impulsos, y para ponderar subsecuentemente (315) los impulsos en el tren de impulsos, o para ponderar únicamente los impulsos negativos o positivos usando los pesos correspondientes asociados con los valores temporales de los impulsos en un tren de impulsos, y para colocar los impulsos ponderados en los valores respectivos temporales para obtener el tren de impulsos.

10. Procesador de audio de una de las reivindicaciones anteriores, donde el sintetizador de señales (300) se configura para derivar (310) los pesos de las amplitudes usando una función de compresión, siendo la función de compresión una función del grupo de funciones que comprende: una función de energía con una energía menor que 1, una función logarítmica, una función de raíz cuadrada y una función no lineal configurada para reducir valores más altos e incrementar los valores más bajos.

11. Procesador de audio de una de las reivindicaciones anteriores, donde el sintetizador de señales (300) se configura para efectuar una función de post-procesamiento (320), comprendiendo la función de post-procesamiento al menos una del grupo de funciones que consiste en adición de ruido, adición de una armónica ausente, filtración inversa y ajuste de envolvente (330).

12. Procesador de audio de una de las reivindicaciones anteriores,

donde el determinador de envolvente (100) se configura para descomponer (160) la envolvente temporal en una porción de baja frecuencia y una porción de alta frecuencia,

donde el analizador (200) se configura para usar la porción de baja frecuencia de la envolvente temporal para el análisis.

13. Procesador de audio de la reivindicación 12, donde el sintetizador de señales (300) se configura para generar (170) ruido ajustado por energía (180) y para sumar (335) el ruido ajustado por energía (180) a una señal que comprende impulsos ponderados o no ponderados para obtener la señal de síntesis.

14. Procesador de audio de una de las reivindicaciones anteriores,

donde el sintetizador de señales (300) se configura para filtrar por paso alto (325) o filtrar por paso de banda (325) una señal que comprende los impulsos colocados y ponderados para obtener al menos una banda de la señal de síntesis que no se incluyó en la señal de audio fuente (30) y para efectuar un ajuste de envolvente espectral (330) con la banda de la señal de síntesis o, donde el ajuste de la envolvente espectral se efectúa usando valores de ajuste de envolvente derivados de información lateral asociada con la señal de audio fuente (30) o usando valores de ajuste de envolvente derivados de la señal de audio fuente (30) o de acuerdo con una función de ajuste de envolvente predeterminada.

15. Procesador de audio de una de las reivindicaciones anteriores, donde la señal de audio fuente (30) es una señal de audio de dominio del tiempo,

donde al menos la banda de la señal de síntesis es una señal de audio de dominio del tiempo, y donde el combinador (400) se configura para efectuar una combinación de dominio del tiempo usando una adición muestra por muestra de muestras de al menos una banda de la señal de síntesis y muestras correspondientes de la señal de audio fuente (30).

16. Procesador de audio de una de las reivindicaciones anteriores, donde la señal de audio fuente (30) es una señal de excitación en un dominio de LPC, Codificación Predictiva Lineal,

donde al menos una banda de la señal de síntesis es una señal de excitación en el dominio de LPC, donde el combinador (400) se configura para combinar (405) la señal de audio fuente (30) y al menos una banda por adición muestra por muestra en el dominio de LPC, donde el combinador (400) se configura para filtrar (410) un resultado de la adición muestra por muestra usando un filtro de síntesis de LPC para obtener la señal de audio de frecuencia mejorada (420), y donde el filtro de síntesis de LPC (410) se controla por medio de datos de LPC (40a) asociados con la señal de audio fuente (30) como información lateral y donde el filtro de síntesis de LPC (410) se controla adicionalmente por información de la envolvente (40b) para al menos una banda de la señal de síntesis.

17. Procesador de audio de una de las reivindicaciones anteriores,

5 donde el analizador (200), el sintetizador de señales (300) y el combinador (400) operan en un dominio del tiempo o un dominio del tiempo de LPC.

18. Procesador de audio de una de las reivindicaciones 1 a 16, donde el determinador de envolvente (100) se configura para aplicar una conversión espectral (105) para extraer una pluralidad de señales de paso de banda de una secuencia de tramas,

10 donde el sintetizador de señales (300) se configura para aplicar una conversión espectral, para extraer (325) al menos una banda de la señal de síntesis y para efectuar (330) un ajuste de envolvente para al menos una banda, y

15 donde el combinador (400) se configura para combinar en el dominio espectral y para aplicar una conversión en el dominio del tiempo para obtener la señal de audio de frecuencia mejorada (420).

19. Procedimiento de generación de una señal de audio de frecuencia mejorada (420) a partir de una señal de audio fuente (30), que comprende:

20 determinar (100) una envolvente temporal de al menos una porción de la señal de audio fuente (30);
analizar (200) la envolvente temporal para determinar valores temporales de ciertas características de la envolvente temporal;

25 colocar (305) impulsos en relación con los valores temporales determinados en una señal de síntesis, donde, en la señal de síntesis, los impulsos colocados se ponderan usando pesos derivados de amplitudes de la envolvente temporal relacionados con los valores temporales, donde se colocan los impulsos; y

combinar (400) al menos una banda de la señal de síntesis que no se incluyó en la señal de audio fuente (30) y la señal de audio fuente (30) para obtener la señal de audio de frecuencia mejorada (420).

20. Programa informático que comprende instrucciones que, cuando son ejecutadas por un ordenador o
30 un procesador, hacen que el ordenador o el procesador lleven a cabo el procedimiento de generar una señal de audio de frecuencia mejorada a partir de una señal de audio fuente de la reivindicación 19.

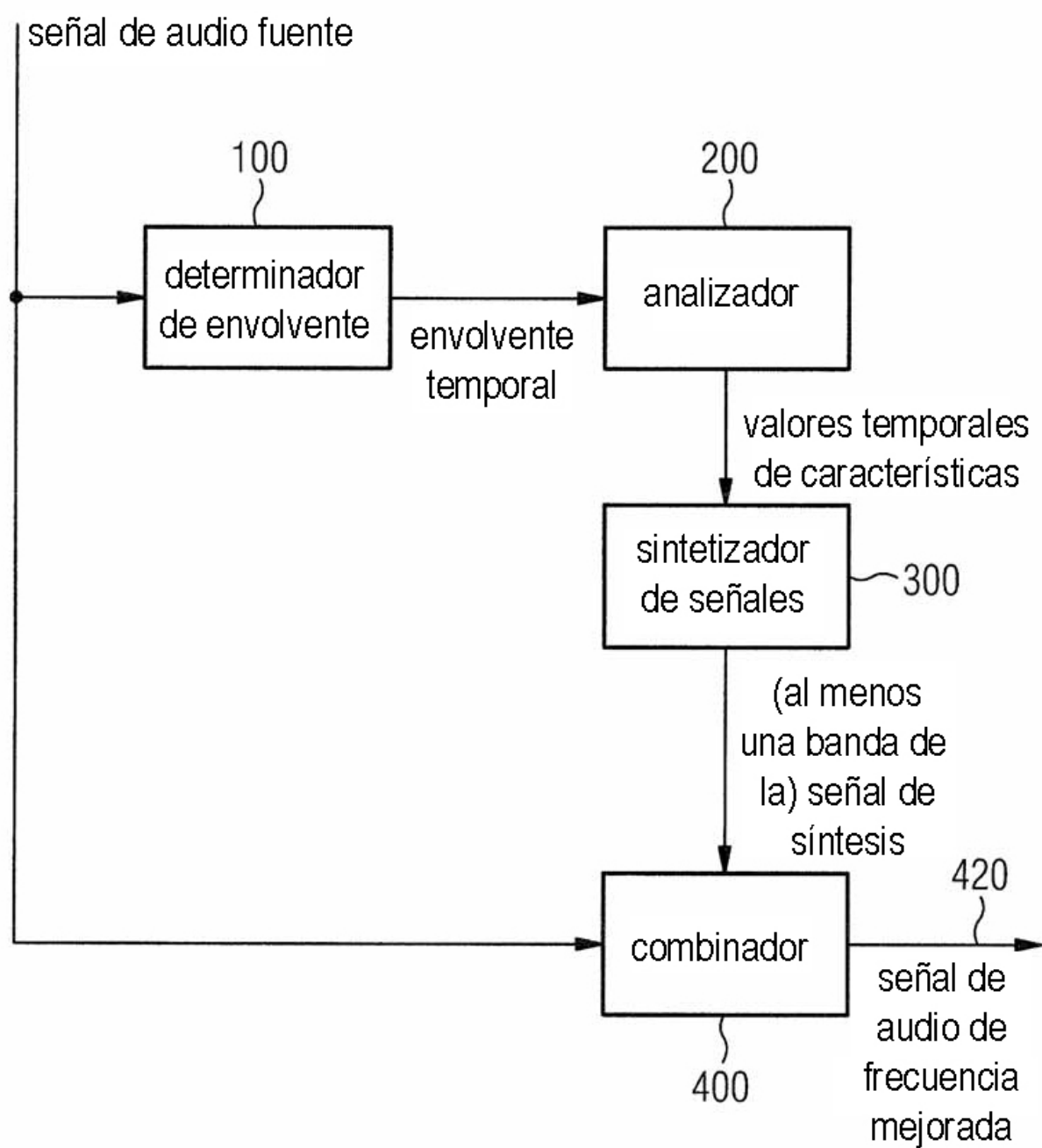


Fig. 1

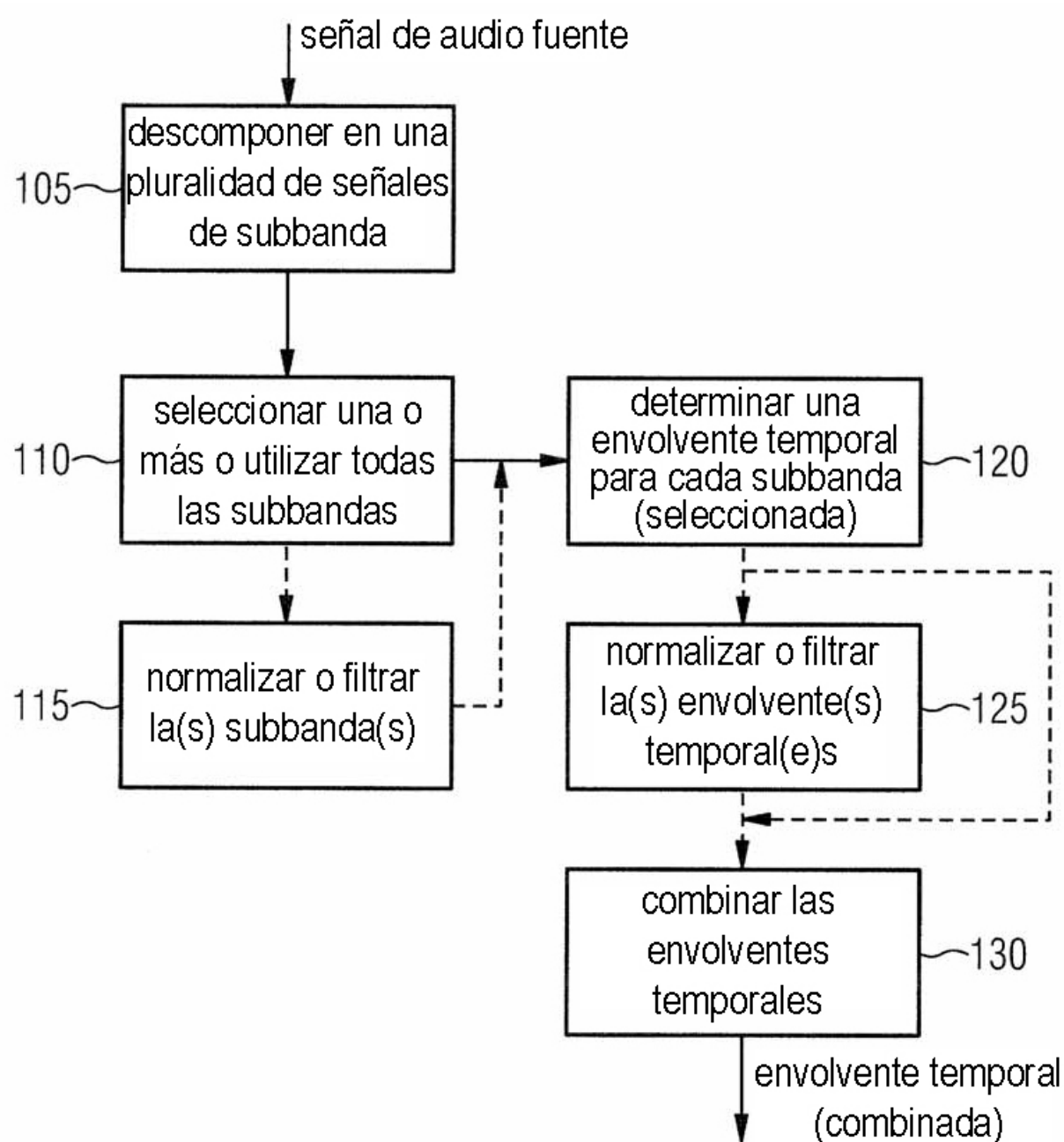


Fig. 2

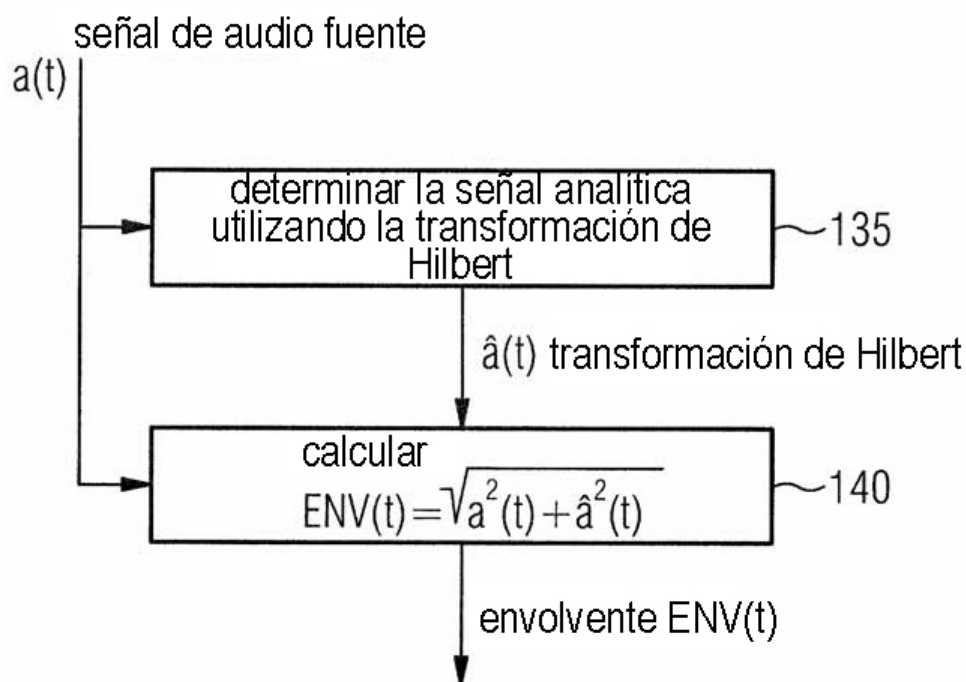


Fig. 3a

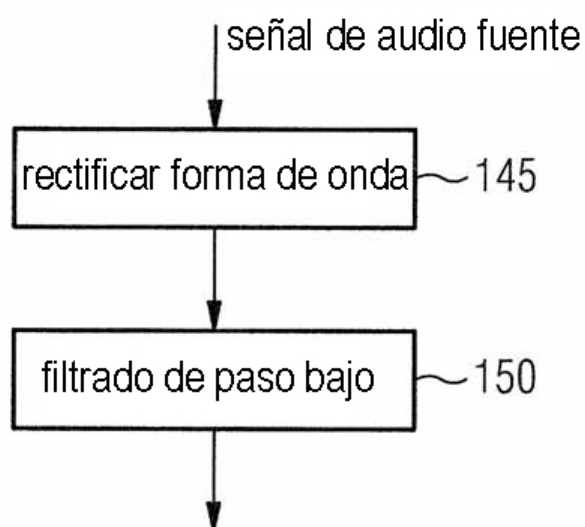


Fig. 3b

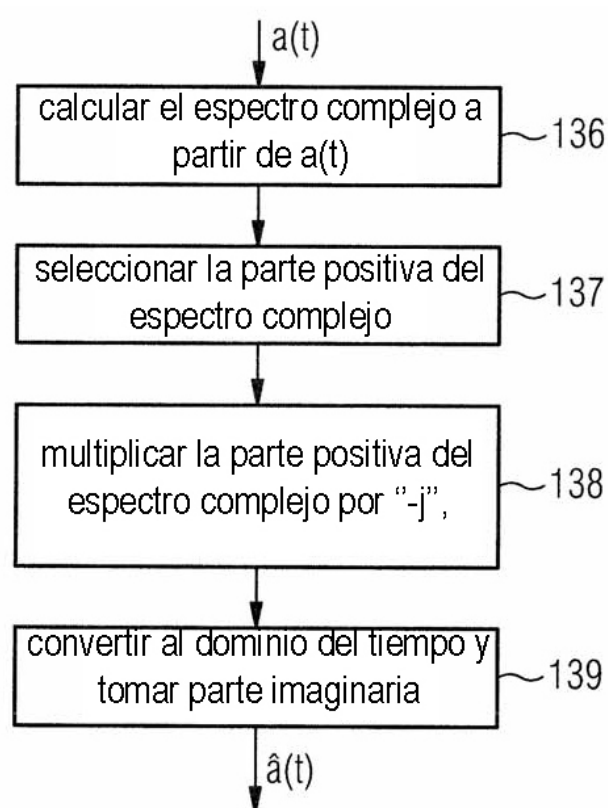


Fig. 3c

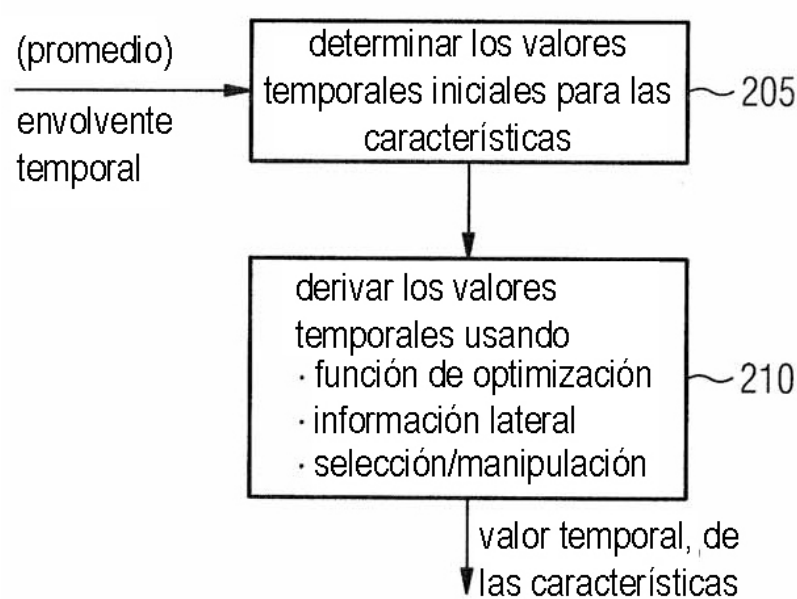


Fig. 4

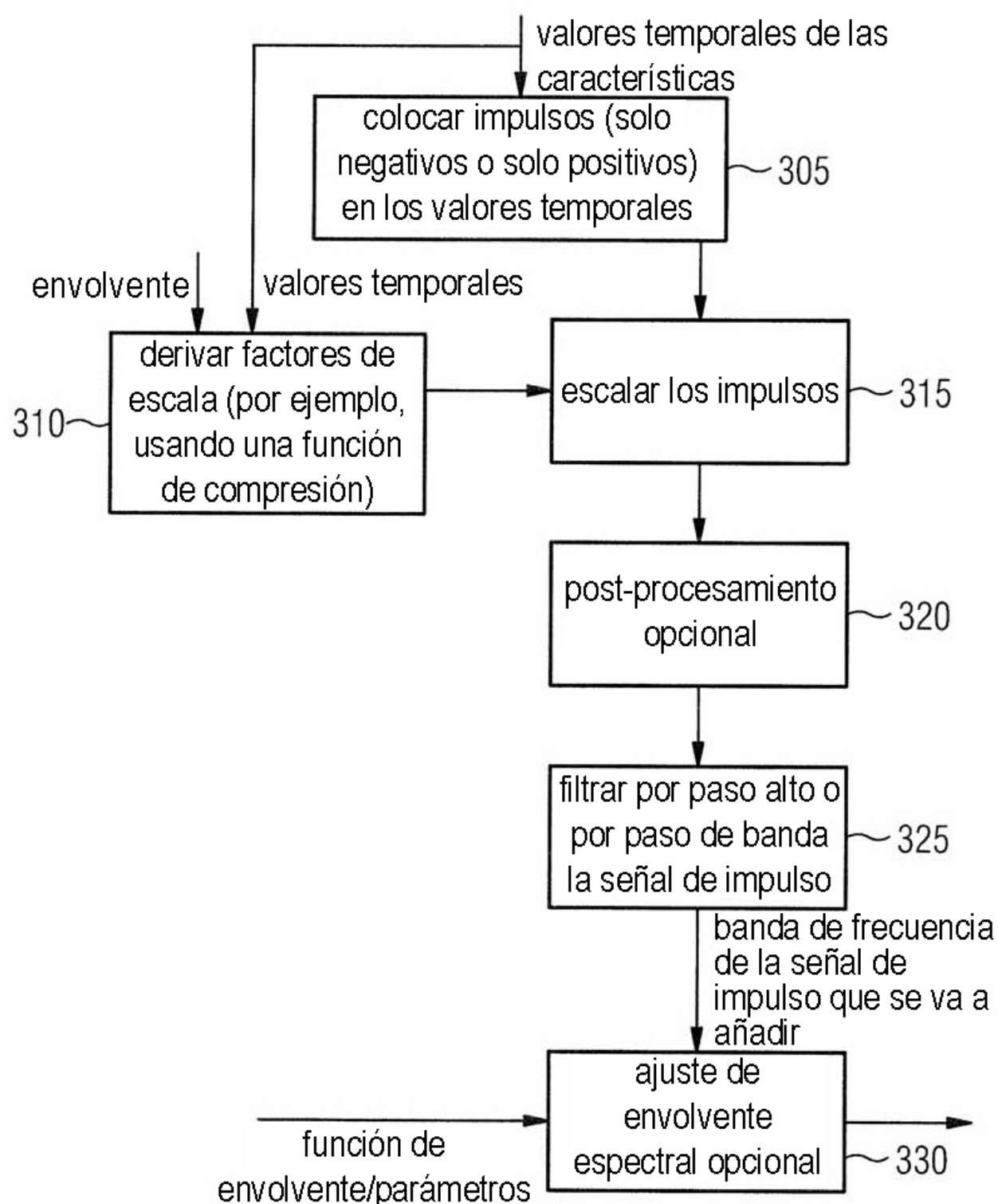


Fig. 5

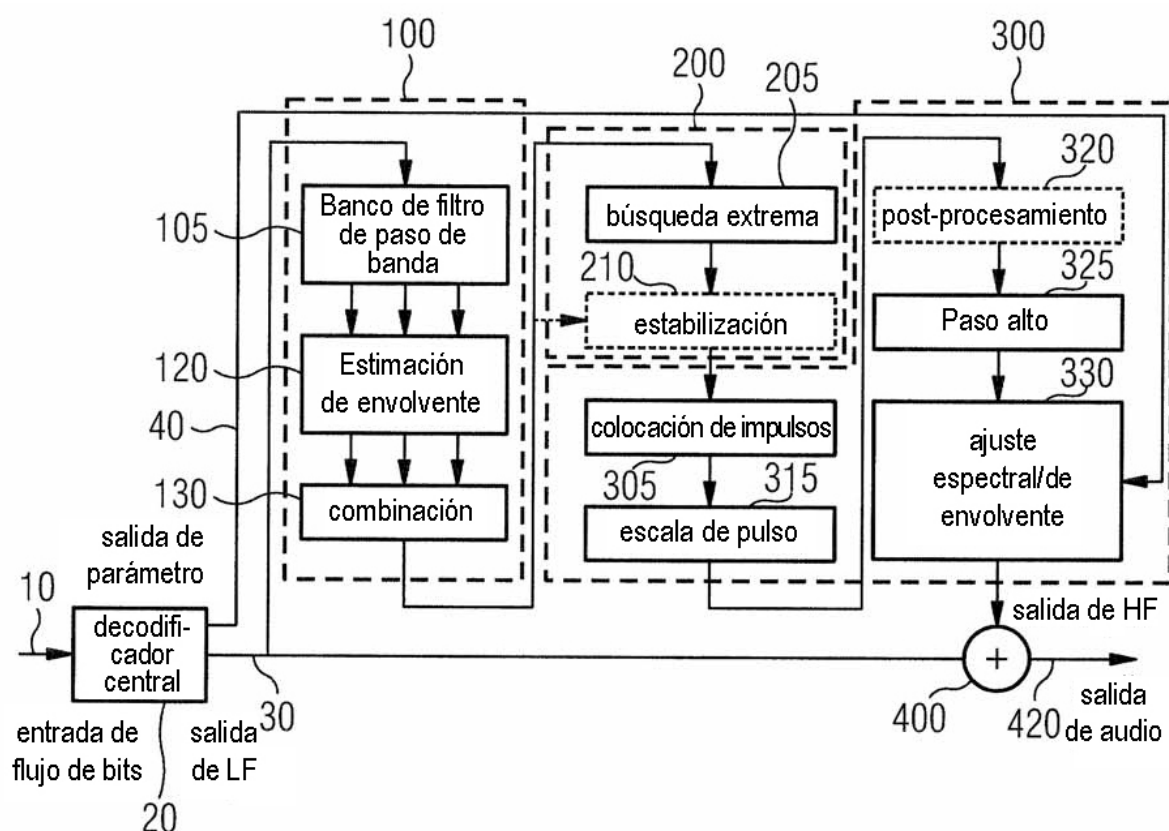


Fig. 6

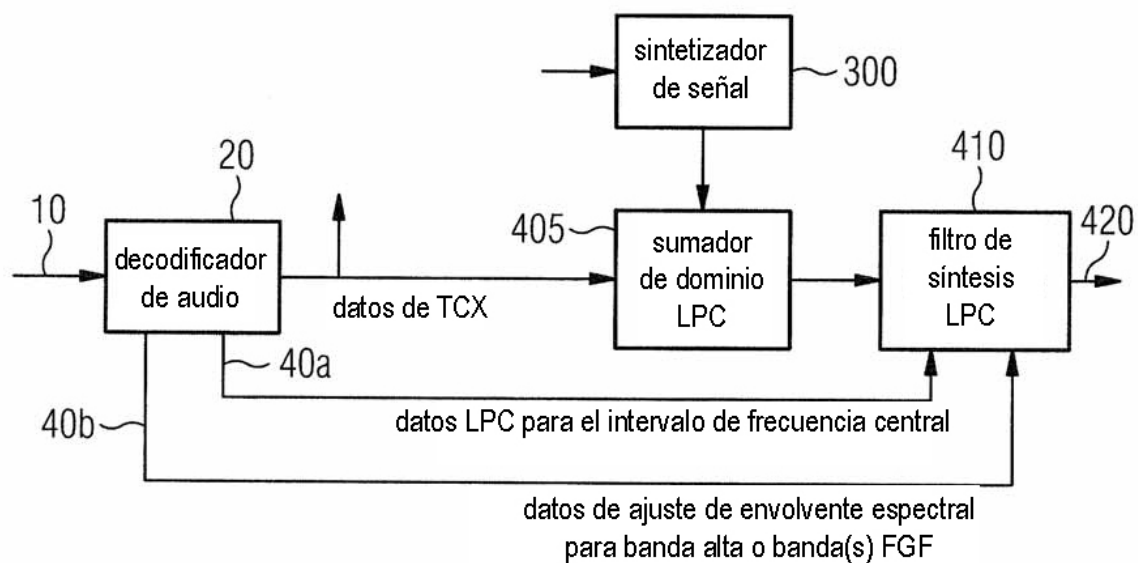


Fig. 7

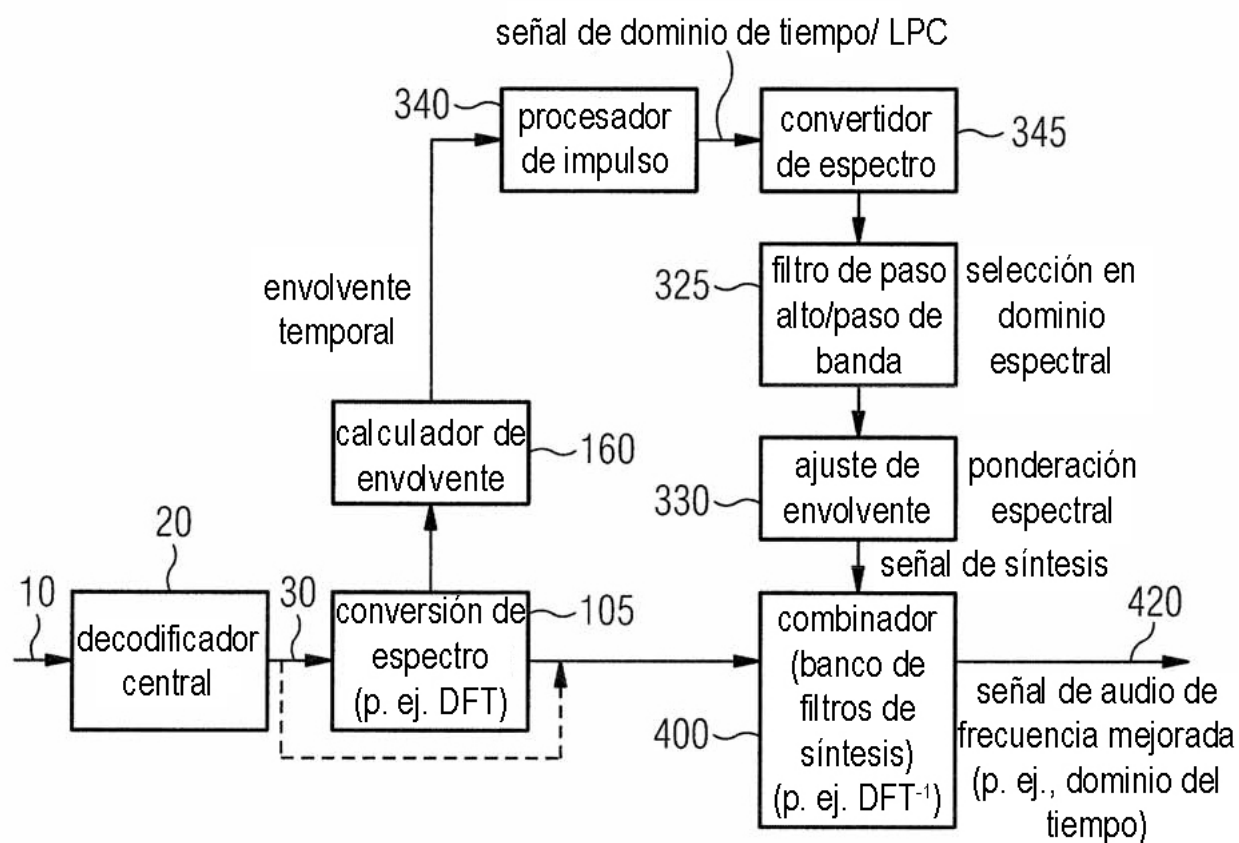


Fig. 8

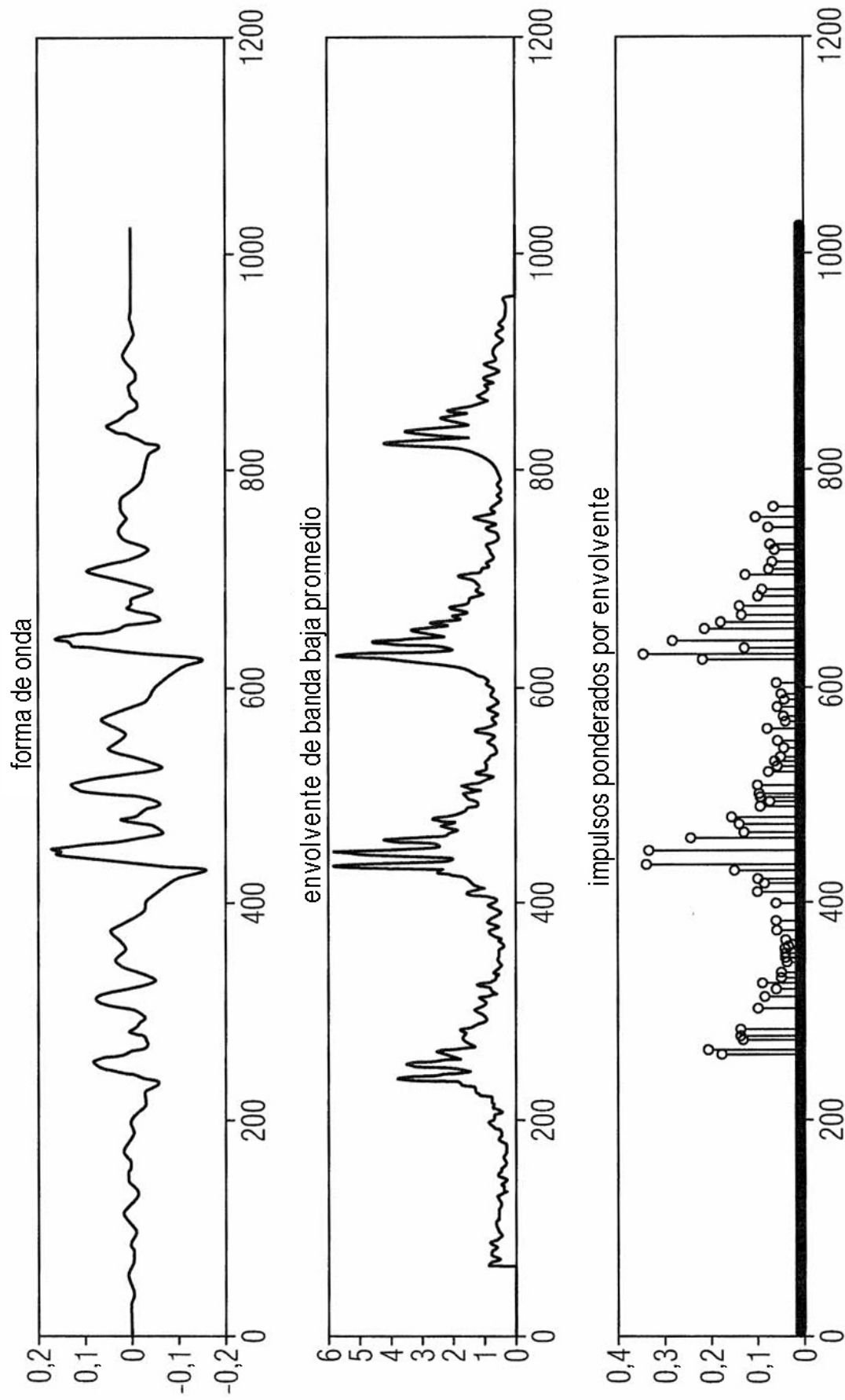


Fig. 9a

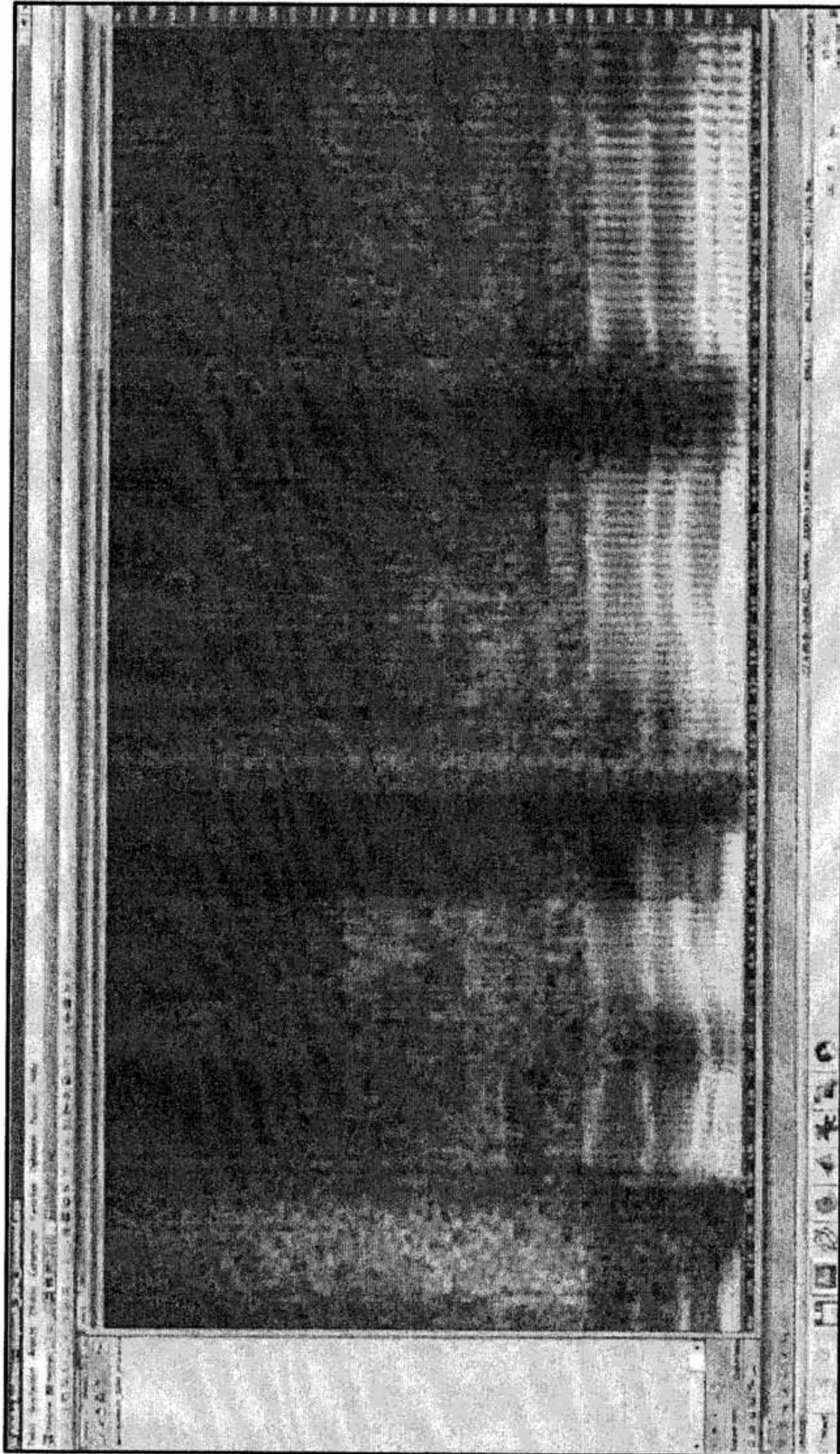


Fig. 9b

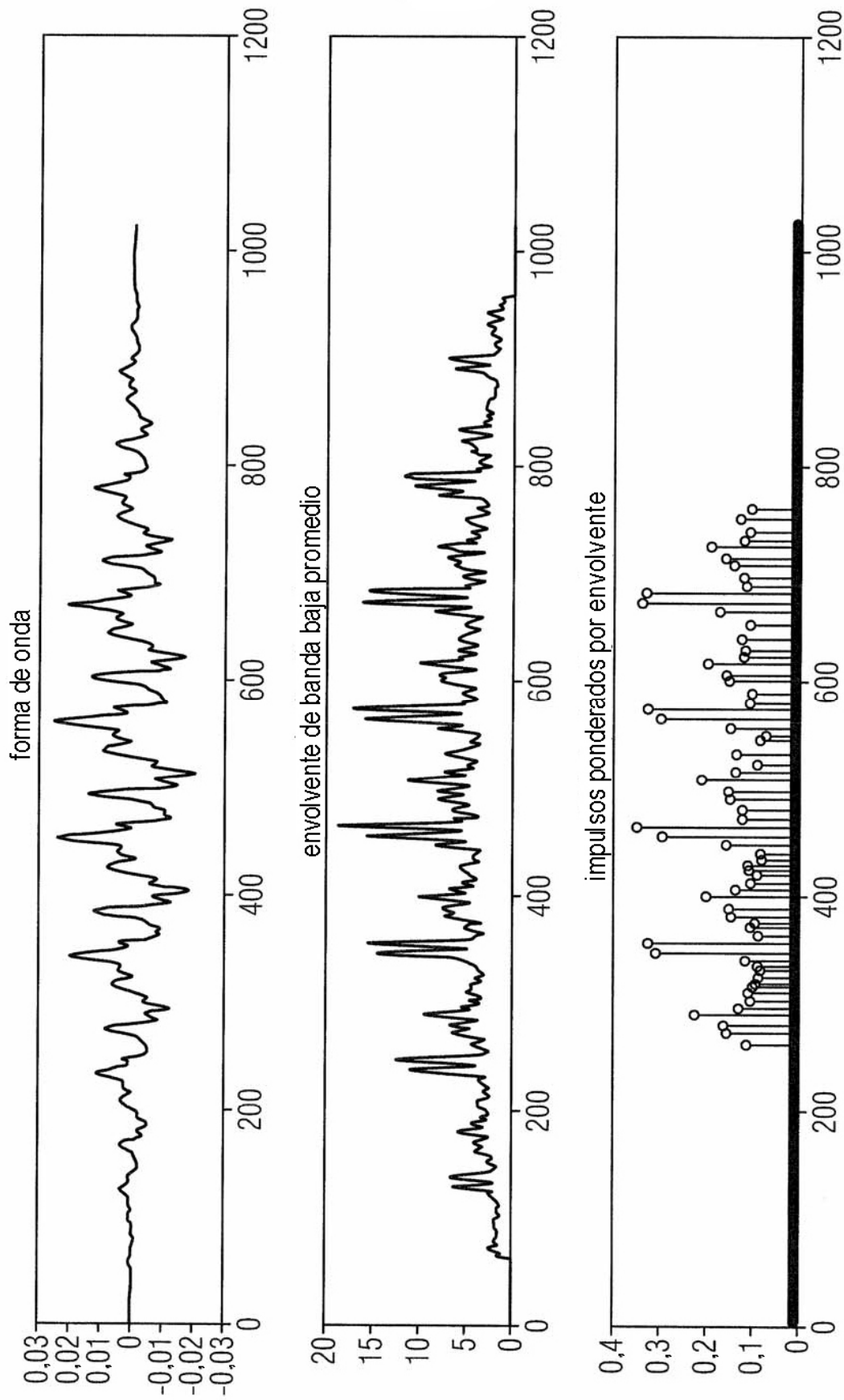


Fig. 10a

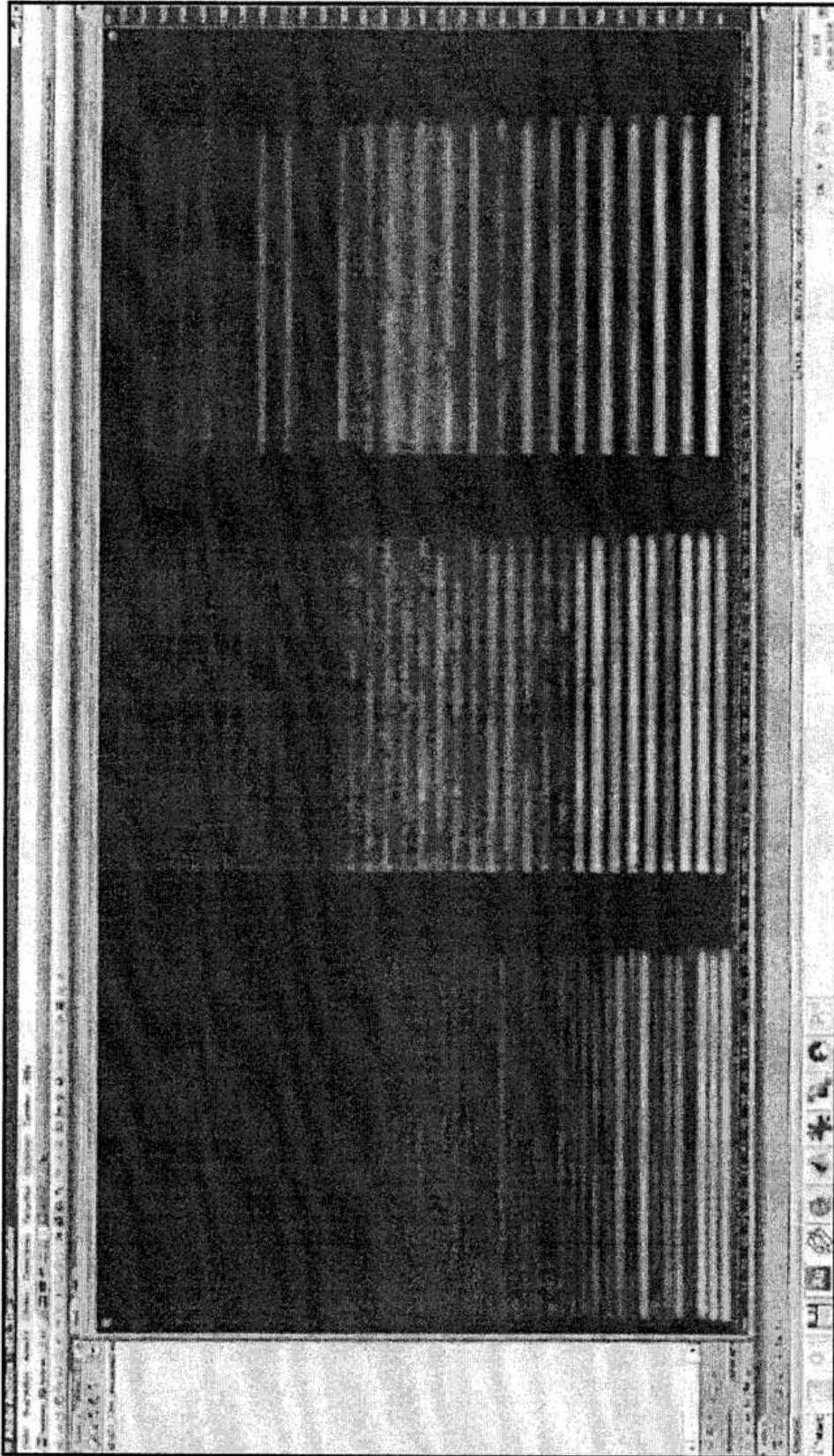


Fig. 10b

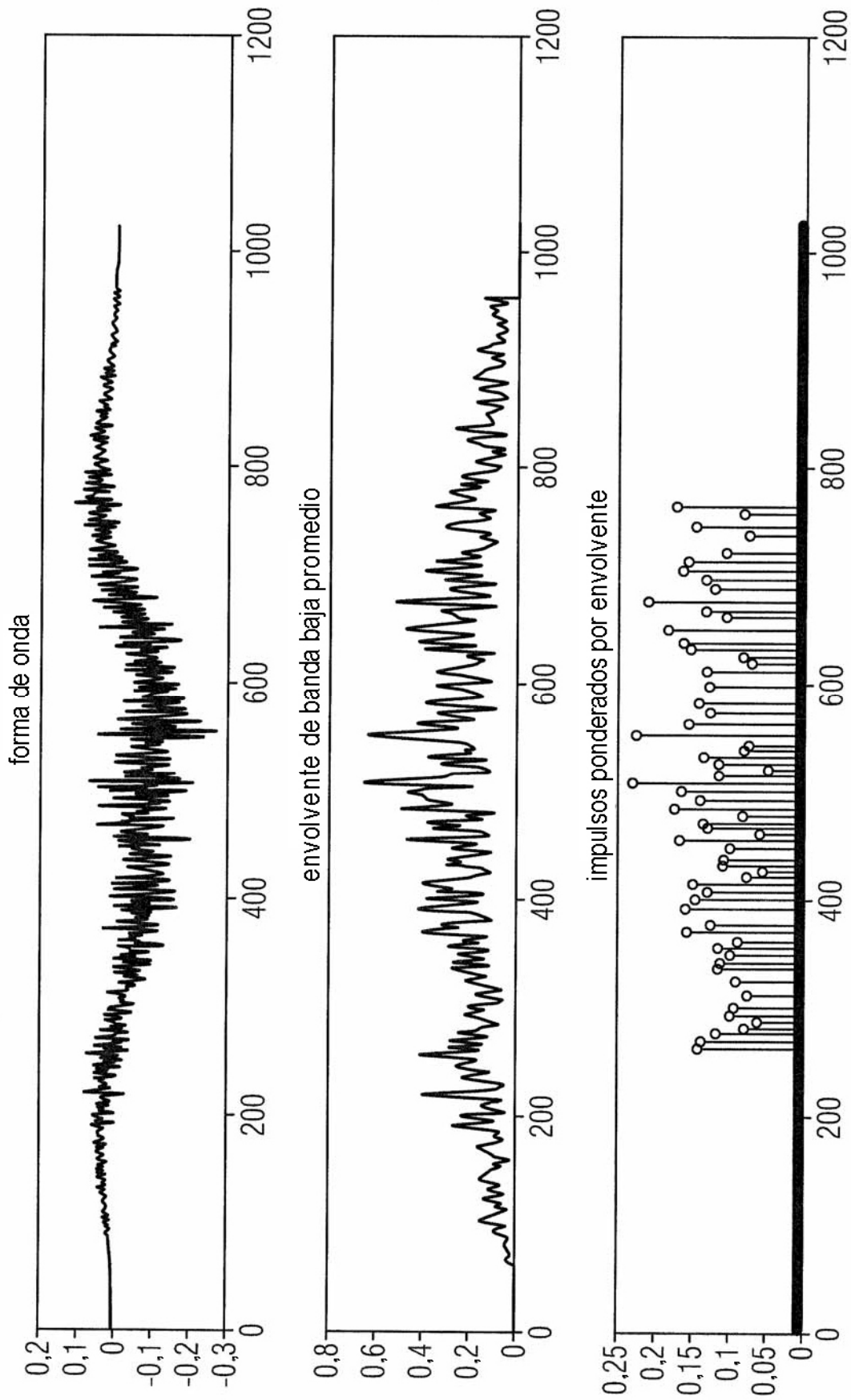


Fig. 11a

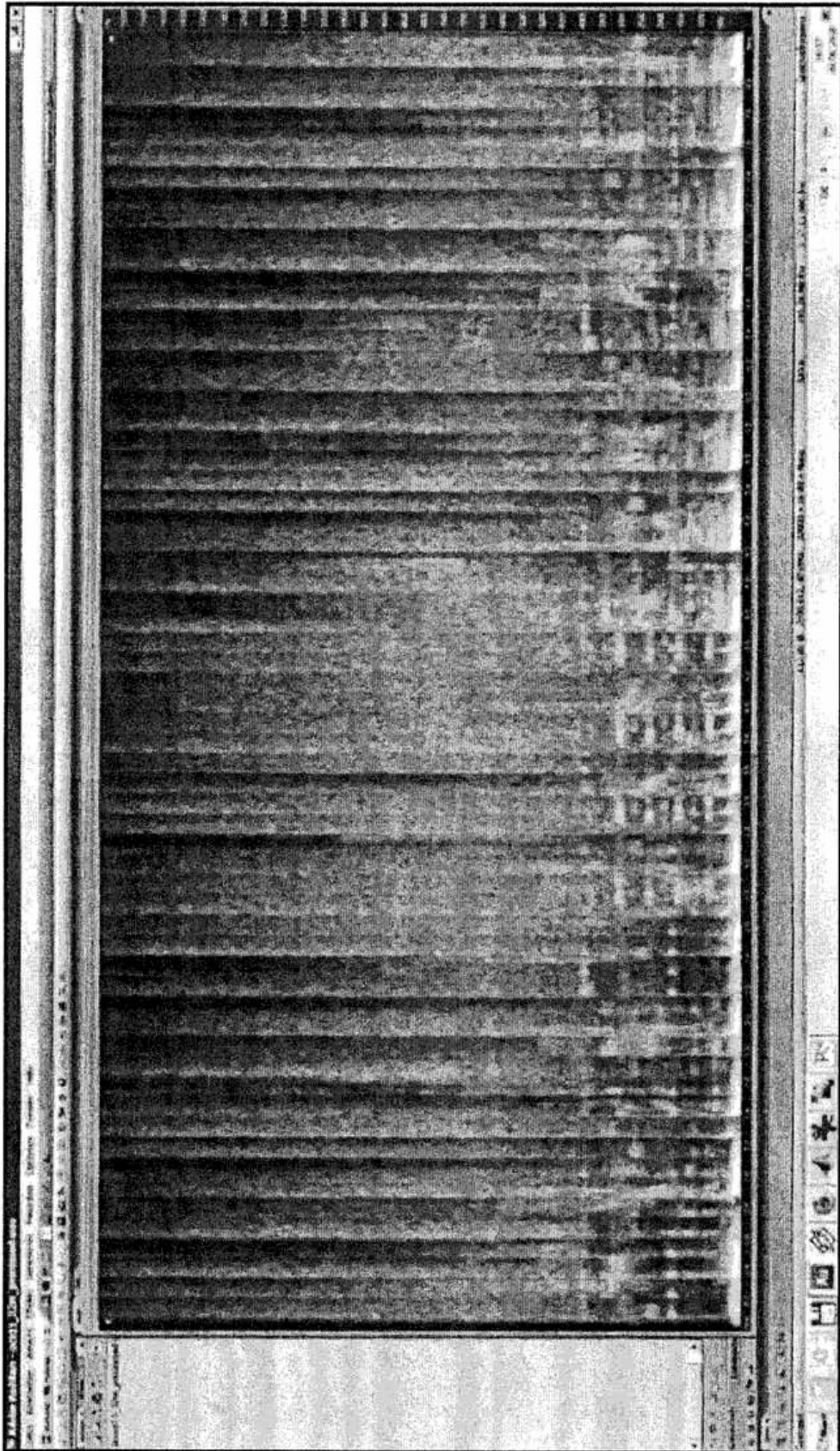


Fig. 11b

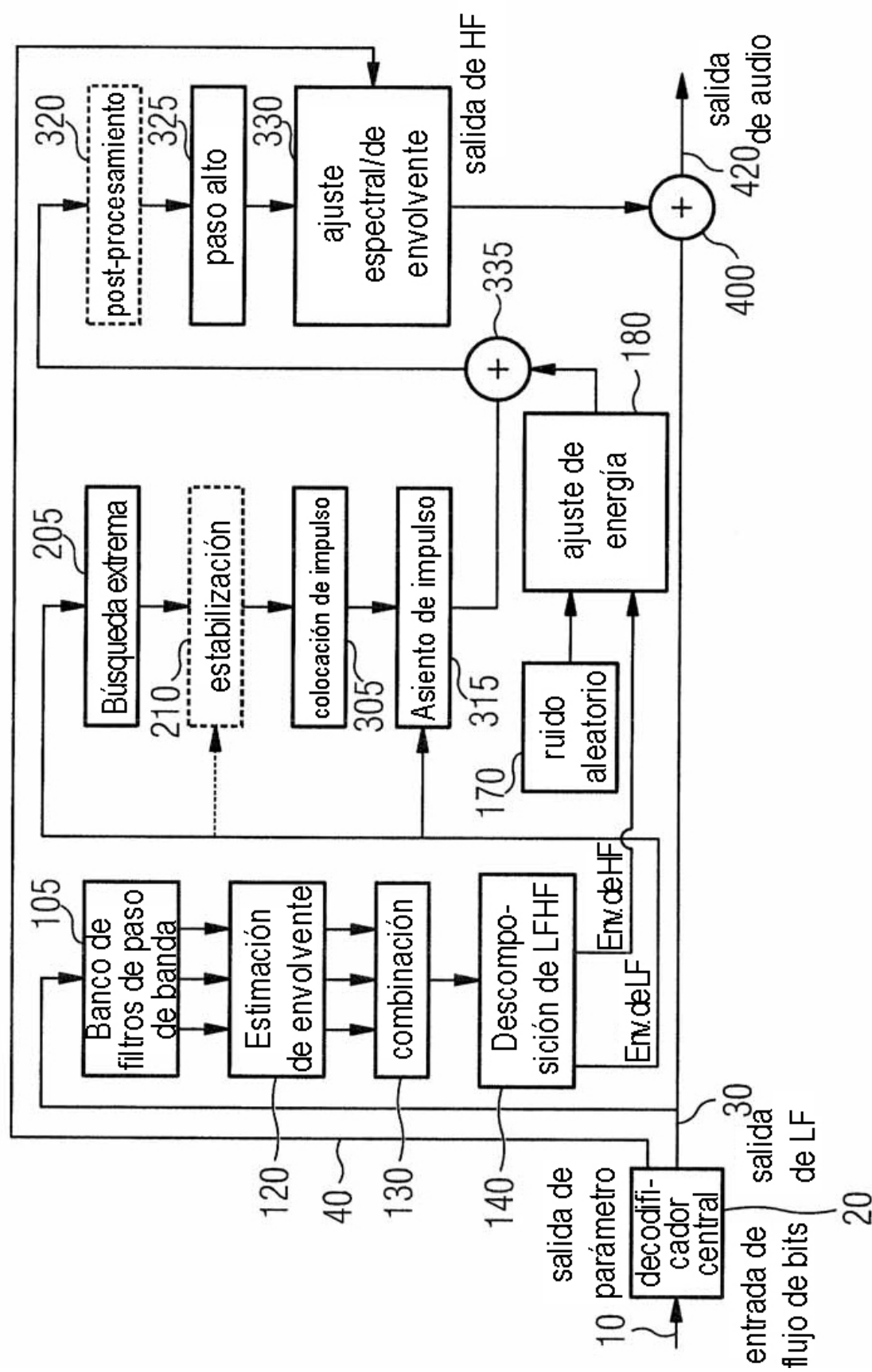


Fig. 12