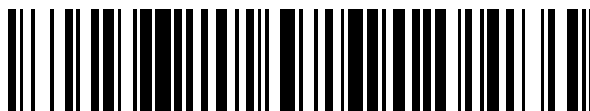


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 693 051**

51 Int. Cl.:

G10L 19/028 (2013.01)

G10L 21/038 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **24.07.2015** **PCT/EP2015/067058**

87 Fecha y número de publicación internacional: **04.02.2016** **WO16016144**

96 Fecha de presentación y número de la solicitud europea: **24.07.2015** **E 15741557 (1)**

97 Fecha y número de publicación de la concesión europea: **15.08.2018** **EP 3175449**

54 Título: **Aparato y procedimiento para generar una señal mejorada mediante el uso de relleno de ruido independiente**

30 Prioridad:

28.07.2014 EP 14178777

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

07.12.2018

73 Titular/es:

**FRAUNHOFER-GESELLSCHAFT ZUR
FÖRDERUNG DER ANGEWANDTEN
FORSCHUNG E.V. (100.0%)
Hansastraße 27c
80686 München, DE**

72 Inventor/es:

**DISCH, SASCHA;
GEIGER, RALF;
NIEDERMEIER, ANDREAS;
NEUSINGER, MATTHIAS;
SCHMIDT, KONSTANTIN;
WILDE, STEPHAN;
SCHUBERT, BENJAMIN y
NEUKAM, CHRISTIAN**

74 Agente/Representante:

SALVÀ FERRER, Joan

ES 2 693 051 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

Aparato y procedimiento para generar una señal mejorada mediante el uso de relleno de ruido independiente

5 **[0001]** La solicitud se refiere al procesamiento de señales y, en particular, al procesamiento de señales de audio.

[0002] La codificación perceptual de señales de audio con fines de reducción de datos para el almacenamiento eficaz o la transmisión de estas señales es una práctica ampliamente usada. En particular, cuando
10 se han de obtener mínimas tasas de bits, la codificación empleada lleva a una reducción de la calidad de audio que a menudo es causada primariamente por una limitación en el lado del codificador del ancho de banda de la señal de audio que se va a transmitir. En códecs contemporáneos, existen procedimientos bien conocidos para la restauración de las señales del lado del decodificador mediante extensión de ancho de banda de señales de audio (BWE), por ejemplo, replicación de banda espectral (SBR).

15 **[0003]** En una codificación de baja tasa de bits, se emplea a menudo también el denominado relleno de ruido. Prominentes regiones espectrales que fueron cuantificadas a cero debido a estrictas restricciones de tasas de bits se rellenan con ruido sintético en el decodificador.

20 **[0004]** Usualmente, ambas técnicas se combinan en aplicaciones de codificación de bajas tasas de bits. Según, existen soluciones integradas tales como relleno inteligente de intervalos (Intelligent Gap Filling, IGF) que combinan codificación de audio, relleno de ruido y relleno de intervalo espectral.

[0005] Sin embargo, todos estos procedimientos tienen en común que, en una primera etapa, la banda de
25 base o señal de audio central se reconstruye mediante el uso de decodificación en forma de onda y relleno de ruido y, en una segunda etapa, el procesamiento de BWE o IGF se lleva a cabo mediante el uso de la señal fácilmente reconstruida. Esto lleva al hecho de que los mismos valores de ruido que fueron rellenos en la banda de base por relleno de ruido durante la reconstrucción se usan para regenerar las partes faltantes en la banda alta (en BWE) o para el relleno de las brechas espectrales restantes (en IGF). Mediante el uso de ruido altamente correlacionado
30 para reconstruir múltiples regiones espectrales en BWE o IGF pueden llevar a alteraciones perceptuales.

[0006] Puntos relevantes en el estado de la técnica comprenden

- SBR como un post-procesador para decodificación en forma de onda [1–3]
- 35 • AAC PNS [4]
- relleno de ruido MPEG–D USAC [5]
- G.719 y G.722.1C [6]
- MPEG–H 3D IGF [8]

40 **[0007]** Los siguientes documentos y solicitudes de patente describen procedimientos que se consideran relevantes para la solicitud:

- [1] M. Dietz, L. Liljeryd, K. Kjörling y O. Kunz, "Spectral Band Replication, a novel approach in audio coding", en 112th AES Convention, Múnich, Alemania, 2002.
- 45 [2] S. Meltzer, R. Böhm y F. Henn, "SBR enhanced audio codecs for digital broadcasting such as "Digital Radio Mondiale" (DRM)," en 112th AES Convention, Múnich, Alemania, 2002.
- [3] T. Ziegler, A. Ehret, P. Ekstrand y M. Lutzky, "Enhancing mp3 with SBR: Features and Capabilities of the new mp3PRO Algorithm", en 112th AES Convention, Múnich, Alemania, 2002.
- [4] J. Herre, D. Schulz, Extending the MPEG–4 AAC Codec by Perceptual Noise Substitution, Audio Engineering
50 Society 104th Convention, Preprint 4720, Ámsterdam, Holanda, 1998.
- [5] Solicitud de patente europea EP2304720 USAC relleno de ruido
- [6] ITU–T Recomendaciones G.719 y G.722.1C
- [7] EP 2704142
- [8] EP 13177350

55 **[0008]** Las señales de audio procesadas con estos procedimientos sufren de artefactos tales como asperezas, distorsiones de modulación y un timbre percibido como desagradable, en particular a una baja tasa de bits y, en consecuencia, bajo ancho de banda y/o la aparición de agujeros espectrales en el rango LF. La razón de ello es, como se explica más abajo, primariamente el hecho de que los componentes reconstruidos del espectro

extendido o de intervalo relleno se basan en una o más copias directas que contienen ruido de la banda de base. Las modulaciones temporales que resultan de dicha correlación no deseada en ruido reconstruido son audibles de una manera perturbadora como aspereza perceptual o distorsión objetable. Todos los procedimientos existentes como mp3+SBR, AAC+SBR, USAC, G.719 y G.722.1C y también MPEG-H 3D IGF realizan primero una
5 decodificación central completa que incluye el relleno de ruido antes de rellenar los intervalos espectrales o la banda alta con datos espectrales copiados o especulares del núcleo.

[0009] El documento US 2014/0188464 A1 describe un aparato para la generación de una señal de ancho de banda ampliado que incluye una unidad de procesamiento anti-escasez para llevar a cabo procesamiento anti-
10 escasez en un espectro de baja frecuencia. El aparato comprende además una unidad de decodificación de extensión de alta frecuencia de dominio de frecuencia para llevar a cabo codificación de extensión de alta frecuencia en el dominio de frecuencia en el espectro de baja frecuencia en el que se lleva a cabo el procesamiento anti-escasez.

[0010] Es un objeto de la presente invención proporcionar un concepto mejorado de generación de una señal mejorada.

[0011] Este objeto se logra por medio de un aparato para la generación de una señal mejorada de la reivindicación 1, un procedimiento para generar una señal mejorada de la reivindicación 13, un sistema de
20 codificación y decodificación de la reivindicación 14, un procedimiento de codificación y decodificación de la reivindicación 15 o un programa informático de la reivindicación 16.

[0012] La presente invención se basa en el hallazgo de que una mejora significativa de la calidad de audio de una señal mejorada generada por extensión del ancho de banda o relleno de intervalo inteligente o cualquier otra vía
25 de generación de una señal mejorada que tiene valores espectrales para una región espectral de mejora no contenida en una señal de entrada se obtiene generando primeros valores de ruido para una región de relleno de ruido en una región espectral de la fuente de la señal de entrada y generando a continuación segundos valores de ruido independientes para una región de ruidos en el destino o la región de destino, es decir, en la región de mejora que ahora tiene valores de ruido, es decir, los segundos valores de ruido que son independientes de los primeros
30 valores de ruido.

[0013] De este modo, el problema de la técnica anterior que tiene ruido dependiente en la banda de base y la banda de mejora debido al mapeo de los valores espectrales se elimina y los problemas relacionados con artefactos
35 tales como aspereza, distorsiones de la modulación y un timbre percibido como desagradable se eliminan en particular a bajas tasas de bits.

[0014] En otras palabras, el relleno de ruido de segundos valores de ruido decorrelacionados de los primeros valores de ruido, es decir, valores de ruido que son al menos parcialmente independientes de los primeros valores de ruido asegura que los artefactos no se produzcan o al menos se reduzcan con respecto a la técnica anterior. Así,
40 el procesamiento de la técnica anterior de los valores espectrales de relleno de ruido en la banda de base por una extensión del ancho de banda directo u operación de relleno de intervalo inteligente no decorrelaciona el ruido de la banda de base, sino que solo cambia el nivel, por ejemplo. Sin embargo, introduciendo valores de ruido decorrelacionados en la banda de fuente, por un lado, y en la banda de destino por el otro lado, preferentemente derivados de un proceso de ruidos separado, proporciona los mejores resultados. Sin embargo, incluso la
45 introducción de valores de ruido no completamente decorrelacionados o no completamente independientes, pero al menos parcialmente decorrelacionados como por un valor de decorrelación de 0,5 o menos cuando el valor de decorrelación de cero indica completamente decorrelacionado, mejora el problema de correlación completo de la técnica anterior.

[0015] De este modo, las formas de realización se refieren a una combinación de codificación en forma de ondas, extensión del ancho de banda o relleno del intervalo y relleno de ruidos en un decodificador perceptual.

[0016] Otras ventajas son que, contrariamente a los conceptos ya existentes, se evita la aparición de distorsiones de señales y artefactos de aspereza perceptual, que corrientemente son típicos para calcular
55 extensiones de ancho de banda o relleno de intervalos posteriores a la decodificación en forma de ondas y relleno de ruido.

[0017] Esto se debe, en algunas formas de realización, a un cambio en el orden de las etapas de procesamiento mencionadas. Se prefiere llevar a cabo la extensión del ancho de banda o el relleno del intervalo

directamente después de la decodificación en forma de ondas y también se prefiere computar el relleno de ruido posteriormente en la señal ya reconstruida mediante el uso de ruido no correlacionado.

[0018] En otras formas de realización, la decodificación en forma de ondas y relleno de ruido se puede llevar a cabo en un orden tradicional y también corriente abajo en el procesamiento, los valores de ruido se pueden reemplazar por ruido no correlacionado apropiadamente escalado.

[0019] Así, la presente invención se refiere a problemas que se producen debido a una operación de copia o una operación especular en espectros rellenos de ruido por desplazamiento de la etapa de relleno de ruido a un extremo de una cadena de procesamiento y mediante el uso de ruido no correlacionado para la aplicación de parches o relleno de intervalos.

[0020] Posteriormente, se tratan las formas de realización preferidas de la presente invención con respecto a los dibujos adjuntos, en los que:

Fig. 1a ilustra un aparato para codificar una señal de audio;
 Fig. 1b ilustra un decodificador para decodificar una señal de audio codificada que concuerda con el codificador de la Fig. 1a;
 Fig. 2a ilustra una implementación preferida del decodificador;
 Fig. 2b ilustra una implementación preferida del codificador;
 Fig. 3a ilustra una representación esquemática de un espectro generado por el decodificador de dominio espectral de la Fig. 1b;
 Fig. 3b ilustra una tabla que indica la relación entre factores de escala para bandas de factores de escala y energías para las bandas de reconstrucción e información de relleno de ruido para una banda de relleno de ruido;
 Fig. 4a ilustra la funcionalidad del codificador del dominio espectral para aplicar la selección de porciones espectrales en el primer y el segundo conjunto de porciones espectrales;
 Fig. 4b ilustra una implementación de la funcionalidad de la Fig. 4a;
 Fig. 5a ilustra una funcionalidad de un codificador MDCT;
 Fig. 5b ilustra una funcionalidad del decodificador con una tecnología MDCT;
 Fig. 5c ilustra una implementación del regenerador de frecuencia;
 Fig. 6 ilustra un diagrama de bloques de un aparato para generar una señal mejorada según la presente invención;
 Fig. 7 ilustra un flujo de señales de relleno de ruido independiente dirigido por una información de selección en un decodificador según una forma de realización de la presente invención;
 Fig. 8 ilustra un flujo de señales de un relleno de ruido independiente implementado por medio de un orden intercambiado de relleno de intervalo o extensión de ancho de banda y relleno de ruido en un decodificador;
 Fig. 9 ilustra un diagrama de flujo de un procedimiento según otra forma de realización de la presente invención;
 Fig. 10 ilustra un diagrama de flujo de un procedimiento según otra forma de realización de la presente invención;
 Fig. 11 ilustra un diagrama de flujo para explicar una escalada de valores aleatorios;
 Fig. 12 ilustra un diagrama de flujo que ilustra una inclusión de la presente invención en una extensión general de ancho de banda o un procedimiento de relleno de intervalo;
 Fig. 13a ilustra un codificador con un cálculo de los parámetros de extensión de ancho de banda; y
 Fig. 13b ilustra un decodificador con una extensión de ancho de banda implementado como un post-procesador más que un procedimiento integrado como en la Fig. 1a o 1b.

[0021] La Fig. 6 ilustra un aparato para la generación de una señal mejorada como una señal de audio a partir de una señal de entrada que también puede ser una señal de audio. La señal mejorada tiene valores espectrales para una región espectral de mejora, en donde los valores espectrales para la región espectral de mejora no están contenidos en la señal de entrada original en una señal de entrada 600. El aparato comprende un mapeador 602 para mapear una región espectral de la fuente de la señal de entrada en una región de destino en la región espectral de mejora, en donde la región espectral de la fuente comprende una región de relleno de ruido.

[0022] Por otra parte, el aparato comprende un relleno de ruido 604 configurado para generar primeros valores de ruido para la región de relleno de ruido en la región espectral de la fuente de la señal de entrada y para generar segundos valores de ruido para una región de ruidos en la región de destino, en donde los segundos valores de ruido, es decir, los valores de ruido en la región de destino son independientes o decorrelacionados de los primeros valores de ruido en la región de relleno de ruido.

[0023] Una forma de realización se refiere a una situación en la que el relleno del ruido se realiza realmente en la banda de base, es decir, en donde los valores de ruido en la región de fuente fueron generados por el relleno

del ruido. En otra alternativa, se asume que un relleno de ruido en la región de fuente no fue realizado. Sin embargo, la región de fuente tiene una región de ruidos realmente rellena con ruido como valores espectrales ilustrativamente codificados como valores espectrales por la fuente o el codificador central. El mapeo de este ruido como región de fuente a la región de mejora también genera ruido dependiente en la fuente y las regiones de destino. Con el fin de abordar este tema, el relleno de ruido solo rellena ruido en la región de destino del mapeador, es decir, genera segundos valores de ruido para la región de ruido en la región de destino, en donde los segundos valores de ruido se decorrelacionan de primeros valores de ruido en la región de fuente. Este reemplazo o relleno de ruido también puede tener lugar ya sea en un búfer mosaico de fuente o puede tener lugar en el destino propiamente dicho. La región de ruido puede ser identificada por el clasificador ya sea analizando la región de fuente o analizando la región de destino.

[0024] Para este fin, se hace referencia a la Fig. 3A. La Fig. 3A ilustra como región de relleno como una banda de factor de escala 301 en la señal de entrada y el relleno de ruido genera los primeros valores espectrales de ruido en esta banda de relleno de ruido 301 en una operación de decodificación de la señal de entrada.

[0025] Por otra parte, esta banda de relleno de ruido 301 se mapea a una región de destino, es decir, según la técnica anterior, los valores de ruido generados se mapean a la región de destino y, en consecuencia, la región de destino tendrá un ruido dependiente o correlacionado con la región de fuente.

[0026] Según la presente invención, sin embargo, el relleno de ruido 604 de la Fig. 6 genera segundos valores de ruido para una región de ruidos en el destino o la región de destino, donde los segundos valores de ruido se decorrelacionan o no correlacionan o independiente de los primeros valores de ruido en la banda de relleno de ruido 301 de la Fig. 3A.

[0027] En general, el relleno de ruido y el mapeador para mapear la región espectral de la fuente a una región de destino pueden estar incluidos dentro de un regenerador de alta frecuencia como se ilustra en el contexto de las Figs. 1A a 5C ilustrativamente dentro de un relleno de intervalo integrado o pueden ser implementados como un post-procesador como se ilustra en la Fig. 13B y el correspondiente codificador en la Fig. 13A.

[0028] En general, una señal de entrada se somete a una cuantificación inversa 700 o cualquier otro procesamiento de decodificador predefinido adicional 700 que significa que, en la salida del bloque 700, se obtiene la señal de entrada de la Fig. 6, de modo que la entrada en el bloque de relleno de ruido del codificador central o el bloque del relleno de ruido 704 es la entrada 600 de la Fig. 6. El mapeador en la Fig. 6 corresponde al bloque de relleno de intervalo o extensión de ancho de banda 602 y el bloque de relleno de ruido independiente 702 también está incluido dentro del relleno de ruido 604 de la Fig. 6. Así, los bloques 704 y 702 se incluyen ambos en el bloque de relleno de ruido 604 de la Fig. 6 y el bloque 704 genera los así llamados primeros valores de ruido para una región de ruidos en la región de relleno de ruido y el bloque 702 genera los segundos valores de ruido para una región de ruidos en el destino o región de destino, que se deriva de la región de relleno de ruido en la banda de base por extensión de ancho de banda realizada por el mapeador o bloque de relleno de intervalo o extensión de ancho de banda 602. Por otra parte, como se trata más abajo, la operación de relleno de ruido independiente realizada por el bloque 702 está controlada por un vector de control PHI ilustrado por una línea de control 706.

1. Etapa: Identificación del ruido

[0029] En una primera etapa, se identifican todas las líneas espectrales que representan ruido en una trama de audio transmitida. El proceso de identificación puede ser controlado por conocimiento de posiciones de ruido transmitido ya existente usado por el relleno de ruido [4] [5] o se puede identificar con un clasificador adicional. El resultado de la identificación de la línea de ruido es un vector que contiene ceros y unos, donde una posición con un uno indica una línea espectral que representa ruido.

[0030] En términos matemáticos, este procedimiento se puede describir como:

[0031] $\hat{X} \in \mathbb{C}^N$ es un espectro transmitido y recuantificado después del relleno de ruido [4] [5] de una señal en ventana codificada por transformada de la longitud $N \in \mathbb{N}$, $m \in \mathbb{N}$, $0 \leq m \leq N$, es la línea de detención de todo el proceso de decodificación.

[0032] El clasificador C_0 determina líneas espectrales donde se usa el relleno de ruido [4] [5] en la región central:

$$C_0: \mathbb{C}^N \rightarrow \{0,1\}^m,$$

$$\varphi[i] := C_0(\hat{X})[i] := \begin{cases} 1 & \text{en } \hat{X}[i] \text{ se usó relleno de ruido,} \\ 0 & \text{entonces} \end{cases}, 0 \leq i < m \leq N,$$

y el resultado $\varphi \in \{0,1\}^m$ es un vector de longitud m.

[0033] Un clasificador C_1 adicional puede identificar otras líneas en \hat{X} que representa ruido. Este clasificador se puede describir como:

$$C_1: \mathbb{C}^N \times \{0,1\}^m \rightarrow \{0,1\}^m,$$

$$\varphi[i] := C_1(\hat{X}, \varphi)[i] := \begin{cases} 1 & \text{if } \varphi[i] = 1 \vee \hat{X}[i] \text{ se clasifica como ruido,} \\ 0 & \text{entonces} \end{cases}, 0 \leq i < m \leq N.$$

[0034] Después del proceso de identificación, el vector de indicación de ruido $\varphi \in \{0,1\}^m$ se define como:

$$\varphi[i] = \begin{cases} 1, & \text{la línea espectral } \hat{X}[i] \text{ se identifica como una línea de ruido} \\ 0, & \text{la línea espectral } \hat{X}[i] \text{ no se identifica como una línea de ruido} \end{cases}, 0 \leq i < m \leq N.$$

20 2. Etapa: ruido independiente

[0035] En la segunda etapa, se selecciona una región específica del espectro transmitido y se copia a un mosaico de fuente. Dentro de este mosaico de fuente, el ruido identificado se reemplaza por ruido aleatorio. La energía del ruido aleatorio insertado se ajusta a la misma energía del ruido original en el mosaico de fuente.

[0036] En términos matemáticos, este procedimiento se puede describir como:

n , $n \leq m$, es la línea de inicio es para la copia del proceso, descrito en la etapa 3. $\hat{X}_{sT} \subset \hat{X}$ es una parte continua de un espectro transmitido \hat{X} , que representa un mosaico de fuente de longitud $v \leq n$, que contiene las líneas espectrales $l_k, l_{k+1}, \dots, l_{k+v-1}$ de \hat{X} , donde k es el índice de la primera línea espectral en el mosaico de fuente \hat{X}_{sT} , de modo que $\hat{X}_{sT}[i] = l_{k+i}$, $0 \leq i < v$. Por otra parte, $\varphi' \subset \varphi$, de modo que $\varphi'[i] = \varphi[k+i]$, $0 \leq i < v$.

[0037] El ruido identificado es reemplazado ahora por ruido sintético generado aleatoriamente. A fin de mantener la energía espectral en el mismo nivel, la energía E de ruido indicada por φ se calcula primero:

$$E := \sum_{i=0}^{v-1} \varphi'[i] |\hat{X}_{sT}[i]|^2.$$

[0038] Si $E = 0$ omite el reemplazo del ruido independiente para el mosaico de fuente \hat{X}_{sT} . entonces reemplazar el ruido indicado por φ' :

$$\hat{X}'_{sT}[i] := \begin{cases} r[i], & \varphi'[i] = 1 \\ \hat{X}_{sT}[i], & \varphi'[i] = 0 \end{cases}, \quad 0 \leq i < v,$$

donde $r[i] \in \mathbb{C}$ es un número aleatorio para todos $0 \leq i < v$.

5

[0039] A continuación, calcular la energía E' de los números aleatorios insertados:

$$E' := \sum_{i=0}^{v-1} \varphi[i] \left| \hat{X}'_{sT}[i] \right|^2.$$

10 Si $E' > 0$ calcula un factor g , fijar entonces $g = 0$:

$$g := \sqrt{\frac{E}{E'}}.$$

[0040] Con g , reescalar el ruido reemplazado:

15

$$\hat{X}'_{sT}[i] := \begin{cases} g \hat{X}'_{sT}[i], & \varphi'[i] = 1 \\ \hat{X}_{sT}[i], & \varphi'[i] = 0 \end{cases}, \quad 0 \leq i < v.$$

[0041] Después del reemplazo del ruido, el mosaico de fuente $\hat{X}'_{sT}[i]$ contiene líneas de ruido que son independientes de las líneas de ruido en \hat{X} .

20

3. Etapa: copiado

[0042] El mosaico de fuente $\hat{X}'_{sT}[i]$ se mapea a su región de destino en \hat{X} :

25

$$\hat{X}[c+i] = \hat{X}'_{sT}[i], \quad 0 \leq i < v, \quad c \geq n, \quad c+i < m < N,$$

o, si se usa el esquema IGF [8]:

$$\hat{X}[c+i] = \begin{cases} \hat{X}'_{sT}[i], & \hat{X}[c+i] = 0 \\ \hat{X}[c+i], & \hat{X}[c+i] < 0 \end{cases}, \quad 0 \leq i < v, \quad c \geq n, \quad c+i < m < N.$$

30

[0043] La Fig. 8 ilustra una forma de realización en la que, después de cualquier post-procesamiento como la decodificación del dominio espectral ilustrado en el bloque 112 en la Fig. 1B o en la forma de realización del post-procesador ilustrado por el bloque 1326 en la Fig. 13B, la señal de entrada se somete a un relleno de intervalo o extensión de ancho de banda primero, es decir, se somete a una operación de mapeo primero y, a continuación, se lleva a cabo un relleno de ruido independiente, es decir, dentro del espectro completo.

35

[0044] El proceso descrito en el contexto anterior de la Fig. 7 se puede llevar a cabo en una orientación in situ, de modo que el búfer intermedio \hat{X}'_{sT} no se necesita. En consecuencia, se adapta el orden de ejecución.

40

[0045] Ejecutar la primera etapa como se describe en el contexto de la Fig. 7, de nuevo el conjunto de líneas espectrales $k, k+1, \dots, k+v-1$ de \hat{X} son la región de fuente. Realizar:

2. Etapa: copiado

$$[0046] \quad \hat{X}[c+i] = \hat{X}[k+i], \quad 0 \leq i < v, \quad c \geq n, \quad 0 < k+i < n, \quad c+i < m < N,$$

o, si se usa el esquema IGF [8]:

5

$$\hat{X}[c+i] = \begin{cases} \hat{X}[k+i], & \hat{X}[c+i] = 0 \\ \hat{X}[c+i], & \hat{X}[c+i] \neq 0 \end{cases} \\ 0 \leq i < v, c \geq n, \quad 0 < k+i < n, \quad c+i < m < N.$$

3. Etapa: relleno de ruido independiente

10

[0047] Realizar el relleno de ruido restante hasta n y calcular la energía de líneas espectrales de ruido en la región de fuente $k, k+1, \dots, k+v-1$:

$$E := \sum_{i=0}^{v-1} \varphi[k+i] |\hat{X}[k+i]|^2.$$

15

[0048] Realizar el relleno de ruido independiente en el relleno de intervalo o la región espectral BWE:

$$\hat{X}[c+i] := \begin{cases} r[i], & \varphi[k+i] = 1 \\ \hat{X}[c+i], & \varphi[k+i] = 0 \end{cases}, \quad 0 \leq i < v,$$

20 donde $r[i], 0 \leq i < v$ es de nuevo un conjunto de números aleatorios.

[0049] Calcular la energía E' de los números aleatorios insertados:

$$E' := \sum_{i=0}^{v-1} \varphi[k+i] |r[i]|^2.$$

25

[0050] De nuevo, si $E' > 0$ calcular el factor g , también establecer $g = 0$:

$$g := \sqrt{\frac{E}{E'}}.$$

30 [0051]

Con g , reescalar el ruido de reemplazo:

$$\hat{X}[c+i] := \begin{cases} g \hat{X}[c+i], & \varphi[k+i] = 1 \\ \hat{X}[c+i], & \varphi[k+i] = 0 \end{cases}, \quad 0 \leq i < v.$$

[0052] El relleno de ruido independiente de la invención también se puede usar en un ambiente de pares de canales estéreo. En consecuencia, el codificador calcula la representación de pares de canales apropiados, L/R o M/S, por banda de frecuencia y coeficientes de predicción opcionales. El decodificador aplica relleno de ruido independiente como se ha descrito con anterioridad a la representación apropiadamente seleccionada de los canales antes de la posterior computación de la conversión final de todas las bandas de frecuencia en la representación L/R.

40

[0053] La invención es aplicable o apropiada para todas las aplicaciones de audio en las que el ancho de banda completo no está disponible o que usa relleno de intervalo para rellenar agujeros espectrales. La invención se puede usar en la distribución de difusión del contenido de audio como, por ejemplo, con radio digital, transmisión por

Internet y aplicaciones de comunicaciones por audio.

[0054] Posteriormente, se tratan formas de realización de la presente invención con respecto a las Figs. 9–12. En la etapa 900, las regiones de ruido se identifican en el rango de la fuente. Este procedimiento que se ha tratado anteriormente con respecto a la “Identificación del ruido” se puede basar en la información lateral del relleno de ruido recibida por completo del lado del codificador o también se puede configurar para basarse alternativa o adicionalmente en el análisis de señales de la señal de entrada ya generada, pero sin valores espectrales para la región espectral de mejora, es decir, sin los valores espectrales para esta región espectral de mejora.

10 **[0055]** A continuación, en la etapa 902, el rango de la fuente que ya se había sometido a un relleno de ruido directo como se conoce en la técnica, es decir, un rango de fuente completo se copia a un búfer mosaico de fuente.

[0056] A continuación, en la etapa 904, los primeros valores de ruido, es decir, los valores de ruido directos generados dentro de la región de relleno de ruido de la señal de entrada se reemplazan en el búfer mosaico de fuente por valores aleatorios. A continuación, en la etapa 906, estos valores aleatorios se escalan en el búfer mosaico de fuente para obtener los segundos valores de ruido para la región de destino. A continuación, en la etapa 908, se realiza la operación de mapeo, es decir, su contenido del búfer mosaico de fuente disponible después de las etapas 904 y 906 se mapea en el rango de destino. De esta manera, por medio de la operación de reemplazo 904 y después de la operación de mapeo 908, se obtuvo la operación de relleno de ruido independiente en el rango de la fuente y en el rango de destino.

[0057] La Fig. 10 ilustra otra forma de realización de la presente invención. De nuevo, en la etapa 900, se identifica el ruido en el rango de la fuente. Sin embargo, la funcionalidad de esta etapa 900 es diferente de la funcionalidad de la etapa 900 en la Fig. 9, dado que la etapa 900 en la Fig. 9 puede operar en un espectro de señales de entrada que tiene valores de ruido ya recibidos, es decir, en donde ya se ha realizado la operación de relleno de ruido.

[0058] Sin embargo, en la Fig. 10, no se ha realizado ninguna operación de relleno de ruido en la señal de entrada y la señal de entrada aún no tiene valores de ruido en la región de relleno de ruido en la entrada en la etapa 902. En la etapa 902, el rango de la fuente se mapea en el destino o el rango de destino, donde los valores de relleno de ruido no se incluyen en el rango de la fuente.

[0059] De esta manera, la identificación del ruido en el rango de la fuente en la etapa 900 se puede llevar a cabo, con respecto a la región de relleno de ruido, por identificación de valores espectrales cero en la señal y/o mediante el uso de esta información lateral de relleno de ruido a partir de la señal de entrada, es decir, la información de relleno de ruido generada del lado del codificador. A continuación, en la etapa 904, la información de relleno de ruido y en particular la información de energía identifica que se lee la energía que se va a introducir en la señal de entrada del lado del decodificador.

40 **[0060]** A continuación, como se ilustra en la etapa 1006, se lleva a cabo un relleno de ruido en el rango de la fuente y, posterior o concurrentemente, se lleva a cabo una etapa 1008, es decir, se insertan valores aleatorios en posiciones en el rango de destino que fueron identificados por la etapa 900 en toda la banda o que fueron identificados mediante el uso de la banda de base o la información de señales de entrada junto con la información de mapeo, es decir, qué rango de fuente (de una pluralidad de fuentes) se mapea con qué rango de destino (de una pluralidad de rangos de destino).

[0061] Finalmente, los valores aleatorios insertados se escalan para obtener los valores de ruido independientes o no correlacionados o decorrelacionados.

50 **[0062]** Posteriormente, la Fig. 11 se trata a fin de ilustrar mayor información acerca de la escala de los valores de relleno de ruido en la región espectral de mejora, es decir, de cómo se obtienen, a partir de los valores aleatorios, los segundos valores de ruido.

[0063] En la etapa 1100, se obtiene una información de energía acerca del ruido en el rango de la fuente. A continuación, se determina una información de energía a partir de los valores aleatorios, es decir, a partir de los valores generados por un proceso aleatorio o pseudoaleatorio, tal como se ilustra en la etapa 1102. Por otra parte, la etapa 1104 ilustra la forma de calcular el factor de escala, es decir, mediante el uso de la información de energía acerca del ruido en el rango de la fuente y mediante el uso de la información de energía acerca de los valores aleatorios. A continuación, en la etapa 1106, los valores aleatorios, es decir, de los que se calculó la energía en la

etapa 1102, se multiplican por el factor de escala generado por la etapa 1104. Así, el procedimiento ilustrado en la Fig. 11 corresponde al cálculo del factor de escala g ilustrado antes en una forma de realización. Sin embargo, todos estos cálculos también se pueden llevar a cabo en un dominio logarítmico o en cualquier otro dominio y la etapa de multiplicación 1106 se puede reemplazar por una adición o sustracción en el rango logarítmico.

5

[0064] Se hace referencia además a la Fig. 12 con el fin de ilustrar la incorporación de la presente invención dentro de un esquema de relleno de intervalo general inteligente o extensión de ancho de banda. En la etapa 1200, se recupera información envolvente espectral de la señal de entrada. La información envolvente espectral puede ser generada, por ejemplo, por un extractor de parámetros 1306 de la Fig. 13A y se puede proporcionar por un decodificador de parámetros 1324 de la Fig. 13B. A continuación, los segundos valores de ruido y los otros valores en el rango de destino se escalan mediante el uso de esta información envolvente espectral como se ilustra en 1202. Posteriormente, cualquier otro post-procesamiento 1204 se puede llevar a cabo para obtener la señal mejorada del dominio de tiempo final que tiene un mayor ancho de banda en el caso de una extensión de ancho de banda o que tiene una cantidad reducida o ningún agujero espectral en el contexto de relleno de intervalo inteligente.

15

[0065] En este contexto, se destaca que, en particular para la forma de realización de la Fig. 9, se pueden aplicar varias alternativas. Para una forma de realización, la etapa 902 se lleva a cabo con el espectro completo de la señal de entrada o al menos con la porción del espectro de la señal de entrada que está por encima de la frecuencia de borde del relleno de ruido. Esta frecuencia asegura que, por debajo de cierta frecuencia, es decir, por debajo de esta frecuencia, no se realiza en absoluto un relleno de ruido.

20

[0066] A continuación, independientemente de cualquier información de rango de fuente/mapeo de rango de destino, todo el espectro de señales de entrada, es decir, el rango de fuente potencial completo se copia al búfer mosaico de fuente 902 y, a continuación, se procesa con la etapa 904 y 906 y la etapa 908, a continuación se selecciona cierta región de fuente específicamente requerida de este búfer mosaico de fuente.

25

[0067] En otras formas de realización, sin embargo, solo los rangos de fuente específicamente requeridos que solo pueden ser partes de la señal de entrada se copian al búfer mosaico de fuente individual o a varios búfers mosaicos de fuentes individuales en base a la información de rango de fuente/rango de destino incluida en la señal de entrada, es decir, asociada como información lateral con esta señal de entrada de audio. Según esta situación, la segunda alternativa, donde solo los rangos de fuente específicamente requeridos son procesados por las etapas 902, 904, 906, se puede reducir la complejidad o al menos los requisitos de memoria en comparación con la situación en la que siempre, independientemente de la situación de mapeo específica, se procesa todo el rango de fuente al menos por encima de la frecuencia de borde del relleno de ruido por medio de las etapas 902, 904, 906.

35

[0068] Posteriormente, se hace referencia a las Figs. 1a – 5e con el fin de ilustrar la implementación específica de la presente invención dentro de un regenerador de frecuencia 116, que se ubica delante del conversor de espectro–tiempo 118.

40

[0069] La Fig. 1a ilustra un aparato para codificar una señal de audio 99. La señal de audio 99 se introduce en un conversor de tiempo–espectro 100 para convertir una señal de audio que tiene una tasa de muestreo en una representación espectral 101 producida por el conversor de tiempo–espectro. El espectro 101 se introduce en un analizador espectral 102 para analizar la representación espectral 101. El analizador espectral 101 se configura para determinar un primer conjunto de primeras porciones espectrales 103 para codificar con una primera resolución espectral y un segundo conjunto diferente de segundas porciones espectrales 105 para codificar con una segunda resolución espectral. La segunda resolución espectral es menor que la primera resolución espectral. El segundo conjunto de segundas porciones espectrales 105 se introduce en una calculadora de parámetros o codificador paramétrico 104 para calcular información envolvente espectral que tiene la segunda resolución espectral. Por otra parte, se proporciona un codificador de audio del dominio espectral 106 para generar una primera representación codificada 107 del primer conjunto de primeras porciones espectrales que tiene la primera resolución espectral. Por otra parte, la calculadora de parámetros/codificador paramétrico 104 se configura para generar una segunda representación codificada 109 del segundo conjunto de segundas porciones espectrales. La primera representación codificada 107 y la segunda representación codificada 109 se introducen en un multiplexador de corriente de bits o formador de corriente de bits 108 y el bloque 108 finalmente produce la señal de audio codificada para transmisión o almacenamiento en un dispositivo de almacenamiento.

55

[0070] Típicamente, una primera porción espectral como 306 de la Fig. 3a se rodeará por dos segundas porciones espectrales como 307a, 307b. Este no es el caso en HE AAC, donde el rango de frecuencia del codificador central tiene banda limitada.

[0071] La Fig. 1b ilustra un decodificador que concuerda con el codificador de la Fig. 1a. La primera representación codificada 107 se introduce en un decodificador de audio del dominio espectral 112 para generar una primera representación decodificada de un primer conjunto de primeras porciones espectrales, donde la representación decodificada tiene una primera resolución espectral. Por otra parte, la segunda representación codificada 109 se introduce en un decodificador paramétrico 114 para generar una segunda representación decodificada de un segundo conjunto de segundas porciones espectrales que tiene una segunda resolución espectral que es inferior a la primera resolución espectral.

10 **[0072]** El decodificador comprende además un regenerador de frecuencia 116 para regenerar una segunda porción espectral reconstruida que tiene la primera resolución espectral mediante el uso de una primera porción espectral. El regenerador de frecuencia 116 realiza una operación de relleno de mosaicos, es decir, usa un mosaico o una porción del primer conjunto de primeras porciones espectrales y copia este primer conjunto de primeras porciones espectrales en el rango de reconstrucción o banda de reconstrucción que tiene la segunda porción espectral y típicamente realiza formación de envolvente espectral u otra operación como se indica por la segunda representación decodificada producida por el decodificador paramétrico 114, es decir, mediante el uso de la información acerca del segundo conjunto de segundas porciones espectrales. El primer conjunto decodificado de primeras porciones espectrales y el segundo conjunto reconstruido de porciones espectrales como se indica en la salida del regenerador de frecuencia 116 en la línea 117 se introduce en un conversor de espectro-tiempo 118 configurado para convertir la primera representación decodificada y la segunda porción espectral reconstruida en una representación temporal 119, donde la representación temporal tiene cierta tasa alta de muestreo.

[0073] La Fig. 2b ilustra una implementación del codificador de la Fig. 1a. Una señal de entrada de audio 99 se introduce en un banco de filtro de análisis 220 correspondiente al conversor de tiempo-espectro 100 de la Fig. 1a. A continuación, una operación de formación de ruido temporal se lleva a cabo en el bloque TNS 222. En consecuencia, la entrada en el analizador espectral 102 de la Fig. 1a correspondiente a la máscara tonal de bloque 226 de la Fig. 2b pueden ser valores espectrales completos, cuando no se aplica la operación de formación de ruido temporal / formación de mosaicos temporales o pueden ser valores residuales espectrales, cuando se aplica la operación de TNS como se ilustra en la Fig. 2b, bloque 222. Para señales de dos canales o señales de múltiples canales, se puede llevar a cabo adicionalmente una codificación de canal conjunto 228, de modo que el codificador del dominio espectral 106 de la Fig. 1a pueda comprender el bloque codificador del canal conjunto 228. Por otra parte, se proporciona un codificador de entropía 232 para realizar una compresión de datos sin pérdidas que también es una porción del codificador del dominio espectral 106 de la Fig. 1a.

35 **[0074]** El analizador espectral / máscara tonal 226 separa la salida de TNS bloque 222 en la banda central y los componentes tonales correspondientes al primer conjunto de primeras porciones espectrales 103 y los componentes residuales correspondientes al segundo conjunto de segundas porciones espectrales 105 de la Fig. 1a. El bloque 224 indicado como codificación de extracción de parámetros IGF corresponde al codificador paramétrico 104 de la Fig. 1a y el multiplexador de la corriente de bits 230 corresponde al multiplexador de la corriente de bits 108 de la Fig. 1a.

[0075] Con preferencia, el banco de filtros de análisis 222 se implementa como MDCT (banco de filtro de transformada de coseno discreta modificada) y la MDCT se usa para transformar la señal 99 en un dominio de tiempo-frecuencia con la transformada de coseno discreta modificada que actúa como la herramienta de análisis de frecuencia.

[0076] El analizador espectral 226 aplica preferentemente una máscara de tonalidad. Esta etapa de estimación de máscara de tonalidad se usa para separar los componentes tonales de los componentes de tipo ruido en la señal. Esto permite que el codificador central 228 codifique todos los componentes tonales con un módulo psicoacústico. La etapa de estimación de la máscara de tonalidad se puede implementar en numerosas vías diferentes y preferentemente se implementa de modo similar en su funcionalidad a la etapa de estimación de rastreo sinusoidal usada en una modelación sinusoidal y de ruido para codificación de habla/audio [8, 9] o un codificador de audio basado en un modelo de HILN descrito en [10]. Con preferencia, se usa una implementación que es fácil de implementar sin necesidad de mantener las trayectorias de nacimiento-muerte, pero también se puede usar cualquier otra tonalidad o detector de ruido.

[0077] El módulo IGF calcula la similitud que existe entre una región de fuente y una región de destino. La región de destino estará representada por el espectro de la región de fuente. La medición de la similitud entre la región de fuente y la región de destino se realiza mediante el uso de una estrategia de correlación cruzada. La

región de destino se divide en mosaicos de frecuencia $nTar$ que no se superponen. Para cada mosaico en la región de destino, se crean mosaicos de fuente $nSrc$ a partir de una frecuencia de inicio fija. Estos mosaicos de fuente se superponen con un factor de entre 0 y 1, donde 0 significa el 0% de superposición y 1 significa el 100% de superposición. Cada uno de estos mosaicos de fuente se correlaciona con el mosaico de destino en diferentes retardos para hallar el mosaico de fuente que concuerde mejor con el mosaico de destino. El número de mosaicos de mejor concordancia se almacena en $tileNum[idx_{tar}]$, en cuyo retardo se correlaciona mejor con el blanco, se almacena en $xcorr_{tag}[idx_{tar}][idx_{src}]$ y el signo de la correlación se almacena en $xcorr_{sign}[idx_{tar}][idx_{src}]$. En el caso de la correlación sea altamente negativa, el mosaico de fuente necesita ser multiplicado por -1 antes del proceso de relleno de mosaicos en el decodificador. El módulo IGF también cuida de no sobrescribir los componentes tonales en el espectro, ya que los componentes tonales se preservan mediante el uso de la máscara de tonalidad. Se usa un parámetro de tipo banda para almacenar la energía de la región de destino que nos permite reconstruir el espectro de forma precisa.

[0078] Este procedimiento tiene ciertas ventajas sobre el clásico SBR [1] porque la rejilla de armónicos de una señal multitonal es preservada por el codificador central, mientras que solo los intervalos entre los sinusoides se rellenan con el “ruido formado” de mejor concordancia de la región de fuente. Otra ventaja de este sistema comparado con ASR (Accurate Spectral Replacement, reemplazo espectral preciso) [2–4] es la ausencia de una etapa de síntesis de señales que crea las porciones importantes de la señal en el decodificador. En vez de ello, esta tarea es realizada por el codificador central, que permite la preservación de importantes componentes del espectro. Otra ventaja del sistema propuesto es la continua escalabilidad que ofrecen las características. Justamente el uso de $tileNum[idx_{tar}]$ y $xcorr_{tag} = 0$ para cada mosaico se denomina concordancia de granularidad gruesa y se puede usar para bajas tasas de bits mientras que el uso de $xcorr_{tag}$ variable para cada mosaico nos permite hacer concordar mejor los espectros de destino y de fuente.

[0079] Además, se propone una técnica de estabilización de selección de mosaicos que elimina los artefactos de dominio de frecuencia como trinos y ruido musical.

[0080] En el caso de pares de canales estéreo, se aplica un procesamiento conjunto adicional. Esto es necesario porque para determinado rango de destino, la señal puede ser una fuente de sonido rotada altamente correlacionada. En el caso de que las regiones de fuente elegidas para esta región particular no se correlacionen bien, a pesar de que las energías concuerdan para las regiones de destino, la imagen espacial puede sufrir debido a las regiones de fuente no correlacionadas. El codificador analiza cada banda de energía de la región de destino, típicamente realizando una correlación cruzada de los valores espectrales y si se excede cierto umbral, fija un indicador conjunto para esta banda de energía. En el decodificador, las bandas de energía del canal izquierdo y derecho se tratan de modo individual si no se fija este indicador estéreo conjunto. En el caso de fijar el indicador estéreo conjunto, tanto las energías como la conexión provisional se realizan en el dominio estéreo conjunto. La información estéreo conjunta para las regiones IGF se señaliza de modo similar a la información estéreo conjunta para la codificación central, incluyendo un indicador que indica en el caso de una predicción si la dirección de la predicción es de mezcla descendente a residual, o viceversa.

[0081] Las energías se pueden calcular a partir de las energías transmitidas en el dominio L/R.

$$\begin{aligned} midNrg[k] &= leftNrg[k] + rightNrg[k]; \\ sideNrg[k] &= leftNrg[k] - rightNrg[k]; \end{aligned}$$

siendo k el índice de frecuencia en el dominio de la transformada.

[0082] Otra solución consiste en calcular y transmitir las energías directamente en el dominio estéreo conjunto para bandas en las que el estéreo conjunto está activo, de modo que no se necesite una transformación adicional de energía en el lado del decodificador.

[0083] Los mosaicos de fuente siempre se crean según la matriz media/lateral:

$$midTile[k] = 0.5 \cdot (leftTile[k] + rightTile[k])$$

$$sideTile[k] = 0,5 \cdot (leftTile[k] - rightTile[k])$$

Ajuste de energía:

5 [0084]

$$\begin{aligned} midTile[k] &= midTile[k] * midNrg[k]; \\ sideTile[k] &= sideTile[k] * sideNrg[k]; \end{aligned}$$

10 Estéreo conjunto → transformación LR:

[0085] Si no se codifica ningún parámetro de predicción adicional:

$$\begin{aligned} leftTile[k] &= midTile[k] + sideTile[k] \\ rightTile[k] &= midTile[k] - sideTile[k] \end{aligned}$$

[0086] Si se codifica un parámetro de predicción adicional y si la dirección señalada va del medio al lateral:

$$\begin{aligned} sideTile[k] &= sideTile[k] - predictionCoeff \cdot midTile[k] \\ leftTile[k] &= midTile[k] + sideTile[k] \\ rightTile[k] &= midTile[k] - sideTile[k] \end{aligned}$$

20

[0087] Si la dirección señalada va del lado al medio:

$$\begin{aligned} midTile[k] &= midTile[k] - predictionCoeff \cdot sideTile[k] \\ leftTile[k] &= midTile[k] - sideTile[k] \\ rightTile[k] &= midTile[k] + sideTile[k] \end{aligned}$$

25 [0088] Este procesamiento asegura que desde los mosaicos usados para regenerar regiones de destino altamente correlacionadas y regiones de destino rotadas, los canales izquierdo y derecho resultantes aún representan una fuente de sonido correlacionada y rotada incluso si las regiones de fuente no se correlacionan, preservando la imagen estéreo para tales regiones.

30 [0089] En otras palabras, en la corriente de bits, los indicadores estéreos conjuntos se transmiten, lo que indica si se debe usar L/R o M/S como un ejemplo para la codificación estéreo conjunta general. En el decodificador, primero se decodifica la señal central como es indicado por los indicadores estéreo conjuntos para las bandas centrales. En segundo lugar, la señal central se almacena en ambas representaciones L/R y M/S. Para el relleno de mosaico IGF, se elige la representación de mosaico de fuente para ajustar la representación de mosaico de destino como se indica por la información estéreo conjunta para las bandas de IGF.

[0090] Temporal Noise Shaping (TNS, configuración de ruido temporal) es una técnica estándar y parte de AAC. TNS puede ser considerada como una extensión del esquema básico de un codificador perceptual, insertando una etapa de procesamiento opcional entre el banco de filtros y la etapa de cuantificación. La tarea principal del módulo de TNS es esconder el ruido de cuantificación producido en la región de enmascaramiento temporal de señales de tipo transitorio y, así, lleva a un esquema de codificación más eficiente. En primer lugar, TNS calcula un conjunto de coeficientes de predicción mediante el uso de "predicción directa" en el dominio de la transformada, por ejemplo, MDCT. Estos coeficientes se usan a continuación para aplanar la envolvente temporal de la señal. Como la cuantificación afecta al espectro filtrado de TNS, también el ruido de la cuantificación es temporalmente plano. Al aplicar el filtrado de TNS inversa en el lado del decodificador, el ruido de cuantificación se conforma según la envolvente temporal del filtro de TNS y, por ello, el ruido de cuantificación se enmascara por el transitorio.

[0091] IGF se basa en una representación de MDCT. Para una codificación más eficaz, se deben usar preferentemente bloques largos de aproximadamente 20 ms. Si la señal dentro de tal bloque largo contiene

transitorios, se producen pre- y post-ecos audibles en las bandas espectrales de IGF debido al relleno de mosaicos.

[0092] Este efecto de pre-eco es reducido mediante el uso de TNS en el contexto de IGF. Aquí, se usa TNS como una herramienta de configuración de mosaicos temporales (TTS) cuando la regeneración espectral en el decodificador se lleva a cabo en la señal residual de TNS. Los coeficientes de predicción de TTS requeridos se calculan y se aplican mediante el uso del espectro total en el lado del codificador como es habitual. Las frecuencias de inicio y detención de TNS/TTS no son afectadas por la frecuencia de inicio de IGF $f_{IGFstart}$ de la herramienta de IGF. En comparación con TNS de herencia, la frecuencia de detención de TTS se aumenta hasta la frecuencia de detención de la herramienta de IGF, que es mayor que $f_{IGFstart}$. En el lado del decodificador, se aplican de nuevo los coeficientes de TNS/TTS en el espectro total, es decir, el espectro central más el espectro regenerado más los componentes tonales del mapa de tonalidad. La aplicación de TTS es necesaria para formar la envolvente temporal del espectro regenerado para hacer concordar la envolvente de la señal original. Así, se reducen los pre-ecos mostrados. Además, también conforma el ruido de cuantificación en la señal por debajo de $f_{IGFstart}$ como es habitual con TNS.

[0093] En decodificadores heredados, la conexión provisional espectral en una señal de audio corrompe la correlación espectral en los bordes del parche y, así, altera la envolvente temporal de la señal de audio al introducir una dispersión. Así, otro beneficio de realizar el relleno de mosaico IGF en la señal residual es que, después de la aplicación del filtro de conformación, los bordes del mosaico se correlacionan perfectamente, dando como resultado una reproducción temporal más fiel de la señal.

[0094] En un codificador de la invención, el espectro que se sometió a filtración TNS/TTS, procesamiento de máscara de tonalidad y estimación paramétrica IGF está desprovisto de cualquier señal por encima de la frecuencia de inicio IGF excepto por los componentes tonales. Este espectro disperso está codificado ahora por el codificador central mediante el uso de los principios de la codificación aritmética y la codificación predictiva. Estos componentes codificados junto con los bits de señalización forman la corriente de bits de audio.

[0095] La Fig. 2a ilustra la correspondiente implementación del decodificador. La corriente de bits en la Fig. 2a correspondiente a la señal de audio codificada se introduce en el demultiplexador/decodificador que se conectaría, con respecto a la Fig. 1B, con los bloques 112 y 114. El demultiplexador de la corriente de bits separa la señal de audio de entrada en la primera representación codificada 107 de la Fig. 1b y la segunda representación codificada 109 de la Fig. 1b. La primera representación codificada que tiene el primer conjunto de primeras porciones espectrales se introduce en el bloque de decodificación del canal conjunto 204 correspondiente al decodificador de dominio espectral 112 de la Fig. 1b. La segunda representación codificada se introduce en el decodificador paramétrico 114 no ilustrado en la Fig. 2a y, a continuación, se introduce en el bloque de IGF 202 correspondiente al regenerador de frecuencia 116 de la Fig. 1b. El primer conjunto de primeras porciones espectrales requerido para la regeneración de frecuencia se introduce en el bloque IGF 202 a través de la línea 203. Por otra parte, después de la decodificación del canal conjunto 204, se aplica la decodificación central específica en el bloque de máscara tonal 206 de modo que la salida de la máscara tonal 206 corresponde a la salida del decodificador de dominio espectral 112. A continuación, se lleva a cabo una combinación por medio del combinador 208, es decir, una formación de trama, donde la salida del combinador 208 tiene ahora el espectro de rango completo, pero aún en el dominio filtrado TNS/TTS. A continuación, en el bloque 210, se lleva a cabo una operación de TNS/TTS inversa mediante el uso de información de filtro TNS/TTS proporcionada por medio de la línea 109, es decir, la información lateral de TTS se incluye preferentemente en la primera representación codificada generada por el codificador del dominio espectral 106 que puede ser, por ejemplo, un codificador central AAC o USAC directo o también se puede incluir en la segunda representación codificada. En la salida del bloque 210, se proporciona un espectro completo hasta la frecuencia máxima, que es la frecuencia de rango completo definida por la tasa de muestreo de la señal de entrada. A continuación, se lleva a cabo una conversión de espectro/tiempo en el banco de filtros de síntesis 212 para obtener finalmente la señal de salida de audio.

[0096] La Fig. 3a ilustra una representación esquemática del espectro. El espectro se subdivide en bandas de factor de escala SCB donde hay siete bandas de factor de escala SCB1 a SCB7 en el ejemplo ilustrado de la Fig. 3a. Las bandas de factor de escala pueden ser bandas de factor de escala AAC que se definen en AAC estándar y tienen un mayor ancho de banda que frecuencias superiores como se ilustra en la Fig. 3a esquemáticamente. Se prefiere llevar a cabo un relleno de intervalo inteligente no desde el principio del espectro, es decir, a bajas frecuencias, sino al comienzo de la operación de IGF a una frecuencia inicial de IGF ilustrada en 309. En consecuencia, la banda de frecuencia central se extiende desde la frecuencia más baja hasta la frecuencia inicial de IGF. Por encima de la frecuencia inicial de IGF, el análisis del espectro se aplica para separar los componentes

espectrales de alta resolución 304, 305, 306, 307 (el primer conjunto de primeras porciones espectrales) de los componentes de baja resolución representados por el segundo conjunto de segundas porciones espectrales. La Fig. 3a ilustra un espectro que se incorpora ilustrativamente en el codificador del dominio espectral 106 o el codificador de canal conjunto 228, es decir, el codificador central opera en el rango completo, pero codifica una cantidad significativa de valores espectrales cero, es decir, estos valores espectrales cero se cuantifican hasta cero o se fijan en cero antes de cuantificar o después de cuantificar. De todos modos, el codificador central opera en todo el rango, es decir, como si el espectro se ilustrara, es decir, el decodificador central no tiene que ser necesariamente consciente de cualquier relleno de intervalo inteligente o la codificación del segundo conjunto de segundas porciones espectrales con una menor resolución espectral.

[0097] Con preferencia, la elevada resolución es definida por una codificación de tipo lineal de líneas espectrales tales como líneas MDCT, mientras que la segunda resolución o baja resolución es definida, por ejemplo, calculando solo un único valor espectral por banda de factor de escala, donde una banda de factor de escala cubre varias líneas de frecuencia. De esta manera, la segunda baja resolución, con respecto a su resolución espectral, es mucho menor que la primera o alta resolución definida por la codificación de tipo lineal típicamente aplicada por el codificador central como un codificador central AAC o USAC.

[0098] Con respecto al cálculo del factor de escala o la energía, la situación se ilustra en la Fig. 3b. Debido al hecho de que el codificador es un codificador central y debido al hecho de que puede haber, aunque no necesariamente, componentes del primer conjunto de porciones espectrales en cada banda, el codificador central calcula un factor de escala para cada banda no solo en el rango central por debajo de la frecuencia inicial de IGF 309, sino también por encima de la frecuencia inicial de IGF hasta la frecuencia máxima ~~fig-top~~ que es menor o igual a la mitad de la frecuencia de muestreo, es decir, $f_{s/2}$. De esta manera, las porciones tonales codificadas 302, 304, 305, 306, 307 de la Fig. 3a y, en esta forma de realización junto con los factores de escala SCB1 a SCB7 corresponden a los datos espectrales de alta resolución. Los datos espectrales de baja resolución se calculan a partir de la frecuencia inicial de IGF y corresponden a la información de los valores de energía E_1 , E_2 , E_3 , E_4 , que se transmiten junto con los factores de escala SF4 a SF7.

[0099] En particular, cuando el codificador central está bajo una condición de baja tasa de bits, se puede aplicar además una operación adicional de relleno de ruido en la banda central, es decir, inferior en frecuencia que la frecuencia inicial de IGF, es decir, en bandas de factor de escala SCB1 a SCB3. En el relleno de ruido, existen varias líneas espectrales adyacentes que fueron cuantificadas a cero. Del lado del decodificador, estos valores espectrales cuantificados a cero se resintetizan y los valores espectrales resintetizados se ajustan en su magnitud mediante el uso de una energía de relleno de ruido como NF_2 ilustrado en 308 en la Fig. 3b. La energía de relleno de ruido, que se puede dar en términos absolutos o en términos relativos particularmente con respecto al factor de escala como en USAC corresponde a la energía del conjunto de valores espectrales cuantificados a cero. Estas líneas espectrales de relleno de ruido también se pueden considerar que son un tercer conjunto de terceras porciones espectrales que son regeneradas por síntesis directa de relleno de ruido sin ninguna operación de IGF que se basa en la regeneración de frecuencia mediante el uso de mosaicos de frecuencia de otras frecuencias para reconstruir mosaicos mediante el uso de valores espectrales de un rango de fuente y la información de energía E_1 , E_2 , E_3 , E_4 .

[0100] Con preferencia, las bandas para las que se calcula la información de energía coinciden con las bandas de factor de escala. En otras formas de realización, se aplica un valor de información de energía agrupados de modo que, por ejemplo, para bandas de factores de escala 4 y 5, solo se transmite un único valor de información de energía, pero incluso en esta forma de realización, los bordes de las bandas de reconstrucción agrupadas coinciden con los bordes de las bandas del factor de escala. Si se aplican diferentes separaciones de bandas, entonces se pueden aplicar determinados recálculos o cálculos de sincronización y esto puede tener sentido dependiendo de cierta implementación.

[0101] Con preferencia, el codificador del dominio espectral 106 de la Fig. 1a es un codificador manejado psico-acústicamente ilustrado en la Fig. 4a. Típicamente, como se ilustra, por ejemplo, en MPEG2/4 AAC estándar o MPEG1/2, capa 3 estándar, la señal de audio por codificar después de ser transformada en el rango espectral (401 en la Fig. 4a) se dirige a un calculador de factores de escala 400. El calculador de factores de escala es controlado por un modelo psico-acústico que adicionalmente recibe la señal de audio por cuantificar o que recibe como en MPEG1/2 capa 3 o MPEG AAC estándar, una representación espectral compleja de la señal de audio. El modelo psico-acústico calcula, para cada banda de factor de escala, un factor de escala que representa el umbral psico-acústico. Adicionalmente, los factores de escala se ajustan a continuación, por cooperación de los bucles de

iteración internos y externos bien conocidos o por cualquier otro procedimiento de codificación apropiado de modo que se satisfacen determinadas condiciones de tasas de bits. A continuación, los valores espectrales por cuantificar, por un lado, y los factores de escala calculados, por otro lado, se introducen en un procesador cuantificador 404. En la operación directa del codificador de audio, los valores espectrales por cuantificar son ponderados por los factores de escala y los valores espectrales ponderados se introducen a continuación en cuantificador fijo que tiene típicamente una funcionalidad de compresión a rangos de amplitud superiores. A continuación, en la salida del procesador cuantificador, existen índices de cuantificación que a continuación se dirigen a un codificador de entropía que típicamente tiene una codificación específica y muy eficaz para un conjunto de índices de cuantificación cero para valores de frecuencia adyacentes o, como también se denomina en la técnica, una “corrida” de valores cero.

[0102] En el codificador de audio de la Fig. 1a, sin embargo, el procesador cuantificador típicamente recibe información acerca de las segundas porciones espectrales del analizador espectral. De esta manera, el procesador cuantificador 404 asegura que, en la salida del procesador cuantificador 404, las segundas porciones espectrales como se identifican por el analizador espectral 102 son cero o tienen una representación reconocida por un codificador o un decodificador como una representación cero que puede codificarse de forma muy eficaz, de modo que existen “corridas” de valores cero en el espectro.

[0103] La Fig. 4b ilustra una implementación del procesador cuantificador. Los valores espectrales de MDCT pueden introducirse en un conjunto de bloque cero 410. A continuación, las segundas porciones espectrales ya se fijan en cero antes de realizar una ponderación por los factores de escala en bloque 412. En una implementación adicional, no se proporciona el bloque 410, pero la fijación a cooperación cero se lleva a cabo en bloque 418 después del bloque de ponderación 412. En otra implementación más, la fijación a operación cero también se puede llevar a cabo en una fijación a bloque cero 422 después de una cuantificación en el bloque de cuantificador 420. En esta implementación, los bloques 410 y 418 no estarían presentes. En general, se proporciona al menos uno de los bloques 410, 418, 422 según la implementación específica.

[0104] A continuación, en la salida del bloque 422, se obtiene un espectro cuantificado correspondiente al que se ilustra en la Fig. 3a. Este espectro cuantificado se introduce a continuación en un codificador de entropía como 232 en la Fig. 2b que puede ser un codificador de Huffman o un codificador aritmético como se define, por ejemplo, en el estándar USAC.

[0105] La fijación a bloques cero 410, 418, 422, que son proporcionados alternativamente entre sí o en paralelo, es controlada por el analizador espectral 424. El analizador espectral comprende preferentemente cualquier implementación de un detector de tonalidad bien conocido o comprende cualquier tipo diferente de detectores operativos para separar un espectro en componentes por codificar con alta resolución y componentes para codificar con una baja resolución. Otros de tales algoritmos implementados en el analizador espectral pueden ser un detector de actividad de voz, un detector de ruidos, un detector del habla o cualquier detector que decide según la información espectral o los metadatos asociados acerca de los requisitos de resolución para diferentes porciones espectrales.

[0106] La Fig. 5a ilustra una implementación preferida del conversor de tiempo–espectro 100 de la Fig. 1A como se implementa, por ejemplo, en AAC o USAC. El conversor de tiempo–espectro 100 comprende un windower 502 controlado por un detector de transitorios 504. Cuando el detector de transitorios 504 detecta un transitorio, entonces se señaliza un cambio de ventanas largas a ventanas cortas al windower. El windower 502 calcula a continuación, para superponer los bloques, tramas de ventana, donde cada trama de ventana tiene típicamente dos valores N tales como los valores 2048. A continuación, se lleva a cabo una transformación dentro de un transformador de bloques 506 y este transformador de bloques proporciona típicamente de modo adicional una pérdida, de modo que una pérdida/transformación combinada se lleva a cabo para obtener una trama espectral con valores N tales como los valores espectrales de MDCT. De esta manera, para una operación de ventana larga, la trama en la entrada del bloque 506 comprende dos valores N tales como los valores 2048 y una trama espectral, entonces tiene valores 1024. A continuación, sin embargo, se lleva a cabo una conmutación a bloques cortos cuando se llevan a cabo ocho bloques cortos, donde cada bloque corto tiene 1/8 valores de dominio temporal de ventana en comparación con una ventana larga y cada bloque espectral tiene 1/8 valores espectrales en comparación con un bloque largo. De esta manera, cuando esta pérdida se combina con una operación de superposición al 50% del windower, el espectro es una versión críticamente muestreada de la señal de audio del dominio de tiempo 99.

[0107] Posteriormente, se hace referencia a la Fig. 5b que ilustra una implementación específica del regenerador de frecuencia 116 y el conversor de espectro–tiempo 118 de la Fig. 1b o de la operación combinada de

bloques 208, 212 de la Fig. 2a. En la Fig. 5b, se considera una banda de reconstrucción específica como una banda del factor de escala 6 de la Fig. 3a. La primera porción espectral en esta banda de reconstrucción, es decir, la primera porción espectral 306 de la Fig. 3a se introduce en el bloque de formador/ajustador de tramas 510. Por otra parte, una segunda porción espectral reconstruida para la banda del factor de escala 6 también se introduce en el formador/ajustador de la trama 510. Por otra parte, también se introduce información de energía como E_3 de la Fig. 3b para una banda de factor de escala 6 en el bloque 510. La segunda porción espectral reconstruida en la banda de reconstrucción ya se generó por relleno de mosaico de frecuencia mediante el uso de un rango de fuente y la banda de reconstrucción se corresponde entonces con el rango de destino. Ahora, un ajuste de energía de la trama se lleva a cabo para obtener después finalmente la trama reconstruida completa que tiene los valores N, por ejemplo, como se obtienen en la salida del combinador 208 de la Fig. 2a. A continuación, en el bloque 512, se lleva a cabo una transformación/interpolación de bloques inversa para obtener 248 valores de dominio de tiempo para, por ejemplo, los 124 valores espectrales en la entrada del bloque 512. A continuación, se lleva a cabo una operación de ventana de síntesis en el bloque 514 que de nuevo es controlada por una indicación de ventana larga/ventana corta transmitida como información lateral en la señal de audio codificada. A continuación, en el bloque 516, se lleva a cabo una operación de superposición/adición con una trama de tiempo previa. Con preferencia, MDCT aplica un 50% de superposición, de modo que, para cada nueva trama de tiempo de valores 2N, finalmente se producen valores de dominio de tiempo N. Se prefiere mucho una superposición del 50% debido al hecho de que proporciona un muestreo crítico y un cruce continuo de una trama a la trama siguiente debido a la operación de superposición/adición en el bloque 516.

[0108] Como se ilustra en 301 en la Fig. 3a, adicionalmente se puede aplicar una operación de relleno de ruido no solo por debajo de la frecuencia inicial de IGF, sino también por encima de la frecuencia inicial de IGF como para la banda de reconstrucción contemplada que coincide con la banda de factor de escala 6 de la Fig. 3a. A continuación, también se pueden introducir valores espectrales de relleno de ruido en el formador/ajustador de trama 510 y el ajuste de los valores espectrales de relleno de ruido también se puede aplicar dentro de este bloque o los valores espectrales de relleno de ruido se pueden ajustar ya mediante el uso de la energía de relleno de ruido antes de ser introducidos en el formador/ajustador de trama 510.

[0109] Con preferencia, se puede aplicar una operación de IGF, es decir, una operación de relleno de mosaico de frecuencia mediante el uso de valores espectrales de otras porciones en el espectro completo. De esta manera, una operación de relleno de mosaico espectral no solo se puede aplicar en la banda alta por encima de una frecuencia inicial de IGF, sino que también se puede aplicar en la banda baja. Por otra parte, el relleno de ruido sin relleno de mosaico de frecuencia también se puede aplicar no solo por debajo de la frecuencia inicial de IGF sino también por encima de la frecuencia inicial de IGF. Sin embargo, se halló que la alta calidad y la codificación de audio muy eficaz se pueden obtener cuando la operación de relleno de ruido está limitada al rango de frecuencias por debajo de frecuencia inicial de IGF y cuando la operación de relleno de mosaico de frecuencia está restringida al rango de frecuencias por encima de la frecuencia inicial de IGF como se ilustra en la Fig. 3a.

[0110] Con preferencia, los mosaicos de destino (TT) (que tienen frecuencias mayores que la frecuencia inicial de IGF) están unidos a bordes de banda de factores de escala del codificador de tasa completa. Los mosaicos de fuente (ST), de los que se extrae la información, es decir, para frecuencias inferiores a la frecuencia inicial de IGF no se unen por bordes de banda de factores de escala. El tamaño de ST debería corresponder con el tamaño de los TT asociados. Esto se ilustra mediante el uso del siguiente ejemplo. TT[0] tiene una longitud de 10 MDCT Bins. Esto corresponde exactamente a la longitud de dos SCBs posteriores (como 4 + 6). A continuación, todos los ST posibles que se han de correlacionar con TT[0], también tienen una longitud de 10 bins. Un segundo mosaico de destino TT[1] que es adyacente a TT[0] tiene una longitud de 15 bins (SCB que tiene una longitud de 7 + 8). A continuación, el ST para ello tiene una longitud de 15 bins más que 10 bins para TT[0].

[0111] Si surgiera el caso de que no se pudiera hallar un TT para un ST con la longitud del mosaico de destino (cuando, por ejemplo, la longitud de TT es mayor que el rango de fuente disponible), entonces no se calcula una correlación y el rango de la fuente se copia una cantidad de veces en este TT (la copia se realiza una tras otra, de modo que una línea de frecuencia para la frecuencia más baja de la segunda copia sigue –en frecuencia– a la línea de frecuencia para la máxima frecuencia de la primera copia), hasta rellenar por completo el mosaico de destino TT.

[0112] Posteriormente, se hace referencia a la Fig. 5c que ilustra otra forma de realización preferida del regenerador de frecuencia 116 de la Fig. 1b o el bloque de IGF 202 de la Fig. 2a. El bloque 522 es un generador de mosaicos de frecuencia que recibe no solo una ID de banda de destino, sino que recibe adicionalmente una ID de banda de fuente. Como ejemplo, se determinó en el lado del codificador que la banda del factor de escala 3 de la

Fig. 3a es muy apropiado para reconstruir la banda del factor de escala 7. De esta manera, la ID de la banda de fuente sería 2 y la ID de la banda de destino sería de 7. En base a esta información, el generador de mosaicos de frecuencia 522 aplica una copia o una operación de relleno de mosaicos de armónicos o cualquier otra operación de relleno de mosaicos para generar la segunda porción en bruto de componentes espectrales 523. La segunda porción en bruto de componentes espectrales tiene una resolución de frecuencia idéntica a la resolución de frecuencia incluida en el primer conjunto de primeras porciones espectrales.

[0113] A continuación, la primera porción espectral de la banda de reconstrucción como 307 de la Fig. 3a se introduce en un formador de tramas 524 y la segunda porción en bruto 523 también se introduce en el formador de tramas 524. A continuación, la trama reconstruida se ajusta por medio del ajustador 526 mediante el uso de un factor de aumento para la banda de reconstrucción calculada por el calculador de factor de aumento 528. De modo importante, sin embargo, la primera porción espectral en la trama no está influida por el ajustador 526, sino que solo la segunda porción en bruto para la trama de reconstrucción es influida por el ajustador 526. Para este fin, el calculador del factor de aumento 528 analiza la banda de fuente o la segunda porción en bruto 523 y según analiza la primera porción espectral en la banda de reconstrucción para hallar finalmente el factor de aumento 527 correcto, de modo que la energía de la trama ajustada producida por el ajustador 526 tenga la energía E_4 cuando se contempla una banda de factor de escala 7.

[0114] En este contexto, es muy importante evaluar la alta precisión de reconstrucción de frecuencia de la presente invención en comparación con HE-AAC. Esto se explica con respecto a la banda del factor de escala 7 en la Fig. 3a. Se asume que un codificador de la técnica anterior tal como se ilustra en la Fig. 13a detectaría la porción espectral 307 por codificar con una alta resolución como una "falta de armónicos". A continuación, la energía de este componente espectral será transmitida junto con una información envolvente espectral para la banda de reconstrucción como banda del factor de escala 7 al decodificador. A continuación, el decodificador recrearía el armónico faltante. Sin embargo, el valor espectral, en el que el armónico faltante 307 se reconstruiría por el decodificador de la técnica anterior de la Fig. 13b estaría en la mitad de la banda 7 a una frecuencia indicada por la frecuencia de reconstrucción 390. De esta manera, la presente invención evita un error de frecuencia 391 que sería introducido por el decodificador de la técnica anterior de la Fig. 13b. En una implementación, el analizador espectral también se implementa para calcular similitudes entre primeras porciones espectrales y segundas porciones espectrales y para determinar, en base a similitudes calculadas, para una segunda porción espectral en un rango de reconstrucción una primera porción espectral que concuerde con la segunda porción espectral lo máximo posible. A continuación, en esta implementación de rango de fuente / rango de destino variable, el codificador paramétrico introducirá adicionalmente en la segunda representación codificada una información concordante que indique para cada rango de destino un rango de fuente concordante. Del lado del decodificador, esta información se usaría entonces por un generador de mosaico de frecuencia 522 de la Fig. 5c que ilustra una generación de una segunda porción en bruto 523 en base a una ID de banda de fuente y una ID de banda de destino. Por otra parte, como se ilustra en la Fig. 3a, el analizador espectral se configura para analizar la representación espectral hasta una frecuencia de análisis máxima que es solo una pequeña cantidad por debajo de la mitad de la frecuencia de muestreo y preferentemente al menos un cuarto de la frecuencia de muestreo o típicamente mayor.

[0115] Como se ilustra, el codificador opera sin muestrear hacia abajo y el decodificador opera sin muestrear hacia arriba. En otras palabras, el codificador de audio del dominio espectral se configura para generar una representación espectral que tiene una frecuencia Nyquist definida por la tasa de muestreo de la señal de audio de entrada original.

[0116] Por otra parte, como se ilustra en la Fig. 3a, el analizador espectral se configura para analizar la representación espectral que comienza con una frecuencia inicial de relleno de intervalo y que termina con una frecuencia máxima representada por una frecuencia máxima incluida en la representación espectral, en donde una porción espectral que se extiende desde una frecuencia mínima hasta la frecuencia inicial de relleno de intervalo pertenece al primer conjunto de porciones espectrales y en donde otra porción espectral como 304, 305, 306, 307 que tiene valores de frecuencia por encima de la frecuencia del relleno de intervalo está incluida adicionalmente en el primer conjunto de primeras porciones espectrales.

[0117] Como se destaca, el decodificador de audio del dominio espectral 112 se configura de modo que una frecuencia máxima representada por un valor espectral en la primera representación decodificada sea igual a una frecuencia máxima incluida en la representación temporal que tiene la tasa de muestreo en donde el valor espectral para la frecuencia máxima en el primer conjunto de primeras porciones espectrales es cero o diferente de cero. De cualquier modo, para esta frecuencia máxima en el primer conjunto de componentes espectrales, existe un factor de escala para la banda del factor de escala, que se genera y transmite independientemente de si todos los valores

espectrales en esta banda del factor de escala se fijan en cero o no, tal como se trata en el contexto de las Figs. 3a y 3b.

[0118] Un ejemplo es ventajoso por tanto con respecto a otras técnicas paramétricas para incrementar la eficacia de compresión, por ejemplo, sustitución de ruido y relleno de ruido (estas técnicas son exclusivamente para la representación eficaz de ruido como contenido de señales locales), la invención permite una reproducción precisa de la frecuencia de componentes tonales. Hasta la fecha, ninguna técnica del estado de la técnica se dirige a una representación paramétrica eficaz de contenido de señales arbitrarias por relleno de intervalo espectral sin la restricción de una división a priori fija en banda baja (LF) y banda alta (HF).

[0119] Las formas de realización del sistema de la invención mejoran las estrategias del estado de la técnica y así proporcionan una alta eficacia de compresión, ninguna o solo una pequeña molestia perceptual y un ancho de banda de audio completo incluso para bajas tasas de bits.

[0120] El sistema general consiste en

- codificación central de banda completa
- relleno de intervalo inteligente (relleno de mosaico o relleno de ruido)
- partes tonales dispersan en el núcleo seleccionadas por máscara tonal
- codificación de pares estéreo conjuntos para banda completa, incluyendo relleno de mosaicos
- TNS en mosaicos
- blanqueo espectral en el rango de IGF

[0121] Una primera etapa hacia un sistema más eficaz consiste en eliminar la necesidad de transformar datos espectrales en un segundo dominio de transformada de un codificador a uno central. Como la mayoría de códecs de audio, como, por ejemplo, AAC, que usa la MDCT como transformada básica, también es de utilidad llevar a cabo el BWE en el dominio de MDCT. Un segundo requisito para el sistema de BWE sería la necesidad de preservar la rejilla tonal preservando incluso los componentes tonales HF y la calidad del audio codificado es así superior a los sistemas existentes. Para cuidar de ambos requisitos antes mencionados, se ha propuesto un sistema llamado Intelligent Gap Filling (IGF). La Fig. 2b muestra el diagrama de bloques del sistema propuesto en el lado del codificador y la Fig. 2a muestra el sistema del lado del decodificador.

[0122] Posteriormente, se describe una trama de post-procesamiento con respecto a la Fig. 13A y la Fig. 13B a fin de ilustrar que la presente invención también se puede implementar en el reconstructor de alta frecuencia 1330 en esta forma de realización de post-procesamiento.

[0123] La Fig. 13a ilustra un diagrama esquemático de un codificador de audio para una tecnología de extensión de ancho de banda, por ejemplo, como se usa en la codificación de audio avanzada de alta eficacia (High Efficiency Advanced Audio Coding, HE-AAC). Una señal de audio en la línea 1300 se introduce en un sistema de filtro que comprende un paso bajo 1302 y un paso alto 1304. La salida de la señal por el filtro de paso alto 1304 se introduce en un extractor/codificador paramétrico 1306. El extractor/codificador paramétrico 1306 se configura para calcular y codificar parámetros tales como un parámetro de envolvente espectral, un parámetro de adición de ruido, un parámetro de armónicos faltantes o un parámetro de filtro inverso, por ejemplo. Estos parámetros extraídos se introducen en un multiplexador de corriente de bits 1308. La señal de salida de paso bajo se introduce en un procesador que típicamente comprende la funcionalidad de un muestreador descendente 1310 y un codificador central 1312. El paso bajo 1302 restringe el ancho de banda que se va a codificar a un ancho de banda significativamente menor al que ocurre en la señal de audio de entrada original en la línea 1300. Esto proporciona un aumento de codificación significativo debido al hecho de que todas las funcionalidades que se producen en el codificador central solo tienen que operar en una señal con un ancho de banda reducido. Cuando, por ejemplo, el ancho de banda de la señal de audio en la línea 1300 es de 20 kHz y cuando el filtro de paso bajo 1302 tiene ilustrativamente un ancho de banda de 4 kHz, a fin de satisfacer el teorema de muestreo, teóricamente es suficiente con que la señal posterior al muestreador descendente tenga una frecuencia de muestreo de 8 kHz, que es una reducción sustancial de la tasa de muestreo requerida para la señal de audio 1300 que tiene que ser de al menos 40 kHz.

[0124] La Fig. 13b ilustra un diagrama esquemático de un correspondiente decodificador de la extensión de ancho de banda. El decodificador comprende un multiplexador de corrientes de bits 1320. El demultiplexador de corrientes de bits 1320 extrae una señal de entrada para un decodificador central 1322 y una señal de entrada para un decodificador paramétrico 1324. Una señal de salida del decodificador central tiene, en el ejemplo anterior, una

tasa de muestreo de 8 kHz y, por ello, un ancho de banda de 4 kHz, mientras que, para una reconstrucción completa del ancho de banda, la señal de salida de un reconstructor de alta frecuencia 1330 debe estar a 20 kHz que requiere una tasa de muestreo de al menos 40 kHz. A fin de hacer esto posible, se requiere un decodificador procesador que tiene la funcionalidad de un muestreador ascendente 1325 y un banco de filtros 1326. El reconstructor de alta frecuencia 1330 recibe a continuación la señal de frecuencia baja analizada producida por el banco de filtros 1326 y reconstruye el rango de frecuencias definido por el filtro de paso alto 1304 de la Fig. 13a mediante el uso de la representación paramétrica de la banda de alta frecuencia. El reconstructor de alta frecuencia 1330 tiene varias funcionalidades tales como la regeneración del rango de frecuencia superior mediante el uso del rango de la fuente en el rango de frecuencia inferior, un ajuste de envolvente espectral, una funcionalidad de adición de ruido y una funcionalidad para introducir los armónicos faltantes en el rango de frecuencia superior y, si se aplica y se calcula en el codificador de la Fig. 13a, una operación de filtrado inversa para contar con el hecho de que el rango de mayor frecuencia no es típicamente tan tonal como el rango de menor frecuencia. En HE-AAC, los armónicos faltantes se resintetizan en el lado del decodificador y se colocan exactamente en la mitad de una banda de reconstrucción. Así, todas las líneas de armónicos faltantes que fueron determinados en cierta banda de reconstrucción no se colocan en los valores de frecuencia donde se ubicaron en la señal original. En vez de ello, esos armónicos faltantes se colocan en frecuencias en el centro de la banda determinada. De esta manera, cuando una línea de armónicos faltantes en la señal original se colocó muy cerca de un borde de la banda de reconstrucción en la señal original, el error en frecuencia introducido al colocar esta línea de armónicos faltantes en la señal reconstruida en el centro de la banda está cerca del 50% de la banda de reconstrucción individual, para la que se generaron y transmitieron parámetros.

[0125] Por otra parte, aunque los codificadores centrales de audio típicos operen en el dominio espectral, el decodificador central genera sin embargo una señal de dominio de tiempo que es convertida de nuevo en un dominio espectral por la funcionalidad del banco de filtros 1326. Esto introduce demoras de procesamiento adicionales, puede introducir artefactos debido al procesamiento en tándem primeramente de la transformación del dominio espectral al dominio de frecuencia y de nuevo la transformación típicamente en un dominio de frecuencia diferente y, de hecho, esto también requiere una cantidad sustancial de complejidad de computación y, así, energía eléctrica, que es específicamente un problema cuando la tecnología de extensión de ancho de banda se aplica en dispositivos móviles tales como teléfonos móviles, tablets o laptops, etc.

[0126] Aunque algunos aspectos han sido descritos en el contexto de un aparato para codificar o decodificar, es evidente que estos aspectos también representan una descripción del correspondiente procedimiento, en donde un recuadro o dispositivo corresponden a una etapa de procedimiento o a un aspecto de una etapa de procedimiento. De manera análoga, los aspectos descritos en el contexto de una etapa de procedimiento también representan una descripción del correspondiente recuadro o ítem o rasgo de un correspondiente aparato. Algunas de las etapas de procedimiento, o todas ellas, pueden ejecutarse mediante un aparato de hardware (o utilizando un aparato de hardware), tal como por ejemplo, un microprocesador, un ordenador programable o un circuito electrónico. En algunas formas de realización algunas o más de las etapas de procedimiento más importantes pueden ser ejecutadas mediante un aparato de este tipo.

[0127] En función de determinados requisitos de implementación, las formas de revisión de la invención pueden ser implementadas en hardware o en software. La implementación puede ser llevada a cabo utilizando un medio de almacenamiento no transitorio como un medio de almacenamiento digital, por ejemplo, un disquete, una unidad de disco duro (HDD), un DVD, un Blu-Ray, un CD, una memoria ROM, PROM y un EPROM, EEPROM o FLASH, que tengan señales de control electrónicamente legibles almacenadas en ellos, que cooperan (o son capaces de cooperar) con un sistema informático programable de tal manera que el procedimiento respectivo sea llevado a cabo. Por lo tanto, el medio de almacenamiento digital puede ser legible por ordenador.

[0128] Algunas formas de realización según la invención comprenden un soporte de datos que tiene señales de control electrónicamente legibles, que son capaces de cooperar con un sistema informático programable, de tal manera que se lleve a cabo uno de los procedimientos descritos en esta invención.

[0129] En términos generales, las formas de realización de la presente invención pueden ser implementadas como un producto programa informático provisto de un código de programa, siendo el código de programa operativo para llevar a cabo uno de los procedimientos cuando el producto programa informático es ejecutado en un ordenador. El código de programa puede estar almacenado, por ejemplo, en un soporte legible por máquina.

[0130] Otras formas de realización comprenden el programa informático para llevar a cabo uno de los procedimientos descritos en esta invención, almacenado en un soporte legible por máquina.

[0131] En otras palabras, una forma de realización del procedimiento inventivo es por lo tanto un programa informático que tiene un código de programa para llevar a cabo uno de los procedimientos descritos en esta invención, cuando el programa informático es ejecutado en un ordenador.

5 **[0132]** Otra forma de realización del procedimiento según la invención es por lo tanto un soporte de datos (o un medio de almacenamiento digital, o un medio legible por ordenador) que comprende, grabado en él, el programa informático para llevar a cabo uno de los procedimientos descritos en esta invención. El soporte de datos, el medio de almacenamiento digital o el medio grabado son típicamente tangibles y/o no transitorios.

10 **[0133]** Otra forma de realización del procedimiento inventivo es una corriente de datos o una secuencia de señales que representan el programa informático para llevar a cabo uno de los procedimientos descritos en esta invención. La corriente de datos o la secuencia de señales, puede estar configurada, por ejemplo, para ser transferida por intermedio de una conexión de comunicación de datos, por ejemplo, a través de Internet.

15 **[0134]** Otra forma de realización comprende un medio de procesamiento, por ejemplo, un ordenador, o un dispositivo lógico programable, configurados para, o adaptados para, llevar a cabo uno de los procedimientos descritos en esta invención.

[0135] Otra forma de realización comprende un ordenador que tiene instalado en él el programa informático para llevar a cabo uno de los procedimientos descritos en esta invención.

[0136] Otra forma de realización según la invención comprende un aparato o un sistema configurados para transferir (por ejemplo, electrónicamente u ópticamente) un programa informático para llevar a cabo uno de los procedimientos distintos en esta invención, a un receptor. El receptor puede ser, por ejemplo, un ordenador, un
25 dispositivo móvil, un dispositivo de memoria o similar. El aparato o sistema puede comprender por ejemplo un servidor de archivos para transferir el programa informático al receptor.

[0137] En algunas formas de realización, puede utilizarse un dispositivo lógico programable (por ejemplo, una disposición de puerta programable en campo, para llevar a cabo algunas, o la totalidad de, las funcionalidades de los
30 procedimientos descritos en esta invención. En algunas formas de realización, una disposición de puerta programable en campo puede cooperar con un microprocesador a efectos de llevar a cabo uno de los procedimientos descritos en esta invención. En términos generales, los procedimientos son preferiblemente llevados a cabo mediante cualquier aparato de hardware.

35 **[0138]** Las formas de realización descritas anteriormente son meramente ilustrativas de los principios de la presente invención. Se da por entendido que modificaciones y variaciones en las disposiciones y detalles descritos en esta invención serán evidentes para otras personas expertas en la técnica; por lo tanto, la invención está destinada a estar limitada solamente por los alcances de las reivindicaciones de patente pendientes y no por los detalles específicos que han sido presentados a modo de descripción y explicación de las formas de realización
40 presentadas en esta invención.

REIVINDICACIONES

1. Aparato para generar una señal mejorada a partir de una señal de entrada (600), en el que la señal mejorada tiene valores espectrales para una región espectral de mejora, donde los valores espectrales para las regiones espectrales de mejora no están contenidos en la señal de entrada (600), que comprende:
un mapeador (602) para mapear una región espectral de la fuente de la señal de entrada en una región de destino en la región espectral de mejora, donde la región espectral de la fuente comprende una región de relleno de ruido (302); y
un rellenador de ruido (604) configurado para generar segundos valores de ruido para una región de ruido en la región de destino, en el que los segundos valores de ruido están decorrelacionados desde los primeros valores de ruido en la región fuente o para la generación de primeros valores de ruido para la región de relleno de ruido (302) en la región espectral de la fuente de la señal de entrada y para la generación de segundos valores de ruido para una región de ruidos en la región de destino, en la que los segundos valores de ruido se decorrelacionan de los primeros valores de ruido, en los que el rellenador de ruido (604) está configurado para:
identificar la región de relleno de ruido que tiene los primeros valores de ruido en la señal de entrada;
copiar (902) al menos una región de la señal de entrada en un búfer mosaico de fuente, comprendiendo la región, la región espectral de la fuente;
reemplazar (904) los primeros valores de ruido tal como se identifica por los segundos valores de ruido; y
en el que el mapeador (602) se configura para mapear el búfer mosaico de fuente que tiene valores de ruido decorrelacionados en la región de destino.
2. Aparato de la reivindicación 1,
en el que la señal de entrada es una señal codificada que comprende parámetros de relleno de ruido para la región espectral de la fuente de la señal de entrada,
en el que el rellenador de ruido está configurado para generar los primeros valores de ruido mediante el uso de los parámetros de relleno de ruido y para generar los segundos valores de ruido mediante el uso de una información de energía en los primeros valores de ruido.
3. Aparato de la reivindicación 1,
en el que el rellenador de ruido (604) se configura para medir una información de energía en los valores de ruido decorrelacionados (1102) y una información de energía acerca de los primeros valores de ruido (1100) y escalar (906) los valores de ruido decorrelacionados mediante el uso de un valor de escala derivado de la información de energía acerca de los valores de ruido decorrelacionados y la información de energía acerca de los primeros valores de ruido.
4. Aparato de una de las reivindicaciones anteriores,
en el que el rellenador de ruido (604) se configura para generar el segundo valor de ruido posterior a una operación del mapeador (602) o para generar (604) los primeros y los segundos valores de ruido posterior a una operación del mapeador (602).
5. Aparato según una de las reivindicaciones anteriores,
en el que el mapeador (602) se configura para mapear la región espectral de la fuente en la región de destino, y
en el que el rellenador de ruido (604) se configura para llevar a cabo relleno de ruido en regiones espectrales generando los primeros valores de ruido mediante el uso de relleno de ruido y parámetros de relleno de ruido transmitidos en la señal de entrada como información lateral y para llevar a cabo relleno de ruido en la región de destino para generar los segundos valores espectrales mediante el uso de información de energía en los primeros valores de ruido.

6. Aparato de una de las reivindicaciones anteriores, que comprende además:

un ajustador de envolvente para ajustar (1202) los segundos valores de ruido en la región espectral de mejora
5 mediante el uso de información envolvente espectral incluida en la señal de entrada como información lateral.

7. Aparato de una de las reivindicaciones anteriores,

en el que el rellenador de ruido (604) se configura para usar solo información lateral de la señal de entrada para
10 identificar posiciones espectrales para relleno de ruido, o

en el que el rellenador de ruido (604) se configura para analizar una característica de tiempo o espectral de la señal
de entrada con o sin valores espectrales en la región de relleno de ruido para identificar posiciones espectrales para
relleno de ruido.

15

8. Aparato de una de las reivindicaciones anteriores,

en el que el rellenador de ruido (604) se configura para identificar posiciones de ruido mediante el uso de un vector
de identificación (706) que tiene entradas para posiciones espectrales en la región espectral de la fuente solo o que
20 tiene entradas para posiciones espectrales en la región espectral de la fuente y en la región espectral de destino.

9. Aparato de la reivindicación 1 ó 3,

en el que el rellenador de ruido (604) se configura para copiar, en la operación de copiado (902), una porción
25 espectral completa de la señal de entrada o una porción espectral completa de la señal de entrada sobre una
frecuencia de borde de relleno de ruido generalmente utilizable por el mapeador para el búfer mosaico de fuente y
para llevar a cabo la operación de reemplazo (904) en el búfer mosaico de fuente completo, o

en el que el rellenador de ruido (604) se configura para copiar, en la operación de copiado (902), solo una región
30 espectral de la señal de entrada identificada por uno o varios identificadores de fuente específicos para una región
de fuente que va a ser utilizada por el mapeador (602) para una región identificada de destino, donde un búfer
mosaico de fuente individual se usa para cada operación de mapeo individual diferente.

10. Aparato de una de las reivindicaciones anteriores,

35

en el que el mapeador (602) se configura para llevar a cabo una operación de relleno de intervalo para generar la
región espectral de destino, que comprende:

un decodificador de audio del dominio espectral (112) para generar una primera representación decodificada de un
40 primer conjunto de primeras porciones espectrales, teniendo la representación decodificada una primera resolución
espectral;

un decodificador paramétrico (114) para generar una segunda representación decodificada de un segundo conjunto
de segundas porciones espectrales que tiene una segunda resolución espectral que es inferior a la primera
45 resolución espectral;

un regenerador de frecuencia (116) para regenerar una segunda porción espectral reconstruida que tiene la primera
resolución espectral mediante el uso de una primera porción espectral e información envolvente espectral para la
segunda porción espectral; y

50

un conversor de espectro-tiempo (118) para convertir la primera representación decodificada en la segunda porción
espectral reconstruida en una representación temporal,

en el que el mapeador (602) y el rellenador de ruido (604) están al menos parcialmente incluidos en el regenerador
55 de frecuencia (116).

11. Aparato de la reivindicación 10,

en el que el decodificador de audio del dominio espectral se configura para producir una secuencia de tramas de

valores espectrales decodificados, donde una trama decodificada es la primera representación decodificada, en el que la trama comprende valores espectrales para el primer conjunto de porciones espectrales y cero indicaciones para el segundo conjunto de segundas porciones espectrales,

5 en el que el aparato para decodificar comprende también un combinador (208) para combinar valores espectrales generados por el regenerador de frecuencia para el segundo conjunto de segundas porciones espectrales y valores espectrales del primer conjunto de primeras porciones espectrales en una banda de reconstrucción para obtener una trama espectral reconstruida que comprende valores espectrales para el primer conjunto de las primeras porciones espectrales y el segundo conjunto de segundas porciones espectrales; y

10 en el que el conversor de espectro-tiempo (118) se configura para convertir la trama espectral reconstruida en la representación temporal.

12. Aparato de una de las reivindicaciones anteriores,

15 que comprende además:

para cada rango de frecuencia de destino, una identificación de región de fuente y

20 en el que el mapeador (602) se configura para seleccionar la región de fuente mediante el uso de la identificación de la región de fuente y para mapear la región de fuente seleccionada en la región de destino.

13. Procedimiento de generación de una señal mejorada a partir de una señal de entrada (600), en el que la señal mejorada tiene valores espectrales para una región espectral de mejora, no estando los valores espectrales para las regiones espectrales de mejora contenidos en la señal de entrada (600), que comprende:

mapear (602) una región espectral de la fuente de la señal de entrada en una región de destino en la región espectral de mejora, donde la región espectral de la fuente comprende una región de relleno de ruido (302); y

30 generar (604) segundos valores de ruido para una región de ruido en la región de destino, en la que los segundos valores de ruido están decorrelacionados desde los primeros valores de ruido en la región fuente, o generar (604) primeros valores de ruido para la región de relleno de ruido (302) en la región espectral de la fuente de la señal de entrada y para generar segundos valores de ruido para una región de ruidos en la región de destino, en la que los segundos valores de ruido se decorrelacionan de los primeros valores de ruido, en la que la generación (604) comprende:

identificar la región de relleno de ruido que tiene los primeros valores de ruido en la señal de entrada;

40 copiar (902) al menos una región de la señal de entrada a un búfer mosaico de fuente, donde la región comprende la región espectral de la fuente; y

reemplazar (904) los primeros valores de ruido tal como se identifica por los valores de ruido independientes; y

45 en el que el mapeo (602) comprende mapear el búfer mosaico de fuente que tiene valores de ruido decorrelacionados a la región de destino.

14. Sistema para procesar una señal de audio, que comprende:

un codificador para generar una señal codificada; y

50 el aparato para generar una señal mejorada según cualquiera de las reivindicaciones 1 a 12, en el que la señal codificada se somete a un procesamiento (700) a fin de generar la señal de entrada (600) en el aparato para generar la señal de audio mejorada.

55 15. Procedimiento para procesar una señal de audio, que comprende:

generar una señal codificada a partir de una señal de entrada; y

un procedimiento para generar una señal mejorada según la reivindicación 13, en el que la señal codificada se

somete a un procesamiento predefinido (700) con el fin de generar la señal de entrada (600) en el aparato para generar la señal de audio mejorada.

16. Programa informático que comprende instrucciones que, cuando el programa es ejecutado por un ordenador, dan lugar a que el ordenador lleve a cabo el procedimiento de la reivindicación 13 o la reivindicación 15.

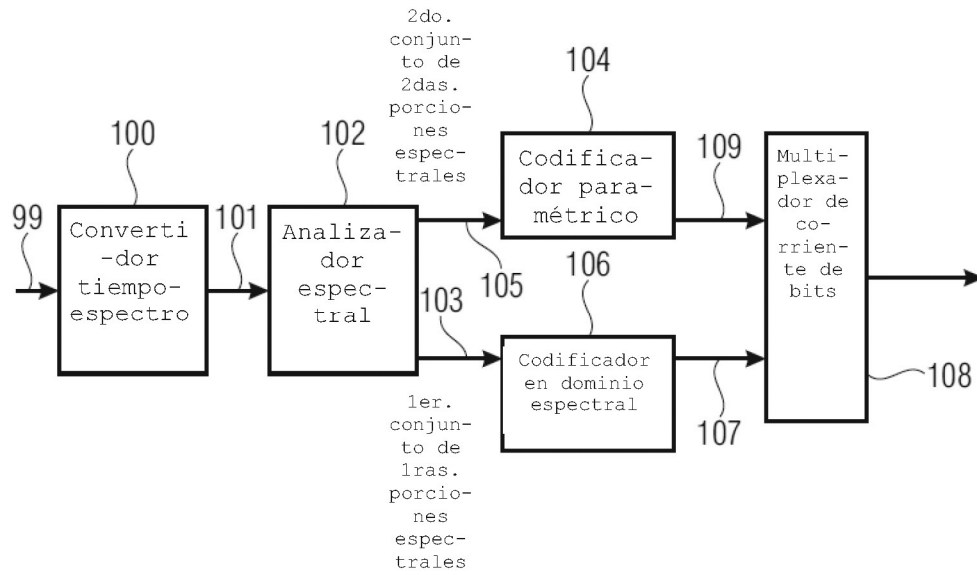


FIG 1A

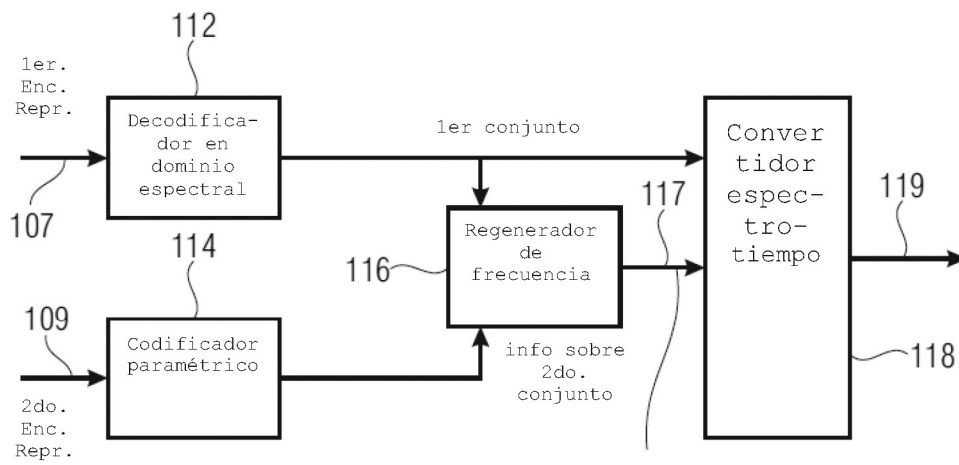


FIG 1B

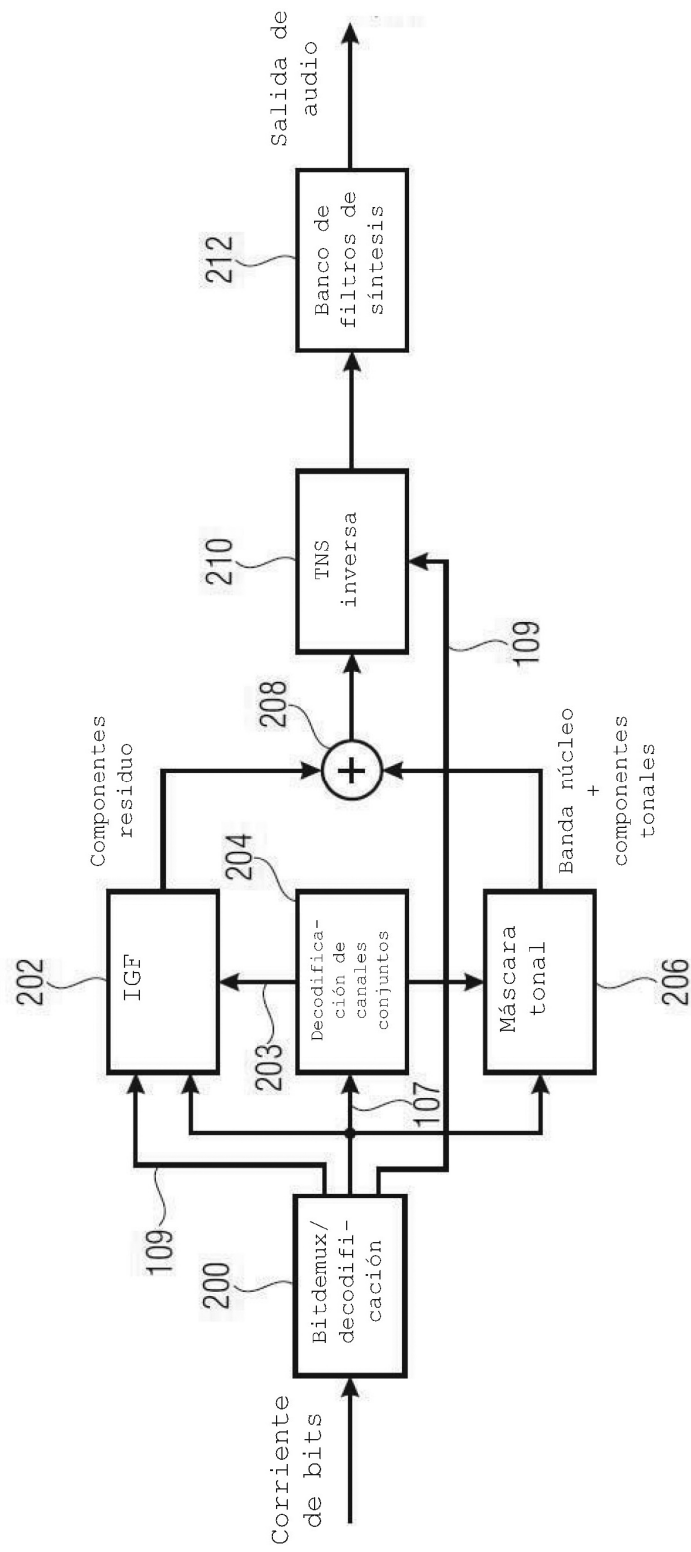


FIG 2A

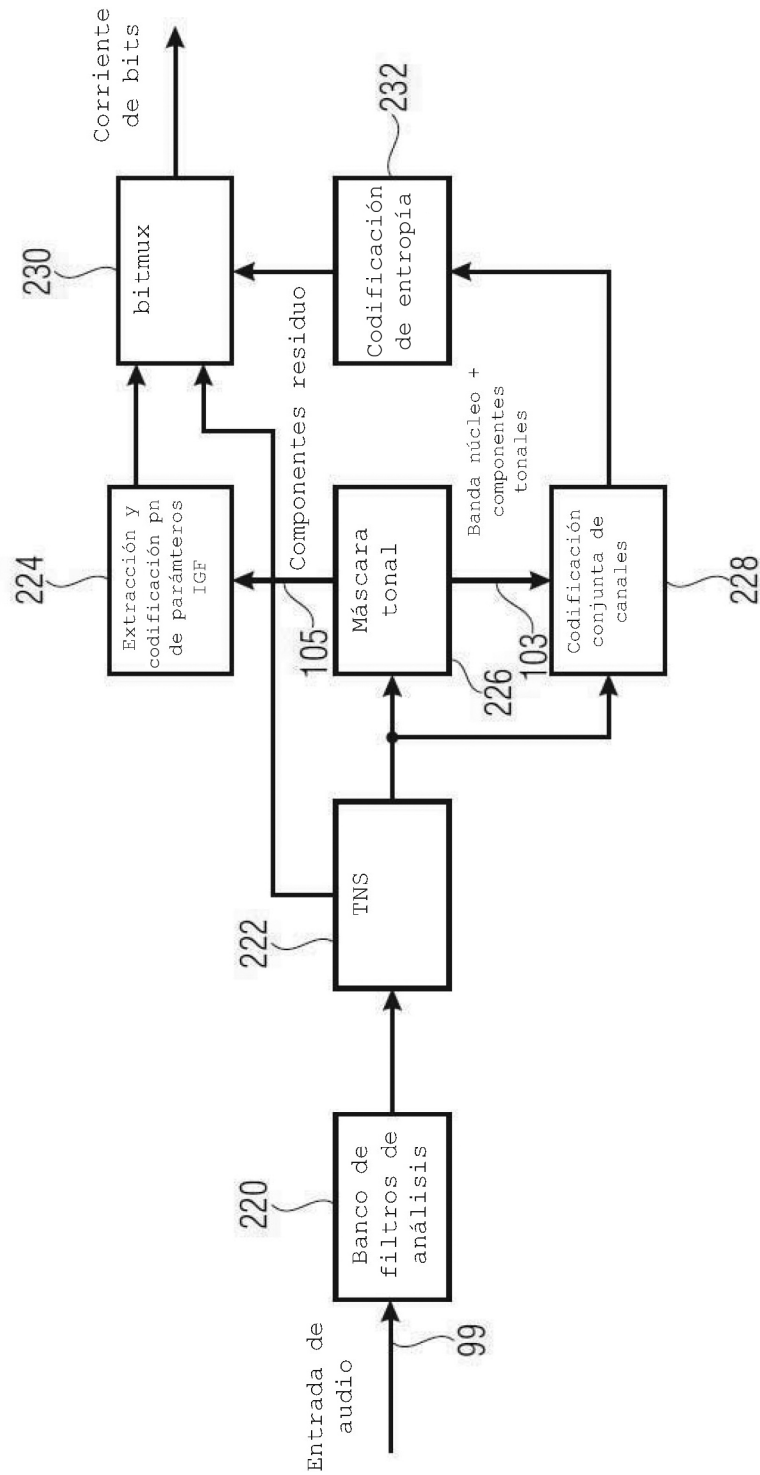


FIG 2B

- 1ra. resolución (elevada resolución) para envoltorio del 1er conjunto (codificación de a líneas)
- 2da. resolución (baja resolución) para envoltorio del 2do conjunto (factor de escala por SCB)

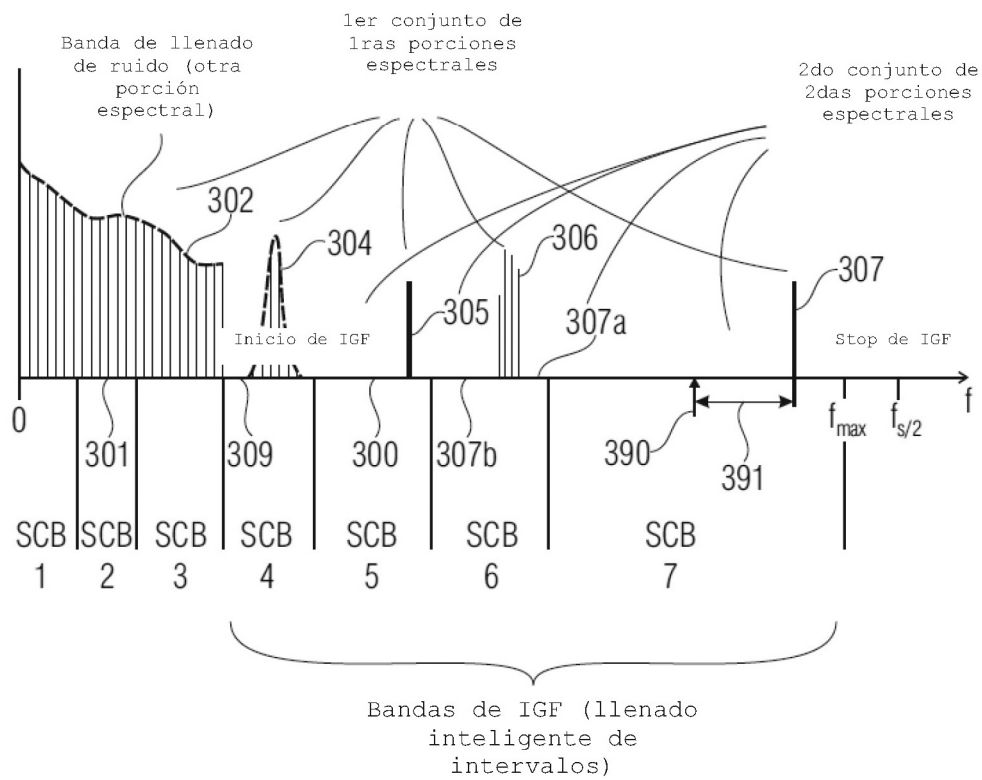


FIG 3A

SCB1	SCB2	SCB3	SCB4	SCB5	SCB6	SCB7
SF1	SF2	SF3	SF4	SF5	SF6	SF7
			E ₁	E ₂	E ₃	E ₄
	NF ₂					

308
310
312

FIG 3B

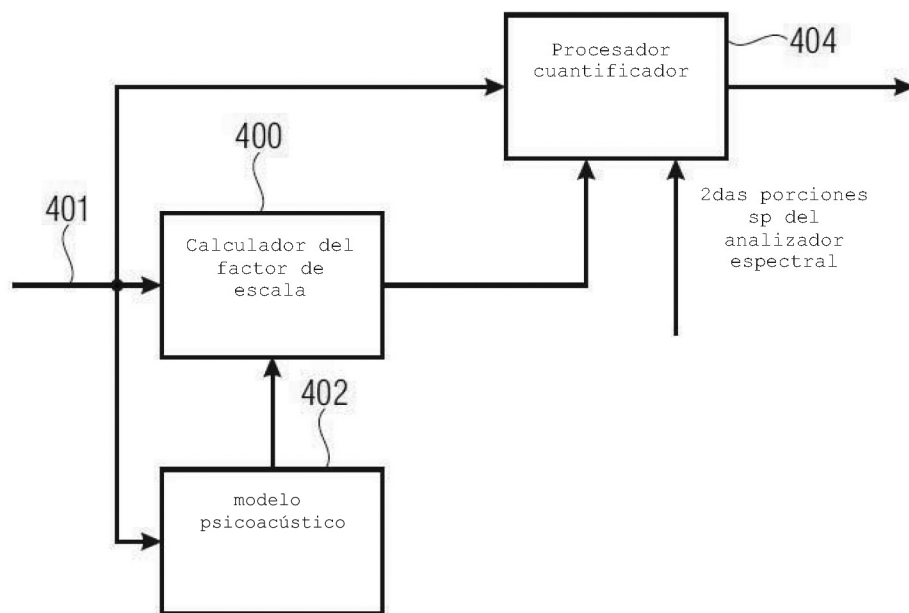


FIG 4A

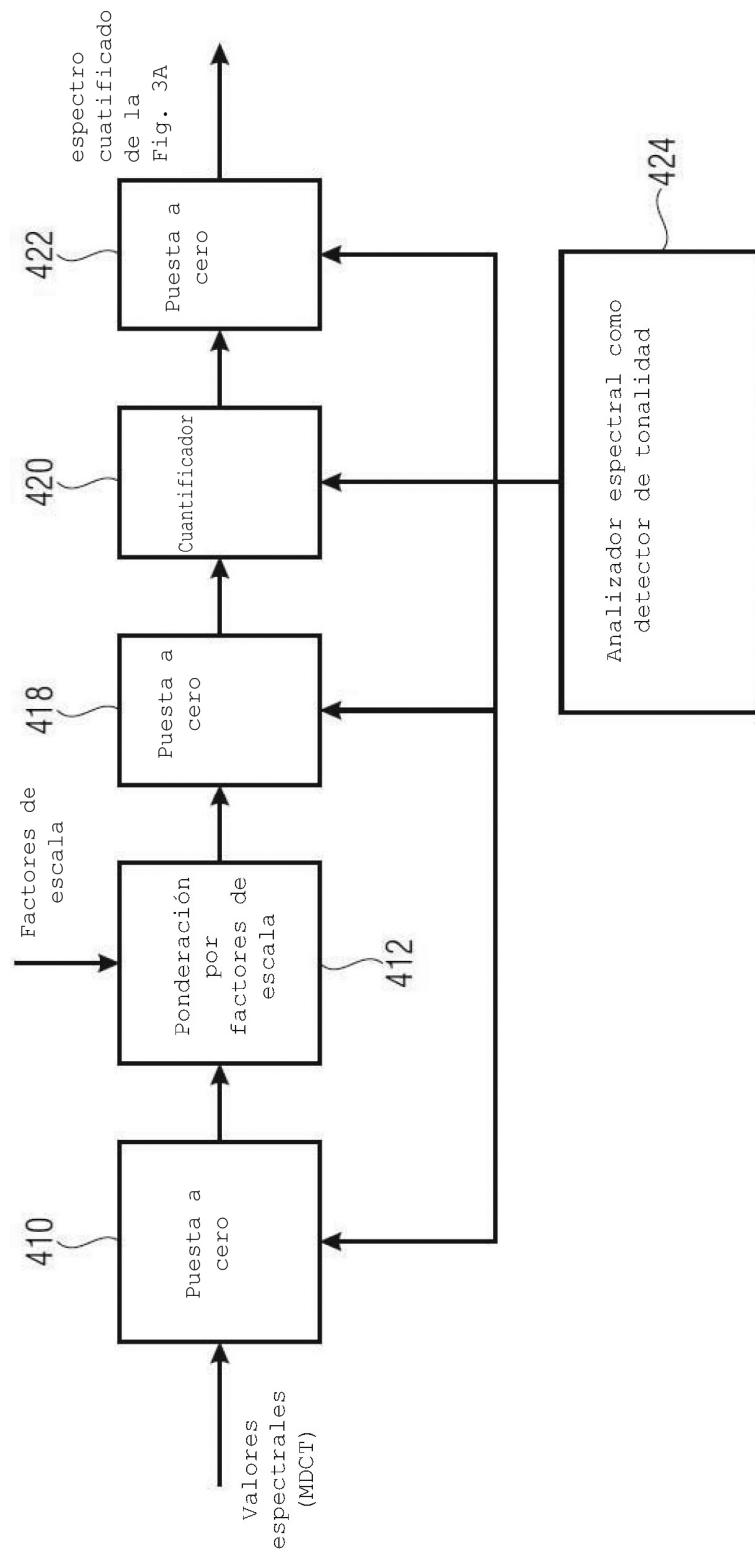


FIG 4B

PROCESADOR CUANTIFICADOR

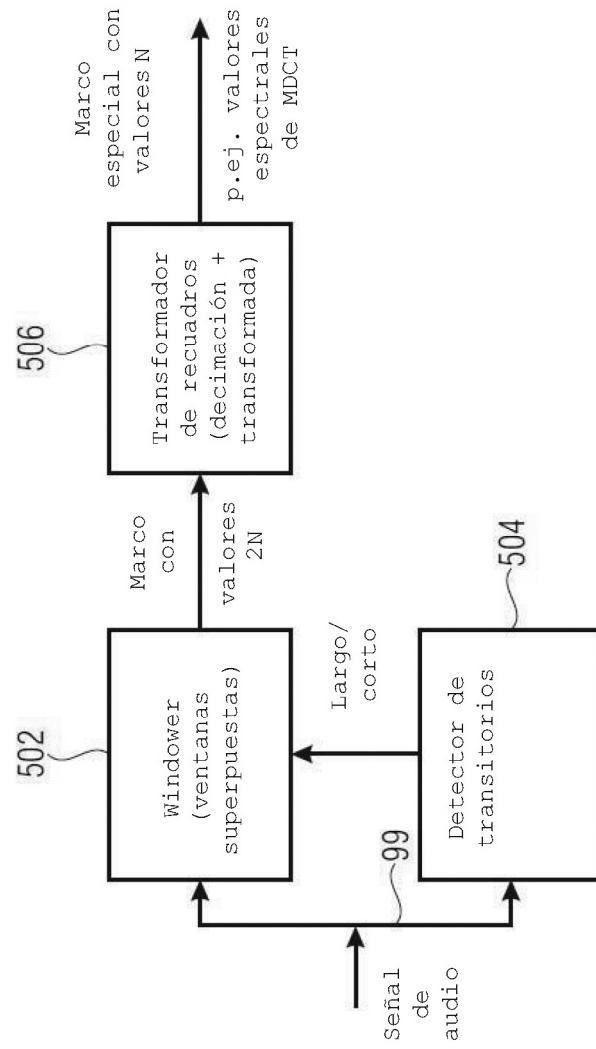


FIG 5A

(otras porciones espectrales)

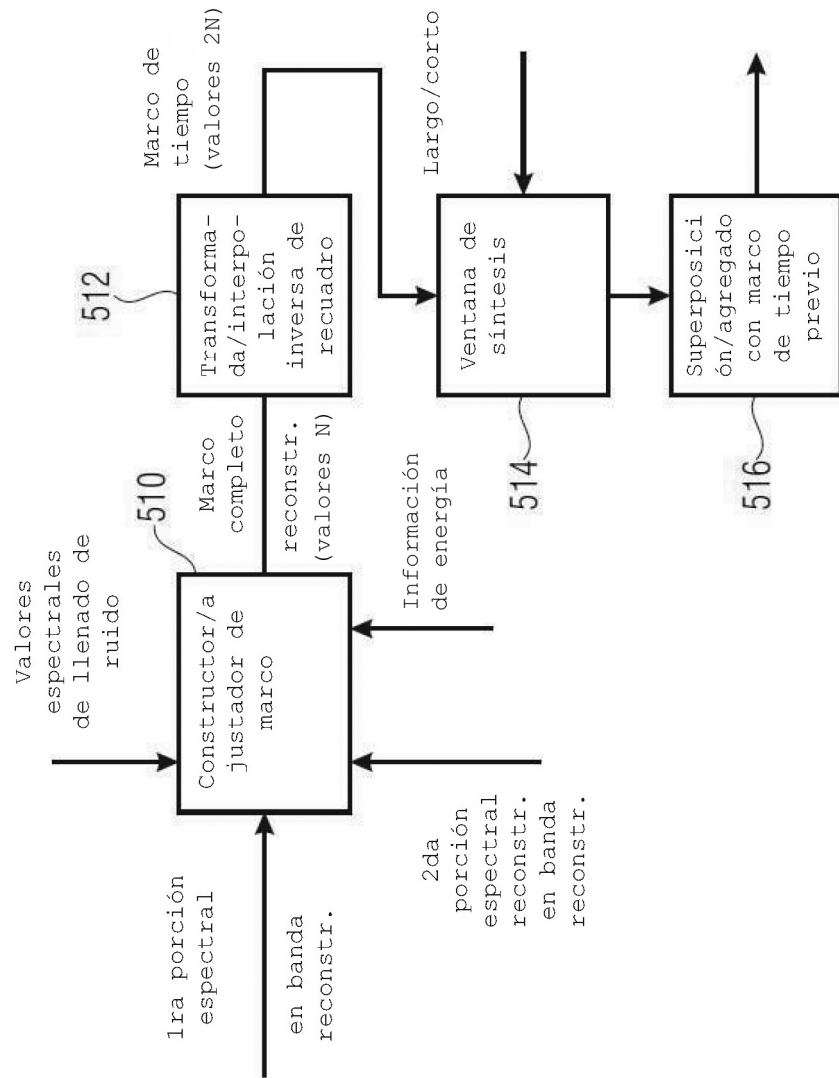


FIG 5B

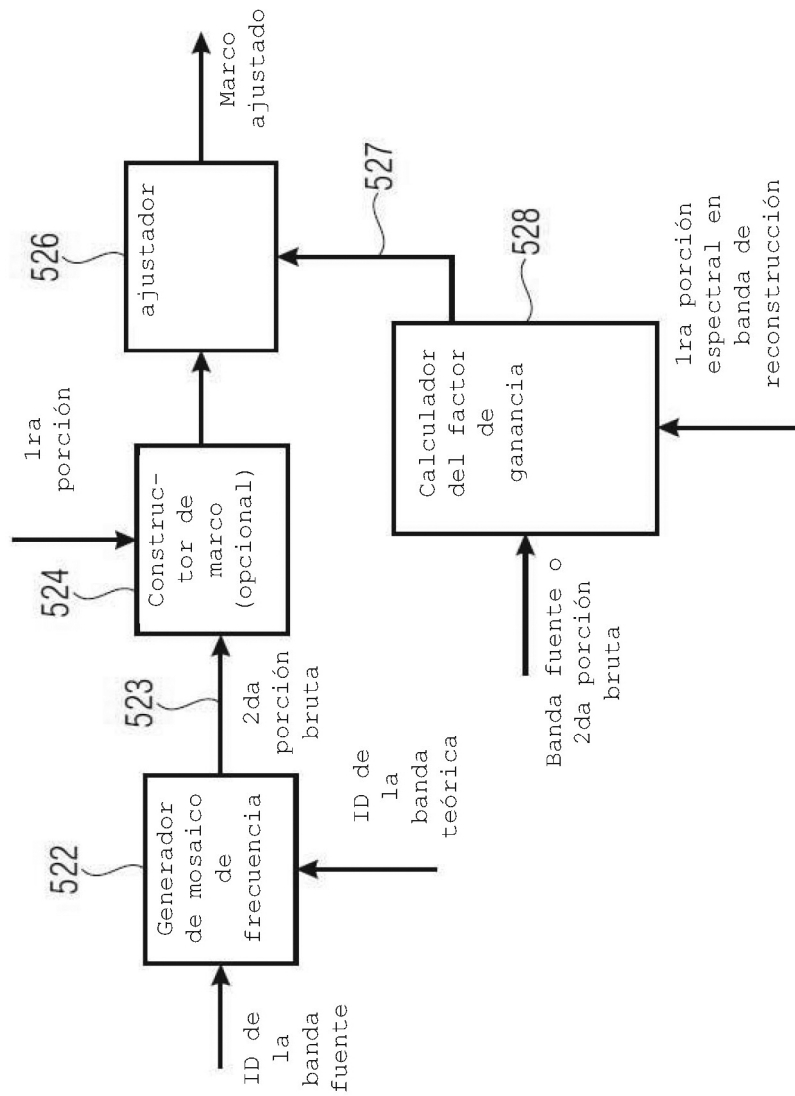


FIG 5C

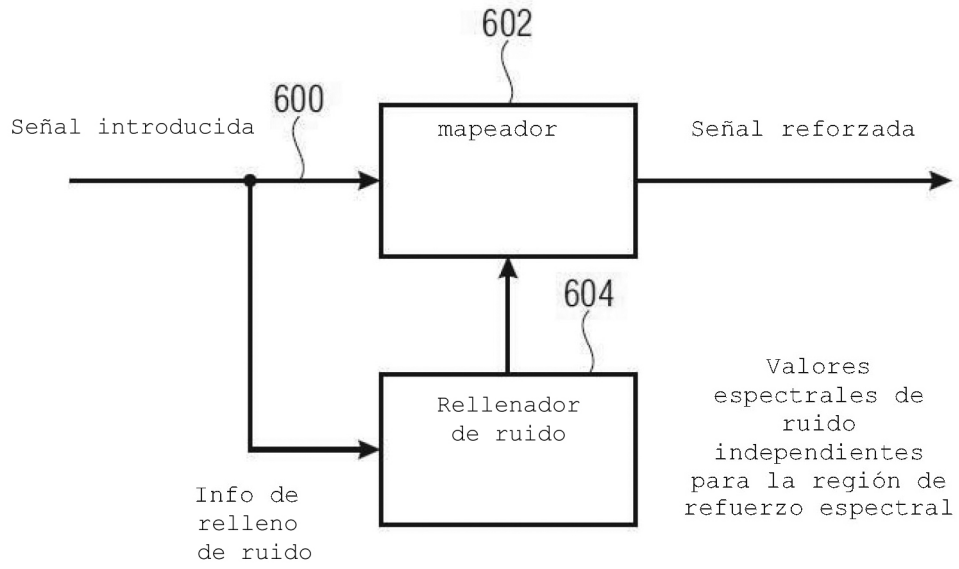


FIG 6

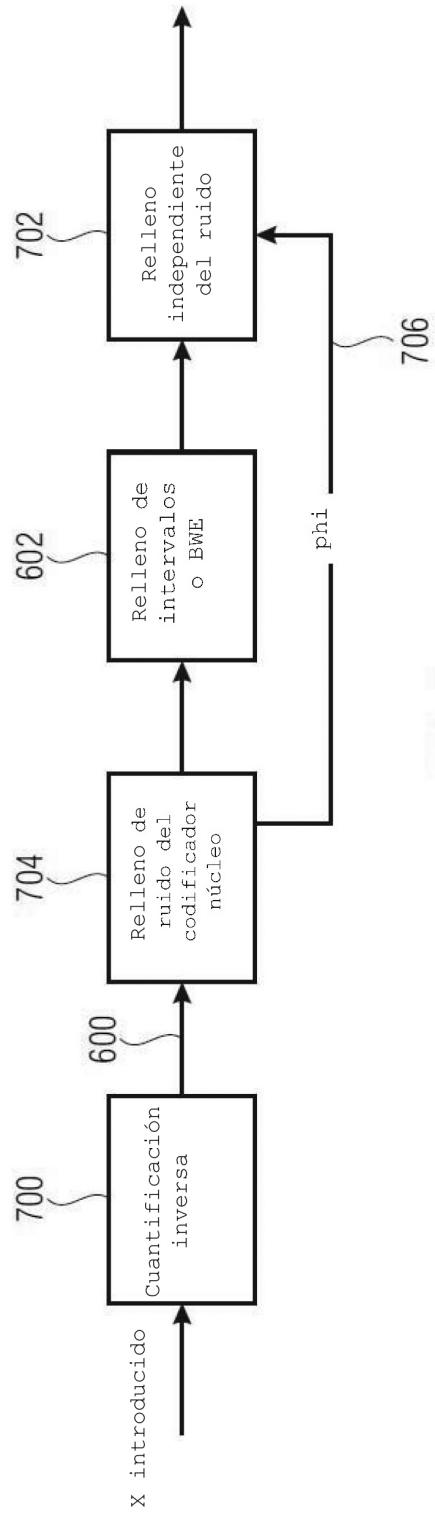


FIG 7

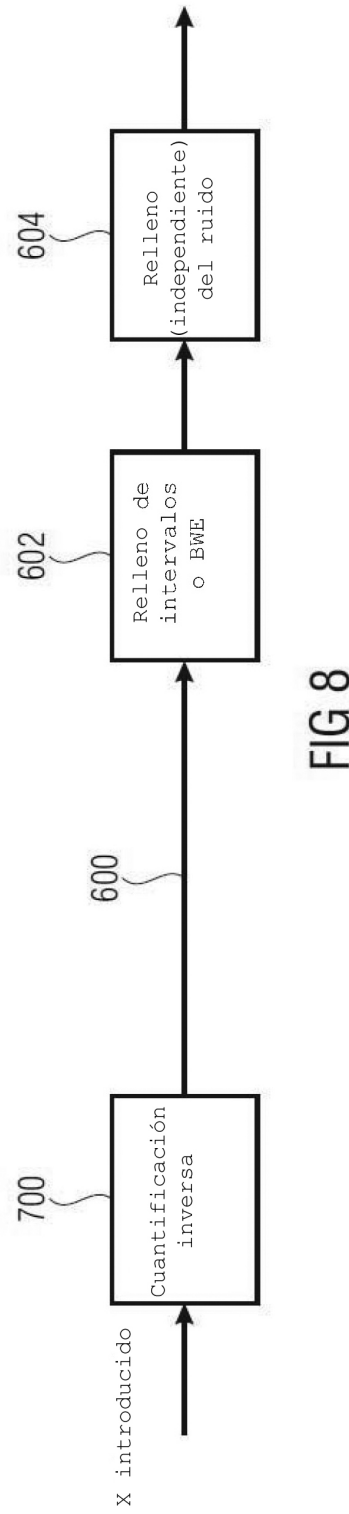


FIG 8

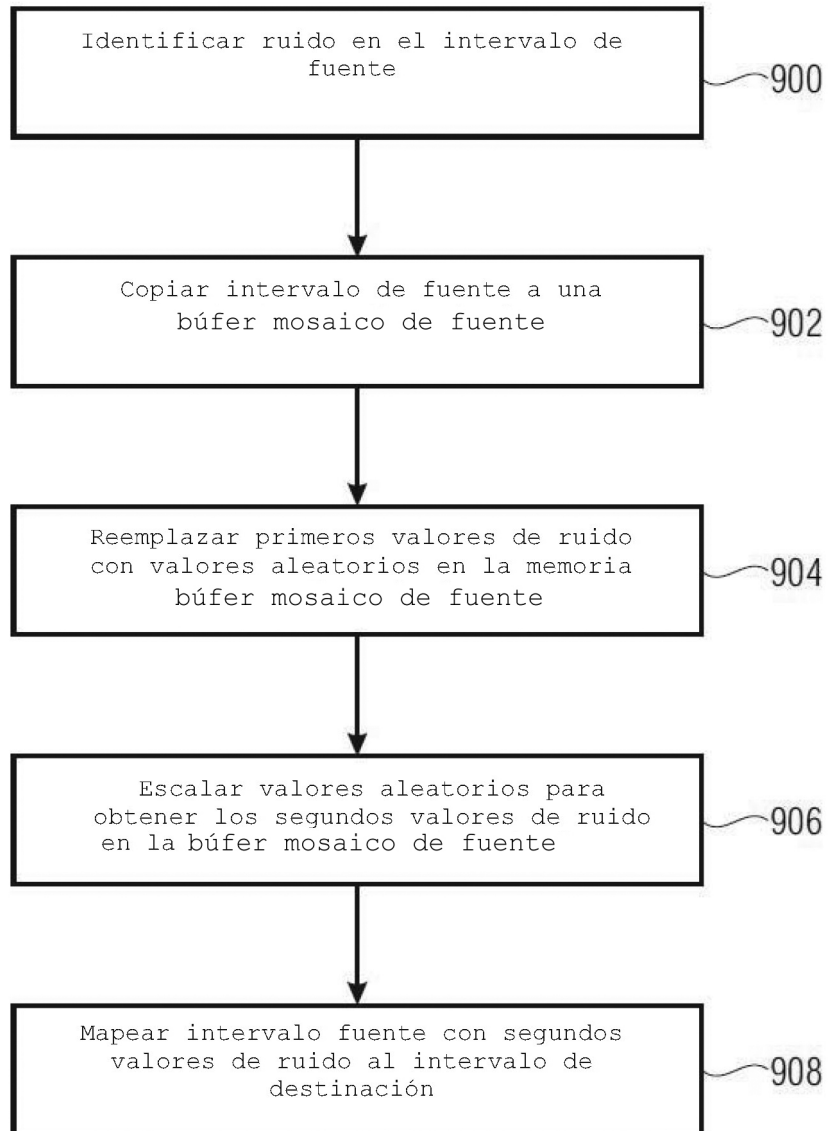


FIG 9

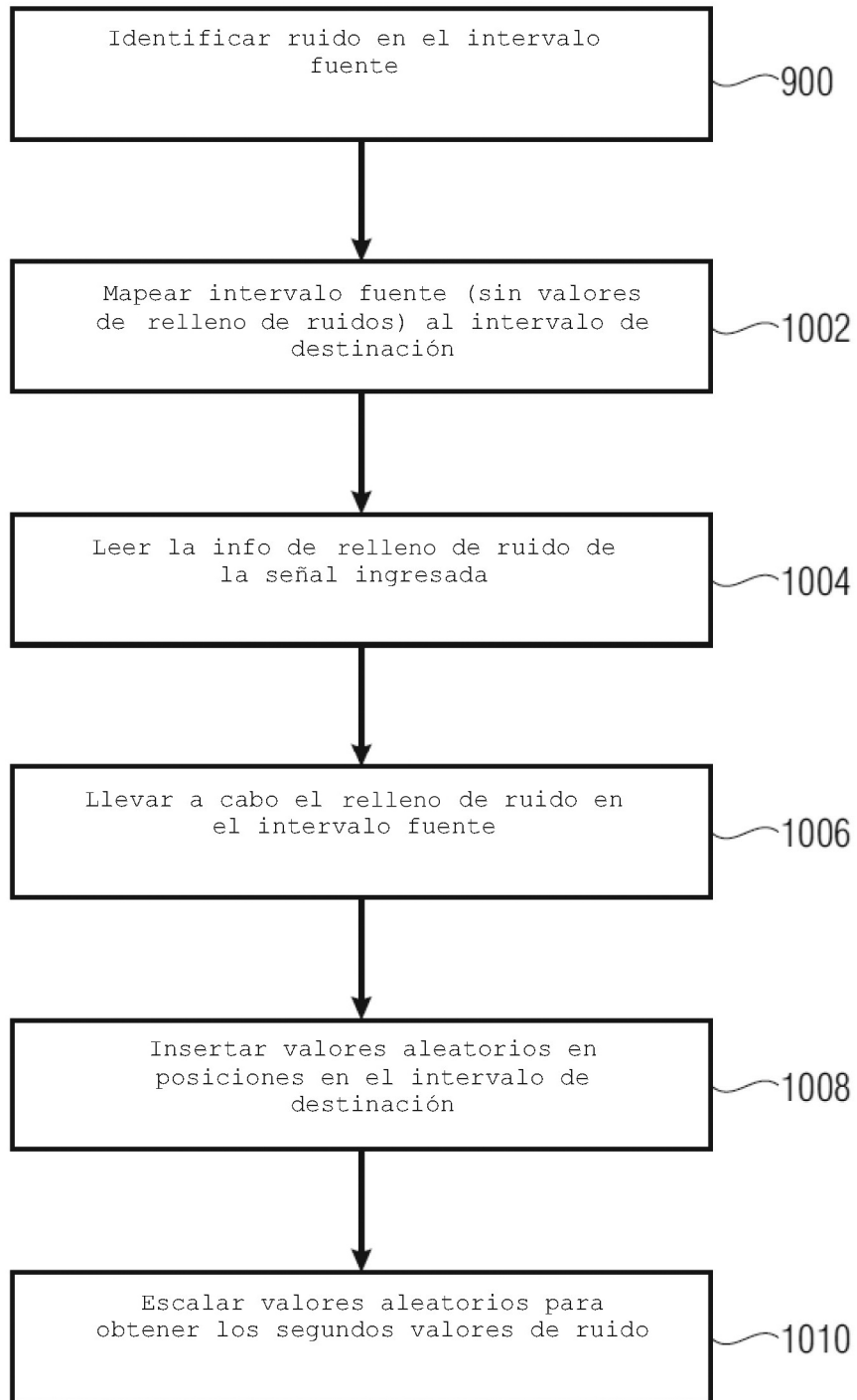


FIG 10

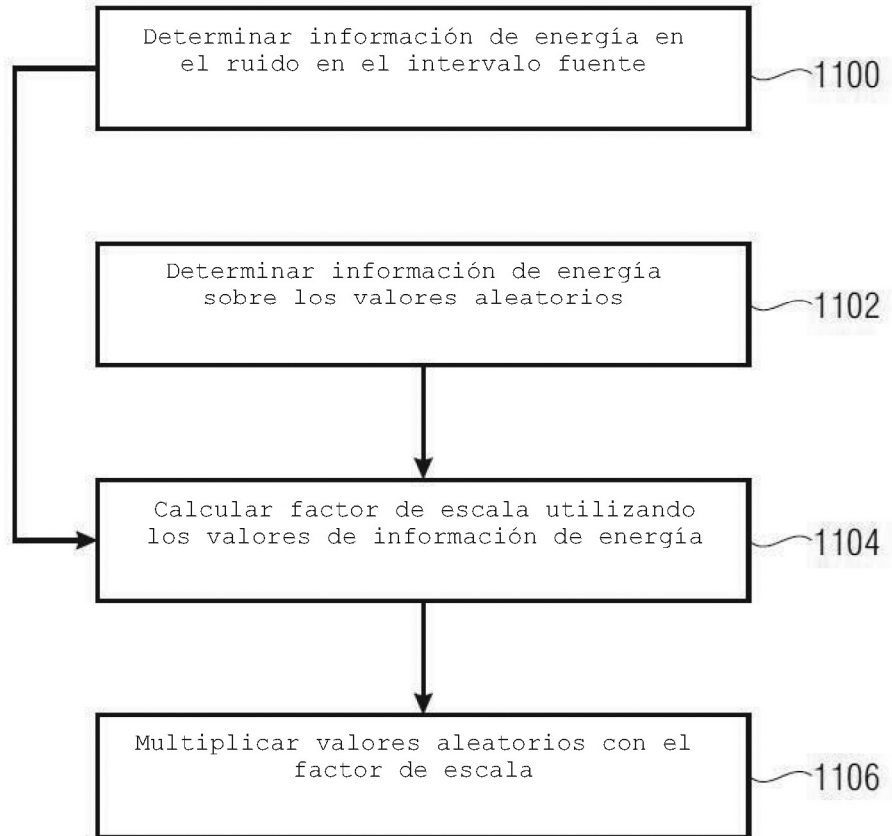


FIG 11

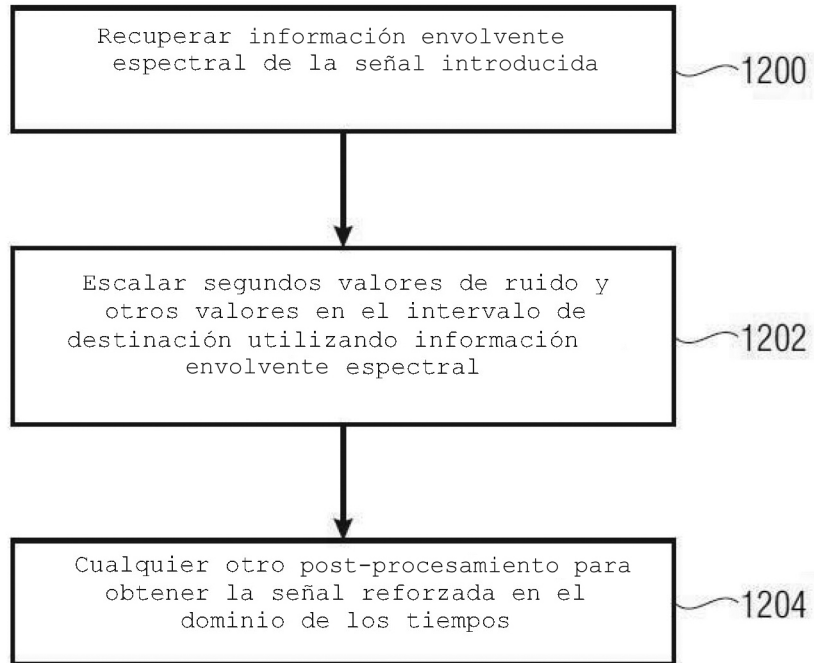


FIG 12

