

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 983 192**

51 Int. Cl.:

G10L 19/032 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **30.12.2008** E 12195829 (2)

97 Fecha y número de publicación de la concesión europea: **26.06.2024** EP 2573765

54 Título: **Codificador y decodificador de audio**

30 Prioridad:

04.01.2008 SE 0800032
24.05.2008 US 55978 P
24.05.2008 EP 08009530

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
22.10.2024

73 Titular/es:

DOLBY INTERNATIONAL AB (100.0%)
77 Sir John Rogerson's Quay, Block C,
Grand Canal Docklands
Dublin, D02 VK60, IE

72 Inventor/es:

HEDELIN, PER HENRIK;
CARLSSON, PONTUS JAN;
SAMUELSSON, JONAS LEIF y
SCHUG, MICHAEL

74 Agente/Representante:

LINAGE GONZÁLEZ, Rafael

ES 2 983 192 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

Codificador y decodificador de audio

5 **Campo técnico**

La presente invención se refiere a la codificación de señales de audio y, en particular, a la codificación de cualquier señal de audio no limitada a la voz, música o una combinación de ellas.

10 **Antecedentes de la invención**

En la técnica anterior, existen codificadores de voz específicamente diseñados para codificar señales de voz basando la codificación en un modelo de fuente de la señal, es decir, el sistema vocal humano. Dichos codificadores no pueden manejar señales de audio arbitrarias, tales como música, o cualquier otra señal diferente de la voz. Además, existen, en la técnica anterior, codificadores de música, a los que comúnmente se hace referencia como codificadores de audio, que basan su codificación en suposiciones del sistema auditivo humano, y no en el modelo de fuente de la señal. Dichos codificadores pueden manejar señales arbitrarias muy bien, a pesar de que, a tasas de bits bajas para las señales de voz, el codificador de voz dedicado proporciona una calidad de audio superior. Por lo tanto, no existe una estructura de codificación general hasta el momento para la codificación de señales de audio arbitrarias que tenga un rendimiento tan bueno como un codificador de voz para la voz y tan bueno como un codificador de música para la música, cuando se hace funcionar a tasas de bits bajas.

Sean A Ramprashad: "The Multimode Transform Predictive Coding Paradigm", IEEE Transactions on Speech and Audio Processing, vol. 11, nº 2, 1 de marzo de 2003, describe un esquema de codificación predictiva de transformada multimodal. Sean A Ramprashad: "High Quality Embedded Wideband Speech Coding Using an Inherently Layered Coding Paradigm", 2000 IEEE International Conference on Acoustics, Estambul, Turquía, 5-9 de junio de 2000, páginas 1145-1148, describe una estrategia de codificación para señales de banda ancha. El documento US 6.826.526 B1 describe un método de codificación de señales de audio.

De este modo, existe la necesidad de un codificador y decodificador de audio potenciados con calidad de audio mejorada y/o tasas de bits reducidas.

35 **Sumario de la invención**

La presente invención está definida por las reivindicaciones independientes. En las reivindicaciones dependientes se describen ejemplos preferidos.

La presente invención se refiere a codificar eficientemente señales de audio arbitrarias con un nivel de calidad igual o mejor que el de un sistema específicamente individualizado para una señal específica.

La presente invención se dirige a algoritmos de códec de audio que contienen tanto una codificación de predicción lineal (CPL) como una parte de codificador de transformada.

La presente invención se refiere además a codificar eficientemente factores de escala en la parte de codificación por transformada de un codificador de audio mediante la explotación de la presencia de datos CPL.

La presente invención se refiere además a un codificador para codificar señales de audio y generar un tren de bits, y a un decodificador para decodificar el tren de bits y generar una señal de audio reconstruida que es perceptualmente indistinguible de la señal de audio de entrada.

Un primer aspecto de la presente invención se refiere a un sistema de codificación de audio de acuerdo con la reivindicación 1.

Otro aspecto de la invención se refiere a un decodificador de audio de acuerdo con la reivindicación 7.

De este modo, la solicitud divulga adicionalmente un método de codificación de audio de acuerdo con la reivindicación 8 y un método de decodificación de audio de acuerdo con la reivindicación 9.

Estos son solo ejemplos de métodos de codificación/decodificación de audio y programas de ordenador preferidos que se enseñan por la presente solicitud y que una persona con experiencia en la técnica puede derivar de la siguiente descripción de realizaciones a modo de ejemplo.

65 **Breve descripción de los dibujos**

La presente invención se describirá ahora mediante ejemplos ilustrativos, sin limitar el alcance o espíritu de la

invención, con referencia a los dibujos que se acompañan, en los cuales:

- 5 la figura 1 ilustra una realización preferida de un codificador y un decodificador de acuerdo con la presente invención;
- la figura 2 ilustra una vista más detallada del codificador y decodificador no de acuerdo con la invención reivindicada;
- 10 la figura 3 ilustra otra realización del codificador no de acuerdo con la invención reivindicada;
- la figura 4 ilustra una realización preferida del codificador no de acuerdo con la invención reivindicada;
- la figura 5 ilustra una realización preferida del decodificador no de acuerdo con la invención reivindicada;
- 15 la figura 6 ilustra una realización preferida de la codificación y decodificación de líneas MDCT no de acuerdo con la invención reivindicada;
- la figura 7 ilustra una realización preferida del codificador y decodificador, y ejemplos de datos de control relevantes transmitidos entre sí, no de acuerdo con la invención reivindicada;
- 20 la figura 7a es otra ilustración de aspectos del codificador no de acuerdo con la invención reivindicada;
- la figura 8 ilustra un ejemplo de una secuencia de ventana y la relación entre datos CPL y datos MDCT de acuerdo con una realización de la presente invención;
- 25 la figura 9 ilustra una combinación de datos de factor de escala y datos CPL de acuerdo con la presente invención;
- la figura 9a ilustra otra realización de la combinación de datos de factor de escala y datos CPL no de acuerdo con la invención reivindicada;
- 30 la figura 9b ilustra otro diagrama de bloques simplificado de un codificador y un decodificador no de acuerdo con la invención reivindicada;
- 35 la figura 10 ilustra una realización preferida de traslación de polinomios CPL a una curva de ganancia MDCT de acuerdo con la presente invención;
- la figura 11 ilustra una realización preferida de mapeo de los parámetros CPL de tasa de actualización constante a los datos de secuencia de ventana MDCT adaptativos, de acuerdo con la presente invención;
- 40 la figura 12 ilustra una realización preferida para adaptar el cálculo de filtro de ponderación perceptual de acuerdo con el tamaño de transformada y tipo de cuantificador, de acuerdo con la presente invención;
- la figura 13 ilustra una realización preferida de adaptación del cuantificador que depende del tamaño de trama, no de acuerdo con la invención reivindicada;
- 45 la figura 14 ilustra una realización preferida de adaptación del cuantificador que depende del tamaño de trama, no de acuerdo con la invención reivindicada;
- 50 la figura 15 ilustra una realización preferida de adaptación del tamaño de escalón de cuantificación como una función de datos CPL y PLP, no de acuerdo con la invención reivindicada;
- la figura 15a ilustra cómo una curva delta se deriva de parámetros CPL y PLP mediante un módulo de adaptación delta;
- 55 la figura 16 ilustra una realización preferida de un cuantificador basado en modelo que utiliza desplazamientos aleatorios, no de acuerdo con la invención reivindicada;
- la figura 17 ilustra una realización preferida de un cuantificador basado en modelo no de acuerdo con la invención reivindicada;
- 60 la figura 17a ilustra otra realización preferida de un cuantificador basado en modelo no de acuerdo con la invención reivindicada;
- 65 la figura 17b ilustra, de forma esquemática, un decodificador 2150 de líneas MDCT basado en modelo no de acuerdo con la invención reivindicada;

la figura 17c ilustra, de forma esquemática, aspectos del preprocesamiento de cuantificador no de acuerdo con la invención reivindicada;

5 la figura 17d ilustra, de forma esquemática, aspectos del cómputo de tamaño de escalón no de acuerdo con la invención reivindicada;

la figura 17e ilustra, de forma esquemática, un codificador de entropía restringida basado en modelo no de acuerdo con la invención reivindicada;

10 la figura 17f ilustra, de forma esquemática, el funcionamiento de un cuantificador escalar uniforme (USQ, por sus siglas en inglés) no de acuerdo con la invención reivindicada;

15 la figura 17g ilustra, de forma esquemática, cálculos de probabilidad no de acuerdo con la invención reivindicada;

la figura 17h ilustra, de forma esquemática, un proceso de descuantificación no de acuerdo con la invención reivindicada;

20 la figura 18 ilustra una realización preferida de un control de depósito de bits, no de acuerdo con la invención reivindicada;

la figura 18a ilustra el concepto básico de un control de depósito de bits;

25 la figura 18b ilustra el concepto de un control de depósito de bits para tamaños de trama variables, no de acuerdo con la invención reivindicada;

la figura 18c muestra una curva de control a modo de ejemplo para el control de depósito de bits no de acuerdo con la invención reivindicada;

30 la figura 19 ilustra una realización preferida del cuantificador inverso mediante el uso de diferentes puntos de reconstrucción, no de acuerdo con la invención reivindicada.

35 Descripción de realizaciones preferidas

Las realizaciones descritas más abajo son meramente ilustrativas de los principios de la presente invención para el codificador y decodificador de audio. Se comprende que las modificaciones y variaciones de las disposiciones y los detalles descritos en la presente memoria serán aparentes para otras personas con experiencia en la técnica. Se pretende, por lo tanto, que se limite solamente por el alcance de las reivindicaciones de patente 40 anexas y no por los detalles específicos presentados a modo de descripción y explicación de las realizaciones en la presente memoria. Los componentes similares de las realizaciones se numeran mediante números de referencia similares.

45 En la figura 1 se visualizan un codificador 101 y un decodificador 102. El codificador 101 toma la señal de entrada de dominio tiempo y produce un tren de bits 103 posteriormente enviado al decodificador 102. El decodificador 102 produce una forma de onda de salida basada en el tren de bits 103 recibido. La señal de salida parece, psicoacústicamente, la señal de entrada original.

50 En la figura 2 se ilustra una realización preferida no de acuerdo con la invención reivindicada del codificador 200 y decodificador 210. La señal de entrada en el codificador 200 atraviesa un módulo CPL (Codificación de Predicción Lineal) 201 que genera una señal residual blanqueada para una trama CPL que tiene una primera longitud de trama, y los parámetros de predicción lineal correspondientes. Además, la normalización de la ganancia puede incluirse en el módulo CPL 201. La señal residual de la CPL se transforma en el dominio frecuencia por un módulo MDCT (Transformada Discreta del Coseno Modificada) 202 que funciona en una 55 segunda longitud de trama variable. En el codificador 200 ilustrado en la figura 2, se incluye un módulo PLP (Predicción a Largo Plazo) 205. PLP se elaborará en un ejemplo adicional. Las líneas MDCT se cuantifican 203 y también descuantifican 204 con el fin de alimentar a una memoria intermedia PLP con una copia de la salida decodificada que estará disponible para el decodificador 210. Debido a la distorsión de cuantificación, dicha copia se llama reconstrucción de la respectiva señal de entrada. En la parte inferior de la figura 2 se ilustra el 60 decodificador 210. El decodificador 210 toma las líneas MDCT cuantificadas, las descuantifica 211, añade la contribución del módulo PLP 214, y realiza una transformada MDCT inversa 212, seguida de un filtro de síntesis CPL 213.

65 Un aspecto importante de la realización de más arriba es que la trama MDCT es la única unidad básica para la codificación, aunque la CPL tiene su propio tamaño de trama (y en una realización, constante) y parámetros CPL también se codifican. La realización comienza desde un codificador de transformada e introduce módulos de

predicción y forma fundamentales de un codificador de voz. Como se discutirá más adelante, el tamaño de trama MDCT es variable y se adapta a un bloque de la señal de entrada mediante la determinación de la secuencia de ventana MDCT óptima para todo el bloque mediante la minimización de una función de coste de entropía perceptual simplista. Ello permite el escalamiento para mantener un control tiempo/frecuencia óptimo. Además, la estructura unificada propuesta evita combinaciones conmutadas o en capas de diferentes paradigmas de codificación.

En la figura 3, partes del codificador 300 se describen, de forma esquemática, en mayor detalle. La señal blanqueada como salida del módulo CPL 201 en el codificador de la figura 2 se ingresa en el banco de filtros MDCT 302. El análisis MDCT puede ser, de manera opcional, un análisis MDCT de tiempo deformado que asegura que la altura de tono de la señal (si la señal es periódica con una altura de tono bien definida) sea constante en la ventana de transformada MDCT.

En la figura 3, el módulo PLP 310 se esboza en mayor detalle. Este comprende una memoria intermedia PLP 311 que tiene muestras de dominio tiempo reconstruidas de los segmentos de señal de salida previos. Un extractor PLP 312 encuentra el segmento que mejor concuerda en la memoria intermedia PLP 311 dado el segmento de entrada actual. Un valor de ganancia apropiado se aplica a dicho segmento por la unidad de ganancia 313 antes de restarlo del segmento que actualmente se ingresa en el cuantificador 303. De manera evidente, con el fin de llevar a cabo la resta antes de la cuantificación, el extractor PLP 312 también transforma el segmento de señal elegido en el dominio MDCT. El extractor PLP 312 busca la mejor ganancia y valores de retardo que minimizan una función de error en el dominio perceptual cuando se combina el segmento de señal de salida previo reconstruido con la trama de entrada de dominio MDCT transformada. Por ejemplo, una función de error cuadrático medio (MSE, por sus siglas en inglés) entre el segmento reconstruido transformado del módulo PLP 310 y la trama de entrada transformada (es decir, la señal residual después de la resta) se optimiza. Dicha optimización puede llevarse a cabo en un dominio perceptual donde los componentes de frecuencia (es decir, líneas MDCT) se ponderan de acuerdo con su importancia perceptual. El módulo PLP 310 funciona en unidades de trama MDCT y el codificador 300 considera una trama MDCT residual por vez, por ejemplo para la cuantificación en el módulo de cuantificación 303. La búsqueda de retardo y ganancia puede llevarse a cabo en un dominio perceptual. De manera opcional, la PLP puede ser selectiva de frecuencia, es decir, adaptar la ganancia y/o el retardo en la frecuencia. Se ilustran una unidad de cuantificación inversa 304 y una unidad MDCT inversa 306. La MDCT puede ser de tiempo deformado como se explica más adelante.

En la figura 4 se ilustra otra realización del codificador 400. Además de la figura 3, el análisis CPL 401 se incluye en aras de la clarificación. Se muestra una transformada DCT-IV 414 usada para transformar un segmento de señal seleccionado en el dominio MDCT. Además, se ilustran varias maneras de calcular el error mínimo para la selección de segmento PLP. Además de la minimización de la señal residual como se muestra en la figura 4 (identificada como PLP2 en la figura 4), se ilustra la minimización de la diferencia entre la señal de entrada transformada y la señal de dominio MDCT descuantificada antes de transformarse inversamente en una señal de dominio tiempo reconstruida para el almacenamiento en la memoria intermedia PLP 411 (indicada como PLP3). La minimización de dicha función MSE dirigirá la contribución PLP hacia una similitud óptima (como sea posible) de la señal de entrada transformada y señal de entrada reconstruida para el almacenamiento en la memoria intermedia PLP 411. Otra función de error alternativa (indicada como PLP1) se basa en la diferencia de dichas señales en el dominio tiempo. En el presente caso, el MSE entre la trama de entrada filtrada CPL y la correspondiente reconstrucción de dominio tiempo en la memoria intermedia PLP 411 se minimiza. El MSE se calcula, de forma ventajosa, en base al tamaño de trama MDCT, que puede ser diferente del tamaño de trama CPL. Además, los bloques de cuantificador y descuantificador se reemplazan por el bloque de codificación de espectro 403 y los bloques de decodificación de espectro 404 ("Cod espec" y "Decod espec") que pueden contener módulos adicionales aparte de la cuantificación como se esbozará en la figura 6. Nuevamente, la MDCT y MDCT inversa pueden ser de tiempo deformado (WMDCT, IWMDCT, por sus siglas en inglés).

En la figura 5 se ilustra un decodificador 500 propuesto. Los datos de espectro del tren de bits recibido se cuantifican inversamente 511 y se añaden con una contribución PLP provista por un extractor PLP de una memoria intermedia PLP 515. El extractor PLP 516 y la unidad de ganancia PLP 517 en el decodificador 500 también se ilustran. Las líneas MDCT sumadas se sintetizan en el dominio tiempo por un bloque de síntesis MDCT, y se da forma espectralmente a la señal de dominio tiempo por un filtro de síntesis CPL 513.

En la figura 6, los bloques "Cod espec" y "Decod espec" 403, 404 de la figura 4 se describen en mayor detalle. El bloque "Cod espec" 603 ilustrado a la derecha en la figura comprende en una realización un módulo de análisis de Predicción Armónica 610, un módulo de análisis TNS (Forma de Ruido Temporal) 611, seguido de un módulo de escalamiento de factor de escala 612 de las líneas MDCT, y finalmente la cuantificación y codificación de las líneas en un módulo de líneas Cod 613. El bloque "Decod espec" 604 de decodificador ilustrado a la izquierda en la figura lleva a cabo el proceso inverso, es decir, las líneas MDCT recibidas se descuantifican en un módulo de líneas Decod 620 y el escalamiento se deshace por un módulo de escalamiento de factor de escala (SCF) 621. La síntesis TNS 622 y síntesis de predicción Armónica 623 se aplican.

En la figura 7 se esboza una ilustración muy general del sistema de codificación. El codificador a modo de

ejemplo toma la señal de entrada y produce un tren de bits que contiene, entre otros datos:

- líneas MDCT cuantificadas;
- 5 • factores de escala;
- representación polinomial CPL;
- energía de segmento de señal (p.ej., varianza de señal);
- 10 • secuencia de ventana;
- datos PLP.

15 El decodificador de acuerdo con la realización lee el tren de bits provisto y produce una señal de salida de audio que parece, psicoacústicamente, la señal original.

La figura 7a es otra ilustración de aspectos de un codificador 700 de acuerdo con una realización no de acuerdo con la invención reivindicada.

20 El codificador 700 comprende un módulo CPL 701, un módulo MDCT 704, un módulo PLP 705 (se muestra solo simplificado), un módulo de cuantificación 703 y un módulo de cuantificación inversa 704 para realimentar señales reconstruidas al módulo PLP 705. Además se proporcionan un módulo de estimación de altura de tono 750, para estimar la altura de tono de la señal de entrada, y un módulo de determinación de secuencia de
 25 ventana 751, para determinar la secuencia de ventana MDCT óptima para un bloque más grande de la señal de entrada (p.ej., 1 segundo). En la presente realización, la secuencia de ventana MDCT se determina en base a un enfoque de bucle abierto donde se determina la secuencia de candidatos de tamaño de ventana MDCT que minimiza una función de coste de codificación, p.ej., una entropía perceptual simplista. La contribución del módulo PLP 705 a la función de coste de codificación que se minimiza por el módulo de determinación de
 30 secuencia de ventana 751 puede considerarse, de forma opcional, cuando se busca la secuencia de ventana MDCT óptima. Preferiblemente, para cada candidato de tamaño de ventana evaluado, se determina la mejor contribución de predicción a largo plazo a la trama MDCT correspondiente al candidato de tamaño de ventana, y se estima el respectivo coste de codificación. En general, los tamaños de trama MDCT cortos son más apropiados para la entrada de voz mientras que las ventanas de transformada largas que tienen una buena
 35 resolución espectral se prefieren para las señales de audio.

Las ponderaciones perceptuales o una función de ponderación perceptual se determinan en base a los parámetros CPL como se calculan por el módulo CPL 701, que se explicará en mayor detalle más abajo. Las ponderaciones perceptuales se suministran al módulo PLP 705 y al módulo de cuantificación 703, ambos
 40 funcionando en el dominio MDCT, para la ponderación de errores o contribuciones de distorsión de componentes de frecuencia de acuerdo con su respectiva importancia perceptual. La figura 7a además ilustra qué parámetros de codificación se transmiten al decodificador, preferiblemente por un esquema de codificación apropiado como se discutirá más adelante. Seguidamente, se discutirán la coexistencia de datos CPL y MDCT y la emulación del efecto de la CPL en la MDCT, tanto para la acción contraria y omisión de filtrado real.

45 De acuerdo con una realización, el módulo PL filtra la señal de entrada de modo que la forma espectral de la señal se elimina, y la salida posterior del módulo PL es una señal espectralmente plana. Ello es ventajoso para el funcionamiento de, p.ej., la PLP. Sin embargo, otras partes del códec que funciona en la señal espectralmente plana pueden beneficiarse de saber qué forma espectral tenía la señal original antes del filtrado PL. Dado que los
 50 módulos de codificador, después del filtrado, funcionan en la transformada MDCT de la señal espectralmente plana, la forma espectral de la señal original antes del filtrado PL puede, si fuera necesario, volver a imponerse a la representación MDCT de la señal espectralmente plana mediante el mapeo de la función de transferencia del filtro PL usado (es decir, la envolvente espectral de la señal original) a una curva de ganancia, o curva de
 55 ecualización, que se aplica en los comportamientos de frecuencia de la representación MDCT de la señal espectralmente plana. En el sentido opuesto, de acuerdo con la invención reivindicada, el módulo PL omite el filtrado real, y solo estima una función de transferencia que se mapea posteriormente a una curva de ganancia que se impone a la representación MDCT de la señal, eliminando de este modo la necesidad de filtrado de dominio tiempo de la señal de entrada.

60 Un aspecto prominente de las realizaciones de la presente invención es que un codificador de transformada basado en MDCT se hace funcionar mediante el uso de una segmentación de ventana flexible, en una señal blanqueada CPL. Esto se esboza en la figura 8, donde se proporciona una secuencia de ventana MDCT a modo de ejemplo, junto con la ventana de la CPL. Por lo tanto, como es claro a partir de la figura, la CPL funciona en un tamaño de trama constante (p.ej., 20 ms), mientras la MDCT funciona en una secuencia de ventana variable
 65 (p.ej., 4 a 128 ms). Ello permite elegir la longitud de ventana óptima para la CPL y la secuencia de ventana óptima para la MDCT de forma independiente. La figura 8 además ilustra la relación entre datos CPL, en

particular los parámetros CPL, generados a una primera tasa de tramas y datos MDCT, en particular las líneas MDCT, generados a una segunda tasa variable. Las flechas hacia abajo en la figura simbolizan datos CPL que se interpolan entre las tramas CPL (círculos) para coincidir con las tramas MDCT correspondientes. Por ejemplo, una función de ponderación perceptual generada por CPL se interpola para instancias de tiempo como se determina por la secuencia de ventana MDCT.

Las flechas hacia arriba simbolizan datos de refinamiento (es decir, datos de control) usados para la codificación de líneas MDCT. Para las tramas AAC, dichos datos son, normalmente, factores de escala, y para las tramas ECQ los datos son, normalmente, datos de corrección de varianza, etc. Las líneas continuas vs. punteadas representan qué datos son los datos más "importantes" para la codificación de líneas MDCT dado cierto cuantificador. Las flechas dobles hacia abajo simbolizan las líneas espectrales de códec.

La coexistencia de datos CPL y MDCT en el codificador puede explotarse, por ejemplo, para reducir los requisitos de bits de codificación de factores de escala MDCT teniendo en cuenta una curva de enmascaramiento perceptual estimada a partir de los parámetros CPL. Además, la ponderación perceptual derivada de CPL puede usarse cuando se determina la distorsión de cuantificación. Como se ilustra y como se discutirá más abajo, el cuantificador funciona en dos modos y genera dos tipos de tramas (tramas ECQ y tramas AAC) dependiendo del tamaño de trama de los datos recibidos, es decir, correspondiente a la trama MDCT o tamaño de ventana.

La figura 11 ilustra una realización preferida de mapeo de los parámetros CPL de tasa constante a los datos de secuencia de ventana MDCT adaptativos. Un módulo de mapeo CPL 1100 recibe los parámetros CPL de acuerdo con la tasa de actualización CPL. Además, el módulo de mapeo CPL 1100 recibe información sobre la secuencia de ventana MDCT. Este entonces genera un mapeo CPL a MDCT, p.ej., para mapear datos psicoacústicos basados en CPL a las respectivas tramas MDCT generadas a la tasa de tramas MDCT variable. Por ejemplo, el módulo de mapeo CPL interpola los polinomios CPL o datos relacionados para instancias de tiempo correspondientes a tramas MDCT para el uso, p.ej., como ponderaciones perceptuales en el módulo PLP o cuantificador. Ahora, detalles específicos del modelo perceptual basado en CPL se discuten mediante referencia a la figura 9. El módulo CPL 901 se adapta en una realización no de acuerdo con la invención reivindicada para producir una señal de salida blanca, mediante el uso de una predicción lineal de, p.ej., orden 16 para una señal de tasa de muestreo de 16 kHz. Por ejemplo, la salida del módulo CPL 201 en la figura 2 es la residual después de la estimación y el filtrado de parámetro CPL. El polinomio CPL estimado $A(z)$, como se visualiza esquemáticamente en la parte inferior izquierda de la figura 9, puede fluctuar por un factor de expansión de ancho de banda, y también inclinarse mediante la modificación del primer coeficiente de reflexión del polinomio CPL correspondiente. La fluctuación expande el ancho de banda de picos en la función de transferencia CPL mediante el movimiento de los polos del polinomio hacia dentro adentro del círculo de la unidad, dando como resultado de este modo picos más suaves. La inclinación permite hacer la función de transferencia CPL más plana con el fin de equilibrar la influencia de frecuencias inferiores y superiores. Dichas modificaciones luchan por generar una curva de enmascaramiento perceptual $A'(z)$ a partir de los parámetros CPL estimados que estarán disponibles en el lado de codificador y decodificador del sistema. Los detalles de la manipulación del polinomio CPL se presentan en la figura 12 más abajo.

La codificación MDCT que funciona en la residual de la CPL tiene, en una implementación de la invención, factores de escala para controlar la resolución del cuantificador o los tamaños de escalón de cuantificación (y, de este modo, el ruido introducido por cuantificación). Dichos factores de escala son estimados por un módulo de estimación de factor de escala 960 en la señal de entrada original. Por ejemplo, los factores de escala se derivan de una curva umbral de enmascaramiento perceptual estimada a partir de la señal original. En una realización, una transformada de frecuencia separada (que tiene, posiblemente, una resolución de frecuencia diferente) puede usarse para determinar la curva umbral de enmascaramiento, pero ello no siempre es necesario. De manera alternativa, la curva umbral de enmascaramiento se estima a partir de las líneas MDCT generadas por el módulo de transformación. La parte inferior derecha de la figura 9 ilustra, de modo esquemático, factores de escala generados por el módulo de estimación de factor de escala 960 para controlar la cuantificación de modo que el ruido de cuantificación introducido se limita a distorsiones inaudibles.

Si un filtro CPL se conecta aguas arriba del módulo de transformación MDCT, una señal blanqueada se transforma en el dominio MDCT. Dado que dicha señal tiene un espectro blanco, no es muy apropiada para derivar una curva de enmascaramiento perceptual de ella. De este modo, una curva de ganancia de equalización de dominio MDCT generada para compensar el blanqueamiento del espectro puede usarse cuando se estima la curva umbral de enmascaramiento y/o los factores de escala. Esto se debe a que los factores de escala necesitan estimarse en una señal que tenga propiedades de espectro absolutas de la señal original, con el fin de estimar perceptualmente, de forma correcta, el enmascaramiento. El cálculo de la curva de ganancia de equalización de dominio MDCT a partir del polinomio CPL se discutirá en mayor detalle con referencia a la figura 10 más abajo.

Una realización del esquema de estimación de factor de escala esbozado más arriba se esboza en la figura 9a. En la presente realización, la señal de entrada se ingresa en el módulo PL 901 que estima la envolvente

espectral de la señal de entrada descrita por $A(z)$, y produce dicho polinomio así como una versión filtrada de la señal de entrada. La señal de entrada se filtra con la inversa de $A(z)$ con el fin de obtener una señal espectralmente blanca como se usa posteriormente por otras partes del codificador. La señal filtrada $\hat{x}(n)$ se ingresa en una unidad de transformación MDCT 902, mientras que el polinomio $A(z)$ se ingresa en una unidad de cálculo de curva de ganancia MDCT 970 (como se esboza en la figura 14). La curva de ganancia estimada a partir del polinomio PL se aplica a los coeficientes o líneas MDCT con el fin de retener la envolvente espectral de la señal de entrada original antes de la estimación de factor de escala. Las líneas MDCT ajustadas por ganancia se ingresan en el módulo de estimación de factor de escala 960 que estima los factores de escala para la señal de entrada.

Mediante el uso del enfoque esbozado más arriba, los datos transmitidos entre el codificador y decodificador contienen tanto el polinomio PL del cual la información perceptual relevante y un modelo de señal pueden derivarse cuando se usa un cuantificador basado en modelo, como los factores de escala comúnmente usados en un códec de transformada.

En mayor detalle, volviendo a la figura 9, el módulo CPL 901 en la figura estima a partir de la señal de entrada una envolvente espectral $A(z)$ de la señal y deriva de ella una representación perceptual $A'(z)$. Además, los factores de escala como se usan normalmente en códecs de audio perceptuales basados en transformada se estiman en la señal de entrada, o pueden estimarse en la señal blanca producida por un filtro PL, si la función de transferencia del filtro PL se tiene en cuenta en la estimación de factor de escala (como se describe en el contexto de la figura 10 más abajo). Los factores de escala pueden entonces adaptarse en el módulo de adaptación de factor de escala 961 dado el polinomio PL, como se esbozará más abajo, con el fin de reducir la tasa de bits requerida para transmitir factores de escala.

Normalmente, los factores de escala se transmiten al decodificador, y lo mismo ocurre con el polinomio PL. Ahora, dado que ambos se estiman a partir de la señal de entrada original y que se correlacionan, de alguna manera, con las propiedades de espectro absolutas de la señal de entrada original, se propone codificar una representación delta entre los dos, con el fin de eliminar cualquier redundancia que pueda ocurrir si ambos se transmiten de forma separada. De acuerdo con una realización, dicha correlación se explota de la siguiente manera. Dado que el polinomio CPL, cuando fluctúa y se inclina correctamente, lucha por representar una curva umbral de enmascaramiento, las dos representaciones pueden combinarse de modo que los factores de escala transmitidos del codificador de transformada representan la diferencia entre los factores de escala deseados y aquellos que pueden derivarse del polinomio CPL transmitido. El módulo de adaptación de factor de escala 961 que se muestra en la figura 9 calcula, por lo tanto, la diferencia entre los factores de escala deseados generados a partir de la señal de entrada original y los factores de escala derivados de CPL. El presente aspecto retiene la capacidad de tener un cuantificador basado en MDCT que tiene la noción de factores de escala como se usa comúnmente en codificadores de transformada, dentro de una estructura CPL, que funciona en una residual de la CPL, y aún tiene la posibilidad de conmutar a un cuantificador basado en modelo que deriva los tamaños de escalón de cuantificación solamente de los datos de predicción lineal.

En la figura 9b, se da un diagrama de bloques simplificado de codificador y decodificador de acuerdo con una realización. La señal de entrada en el codificador atraviesa el módulo CPL 901 que genera una señal residual blanqueada y los correspondientes parámetros de predicción lineal. Además, la normalización de la ganancia puede incluirse en el módulo CPL 901. La señal residual de la CPL se transforma en el dominio frecuencia por una transformada MDCT 902. En la parte derecha de la figura 9b se representa el decodificador. El decodificador toma las líneas MDCT cuantificadas, las descuantifica 911 y aplica una transformada MDCT inversa 912, seguida de un filtro de síntesis CPL 913.

La señal blanqueada como salida del módulo CPL 901 en el codificador de la figura 9b se ingresa en el banco de filtros MDCT 902. Las líneas MDCT como resultado del análisis MDCT se codifican mediante transformada con un algoritmo de codificación por transformada que consiste en un modelo perceptual que guía el tamaño de escalón de cuantificación deseado para diferentes partes del espectro MDCT. Los valores que determinan el tamaño de escalón de cuantificación se llaman factores de escala y hay un valor de factor de escala que se necesita para cada partición, llamada banda de factor de escala, del espectro MDCT. En los algoritmos de codificación por transformada de la técnica anterior, los factores de escala se transmiten mediante el tren de bits al decodificador.

De acuerdo con un aspecto de la invención, la curva de enmascaramiento perceptual estimada a partir de los parámetros CPL, como se explica con referencia a la figura 9, se usa cuando se codifican los factores de escala usados en la cuantificación. Otra posibilidad para estimar una curva de enmascaramiento perceptual es usar los coeficientes de filtro CPL no modificados para una estimación de la distribución de la energía en las líneas MDCT. Con esta estimación de energía, un modelo psicoacústico, como se usa en esquemas de codificación por transformada, puede aplicarse tanto en el codificador como en el decodificador para obtener una estimación de una curva de enmascaramiento.

Las dos representaciones de una curva de enmascaramiento se combinan entonces de modo que los factores de

escala que se transmitirán del codificador de transformada representan la diferencia entre los factores de escala deseados y aquellos que pueden derivarse del polinomio CPL transmitido o del modelo psicoacústico basado en CPL. La presente característica retiene la capacidad de tener un cuantificador basado en MDCT que tiene la noción de factores de escala como se usa comúnmente en codificadores de transformada, dentro de una estructura CPL, que funciona en una residual de la CPL, y aún tiene la posibilidad de controlar el ruido de cuantificación por cada banda de factor de escala, de acuerdo con el modelo psicoacústico del codificador de transformada. La ventaja es que la transmisión de la diferencia de los factores de escala costará menos bits en comparación con la transmisión de los valores de factores de escala absolutos sin tener en cuenta los datos CPL ya presentes. Dependiendo de la tasa de bits, el tamaño de trama u otros parámetros, la cantidad de residual de factor de escala que se transmitirá puede seleccionarse. Para tener control total de cada banda de factor de escala, un delta de factor de escala puede transmitirse con un esquema de codificación sin ruido apropiado. En otros casos, el coste de transmisión de factores de escala puede reducirse más por una representación más tosca de las diferencias de factores de escala. El caso especial con la sobrecarga más baja ocurre cuando la diferencia de factor de escala se establece en 0 para todas las bandas y no se transmite ninguna información adicional.

La figura 10 ilustra una realización preferida de traslación de polinomios CPL hacia una curva de ganancia MDCT. Como se esboza en la figura 2, la MDCT funciona en una señal blanqueada, blanqueada por el filtro CPL 1001. Con el fin de retener la envolvente espectral de la señal de entrada original, una curva de ganancia MDCT se calcula por el módulo de curva de ganancia MDCT 1070. La curva de ganancia de equalización de dominio MDCT puede obtenerse mediante la estimación de la respuesta de magnitud de la envolvente espectral descrita por el filtro CPL, para las frecuencias representadas por los comportamientos en la transformada MDCT. La curva de ganancia puede entonces aplicarse en los datos MDCT, p.ej., cuando se calcula la señal de error cuadrático medio mínimo como se esboza en la figura 3, o cuando se estima una curva de enmascaramiento perceptual para la determinación de factor de escala como se esboza con referencia a la figura 9 más arriba.

La figura 12 ilustra una realización preferida de adaptación del cálculo de filtro de ponderación perceptual basándose en tipo y/o tamaño de transformada de cuantificador. El polinomio $PLA(z)$ es estimado por el módulo CPL 1201 en la figura 16. Un módulo de modificación de parámetro CPL 1271 recibe parámetros CPL tales como el polinomio $CPL A(z)$, y genera un filtro de ponderación perceptual $A'(z)$ mediante la modificación de los parámetros CPL. Por ejemplo, el ancho de banda del polinomio $CPL A(z)$ se expande y/o el polinomio se inclina. Los parámetros de entrada para el módulo de fluctuación e inclinación 1272 adaptado son los valores de fluctuación e inclinación por defecto ρ y γ . Estos se modifican dadas las normas predeterminadas, basándose en el tamaño de transformada usado y/o la estrategia de cuantificación Q usada. Los parámetros de fluctuación e inclinación modificados ρ' y γ' se ingresan en el módulo de modificación de parámetros CPL 1271 mediante la traslación de la envolvente espectral de la señal de entrada, representada por $A(z)$, a una curva de enmascaramiento perceptual representada por $A'(z)$.

A continuación, se explicarán la estrategia de cuantificación condicionada al tamaño de trama, y la cuantificación basada en modelo condicionada a parámetros variados no de acuerdo con la invención reivindicada. Un aspecto no de acuerdo con la invención reivindicada es que utiliza diferentes estrategias de cuantificación para diferentes tamaños de transformada o tamaños de trama. Esto se ilustra en la figura 13, donde el tamaño de trama se usa como un parámetro de selección para usar un cuantificador basado en modelo o un cuantificador no basado en modelo. Se debe apreciar que este aspecto de cuantificación es independiente de otros aspectos del codificador/decodificador descrito y también puede aplicarse en otros códecs. Un ejemplo de un cuantificador no basado en modelo es el cuantificador basado en la tabla de Huffman usado en el estándar de codificación de audio AAC. El cuantificador basado en modelo puede ser un Cuantificador de Restricción de Entropía (ECQ) que emplea la codificación aritmética. Sin embargo, otros cuantificadores también pueden usarse en realizaciones de la presente invención.

De acuerdo con un aspecto independiente, se sugiere conmutar entre diferentes estrategias de cuantificación como función de tamaño de trama con el fin de poder usar la estrategia de cuantificación óptima dado un tamaño de trama particular. A modo de ejemplo, la secuencia de ventana puede dictar el uso de una transformada larga para un segmento de música tonal estacionario de la señal. Para dicho tipo de señal particular, mediante el uso de una transformada larga, es altamente beneficioso emplear una estrategia de cuantificación que pueda aprovechar el carácter "disperso" (es decir, tonos discretos bien definidos) en el espectro de la señal. Un método de cuantificación como se usa en AAC en combinación con tablas de Huffman y el agrupamiento de líneas espectrales, también como se usa en AAC, es muy beneficioso. Sin embargo, y por el contrario, para segmentos de voz, la secuencia de ventana puede, dada la ganancia de codificación de la PLP, ordenar el uso de transformadas cortas. Para dicho tipo de señal y tamaño de transformada, es beneficioso emplear una estrategia de cuantificación que no intente descubrir o introducir dispersión en el espectro, pero que, en cambio, mantenga una energía de banda ancha que, dada la PLP, retenga el carácter tipo pulso de la señal de entrada original.

Una visualización más general del presente concepto se da en la figura 14, donde la señal de entrada se transforma en el dominio MDCT, y posteriormente se cuantifica por un cuantificador controlado por el tamaño de transformada o tamaño de trama usado para la transformada MDCT.

De acuerdo con otro aspecto, el tamaño de escalón de cuantificador se adapta como función de datos CPL y/o PLP. Ello permite una determinación del tamaño de escalón dependiendo de la dificultad de una trama y controla el número de bits que se asignan para codificar la trama. En la figura 15, se proporciona una ilustración sobre cómo la cuantificación basada en modelo puede controlarse por datos CPL y PLP. En la parte superior de la figura 15, se proporciona una visualización esquemática de líneas MDCT. Más abajo, se representa la delta de tamaño de escalón de cuantificación Δ como una función de frecuencia. Está claro a partir del presente ejemplo particular que el tamaño de escalón de cuantificación aumenta con la frecuencia, es decir, se incurre en una mayor distorsión de cuantificación para frecuencias más altas. La curva delta se deriva de los parámetros CPL y PLP por medio de un módulo de adaptación de delta representado en la figura 15a. La curva delta puede además derivarse del polinomio de predicción $A(z)$ mediante la fluctuación y/o inclinación como se explica con referencia a la figura 13.

Una función de ponderación perceptual preferida derivada de los datos CPL se proporciona en la siguiente ecuación:

$$P(z) = \frac{1 - (1 - \tau)r_1 z^{-1}}{A(z/\rho)}$$

donde $A(z)$ es el polinomio CPL, τ es un parámetro de inclinación, ρ controla la fluctuación y r_1 es el primer coeficiente de reflexión calculado a partir del polinomio $A(z)$. Se ha de apreciar que el polinomio $A(z)$ puede volver a calcularse para una diversidad de representaciones diferentes con el fin de extraer información relevante del polinomio. Si uno está interesado en la pendiente espectral con el fin de aplicar una "inclinación" para contrarrestar la pendiente del espectro, se prefiere el recálculo del polinomio para los coeficientes de reflexión, dado que el primer coeficiente de reflexión representa la pendiente del espectro.

Además, los valores delta Δ pueden adaptarse como una función de la varianza de señal de entrada σ , la ganancia PLP g , y el primer coeficiente de reflexión r_1 derivado del polinomio de predicción. Por ejemplo, la adaptación puede basarse en la siguiente ecuación:

$$\Delta' = \Delta(1 + r_1(1 - g^2))$$

A continuación, se esbozan aspectos de un cuantificador basado en modelo no de acuerdo con la invención reivindicada. En la figura 16, se visualiza uno de los aspectos del cuantificador basado en modelo. Las líneas MDCT se ingresan en un cuantificador que emplea cuantificadores escalares uniformes. Además, los desplazamientos aleatorios se ingresan en el cuantificador, y se usan como valores de desplazamiento para los intervalos de cuantificación que trasladan los límites del intervalo. El cuantificador propuesto proporciona ventajas de cuantificación de vector mientras mantiene la capacidad de búsqueda de cuantificadores escalares. El cuantificador se repite en un conjunto de diferentes valores de desplazamiento, y calcula el error de cuantificación para estos. El valor de desplazamiento (o vector de valor de desplazamiento) que minimiza la distorsión de cuantificación para las líneas MDCT particulares que se cuantifican se usa para la cuantificación. El valor de desplazamiento se transmite entonces al decodificador junto con las líneas MDCT cuantificadas. El uso de desplazamientos aleatorios introduce un relleno de ruido en la señal decodificada descuantificada y, mediante ello, evita los agujeros espectrales en el espectro cuantificado. Esto es particularmente importante para tasas de bits bajas donde muchas líneas MDCT se cuantifican, de otra forma, en un valor cero que llevará a agujeros audibles en el espectro de la señal reconstruida.

La figura 17 ilustra, de forma esquemática, un Cuantificador de Líneas MDCT basado en Modelo (MBMLQ, por sus siglas en inglés) no de acuerdo con la invención reivindicada. La parte superior de la figura 17 ilustra un codificador MBMLQ 1700. El codificador MBMLQ 1700 toma como entrada las líneas MDCT en una trama MDCT o las líneas MDCT de la residual de la PLP si una PLP está presente en el sistema. El MBMLQ emplea modelos estadísticos de las líneas MDCT, y códigos de fuente se adaptan a las propiedades de señal trama por trama de la MDCT y producen una compresión eficaz para un tren de bits.

Una ganancia local de las líneas MDCT puede estimarse como el valor RMS de las líneas MDCT, y las líneas MDCT pueden normalizarse en el módulo de normalización de ganancia 1720 antes de la entrada en el codificador MBMLQ 1700. La ganancia local normaliza las líneas MDCT y es un complemento para la normalización de ganancia PL. Mientras la ganancia PL se adapta a variaciones en el nivel de señal en una escala de tiempo mayor, la ganancia local se adapta a variaciones en una escala de tiempo menor, y produce una calidad mejorada de sonidos transitorios e inicios en la voz. La ganancia local se codifica mediante codificación de tasa fija o de tasa variable y se transmite al decodificador.

Un módulo de control de tasa 1710 puede emplearse para controlar el número de bits usados para codificar una

trama MDCT. Un índice de control de tasa controla el número de bits usados. El índice de control de tasa señala una lista de tamaños de escalón de cuantificador nominales. La tabla puede ordenarse con tamaños de escalón en orden descendente (véase la figura 17g).

5 El codificador MBMLQ se ejecuta con un conjunto de diferentes índices de control de tasa, y el índice de control de tasa que produce un recuento de bits que es menor que el número de bits otorgados provistos por el control de depósito de bits, se usa para la trama. El índice de control de tasa varía lentamente y ello puede explotarse para reducir la complejidad de búsqueda y para codificar el índice de forma eficaz. El conjunto de índices que se prueban puede reducirse si la prueba se inicia alrededor del índice de la trama MDCT previa. Asimismo, la
10 codificación de entropía eficaz del índice se obtiene si las probabilidades alcanzan un pico de alrededor del valor previo del índice. Por ejemplo, para una lista de 32 tamaños de escalón, el índice de control de tasa puede codificarse mediante el uso de 2 bits por trama MDCT en promedio.

15 La figura 17 además ilustra, de forma esquemática, el decodificador MBMLQ 1750 donde la trama MDCT tiene la ganancia renormalizada si en el codificador 1700 se estimó una ganancia local.

La figura 17a ilustra, de forma esquemática, el codificador de líneas MDCT basado en modelo 1700 de acuerdo con una realización en mayor detalle. La figura comprende un módulo de preprocesamiento de cuantificador 1730 (véase la figura 17c), un codificador de entropía restringida basado en modelo 1740 (véase la figura 17e), y
20 un codificador aritmético 1720 que puede ser un codificador aritmético de la técnica anterior. La tarea del módulo de preprocesamiento de cuantificador 1730 es adaptar el codificador MBMLQ a las estadísticas de señal, trama a trama de MDCT. Este toma como entrada otros parámetros de códec y deriva de ellos estadísticas útiles sobre la señal que pueden usarse para modificar el comportamiento del codificador de entropía restringida basado en modelo 1740. El codificador de entropía restringida basado en modelo 1740 se controla, p.ej., por un conjunto de
25 parámetros de control: un tamaño de escalón de cuantificador Δ (delta, longitud de intervalo), un conjunto de estimaciones de varianza de las líneas MDCT V (un vector; un valor estimado por línea MDCT), una curva de enmascaramiento perceptual P_{mod} , una matriz o tabla de desplazamientos (aleatorios), y un modelo estadístico de las líneas MDCT que describen la forma de la distribución de las líneas MDCT y sus interdependencias. Todos los parámetros de control mencionados más arriba pueden variar entre las tramas MDCT.

30 La figura 17b ilustra, de forma esquemática, un decodificador de líneas MDCT basado en modelo 1750 no de acuerdo con la invención reivindicada. Este toma como entrada bits de información conexa del tren de bits y los decodifica en parámetros que se ingresan en el módulo de preprocesamiento de cuantificador 1760 (véase la figura 17c). El módulo de preprocesamiento de cuantificador 1760 tiene, preferiblemente, exactamente la misma funcionalidad en el codificador 1700 que en el decodificador 1750. Los parámetros que se ingresan en el módulo de preprocesamiento de cuantificador 1760 son exactamente iguales en el codificador que en el decodificador. El
35 módulo de preprocesamiento de cuantificador 1760 produce un conjunto de parámetros de control (igual que en el codificador 1700) y estos se ingresan en el módulo de cálculos de probabilidad 1770 (véase la figura 17g; igual que en el codificador, véase la figura 17e) y en el módulo de descuantificación 1780 (véase la figura 17h; igual que en el codificador, véase la figura 17e). Las tablas cdf del módulo de cálculos de probabilidad 1770 que representan las funciones de densidad de probabilidad para todas las líneas MDCT dada la delta usada para la cuantificación y la varianza de la señal, se ingresan en el decodificador aritmético (que puede ser cualquier
40 codificador aritmético conocido por las personas con experiencia en la técnica) que luego decodifica los bits de líneas MDCT en índices de líneas MDCT. Los índices de líneas MDCT se descuantifican entonces en líneas MDCT por el módulo de descuantificación 1780.

45 La figura 17c ilustra, de manera esquemática, aspectos del preprocesamiento de cuantificador no de acuerdo con la invención reivindicada que consiste en i) cómputo de tamaño de escalón, ii) modificación de curva de enmascaramiento perceptual, iii) estimación de varianza de líneas MDCT, iv) construcción de tabla de desplazamiento.

50 El cómputo de tamaño de escalón se explica en mayor detalle en la figura 17d. Esta comprende i) una tabla de consulta donde los puntos de índice de control de tasa en una tabla de tamaños de escalón producen un tamaño de escalón nominal Δ_{nom} (delta_nom), ii) adaptación de energía baja, y iii) adaptación de paso alto.

55 La normalización de la ganancia normalmente resulta en que los sonidos de energía alta y los sonidos de energía baja se codifican con la misma SNR segmental. Ello puede llevar a un número excesivo de bits usados en los sonidos de energía baja. La adaptación de energía baja propuesta permite la buena sintonización de un compromiso entre sonidos de energía baja y sonidos de energía alta. El tamaño de escalón puede aumentarse cuando la energía de señal se convierte en baja como se representa en la figura 17d-ii) donde se muestra una
60 curva a modo de ejemplo para la relación entre la energía de señal (ganancia g) y un factor de control q_{Le} . La ganancia de señal g puede computarse como el valor RMS de la propia señal de entrada o de la residual de la PL. La curva de control en la figura 17d-ii) solo es un ejemplo y pueden emplearse otras funciones de control para aumentar el tamaño de escalón para señales de energía baja. En el ejemplo ilustrado, la función de control se determina por secciones lineales escalonadas que se definen por umbrales T_1 y T_2 y el factor de tamaño de
65 escalón L .

Los sonidos paso alto son perceptualmente menos importantes que los sonidos paso bajo. La función de adaptación de paso alto aumenta el tamaño de escalón cuando la trama MDCT es de paso alto, es decir, cuando la energía de la señal en la presente trama MDCT se concentra en las frecuencias más altas, lo cual resulta en menos bits empleados en dichas tramas. Si PLP está presente y si la ganancia PLP g_{PLP} es cercana a 1, la residual de la PLP puede convertirse en un paso alto; en dicho caso, es ventajoso no aumentar el tamaño de escalón. El presente mecanismo se representa en la figura 17d-iii) donde r es el 1^{er} coeficiente de reflexión de CPL. La adaptación de paso alto propuesta puede usar la siguiente ecuación:

$$q_{hp} = \begin{cases} 1+r(1-g^2) & \text{si } r > 0 \\ 1 & \text{si } r \leq 0 \end{cases}$$

La figura 17c-ii) ilustra, de manera esquemática, la modificación de curva de enmascaramiento perceptual que emplea un aumento de frecuencia baja (LF, por sus siglas en inglés) para eliminar artefactos de codificación "tipo ronroneo". El aumento LF puede fijarse o ser adaptativo de modo que solo se aumenta una parte por debajo del primer pico espectral. El aumento LF puede adaptarse mediante el uso de los datos de envolvente CPL.

La figura 17c-iii) ilustra, de manera esquemática, la estimación de varianza de líneas MDCT. Con un filtro de blanqueamiento CPL activo, las líneas MDCT tienen, todas, varianza de unidad (de acuerdo con la envolvente CPL). Después de la ponderación perceptual en el codificador de entropía restringida basado en modelo 1740 (véase la figura 17e), las líneas MDCT tienen varianzas que son la inversa de la curva de enmascaramiento perceptual cuadrática, o la curva de enmascaramiento modificada cuadrática P_{mod} . Si una PLP está presente, esta puede reducir la varianza de las líneas MDCT. En la figura 17c-iii) se representa un mecanismo que adapta las varianzas estimadas a la PLP. La figura muestra una función de modificación q_{PLP} en la frecuencia f . Las varianzas modificadas pueden determinarse por $V_{PLPmod} = V \cdot q_{PLP}$. El valor L_{PLP} puede ser una función de la ganancia PLP de modo que L_{PLP} es más cercana a 0 si la ganancia PLP es de alrededor de 1 (lo cual indica que la PLP ha encontrado una buena concordancia), y L_{PLP} es más cercana a 1 si la ganancia PLP es de alrededor de 0. La adaptación PLP propuesta de las varianzas $V = \{v_1, v_2, \dots, v_j, \dots, v_N\}$ solo afecta a las líneas MDCT por debajo de cierta frecuencia ($f_{PLPcorte}$). Como resultado, las varianzas de líneas MDCT por debajo de la frecuencia de corte $f_{PLPcorte}$ se reducen, la reducción dependiendo de la ganancia PLP.

La figura 17c-iv) ilustra, de manera esquemática, la construcción de la tabla de desplazamiento. La tabla de desplazamiento nominal es una matriz rellena de números pseudoaleatorios distribuidos entre -0,5 y 0,5. El número de columnas en la matriz es igual al número de líneas MDCT que se codifican por el MBMLQ. El número de filas es ajustable y es igual al número de vectores de desplazamiento que se prueban en la optimización RD en el codificador de entropía restringida basado en modelo 1740 (véase la figura 17e). La función de construcción de tabla de desplazamiento escala la tabla de desplazamiento nominal con el tamaño de escalón de cuantificador de modo que los desplazamientos se distribuyen entre $-\Delta/2$ y $+\Delta/2$.

La figura 17g ilustra, de manera esquemática, una realización para una tabla de desplazamiento. El índice de desplazamiento es un puntero en la tabla y selecciona un vector de desplazamiento elegido $O = \{o_1, o_2, \dots, o_n, \dots, o_N\}$, donde N es el número de líneas MDCT en la trama MDCT.

Como se describirá más abajo, los desplazamientos proporcionan un medio para el llenado de ruido. Una mejor calidad perceptual y objetiva se obtiene si la dispersión de los desplazamientos se limita para las líneas MDCT que tienen baja varianza v_j en comparación con el tamaño de escalón de cuantificador Δ . Un ejemplo de dicha limitación se describe en la figura 17c-iv) donde k_1 y k_2 son parámetros de sintonización. La distribución de los desplazamientos puede ser uniforme y distribuirse entre $-s$ y $+s$. Los límites s pueden determinarse de acuerdo con

$$s = \begin{cases} k_2 \sqrt{v_j} & \text{si } \sqrt{v_j} < k_1 \Delta \\ \frac{\Delta}{2} & \text{de lo contrario} \end{cases}$$

Para líneas MDCT de varianza baja (donde v_j es pequeño en comparación con Δ) puede ser ventajoso hacer que la distribución de desplazamiento no sea uniforme y sea dependiente de la señal.

La figura 17e ilustra, de manera esquemática, el codificador de entropía restringida basado en modelo 1740 en mayor detalle. Las líneas MDCT de entrada se ponderan perceptualmente dividiéndolas con los valores de la curva de enmascaramiento perceptual, preferiblemente derivados del polinomio CPL, lo cual resulta en el vector de líneas MDCT ponderadas $y = (y_1, \dots, y_N)$. El objetivo de la siguiente codificación es introducir ruido de cuantificación blanco a las líneas MDCT en el dominio perceptual. En el decodificador, se aplica la inversa de la

ponderación perceptual, lo cual resulta en ruido de cuantificación que sigue a la curva de enmascaramiento perceptual.

5 Primero, se esboza la iteración en los desplazamientos aleatorios. Las siguientes operaciones se llevan a cabo para cada fila j en la matriz de desplazamiento: Cada línea MDCT se cuantifica por un cuantificador escalar uniforme (USQ) de desplazamiento, en donde cada cuantificador se desplaza por su propio valor de desplazamiento único tomado del vector de fila de desplazamiento.

10 La probabilidad de intervalo de distorsión mínima de cada USQ se computa en el módulo de cómputos de probabilidad 1770 (véase la figura 17g). Los índices USQ se codifican por entropía. El coste en términos del número de bits requeridos para codificar los índices se computa como se muestra en la figura 17e y produce una longitud de palabra de código teórica R_j . El borde de sobrecarga del USQ de la línea MDCT j puede computarse como $k_3 \cdot \sqrt{v_j}$ donde k_3 puede elegirse para que sea cualquier número apropiado, p.ej., 20. El borde de sobrecarga es el límite para el cual el error de cuantificación es mayor que la mitad del tamaño de escalón de cuantificación en magnitud.

15 Un valor de reconstrucción escalar para cada línea MDCT se computa por el módulo de descuantificación 1780 (véase la figura 17h) que produce el vector MDCT cuantificado \bar{y} . En el módulo de optimización RD 1790 se computa una distorsión $D_j = d(y, \bar{y})$. $d(y, \bar{y})$ puede ser el error cuadrático medio (MSE), u otra medida de distorsión perceptualmente más relevante, p.ej., basándose en una función de ponderación perceptual. En particular, una medida de distorsión que pondera juntos MSE y la discordancia en energía entre y e \bar{y} puede ser útil.

20 En el módulo de optimización RD 1790, se computa un coste C , preferiblemente basándose en la distorsión D_j y/o la longitud de palabra de código teórica R_j para cada fila j en la matriz de desplazamiento. Un ejemplo de una función de coste es $C = 10 \cdot \log_{10}(D_j) + \lambda \cdot R_j/N$. El desplazamiento que minimiza C se elige y los índices USQ correspondientes y las probabilidades se producen desde el codificador de entropía restringida basado en modelo 1780.

25 La optimización RD puede mejorarse más, de forma opcional, mediante la variación de otras propiedades del cuantificador junto con el desplazamiento. Por ejemplo, en lugar de usar la misma estimación de varianza fija V para cada vector de desplazamiento que se prueba en la optimización RD, el vector de estimación de varianza V se puede variar. Para el vector de fila de desplazamiento m , puede entonces usarse una estimación de varianza $k_m \cdot V$ donde k_m puede abarcar, por ejemplo, el rango de 0,5 a 1,5 mientras m varía desde $m=1$ a $m=(\text{número de filas en la matriz de desplazamiento})$. Ello hace a la codificación de entropía y al cómputo de MMSE menos sensibles a las variaciones en las estadísticas de señal de entrada que el modelo estadístico no puede capturar. Ello resulta en un coste C más bajo en general.

30 Las líneas MDCT descuantificadas pueden refinarse más mediante el uso de un cuantificador residual como se representa en la figura 17e. El cuantificador residual puede ser, p.ej., un cuantificador de vector aleatorio de tasa fija.

35 El funcionamiento del Cuantificador Escalar Uniforme (USQ) para la cuantificación de línea MDCT n se ilustra, de manera esquemática, en la figura 17f que muestra el valor de línea MDCT n en el intervalo de distorsión mínima que tiene el índice i_n . Las marcas 'x' indican el centro (punto medio) de los intervalos de cuantificación con tamaño de escalón Δ . El origen del cuantificador escalar es trasladado el desplazamiento o_n del vector de desplazamiento $O = \{o_1, o_2, \dots, o_n, \dots, o_N\}$. De este modo, los límites de intervalo y los puntos medios son trasladados por el desplazamiento.

40 El uso de desplazamientos introduce un relleno de ruido controlado por el codificador en la señal cuantificada y, mediante ello, evita los agujeros espectrales en el espectro cuantificado. Además, los desplazamientos aumentan la eficacia de codificación mediante la provisión de un conjunto de alternativas de codificación que llenan el espacio de manera más eficaz que una retícula cúbica. Asimismo, los desplazamientos proporcionan variación en las tablas de probabilidad que se computan por el módulo de cómputos de probabilidad 1770, lo cual lleva a una codificación de entropía más eficaz de los índices de líneas MDCT (es decir, se requieren menos bits).

45 El uso de un tamaño de escalón variable Δ (delta) permite la exactitud variable en la cuantificación de modo que más exactitud puede usarse para sonidos perceptualmente importantes, y menos exactitud puede usarse para sonidos perceptualmente menos importantes.

50 La figura 17g ilustra, de manera esquemática, los cómputos de probabilidad en el módulo de cómputos de probabilidad 1770. Las entradas en dicho módulo son el modelo estadístico aplicado para las líneas MDCT, el tamaño de escalón de cuantificador Δ , el vector de varianza V , el índice de desplazamientos, y la tabla de desplazamientos. La salida del módulo de cómputos de probabilidad 1770 son las tablas cdf. Para cada línea MDCT x_j se evalúa el modelo estadístico (es decir, una función de densidad de probabilidad, pdf, por sus siglas en inglés). El área debajo de la función pdf para un intervalo i es la probabilidad $p_{i,j}$ del intervalo. Dicha

probabilidad se usa para la codificación aritmética de las líneas MDCT.

La figura 17h ilustra, de manera esquemática, el proceso de descuantificación como se realiza, p.ej., en el módulo de descuantificación 1780. El centro de la masa (valor MMSE) x_{MMSE} para el intervalo de distorsión mínima de cada línea MDCT se computa junto con el punto medio x_{PM} del intervalo. Teniendo en cuenta que un vector N-dimensional de las líneas MDCT se cuantifica, el valor MMSE escalar es subóptimo y, en general, demasiado bajo. Ello resulta en una pérdida de varianza y desequilibrio espectral en la salida decodificada. Este problema puede mitigarse mediante decodificación de preservación de varianza como se describe en la figura 17h donde el valor de reconstrucción se computa como una suma ponderada del valor MMSE y el valor de punto medio. Una mejora opcional adicional es adaptar la ponderación de modo que el valor MMSE domina la voz y el punto medio domina los sonidos diferentes de la voz. Ello produce voz más limpia mientras el equilibrio espectral y la energía se preservan para sonidos diferentes de la voz.

La decodificación de preservación de varianza no de acuerdo con la invención reivindicada se consigue mediante la determinación del punto de reconstrucción de acuerdo con la siguiente ecuación:

$$x_{descuant} = (1 - \chi)x_{MMSE} + \chi x_{PM}$$

La decodificación adaptativa para preservar la varianza puede basarse en la siguiente regla para determinar el factor de interpolación:

$$\chi = \begin{cases} 0 & \text{si son sonidos de voz} \\ 1 & \text{si son sonidos diferentes de la voz} \end{cases}$$

La ponderación adaptativa puede además ser una función de, por ejemplo, la ganancia de predicción PLP g_{PLP} : $\chi = f(g_{PLP})$. La ponderación adaptativa varía lentamente y puede codificarse de manera eficaz por un código de entropía recursivo.

El modelo estadístico de las líneas MDCT que se usa en los cálculos de probabilidad (figura 17g) y en la descuantificación (figura 17h) debe reflejar las estadísticas de la señal real. En una versión, el modelo estadístico supone que las líneas MDCT son independientes y distribuidas de manera Laplaciana. Otra versión modela las líneas MDCT como Gaussianas independientes. Una versión modela las líneas MDCT como modelos de mezcla Gaussianos, incluidas las interdependencias entre líneas MDCT dentro y entre tramas MDCT. Otra versión adapta el modelo estadístico a estadísticas de señal en línea. Los modelos estadísticos adaptativos pueden adaptarse hacia delante y/o hacia atrás.

Otro aspecto relacionado con los puntos de reconstrucción modificados del cuantificador se ilustra, de manera esquemática, en la figura 19 donde se representa un cuantificador inverso como se usa en el decodificador de una realización. El módulo tiene, aparte de las entradas normales de un cuantificador inverso, es decir, las líneas cuantificadas e información sobre el tamaño de escalón de cuantificación (tipo de cuantificación), también información sobre el punto de reconstrucción del cuantificador. El cuantificador inverso de esta realización puede usar múltiples tipos de puntos de reconstrucción cuando determina un valor reconstruido \bar{y}_n del correspondiente índice de cuantificación i_n . Como se menciona más arriba, valores de reconstrucción \bar{y} se usan además, p.ej., en el codificador de líneas MDCT (véase la figura 17) para determinar la residual de cuantificación para la entrada en el cuantificador residual. Además, la reconstrucción de cuantificación se lleva a cabo en el cuantificador inverso 304 para reconstruir una trama MDCT codificada para su uso en la memoria intermedia PLP (véase la figura 3) y, naturalmente, en el decodificador.

El cuantificador inverso puede, p.ej., elegir el punto medio de un intervalo de cuantificación como el punto de reconstrucción, o el punto de reconstrucción MMSE. En una realización no de acuerdo con la invención reivindicada, el punto de reconstrucción del cuantificador se elige para que sea el valor medio entre el punto central y el punto de reconstrucción MMSE. En general, el punto de reconstrucción puede interpolarse entre el punto medio y el punto de reconstrucción MMSE, p.ej., dependiendo de propiedades de la señal tales como la periodicidad de la señal. Información sobre la periodicidad de la señal puede derivarse del módulo PLP, por ejemplo. La presente característica permite al sistema controlar la distorsión y preservación de energía. El punto de reconstrucción central asegurará la preservación de energía, mientras que el punto de reconstrucción MMSE asegurará una distorsión mínima. Dada la señal, el sistema puede entonces adaptar el punto de reconstrucción a donde se proporciona el mejor compromiso.

La presente invención además incorpora un nuevo formato de codificación de secuencia de ventana. De acuerdo con una realización de la invención, las ventanas usadas para la transformación MDCT son de tamaños diádicos, y solo pueden variar un factor dos en tamaño de ventana a ventana. Los tamaños de transformada diádicos son, p.ej., 64, 128, ..., 2048 muestras correspondientes a 4, 8, ..., 128 ms a 16 kHz de tasa de muestreo. En general,

se proponen ventanas de tamaño variable que pueden tomar múltiples tamaños de ventana entre un tamaño de ventana mínimo y un tamaño máximo. En una secuencia, los tamaños de ventana consecutivos pueden variar solamente por un factor de dos de modo que se desarrollan secuencias suaves de tamaños de ventana sin cambios abruptos. Las secuencias de ventana como se definen por una realización, es decir, limitadas a tamaños

5 diádicos y que solamente se permite que varíen un factor dos en tamaño de ventana a ventana, tienen varias ventajas. En primer lugar, no se necesitan ventanas específicas de inicio o fin, es decir, ventanas con bordes afilados. Ello mantiene una buena resolución tiempo/frecuencia. En segundo lugar, la secuencia de ventana se convierte en muy eficaz para codificar, es decir, para señalar a un decodificador qué secuencia de ventana particular se usa. Finalmente, la secuencia de ventana siempre encajará bien en una estructura de hipertrama.

10 La estructura de hipertrama es útil cuando se hace funcionar un codificador en un sistema del mundo real, donde ciertos parámetros de configuración de decodificador necesitan transmitirse con el fin de poder iniciar el decodificador. Dichos datos se almacenan comúnmente en un campo de encabezamiento en el tren de bits y describen la señal de audio codificada. Con el fin de minimizar la velocidad binaria, el encabezamiento no se transmite para cada trama de datos codificados, en particular en un sistema como el propuesto por la presente invención, donde los tamaños de trama MDCT pueden variar de muy corto a muy largo. Por lo tanto, se propone por la presente invención agrupar cierta cantidad de tramas MDCT juntas en una hipertrama, donde los datos de encabezamiento se transmiten en el inicio de la hipertrama. La hipertrama se define, normalmente, como una longitud específica en el tiempo. Por lo tanto, es preciso tener cuidado de modo que las variaciones de los

15 tamaños de trama MDCT encajen en una longitud constante, longitud de hipertrama predefinida. La secuencia de ventana inventiva esbozada más arriba asegura que la secuencia de ventana seleccionada siempre encaje en una estructura de hipertrama.

20 De acuerdo con una realización no de acuerdo con la invención reivindicada, el retardo PLP y la ganancia PLP se codifican de manera que la tasa sea variable. Ello es ventajoso dado que, debido a la eficacia PLP para señales periódicas estacionarias, el retardo PLP tiende a ser el mismo en segmentos de alguna forma largos. Por lo tanto, ello puede explotarse por medio de la codificación aritmética, lo cual resulta en una codificación de ganancia PLP y retardo PLP de tasa variable.

25 De manera similar, una realización no de acuerdo con la invención reivindicada se aprovecha de un depósito de bits y codificación de tasa variable también para la codificación de los parámetros PL. Además, la codificación PL recursiva se enseña por la presente invención.

30 Otro aspecto es el manejo de un depósito de bits para tamaños de trama variables en el codificador. En la figura 18 se esboza una unidad de control de depósito de bits 1800. Además de una medida de dificultad prevista como entrada, la unidad de control de depósito de bits también recibe información sobre la longitud de trama de la trama actual. Un ejemplo de medida de dificultad para su uso en la unidad de control de depósito de bits es la entropía perceptual, o el logaritmo del espectro de energía. El control de depósito de bits es importante en un sistema donde las longitudes de trama pueden variar en un conjunto de longitudes de trama diferentes. La

35 unidad de control de depósito de bits 1800 sugerida toma en cuenta la longitud de trama cuando calcula el número de bits otorgados para que la trama se codificará como se esbozará más abajo.

40 El depósito de bits se define aquí como cierta cantidad fija de bits en una memoria intermedia que tiene que ser más grande que el número promedio de bits que una trama puede usar para una tasa de bits dada. Si es del mismo tamaño, ninguna variación en el número de bits para una trama es posible. El control de depósito de bits siempre observa el nivel del depósito de bits antes de extraer bits que se otorgarán al algoritmo de codificación como número permitido de bits para la trama real. De este modo, un depósito de bits completo significa que el número de bits disponible en el depósito de bits es igual al tamaño del depósito de bits. Después de codificar la trama, el número de bits usados se restará de la memoria intermedia y el depósito de bits se actualiza mediante

45 la adición del número de bits que representan la tasa de bits constante. Por lo tanto, el depósito de bits está vacío, si el número de bits en el depósito de bits antes de codificar una trama es igual al número promedio de bits por trama.

50 En la figura 18a se representa el concepto básico del control de depósito de bits. El codificador proporciona medios para calcular lo difícil que es codificar la trama real en comparación con la trama previa. Para una dificultad promedio de 1,0, el número de bits otorgados depende del número de bits disponible en el depósito de bits. De acuerdo con una línea de control dada, más bits que los correspondientes a una tasa de bits promedio se extraerán del depósito de bits si el depósito de bits está bastante lleno. En el caso de un depósito de bits vacío, menos bits en comparación con los bits promedio se usarán para codificar la trama. Dicho comportamiento produce un nivel de depósito de bits promedio para una secuencia más larga de tramas con dificultad promedio.

55 Para tramas con una dificultad más alta, la línea de control puede ser trasladada hacia arriba, teniendo el efecto de que las tramas difíciles de codificar pueden usar más bits en el mismo nivel de depósito de bits. De acuerdo con ello, por facilidad para la codificación de tramas, el número de bits permitido para una trama será más bajo simplemente por trasladar hacia abajo la línea de control en la figura 18a desde el caso de dificultad promedio al caso de dificultad fácil. También son posibles otras modificaciones aparte de un simple traslado de la línea de control. Por ejemplo, como se muestra en la figura 18a, la pendiente de la curva de control puede cambiarse

60

65

dependiendo de la dificultad de la trama.

5 Cuando se calcula el número de bits otorgados, los límites del extremo inferior del depósito de bits tienen que obedecerse con el fin de no extraer más bits de la memoria intermedia que los permitidos. Un esquema de control de depósito de bits que incluye el cálculo de los bits otorgados por una línea de control como se muestra en la figura 18a es solo un ejemplo de las posibles relaciones del nivel de depósito de bits y medida de dificultad con los bits otorgados. Asimismo, otros algoritmos de control tendrán en común los límites duros en el extremo inferior del nivel de depósito de bits que previenen que un depósito de bits viole la restricción del depósito de bits vacío, así como los límites en el extremo superior, donde se forzará al codificador que escriba bits llenos, si un número de bits demasiado bajo se consumirá por el codificador.

15 Para que dicho mecanismo de control pueda manejar un conjunto de tamaños de trama variables, el presente algoritmo de control simple tiene que adaptarse. La medida de dificultad que se usará tiene que normalizarse de modo que los valores de dificultad de diferentes tamaños de trama sean comparables. Para cada tamaño de trama, habrá un rango diferente permitido para los bits otorgados, y dado que el número promedio de bits por trama es diferente para un tamaño de trama variable, como resultado, cada tamaño de trama tiene su propia ecuación de control con sus propias limitaciones. Un ejemplo se muestra en la figura 18b. Una modificación importante al caso de tamaño de trama fijo es el borde inferior permitido del algoritmo de control. En lugar del número promedio de bits para el tamaño de trama real, que corresponde al caso de tasa de bits fija, ahora el número promedio de bits para el tamaño de trama más grande permitido es el valor más bajo permitido para el nivel de depósito de bits antes de extraer los bits para la trama real. Esta es una de las principales diferencias con el control de depósito de bits para tamaños de trama fijos. Esta restricción garantiza que una trama siguiente con el tamaño de trama más grande posible pueda utilizar al menos el número promedio de bits para dicho tamaño de trama.

25 La medida de dificultad puede basarse, p.ej., en un cálculo de entropía perceptual (EP) que se deriva de umbrales de enmascaramiento de un modelo psicoacústico como se realiza en AAC, o como una alternativa del recuento de bits de una cuantificación con tamaño de escalón fijo como se realiza en la parte ECQ de un codificador no de acuerdo con la invención reivindicada. Estos valores pueden normalizarse con respecto a los tamaños de trama variables, lo cual se puede lograr por una simple división por la longitud de trama, y el resultado será una EP respectivamente, un recuento de bits por muestra. Puede tener lugar otro paso de normalización con respecto a la dificultad promedio. A tales fines, un promedio móvil de las tramas pasadas puede usarse, lo cual resulta en un valor de dificultad mayor que 1,0 para tramas difíciles o menor que 1,0 para tramas fáciles. En el caso de un codificador de dos pasadas o de un registro de anticipación grande, también los valores de dificultad de tramas futuras pueden tenerse en cuenta para dicha normalización de la medida de dificultad.

40 Otro aspecto se refiere a detalles específicos del manejo del depósito de bits para ECQ. La gestión del depósito de bits para ECQ trabaja suponiendo que ECQ produce una calidad aproximadamente constante cuando usa un tamaño de escalón de cuantificador constante para la codificación. El tamaño de escalón de cuantificador constante produce una tasa variable y el objetivo del depósito de bits es mantener la variación en el tamaño de escalón de cuantificador entre diferentes tramas lo más pequeño posible, mientras no se violan las restricciones de la memoria intermedia del depósito de bits. Además de la tasa producida por el ECQ, la información adicional (p.ej., ganancia y retardo PLP) se transmite trama a trama MDCT. La información adicional, en general, también se codifica con entropía y de este modo consume una tasa diferente de trama a trama.

En una realización no de acuerdo con la invención reivindicada, un control de depósito de bits propuesto intenta minimizar la variación del tamaño de escalón ECQ introduciendo tres variables (véase la figura 18c):

- 50 - R_{ECQ_AVG} : tasa media de ECQ por muestra usada previamente;
 - Δ_{ECQ_AVG} : tamaño medio de escalón de cuantificador usado previamente.

Estas variables se actualizan ambas dinámicamente para reflejar las últimas estadísticas de codificación.

- 55 - $R_{ECQ_AVG_DES}$: La tasa de ECQ correspondiente a la velocidad binaria total media.

60 Este valor se diferenciará de R_{ECQ_AVG} en caso de que el nivel de depósito de bits haya cambiado durante la trama de tiempo de la ventana de promediación, por ejemplo se ha usado durante esta trama de tiempo una velocidad binaria mayor o inferior que la velocidad binaria media especificada. También se actualiza a medida que la tasa de la información conexas cambia, de modo que la tasa total se iguale a la velocidad binaria especificada.

65 El control de depósito de bits usa estos tres valores para determinar una suposición inicial sobre la delta a usar para la trama actual. Lo hace hallando $\Delta_{ECQ_AVG_DES}$ en la curva $R_{ECQ-\Delta}$ mostrada en la figura 18c que corresponde a $R_{ECQ_AVG_DES}$. En una segunda etapa, este valor se modifica posiblemente si la tasa no está

acorde con las restricciones de depósito de bits. La curva de ejemplo $R_{ECQ-\Delta}$ en la figura 18c se basa en la siguiente ecuación:

$$R_{ECQ} = \frac{1}{2} \log_2 \frac{\alpha}{\Delta^2}$$

5

Por supuesto, también se pueden usar otras relaciones matemáticas entre R_{ECQ} y Δ .

En el caso estacionario, R_{ECQ_AVG} será cercano a $R_{ECQ_AVG_DES}$ y la variación en Δ será muy pequeña. En el caso no estacionario, la operación de promediación asegurará una variación suave de Δ .

10

Aunque lo que antecede se ha divulgado con referencia a realizaciones particulares de la presente invención, se ha de entender que el concepto inventivo no está limitado a las realizaciones descritas. Por otro lado, la divulgación presentada en esta solicitud posibilitará que un experto en la técnica entienda y lleve a cabo la invención. Se entenderá por parte de los expertos en la técnica que se pueden hacer diversas modificaciones sin salir del alcance de la invención como se establece exclusivamente mediante las reivindicaciones que se acompañan.

15

REIVINDICACIONES

1. Sistema de codificación de audio que comprende:

5 una unidad de predicción lineal (PL) (201) para analizar tramas de una primera longitud fija de una señal de audio para determinar una secuencia de polinomios CPL;

10 una unidad de transformación de longitud adaptativa (202) para transformar tramas de una segunda longitud variable de la señal de audio en tramas con coeficientes de Transformada Discreta del Coseno Modificada (MDCT);

15 una unidad de generación de curva de ganancia (970, 1070) para generar curvas de ganancia de dominio MDCT para las tramas de coeficientes MDCT basándose en respuestas de magnitud determinadas a partir de la secuencia de polinomios CPL, en la que la generación de curvas de ganancia de dominio MDCT comprende el mapeo, por una unidad de mapeo (1100), de los polinomios CPL a correspondientes tramas de coeficientes MDCT;

20 una unidad de aplicación de curva de ganancia para aplicar las curvas de ganancia de dominio MDCT a las tramas de coeficientes MDCT para generar tramas de coeficientes MDCT ajustados por ganancia; y

una unidad de cuantificación (203) para cuantificar las tramas de coeficientes MDCT ajustados por ganancia.

2. Sistema de codificación de audio de la reivindicación 1, que comprende:

25 una unidad de control de secuencia de ventana para determinar, para un bloque de la señal de audio, las segundas longitudes de trama para ventanas MDCT solapadas, mediante la minimización de una función de coste de codificación para un bloque completo de la señal de audio que incluye varias tramas.

30 3. Sistema de codificación de audio de acuerdo con cualquier reivindicación anterior, que comprende una unidad de modelado perceptual que modifica una característica de un filtro PL mediante fluctuación y/o inclinación del polinomio CPL generado por la unidad de predicción lineal para una trama CPL.

4. Sistema de codificación de audio de acuerdo con cualquier reivindicación anterior, que comprende:

35 una unidad de estimación de factor de escala (1360) para estimar factores de escala para controlar el ruido de cuantificación de la unidad de cuantificación (203), de modo que el ruido de cuantificación introducido está limitado a distorsiones inaudibles.

40 5. Sistema de codificación de audio de la reivindicación 4, en el que los factores de escala se determinan basándose en las curvas de ganancia de dominio MDCT.

45 6. Sistema de codificación de audio de acuerdo con cualquier reivindicación anterior, en el que la unidad de mapeo (1500) interpola polinomios PL generados a una tasa correspondiente a la primera longitud de trama como para coincidir con tramas de la señal de dominio MDCT generada a una tasa correspondiente a la segunda longitud de trama.

7. Decodificador de audio que comprende:

50 una unidad de descuantificación (211) para descuantificar tramas cuantificadas de coeficientes MDCT decodificados a partir de un tren de bits de entrada, en la que las tramas de coeficientes MDCT representan una señal de audio;

55 una unidad de generación de curva de ganancia (970, 1070) para generar curvas de ganancia de dominio MDCT para las tramas de coeficientes MDCT basándose en respuestas de magnitud determinadas a partir de polinomios CPL, en la que los polinomios CPL se derivan del tren de bits de entrada, y en el que

60 los polinomios CPL se han determinado mediante análisis de tramas de una primera longitud fija de la señal de audio, y en el que la generación de las curvas de ganancia de dominio MDCT comprende el mapeo, por una unidad de mapeo (1100), de los polinomios CPL a correspondientes tramas de coeficientes MDCT;

una unidad de aplicación de curva de ganancia para aplicar las curvas de ganancia de dominio MDCT a las tramas de coeficientes MDCT para generar tramas de coeficientes MDCT ajustados por ganancia; y

65 una unidad de transformación MDCT inversa de longitud adaptativa (212) para transformar inversamente las tramas de coeficientes MDCT ajustados por ganancia en una señal de audio de dominio tiempo, funcionando la unidad de transformación MDCT inversa en una segunda longitud de trama variable.

8. Método de codificación de audio que comprende los pasos:

- 5 realizar un análisis de predicción lineal (PL) en tramas de una primera longitud fija de una señal de audio para determinar una secuencia de polinomios CPL;
- transformar tramas de una segunda longitud variable de la señal de audio en tramas de coeficientes de Transformada Discreta del Coseno Modificada (MDCT);
- 10 generar curvas de ganancia de dominio MDCT para las tramas de coeficientes MDCT basándose en respuestas de magnitud determinadas a partir de la secuencia de polinomios CPL, en el que la generación de curvas de ganancia de dominio MDCT comprende el mapeo de los polinomios CPL a correspondientes tramas de coeficientes MDCT;
- 15 aplicar las curvas de ganancia de dominio MDCT a las tramas de coeficientes MDCT para obtener tramas de coeficientes MDCT ajustados por ganancia; y
- cuantificar las tramas de coeficientes MDCT ajustados por ganancia usando una unidad de cuantificación (203).

20 9. Método de decodificación de audio que comprende los pasos:

- descuantificar tramas cuantificadas de coeficientes MDCT decodificados a partir de un tren de bits de entrada, en el que las tramas de coeficientes MDCT representan una señal de audio;
- 25 generar curvas de ganancia de dominio MDCT para las tramas de coeficientes MDCT basándose en respuestas de magnitud determinadas a partir de polinomios CPL, en las que los polinomios CPL se derivan del tren de bits de entrada, y en las que los polinomios CPL se han determinado analizando tramas de una primera longitud fija de la señal de audio, y en las que la generación de las curvas de ganancia de dominio MDCT comprende el mapeo de los polinomios CPL a correspondientes tramas de coeficientes MDCT;
- 30 aplicar las curvas de ganancia de dominio MDCT a las tramas de coeficientes MDCT para generar tramas de coeficientes MDCT ajustados por ganancia;
- 35 transformar inversamente las tramas de coeficientes MDCT ajustados por ganancia en una señal de audio de dominio tiempo usando una MDCT inversa que funciona en una segunda longitud de trama variable.
10. Programa de ordenador que comprende instrucciones que, cuando el programa es ejecutado por un dispositivo programable, hace que el dispositivo programable realice un método de codificación o decodificación de audio de acuerdo con la reivindicación 8 o 9.

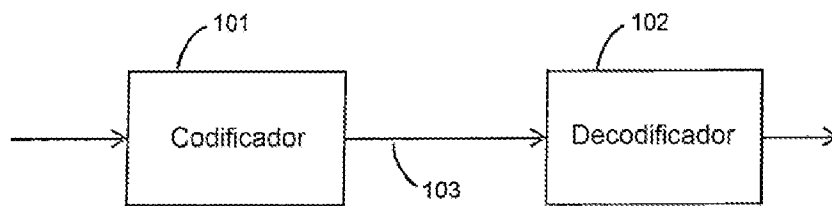


Fig. 1

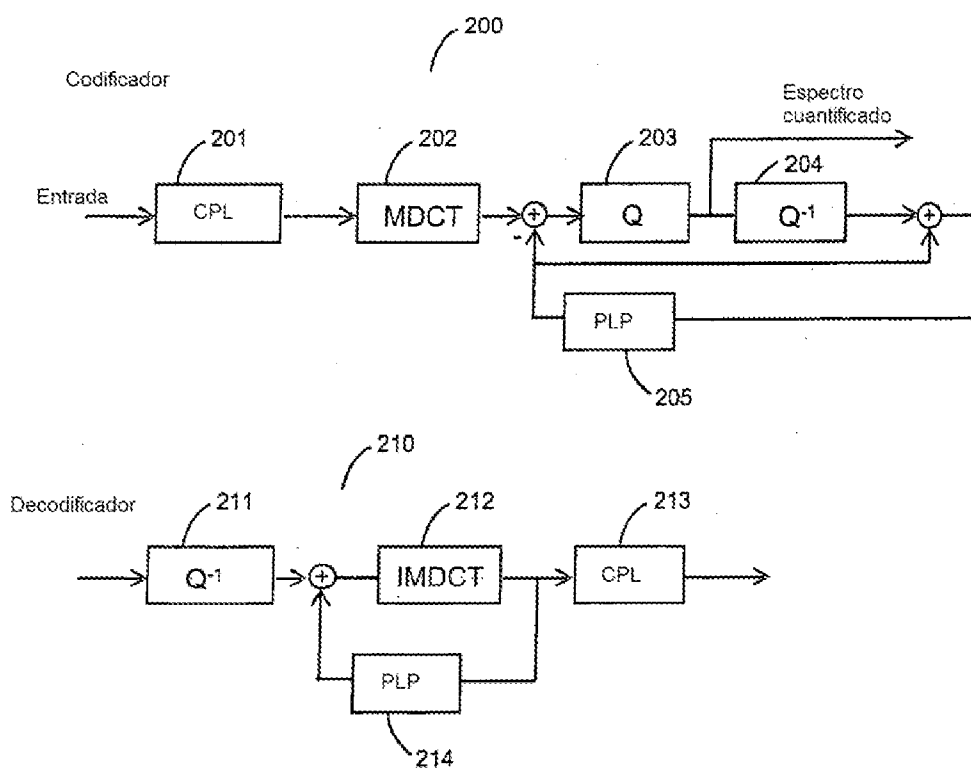


Fig. 2

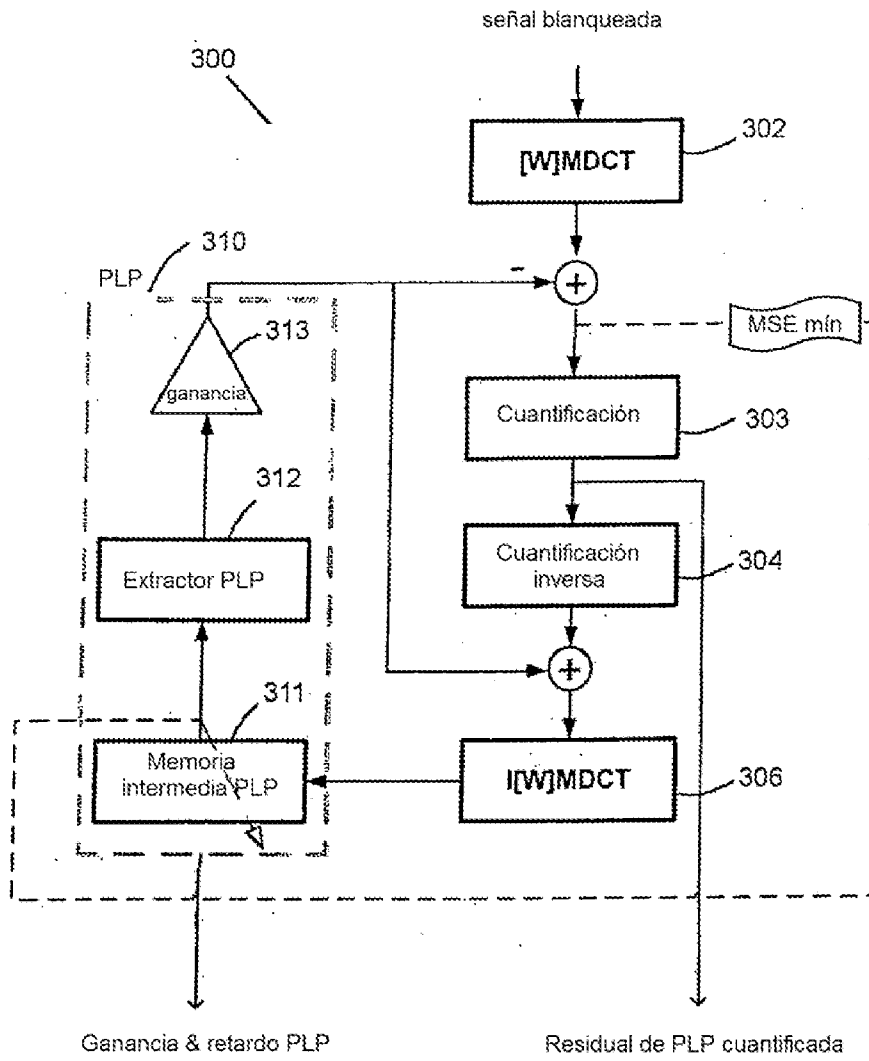


Fig. 3

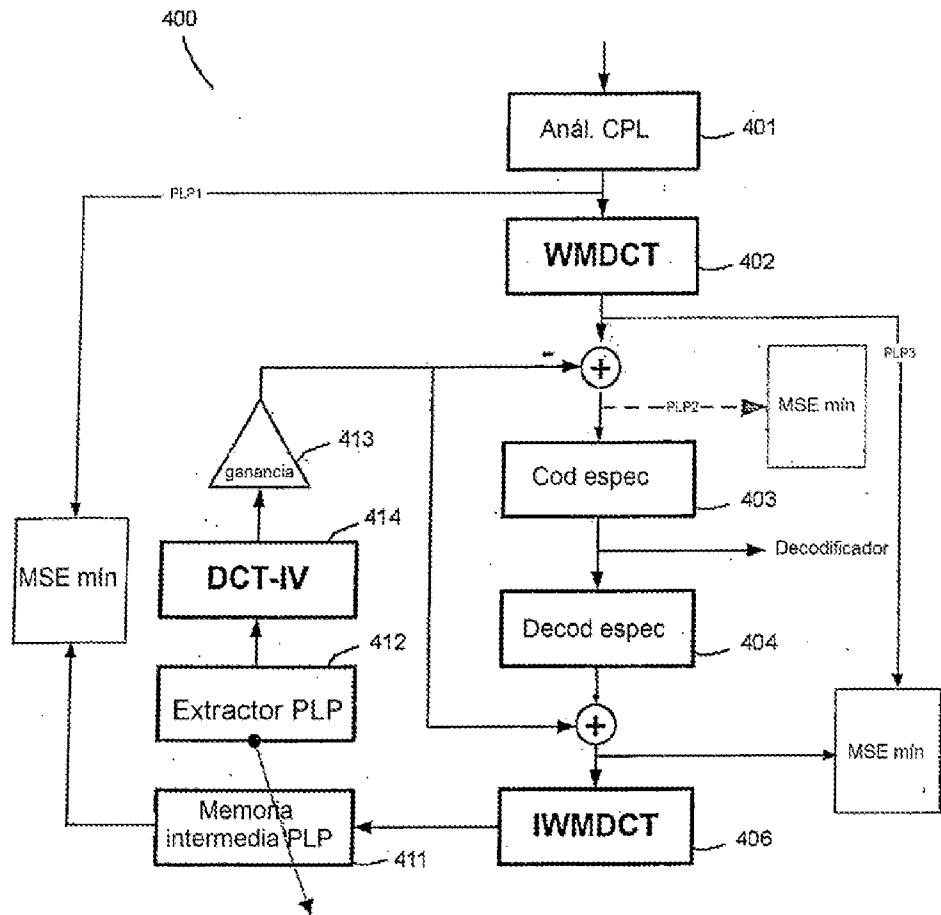


Fig. 4

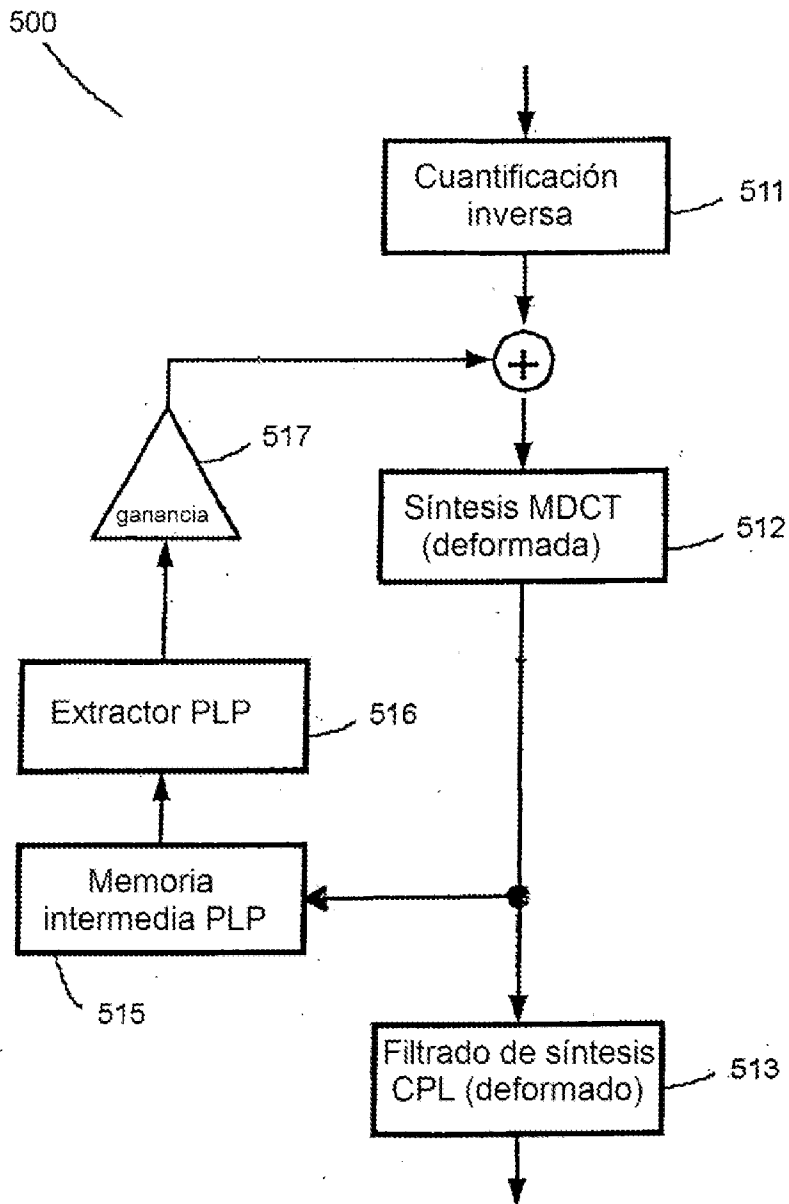


Fig. 5

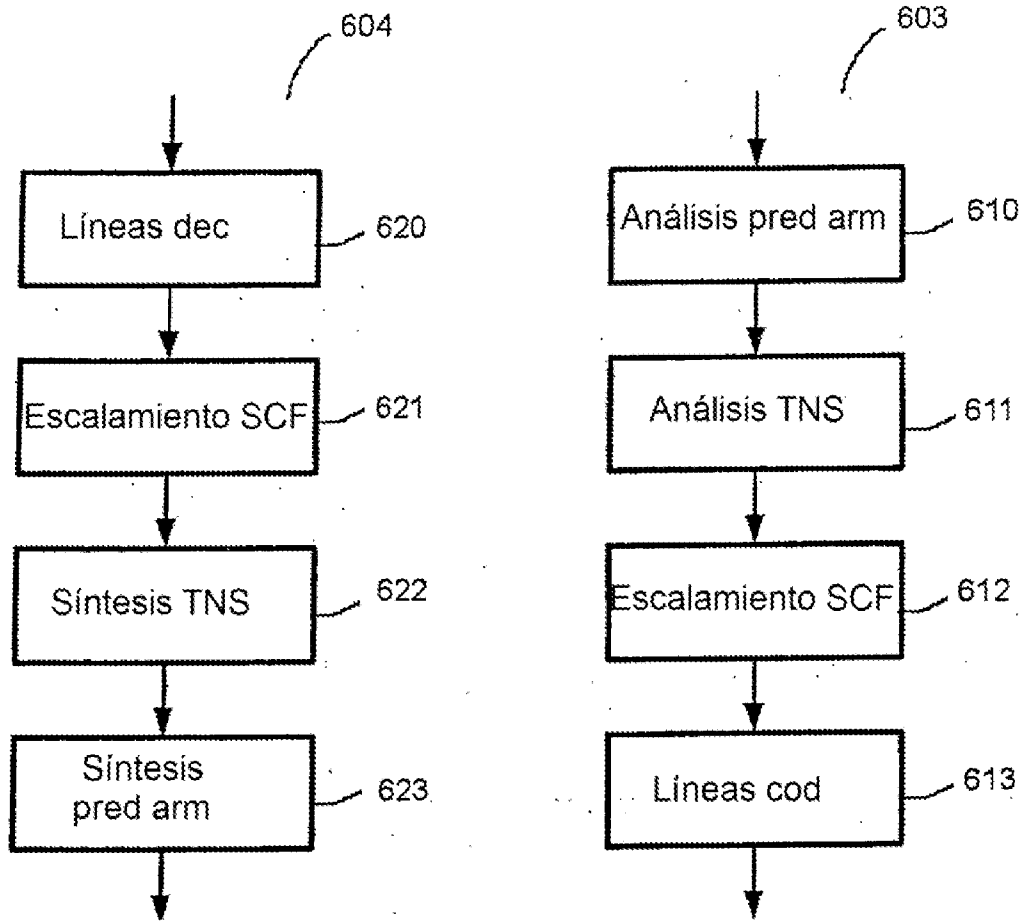


Fig. 6

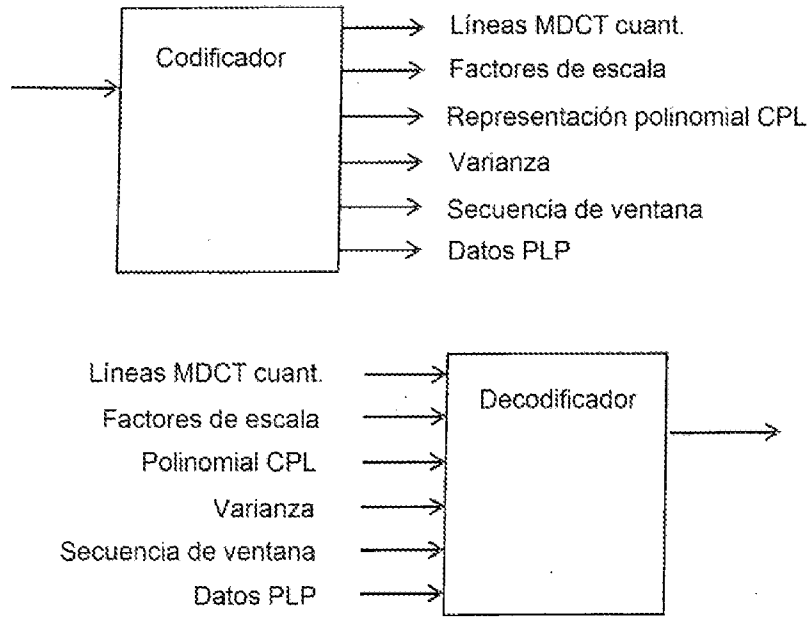


Fig. 7

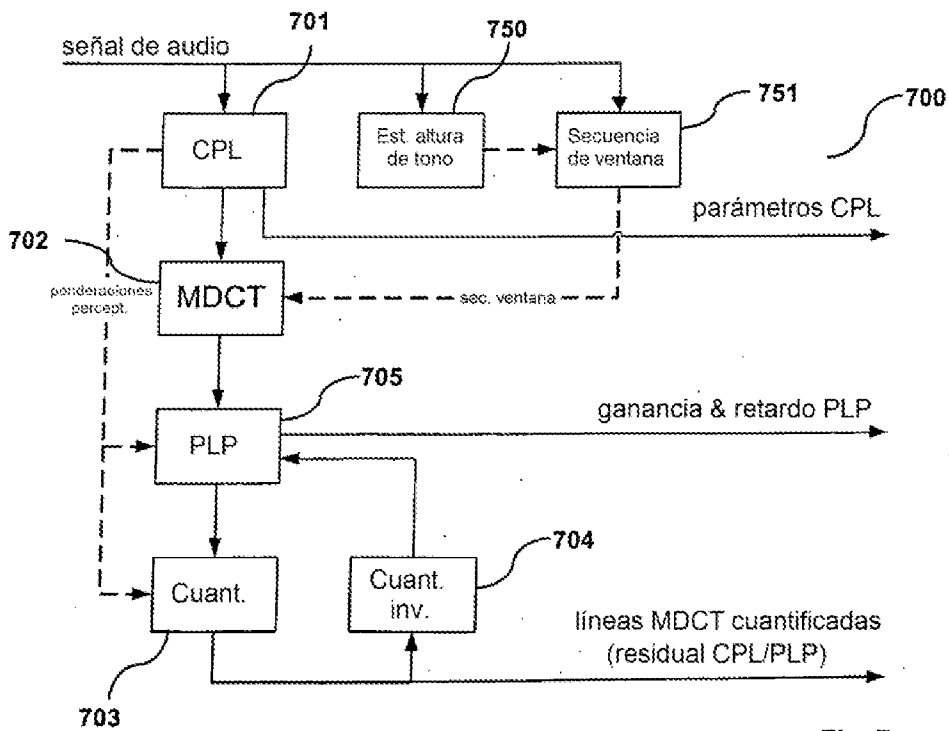


Fig. 7a

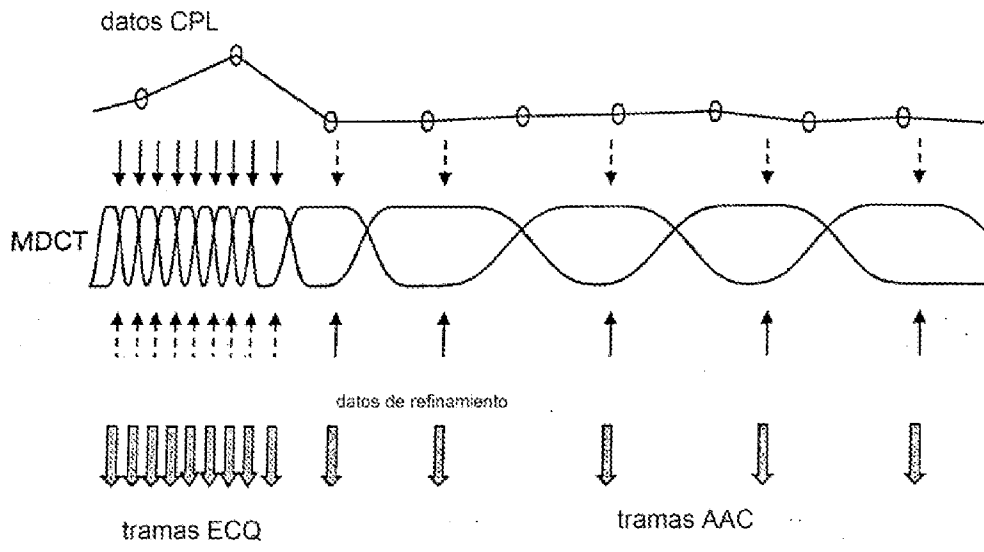
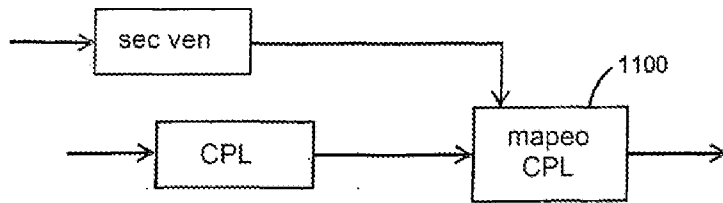
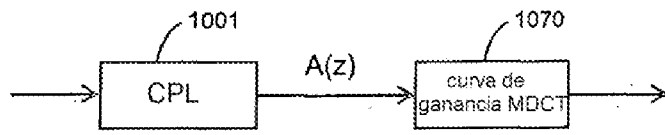
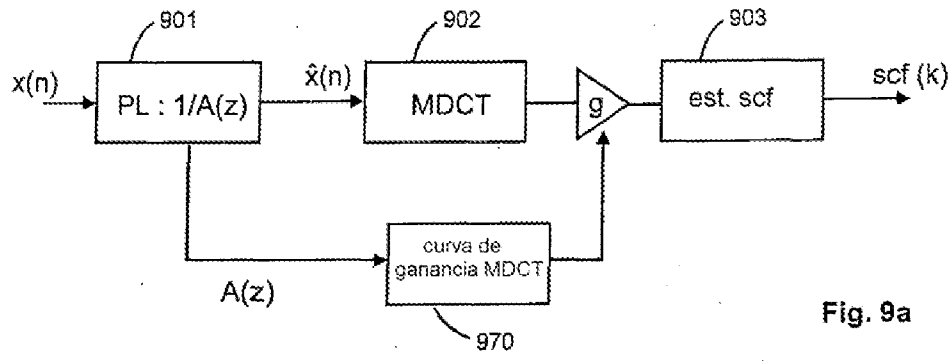
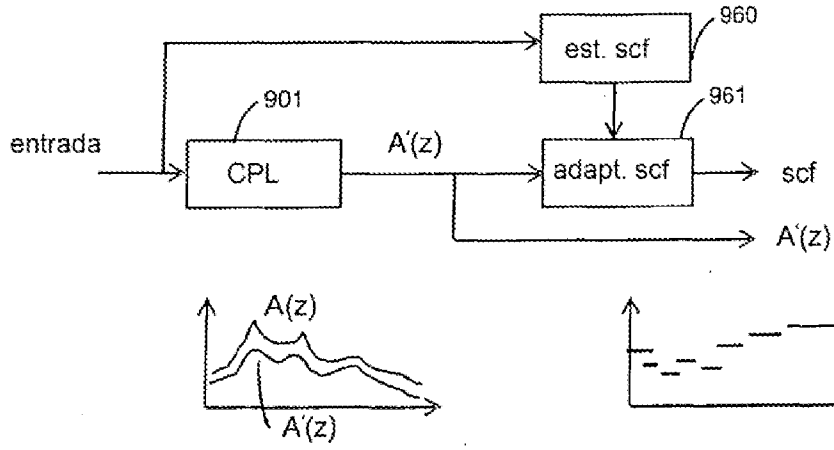


Fig. 8



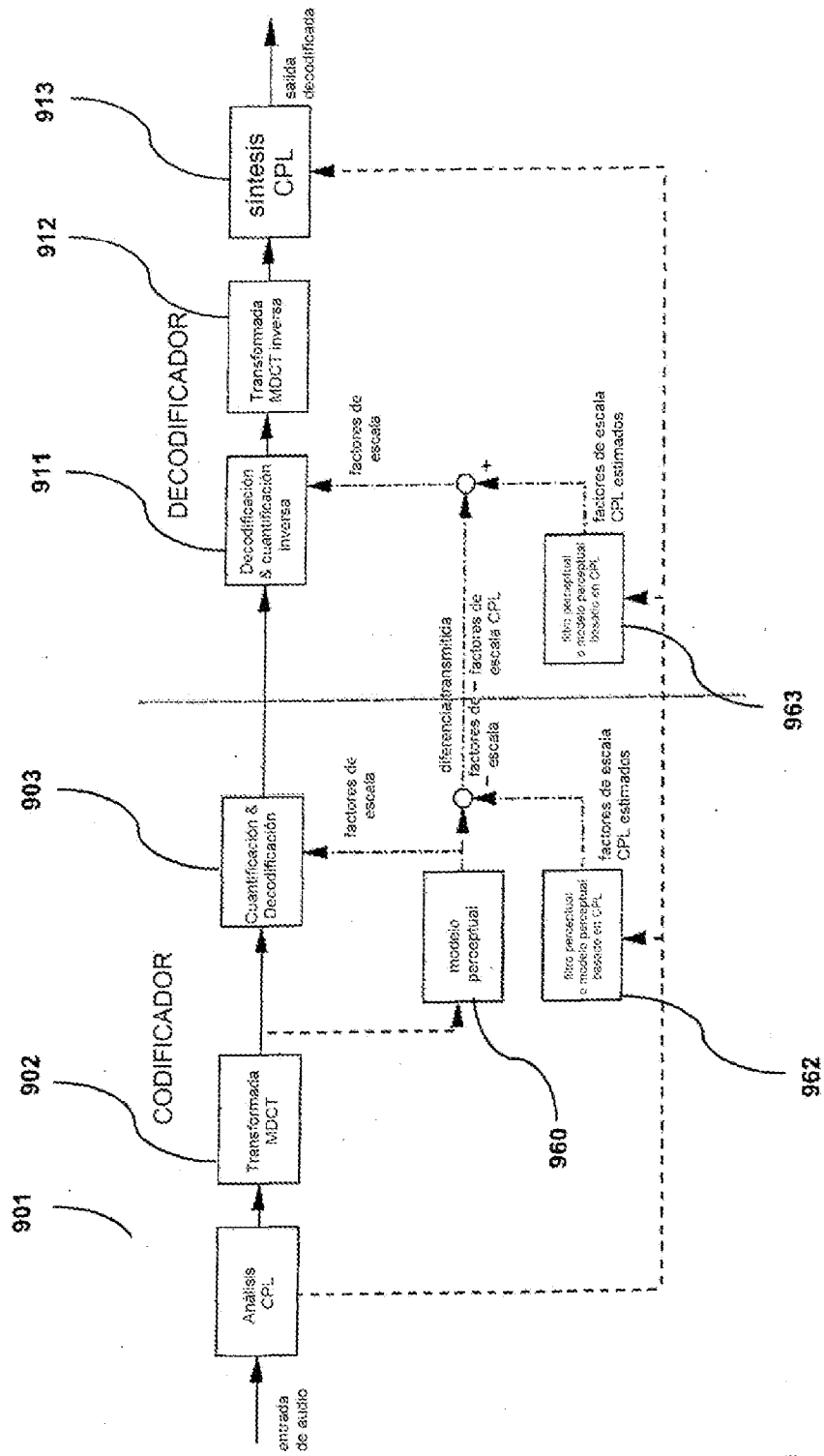


Fig. 9b

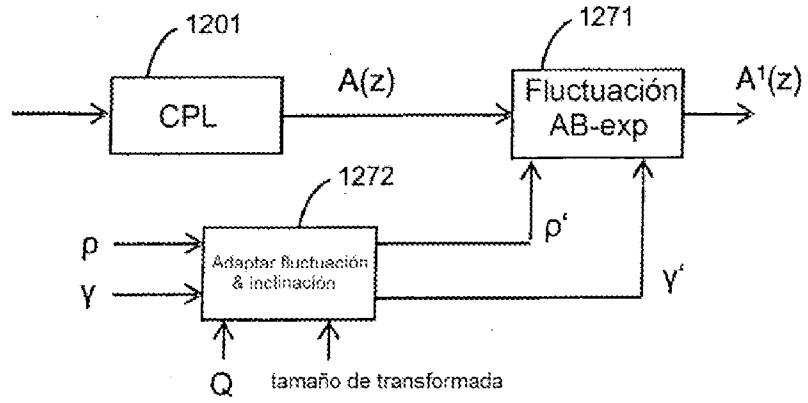


Fig. 12

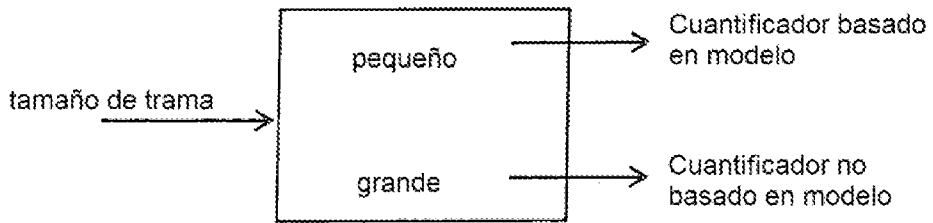


Fig. 13

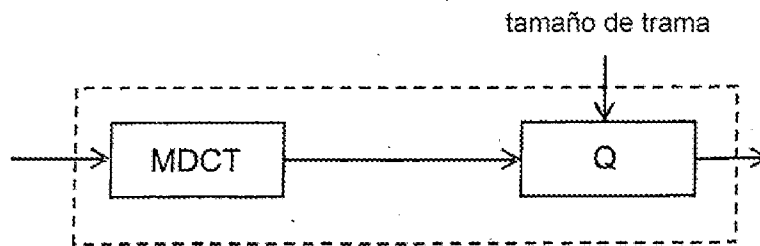


Fig. 14

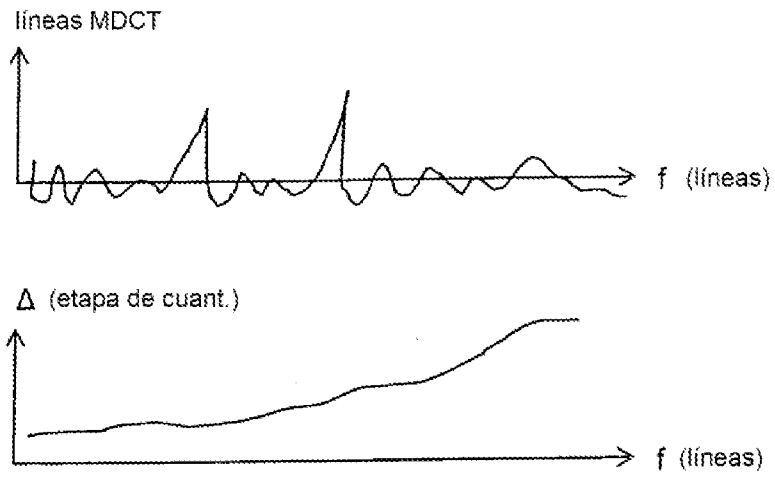


Fig. 15

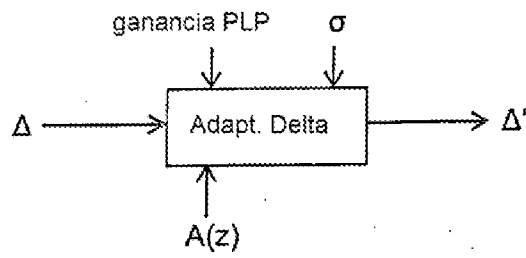


Fig. 15a

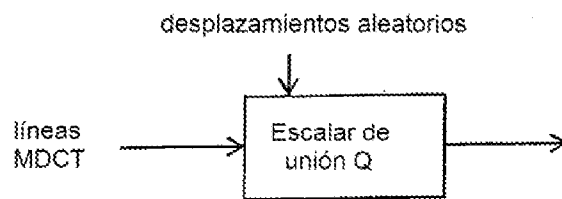


Fig. 16

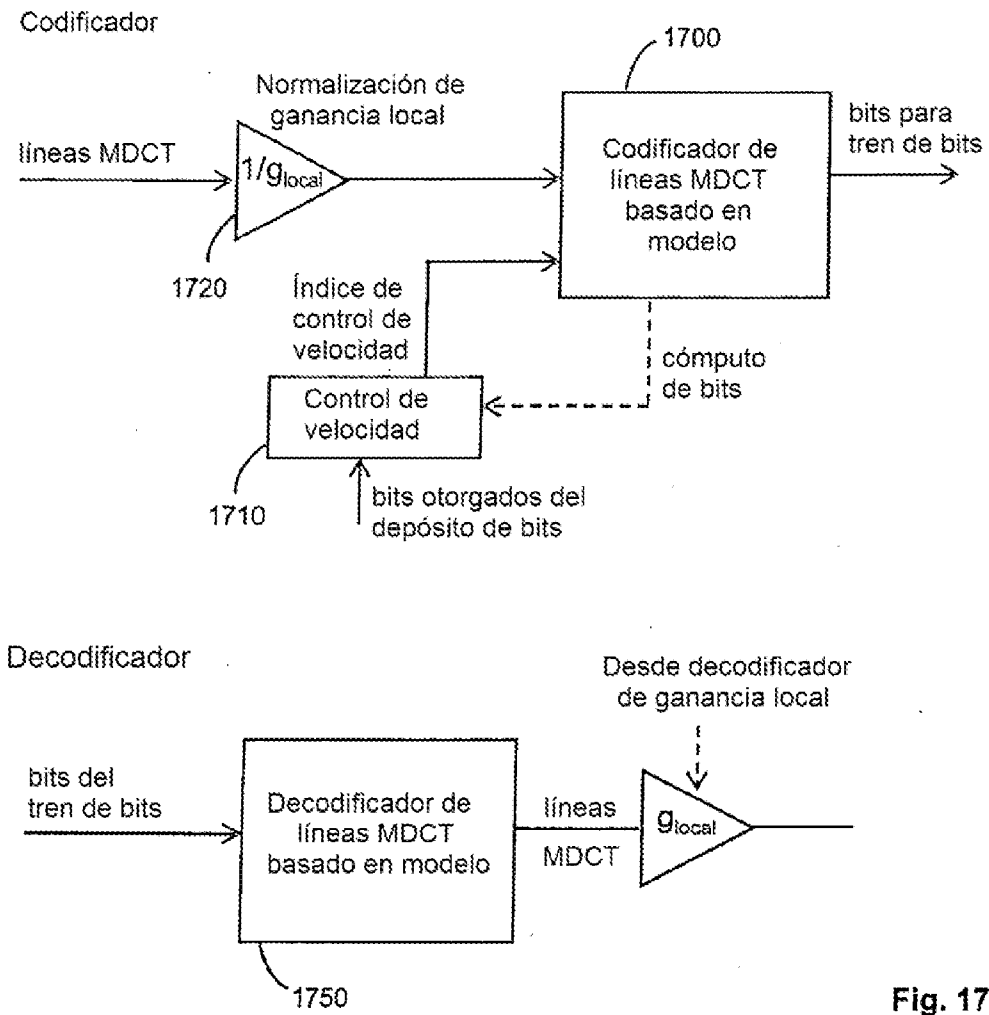


Fig. 17

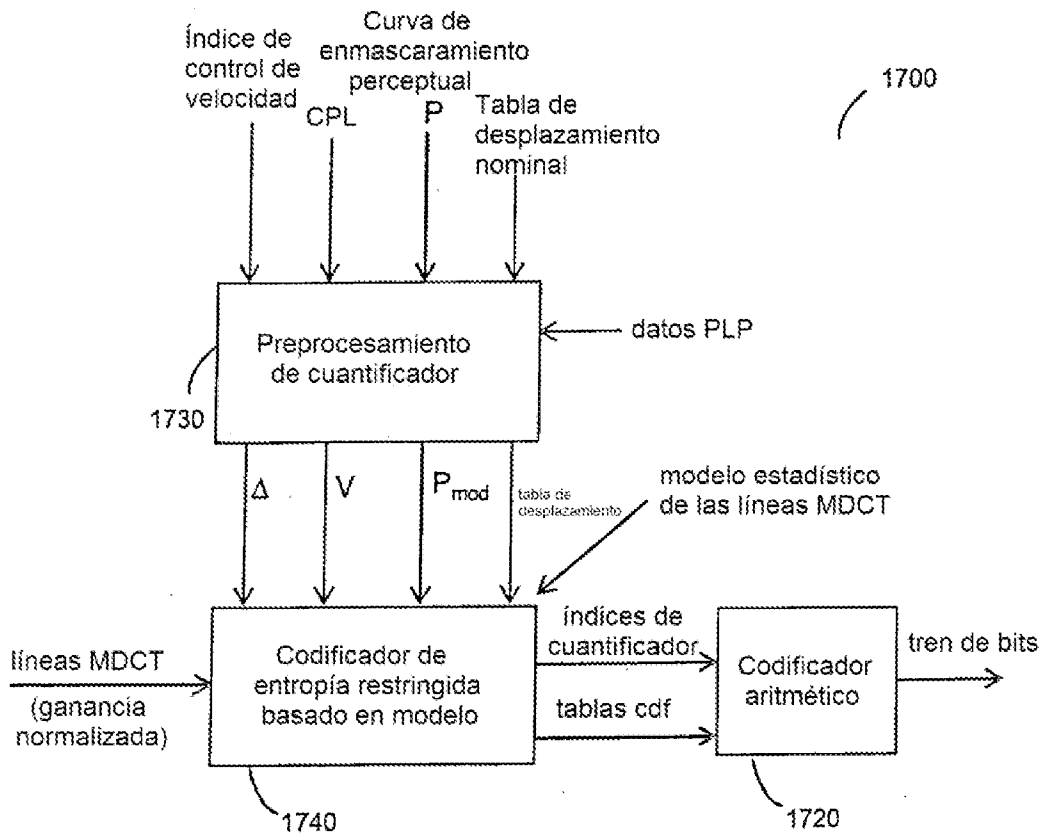


Fig. 17a

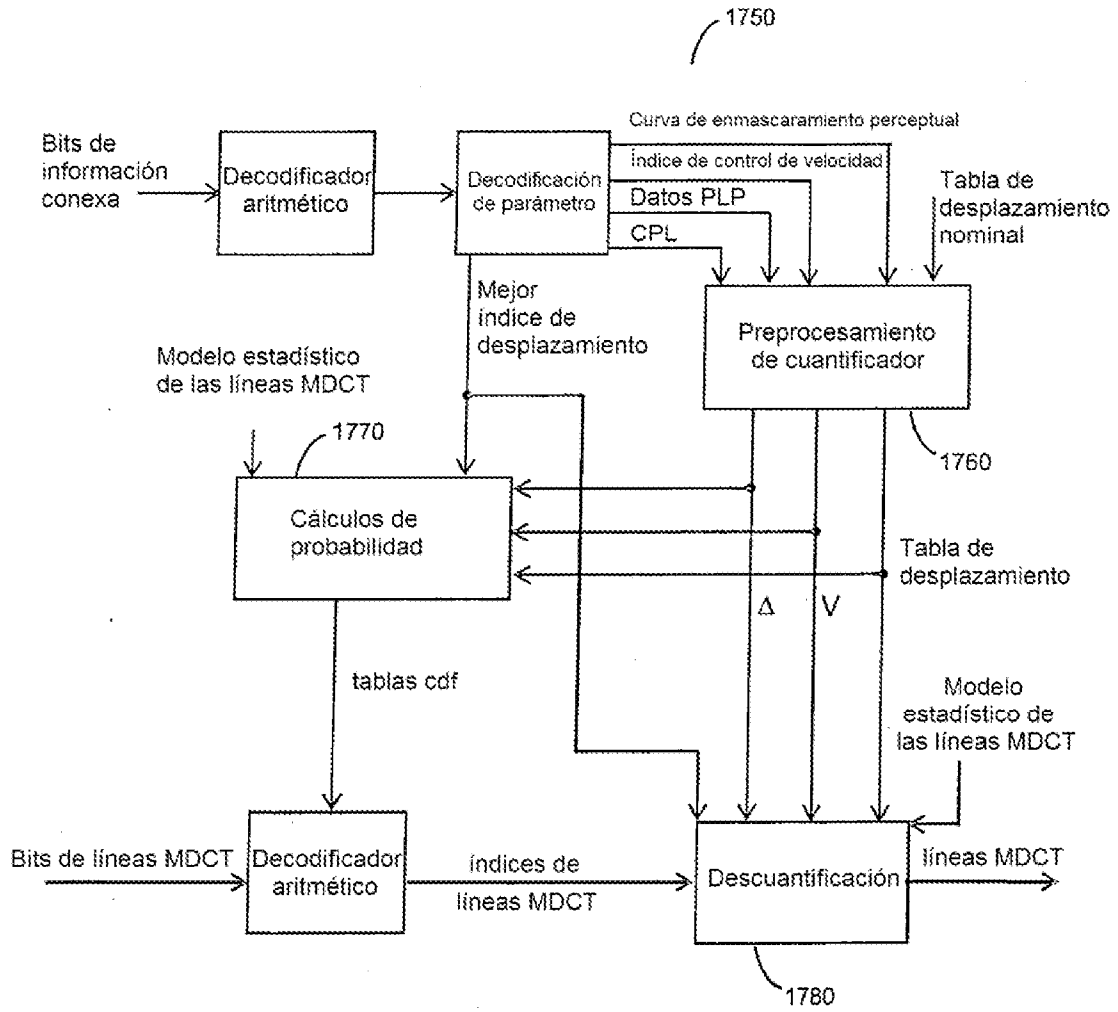


Fig. 17b

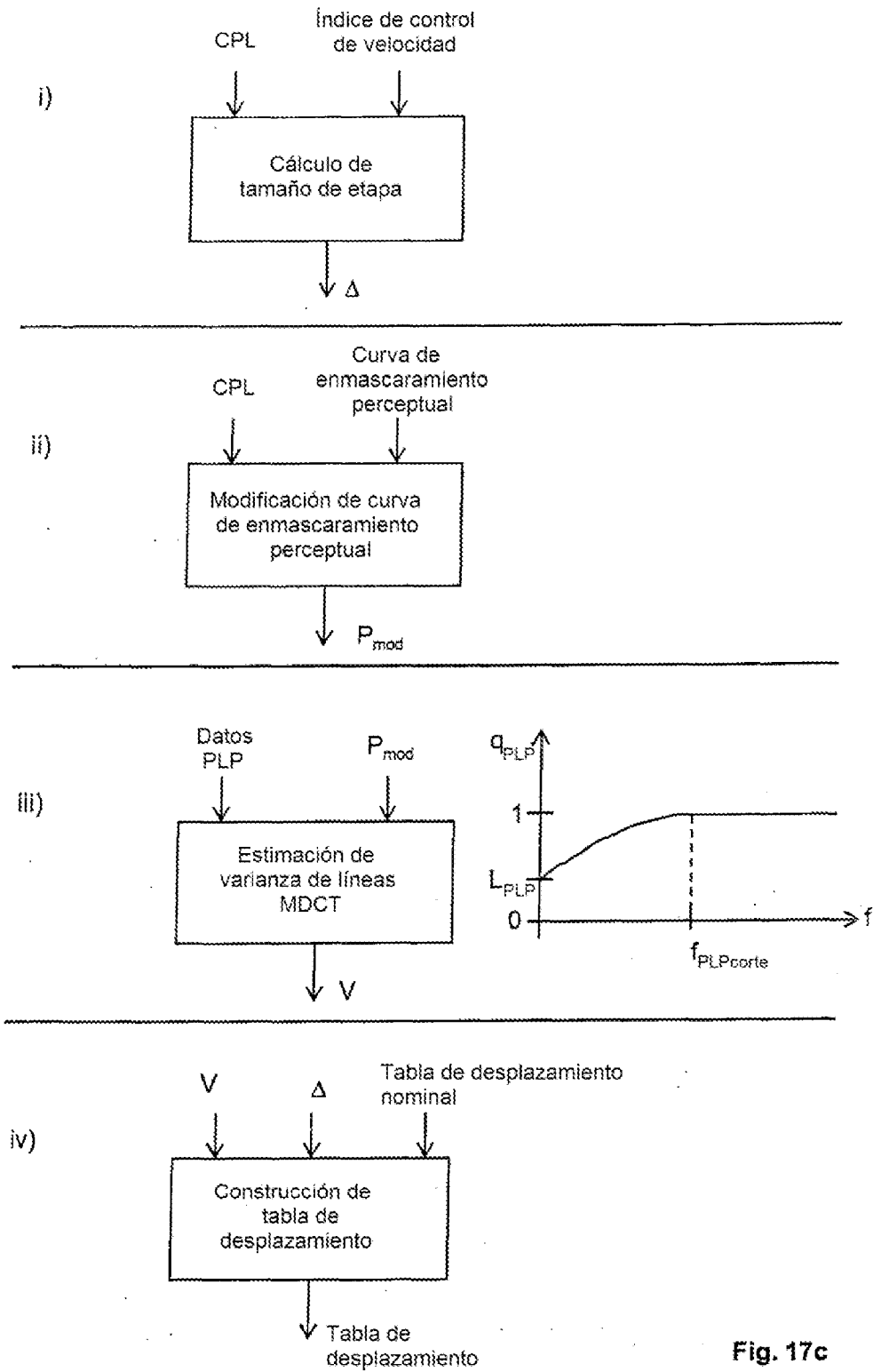
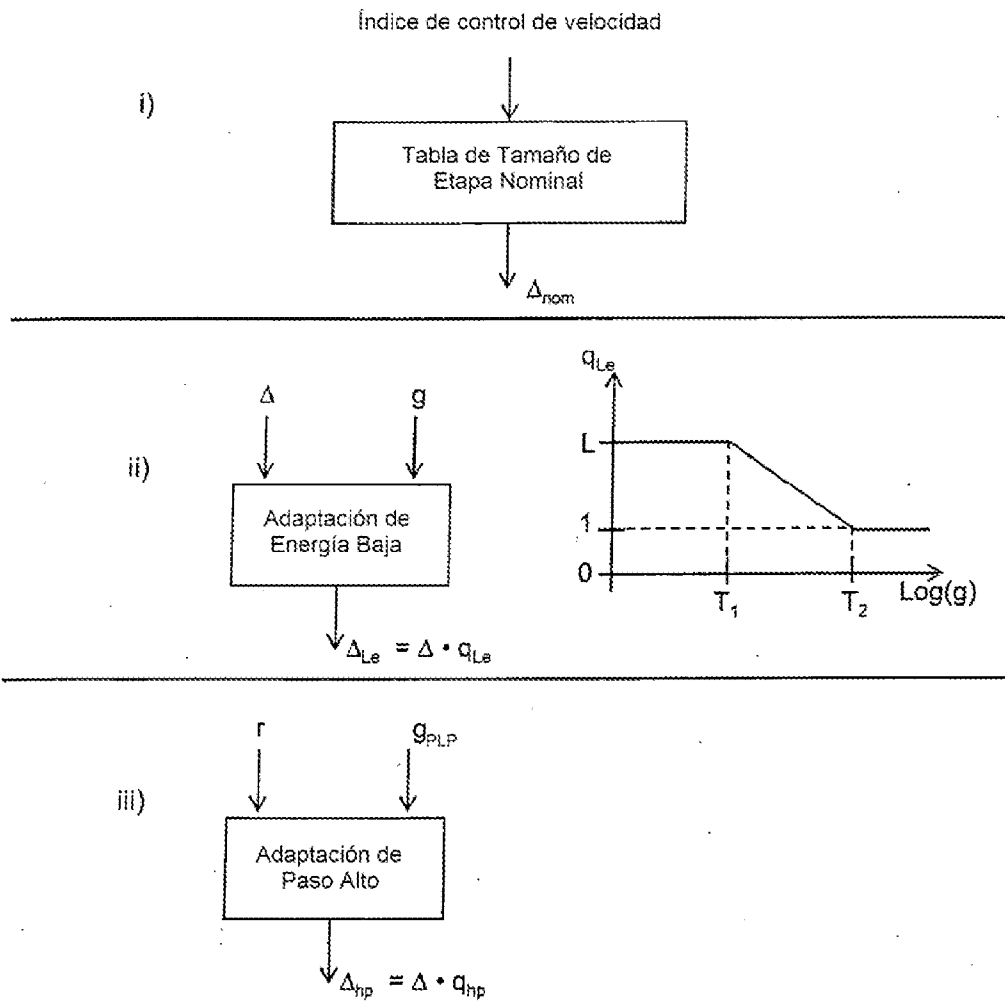


Fig. 17c



r
↓
 g_{PLP}
↓

Adaptación de Paso Alto

 ↓
 $\Delta_{hp} = \Delta \cdot q_{hp}$

Fig. 17d

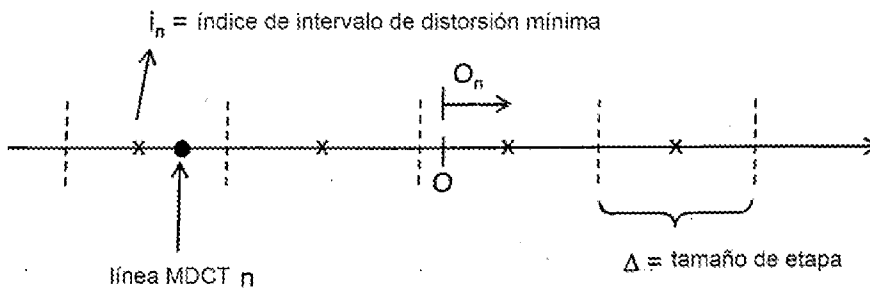


Fig. 17f

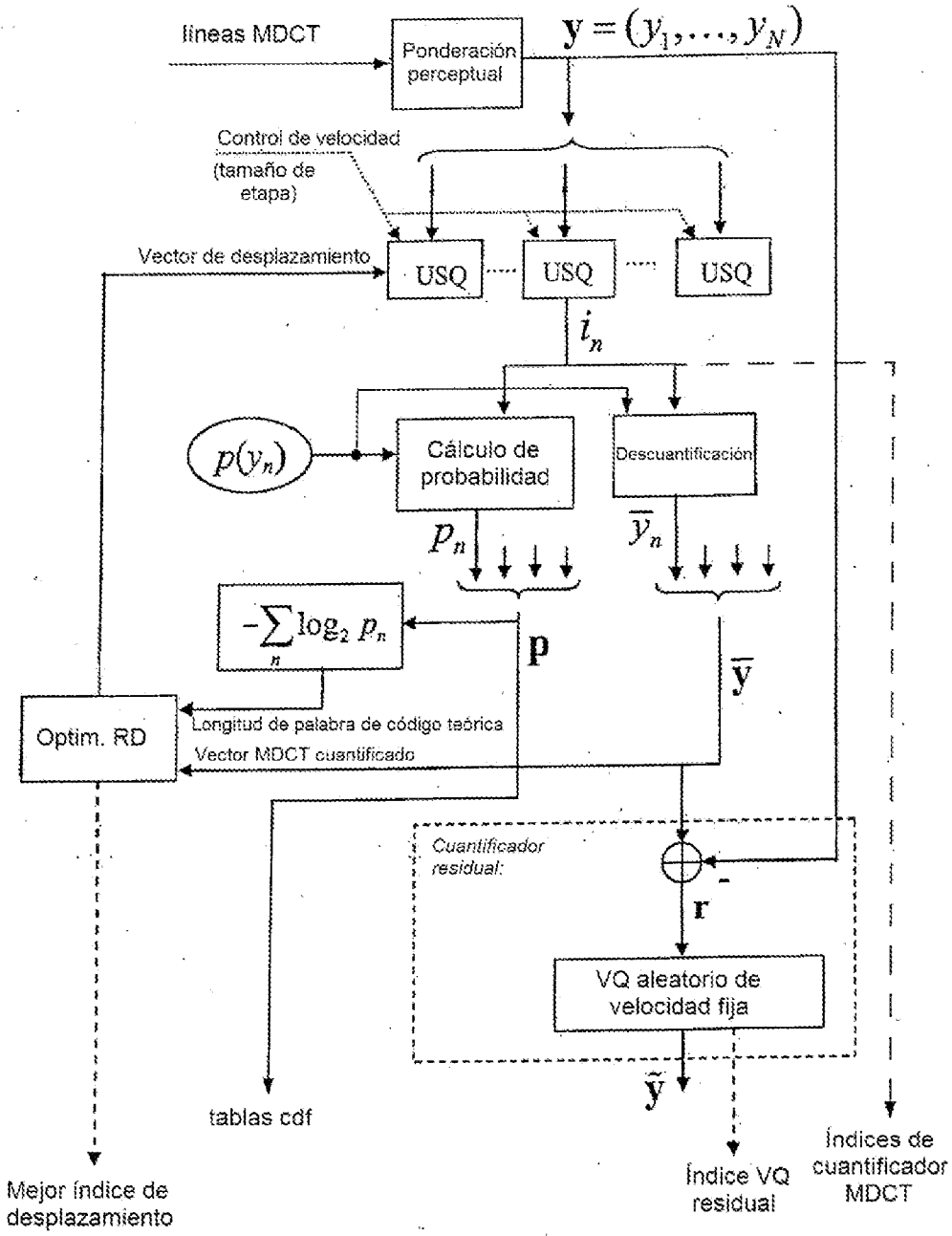


Fig. 17e

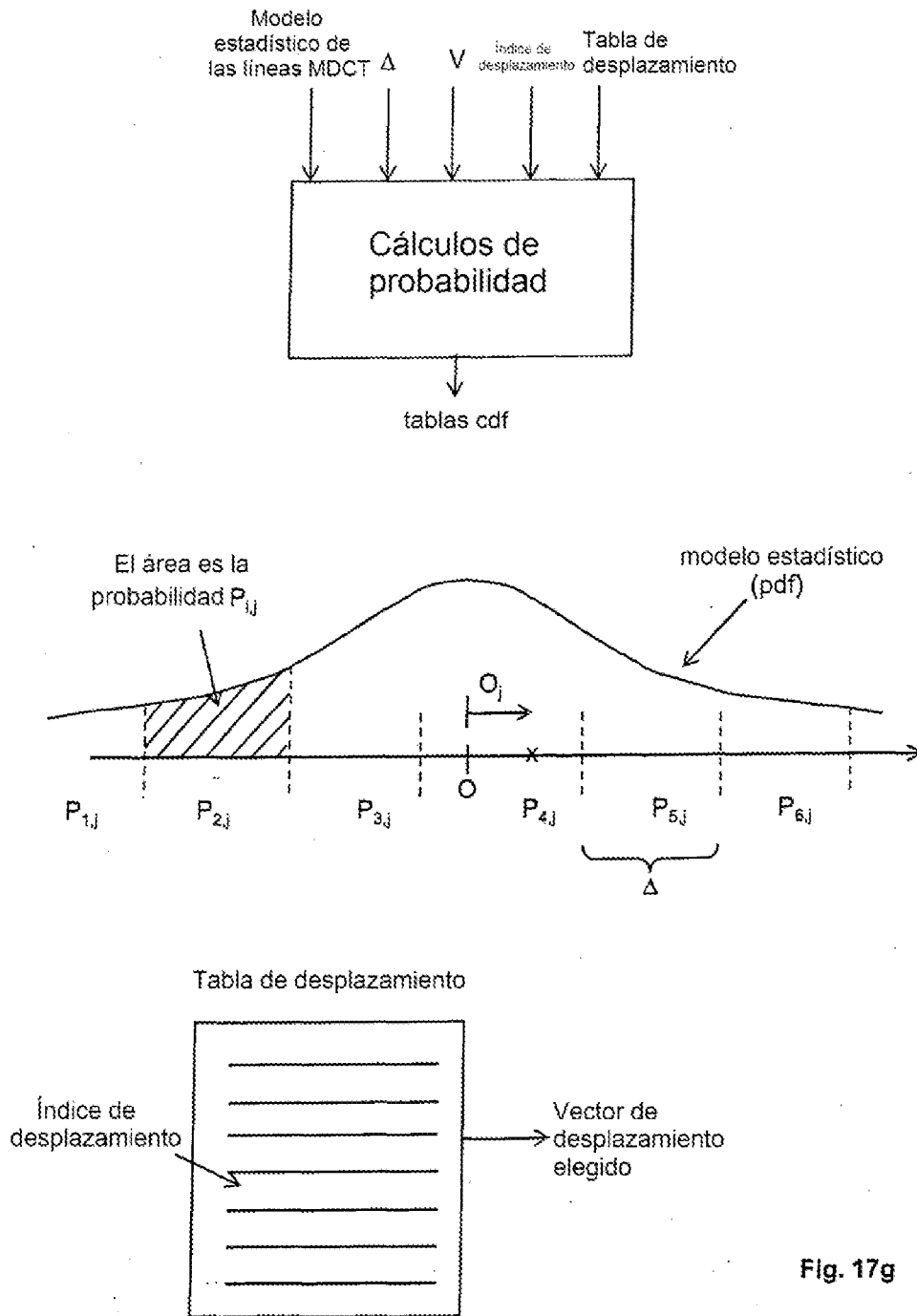


Fig. 17g

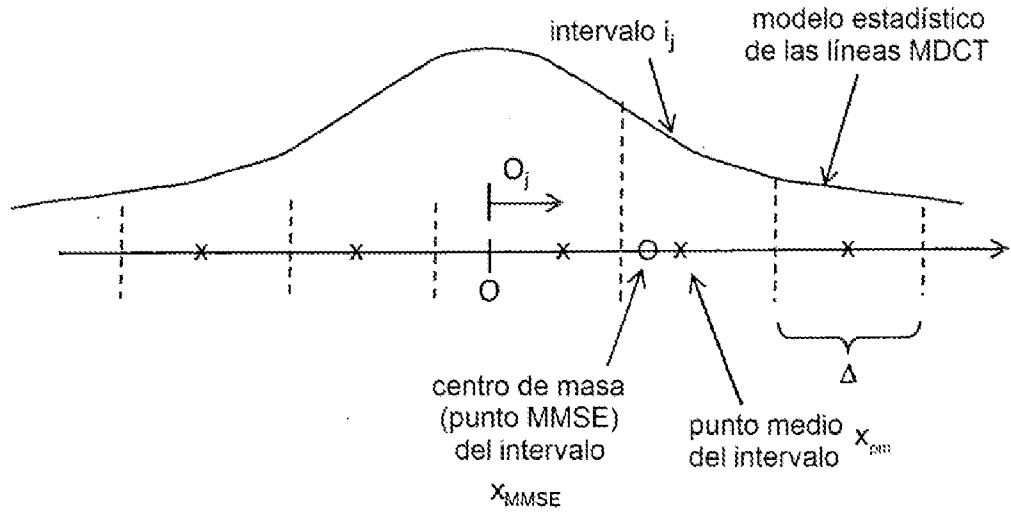


Fig. 17h

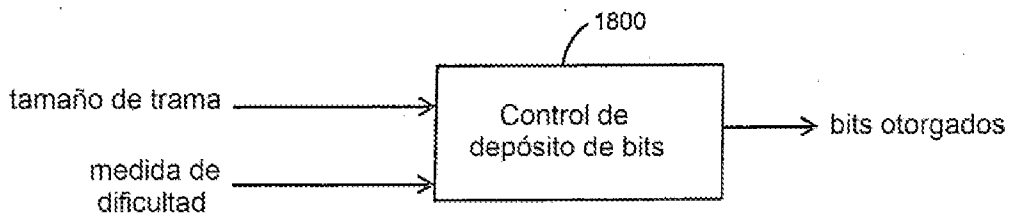


Fig. 18

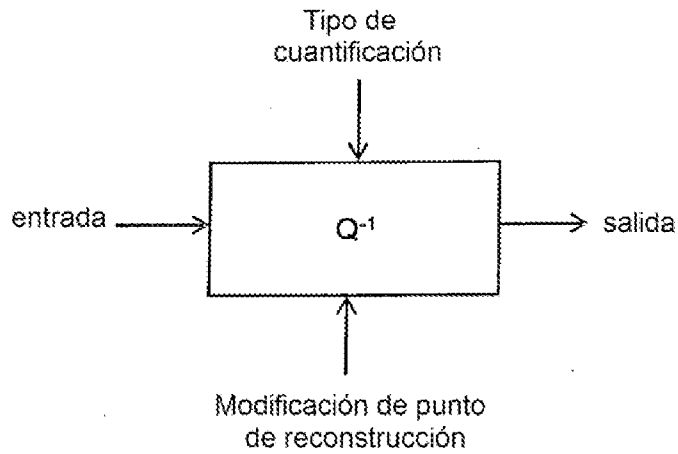


Fig. 19

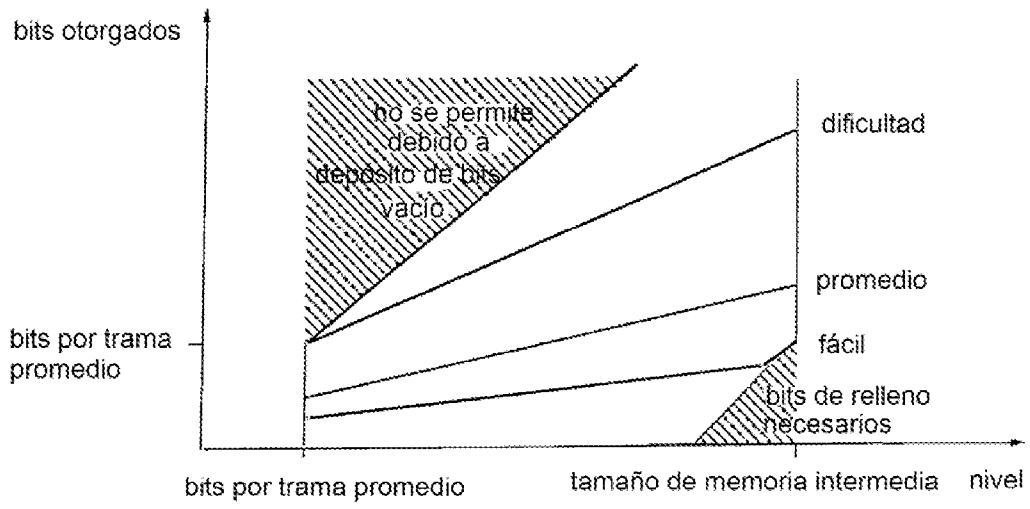


Fig. 18a

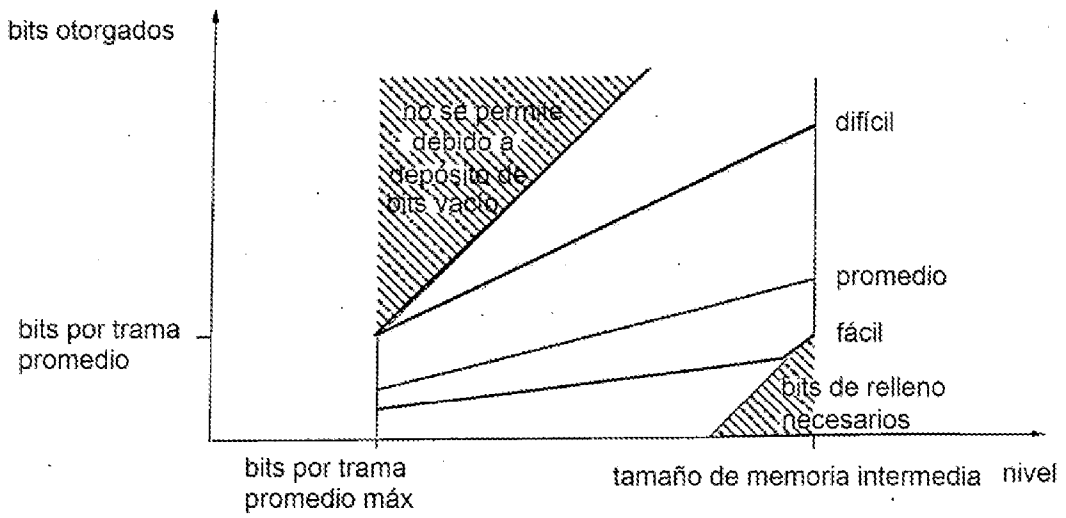


Fig. 18b

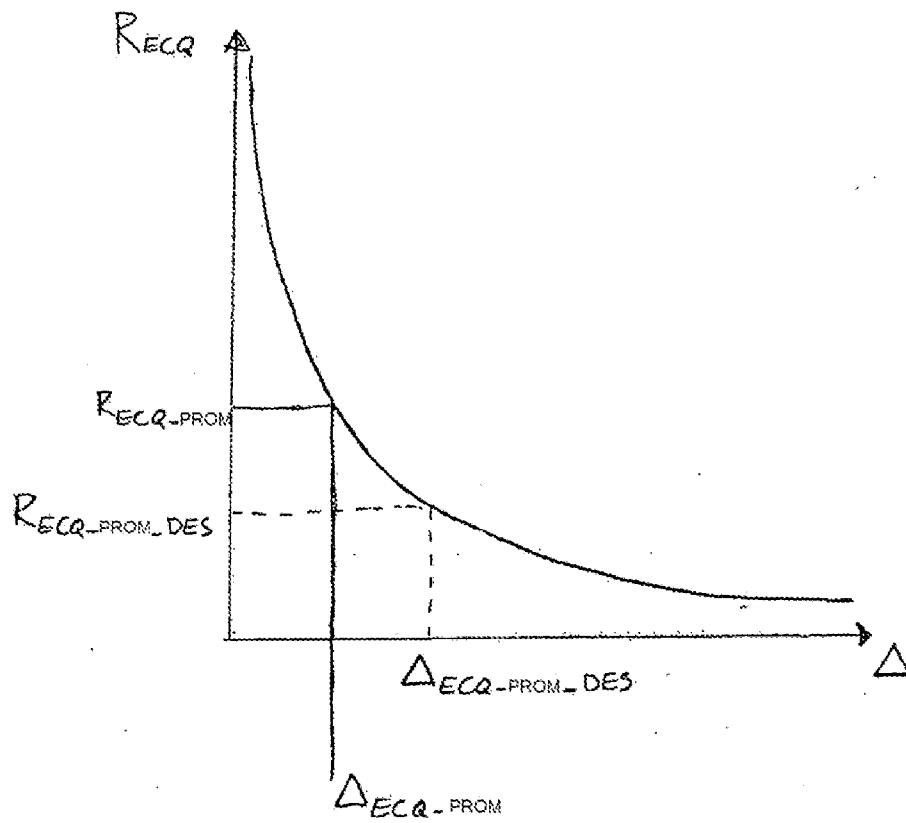


Fig. 18c