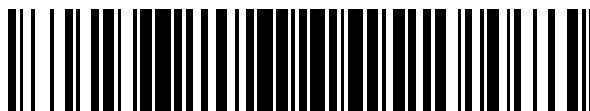


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 677 900**

51 Int. Cl.:

G10L 19/032 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **30.12.2008 PCT/EP2008/011144**

87 Fecha y número de publicación internacional: **16.07.2009 WO09086918**

96 Fecha de presentación y número de la solicitud europea: **30.12.2008 E 08870326 (9)**

97 Fecha y número de publicación de la concesión europea: **30.05.2018 EP 2235719**

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

30 Prioridad:

04.01.2008 SE 0800032
24.05.2008 EP 08009530
24.05.2008 US 55978 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
07.08.2018

73 Titular/es:

DOLBY INTERNATIONAL AB (100.0%)
ATLAS COMPLEX, AFRICA BUILDING
HOOGOORDDREEF 9
1101 BA AMSTERDAM, NL

72 Inventor/es:

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

74 Agente/Representante:

LEHMANN NOVO, María Isabel

ES 2 677 900 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.

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.

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, a saber, el sistema vocal humano. Dichos codificadores no pueden manejar señales de audio arbitrarias como, por ejemplo, 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, sin embargo, a velocidades binarias bajas para las señales de voz, el codificador de voz dedicado provee 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 funcionan a velocidades binarias bajas.

El documento EP-1278184-A2 describe un método de codificación por transformada eficaz para señales de música que es apropiado para su uso en un códec híbrido, mediante el cual un filtro de síntesis Predictivo Lineal (PL) común se emplea tanto para las señales de voz como de música. El filtro de síntesis PL conmuta entre un generador de excitación de voz y un generador de excitación de transformada, según la codificación de una señal de voz o música, respectivamente. Para codificar señales de voz, puede usarse la técnica CELP convencional, mientras una técnica de transformada de superposición añadida asimétrica innovadora se aplica para codificar señales de música. Al llevar a cabo el filtrado de síntesis PL común, la interpolación de los coeficientes PL se lleva a cabo para señales en regiones de funcionamiento de superposición añadida. La invención permite las transiciones suaves cuando el decodificador conmuta entre modos de decodificación de voz y música.

El documento US-2002/0010577-A1 describe un aparato y un método para codificar una señal de entrada en la base de tiempo a través de la transformada ortogonal, que comprende una etapa de eliminación de la correlación de la forma de onda de señal según los parámetros obtenidos por medio del análisis de codificación predictiva lineal (CPL) y del análisis de altura de tono de la señal de entrada en la base de tiempo con anterioridad a la transformada ortogonal. La señal de entrada de base de tiempo del terminal de entrada 10 se envía a la sección de circuito de normalización 11 y al circuito de análisis (CPL) 39. La sección de circuito de normalización 11 elimina la correlación de la forma de onda de señal y extrae los residuos mediante el filtro inverso CPL 12 y filtro inverso de altura de tono 13 y envía los residuos a la sección de circuito de transformada ortogonal 25. Los parámetros CPL del circuito de análisis superior 39 y los parámetros de altura de tono del circuito de análisis de altura de tono 15 se envían al circuito de cálculo de asignación de bits 41; la sección de cuantificación de coeficientes 40 cuantifica los coeficientes a partir de la sección de circuito de transformada ortogonal 25 según el número de bits asignados de la sección de cálculo de asignación de bits 41.

El documento de M. Oger, S. Ragot, M. Antonini, "Transform Audio Coding with Arithmetic-Coded Scalar Quantization and Model-Based Bit Allocation" *Proceedings of ICASSP 2007*, vol. 4, abril 15-20, 2007, páginas 5454-548 describe un método basado en modelo para codificar coeficientes de transformada de señales de audio. El histograma de coeficientes de transformada se aproxima por un modelo Gaussiano generalizado para la asignación de bits basada en modelo de coeficientes y el espectro se codifica por la cuantificación escalar seguida de la codificación aritmética.

Por consiguiente, existe la necesidad de un codificador y decodificador de audio mejorados con calidad de audio mejorada y/o velocidades binarias reducidas.

Compendio de la invención

La presente invención se refiere a la codificación, de manera eficaz, de señales de audio arbitrarias con un nivel de calidad igual a o mejor que el de un sistema específicamente adaptado a 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 que funciona en una señal procesada CPL.

La presente invención se refiere además a la codificación, de forma eficaz, de 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 al uso, de forma eficaz, de un depósito de bits en un codificador de audio con un tamaño de trama variable.

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 la cuantificación en un codificador de transformada que, p.ej., aplica una Transformada Discreta del Coseno Modificada (MDCT, por sus siglas en inglés). El cuantificador propuesto cuantifica, preferiblemente, líneas MDCT. El presente aspecto es aplicable independientemente de si el codificador además usa un análisis de codificación de predicción lineal (CPL) o predicción a largo plazo (PLP) adicional.

La presente invención provee un sistema de codificación de audio como en la reivindicación 1. Preferiblemente, la decisión se basa en el tamaño de trama aplicado por la unidad de transformación. Sin embargo, también se conciben otros criterios dependientes de la señal de entrada para la conmutación de la estrategia de cuantificación, los cuales se encuentran dentro del alcance de la presente solicitud.

Otro aspecto importante de la invención es que el cuantificador puede ser adaptativo. En particular, el modelo en el cuantificador basado en modelo puede ser adaptativo para ajustarse a la señal de audio de entrada. El modelo puede variar con el tiempo, p.ej., según las características de la señal de entrada. Ello permite una distorsión reducida de la cuantificación y, por consiguiente, una calidad mejorada de la codificación.

Según una realización, la estrategia de cuantificación propuesta se encuentra condicionada al tamaño de trama. Se sugiere que la unidad de cuantificación puede decidir, según el tamaño de trama aplicado por la unidad de transformación, codificar la señal de dominio de transformada con un cuantificador basado en modelo o un cuantificador no basado en modelo. Preferiblemente, la unidad de cuantificación se configura para codificar una señal de dominio de transformada para una trama con un tamaño de trama más pequeño que un valor umbral mediante una cuantificación de entropía restringida basada en modelo. La cuantificación basada en modelo puede condicionarse a parámetros variados. Las tramas grandes pueden cuantificarse, p.ej., por un cuantificador escalar con, p.ej., codificación de entropía basada en Huffman, como se usa en, p.ej., el códec AAC.

El sistema de codificación de audio puede además comprender una unidad de predicción a largo plazo (PLP) para calcular la trama de la señal de entrada filtrada según una reconstrucción de un segmento previo de la señal de entrada filtrada y una unidad de combinación de señales de dominio de transformada para combinar, en el dominio de transformada, la estimación de predicción a largo plazo y la señal de entrada transformada para generar la señal de dominio de transformada que se ingresa en la unidad de cuantificación.

La conmutación entre diferentes métodos de cuantificación de las líneas MDCT es otro aspecto de una realización preferida de la invención. Mediante el empleo de diferentes estrategias de cuantificación para diferentes tamaños de transformada, el códec puede hacer toda la cuantificación y codificación en el dominio MDCT sin tener la necesidad de tener un codificador de voz de dominio temporal específico ejecutándose en paralelo o en forma serial al códec de dominio de transformada. La presente invención enseña que para señales tipo voz, donde hay una ganancia PLP, la señal se codifica preferiblemente mediante el uso de una transformada corta y un cuantificador basado en modelo. El cuantificador basado en modelo es particularmente apropiado para la transformada corta y provee, como se describirá más adelante, las ventajas de un cuantificador de vector (CV) específico para la voz de dominio temporal, mientras aún funciona en el dominio MDCT, y sin requisitos de que la señal de entrada sea una señal de voz. En otras palabras, cuando el cuantificador basado en modelo se usa para los segmentos de transformada corta en combinación con la PLP, la eficacia del codificador CV de voz de dominio temporal dedicado se retiene sin pérdida de generalidad y sin abandonar el dominio MDCT.

Además, para señales de música más estacionarias, se prefiere usar una transformada de tamaño relativamente grande como se usa comúnmente en códecs de audio, y un esquema de cuantificación que puede tomar ventaja de líneas espectrales dispersas discriminadas por la transformada grande. Por lo tanto, la presente invención enseña a usar el presente tipo de esquema de cuantificación para transformadas largas.

Por consiguiente, la conmutación de estrategia de cuantificación como una función de tamaño de trama permite al códec retener tanto las propiedades de un códec de voz dedicado como las propiedades de un códec de audio dedicado, simplemente mediante la elección del tamaño de transformada. Ello evita todos los problemas en los sistemas de la técnica anterior que luchan por manejar señales de voz y audio igualmente bien a velocidades bajas, dado que dichos sistemas inevitablemente se encuentran con los problemas y dificultades de combinar, de forma eficaz, la codificación de dominio temporal (el codificador de voz) con la codificación de dominio de la frecuencia (el codificador de audio).

Según otro aspecto de la invención, la cuantificación usa tamaños de etapa adaptativos. Preferiblemente, el tamaño de etapa de cuantificación para componentes de la señal de dominio de transformada se adapta según la predicción

lineal y/o parámetros de predicción a largo plazo. El tamaño de etapa de cuantificación puede además configurarse para que dependa de la frecuencia. En realizaciones de la invención, el tamaño de etapa de cuantificación se determina según al menos uno de: el polinomio del filtro adaptativo, un parámetro de control de velocidad de codificación, un valor de ganancia de predicción a largo plazo, y una varianza de señal de entrada.

5 Preferiblemente, la unidad de cuantificación comprende cuantificadores escalares uniformes para cuantificar los componentes de señal de dominio de transformada. Cada cuantificador escalar aplica una cuantificación uniforme, p.ej., basada en un modelo de probabilidad, a una línea MDCT. El modelo de probabilidad puede ser un modelo Laplaciano o Gaussiano, o cualquier otro modelo de probabilidad que sea apropiado para las características de señal. La unidad de cuantificación puede además insertar un desplazamiento aleatorio en los cuantificadores
10 escalares uniformes. La inserción de un desplazamiento aleatorio provee ventajas de cuantificación de vector a los cuantificadores escalares uniformes. Según una realización, los desplazamientos aleatorios se determinan según una optimización de una distorsión de cuantificación, preferiblemente en un dominio perceptual y/o teniendo en cuenta el coste en términos del número de bits requeridos para codificar los índices de cuantificación.

15 La unidad de cuantificación puede además comprender un codificador aritmético para codificar índices de cuantificación generados por los cuantificadores escalares uniformes. Ello logra una velocidad binaria baja que se aproxima a la mínima posible provista por la entropía de señal.

La unidad de cuantificación puede además comprender un cuantificador residual para cuantificar una señal de cuantificación residual que resulta de los cuantificadores escalares uniformes con el fin de reducir más la distorsión total. El cuantificador residual es, preferiblemente, un cuantificador de vector de velocidad fija.

20 Pueden usarse múltiples puntos de reconstrucción de cuantificación en la unidad de descuantificación del codificador y/o cuantificador inverso en el decodificador. Por ejemplo, pueden usarse puntos de reconstrucción de error cuadrático medio mínimo (MMSE, por sus siglas en inglés) y/o punto central (punto medio) para reconstruir un valor cuantificado según su índice de cuantificación. Un punto de reconstrucción de cuantificación puede además basarse en una interpolación dinámica entre un punto central y un punto MMSE, posiblemente controlado por las
25 características de los datos. Ello permite controlar la inserción de ruido y evitar agujeros espectrales debido a la asignación de líneas MDCT a un segmento de cuantificación cero para velocidades binarias bajas.

Una ponderación perceptual en el dominio de transformada se aplica, preferiblemente, cuando se determina la distorsión de cuantificación con el fin de poner diferentes ponderaciones a componentes de frecuencia específicos. Las ponderaciones perceptuales pueden derivarse, de manera eficaz, de los parámetros de predicción lineal.

30 Otro aspecto independiente de la invención se refiere al concepto general de hacer uso de la coexistencia de datos CPL y SCF (Factor de Escala). En un codificador basado en transformada, p.ej., mediante la aplicación de una Transformada Discreta del Coseno Modificada (MDCT), los factores de escala pueden usarse en la cuantificación para controlar el tamaño de etapa de cuantificación. En la técnica anterior, dichos factores de escala se calculan a partir de la señal original para determinar una curva de enmascaramiento. Ahora se sugiere calcular un segundo
35 conjunto de factores de escala con la ayuda de un filtro perceptual o modelo psicoacústico que se calcula a partir de datos CPL. Ello permite una reducción del coste para transmitir/almacenar los factores de escala mediante la transmisión/almacenamiento solamente de la diferencia de los factores de escala realmente aplicados con respecto a los factores de escala de calculados por CPL en lugar de transmitir/almacenar los factores de escala reales. Por consiguiente, en un sistema de codificación de audio que contiene elementos de codificación de voz como, por
40 ejemplo, una CPL, y elementos de codificación por transformada como, por ejemplo, una MDCT, la presente invención reduce el coste de transmisión de información de factores de escala que se necesita para la parte de codificación por transformada del códec mediante la explotación de datos provistos por la CPL. Debe notarse que el presente aspecto es independiente de otros aspectos del sistema de codificación de audio propuesto y también puede implementarse en otros sistemas de codificación de audio.

45 Por ejemplo, una curva de enmascaramiento perceptual puede calcularse según los parámetros del filtro adaptativo. El segundo conjunto de factores de escala basado en la predicción lineal puede determinarse según la curva de enmascaramiento perceptual estimada. La información de factor de escala almacenada/transmitida se determina entonces según la diferencia entre los factores de escala realmente usados en la cuantificación y los factores de escala que se calculan a partir de la curva de enmascaramiento perceptual basada en CPL. Ello elimina la dinámica
50 y redundancia de la información almacenada/transmitida de modo que menos bits son necesarios para almacenar/transmitir los factores de escala.

En el caso en el que la CPL y MDCT no funcionan a la misma velocidad de trama, a saber, tienen diferentes tamaños de trama, los factores de escala basados en la predicción lineal para una trama de la señal de dominio de transformada pueden estimarse según los parámetros de predicción lineal interpolados para coincidir con la ventana
55 de tiempo cubierta por la trama MDCT.

La presente descripción provee, por lo tanto, un sistema de codificación de audio que se basa en un codificador de transformada e incluye módulos fundamentales de predicción y forma a partir de un codificador de voz. El sistema

comprende una unidad de predicción lineal para filtrar una señal de entrada según un filtro adaptativo; una unidad de transformación para transformar una trama de la señal de entrada filtrada en un dominio de transformada; una unidad de cuantificación para cuantificar una señal de dominio de transformada; una unidad de determinación de factor de escala para generar factores de escala, según una curva umbral de enmascaramiento, para su uso en la unidad de cuantificación cuando se cuantifica la señal de dominio de transformada; una unidad de estimación de factor de escala de predicción lineal para estimar factores de escala basados en la predicción lineal según parámetros del filtro adaptativo; y un codificador de factor de escala para codificar la diferencia entre los factores de escala basados en la curva umbral de enmascaramiento y los factores de escala basados en la predicción lineal. Mediante la codificación de la diferencia entre los factores de escala aplicados y los factores de escala que pueden determinarse en el decodificador según la información de predicción lineal disponible, la eficacia de codificación y almacenamiento puede mejorarse y menos bits necesitan almacenarse/transmitirse.

Otro aspecto específico del codificador independiente de la descripción se refiere al manejo del depósito de bits para tamaños de trama variables. En un sistema de codificación de audio que puede codificar tramas de longitud variable, el depósito de bits se controla mediante la distribución de los bits disponibles entre las tramas. Dada una medida de dificultad razonable para las tramas individuales y un depósito de bits de un tamaño definido, cierto desvío de una velocidad binaria constante requerida permite una mejor calidad total sin violar los requisitos de memoria intermedia que se imponen por el tamaño de depósito de bits. La presente descripción extiende el concepto de uso de un depósito de bits a un control de depósito de bits para un códec de audio generalizado con tamaños de trama variables. Un sistema de codificación de audio puede, por lo tanto, comprender una unidad de control de depósito de bits para determinar el número de bits otorgado para codificar una trama de la señal filtrada según la longitud de la trama y una medida de dificultad de la trama. Preferiblemente, la unidad de control de depósito de bits tiene ecuaciones de control separadas para diferentes medidas de dificultad de trama y/o diferentes tamaños de trama. Las medidas de dificultad para diferentes tamaños de trama pueden normalizarse de modo que pueden compararse más fácilmente. Con el fin de controlar la asignación de bits para un codificador de velocidad variable, la unidad de control de depósito de bits preferiblemente establece el límite inferior permitido del algoritmo de control de bits otorgado en el número de bits promedio para el tamaño de trama más grande permitido.

Un aspecto adicional de la descripción se refiere al manejo de un depósito de bits en un codificador mediante el empleo de un cuantificador basado en modelo, p.ej., un Cuantificador de Restricción de Entropía (ECQ, por sus siglas en inglés). Se sugiere minimizar la variación del tamaño de etapa ECQ. Se sugiere una ecuación de control particular que relaciona el tamaño de etapa de cuantificador a la velocidad ECQ.

El filtro adaptativo para filtrar la señal de entrada se basa, preferiblemente, en un análisis de Codificación de Predicción Lineal (CPL) que incluye un filtro CPL que produce una señal de entrada blanqueada. Los parámetros CPL para la presente trama de datos de entrada pueden determinarse por algoritmos conocidos en la técnica. Una unidad de estimación de parámetros CPL puede calcular, para la trama de datos de entrada, cualquier representación de parámetros CPL apropiada como, por ejemplo, polinomios, funciones de transferencia, coeficientes de reflexión, frecuencias espectrales de línea, etc. El tipo particular de representación de parámetros CPL que se usa para la codificación u otros procesamiento depende de los respectivos requisitos. Como una persona con experiencia en la técnica sabe, algunas representaciones son más apropiadas para ciertas funciones que otras y, por lo tanto, se prefieren para llevar a cabo dichas funciones. La unidad de predicción lineal puede funcionar en una primera longitud de trama que es fija, p.ej., 20 mseg. El filtrado de predicción lineal puede además funcionar en un eje de frecuencia deformado para enfatizar, de forma selectiva, ciertos rangos de frecuencia como, por ejemplo, frecuencias bajas, con respecto a otras frecuencias.

La transformación aplicada a la trama de la señal de entrada filtrada es, preferiblemente, una Transformada Discreta del Coseno Modificada (MDCT) que funciona en una segunda longitud de trama variable. El sistema de codificación de audio puede comprender una unidad de control de secuencia de ventana que determina, para un bloque de la señal de entrada, las longitudes de trama para ventanas MDCT superpuestas mediante la minimización de una función de coste de codificación, preferiblemente una entropía perceptual simplista, para todo el bloque de señal de entrada incluidas varias tramas. Por consiguiente, se deriva una segmentación óptima del bloque de señal de entrada en ventanas MDCT que tienen respectivas segundas longitudes de trama. Como resultado, se propone una estructura de codificación de dominio de transformada, incluidos los elementos de codificador de voz, con una trama MDCT de longitud adaptativa como la única unidad básica para todo el procesamiento excepto la CPL. Dado que las longitudes de trama MDCT pueden asumir muchos valores diferentes, puede encontrarse una secuencia óptima y los cambios de tamaño de trama abruptos pueden evitarse, como son comunes en la técnica anterior donde solo un tamaño pequeño de ventana y un tamaño grande de ventana se aplican. Además, las ventanas de transformada de transición que tienen bordes afilados, según su uso en algunos enfoques de la técnica anterior para la transición entre tamaños de ventana pequeño y grande, no son necesarias.

Preferiblemente, las longitudes de ventana MDCT consecutivas cambian, como máximo, por un factor de dos (2) y/o las longitudes de ventana MDCT son valores diádicos. De manera más particular, las longitudes de ventana MDCT puede ser particiones diádicas del bloque de señal de entrada. La secuencia de ventana MDCT se encuentra, por lo tanto, limitada a secuencias predeterminadas que son fáciles de codificar con un número pequeño de bits. Además,

la secuencia de ventana tiene suaves transiciones de tamaños de trama y, de esta manera, excluye cambios abruptos de tamaño de trama.

La unidad de control de secuencia de ventana puede además configurarse para considerar estimaciones de predicción a largo plazo, generadas por la unidad de predicción a largo plazo, para candidatos de longitud de ventana cuando se busca la secuencia de longitudes de ventana MDCT que minimiza la función de coste de codificación para el bloque de señal de entrada. En la presente realización, el bucle de predicción a largo plazo se cierra cuando se determinan las longitudes de ventana MDCT, lo cual resulta en una secuencia mejorada de ventanas MDCT aplicada para la codificación. El sistema de codificación de audio puede además comprender un codificador CPL para codificar, de manera recursiva, a una velocidad variable, frecuencias espectrales de línea u otras representaciones de parámetro CPL apropiadas generadas por la unidad de predicción lineal para el almacenamiento y/o transmisión a un decodificador. Según un ejemplo, se provee una unidad de interpolación de predicción lineal para interpolar parámetros de predicción lineal generados en una velocidad correspondiente a la primera longitud de trama para coincidir con las longitudes de trama variables de la señal de dominio de transformada.

Según un aspecto de la descripción, el sistema de codificación de audio puede comprender una unidad de modelado perceptual que modifica una característica del filtro adaptativo mediante la fluctuación y/o inclinación de un polinomio CPL generado por la unidad de predicción lineal para una trama CPL. El modelo perceptual recibido por la modificación de las características del filtro adaptativo puede usarse para muchos propósitos en el sistema. Por ejemplo, puede aplicarse como una función de ponderación perceptual en la cuantificación o predicción a largo plazo.

Otro aspecto de la descripción se refiere a la predicción a largo plazo (PLP), en particular a la predicción a largo plazo en el dominio MDCT, PLP adaptada a la trama MDCT y búsqueda PLP ponderada de MDCT. Dichos aspectos son aplicables independientemente de si un análisis CPL está presente corriente arriba del codificador de transformada.

Según un ejemplo, el sistema de codificación de audio además comprende una cuantificación inversa y unidad de transformación inversa para generar una reconstrucción de dominio temporal de la trama de la señal de entrada filtrada. Además, puede proveerse una memoria intermedia de predicción a largo plazo para almacenar reconstrucciones de dominio temporal de tramas previas de la señal de entrada filtrada. Dichas unidades pueden disponerse en un bucle de realimentación de la unidad de cuantificación a una unidad de extracción de predicción a largo plazo que busca, en la memoria intermedia de predicción a largo plazo, el segmento reconstruido que mejor concuerda con la trama presente de la señal de entrada filtrada. Además, puede proveerse una unidad de estimación de ganancia de predicción a largo plazo que ajusta la ganancia del segmento seleccionado de la memoria intermedia de predicción a largo plazo de modo que coincida mejor con la trama presente. Preferiblemente, la estimación de predicción a largo plazo se resta de la señal de entrada transformada en el dominio de transformada. Por lo tanto, puede proveerse una segunda unidad de transformada para transformar el segmento seleccionado en el dominio de transformada. El bucle de predicción a largo plazo puede además incluir añadir la estimación de predicción a largo plazo en el dominio de transformada a la señal de realimentación después de la cuantificación inversa y antes de la transformación inversa en el dominio temporal. Por consiguiente, puede usarse un esquema de predicción a largo plazo adaptativo hacia atrás que predice, en el dominio de transformada, la trama presente de la señal de entrada filtrada según tramas previas. Con el fin de ser más eficaz, el esquema de predicción a largo plazo puede además adaptarse de diferentes maneras, según se establece más abajo para algunos ejemplos.

Según un ejemplo, la unidad de predicción a largo plazo comprende un extractor de predicción a largo plazo para determinar un valor de retardo que especifica el segmento reconstruido de la señal filtrada que mejor encaja en la trama actual de la señal filtrada. Un estimador de ganancia de predicción a largo plazo puede calcular un valor de ganancia aplicado a la señal del segmento seleccionado de la señal filtrada. Preferiblemente, el valor de retardo y el valor de ganancia se determinan para minimizar un criterio de distorsión relacionado con la diferencia, en un dominio perceptual, de la estimación de predicción a largo plazo con respecto a la señal de entrada transformada. Un polinomio de predicción lineal modificado puede aplicarse como curva de ganancia de ecualización de dominio MDCT cuando se minimiza el criterio de distorsión.

La unidad de predicción a largo plazo puede comprender una unidad de transformación para transformar la señal reconstruida de segmentos de la memoria intermedia PLP en el dominio de transformada. Para una implementación eficaz de una transformación MDCT, la transformación es, preferiblemente, una Transformación Discreta del Coseno tipo IV.

Otro aspecto de la descripción se refiere a un decodificador de audio para decodificar el tren de bits generado por las realizaciones del codificador de más arriba. Un decodificador según una realización comprende una unidad de descuantificación para descuantificar una trama de un tren de bits de entrada según los factores de escala; una unidad de transformación inversa para transformar, de manera inversa, una señal de dominio de transformada; una

- unidad de predicción lineal para filtrar la señal de dominio de transformada inversamente transformada; y una unidad de decodificación de factor de escala para generar los factores de escala usados en la descuantificación según información delta de factor de escala recibida que codifica la diferencia entre los factores de escala aplicados en el codificador y factores de escala que se generan según los parámetros del filtro adaptativo. El decodificador puede además comprender una unidad de determinación de factor de escala para generar factores de escala según una curva umbral de enmascaramiento que se deriva de los parámetros de predicción lineal para la presente trama. La unidad de decodificación de factor de escala puede combinar la información delta de factor de escala recibida y los factores de escala generados basados en la predicción lineal para generar factores de escala para la entrada en la unidad de descuantificación.
- Un decodificador según otra realización comprende una unidad de descuantificación basada en modelo para descuantificar una trama de un tren de bits de entrada; una unidad de transformación inversa para transformar inversamente una señal de dominio de transformada; y una unidad de predicción lineal para filtrar la señal de dominio de transformada inversamente transformada. La unidad de descuantificación comprende un descuantificador no basado en modelo y un descuantificador basado en modelo.
- Preferiblemente, la unidad de descuantificación comprende al menos un modelo de probabilidad adaptativo. La unidad de descuantificación puede configurarse para adaptar la descuantificación como una función de las características de señales transmitidas.
- La unidad de descuantificación puede además decidir una estrategia de descuantificación según los datos de control para la trama decodificada. Preferiblemente, los datos de control de descuantificación se reciben con el tren de bits o se derivan de los datos recibidos. Por ejemplo, la unidad de descuantificación decide la estrategia de descuantificación según el tamaño de transformada de la trama.
- Según otro aspecto, la unidad de descuantificación comprende puntos de reconstrucción adaptativos.
- La unidad de descuantificación puede comprender descuantificadores escalares uniformes que se configuran para usar dos puntos de reconstrucción de descuantificación por intervalo de cuantificación, en particular un punto medio y un punto de reconstrucción MMSE.
- Según una realización, la unidad de descuantificación usa un cuantificador basado en modelo en combinación con una codificación aritmética.
- Además, el decodificador puede comprender muchos de los aspectos según se describe más arriba para el codificador. En general, el decodificador reflejará las funciones del codificador, aunque algunas funciones solo se llevan a cabo en el codificador y no tendrán componentes correspondientes en el decodificador. Por consiguiente, se considera que aquello que se describe para el codificador es aplicable al decodificador también, si no se establece lo contrario.
- Los aspectos de más arriba de la invención pueden implementarse como un dispositivo, aparato, método o programa de ordenador que funcionan en un dispositivo programable. Los aspectos inventivos pueden además realizarse en señales, estructuras de datos y trenes de bits.
- Por consiguiente, la solicitud además describe un método de codificación de audio y un método de decodificación de audio. Un método de codificación de audio a modo de ejemplo comprende las etapas de: filtrar una señal de entrada según un filtro adaptativo; transformar una trama de la señal de entrada filtrada en un dominio de transformada; cuantificar una señal de dominio de transformada; generar factores de escala, según una curva umbral de enmascaramiento, para su uso en la unidad de cuantificación cuando se cuantifica la señal de dominio de transformada; calcular factores de escala basados en la predicción lineal según parámetros del filtro adaptativo; y codificar la diferencia entre los factores de escala basados en la curva umbral de enmascaramiento y los factores de escala basados en la predicción lineal.
- Otro método de codificación de audio comprende las etapas de: filtrar una señal de entrada según un filtro adaptativo; transformar una trama de la señal de entrada filtrada en un dominio de transformada; y cuantificar la señal de dominio de transformada; en donde la unidad de cuantificación decide, según las características de la señal de entrada, codificar la señal de dominio de transformada con un cuantificador basado en modelo o un cuantificador no basado en modelo.
- Un método de decodificación de audio a modo de ejemplo comprende las etapas de: descuantificar una trama de un tren de bits de entrada según factores de escala; transformar inversamente una señal de dominio de transformada; filtrado de predicción lineal de la señal de dominio de transformada inversamente transformada; calcular segundos factores de escala según los parámetros del filtro adaptativo; y generar los factores de escala usados en la descuantificación según la información de diferencia de factores de escala recibida y los segundos factores de escala estimados.

Otro método de codificación de audio comprende las etapas de: descuantificar una trama de un tren de bits de entrada; transformar inversamente una señal de dominio de transformada; y filtrado de predicción lineal de la señal de dominio de transformada inversamente transformada; en donde la descuantificación usa un cuantificador no basado en modelo y un cuantificador basado en modelo.

- 5 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.

Breve descripción de los dibujos

- 10 La presente invención se describirá ahora mediante ejemplos ilustrativos, sin limitar el alcance de la invención, con referencia a los dibujos anexos, en los cuales:

La Figura 1 ilustra una realización preferida de un codificador y un decodificador según la presente invención;

la Figura 2 ilustra una vista más detallada del codificador y decodificador según la presente invención;

la Figura 3 ilustra otra realización del codificador según la presente invención;

la Figura 4 ilustra una realización preferida del codificador según la presente invención;

- 15 la Figura 5 ilustra una realización preferida del decodificador según la presente invención;

la Figura 6 ilustra una realización preferida de la codificación y decodificación de líneas MDCT según la presente invención;

la Figura 7 ilustra una realización preferida del codificador y decodificador, y ejemplos de datos de control relevantes transmitidos entre sí, según la presente invención;

- 20 la Figura 7a es otra ilustración de aspectos del codificador según una realización de la invención;

la Figura 8 ilustra un ejemplo de una secuencia de ventana y la relación entre datos CPL y datos MDCT según una realización de la presente invención;

la Figura 9 ilustra una combinación de datos de factor de escala y datos CPL según la presente invención;

- 25 la Figura 9a ilustra otra realización de la combinación de datos de factor de escala y datos CPL según la presente invención;

la Figura 9b ilustra otro diagrama de bloques simplificado de un codificador y un decodificador según la presente invención;

la Figura 10 ilustra una realización preferida de traslación de polinomios CPL a una curva de ganancia MDCT según la presente invención;

- 30 la Figura 11 ilustra una realización preferida de mapeo de los parámetros CPL de velocidad de actualización constante hacia los datos de secuencia de ventana MDCT adaptativos, según la presente invención;

la Figura 12 ilustra una realización preferida para adaptar el cálculo de filtro de ponderación perceptual según el tamaño de transformada y tipo de cuantificador, según la presente invención;

- 35 la Figura 13 ilustra una realización preferida de adaptación del cuantificador que depende del tamaño de trama, según la presente invención;

la Figura 14 ilustra una realización preferida de adaptación del cuantificador que depende del tamaño de trama, según la presente invención;

la Figura 15 ilustra una realización preferida de adaptación del tamaño de etapa de cuantificación como una función de datos CPL y PLP, según la presente invención;

- 40 la Figura 15a ilustra cómo una curva delta se deriva de parámetros CPL y PLP mediante un módulo de adaptación delta;

la Figura 16 ilustra una realización preferida de un cuantificador basado en modelo que utiliza desplazamientos aleatorios, según la presente invención;

la Figura 17 ilustra una realización preferida de un cuantificador basado en modelo según la presente invención;

la Figura 17a ilustra otra realización preferida de un cuantificador basado en modelo según la presente invención;

la Figura 17b ilustra, de forma esquemática, un decodificador 2150 de líneas MDCT basado en modelo según una realización de la invención;

5 la Figura 17c ilustra, de forma esquemática, aspectos del preprocesamiento de cuantificador según una realización de la invención;

la Figura 17d ilustra, de forma esquemática, aspectos del cálculo de tamaño de etapa según una realización de la invención;

la Figura 17e ilustra, de forma esquemática, un codificador de entropía restringida basado en modelo según una realización de la invención;

10 la Figura 17f ilustra, de forma esquemática, el funcionamiento de un cuantificador escalar uniforme (USQ, por sus siglas en inglés) según una realización de la invención;

la Figura 17g ilustra, de forma esquemática, cálculos de probabilidad según una realización de la invención;

la Figura 17h ilustra, de forma esquemática, un proceso de descuantificación según una realización de la invención;

la Figura 18 ilustra una realización preferida de un control de depósito de bits, según la presente invención;

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

la Figura 18b ilustra el concepto de un control de depósito de bits para tamaños de trama variables, según la presente invención;

la Figura 18c muestra una curva de control a modo de ejemplo para el control de depósito de bits según una realización;

20 la Figura 19 ilustra una realización preferida del cuantificador inverso mediante el uso de diferentes puntos de reconstrucción, según la presente invención.

Descripción de realizaciones preferidas

25 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 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.

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

35 En la Figura 2 se ilustra una realización preferida 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 de la frecuencia por un módulo MDCT (Transformada Discreta del Coseno Modificada) 202 que funciona en una 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 una realización adicional de la presente invención. 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 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.

40 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. Según se describirá 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 describe en mayor detalle. Este comprende una memoria intermedia PLP 311 que tiene muestras de dominio temporal 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 (a saber, la señal residual después de la resta) se optimiza. Dicha optimización puede llevarse a cabo en un dominio perceptual donde los componentes de frecuencia (a saber, líneas MDCT) se ponderan según 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, a saber, 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 según 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 según 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 temporal 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 (según 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 temporal. En el presente caso, el MSE entre la trama de entrada filtrada CPL y la reconstrucción de dominio temporal correspondiente en la memoria intermedia PLP 411 se minimiza. El MSE se calcula, de forma ventajosa, según el 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 según se describirá 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 temporal por un bloque de síntesis MDCT, y se da forma espectralmente a la señal de dominio temporal 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, a saber, 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 describe una ilustración muy general del sistema de codificación inventivo. 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;

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

El decodificador según 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 según una realización de la invención. 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 proveen un módulo de estimación de altura de tono 750 para calcular la altura de tono de la señal de entrada, y un módulo de determinación de secuencia de 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 según 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 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 el respectivo coste de codificación se calcula. 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 resolución espectral se prefieren para las señales de audio.

Las ponderaciones perceptuales o una función de ponderación perceptual se determinan según los parámetros CPL según 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 funcionando en el dominio MDCT, para la ponderación de errores o contribuciones de distorsión de componentes de frecuencia según su respectiva importancia perceptual. La Figura 7a además ilustra qué parámetros de codificación se transmiten al decodificador, preferiblemente por un esquema de codificación apropiado, según se describe más adelante.

A continuación, se describirán 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.

Según 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 módulos de codificador, después del filtrado, funcionan en la transformada MDCT de la señal espectralmente plana, la presente invención enseña que 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 (a saber, la envolvente espectral de la señal original) hacia una curva de ganancia, o curva de ecualización, que se aplica en los comportamientos de frecuencia de la representación MDCT de la señal espectralmente plana. Por el contrario, el módulo PL puede omitir el filtrado real, y solo calcular una función de transferencia que se mapea posteriormente hacia una curva de ganancia que puede imponerse a la representación MDCT de la señal y, por consiguiente, eliminar la necesidad de filtrado de dominio temporal de la señal de entrada.

Un aspecto prominente de las realizaciones de la presente invención es que un codificador de transformada basado en MDCT funciona mediante el uso de una segmentación de ventana flexible, en una señal blanqueada CPL. Ello se ilustra en la Figura 8, donde se provee 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 (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 velocidad de trama y datos MDCT, en particular las líneas MDCT, generados a una segunda velocidad 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 según se determina por la secuencia de ventana MDCT.

Las flechas hacia arriba simbolizan datos de refinamiento (a saber, 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 calculada a partir de los parámetros CPL. Además, la ponderación perceptual derivada de CPL puede usarse cuando se determina la distorsión de cuantificación. Según se ilustra y según se describe 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, a saber, correspondiente a la trama MDCT o tamaño de ventana.

La Figura 11 ilustra una realización preferida de mapeo de los parámetros CPL de velocidad constante hacia los datos de secuencia de ventana MDCT adaptativos. Un módulo de mapeo CPL 1100 recibe los parámetros CPL según la velocidad de actualización CPL. Además, el módulo de mapeo CPL 1100 recibe información sobre la secuencia de ventana MDCT. Este entonces genera un mapeo CPL hacia MDCT, p.ej., para mapear datos psicoacústicos basados en CPL hacia las respectivas tramas MDCT generadas a la velocidad de trama 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 describen mediante referencia a la Figura 9. El módulo CPL 901 se adapta en una realización de la presente invención para producir una señal de salida blanca, mediante el uso de una predicción lineal de, p.ej., orden 16 para una señal de velocidad de muestreo de 16 kHz. Por ejemplo, la salida del módulo CPL 201 en la Figura 2 es la residual después de la estimación de parámetro CPL y filtrado. El polinomio CPL estimado $A(z)$, según se visualiza esquemáticamente en la parte izquierda inferior de la Figura 9, puede fluctuar por un factor de expansión de ancho de banda, y también inclinarse, en una implementación de la invención, 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 adentro hacia el círculo de la unidad y, por consiguiente, resulta en 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 la etapa de cuantificación (y, por consiguiente, el ruido introducido por la cuantificación). Dichos factores de escala se calculan 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 calculada 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 calcula a partir de las líneas MDCT generadas por el módulo de transformación. La parte derecha inferior 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 corriente 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. Por consiguiente, una curva de ganancia de equalización de dominio MDCT generada para compensar el blanqueamiento del espectro puede usarse cuando se calcula la curva umbral de enmascaramiento y/o los factores de escala. Ello se debe a que los factores de escala necesitan calcularse en una señal que tenga propiedades de espectro absolutas de la señal original, con el fin de calcular perceptualmente, de forma correcta, el enmascaramiento. El cálculo de la curva de ganancia de equalización de dominio MDCT a partir del polinomio CPL se describe en mayor detalle con referencia a la Figura 10 más abajo.

Una realización del esquema de estimación de factor de escala descrito más arriba se describe en la Figura 9a. En la presente realización, la señal de entrada se ingresa en el módulo PL 901 que calcula 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 según su uso posterior 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 (según se representa en la Figura 14). La curva de ganancia calculada 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 de ganancia se ingresan en el módulo de estimación de factor de escala 960 que calcula los factores de escala para la señal de entrada.

Mediante el uso del enfoque descrito 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 calcula 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 según se usan normalmente en códecs de audio perceptuales basados en transformada se calculan en la señal de entrada, o pueden calcularse 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 (según 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, según se describirá más abajo, con el fin de reducir la velocidad binaria 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 estos se calculan 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. Según 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 según su uso común 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 etapa de cuantificación solamente de los datos de predicción lineal.

En la Figura 9b, se provee un diagrama de bloques simplificado de codificador y decodificador según 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 de la 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 etapa de cuantificación deseado para diferentes partes del espectro MDCT. Los valores que determinan el tamaño de etapa 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.

Según un aspecto de la invención, la curva de enmascaramiento perceptual calculada a partir de los parámetros CPL, según se explica con referencia a la Figura 9, se usa cuando se codifican los factores de escala usados en la cuantificación. Otra posibilidad para calcular 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 dicha estimación de energía, un modelo psicoacústico, según su uso 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 según su uso común 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 según la banda de factor de escala, según 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. Según la velocidad binaria, el tamaño de trama u otros parámetros, la cantidad de residual de factor de escala que se transmitirá puede seleccionarse. Con el fin de tener un 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. Según se representa 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 ecualización 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 según se representa en la Figura 3, o cuando se calcula una curva de enmascaramiento perceptual para la determinación de factor de escala según se describe con referencia a la Figura 9 más arriba.

La Figura 12 ilustra una realización preferida de adaptación del cálculo de filtro de ponderación perceptual según el tamaño de transformada y/o el tipo de cuantificador. El polinomio PL $A(z)$ se calcula por el módulo CPL 1201 en la Figura 16. Un módulo de modificación de parámetro CPL 1271 recibe parámetros CPL como, por ejemplo, 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 p e γ . Estos se modifican dadas las normas predeterminadas, según el tamaño de transformada usado y/o la estrategia de cuantificación Q usada. Los parámetros de fluctuación e inclinación modificados p' e 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 según una realización de la invención. Un aspecto de la presente invención es que utiliza diferentes estrategias de cuantificación para diferentes tamaños de transformada o tamaños de trama. Ello 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. Debe notarse que el presente 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.

Según un aspecto independiente de la presente invención, 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 tomar ventaja del carácter "disperso" (a saber, tonos discretos bien definidos) en el espectro de la señal. Un método de cuantificación según su uso en AAC en combinación con tablas de Huffman y el agrupamiento de líneas espectrales, también según su uso 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 provee 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.

Según otro aspecto de la invención, el tamaño de etapa de cuantificador se adapta como función de datos CPL y/o PLP. Ello permite una determinación del tamaño de etapa 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 provee 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 provee una visualización esquemática de líneas MDCT. Más abajo, se representa la delta de tamaño de etapa de cuantificación Δ como una función de frecuencia. Está claro a partir del presente ejemplo particular que el tamaño de etapa de cuantificación aumenta con la frecuencia, a saber, 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 según se explica con referencia a la Figura 13.

Una función de ponderación perceptual preferida derivada de los datos CPL se provee 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 $A(z)$. Debe notarse 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 describen aspectos de un cuantificador basado en modelo según una realización de la presente invención. 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 desplazan a los límites del intervalo. El cuantificador propuesto provee 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. Ello es particularmente importante para velocidades binarias 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) según una realización de la invención. 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 calcularse 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 la codificación de velocidad fija o variable y se transmite al decodificador.

Un módulo de control de velocidad 1710 puede emplearse para controlar el número de bits usados para codificar una trama MDCT. Un índice de control de velocidad controla el número de bits usados. El índice de control de velocidad señala una lista de tamaños de etapa de cuantificador nominales. La tabla puede ordenarse con tamaños de etapa en orden descendente (es preciso ver la Figura 17g).

El codificador MBMLQ se ejecuta con un conjunto de diferentes índices de control de velocidad, y el índice de control de velocidad que produce un cómputo 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 velocidad 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 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 etapa, el índice de control de velocidad puede codificarse mediante el uso de 2 bits por trama MDCT en promedio.

La Figura 17 además ilustra, de forma esquemática, el decodificador MBMLQ 1750 donde la trama MDCT tiene la ganancia renormalizada si una ganancia loca se ha calculado en el codificador 1700.

5 La Figura 17a ilustra, de forma esquemática, el codificador de líneas MDCT basado en modelo 1700 según una realización en mayor detalle. La figura comprende un módulo de preprocesamiento de cuantificador 1730 (es preciso ver la Figura 17c), un codificador de entropía restringida basado en modelo 1740 (es preciso ver la Figura 17e), y 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, según la
10 trama de la 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 parámetros de control: un tamaño de etapa de cuantificador Δ (delta, longitud de intervalo), un conjunto de cálculos 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.

La Figura 17b ilustra, de forma esquemática, un decodificador de líneas MDCT basado en modelo 1750 según una realización de la invención. Este toma como entrada bits de información conexa del tren de bits y los decodifica en
20 parámetros que se ingresan en el módulo de preprocesamiento de cuantificador 1760 (es preciso ver 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 módulo de preprocesamiento de cuantificador 1760 produce un conjunto de parámetros de control (igual que en el
25 codificador 1700) y estos se ingresan en el módulo de cálculos de probabilidad 1770 (es preciso ver la Figura 17g; igual que en el codificador, es preciso ver la Figura 17e) y en el módulo de descuantificación 1780 (es preciso ver la Figura 17h; igual que en el codificador, es preciso ver 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
30 cualquier 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.

La Figura 17c ilustra, de manera esquemática, aspectos del preprocesamiento de cuantificador según una realización de la invención que consiste en i) cómputo de tamaño de etapa, ii) modificación de curva de
35 enmascaramiento perceptual, iii) estimación de varianza de líneas MDCT, iv) construcción de tabla de desplazamiento.

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

40 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 etapa puede aumentarse cuando la energía de señal se convierte en baja según se representa en la Figura 17d-ii) donde se muestra una curva a modo de
45 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 etapa 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 etapa L .

50 Los sonidos paso alto son perceptualmente menos importantes que los sonidos paso bajo. La función de adaptación de paso alto aumenta el tamaño de etapa cuando la trama MDCT es de paso alto, a saber, 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 etapa. El presente mecanismo se representa en la Figura 17d-iii) donde r es el 1^{er} coeficiente de reflexión de CPL. La adaptación de
55 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 (según la envolvente CPL). Después de la ponderación perceptual en el codificador de entropía restringida basado en modelo 1740 (es preciso ver 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 ilustra 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 (es preciso ver la Figura 17e). La función de construcción de tabla de desplazamiento escala la tabla de desplazamiento nominal con la etapa de tamaño 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.

Según se describe más abajo, los desplazamientos proveen 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 etapa 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 según

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

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.

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

La probabilidad de intervalo de distorsión mínima de cada USQ se computa en el módulo de cálculos de probabilidad 1770 (es preciso ver 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

5 $k_3 \cdot \sqrt{v_j}$, 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 etapa de cuantificación en magnitud.

Un valor de reconstrucción escalar para cada línea MDCT se computa por el módulo de descuantificación 1780 (es preciso ver la Figura 17h) que produce el vector MDCT cuantificado $\bar{y}$. En el módulo de optimización RD 1790 se

10 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., según 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.

En el módulo de optimización RD 1790, se computa un coste C , preferiblemente según 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

15 probabilidades se producen desde el codificador de entropía restringida basado en modelo 1780.

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 el mismo cálculo de varianza fija V para cada vector de desplazamiento que se prueba en la optimización RD, el vector de cálculo de varianza V puede

20 variar. Para el vector de fila de desplazamiento m , puede entonces usarse un cálculo de varianza $k_m \cdot V$ donde k_m puede abarcar, por ejemplo, el rango de 0,5 a 1,5 mientras m varía de $m=1$ 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.

25 Las líneas MDCT descuantificadas pueden refinarse más mediante el uso de un cuantificador residual según se ilustra en la Figura 17e. El cuantificador residual puede ser, p.ej., un cuantificador de vector aleatorio de velocidad fija.

La función 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

30 índice i_n . Las marcas 'x' indican el centro (punto medio) de los intervalos de cuantificación con tamaño de etapa Δ . El origen del cuantificador escalar se desplaza por el desplazamiento o_n del vector de desplazamiento $O = \{o_1, o_2, \dots, o_n, \dots, o_N\}$. Por consiguiente, los límites de intervalo y los puntos medio se desplazan por el desplazamiento.

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

35 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 proveen variación en las tablas de probabilidad que se computan por el módulo de cálculos de probabilidad 1770, lo cual lleva a una codificación de entropía más eficaz de los índices de líneas MDCT (a saber, se requieren menos bits).

El uso de un tamaño de etapa variable Δ (delta) permite la exactitud variable en la cuantificación de modo que más

40 exactitud puede usarse para sonidos perceptualmente importantes, y menos exactitud puede usarse para sonidos perceptualmente menos importantes.

La Figura 17g ilustra, de manera esquemática, los cálculos de probabilidad en el módulo de cálculos de probabilidad 1770. Las entradas en dicho módulo son el modelo estadístico aplicado para las líneas MDCT, el tamaño de etapa de cuantificador Δ , el vector de varianza V , el índice de desplazamientos, y la tabla de desplazamientos. La salida

45 del módulo de cálculos de probabilidad 1770 son las tablas cdf. Para cada línea MDCT x_j se evalúa el modelo estadístico (a saber, 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 según se lleva a cabo, p.ej., en el

50 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. Dicho problema puede mitigarse mediante la decodificación para preservar la varianza según 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.

- 5 La decodificación para preservar la varianza según una realización de la invención se logra mediante la determinación del punto de reconstrucción según la siguiente ecuación:

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

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} : $\chi = f(g_{\text{PLP}})$. 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 adelante y/o hacia atrás.

Otro aspecto de la invención relacionado con los puntos de reconstrucción modificados del cuantificador se ilustra, de manera esquemática, en la Figura 19 donde se ilustra un cuantificador inverso según su uso en el decodificador de una realización. El módulo tiene, aparte de las entradas normales de un cuantificador inverso, a saber, las líneas cuantificadas e información sobre el tamaño de etapa de cuantificación (tipo de cuantificación), también información sobre el punto de reconstrucción del cuantificador. El cuantificador inverso de la presente realización puede usar múltiples tipos de puntos de reconstrucción cuando determina un valor reconstruido $\hat{y}_n$ del índice de cuantificación i_n correspondiente. Según se menciona más arriba, los valores de reconstrucción $\hat{y}$ se usan además, p.ej., en el codificador de líneas MDCT (es preciso ver 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 (es preciso ver 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 de la presente invención, 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., según las propiedades de la señal como, por ejemplo, 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 la distorsión mínima. Dada la señal, el sistema puede entonces adaptar el punto de reconstrucción a donde se provee el mejor compromiso.

La presente invención además incorpora un nuevo formato de codificación de secuencia de ventana. Según 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 velocidad 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 según se define por una realización, a saber, limitadas a tamaños 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, a saber, 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, a saber, 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.

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

- Según una realización de la presente invención, el retardo PLP y la ganancia PLP se codifican de manera que la velocidad 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 un retardo PLP y una codificación de ganancia PLP de velocidad variable.

De manera similar, una realización de la presente invención toma ventaja de un depósito de bits y codificación de velocidad variable también para la codificación de los parámetros PL. Además, la codificación PL recursiva se enseña por la presente invención.

- 25 Otro aspecto de la presente invención es el manejo de un depósito de bits para tamaños de trama variables en el codificador. En la Figura 18 se describe una unidad de control de depósito de bits 1800 según la presente invención. Además de una medida de dificultad provista 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 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 según se describirá más abajo.

- 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 velocidad binaria 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. Por consiguiente, 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 la adición del número de bits que representan la velocidad binaria 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.

- En la Figura 18a se representa el concepto básico del control de depósito de bits. El codificador provee medios para calcular cuán difícil 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. Según una línea de control dada, más bits que los correspondientes a una velocidad binaria 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. Para tramas con una dificultad más alta, la línea de control puede desplazarse hacia arriba, y tener el efecto de que las tramas difíciles de codificar pueden usar más bits en el mismo nivel de depósito de bits. Por consiguiente, en aras de la facilidad de codificación de tramas, el número de bits permitido para una trama será más bajo simplemente desplazando hacia abajo la línea de control en la Figura 18a del caso de dificultad promedio al caso de dificultad fácil. Otras modificaciones diferentes del desplazamiento simple de la línea de control son también posibles. Por ejemplo, según se muestra en la Figura 18a, la pendiente de la curva de control puede cambiarse dependiendo de la dificultad de la trama.

- 55 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 según 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.

- 5 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 velocidad binaria 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. Dicha 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.

- 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 cómputo de bits de una cuantificación con tamaño de etapa fijo como se realiza en la parte ECQ de un codificador según una realización de la presente invención. Dichos valores pueden normalizarse con respecto a los tamaños de trama variables, los cuales pueden acompañarse de una simple división por la longitud de trama, y el resultado será una EP respectivamente, un cómputo de bits por muestra. Otra etapa de normalización puede tener lugar 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.

- Otro aspecto de la invención se refiere a los detalles específicos del manejo del depósito de bits para ECQ. La gestión del depósito de bits para ECQ funciona en la suposición de que ECQ produce una calidad aproximadamente constante cuando usa un tamaño de etapa de cuantificador constante para la codificación. El tamaño de etapa de cuantificador constante produce una velocidad variable y el objetivo del depósito de bits es mantener la variación en el tamaño de etapa 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 velocidad producida por el ECQ, la información adicional (p.ej., ganancia y retardo PLP) se transmite según la trama MDCT. La información adicional, en general, también se codifica con entropía y, por consiguiente, consume una velocidad diferente de trama a trama.

- 35 La presente invención se refiere además a una estrategia de cuantificación según el tamaño de trama de transformada. Además, se propone un cuantificador de restricción de entropía basado en modelo que emplea una codificación aritmética. Además, se provee la inserción de desplazamientos aleatorios en un cuantificador escalar uniforme. La invención además sugiere un cuantificador basado en modelo, p.ej., un Cuantificador de Restricción de Entropía (ECQ), mediante el empleo de la codificación aritmética.

40

REIVINDICACIONES

1. El sistema de codificación de audio (200, 300, 400, 700) que comprende:
 - una unidad de predicción lineal (201, 401, 701) para filtrar una señal de entrada según un filtro adaptativo;
 - una unidad de transformación (202, 302, 402, 702) para transformar una trama de la señal de entrada filtrada en una señal de dominio de transformada; y
 - una unidad de cuantificación (203, 303, 403, 703) para cuantificar la señal de dominio de transformada, caracterizado por que la unidad de cuantificación (203, 303, 403, 703) decide, según la estacionariedad de la señal de entrada, codificar la señal de dominio de transformada con un cuantificador basado en modelo o un cuantificador no basado en modelo, el modelo siendo un modelo de probabilidad o un modelo estadístico.
2. El sistema de codificación de audio según la reivindicación 1, en donde el modelo en el cuantificador basado en modelo es adaptativo y variable con el tiempo.
3. El sistema de codificación de audio según la reivindicación 1 o 2, que comprende
 - una unidad de predicción a largo plazo (205, 310, 705) para determinar una estimación de la trama de la señal de entrada filtrada según una reconstrucción de un segmento previo de la señal de entrada filtrada, y
 - una unidad de control de tamaño de etapa de cuantificación para determinar tamaños de etapa de cuantificación para componentes de la señal de dominio de transformada según la predicción lineal y parámetros de predicción a largo plazo.
4. El sistema de codificación de audio según la reivindicación 3, en donde los tamaños de etapa de cuantificación se determinan según la frecuencia, y la unidad de control de tamaño de etapa de cuantificación determina los tamaños de etapa de cuantificación según al menos uno de: un polinomio del filtro adaptativo, un parámetro de control de velocidad de codificación, un valor de ganancia de predicción a largo plazo, y una varianza de señal de entrada.
5. El sistema de codificación de audio según cualquiera de las reivindicaciones 1 a 4, en donde la unidad de cuantificación (203, 303, 403, 703) comprende cuantificadores escalares uniformes para cuantificar componentes de señal de dominio de transformada, cada cuantificador escalar aplicando una cuantificación uniforme, según un modelo de probabilidad, a una línea de Transformada Discreta del Coseno Modificada generada por la unidad de transformada (202, 302, 402, 702).
6. El sistema de codificación de audio según la reivindicación 5, en donde la unidad de cuantificación (203, 303, 403, 703) comprende un cuantificador residual para cuantificar una señal de cuantificación residual que resulta de los cuantificadores escalares uniformes.
7. El sistema de codificación de audio según cualquiera de las reivindicaciones 5 a 6, en donde la unidad de cuantificación (203, 303, 403, 703) comprende una unidad de punto de reconstrucción dinámica que determina un punto de reconstrucción de cuantificación según una interpolación entre un punto central de modelo de probabilidad y un punto de error cuadrático medio mínimo.
8. El sistema de codificación de audio según cualquiera de las reivindicaciones 5 a 7, en donde la unidad de cuantificación (203, 303, 403, 703) aplica una ponderación perceptual en el dominio de transformada cuando determina la distorsión de cuantificación, las ponderaciones perceptuales derivándose de parámetros de predicción lineal.
9. Un decodificador de audio (210, 500) que comprende:
 - una unidad de descuantificación (211, 511) para descuantificar una trama de un tren de bits de entrada en un dominio de transformada;
 - una unidad de transformación inversa (212, 512) para transformar una señal de dominio de transformada en una señal de dominio temporal; y
 - una unidad de predicción lineal (213, 513) para filtrar la señal de dominio temporal; caracterizado por que la unidad de descuantificación (211, 511) comprende un descuantificador no basado en modelo y un descuantificador basado en modelo, el modelo siendo un modelo de probabilidad o un modelo estadístico.
10. El decodificador de audio (210, 500) de la reivindicación 9, en donde la unidad de descuantificación (211, 511) decide una estrategia de descuantificación según los datos de control para la trama.

11. El decodificador de audio (210, 500) de la reivindicación 10, en donde los datos de control de descuantificación se reciben con el tren de bits o se derivan de los datos recibidos.
12. El decodificador de audio (210, 500) de cualquiera de las reivindicaciones 9 a 11, en donde la unidad de descuantificación (211, 511) aplica puntos de reconstrucción adaptativos para la descuantificación de la trama, y la
5 unidad de descuantificación (211, 511) comprende descuantificadores escalares uniformes que se configuran para usar dos puntos de reconstrucción de descuantificación por intervalo de cuantificación, en particular un punto medio y un punto de reconstrucción de error cuadrático medio mínimo.
13. El decodificador de audio (210, 500) de la cualquiera de las reivindicaciones 9 a 12, en donde la unidad de descuantificación (211, 511) comprende al menos un modelo de probabilidad adaptativo.
- 10 14. El decodificador de audio (210, 500) de la cualquiera de las reivindicaciones 9 a 13, en donde la unidad de descuantificación (211, 511) usa un cuantificador basado en modelo en combinación con la codificación aritmética.
- 15 15. El decodificador de audio (210, 500) de la cualquiera de las reivindicaciones 9 a 14, en donde la unidad de descuantificación (211, 511) se configura para adaptar la estrategia de descuantificación como una función de las características de señal transmitida.
- 15 16. El método de codificación de audio que comprende las etapas de:
filtrar una señal de entrada según un filtro adaptativo;
transformar una trama de la señal de entrada filtrada en una señal de dominio de transformada;
cuantificar la señal de dominio de transformada; y
caracterizado por que la señal de dominio de transformada se codifica con un cuantificador basado en modelo o un
20 cuantificador no basado en modelo según la estacionariedad de la señal de entrada, el modelo siendo un modelo de probabilidad o un modelo estadístico.
17. El método de decodificación de audio que comprende las etapas de:
descuantificar una trama de un tren de bits de entrada en un dominio de transformada;
transformar una señal de dominio de transformada en una señal de dominio temporal; y
25 el filtrado de predicción lineal de la señal de dominio temporal;
caracterizado por la decisión de usar un cuantificador basado en modelo o no basado en modelo para la descuantificación de la trama, el modelo siendo un modelo de probabilidad o un modelo estadístico.
18. El programa de ordenador para hacer que un dispositivo programable lleve a cabo un método de codificación o decodificación de audio según la reivindicación 16 o 17.

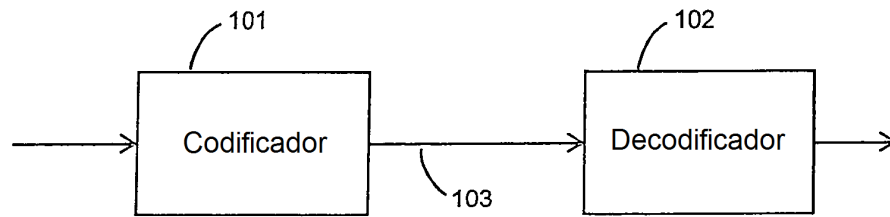


Fig. 1

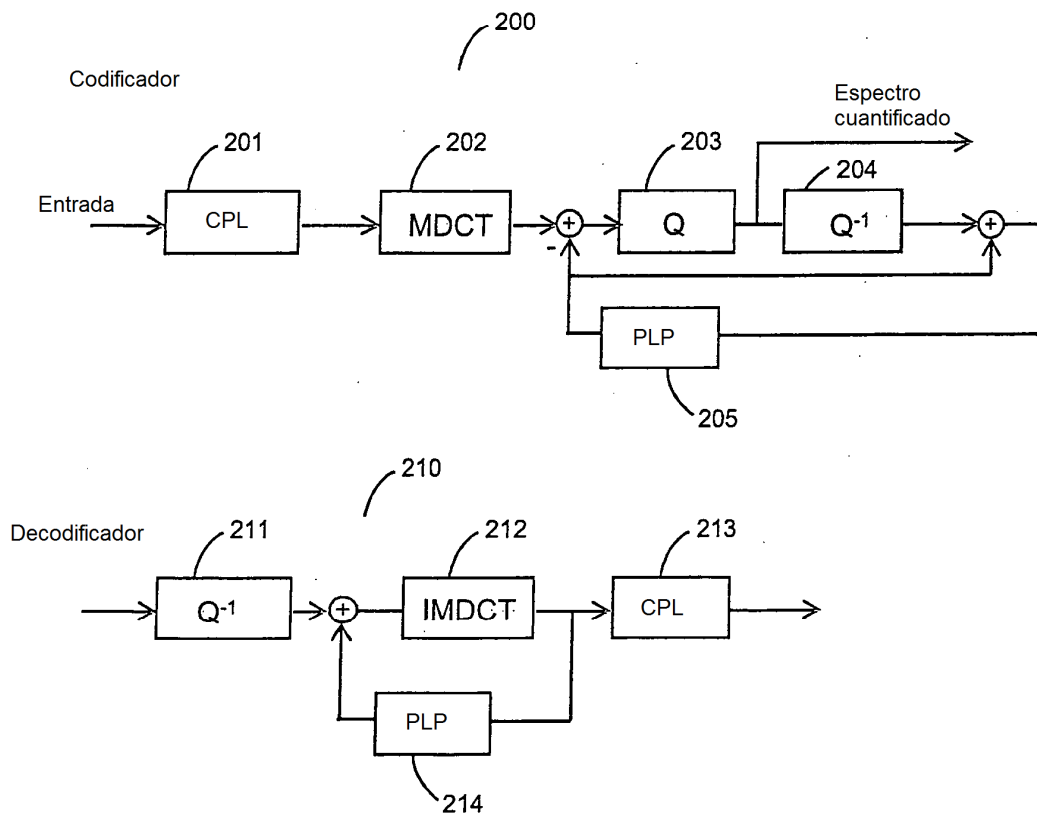


Fig. 2

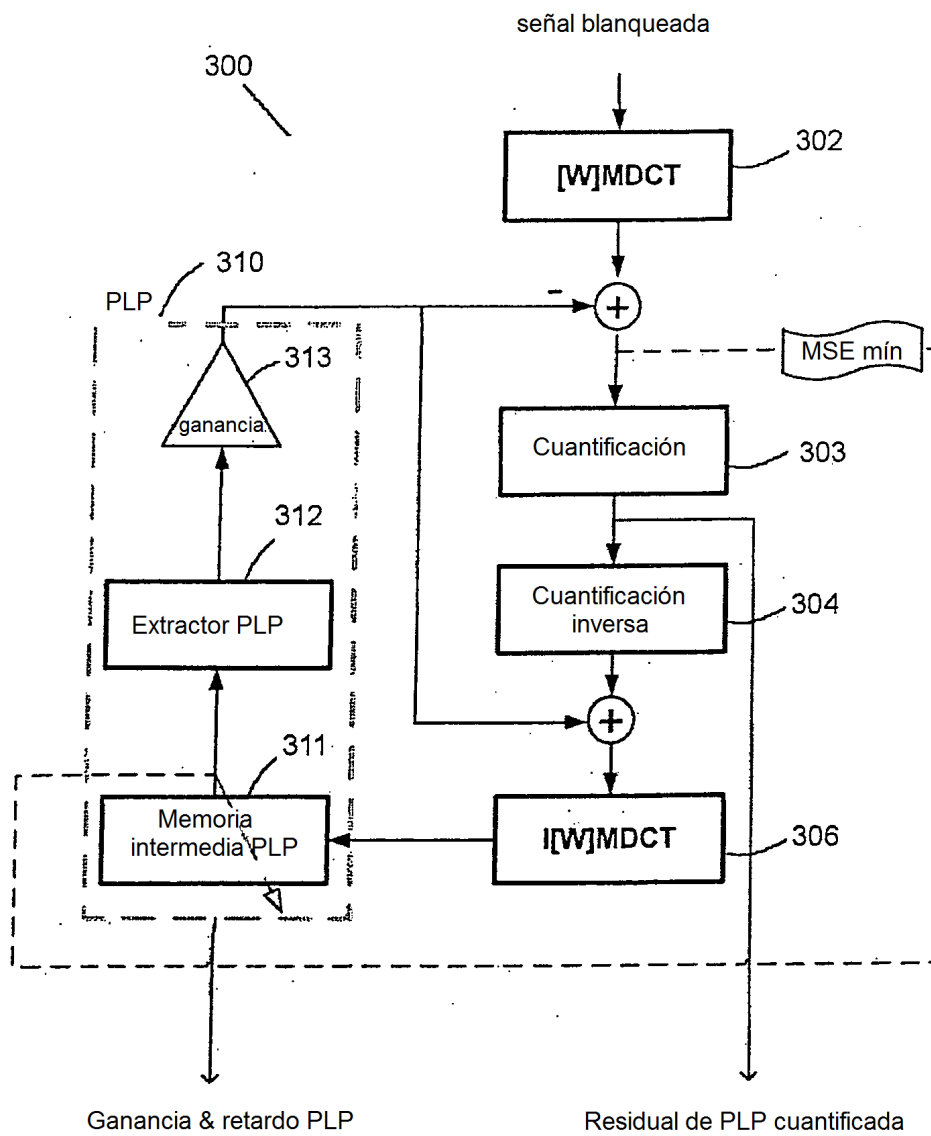


Fig. 3

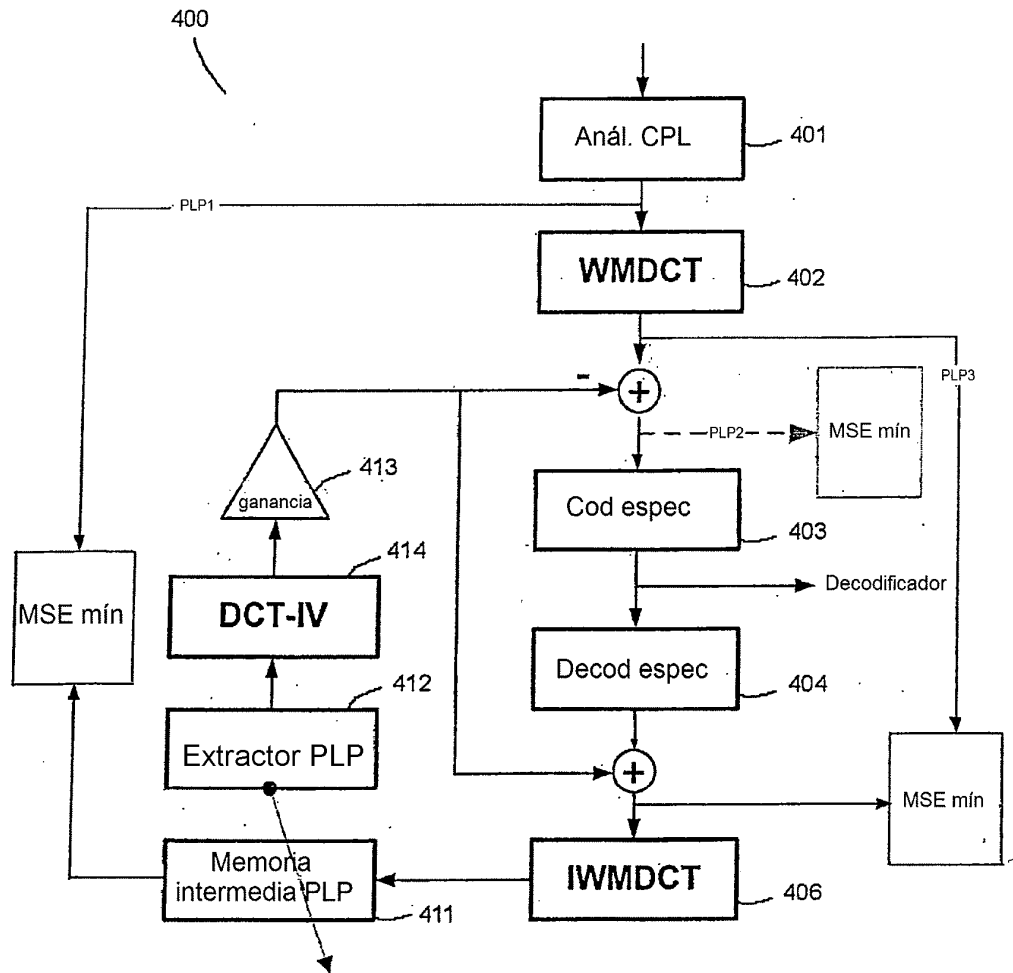


Fig. 4

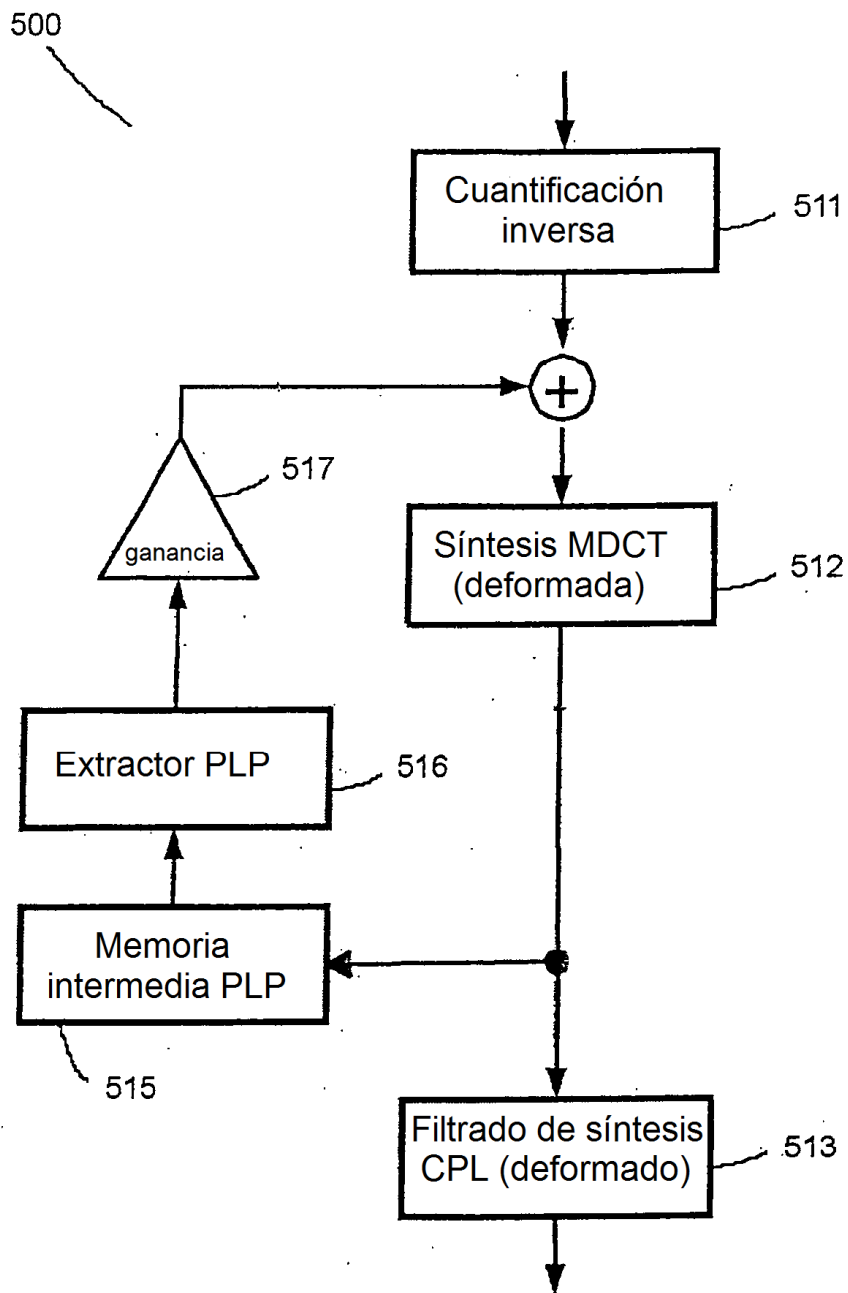


Fig. 5

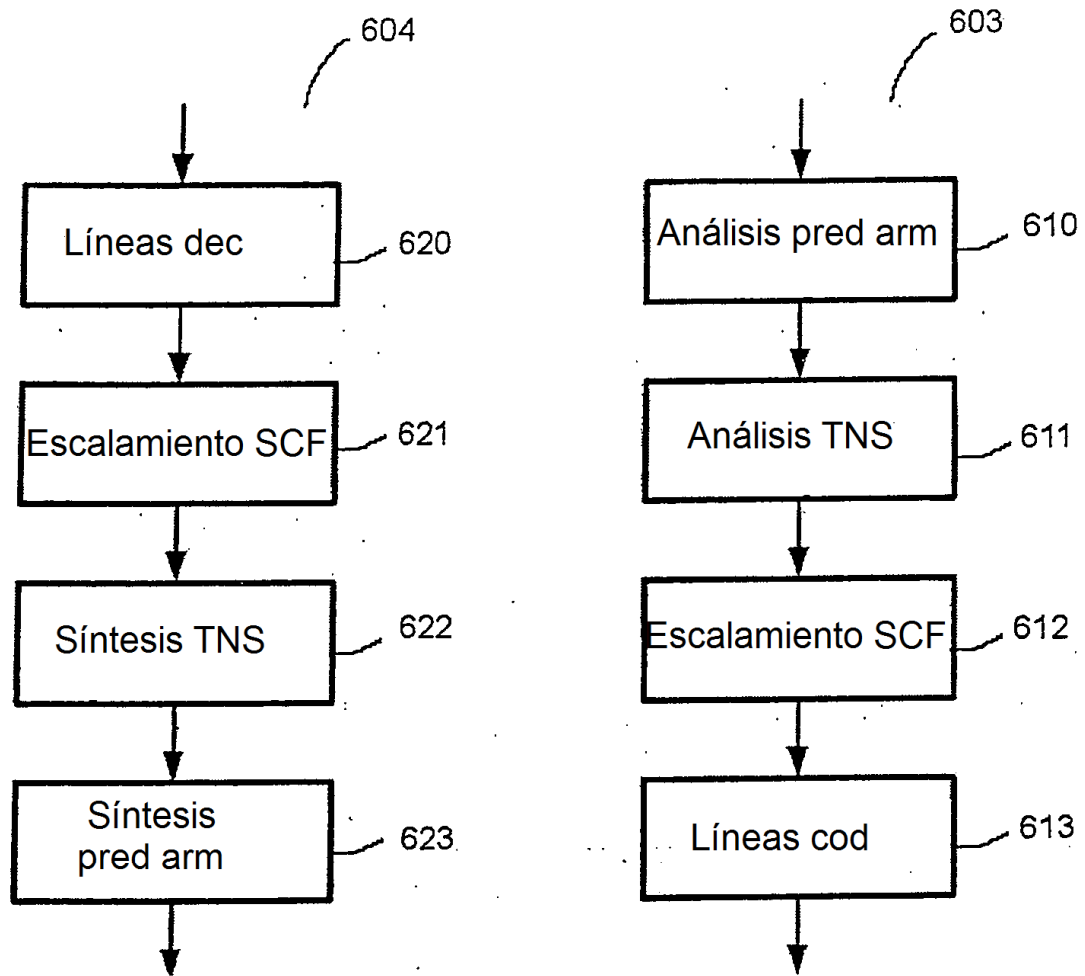


Fig. 6

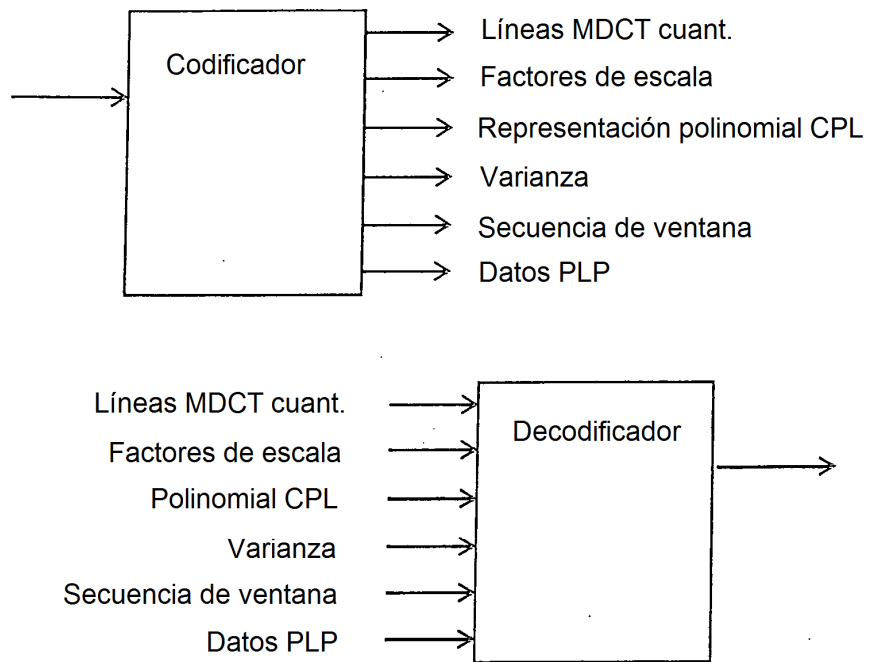


Fig. 7

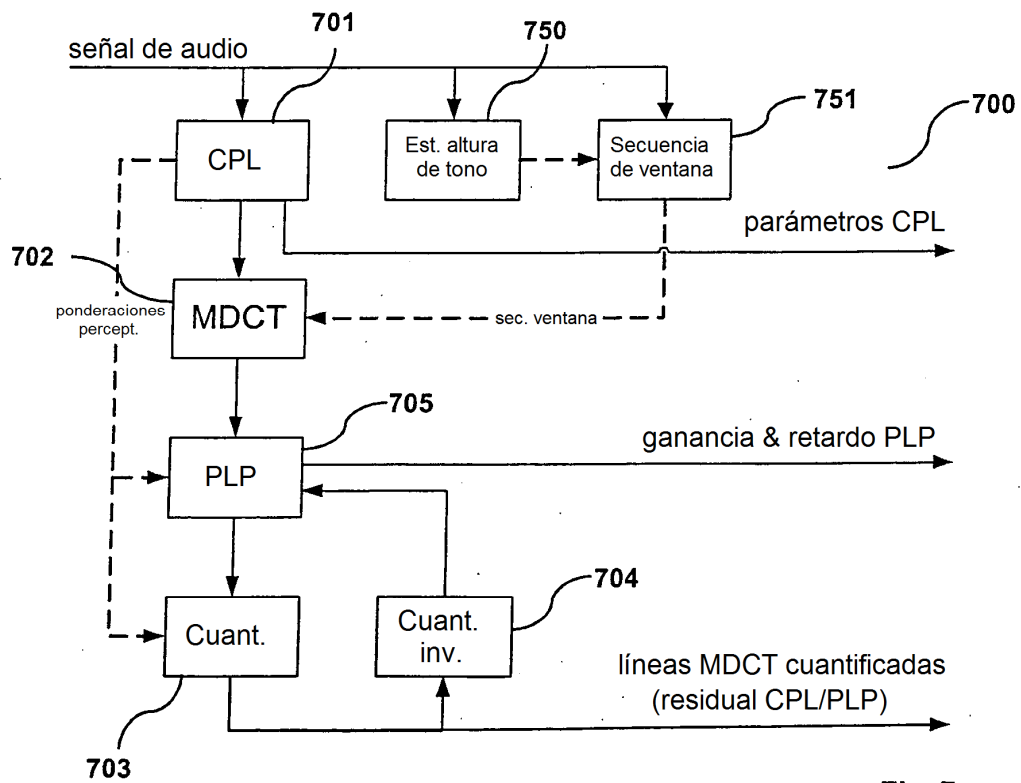


Fig. 7a

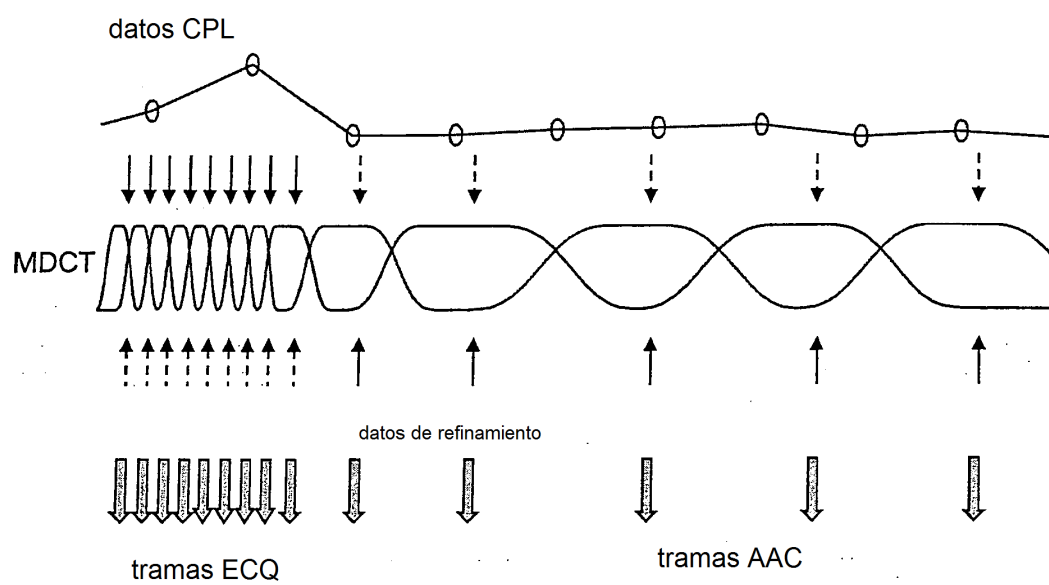


Fig. 8

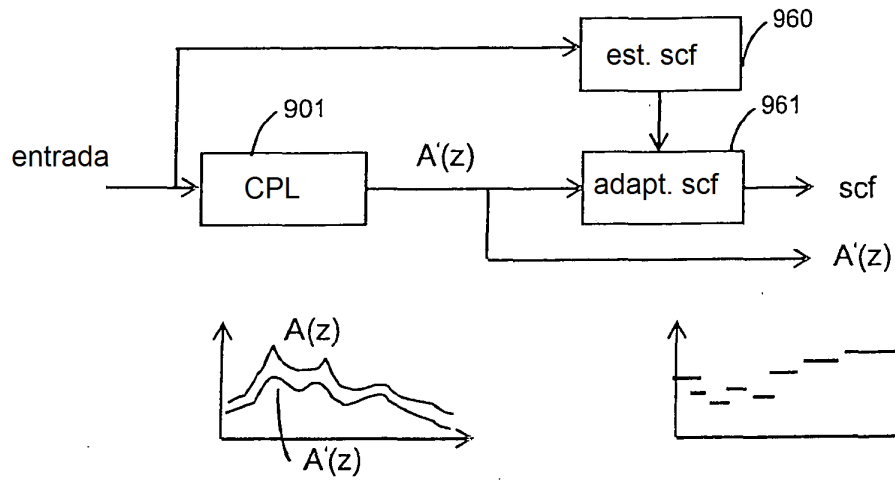


Fig. 9

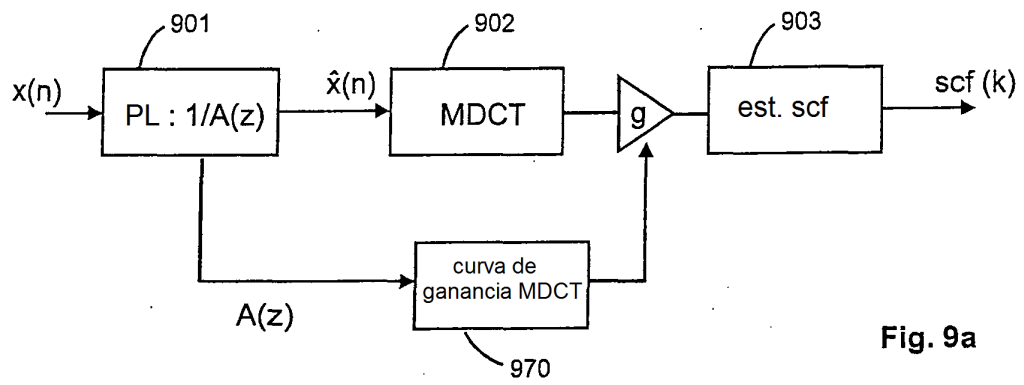


Fig. 9a

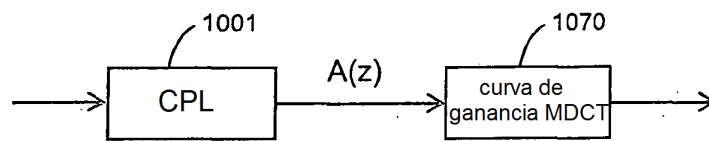


Fig. 10

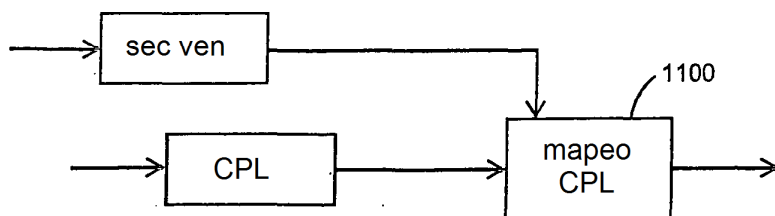


Fig. 11

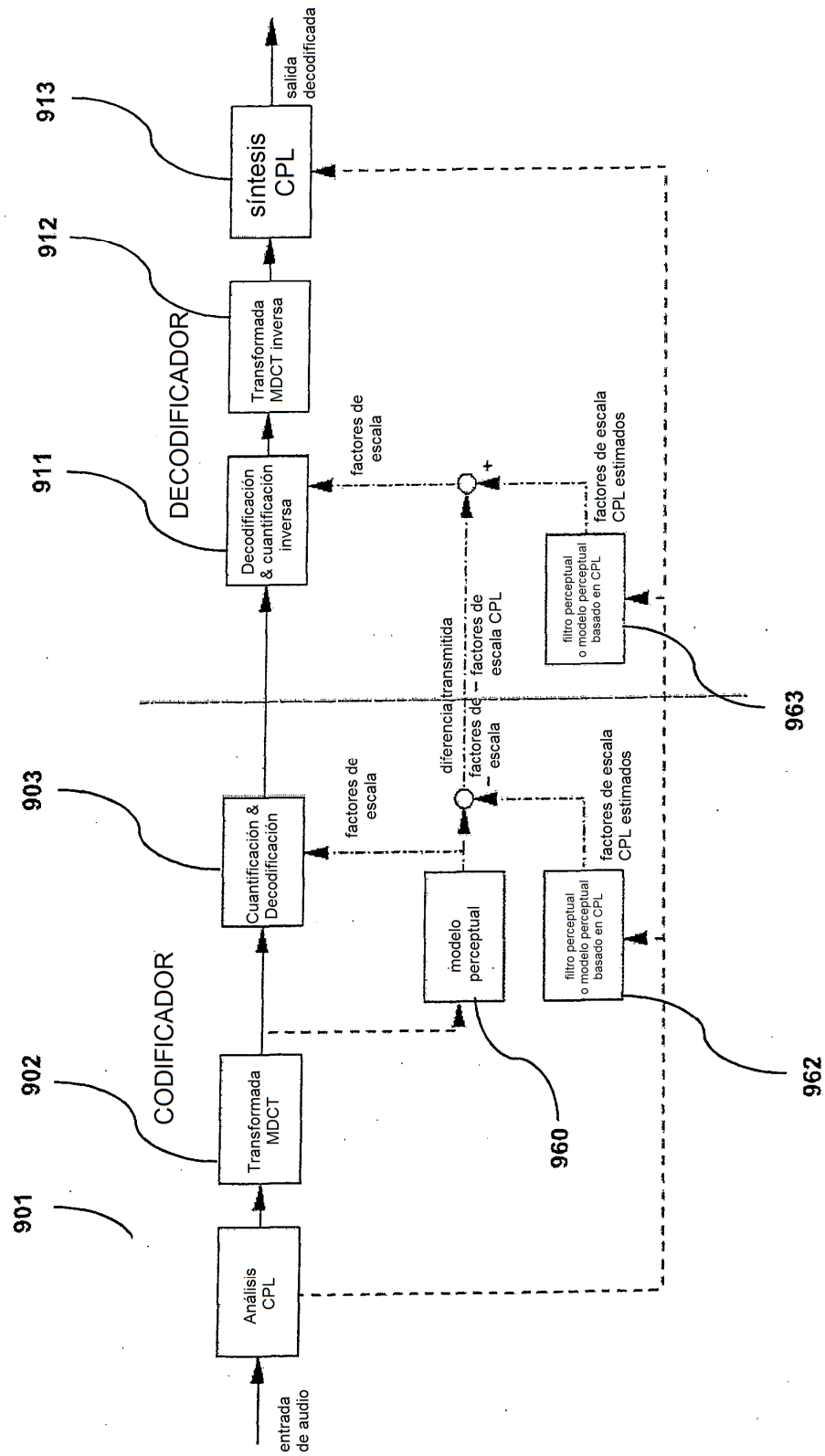


Fig. 9b

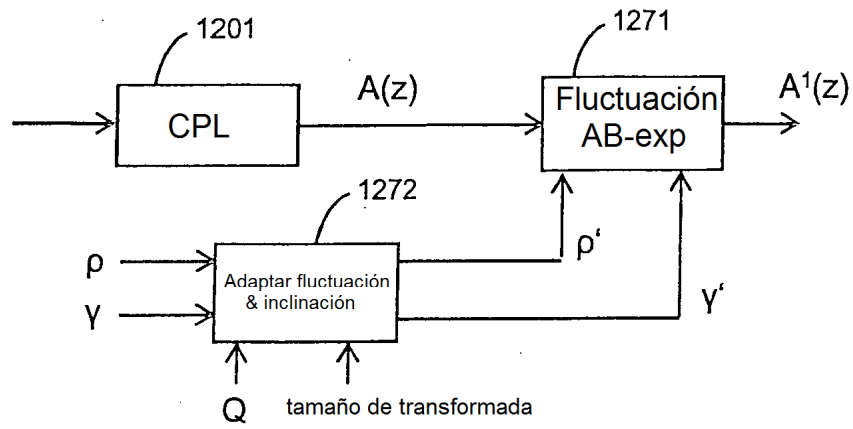


Fig. 12

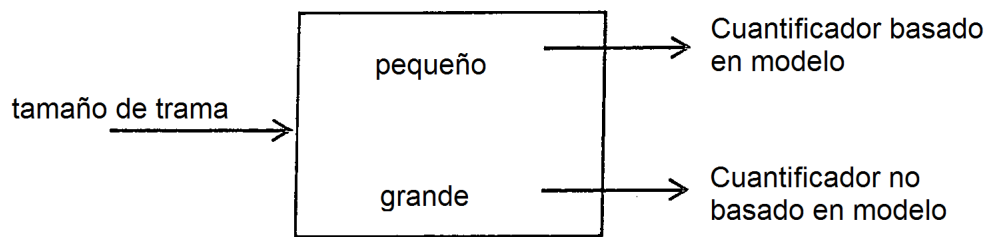


Fig. 13

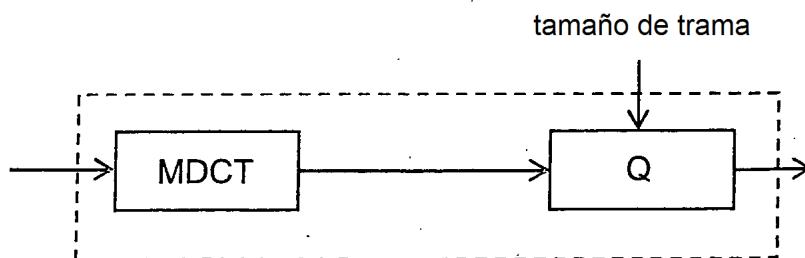


Fig. 14

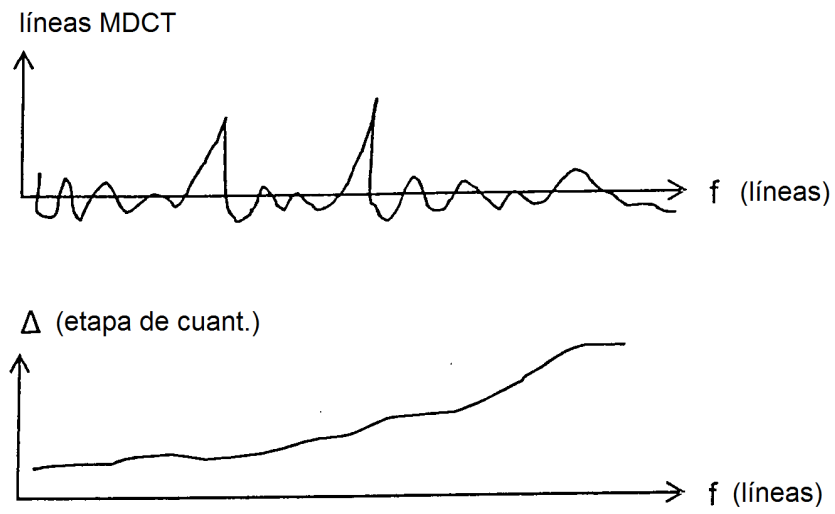


Fig. 15

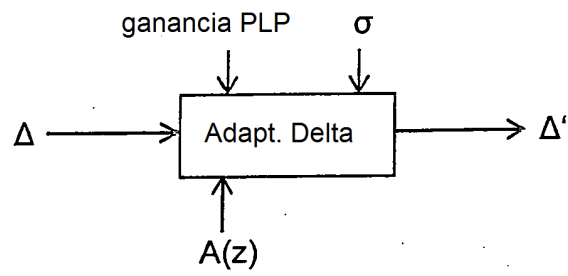


Fig. 15a

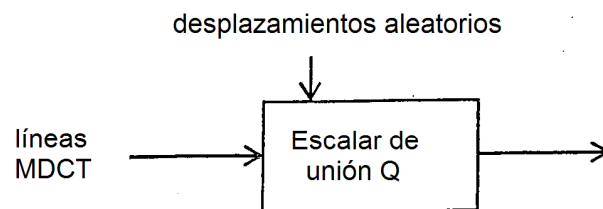


Fig. 16

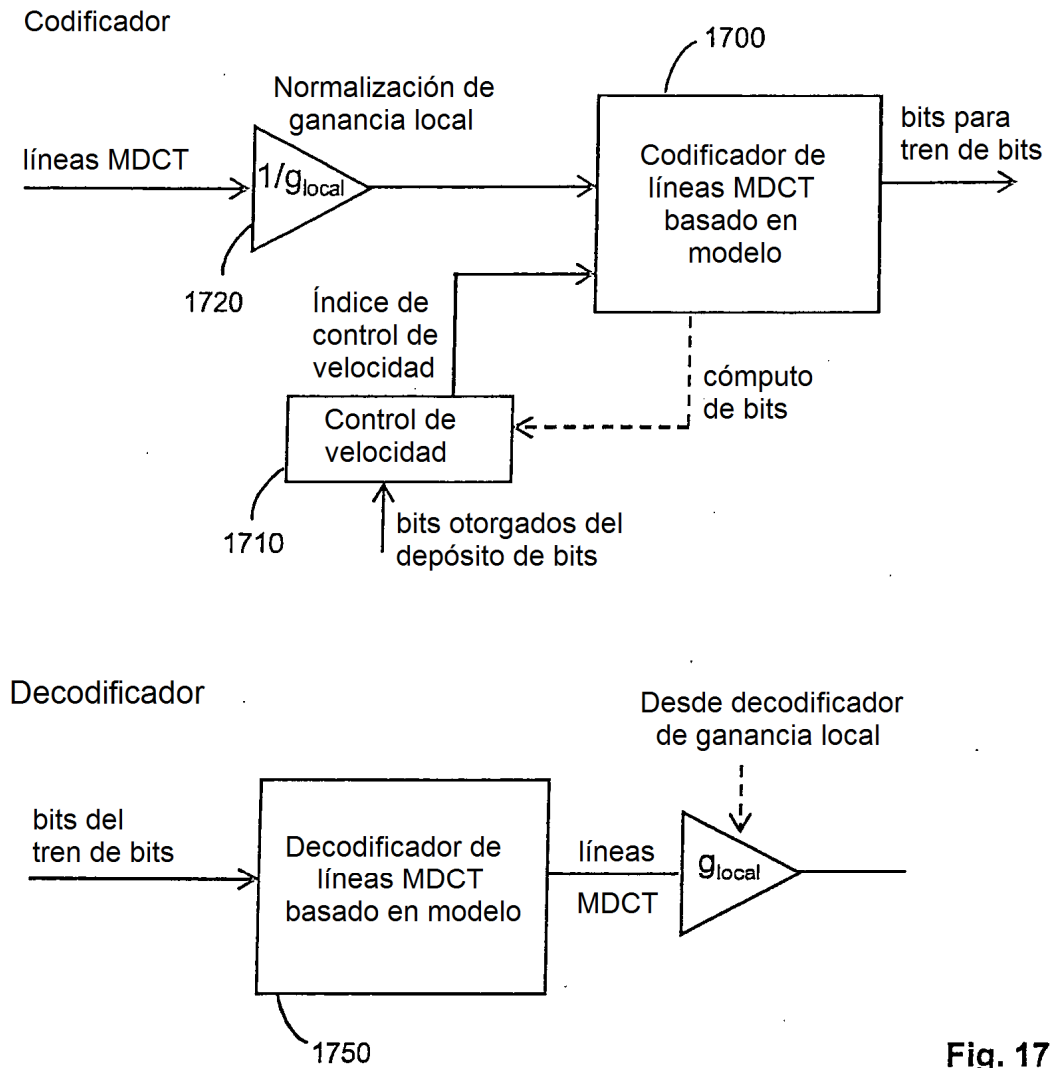


Fig. 17

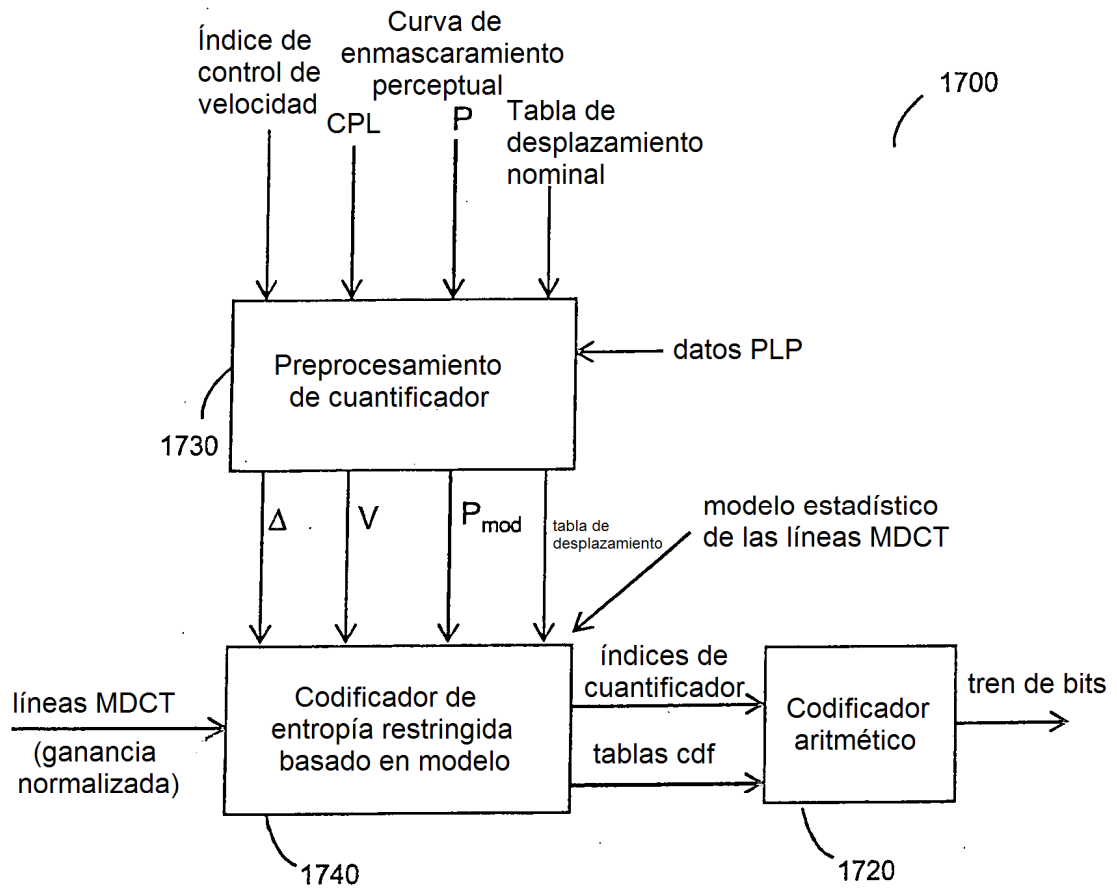


Fig. 17a

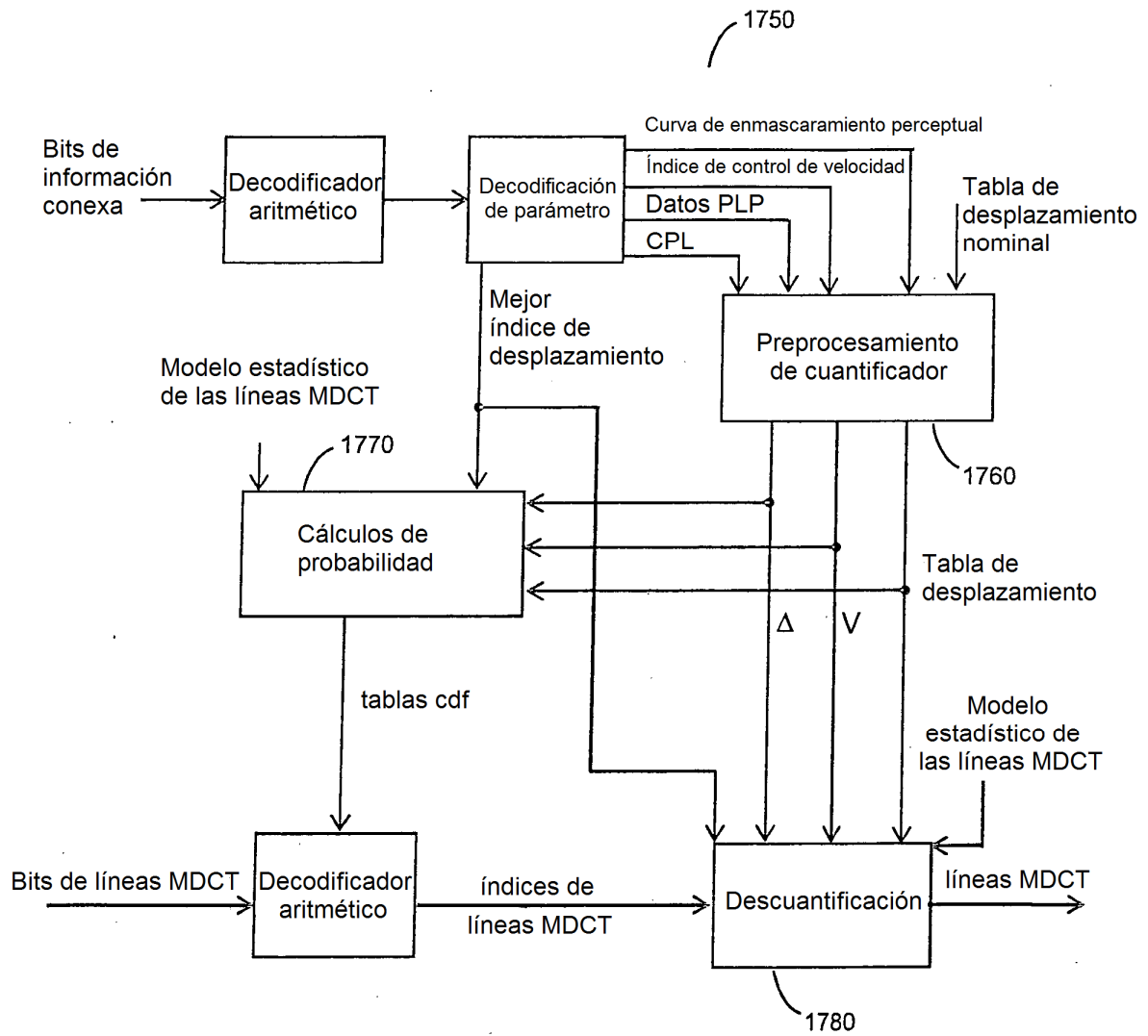


Fig. 17b

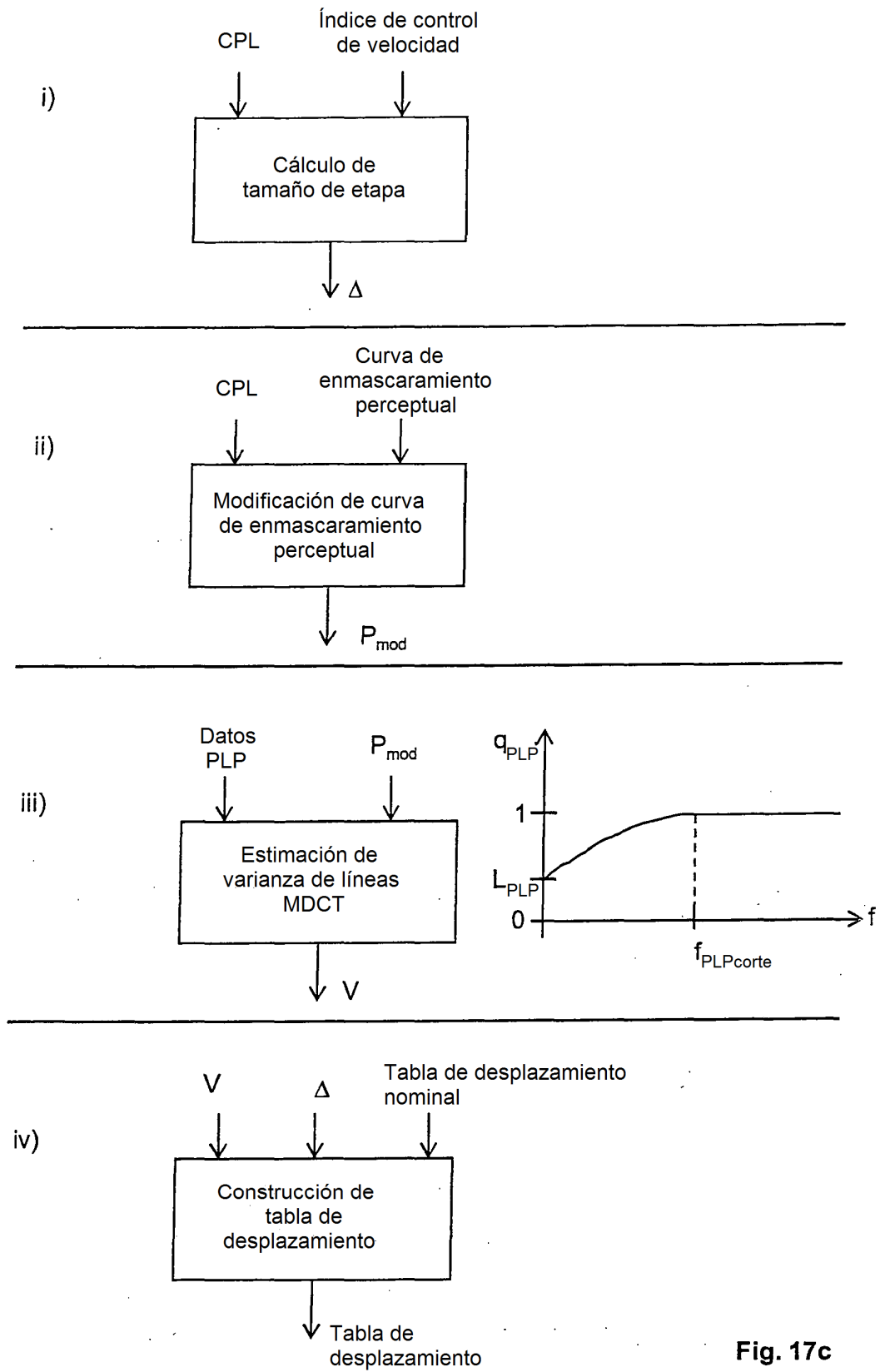


Fig. 17c

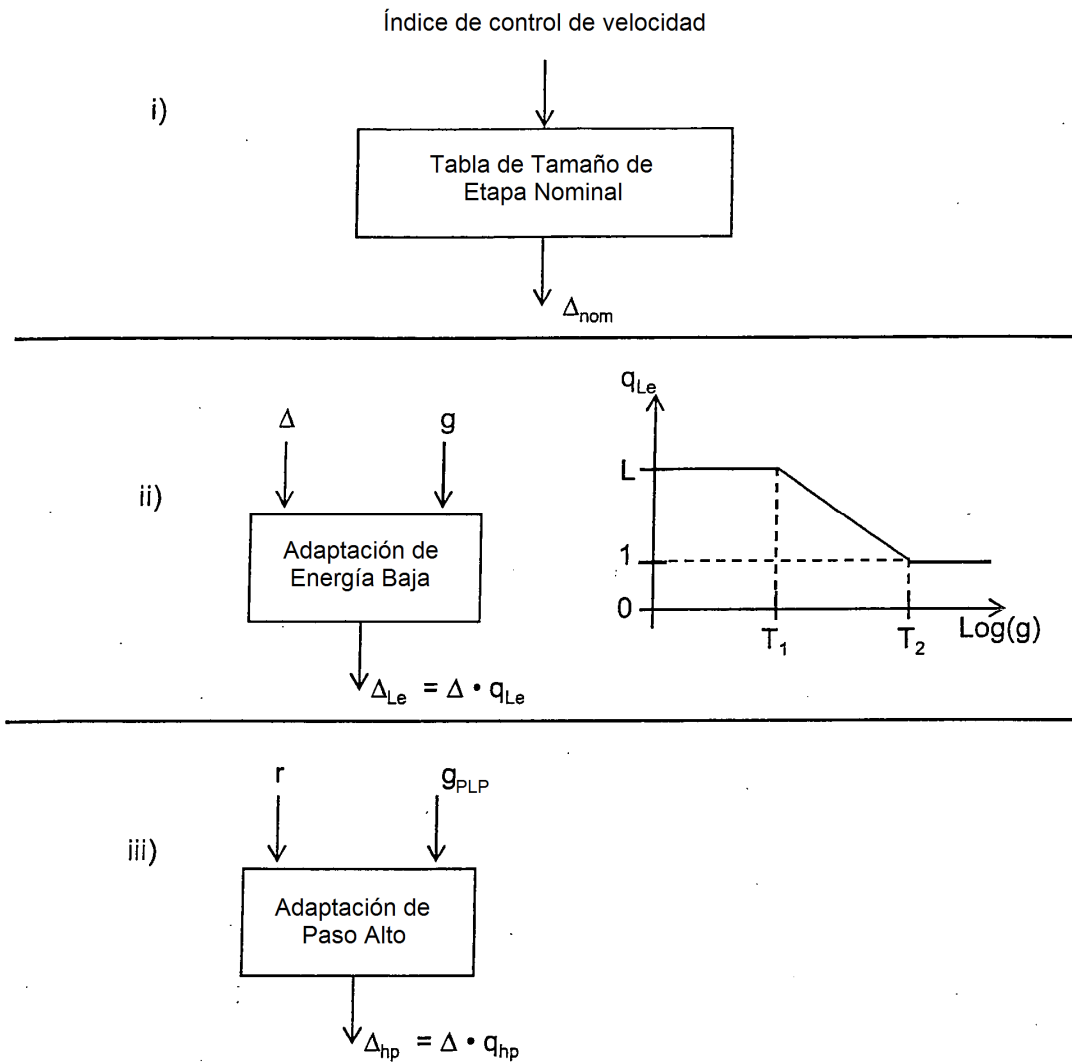


Fig. 17d

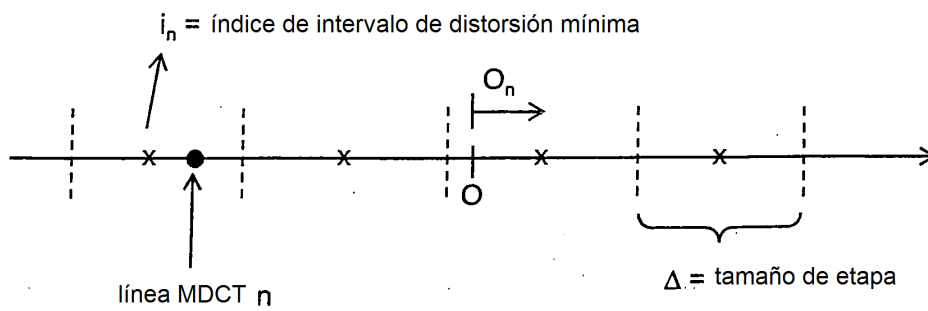


Fig. 17f

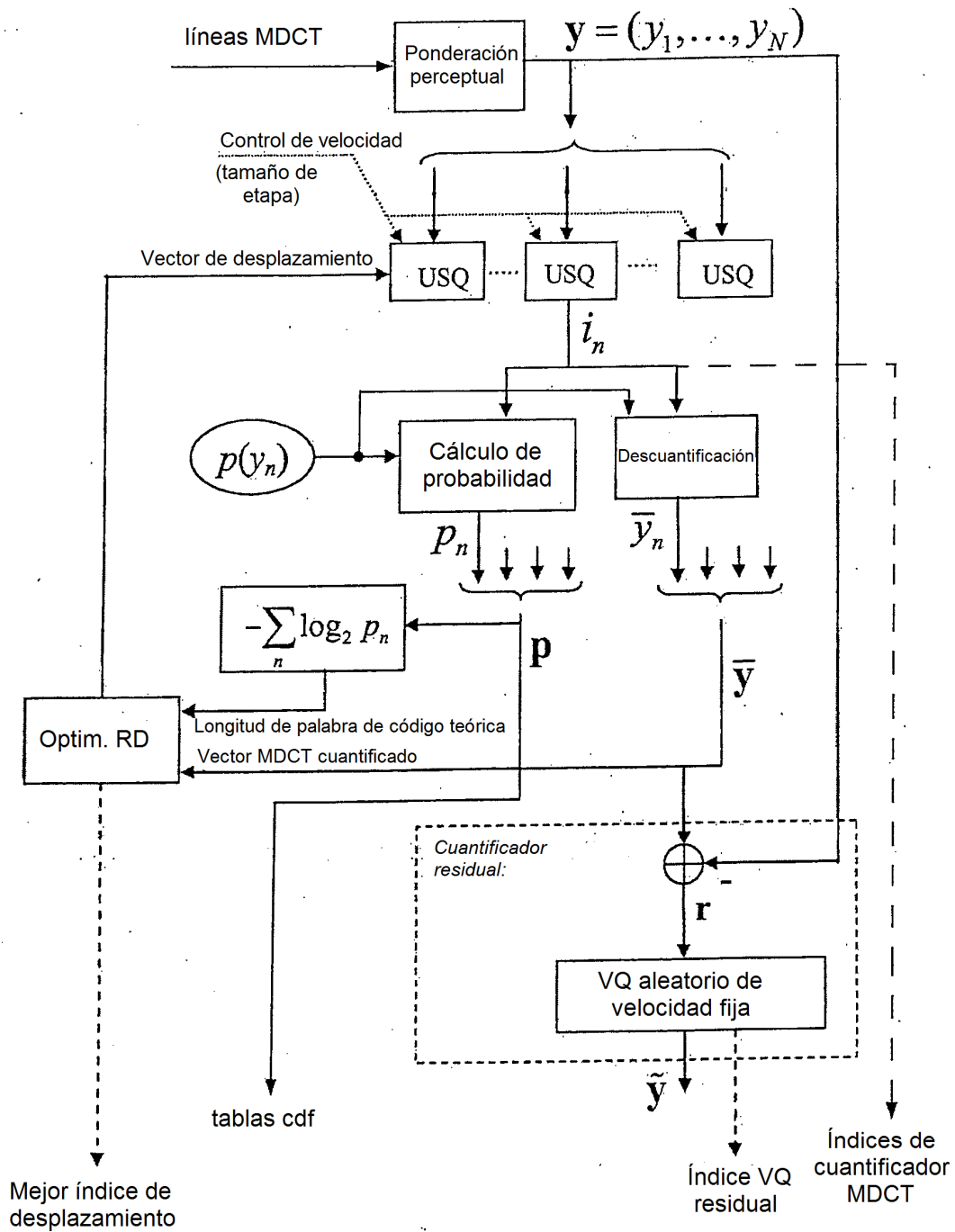


Fig. 17e

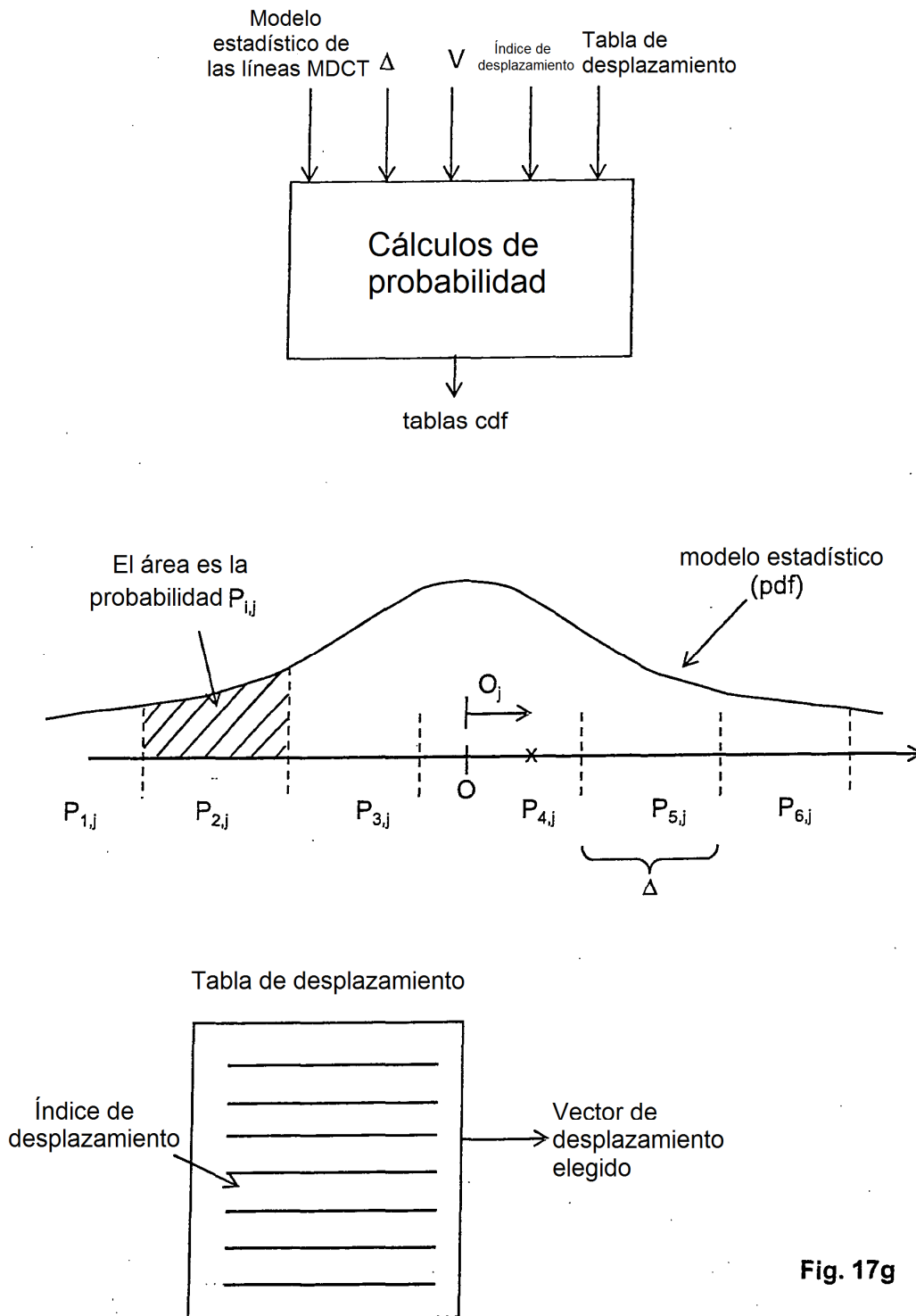


Fig. 17g

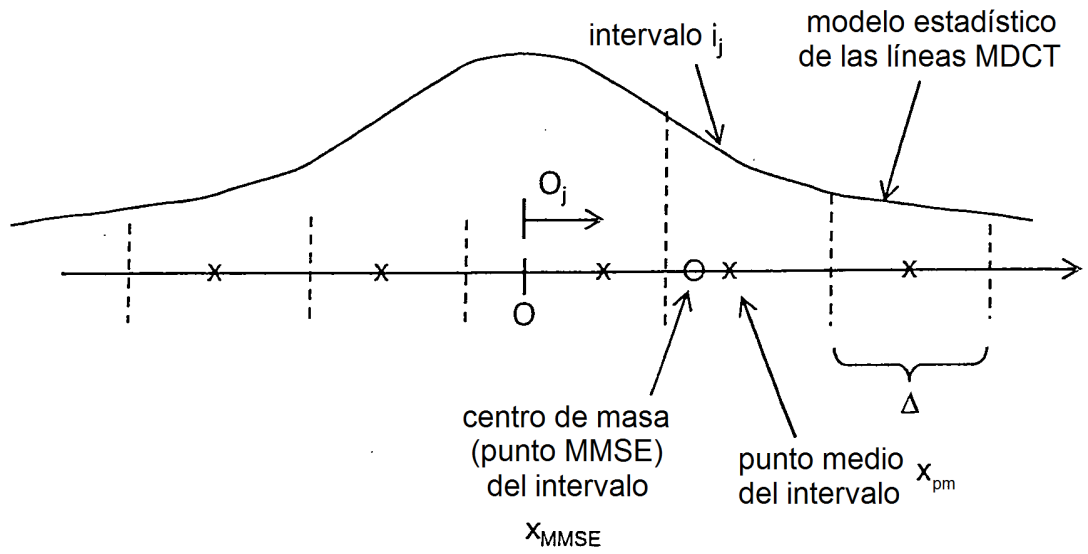


Fig. 17h

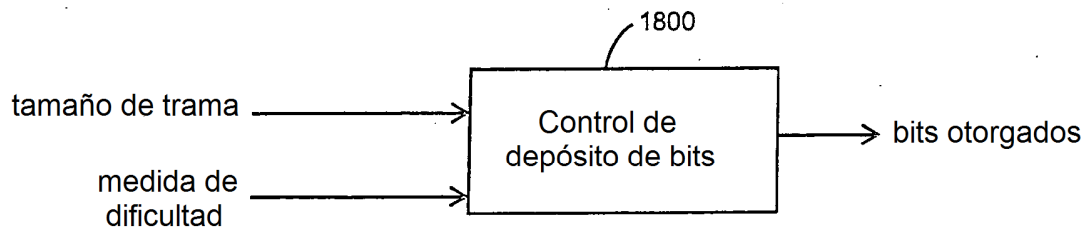


Fig. 18

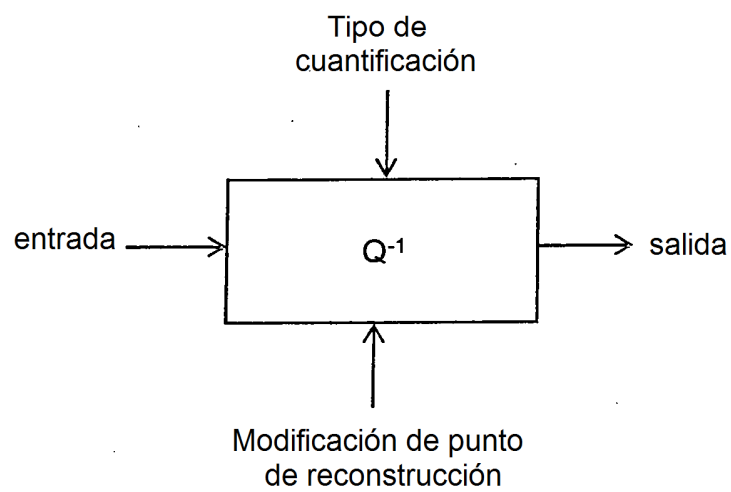


Fig. 19

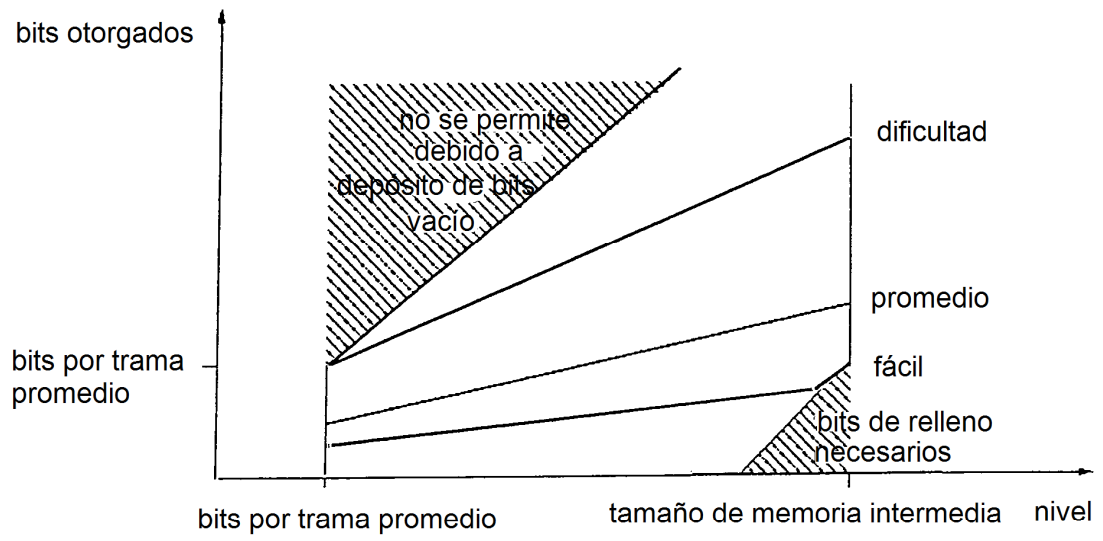


Fig. 18a

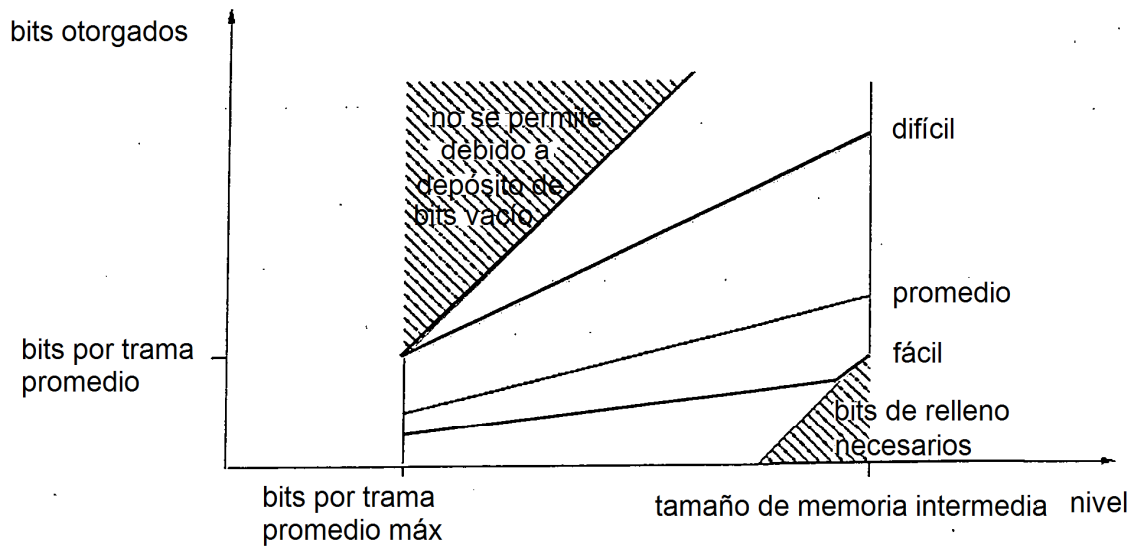


Fig. 18b

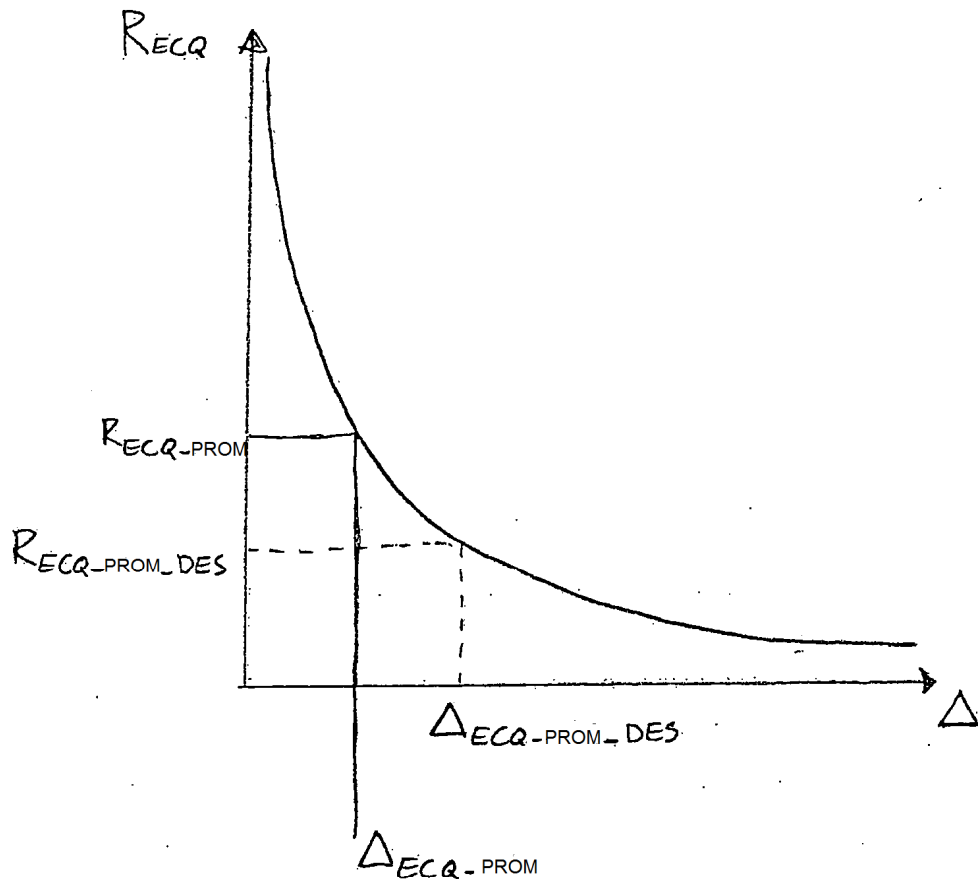


Fig. 18c