

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 592 416**

51 Int. Cl.:

G10L 19/18 (2013.01)

G10L 19/16 (2013.01)

G10L 19/02 (2013.01)

G10L 19/00 (2013.01)

G10L 19/008 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **18.02.2009 E 09002270 (8)**

97 Fecha y número de publicación de la concesión europea: **06.07.2016 EP 2146344**

54 Título: **Esquema de codificación/decodificación de audio que tiene una derivación conmutable**

30 Prioridad:

17.07.2008 US 81586 P

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

30.11.2016

73 Titular/es:

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

72 Inventor/es:

**GRILL, BERNHARD;
BAYER, STEFAN;
FUCHS, GUILLAUME;
GEYERSBERGER, STEFAN;
GEIGER, RALF;
HILPERT, JOHANNES;
KRÄMER, ULRICH;
LECOMTE, JEREMIE;
MULTRUS, MARKUS;
NEUENDORF, MAX;
POPP, HARALD;
RETTTELBACH, NIKOLAUS;
LEFEBVRE, ROCH;
BESSETTE, BRUNO;
LAPIERRE, JIMMY;
GOURNAY, PHILIPPE y
SALAMI, REDWAN**

74 Agente/Representante:

ARIZTI ACHA, Monica

ES 2 592 416 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).

Esquema de codificación/decodificación de audio que tiene una derivación conmutable

DESCRIPCIÓN

5 La presente invención se refiere a la codificación de audio y, en particular, a esquemas de codificación de audio de baja tasa de bits.

10 En la técnica se conocen esquemas de codificación en el dominio de frecuencia tales como MP3 o AAC. Estos codificadores en el dominio de frecuencia se basan en una conversión de dominio del tiempo/dominio de frecuencia, una etapa de cuantificación posterior en la que se controla el error de cuantificación mediante el uso de información de un módulo psicoacústico, y una etapa de codificación en la que los coeficientes espectrales cuantificados y la correspondiente información secundaria se codifican por entropía mediante el uso de tablas de código.

15 Por otra parte, hay codificadores que son muy adecuados para el procesamiento del habla (speech) tal como el AMR-WB+ que se ha descrito en el documento 3GPP TS 26.290. Tales esquemas de codificación del habla llevan a cabo un filtrado predictivo lineal de una señal en el dominio del tiempo. Un filtrado LP de este tipo se deriva de un análisis predictivo lineal de la señal en el dominio del tiempo. Los coeficientes de filtro de LP resultantes se codifican y transmiten a continuación como información secundaria. Este proceso se conoce como LPC (Linear Prediction Coding, Codificación de Predicción Lineal). En la salida del filtro, la señal residual de predicción o la señal de error de predicción que también se conoce como la señal de excitación, se codifica mediante las etapas de análisis por síntesis del codificador ACELP, o como alternativa, se codifica mediante un codificador de transformada, que utiliza una transformada de Fourier con una superposición. La decisión entre la codificación ACELP y la codificación de excitación codificada por transformada, que también recibe la denominación de codificación TCX, se lleva a cabo mediante el uso de un algoritmo de bucle cerrado o de bucle abierto.

20 Los esquemas de codificación de audio en el dominio de la frecuencia, tales como el esquema de codificación AAC de alta eficiencia, que combina un esquema de codificación AAC con una técnica de replicación de ancho de banda espectral, también pueden combinarse para dar una herramienta de codificación conjunta estéreo o de múltiples canales, conocida con el término "MPEG surround".

Por otra parte, los codificadores del habla tales como el AMR-WB+ también tienen una etapa de refuerzo de alta frecuencia y una funcionalidad estéreo.

35 Los esquemas de codificación en el dominio de la frecuencia son ventajosos porque muestran una elevada calidad a bajas tasas de bits para las señales de música. Sin embargo, es problemática la calidad de las señales del habla a bajas tasas de bits.

40 Los esquemas de codificación del habla muestran una elevada calidad para las señales del habla, aún a bajas tasas de bits, pero muestran una pobre calidad para las señales de música a bajas tasas de bits.

45 El documento WO 2008/071353 A2 y RAMPRASHAD, S. A. The Multimode Transform Predictive Coding Paradigm. IEEE Transactions on Speech and Audio Processing, Vol. 11, N.º 2, marzo de 2003, páginas 117-129, XP011079700, desvelan ejemplos de esquemas híbridos para codificación y decodificación de señales del habla y de audio.

Un objeto de la presente invención es proporcionar un concepto mejorado de codificación/decodificación.

50 Se logra este objeto mediante un aparato para codificar una señal de audio según la reivindicación 1, un método para codificar una señal de audio según la reivindicación 11, un aparato para decodificar una señal de audio codificada según la reivindicación 12, un método para decodificar una señal de audio codificada según la reivindicación 18, o un programa informático según la reivindicación 19.

55 En un codificador según la presente invención, se utilizan dos conversores de dominio, en el que el primer convertidor de dominio convierte una señal de audio desde el primer dominio tal como el dominio del tiempo en un segundo dominio tal como un dominio de LPC. El segundo convertidor de dominio está operativo para convertir desde un dominio de entrada en un dominio de salida, y el segundo convertidor de dominio recibe, como una entrada, una señal de salida del primer convertidor de dominio o una señal de salida de una derivación conmutable, que está conectada para puentear el primer convertidor de dominio. En otras palabras, esto significa que el segundo convertidor de dominio recibe, como una entrada, la señal de audio en el primer dominio tal como el dominio del tiempo, o como alternativa, la señal de salida del primer convertidor de dominio, es decir, una señal de audio, que ya se ha convertido desde uno de los dominios en un dominio diferente. La salida del segundo convertidor de dominio se procesa por un primer procesador para generar una primera señal procesada, y la salida del primer convertidor de dominio se procesa por un segundo procesador para generar una segunda señal procesada. Se prefiere que la derivación

conmutable también pueda adicionalmente conectarse al segundo procesador de manera tal que la entrada en el segundo procesador sea la señal de audio en el dominio del tiempo en lugar de una salida del primer conversor de dominio.

5 Este concepto de codificación, sumamente flexible, es especialmente útil para la codificación de audio de alta calidad y de elevada eficiencia de bit, ya que permite codificar una señal de audio en al menos tres dominios diferentes y, cuando la derivación conmutable también está adicionalmente conectada al segundo procesador, incluso en cuatro dominios. Esto puede lograrse mediante la conmutación controlable de la derivación conmutable para derivar o puentear o no el primer conversor de dominio para una determinada parte de la señal de audio en el dominio del tiempo. Aunque se puentee el primer conversor de dominio, sigue habiendo dos posibilidades diferentes para codificar la señal de audio del dominio del tiempo, es decir por medio del primer procesador conectado a un 10 segundo conversor de dominio o al segundo procesador.

Se prefiere que el primer procesador y el segundo conversor de dominio formen conjuntamente un codificador de modelo de sumidero de información, tal como el codificador de audio accionado psicoacústicamente según se conoce a partir de la capa 3 de MPEG 1 o MPEG 4 (AAC). 15

Se prefiere que el otro codificador, es decir, el segundo procesador sea un codificador del dominio del tiempo, que es por ejemplo el codificador residual conocido de un codificador ACELP, en el que la señal residual de LPC se codifica mediante el uso de un codificador residual tal como un codificador de cuantificación de vectores para la 20 señal residual de LPC o una señal del dominio del tiempo. En una forma de realización, este codificador del dominio del tiempo recibe como una entrada, una señal del dominio de LPC, cuando la derivación está abierta. Un codificador de este tipo es un codificador modelo de fuentes de información ya que, a diferencia del codificador modelo de sumideros de información, el codificador modelo de fuentes de información se ha diseñado específicamente para utilizar detalles específicos de un modelo de generación del habla. Sin embargo, si la 25 derivación está cerrada, la señal introducida en el segundo procesador será una señal del dominio del tiempo en lugar de una señal de dominio de LPC.

Sin embargo, si la derivación conmutable está desactivada, lo que significa que la señal de audio procedente del primer dominio se convierte en un segundo dominio antes de que se procese adicionalmente, nuevamente quedan 30 dos posibilidades diferentes, es decir o bien codificar la salida del primer conversor de dominio en el segundo dominio, que puede ser por ejemplo un dominio de LPC, o bien, como alternativa, transformar la señal del segundo dominio en un tercer dominio, que puede ser por ejemplo un dominio espectral.

Es ventajoso que el conversor de dominio espectral, es decir el segundo conversor de dominio, esté adaptado para implementar el mismo algoritmo independientemente de si la señal introducida en el segundo conversor de dominio está en el primer dominio tal como el dominio del tiempo o de si está en el segundo dominio tal como el dominio 35 LPC.

En el lado del decodificador, existen dos ramales de decodificación diferentes, uno de los ramales de decodificación incluye un conversor de dominio, es decir el segundo conversor de dominio, mientras que el otro ramal de decodificación incluye solamente un procesador inverso, pero no incluye un conversor de dominio. En función del ajuste real de la derivación en el lado del codificador, es decir de si la derivación estaba activa o no, se desvía o no un primer conversor en un decodificador. En particular, el primer conversor en un decodificador se puentea en un decodificador cuando la salida del segundo conversor ya está en el dominio objetivo deseado tal como el primer 40 dominio o del tiempo. Sin embargo, si la salida del segundo conversor en el decodificador está en un dominio distinto del primer dominio, entonces se desactiva la derivación del decodificador y la señal se convierte desde el dominio diferente al dominio objetivo, es decir, el primer dominio en la forma de realización preferida. En una forma de realización, la segunda señal procesada está en el mismo dominio, es decir, en el segundo dominio, pero en otras formas de realización en las que una derivación conmutable en el lado del codificador también puede conectarse al segundo procesador, la salida del segundo procesador inverso en el lado del decodificador ya puede estar también en el primer dominio. En este caso, el primer conversor se puentea mediante el uso de la derivación conmutable en el lado del decodificador de manera tal que un combinador de salida del decodificador recibe señales de entrada, que representan diferentes partes de una señal de audio y que están en el mismo dominio. Estas señales pueden multiplexarse en tiempo por el combinador, o pueden ser objeto de desvanecimiento cruzado (cross-fade) por el 45 combinador de salida del decodificador. 50

En una forma de realización preferida, el aparato para codificar comprende una etapa de procesamiento previo común para comprimir una señal de entrada. Esta etapa de procesamiento previo común puede incluir el procesador de múltiples canales y/o un procesador de replicación de ancho de banda espectral de manera tal que la salida de la 60 etapa de procesamiento previo común para todos los diferentes modos de codificación, es una versión comprimida con respecto a una entrada en la etapa de procesamiento previo común. De manera correspondiente, la señal de salida del combinador del lado del decodificador puede procesarse posteriormente por una etapa de procesamiento posterior común que, por ejemplo, está operativa para llevar a cabo una síntesis de replicación de ancho de banda espectral y/o una operación de expansión de múltiples canales tal como una operación de mezclado ascendente de

múltiples canales, preferentemente guiada mediante el uso de información paramétrica de múltiples canales transmitida desde el lado del codificador al lado del decodificador.

5 En una forma de realización preferida, el primer dominio en el que están situadas la señal de audio introducida en el codificador y la señal de audio emitida por el decodificador, es el dominio del tiempo. En una forma de realización preferida, el segundo dominio en el que está situada la salida del primer conversor de dominio, es un dominio de LPC, por lo que el primer conversor de dominio es una etapa de análisis de LPC. En otra forma de realización, el tercer dominio, es decir, aquel en el que está situada la salida del segundo conversor de dominio, es un dominio espectral o es un dominio espectral de la señal de dominio de LPC generada por el primer conversor de dominio. El primer procesador, conectado al segundo conversor de dominio se implementa preferentemente como un codificador sumidero de información tal como un cuantificador/escalador junto con un código reductor de la entropía tal como un cuantificador accionado psicoacústicamente conectado a un codificador de Huffman o a un codificador aritmético, que lleve a cabo las mismas funcionalidades, independientemente de si la señal de entrada está en el dominio espectral o en el dominio espectral de LPC.

15 En una forma de realización adicional preferida, el segundo procesador para procesar la salida del primer conversor de dominio o para procesar la salida de la derivación conmutable en un dispositivo de funcionalidad completa, es un codificador del dominio del tiempo tal como un codificador de señales residuales utilizado en el codificador ACELP o en cualquier otro codificador CELP.

20 Posteriormente se describen formas de realización preferidas de la presente invención, con respecto a los dibujos adjuntos, en los que:

25 La figura 1a es un diagrama de bloques de un esquema de codificación según un primer aspecto de la presente invención;

La figura 1b es un diagrama de bloques de un esquema de decodificación según el primer aspecto de la presente invención;

30 La figura 1c es un diagrama de bloques de un esquema de codificación según otro aspecto de la presente invención;

35 La figura 1d es un diagrama de bloques de un esquema de decodificación según el otro aspecto de la presente invención;

La figura 2a es un diagrama de bloques de un esquema de codificación según un segundo aspecto de la presente invención; y

40 La figura 2b es un diagrama esquemático de un esquema de decodificación según el segundo aspecto de la presente invención;

La figura 2c es un diagrama de bloques de un procesamiento previo común preferido de la figura 2a; y

45 La figura 2d es un diagrama de bloques de un procesamiento posterior común preferido de la figura 2b;

La figura 3a ilustra un diagrama de bloques de un esquema de codificación según otro aspecto de la presente invención;

50 La figura 3b ilustra un diagrama de bloques de un esquema de decodificación según el otro aspecto de la presente invención;

La figura 3c ilustra una representación esquemática de un aparato/método de codificación con conmutadores en cascada;

55 La figura 3d ilustra un diagrama esquemático de un aparato o método para decodificar, en el que se utilizan combinadores en cascada;

60 La figura 3e ilustra una ilustración de una señal del dominio del tiempo y una correspondiente representación de la señal codificada, ilustrándose regiones de desvanecimiento cruzado breve que están incluidas en ambas señales codificadas;

La figura 4a ilustra un diagrama de bloques con un conmutador situado antes de los ramales de codificación;

- La figura 4b ilustra un diagrama de bloques de un esquema de codificación con el conmutador situado después de la codificación de los ramales;
- 5 La figura 4c ilustra un diagrama de bloques para una forma de realización preferida para un combinador;
- La figura 5a ilustra una forma de onda de un segmento del habla de dominio del tiempo como un segmento de señal casi-periódico o similar a un impulso;
- La figura 5b ilustra un espectro del segmento de la figura 5a;
- 10 La figura 5c ilustra un segmento del habla en el dominio del tiempo de habla no vocalizada como un ejemplo de un segmento similar a ruido o estacionario;
- La figura 5d ilustra un espectro de la forma de onda de dominio del tiempo de la figura 5c;
- 15 La figura 6 ilustra un diagrama de bloques de un análisis mediante codificador CELP de síntesis;
- Las figuras 7a a 7d ilustran señales de excitación vocalizadas/no vocalizadas como un ejemplo de señales similares a impulsos o estacionarias;
- 20 La figura 7e ilustra una etapa de LPC del lado del codificador, que proporciona información de predicción a corto plazo y la señal de error de predicción;
- La figura 7f ilustra otra forma de realización de un dispositivo de LPC para generar una señal ponderada;
- 25 La figura 7g ilustra una implementación para transformar una señal ponderada en una señal de excitación mediante la aplicación de una operación de ponderación inversa y un posterior análisis de excitación, requerido en el conversor 537 de la figura 2b;
- 30 La figura 8 ilustra un diagrama de bloques de un algoritmo conjunto de múltiples canales según una forma de realización de la presente invención;
- La figura 9 ilustra una forma de realización preferida de un algoritmo de ampliación de ancho de banda;
- 35 La figura 10a ilustra una descripción detallada del conmutador cuando se lleva a cabo una decisión de bucle abierto; y
- La figura 10b muestra una ilustración del conmutador cuando se opera en un modo de decisión de bucle cerrado.
- 40 La figura 1a ilustra una forma de realización de la invención en la que hay dos conversores 510, 410 de dominio, y la derivación 50 conmutable. La derivación 50 conmutable está adaptada para estar activa o inactiva en respuesta a una señal 51 de control, que se introduce en una entrada de control de conmutación de la derivación 50 conmutable. Si la derivación conmutable está activa, la señal de audio en una entrada 99, 195 de señal de audio no se alimenta en el primer conversor 510 de dominio, sino que se alimenta en la derivación 50 conmutable de manera que el
- 45 segundo conversor 410 de dominio recibe la señal de audio en la entrada 99, 195 directamente. En una forma de realización, que se analizará en relación con las figuras 1c y 1d, la derivación 50 conmutable puede conectarse de manera alterna al segundo procesador 520 sin estar conectado al segundo conversor 410 de dominio de manera tal que la señal de salida de la derivación 50 conmutable se procesa por medio del segundo procesador 520 solamente.
- 50 Sin embargo, si la derivación 50 conmutable se establece en un estado inactivo por la señal 51 de control, la señal de audio en la entrada 99 o 195 de señal de audio se introduce en el primer conversor 510 de dominio y a la salida del primer conversor 510 de dominio, y se introduce o bien en el segundo conversor 410 de dominio o bien en el segundo procesador 520. La decisión en cuanto si la señal de salida del primer conversor de dominio se introduce en el segundo conversor 410 de dominio o en el segundo procesador 520 se toma preferentemente basándose
- 55 también en una señal de control del conmutador, pero puede, como alternativa, efectuarse mediante otros medios tales como metadatos o basándose en un análisis de la señal. Como alternativa, la señal del primer conversor 510 de dominio, puede incluso introducirse en ambos dispositivos 410, 520 y la selección, de qué señal de proceso se introduce en la interfaz de salida para representar la señal de audio en una determinada parte de tiempo, se efectúa mediante un conmutador conectado entre los procesadores y la interfaz de salida, como se analiza en relación con
- 60 la figura 4b. Por otra parte, la decisión en cuanto a qué señal se introduce en la corriente de datos de salida, también puede efectuarse dentro de la propia interfaz 800 de salida.

Tal como se ilustra en la figura 1a, el aparato según la invención para codificar una señal de audio para obtener una señal de audio codificada en la entrada 99/195 en el primer dominio comprende el primer conversor de dominio para

convertir la señal de audio desde el primer dominio en un segundo dominio. Además, se proporciona la derivación 54 conmutable que puentea el primer conversor 510 de dominio o para producir una conversión de la señal de audio por el primer conversor de dominio en respuesta a una señal de control de conmutación de derivación. Por lo tanto, en el estado activo la derivación conmutable puentea el primer conversor de dominio y, en el estado no activo, la señal de audio se introduce en el primer conversor de dominio.

Además, se proporciona el segundo conversor 410 de dominio para convertir la señal de audio recibida desde la derivación 50 conmutable o el primer conversor de dominio en un tercer dominio. El tercer dominio es diferente del segundo dominio. Además, se proporciona un primer procesador 420 para codificar la señal de audio del tercer dominio según un primer algoritmo de codificación para obtener una primera señal procesada. Además, se proporciona un segundo procesador 520 para codificar la señal de audio recibida desde el primer conversor de dominio según un segundo algoritmo de codificación, donde el segundo algoritmo de codificación es diferente del primer algoritmo de codificación. El segundo procesador proporciona la segunda señal procesada. En particular, el aparato está adaptado para tener un señal de audio codificada en la salida del mismo para una parte de la señal de audio donde esta señal codificada incluye o bien la primera señal procesada o bien la segunda señal procesada. Evidentemente, puede haber regiones de cruce, pero en vista de una mayor eficiencia en la codificación, el objetivo es mantener las regiones de cruce lo más pequeñas posibles y eliminarlas siempre que sea posible de manera que se obtenga una máxima compresión de la tasa de bits.

La figura 1b ilustra un decodificador correspondiente al codificador en la figura 1a en una forma de realización preferida. El aparato para decodificar una señal de audio codificada en la figura 1b recibe, como una entrada, una señal de audio codificada que comprende una primera señal procesada que está en un tercer dominio y una segunda señal procesada que está en un segundo dominio, donde el segundo dominio y el tercer dominio son diferentes entre sí. En particular, la señal introducida en una interfaz 900 de entrada, es similar a la salida procedente de la interfaz 800 de la figura 1a. El aparato para decodificar comprende un primer procesador 430 inverso para el procesamiento inverso de la primera señal procesada y un segundo 530 procesador inverso para el procesamiento inverso de la segunda señal procesada. Adicionalmente, se proporciona un segundo conversor 440 para convertir el dominio de la primera señal procesada inversa procedente del tercer dominio en un dominio diferente. Además, se proporciona un primer conversor 540 para convertir la segunda señal procesada inversa en un primer dominio o para convertir la primera señal procesada inversa en el primer dominio cuando el dominio diferente no es el primer dominio. Esto significa que la primera señal procesada inversa sólo se convierte por el primer conversor cuando la primera señal procesada no está ya en el primer dominio, es decir, en un dominio objetivo en el que ha de hallarse la señal de audio decodificada o la señal de audio intermedia en el caso de un circuito de procesamiento previo/procesamiento posterior. Además, el decodificador comprende una derivación 52 para puentear el primer conversor 540 cuando el dominio diferente es el primer dominio. El circuito en la figura 1b comprende además un combinador 600 para combinar una salida del primer conversor 540 y una salida de derivación, es decir, una salida de señal por la derivación 52 para obtener una señal 699 de audio decodificada combinada, que puede utilizarse como tal o que puede incluso descomprimirse mediante el uso de una etapa de procesamiento posterior común, como se analizará más adelante.

La figura 1c ilustra una forma de realización preferida del codificador de audio según la invención en la que se ha proporcionado el clasificador de señales en el modelo 300 psicoacústico, para clasificar la señal de audio introducida en una etapa de procesamiento previo común formada por un codificador 101 MPEG Surround y un procesador 102 de replicación de banda espectral reforzada. Además, el primer conversor 510 de dominio es una etapa de análisis de LPC, y la derivación conmutable está conectada entre una entrada y una salida de la etapa 510 de análisis de LPC, que es el primer conversor de dominio.

Por lo general, el dispositivo de LPC emite una señal de dominio de LPC, que puede ser cualquier señal de dominio de LPC tal como la señal de excitación en la figura 7e o una señal ponderada en la figura 7f, o cualquier otra señal, que se haya generado mediante la aplicación de coeficientes de filtro de LPC a una señal de audio. Adicionalmente, un dispositivo de LPC también puede determinar estos coeficientes y también puede cuantificar/codificar estos coeficientes.

Adicionalmente, se proporciona un conmutador 200 en la salida del primer conversor de dominio de manera tal que se reenvía una señal en la salida en común de la derivación 50 y de la etapa 510 de LPC o bien a un primer ramal 400 de codificación o bien a un segundo ramal 500 de codificación. El primer ramal 400 de codificación comprende el segundo conversor 410 de dominio y el primer procesador 420 de la figura 1a y el segundo ramal 500 de codificación comprende el segundo procesador 520 de la figura 1a. En la forma de realización del codificador de la figura 1c, la entrada del primer conversor 510 de dominio está conectada a la entrada de la derivación 50 conmutable y la salida de la derivación 50 conmutable está conectada a la salida del primer conversor 510 de dominio para formar una salida en común, y esta salida en común es la entrada en el conmutador 200 donde el conmutador comprende dos salidas, pero puede comprender incluso salidas adicionales para procesadores de codificación adicionales.

Preferentemente, el segundo conversor 410 de dominio en el primer ramal 400 de codificación comprende una transformada de MDCT, que, adicionalmente, está combinada con una funcionalidad TW (time warp, distorsión de tiempo) conmutable. El espectro de MDCT se codifica mediante el uso de un escalar/cuantificador, que lleva a cabo una cuantificación de valores de entrada basándose en la información proporcionada desde el modelo psicoacústico situado dentro del bloque 300 clasificador de señales. Por otra parte, el segundo procesador comprende un codificador del dominio del tiempo para la codificación en el dominio del tiempo de la señal de entrada. En una forma de realización, el conmutador 200 se controla de manera que en el caso de una derivación 50 activa/cerrada, el conmutador 200 se establece automáticamente al ramal 400 de codificación superior. Sin embargo, en otra forma de realización, el conmutador 200 también puede controlarse independientemente de la derivación 50 conmutable incluso cuando la derivación está activa/cerrada, por lo que el codificador 520 del dominio del tiempo puede recibir directamente la señal de entrada de audio del dominio del tiempo.

La figura 1d ilustra un decodificador correspondiente en el que el bloque 540 de síntesis de LPC corresponde al primer conversor de la figura 1b y puede puentearse por medio de la derivación 52, que es preferentemente una derivación conmutable controlada mediante una señal de derivación generada por el demultiplexor 900 de la corriente de bits. El demultiplexor 900 de la corriente de bits puede generar esta señal y todas las otras señales de control para los ramales 430, 530 de codificación o del bloque 701 de síntesis de SBR o del bloque 702 decodificador MPEG Surround a partir de una corriente 899 de bits de entrada, o puede recibir los datos para estas líneas de control de un análisis de la señal o según otra fuente de información separada.

A continuación se dará una descripción más detallada de la forma de realización en la figura 1c para el codificador y en la figura 1d para el decodificador.

La forma de realización preferida consiste en un codificador híbrido de audio, que combina las ventajas de una tecnología MPEG exitosa, tal como AAC, SBR y MPEG Surround con la tecnología exitosa del codificador del habla. El códec resultante comprende un procesamiento previo común para todas las categorías de señal, consistente en MPEG Surround y un SBR reforzado (eSBR). Controlado por un modelo psicoacústico y basándose en la categoría de la señal, se selecciona una arquitectura de codificador derivada de sumidero o fuente de información en una base trama a trama.

El códec propuesto utiliza ventajosamente herramientas de codificación, como el MPEG Surround, SBR y el codificador de base de AAC. Estas mismas han experimentado modificaciones y mejoras para mejorar el rendimiento para el habla y con muy bajas tasas de bit. Con mayores tasas de bit se consigue el rendimiento del AAC, ya que el nuevo códec puede volver a un modo muy cercano al AAC. Se implementa un modo de codificación insonoro reforzado que proporciona de media un rendimiento de codificación insonora ligeramente mejor. Para tasas de bit de aproximadamente 32 kbps y menores, se activan herramientas adicionales para mejorar el rendimiento del codificador de base para el habla y otras señales. Los principales componentes de estas herramientas son una conformación de frecuencia basada en LPC, más opciones de longitud de ventana alternativas para el codificador basado en MDCT, y un codificador del dominio del tiempo. Se utiliza una nueva técnica de ampliación de ancho de banda como una ampliación a la herramienta de SBR, que es más adecuada para bajas frecuencias de cruce y para el habla. La herramienta MPEG Surround proporciona una representación paramétrica de una señal estéreo o multicanal al proporcionar una mezcla descendente ("down mix") y una imagen estéreo parametrizada. Para los casos de prueba dados, se utiliza para codificar señales estéreo solamente, pero también es adecuado para señales de entrada multicanal al hacer uso de la funcionalidad MPEG Surround existente a partir del MPEG-D.

Todas las herramientas en la cadena de códec, con la excepción del codificador MDCT, se utilizan preferentemente sólo a baja tasa de bits.

La tecnología MPEG Surround se utiliza para transmitir N canales de entrada de audio mediante M canales de transmisión de audio. Por lo tanto, el sistema puede inherentemente manejar múltiples canales. La tecnología MPEG Surround ha recibido mejoras para incrementar el rendimiento a bajas tasas de bit y para señales similares al habla.

El modo de operación básico es la creación de una mezcla descendente mono de alta calidad a partir de la señal estéreo de entrada. Adicionalmente, se extrae un conjunto de parámetros espaciales. En el lado del decodificador se genera una señal estéreo de salida mediante el uso de la mezcla descendente mono decodificada en combinación con los parámetros espaciales extraídos y transmitidos. Se ha añadido un modo 2-1-2 de baja tasa de bits a los puntos operativos 5-x-5 o 7-x-7 existentes en MPEG Surround, que usan una estructura de árbol sencilla que consiste en una única caja de OTT (one-to-two, de uno a dos) en el mezclado ascendente de MPEG Surround. Algunos de los componentes han recibido modificaciones para adaptarse mejor a la reproducción del habla. Para mayores velocidades de datos, tales como 64 kbps y superiores, el codificador principal está utilizando codificación estéreo discreta (Media/Lateral o L/R), no se utiliza el MPEG Surround para este punto de operación.

La ampliación de ancho de banda propuesta en esta propuesta tecnológica se basa en la tecnología SBR MPEG. El banco de filtros utilizado es idéntico al banco de filtros QMF en MPEG Surround y SBR, con lo que ofrece la posibilidad de compartir muestras en el dominio de QMF entre MPEG Surround y SBR sin síntesis/análisis

adicionales. En comparación con la herramienta SBR estandarizada, eSBR introduce un algoritmo de procesamiento reforzado, que es óptimo para el contenido tanto del habla como de audio. Se incluye una ampliación a SBR, que es más adecuada para muy bajas tasas de bit y bajas frecuencias de cruce.

5 Como es conocido a partir de una combinación de SBR y AAC, esta característica puede desactivarse globalmente, dejando la decodificación de la totalidad del intervalo de frecuencias al codificador principal.

La parte del codificador principal del sistema propuesto puede observarse como la combinación de un filtro LPC opcional y un codificador principal del dominio de la frecuencia/dominio del tiempo, conmutable.

10 Como es conocido a partir de las arquitecturas de codificadores del habla, el filtro LPC proporciona la base de un modelo de fuente para el habla humana. El procesamiento de LPC puede habilitarse o deshabilitarse (puentearse) globalmente en una base trama a trama.

15 Después del filtro de LPC, se codifica la señal de dominio de LPC utilizándose o bien una arquitectura de codificador de dominio del tiempo o bien de dominio de frecuencia basado en transformada. La conmutación entre estos dos ramales se controla mediante un modelo psicoacústico ampliado.

La arquitectura de codificador del dominio del tiempo se basa en la tecnología ACELP, que proporciona un óptimo rendimiento de codificación especialmente para señales del habla con bajas tasas de bit.

20 El ramal del códec basado en el dominio de la frecuencia se basa en una arquitectura MDCT con cuantificador escalar y codificación de entropía.

25 Opcionalmente, hay una herramienta de distorsión de tiempo disponible para reforzar la eficiencia de la codificación para señales del habla a tasas de bit superiores (tales como 64 kbps y superiores) por medio de una representación más compacta de la señal.

30 La arquitectura basada en MDCT proporciona una buena calidad para tasas de bit más bajas y escala hacia la transparencia, como es conocido a partir de las tecnologías de MPEG existentes. Puede converger a un modo de AAC, con tasas de bit superiores.

35 Los requisitos de memoria intermedia son idénticos a los de AAC, es decir, la cantidad máxima de bits en la memoria intermedia de entrada es de 6.144 por canal del codificador principal: 6.144 bits por elemento de canal mono, 12.288 bits por elemento de par de canal estéreo.

En el codificador se controla una reserva de bits, que permite la adaptación del proceso de codificación a la demanda de bits actual. Las características de la reserva de bits son idénticas que para el AAC.

40 El codificador y decodificador pueden controlarse para operar sobre diferentes tasas de bit entre 12 kbps mono y 64 kbps estéreo.

45 La complejidad del decodificador se especifica en términos de PCU. Para el decodificador de base se requiere una complejidad de aproximadamente 11,7 PCU. En el caso de utilizarse la herramienta de distorsión del tiempo, como para el modo de prueba de 64 kbps, se incrementa la complejidad del decodificador a 22.2 PCU.

Los requisitos para RAM y ROM para un decodificador estéreo preferido son:

RAM: ~24 kPalabras
ROM: ~150 kPalabras

50 Mediante la notificación del codificador de entropía, puede obtenerse un tamaño de ROM global de solamente ~98 kPalabras.

55 En el caso en que se utiliza la herramienta de distorsión del tiempo, la demanda de RAM se incrementa en ~3 kPalabras, y la demanda de ROM se incrementa en ~40 kPalabras.

60 El retardo algorítmico teórico depende de las herramientas utilizadas en la cadena de códec (por ejemplo MPEG Surround, etc.): el retardo algorítmico de la tecnología propuesta se visualiza por punto de operación en la tasa de muestreo de códec. Los valores dados a continuación no incluyen un retardo de trama, es decir, el retardo necesario para llenar la memoria intermedia de entrada del codificador con la cantidad de muestras necesaria para procesar la primera trama. Este retardo de trama es de 2.048 muestras para todos los modos operativos especificados. Las siguientes tablas contienen tanto el retardo algorítmico mínimo como el retardo para la implementación utilizada. El retardo adicional para muestrear nuevamente archivos de PCM introducidos de 48 kHz a la tasa de muestreo de códec, se especifica en '(.)'.

ID del ensayo	Retardo algorítmico mínimo teórico (muestras)	Retardo algorítmico según la implementación (muestras)
Ensayo 1, 64 kbps estéreo	8278	8278 (+44)
Ensayo 2, 32 kbps estéreo	9153	11201 (+44)
Ensayo 3, 24 kbps estéreo	9153	11200 (+45)
Ensayo 4, 20 kbps estéreo	9153	9153 (+44)
Ensayo 5, 16 kbps estéreo	11201	11201 (+44)
Ensayo 6, 24 kbps mono	4794	5021 (+45)
Ensayo 7, 20 kbps mono	4794	4854 (+44)
Ensayo 8, 16 kbps mono	6842	6842 (+44)
Ensayo 9, 12 kbps mono	6842	6842 (+44)

Los atributos principales de este códec pueden resumirse como sigue:

- 5 La tecnología propuesta utiliza de manera ventajosa una tecnología de codificación de audio y del habla según el estado de la técnica, sin sacrificar rendimiento para codificar contenido o bien del habla o bien de música. Esto tiene como resultado un códec que puede entregar calidad para contenido del habla, música y mixto según el estado de la técnica, para un intervalo de tasa de bits que empieza con tasas de bit muy bajas (12 kbps) y que se eleva hasta elevadas velocidades de datos tales como 128 kbps y más, a la que el códec obtiene una calidad transparente.
- 10 En la figura 2a, en una etapa 100 de procesamiento previo común se introduce una señal mono, una señal estéreo o una señal de múltiples canales. El esquema de procesamiento previo común puede tener una de funcionalidad estéreo conjunto, una funcionalidad envolvente, y/o una funcionalidad de ampliación de ancho de banda. En la salida de bloque 100 hay un canal mono, un canal estéreo o múltiples canales que se introducen en un conjunto de derivación 50 y conversor 510 o en múltiples conjuntos de este tipo.
- 15 El conjunto de derivación 50 y conversor 510 puede existir para cada salida de etapa 100, cuando la etapa 100 tiene dos o más salidas, es decir, cuando la etapa 100 emite una señal estéreo o una señal de múltiples canales. A modo de ejemplo, el primer canal de una señal estéreo podría ser un canal del habla y el segundo canal de la señal estéreo podría ser un canal de música. En esta situación, la decisión en la etapa de decisión puede ser diferente entre los dos canales para el mismo instante de tiempo.
- 20 La derivación 50 se controla por una etapa 300 de decisión. La etapa de decisión recibe, como una entrada, una señal de entrada en el bloque 100 o una señal emitida por el bloque 100. Como alternativa, la etapa 300 de decisión puede también recibir una información secundaria que está incluida en la señal mono, la señal estéreo o la señal de múltiples canales o que está al menos asociada a una señal de este tipo, donde la información existe como se generó, por ejemplo, cuando se produjo originalmente la señal mono, la señal estéreo o la señal de múltiples canales.
- 25 En una forma de realización, la etapa de decisión no controla la etapa 100 de procesamiento previo, y la flecha entre los bloques 300 y 100 no existe. En otra forma de realización, el procesamiento en el bloque 100 se controla en cierta medida por la etapa 300 de decisión a efectos de establecer uno o más parámetros en el bloque 100 basándose en la decisión. Sin embargo, esto no influirá al algoritmo general en el bloque 100, de modo que la principal funcionalidad en el bloque 100 está activa, independientemente de la decisión en la etapa 300.
- 30 La etapa de decisión 300 acciona la derivación 50 para alimentar la salida de la etapa de procesamiento previo común o bien en una parte 400 de codificación de la frecuencia ilustrada en un ramal superior de la figura 1a, o bien en el conversor 510 del dominio de LPC que puede ser parte de la segunda parte 500 de codificación ilustrada en un ramal inferior en la figura 2a y que tiene los elementos 510, 520.
- 35 En una forma de realización, la derivación puentea un único conversor de dominio. En otra forma de realización, puede haber conversores de dominio adicionales para diferentes ramales de codificación tales como un tercer ramal de codificación o incluso un cuarto ramal de codificación o incluso más ramales de codificación. En una forma de realización con tres ramales de codificación, el tercer ramal de codificación podría ser similar al segundo ramal de codificación, pero podría incluir un codificador de excitación diferente del codificador 520 de excitación en el segundo ramal 500. En esta forma de realización, el segundo ramal comprende la etapa 510 de LPC y un codificador de excitación basado en un libro de código tal como en el ACELP, y el tercer ramal comprende una etapa de LPC y un codificador de excitación que opera en una representación espectral de la señal de salida de la etapa de LPC.
- 40 Un elemento clave del ramal de codificación del dominio de la frecuencia es un bloque 410 de conversión espectral
- 45 que está operativo para convertir la señal de salida de la etapa de procesamiento previo común en un dominio
- 50

espectral. El bloque de conversión espectral puede incluir un algoritmo de MDCT, un QMF, un algoritmo de FFT, un análisis de ondícula o un banco de filtros tal como un banco de filtros críticamente muestreado que tiene una determinada cantidad de canales de bancos de filtros, donde las señales de subbanda en este banco de filtros pueden ser señales con valor real o señales con valor complejo. La salida del bloque 410 de conversión espectral se codifica mediante el uso de un codificador 420 de audio espectral, que puede incluir bloques de procesamiento como es conocido a partir del esquema de codificación de AAC.

En el ramal 500 inferior de codificación, un elemento clave es un analizador de modelo de fuente tal como el LPC 510, que en esta forma de realización es el conversor 510 de dominio, y que emite dos tipos de señales. Una señal es una señal de información de LPC que se utiliza para controlar la característica de filtro de un filtro de síntesis de LPC. Esta información de LPC se transmite a un decodificador. La otra señal de salida de etapa 510 de LPC es una señal de excitación o una señal de dominio de LPC, que se introduce en un codificador 520 de excitación. El codificador 520 de excitación puede provenir de cualquier codificador de modelo de fuente-filtro tal como un codificador CELP, un codificador ACELP o cualquier otro codificador que procese una señal de dominio de LPC.

Otra implementación preferida de codificador de excitación es una codificación de transformada de la señal de excitación o una señal de dominio de LPC. En esta forma de realización, la señal de excitación no se codifica mediante el uso de un mecanismo de libro de código de ACELP, sino que la señal de excitación se convierte en una representación espectral y los valores de la representación espectral tales como las señales de subbanda en el caso de un banco de filtros o de coeficientes de frecuencia en el caso de una transformada tales como una FFT se codifican para obtener una compresión de los datos. Una implementación de este tipo de codificador de excitación es el modo de codificación de TCX conocido a partir del AMR-WB+. Este modo se obtiene conectando la salida de la etapa 510 de LPC al conversor 410 espectral. El modo de TCX conocido del documento 3GPP TS 26.290 causa un procesamiento de la señal perceptualmente ponderada en el dominio de la transformada. Una señal ponderada transformada de Fourier se cuantifica mediante el uso de una cuantificación en malla dividida de múltiples coeficientes (VQ algebraico) con cuantificación del factor de ruido. Se calcula una transformada en 1024, 512 o 256 ventanas de muestras. La señal de excitación se recupera mediante filtración inversa de la señal ponderada cuantificada, por medio de un filtro de ponderación inverso.

En la figura 1a o en la figura 1c, al bloque 510 de LPC le sigue un codificador del dominio del tiempo, que puede ser un bloque de ACELP o un codificador del dominio de la transformada, que puede ser un bloque 527 de TCX. El ACELP se describe en el documento 3GPP TS 26.190, y el TCX se describe en el documento 3GPP TS 26.290. En términos generales, el bloque de ACELP recibe una señal de excitación de LPC calculada mediante un procedimiento como se describe en la figura 7e. El bloque 527 de TCX recibe una señal ponderada generada como se genera en la figura 7f.

En el TCX, se aplica la transformada a la señal ponderada calculada filtrando la señal de entrada por medio de un filtro de ponderación basado en LPC. El filtro de ponderación utilizado en las formas de realización preferidas de la invención, está dado por $(1 - A(z/\gamma))/(1 - \mu z^{-1})$. Por lo tanto, la señal ponderada es una señal de dominio de LPC y su transformada es un dominio espectral de LPC. La señal procesada por el bloque 526 ACELP es la señal de excitación y es diferente de la señal procesada por el bloque 527, pero ambas señales se hallan en el dominio de LPC.

En el lado del decodificador, después de la transformada espectral inversa, se aplica la inversa del filtro de ponderación, es decir $(1 - \mu z^{-1})/A(z/\gamma)$. A continuación se filtra la señal a través de $(1-A(z))$ para pasar al dominio de excitación del LPC. Por lo tanto, la conversión al dominio de LPC y una operación de TCX^{-1} incluyen una

transformada inversa y a continuación una filtración a través de $\frac{(1 - \mu z^{-1})}{(1 - A(z/\gamma))} (1 - A(z))$ para convertir desde el dominio de la señal ponderada al dominio de la excitación.

Aunque el elemento 510 ilustra un único bloque, el bloque 510 puede emitir diferentes señales siempre y cuando estas señales se encuentren en el dominio de LPC. El modo real del bloque 510 tal como el modo de la señal de excitación o el modo de señal ponderada, puede depender del estado de conmutación real. Como alternativa, el bloque 510 puede tener dos dispositivos de procesamiento paralelos, donde un dispositivo se implementa similar a la figura 7e, y el otro dispositivo se implementa como la figura 7f. Por ello, el dominio de LPC a la salida de 510 puede representar o bien la señal de excitación de LPC o bien la señal ponderada de LPC o bien cualquier otra señal de dominio de LPC.

En el modo de LPC, cuando la derivación está inactiva, es decir, cuando hay una codificación de ACELP/TCX, se prefiere que la señal se enfatice previamente a través de un filtro $1-0,68z^{-1}$ antes de la codificación. En el decodificador de ACELP/TCX se desenfatisa la señal sintetizada mediante el filtro $1/(1-0,68z^{-1})$. El énfasis previo puede ser parte del bloque 510 de LPC donde la señal se enfatiza previamente antes del análisis y cuantificación de

LPC. De manera similar, el desénfasis puede ser parte del bloque de síntesis de LPC LPC^{-1} 540.

Existen varios dominios de LPC. Un primer dominio de LPC representa la excitación de LPC, y el segundo dominio de LPC representa la señal ponderada de LPC. Es decir, la primera señal de dominio de LPC se obtiene mediante la filtración a través de $(1-A(z))$ para convertir al dominio residual/de excitación de LPC, mientras que la segunda señal de dominio de LPC se obtiene filtrando a través del filtro $(1-A(z/\gamma))/(1-\mu z^{-1})$ para convertir al dominio ponderado de LPC.

La decisión en la etapa de decisión puede adaptarse a la señal de manera que la etapa de la decisión lleve a cabo una discriminación entre música y habla y controle la derivación 50 y si está presente, el conmutador 200 en la figura 1c de manera tal que las señales de música se introducen en el ramal 400 superior, y las señales del habla se introducen en el ramal 500 inferior. En una forma de realización, la etapa de decisión alimenta su información de decisión hacia una corriente de bits de salida de manera tal que un decodificador puede utilizar esta información de decisión para llevar a cabo las operaciones de decodificación correctas.

En la figura 2b se ilustra un decodificador de este tipo. La señal emitida por el codificador 420 de audio espectral se introduce, después de la transmisión, en un decodificador 430 de audio espectral. La salida del decodificador 430 de audio espectral se introduce en un conversor 440 del dominio del tiempo. De manera análoga, la salida del codificador 520 de excitación de la figura 2a se introduce en un decodificador 530 de excitación que emite una señal del dominio de LPC. La señal del dominio de LPC se introduce en una etapa 540 de síntesis de LPC, que recibe, como entrada adicional, la información de LPC generada por la correspondiente etapa 510 de análisis de LPC. La salida del conversor 440 del dominio del tiempo y/o la salida de la etapa 540 de síntesis de LPC se introducen en una derivación 52 conmutable. La derivación 52 se controla por medio de una señal de control de derivación que se ha generado, por ejemplo, por la etapa 300 de decisión, o que se había proporcionado externamente tal como por un creador de la señal mono, señal estéreo o señal de múltiples canales, original.

La salida de la derivación 540 o de la etapa 540 que se introduce en el combinador 600 es una señal mono completa que se introduce posteriormente en una etapa 700 de procesamiento posterior común, que puede llevar a cabo un procesamiento de estéreo conjunto o un procesamiento de ampliación de ancho de banda, etc. En función de la funcionalidad específica de la etapa de procesamiento posterior común, se emite una señal mono, una señal estéreo o una señal de múltiples canales, que tiene, cuando la etapa 700 de procesamiento posterior común lleva a cabo una operación de ampliación de ancho de banda, un ancho de banda más grande que la señal introducida en el bloque 700.

En una forma de realización, la derivación 52 está adaptada para puentear el conversor 540 individual. En otra forma de realización, puede haber conversores adicionales que definen ramales de decodificación adicionales tales como un tercer ramal de decodificación o incluso un cuarto ramal de decodificación o incluso más ramales de decodificación. En una forma de realización con tres ramales de decodificación, el tercer ramal de decodificación podría ser similar al segundo ramal de decodificación, pero podría incluir un decodificador de excitación diferente del decodificador 530 de excitación en el segundo ramal, 530, 540. En esta forma de ramificación, el segundo ramal comprende la etapa 540 de LPC y un decodificador de excitación basado en libro de código tal como en el ACELP, y el tercer ramal comprende una etapa de LPC y un decodificador de excitación que opera en una representación espectral de la señal emitida de la etapa 540 de LPC.

Como se enunció anteriormente, en la figura 2c se ilustra un esquema de codificación preferido según un segundo aspecto de la invención. El esquema de procesamiento previo común en 100 de la figura 1a comprende ahora un bloque 101 de estéreo conjunto/envolvente que genera, como salida, parámetros de estéreo conjunto y una señal de salida mono, que se genera mediante el mezclado descendente de la señal de entrada que es una señal que tiene dos o más canales. En términos generales, la señal en la salida del bloque 101 también puede ser una señal que tiene más canales, pero debido a la funcionalidad del mezclado descendente del bloque 101, el número de canales en la salida del bloque 101 será menor que el número de canales introducidos en el bloque 101.

La salida de bloque 101 se introduce en un bloque 102 de ampliación de ancho de banda que, en el codificador de la figura 2c, emite una señal limitada en banda tal como la señal de banda baja o la señal de paso bajo en su salida. Por otra parte, para la banda alta de la señal introducida en el bloque 102, se generan parámetros de ampliación de ancho de banda tales como parámetros de envolvente espectral, parámetros de filtración inversa, parámetros de margen de ruido, etc., como es conocido a partir del perfil HE-AAC de MPEG-4, y se reenvían a un multiplexor 800 de corriente de bits.

Se prefiere que la etapa 300 de decisión reciba la señal introducida en el bloque 101 o introducida en el bloque 102 para decidir entre, por ejemplo, un modo de música o un modo del habla. En el modo de música, se selecciona el ramal 400 superior de codificación, mientras que en el modo del habla, se selecciona el ramal 500 inferior de codificación. Se prefiere que la etapa de decisión adicionalmente controle el bloque 101 de estéreo conjunto y/o el bloque 102 de ampliación de ancho de banda para adaptar la funcionalidad de estos bloques a la señal específica.

Por lo tanto, cuando la etapa de decisión determina que una determinada parte de tiempo de la señal de entrada es del primer modo tal como el modo de música, entonces pueden controlarse características específicas del bloque 101 y/o del bloque 102 mediante la etapa 300 de decisión. Como alternativa, cuando la etapa 300 de decisión determina que la señal está en un modo del habla o, en términos generales, en un modo de codificación del dominio de LPC, entonces las características específicas de los bloques 101 y 102 pueden controlarse según la salida de la etapa de decisión.

En función de la decisión del conmutador, que puede derivarse de la señal de entrada del conmutador 200 o a partir de cualquier fuente externa tal como un productor de la señal de audio original subyacente a la señal introducida en la etapa 200, el conmutador conmuta entre el ramal 400 de codificación de frecuencia y el ramal 500 de codificación de LPC. El ramal 400 de codificación de frecuencia comprende una etapa de conversión espectral y una posterior etapa de cuantificación/codificación relacionada. La etapa de cuantificación/codificación puede incluir cualquiera de las funcionalidades conocidas a partir de los codificadores del dominio de la frecuencia modernos, tales como el codificador AAC. Por otra parte, la operación de cuantificación en la etapa de cuantificación/codificación puede controlarse mediante un módulo psicoacústico que genera información psicoacústica tal como un umbral de enmascaramiento psicoacústico sobre la frecuencia, donde esta información se introduce en la etapa.

Se prefiere que la conversión espectral se efectúe mediante el uso de una operación de MDCT que, incluso más preferentemente es la operación de MDCT de distorsión del tiempo, donde la intensidad o en términos generales, la intensidad de la distorsión puede controlarse entre cero y una elevada intensidad de distorsión. En una intensidad de distorsión cero, la operación de MDCT en el bloque 400 en la figura 1c es una operación de MDCT directa conocida en la técnica. La intensidad de distorsión de tiempo, junto con la información secundaria de la distorsión del tiempo, puede transmitirse/introducirse en el multiplexor 800 de la corriente de bits como información secundaria. Por ello, si se utiliza TW-MDCT, debería enviarse información secundaria de distorsión de tiempo a la corriente de bits, como se ilustra mediante 424 en la figura 1c, y - en el lado del decodificador - debería recibirse información secundaria de distorsión de tiempo a partir de la corriente de bits, como se ilustra mediante el elemento 434 en la figura 1d.

En el ramal de codificación de LPC, el codificador en el dominio de LPC puede incluir un núcleo de ACELP que calcule una ganancia de tono, un retardo de tono y/o información de libro de código tal como un índice de libro de código y una ganancia de código.

En el primer ramal 400 de codificación, se prefiere que un conversor espectral comprenda una operación de MDCT específicamente adaptada que tiene determinadas funciones de ventana seguido por una etapa de codificación de cuantificación/entropía que puede ser una etapa de cuantificación de vectores, pero se prefiere que sea un cuantificador/codificador similar al cuantificador/codificador en el ramal de codificación del dominio de la frecuencia.

La figura 2d ilustra un esquema de decodificación correspondiente al esquema de codificación de la figura 2c. La corriente de bits generada por un multiplexor de corrientes de bits se introduce en un demultiplexor de corrientes de bits. En función de una información derivada por ejemplo de la corriente de bits por medio de un bloque de detección de modo, se controla un conmutador del lado de decodificador para o bien reenviar señales desde el ramal superior o bien señales desde el ramal inferior al bloque 701 de ampliación de ancho de banda. El bloque 701 de ampliación de ancho de banda recibe información secundaria procedente del demultiplexor de la corriente de bits, y, basándose en esta información secundaria y la salida de la decisión de modo, reconstruye la banda alta basándose en la banda baja emitida por el combinador 600 de la figura 1d, por ejemplo.

La señal de banda completa generada por el bloque 701 se introduce en la etapa 702 de procesamiento de estéreo conjunto/envolvente, que reconstruye dos canales estéreo o varios multicanales. En términos generales, el bloque 702 emitirá más canales que los que introdujeron en este bloque. En función de la aplicación, la entrada en el bloque 702 puede incluir incluso dos canales tales como en un modo estéreo y puede incluir incluso más canales siempre y cuando la salida por este bloque tenga más canales que la entrada en este bloque.

Se ha mostrado que el conmutador 200 en la figura 1c conmuta entre ambos ramales de manera que solamente un ramal recibe una señal para procesar y el otro ramal no recibe una señal para procesar, como se muestra en términos generales en la figura 4a. Sin embargo, en una forma de realización alternativa ilustrada en la figura 4b, el conmutador también puede estar dispuesto después de, por ejemplo, el codificador 420 de audio y el codificador 520 de excitación, lo que significa que ambos ramales 400, 500 procesan la misma señal en paralelo. Sin embargo, para no duplicar la tasa de los bits, se selecciona solamente la señal emitida por uno de los ramales 400 o 500 de codificación para escribirse en la corriente emitida de bits. La etapa de decisión operará entonces de manera tal que la señal escrita en la corriente de bits minimiza una determinada función de coste, en la que la función de coste puede ser la tasa de bits generada o la distorsión perceptual generada o una función de coste de tasa/distorsión combinada. Por ello, sea en este modo sea en el modo ilustrado en las figuras, la etapa de decisión también puede operar en un modo de bucle cerrado para asegurar que finalmente sólo se escribe la salida de ramal de codificación en la corriente de bits que tiene la tasa de bits más baja para una distorsión perceptual o, para una tasa de bits dada, tiene la distorsión perceptual más baja.

En términos generales, el procesamiento en el ramal 400 es un procesamiento en un modelo basado en la percepción, o en un modelo de sumidero de información. Por lo tanto, este ramal modela el sistema auditivo humano que recibe sonido. En cambio, el procesamiento en el ramal 500 sirve para generar una señal en la excitación, residual o dominio de LPC. En términos generales, el procesamiento en el ramal 500 es un procesamiento en un modo del habla o un modelo de generación de información. Para las señales del habla, este modelo es un modelo del sistema de generación del habla humana/sonido que genera sonido. Sin embargo, si ha de codificarse un sonido procedente de una fuente diferente que requiere un modelo diferente de generación de sonido, entonces el procesamiento en el ramal 500 puede ser diferente.

Si bien las figuras 1a a 4c se ilustran como diagramas de bloques de un aparato, estas figuras son al mismo tiempo una ilustración de un método, en el que las funcionalidades de bloque corresponden a las etapas del método.

La figura 3c ilustra un codificador de audio para codificar una señal 195 de entrada de audio. La señal 195 de entrada de audio está presente en un primer dominio que puede ser, por ejemplo, el dominio de tiempo pero que también puede ser cualquier otro dominio tal como un dominio de la frecuencia, un dominio de LCP, un dominio espectral de LPC o cualquier otro dominio. En términos generales, la conversión desde un dominio al otro dominio se lleva a cabo mediante un tipo de algoritmo de conversión tal como cualquiera de los algoritmos de conversión de tiempo/frecuencia o de los algoritmos de conversión frecuencia/tiempo, bien conocidos.

Una transformada alternativa del dominio del tiempo, por ejemplo en el dominio de LPC, es el resultado de la filtración basada en LPC de una señal del dominio del tiempo que da como resultado una señal residual de LPC o una señal de excitación, u otra señal de dominio de LPC. Cualquier otra operación de filtrado que produzca una señal filtrada que tenga un impacto en una cantidad sustancial de muestras de señal antes de la transformada puede utilizarse como un algoritmo de transformada, cualquiera que sea el caso. Por ello la ponderación de una señal de audio mediante el uso de un filtro de ponderación basado en LPC, es otra transformada, que genera una señal en el dominio de LPC. En una transformada de tiempo/frecuencia, la modificación de un sólo valor espectral tendrá un impacto sobre todos los valores de dominio del tiempo antes de la transformada. De forma análoga, una modificación de cualquier muestra en el dominio del tiempo tendrá un impacto en cada muestra en el dominio de la frecuencia. De manera similar, una modificación de una muestra de la señal de excitación en una situación de dominio de LPC tendrá, debido a la longitud del filtro de LPC, un impacto sobre una cantidad sustancial de muestras antes del filtrado de LPC. De manera similar, una modificación de una muestra antes de una transformación de LPC tendrá un impacto sobre muchas muestras obtenidas mediante esta transformación de LPC debido al efecto de memoria inherente del filtro de LPC.

El codificador de audio de la figura 3c incluye un primer ramal 522 de codificación que genera una primera señal codificada. Esta primera señal codificada puede estar en un cuarto dominio que es, en la forma de realización preferida, el dominio de tiempo-espectral, es decir, el dominio que se obtiene cuando una señal del dominio del tiempo se procesa mediante una conversión de tiempo/frecuencia.

Por ello, el primer ramal 522 de codificación para codificar una señal de audio utiliza un primer algoritmo de codificación para obtener una primera señal codificada, donde este primer algoritmo de codificación puede incluir o no un algoritmo de conversión de tiempo/frecuencia.

Por otra parte el codificador de audio incluye un segundo ramal 523 de codificación para codificar una señal de audio. El segundo ramal 523 de codificación utiliza un segundo algoritmo de codificación para obtener una segunda señal codificada, que es diferente del primer algoritmo de codificación.

Además el codificador de audio incluye un primer conmutador 521 para conmutar entre el primer ramal 522 de codificación y el segundo ramal 523, 524 de codificación por lo que para una parte de la señal de audio introducida, sea la primera señal codificada en la salida de bloque 522 o sea la segunda señal codificada en la salida del segundo ramal de codificación está incluida en una señal de salida de codificador. Por lo tanto, cuando para una determinada parte de la señal 195 de audio de entrada, la primera señal codificada en el cuarto dominio está incluida en la señal de salida de codificador, la segunda señal codificada que es o bien la primera señal procesada en el segundo dominio o bien la segunda señal procesada en el tercer dominio no está incluida en la señal de salida del codificador. Esto asegura que este codificador es eficiente en cuanto a tasa de bits. En algunas formas de realización cualquier parte de tiempo de la señal de audio que se incluya en dos señales codificadas diferentes es pequeña en comparación con una longitud de trama, como se analizará en relación con la figura 3e. Estas pequeñas partes son útiles para un desvanecimiento cruzado desde una señal codificada a la otra señal codificada en el caso de un acontecimiento de conmutador para reducir los artefactos que podrían presentarse sin ningún desvanecimiento cruzado. Por ello, aparte de la región de desvanecimiento cruzado, cada bloque de dominio en el tiempo está representado por una señal codificada de un sólo dominio único.

Como se ilustra en la figura 3c, el segundo ramal 523 de codificación sigue a un conversor 521 para convertir la

señal de audio en el primer dominio, es decir, la señal 195 en un segundo dominio, y la derivación 50. Por otra parte, el primer ramal 522 de procesamiento obtiene una primera señal procesada que preferentemente también está en el segundo dominio de manera tal que el primer ramal 522 de procesamiento no lleve a cabo un cambio de dominio, o que esté en el primer dominio.

5 El segundo ramal 523, 524 de codificación convierte la señal de audio en un tercer dominio o en un cuarto dominio, que es diferente del primer dominio y que es también diferente del segundo dominio para obtener una segunda señal procesada en la salida del segundo ramal 523, 524 de procesamiento.

10 Además, el codificador comprende un conmutador 521 para conmutar entre el primer ramal 522 de procesamiento y el segundo ramal 523, 524 de procesamiento, donde este conmutador corresponde al conmutador 200 de la figura 1c.

15 La figura 3 ilustra un decodificador correspondiente para decodificar una señal de audio codificada generada por el codificador de la figura 3c. En términos generales, cada bloque de la señal de audio del primer dominio está representado por o bien una señal de segundo o primer dominio, o bien una señal de tercer o cuarto dominio codificada aparte de una región de desvanecimiento opcional que es preferentemente corta en comparación con la longitud de una trama para obtener un sistema que esté lo más cercano posible al límite de muestreo crítico. La señal de audio codificada incluye la primera señal codificada, una segunda señal codificada, refiriéndose la primera
20 señal codificada y la segunda señal codificada a diferentes partes de tiempo de la señal de audio decodificada y en el que el segundo dominio, el tercer dominio y el primer dominio para una señal de audio decodificada son diferentes entre sí.

25 El decodificador comprende un primer ramal de decodificación para decodificar basado en el primer algoritmo de codificación. El primer ramal de decodificación se ilustra en 531 en la figura 3d. El decodificador de la figura 3d comprende además un segundo ramal 533, 534 de decodificación, que comprende varios elementos.

30 El decodificador comprende además un primer combinador 532 para combinar la primera señal procesada inversa y la segunda señal procesada inversa para obtener una señal en el primer o segundo dominio, donde esta señal combinada está, en el primer instante de tiempo, influida solamente por la primera señal procesada inversa y está, en un instante de tiempo posterior, influenciada únicamente por la segunda señal procesada inversa.

35 El decodificador comprende además el conversor 540 para convertir la señal combinada en el primer dominio y la derivación 52 conmutable.

40 Finalmente, el decodificador ilustrado en la figura 3d comprende un segundo combinador 600 para combinar la primera señal decodificada a partir de la derivación 52 y la señal de salida del conversor 540 para obtener una señal de salida decodificada en el primer dominio. Nuevamente, la señal de salida decodificada en el primer dominio está, en el primer instante de tiempo, solamente influenciada por la señal emitida por el conversor 540 y está, en un instante de tiempo posterior, solamente influenciada por la señal puenteadada.

45 Esta situación se ilustra, desde una perspectiva del codificador, en la figura 3e. La parte superior en la figura 3e ilustra en la representación esquemática, una señal de audio del primer dominio tal como una señal de audio en el dominio del tiempo, donde el índice de tiempo aumenta de izquierda a derecha y el elemento 3 podría considerarse como una corriente de muestras de audio que representan la señal 195 en la figura 3c. La figura 3e ilustra las tramas 3a, 3b, 3c, 3d que pueden generarse mediante la conmutación entre la primera señal codificada y la segunda señal codificada como se ilustra en el elemento 4 en la figura 3e. La primera señal codificada y la segunda señal codificada están, todas ellas, en dominios diferentes. Para asegurar que la conmutación entre los dominios
50 diferentes no dé como resultado un artefacto en el lado del decodificador, las tramas 3a, 3b, 3c, ... de la señal del dominio del tiempo tienen un intervalo solapante que se indica como una región de desvanecimiento cruzado. Sin embargo, no existe ninguna región de desvanecimiento cruzado entre las tramas 3d, 3c lo que significa que la trama 3d podría también representarse mediante una señal en el mismo dominio que la señal 3c anterior, y no hay un cambio de dominio entre las tramas 3c y 3d.

55 Por ello, en términos generales, se prefiere no proporcionar una región de desvanecimiento cruzado donde no haya un cambio de dominio y proporcionar una región de desvanecimiento cruzado, es decir, una parte de la señal de audio que se codifica mediante dos señales codificadas/procesadas posteriores cuando hay un cambio de dominio, es decir, una acción de conmutación de cualquier de los dos conmutadores.

60 En la forma de realización en la que la primera señal codificada o la segunda señal procesada se ha generado por un procesamiento de MCDT que tiene por ejemplo un solapamiento del 50 por ciento, cada muestra del dominio del tiempo está incluida en dos tramas posteriores. Sin embargo, debido a las características del MDCT, esto no da como resultado un encabezamiento, ya que el MDCT es un sistema críticamente muestreado. En este contexto,

críticamente muestreadas significa que la cantidad de valores espectrales es igual a la cantidad de los valores en el dominio del tiempo. El MDCT es ventajoso en que se proporciona el efecto del cruce sin una región de cruce específica, por lo que se proporciona un cruce desde un bloque de MDCT al siguiente bloque de MDCT sin ningún encabezamiento que infringiría el requisito crítico del muestreo.

5 Se prefiere que el primer algoritmo de codificación en el primer ramal de codificación esté basado en un modelo de sumidero de información, y que el segundo algoritmo de codificación en el segundo ramal de codificación esté basado en una fuente de información o en un modelo de SNR. Un modelo de SNR es un modelo que no está específicamente relacionado con un mecanismo de generación específico de sonido sino que es un modo de
10 codificación que puede seleccionarse entre una pluralidad de modos de codificación basados, por ejemplo, en una decisión de bucle cerrado. Por lo tanto, un modelo de SNR es cualquier modelo de codificación disponible pero que no ha de estar relacionado necesariamente con la constitución física del generador de sonido sino que es cualquier módulo de codificación parametrizado diferente del modelo de sumidero de información, que puede seleccionarse mediante una decisión de bucle cerrado y, específicamente, mediante la comparación de diferentes resultados de
15 SNR de diferentes modelos.

Como se ilustra en la figura 3c, se proporciona un controlador 300, 525. Este controlador puede incluir las funcionalidades de la etapa 300 de decisión de la figura 1c. En términos generales, el controlador es para controlar la derivación y el conmutador 200 en la figura 1c de una manera adaptativa a las señales. El controlador está
20 operativo para analizar una señal de entrada en la derivación o emitida por el primer o segundo ramal de codificación o las señales obtenidas mediante la codificación y decodificación a partir del primer y segundo ramales de codificación con respecto a una función objetivo. Como alternativa o adicionalmente, el controlador está operativo para analizar la señal introducida en el conmutador o emitida por el primer ramal de procesamiento o por el segundo ramal de procesamiento u obtenida por procesamiento y procesamiento inverso a partir del primer ramal de
25 procesamiento y del segundo ramal de procesamiento, nuevamente con respecto a una función objetivo.

En una forma de realización, el primer ramal de codificación o el segundo ramal de codificación comprende un algoritmo de conversión de tiempo/frecuencia que introduce repliegue tal como un algoritmo de MDCT o de MDST, que es diferente de una transformada FFT directa, que no introduce un efecto de repliegue. Por otra parte, uno o
30 ambos ramales comprende un bloque codificador cuantificador/de entropía. Específicamente, solamente el segundo ramal de procesamiento del segundo ramal de codificación incluye el conversor de tiempo/frecuencia que introduce una operación de repliegue y el primer ramal de procesamiento del segundo ramal de codificación comprende un codificador cuantificador y/o de entropía y no introduce ningún efecto de repliegue. Se prefiere que el conversor de tiempo/frecuencia que introduce el repliegue comprenda un generador de ventanas para aplicar una ventana de
35 análisis y un algoritmo de transformada de MDCT. Específicamente, el generador de ventanas está operativo para aplicar la función de ventana a tramas posteriores en un modo de solapamiento de manera que tiene lugar una muestra de una señal de ventana en al menos dos tramas de ventana posteriores.

En una forma de realización, el primer ramal de procesamiento comprende un codificador de ACELP y un segundo
40 ramal de procesamiento comprende un conversor espectral de MDCT y el cuantificador para cuantificar componentes espectrales para obtener componentes espectrales cuantificados, donde cada componente espectral cuantificado es cero, o se define por un índice de cuantificador de la pluralidad de índices cuantificadores diferentes posibles.

45 Como se mencionó anteriormente, ambos ramales de codificación operan para codificar la señal de audio en una manera a nivel de bloques, en la que la derivación o el conmutador operan de una manera a nivel de bloques de manera tal que tiene lugar una conmutación o puenteo, como mínimo, después de un bloque de una cantidad predefinida de muestras de una señal, formando la cantidad predefinida una longitud de trama para el conmutador correspondiente. Por lo tanto, el gránulo para puentear mediante la derivación puede ser, por ejemplo, un bloque de
50 2048 o 1028 muestras, y la longitud de trama, basándose en la cual la derivación está conmutando puede ser variable, pero se prefiere que esté fijado a un periodo bastante largo.

Contrario a lo mismo, la longitud de bloque para el conmutador 200, es decir, cuando el conmutador 200 conmuta de un modo al otro, es sustancialmente menor que la longitud de bloque para el primer conmutador. Se prefiere que
55 ambas longitudes de bloque para los conmutadores se seleccionen de manera que la mayor longitud de bloque sea un número entero múltiplo de la longitud de bloque más pequeña. En la forma de realización preferida, la longitud de bloque del primer conmutador es de 2048 y la longitud de bloque del segundo conmutador es de 1024 o más, preferentemente 512 e incluso más preferentemente 256 e incluso más preferentemente 256 o incluso de 128 muestras por lo que, como máximo, el conmutador puede conmutar 16 veces cuando la derivación cambia
60 solamente una vez.

En otra forma de realización, el controlador 300 está operativo para llevar a cabo una discriminación entre música y habla para el primer conmutador, de una manera tal que se favorece una decisión a favor del habla con respecto a una decisión para música. En esta forma de realización, se toma una decisión para el habla incluso aunque una

parte inferior al 50 % de una trama para el primer conmutador sea habla y la parte de más del 50 % de la trama sea música.

5 Además, el controlador ya está operativo para conmutar al modo del habla, cuando una parte bastante pequeña de la primera trama sea habla y, específicamente, cuando una parte de la primera trama sea habla, que es el 50 % de la longitud de la segunda trama menor. Por lo tanto, una decisión de conmutación de favorecer/habla, preferida, ya se conmuta a habla incluso si, por ejemplo, sólo el 6 % o el 12 % de un bloque correspondiente a la longitud de trama del primer conmutador es habla.

10 Se prefiere que este procedimiento, para aprovechar por completo la capacidad de ahorrar la tasa de bits del primer ramal de procesamiento, que tenga un núcleo del habla vocalizada en una forma de realización y no pierda ninguna calidad ni siquiera para el resto de la primera trama grande, que es de no habla debido al hecho de que el segundo ramal de procesamiento incluye un conversor y, por ello es útil para señales de audio que también tengan señales de no habla. Se prefiere que este segundo ramal de procesamiento incluya un MDCT de solapamiento, que está críticamente muestreado, y que incluso con pequeños tamaños de ventana proporciona una operación altamente eficiente y sin repliegue debido al procesamiento de cancelación de repliegue del dominio del tiempo tal como solapar y añadir en el lado del decodificador. Además, una gran longitud de bloque para el primer ramal de codificación, que es preferentemente un ramal de codificación de MDCT similar a AAC, es útil, ya que las señales de no habla son normalmente bastante estacionarias y una ventana de transformada grande proporciona una resolución de alta frecuencia, y por ello una alta calidad y, adicionalmente, proporciona una eficiencia en tasa de bits debido a un módulo de cuantificación psicoacústicamente controlado, que también puede aplicarse al modo de codificación basado en transformada en el segundo ramal de procesamiento del segundo ramal de codificación.

25 Con respecto a la ilustración del decodificador de la figura 3d, se prefiere que la señal transmitida incluya un indicador explícito como información 4a secundaria, como se ilustra en la figura 3e. Esta información 4a secundaria se extrae mediante un analizador de corrientes de bits, no ilustrado en la figura 3d, para reenviar la correspondiente primera señal procesada o segunda señal procesada al procesador correcto tal como el primer ramal de procesamiento inverso o el segundo ramal de procesamiento inverso en la figura 3d. Por ello, una señal codificada no sólo tiene señales codificadas/procesadas sino que también incluye información secundaria relacionada con estas señales. Sin embargo, en otras formas de realización puede haber una señalización implícita que permita que un analizador de la corriente de bits del lado del decodificador distinga entre las señales determinadas. Con respecto a la figura 3e, se señala que la primera señal procesada o la segunda señal procesada sea la salida del segundo ramal de codificación y por ello la segunda señal codificada.

35 Se prefiere que el primer ramal de decodificación y/o el segundo ramal de procesamiento inverso incluyan una transformada de MDCT para convertir desde el dominio espectral al dominio del tiempo. Para este fin se proporciona un solapamiento-agregador para llevar a cabo una funcionalidad de cancelación de repliegue en el dominio del tiempo que, al mismo tiempo, proporciona un efecto de desvanecimiento cruzado para evitar artefactos de bloqueo. En términos generales, el primer ramal de decodificación convierte una señal codificada en el cuarto dominio en el primer dominio, mientras que el segundo ramal de procesamiento inverso lleva a cabo una conversión desde el tercer dominio al segundo dominio y el conversor posteriormente conectado al primer combinador proporciona una conversión desde el segundo dominio al primer dominio de manera tal que, en la entrada del combinador 600, solo hay señales del primer dominio, que en la forma de realización de la figura 3d representan la señal de salida decodificada.

45 La figura 4c ilustra otro aspecto de una implementación de decodificador preferida. Para evitar artefactos audibles específicamente en una situación en la que el primer decodificador es un decodificador de generación de repliegue de tiempo o indicado en términos generales un decodificador en el dominio de la frecuencia y el segundo decodificador es un dispositivo en el dominio del tiempo, los límites entre los bloques o tramas emitidos por el primer decodificador 450 y el segundo decodificador 550 no deberían ser completamente continuos, específicamente en una situación de conmutación. Por lo tanto, cuando se emite el primer bloque del primer decodificador 450 y, cuando para la siguiente parte de tiempo se emite un bloque del segundo decodificador, se prefiere llevar a cabo una operación de desvanecimiento cruzado como se ilustra en el bloque 607 de desvanecimiento cruzado. Para este fin, el bloque 607 de desvanecimiento cruzado podría implementarse como se ilustra en la figura 4c en 607a, 607b y 55 607c. Cada ramal podría tener un ponderador que tiene un factor de ponderación m_1 entre 0 y 1 en la escala normalizada, donde el factor de ponderación puede variar como se indica en el gráfico 609, de manera que una regla de desvanecimiento cruzado de este tipo asegura que tiene lugar un desvanecimiento cruzado continuo y suave, que adicionalmente asegura que un usuario no percibirá ninguna variación de sonoridad. Pueden aplicarse reglas de desvanecimiento cruzado no lineales tales como una regla de desvanecimiento cruzado sen^2 , en lugar de una regla de desvanecimiento cruzado lineal.

60 En determinados casos, el último bloque del primer decodificador se generó mediante el uso de una ventana donde la ventana realmente llevó a cabo un desvanecimiento de este bloque. En este caso, el factor de ponderación m_1 en el bloque 607a es igual a 1, y realmente, no se requiere ninguna ponderación en absoluto para este ramal. Cuando

tiene lugar una conmutación desde el segundo decodificador hacia el primer decodificador, y cuando el segundo decodificador incluye una ventana que realmente desvanece la salida hacia el extremo del bloque, entonces el ponderador indicado como "m₂" no sería necesario, o el parámetro de ponderación puede establecerse a 1 en la totalidad de la región de desvanecimiento.

5 Cuando se generó el primer bloque después de una conmutación mediante el uso de una operación de ventana, y cuando esta ventana realmente llevó a cabo una operación de aparición gradual, entonces también es posible establecer el correspondiente factor de ponderación a 1, por lo que un ponderador no es realmente necesario. Por
10 ello, cuando el último bloque es objeto de una operación de ventana para desvanecerse mediante el decodificador y cuando el primer bloque después de la conmutación es objeto de una operación de ventana mediante el uso del decodificador para proporcionar una aparición gradual, entonces los ponderadores 607a, 607b no se requieren en absoluto y es suficiente una operación de adición por el agregador 607c.

15 En este caso, la parte de desvanecimiento de la última trama y la parte de aparición gradual de la siguiente trama, definen la región de desvanecimiento cruzado indicada en el bloque 609. Por otra parte, se prefiere en una situación de este tipo que el último bloque de un decodificador tenga un determinado solapamiento de tiempo con el primer bloque del otro decodificador.

20 Si no se requiere una operación de desvanecimiento cruzado, o no es posible o no es deseada, y si sólo hay una conmutación definitiva de uno de los decodificadores al otro decodificador, se prefiere llevar a cabo una conmutación de este tipo en los pasajes silenciosos de la señal de audio o al menos en los pasajes de la señal de audio en los que hay una energía baja, es decir, que se perciben como silenciosos o casi silenciosos. En una forma de realización de este tipo, se prefiere que la etapa 300 de decisión asegure que el conmutador 200 se active
25 solamente cuando la correspondiente parte de tiempo que sigue al acontecimiento de conmutación tenga una energía que sea, por ejemplo, inferior a la energía media de la señal de audio, y sea preferentemente inferior al 50 % de la energía media de la señal de audio relacionada con, por ejemplo, dos o incluso más partes/tramas de tiempo de la señal de audio.

30 Se prefiere que la segunda regla de codificación/regla de decodificación sea un algoritmo de codificación basado en LPC. En la codificación del habla basada en LPC, se efectúa una diferenciación entre partes de señales o segmentos de señales de excitación similares a impulsos casi periódicos, y partes de señales o segmentos de señales de excitación similares a ruido. Esto se lleva a cabo para codificadores de voz de LPC con una tasa de bits muy baja (2,4 kbps) como en la figura 7b. Sin embargo, en los codificadores de CELP de tasa media, la excitación se obtiene para la adición de vectores escalados a partir de un libro de código adaptativo y un libro de código fijo.
35

Los segmentos de señales de excitación similares a impulsos casi periódicos, es decir, los segmentos de señales que tienen un tono específico, se codifican con mecanismos que son diferentes de los de las señales de excitación similares a ruido. Mientras que las señales de excitación similares a impulsos casi periódicos se conectan al habla vocalizada, las señales similares a ruido están relacionadas con el habla no vocalizada.
40

A modo de ejemplo se hace referencia a las figuras 5a a 5d. En este punto, se analizan a modo de ejemplo partes de señales o segmentos de señales similares a impulsos casi periódicos y partes de señales o segmentos de señales similares a ruido. Específicamente, un habla vocalizada como se ilustra en la figura 5a en el dominio del tiempo y en la figura 5b en el dominio de la frecuencia, se analiza como un ejemplo para una parte de señal similar a un impulso casi periódico, y se analiza un segmento del habla no vocalizado como un ejemplo de una parte de señal similar a ruido en relación con las figuras 5c y 5d. En términos generales, el habla puede clasificarse como vocalizada, no vocalizada, o mixta. En las figuras 5a a 5d se muestran gráficos de dominio de tiempo y frecuencia para segmentos vocalizados y no vocalizados muestreados. El habla vocalizada es casi periódica en el dominio del tiempo y está estructurada armónicamente en el dominio de la frecuencia, mientras que el habla no vocalizada es aleatoria y de banda ancha. El espectro de tiempo corto del habla vocalizada se caracteriza por su estructura fina y formante. La estructura armónica fina es una consecuencia de la casi periodicidad del habla, y puede atribuirse a las cuerdas vocales vibratorias. La estructura formante (envolvente espectral) se debe a la interacción de la fuente y al tracto vocal. Los tractos vocales consisten en la faringe y la cavidad de la boca. La forma de la envolvente espectral que "encaja" el espectro de tiempo breve del habla vocalizada, está asociada con las características de transferencia del tracto vocal y de la pendiente espectral (6 dB /octava) debido al pulso glótico. La envolvente espectral se caracteriza por un conjunto de picos que se denominan formantes. Los formantes son los modos resonantes del tracto vocal. Para el tracto vocal promedio hay de tres a cinco formantes inferiores a 5 kHz. Las amplitudes y localizaciones de los tres primeros formantes, que habitualmente tienen lugar por debajo de los 3 kHz, son ambas bastante importantes, en la síntesis y percepción del habla. Los formantes superiores son también importantes para representaciones de banda ancha y del habla no vocalizada. Las propiedades del habla están relacionadas con el sistema de producción del habla física, como sigue. El habla vocalizada se produce excitando el tracto vocal con pulsos de aire glóticos casi periódicos generados por las cuerdas vocales vibratorias. La frecuencia de los pulsos periódicos se denomina frecuencia o tono fundamental. El habla no vocalizada se produce forzando el paso del aire a través de una constricción en el tracto vocal. Los sonidos nasales se deben al acoplamiento acústico del tracto
50
55
60

nasal al tracto vocal, y los sonidos plosivos se producen mediante la liberación abrupta de la presión de aire que se había acumulado detrás del cierre en el tracto.

Por lo tanto, una parte similar a ruido de la señal de audio no muestra ni estructura en el dominio del tiempo similar a impulsos ni estructura en el dominio de la frecuencia armónica, como se ilustra en la figura 5c y en la figura 5d, que es diferente de la parte similar a impulso casi periódica como se ilustra por ejemplo en la figura 5a y en la figura 5b. Sin embargo, y como se señalará más adelante, la diferenciación entre las partes similares a ruido y las partes similares a pulsos casi periódicos, también puede observarse después de un LPC para la señal de excitación. El LPC es un método que modela el tracto vocal y extrae de la señal la excitación de los tractos vocales.

Por otro lado, las partes similares a impulsos casi periódicos y las partes similares a ruido pueden presentarse de una manera oportuna, es decir, lo que significa que una parte de la señal de audio en el tiempo es ruidosa y otra parte de la señal de audio en el tiempo es casi periódica, es decir, tonal. Como alternativa o adicionalmente, la característica de una señal puede ser diferente en diferentes bandas de frecuencia. Por lo tanto, la determinación de si la señal de audio es ruidosa o tonal, también puede llevarse a cabo de manera selectiva en frecuencia de manera que una determinada banda de frecuencia o varias bandas de frecuencia se consideran ruidosas y otras bandas de frecuencia se consideran tonales. En este caso, una determinada parte de tiempo de la señal de audio podría incluir componentes tonales y componentes ruidosos.

La figura 7a ilustra un modelo lineal de un sistema de producción del habla. Este sistema supone una excitación en dos etapas, es decir, un tren de impulsos para el habla vocalizada como se indica en la figura 7c, y un ruido aleatorio para el habla no vocalizada como se indica en la figura 7d. El tracto vocal se modela como un filtro 70 todo polo que procesa los pulsos de la figura 7c o de la figura 7d, generados por el modelo 72 glótico. Por lo tanto, el sistema de la figura 7a puede reducirse a un modelo de filtro todo polo de la figura 7b que tiene una etapa 77 de ganancia, una trayectoria 78 hacia delante, una trayectoria 79 de retroalimentación, y una etapa 80 de adición. En la trayectoria 79 de retroalimentación, hay un filtro 81 de predicción, y la totalidad del sistema de síntesis del modelo de fuente, ilustrado en la figura 7b, puede representarse mediante el uso de funciones en el dominio de z, como sigue:

$$S(z)=g/(1-A(z))\cdot X(z),$$

donde g representa la ganancia, A(z) es el filtro de predicción determinado mediante un análisis de LP, X(z) es la señal de excitación, y S(z) es la salida del habla de la síntesis.

Las figuras 7c y 7d dan una descripción gráfica del dominio del tiempo de la síntesis del habla, vocalizada y no vocalizada, mediante el uso del modelo del sistema de fuente lineal. Este sistema y los parámetros de excitación en la ecuación anterior, son desconocidos y han de determinarse a partir de un conjunto finito de muestras del habla. Los coeficientes de A(z) se obtienen mediante una predicción lineal de la señal de entrada y una cuantificación de los coeficientes de filtro. En un predictor lineal directo de orden p, la muestra presente de la secuencia del habla se predice a partir de una combinación lineal de p muestras pasadas. Los coeficientes predictores pueden determinarse mediante algoritmos bien conocidos tales como el algoritmo de Levinson-Durbin, o en términos generales mediante un método de autocorrelación o un método de reflexión.

La figura 7e ilustra una implementación más detallada del bloque 510 de análisis de LPC. La señal de audio se introduce en un bloque de determinación de filtro que determina la información de filtro A(z). Esta información se emite como la información de predicción a corto plazo requerida para un decodificador. Esta información se cuantifica mediante un cuantificador 81 como es conocido, por ejemplo a partir de la especificación AMR-WB+. La información de predicción a corto plazo se requiere por el filtro 85 de predicción real. En un restador 86, se introduce una muestra actual de la señal de audio, y se resta un valor predicho para la muestra actual de modo que para esta muestra, la señal de error de predicción se genera en la línea 84. En la figura 7c o 7d se ilustra de manera muy esquemática una secuencia de tales muestras de señal de error de predicción. Por ello, las figuras 7c, 7d pueden considerarse como un tipo de señal similar a impulso rectificadas.

Mientras que la figura 7e ilustra una manera preferida para calcular la señal de excitación, la figura 7f ilustra una manera preferida para calcular la señal ponderada. A diferencia de la figura 7e, el filtro 85 es diferente, cuando γ es diferente de 1. Para γ se prefiere un valor inferior a 1. Por otra parte, el bloque 87 está presente, y se prefiere que μ sea un valor inferior a 1. En términos generales, los elementos en las figuras 7e y 7f pueden implementarse como en 3GPP TS 26.190 o en 3GPP TS 26.290.

La figura 7g ilustra un procesamiento inverso que puede aplicarse en el lado del decodificador tal como en el elemento 537 de la figura 2b. En particular, el bloque 88 genera una señal no ponderada a partir de la señal ponderada, y el bloque 89 calcula una excitación a partir de la señal no ponderada. En términos generales, todas las señales, con la excepción de la señal no ponderada en la figura 7g se hallan en el dominio de LPC, pero la señal de excitación y la señal ponderada son señales diferentes en el mismo dominio. El bloque 89 emite una señal de excitación que puede a continuación utilizarse junto con la salida del bloque 536. A continuación puede llevarse a cabo la transformada de LPC inversa común en el bloque 540 de la figura 2b.

Posteriormente se analizará un codificador de CELP de análisis por síntesis en relación con la figura 6 para ilustrar las modificaciones aplicadas a este algoritmo. Este codificador CELP se analiza con detalle en: "Speech Coding: A Tutorial Review", Andreas Spanias, Proceedings of the IEEE, Vol. 82, n.º 10, octubre de 1994, páginas 1541-1582.

5 El codificador CELP como se ilustra en la figura 6 incluye un componente 60 de predicción a largo plazo y un componente 62 de predicción a corto plazo. Además, se utiliza un libro de código que se indica en 64. Un filtro de ponderación perceptual $W(z)$ se ha implementado en 66, y se ha proporcionado un controlador de la minimización de errores en 68. $s(n)$ es la señal de entrada en el dominio del tiempo. Después de haberse ponderado perceptualmente, la señal ponderada se introduce en un restador 69, que calcula el error entre la señal de síntesis ponderada en la salida del bloque 66 y la señal original ponderada $s_w(n)$. En términos generales, los coeficientes de filtro de predicción a corto plazo $A(z)$ se calculan mediante una etapa de análisis de LP, y sus coeficientes se cuantifican en $\hat{A}(z)$, como se indica en la figura 7e. La información de predicción a largo plazo $A_L(z)$ que incluye la ganancia de predicción a largo plazo g y el índice de cuantificación de vector, es decir, las referencias del libro de código se calculan en la señal de error de predicción en la salida de la etapa de análisis de LPC indicada como 10a en la figura 7e. Los parámetros de LTP son el retardo de tono y la ganancia. En el CELP esto se implementa habitualmente en forma de un libro de código adaptativo que contiene la señal de excitación anterior (no la residual). El retardo de CB adaptativo y la ganancia se encuentran mediante la minimización del error ponderado cuadrático medio (búsqueda de tono de bucle cerrado).

20 El algoritmo de CELP codifica a continuación la señal residual obtenida después de las predicciones a corto plazo y a largo plazo mediante el uso de un libro de código de, por ejemplo, secuencias gaussianas. El algoritmo de ACELP, donde "A" representa "Algebraico" tiene un libro de código específico diseñado algebraicamente.

Un libro de código puede contener más o menos vectores donde cada vector tiene una longitud de varias muestras. 25 Un factor de ganancia, g , escala el vector de código, y el código con ganancia se filtra mediante el filtro de síntesis de predicción a largo plazo y el filtro de síntesis de predicción a corto plazo. El vector de código "óptimo" se selecciona de manera que se minimiza el error cuadrático medio perceptualmente ponderado en la salida del restador 69. El proceso de búsqueda en el CELP se efectúa mediante una optimización de análisis por síntesis como se ilustra en la figura 6.

30 Para casos específicos, cuando una trama es una mezcla del habla no vocalizada y vocalizada o cuando tiene lugar un habla sobre la música, una codificación de TCX puede ser más adecuada para codificar la excitación en el dominio de LPC. La codificación de TCX procesa la señal ponderada en el dominio de la frecuencia sin efectuar ninguna suposición en cuanto a la producción de la excitación. El TCX es por lo tanto más genérico que la codificación de CELP, y no se restringe a un modelo de fuente vocalizado o no vocalizado de la excitación. El TCX sigue siendo una codificación de modelo de fuente-filtro que utiliza un filtro predictivo lineal para modelar los formantes de las señales similares al habla.

40 En la codificación similar a AMR-WB+, tiene lugar una selección entre diferentes modos de TCX y ACELP, como es conocido a partir de la descripción del AMR-WB+. Los modos de TCX son diferentes, ya que la longitud de la Transformada Discreta de Fourier a nivel de bloques, es diferente para diferentes modos, y puede seleccionarse el mejor modo mediante un enfoque de análisis por síntesis o mediante un modo de "alimentación hacia delante" directa.

45 Como se analizó en relación con las figuras 2c y 2d, se prefiere que la etapa 100 de procesamiento previo común incluya un multicanal 101 conjunto (dispositivo estéreo conjunto/envolvente) y, adicionalmente, una etapa 102 de ampliación de ancho de banda. De manera correspondiente, el decodificador incluye una etapa 701 de ampliación de ancho de banda y una etapa 702 multicanal conjunta conectada posteriormente. Se prefiere que la etapa 101 de multicanal conjunta esté conectada, con respecto al codificador, antes de la etapa 102 de ampliación del ancho de banda, y, en el lado del decodificador, que la etapa 701 de ampliación de ancho de banda esté conectada antes de la etapa 702 de multicanal conjunta con respecto a la dirección de procesamiento de la señal. Sin embargo, como alternativa, la etapa de procesamiento previo común puede incluir una etapa multicanal conjunta sin la etapa de ampliación de ancho de banda posteriormente conectada o una etapa de ampliación de ancho de banda sin una etapa multicanal conjunta conectada.

55 En el contexto de la figura 8 se ilustra un ejemplo preferido para una etapa multicanal conjunta en el lado del codificador 101a, 101b y en el lado del decodificador 702a y 702b. Se introduce un número de E canales de entrada originales en el mezclador 101a descendente de manera que el mezclador descendente genere un número de K canales transmitidos, donde el número K es mayor o igual a uno, y es inferior o igual a E .

60 Se prefiere que los E canales de entrada se introduzcan en un analizador 101b conjunto de parámetros multicanal, que genera información paramétrica. Se prefiere que esta información paramétrica se codifique por entropía tal como mediante una codificación diferente y posterior codificación de Huffman o, como alternativa, posterior codificación aritmética. La información paramétrica codificada emitida por el bloque 101d se transmite a un

decodificador 702b de parámetros que puede ser parte del elemento 702 en la figura 2b. El decodificador 702b de parámetros decodifica la información paramétrica transmitida, y reenvía la información paramétrica decodificada al mezclador 702a ascendente. El mezclador 702a ascendente recibe los K canales transmitidos y genera un número de L canales de salida, donde el número L mayor que o igual a K, e inferior o igual a E.

5 La información paramétrica puede incluir diferencias de nivel entre canales, diferencias de tiempo entre canales, diferencias de fase entre canales y/o medidas de coherencia entre canales, como es conocido a partir de la técnica de BCC o como es conocido y describe en detalle en la norma de MPEG surround. La cantidad de canales transmitidos puede ser un sólo canal mono para aplicaciones de tasa de bits ultra baja, o puede incluir una aplicación estéreo compatible, o puede incluir una señal estéreo compatible, es decir, dos canales. Normalmente, el número de E canales de entrada puede ser de cinco o tal vez incluso mayor. Como alternativa, el número de E canales de salida también puede ser E objetos de audio, como es conocido en el contexto de la codificación de objeto audio espacial (SAOC).

15 En una implementación, el mezclador descendente lleva a cabo una adición, ponderada o sin ponderar, de los E canales de entrada originales o una adición de los E objetos de audio de entrada. En el caso de los objetos de audio como canales de entrada, el analizador 101b conjunto de los parámetros multicanal, calculará preferentemente los parámetros de objetos de audio tales como una matriz de correlación entre los objetos de audio, preferentemente para cada parte de tiempo e incluso más preferentemente para cada banda de frecuencia. Para este fin, es posible dividir la totalidad del intervalo de frecuencia en al menos 10 y preferentemente 32 o 64 bandas de frecuencia.

25 La figura 9 ilustra una forma de realización preferida para la implementación de la etapa 102 de ampliación del ancho de banda, en la figura 2a y la correspondiente etapa 701 de ampliación de ancho de banda, en la figura 2b. En el lado del codificador, se prefiere que el bloque 102 de ampliación de ancho de banda incluya un bloque 102b de filtración de paso bajo, un bloque de muestreador descendente, que sigue al paso bajo, o que es parte del QMF inverso, que actúa sobre solamente la mitad de las bandas de QMF, y un analizador 102a de banda alta. La señal de audio original introducida en el bloque 102 de ampliación de ancho de banda es objeto de filtrado de paso bajo para generar la señal de banda baja que se introduce a continuación en los ramales de codificación y/o el conmutador. El filtro de paso bajo tiene una frecuencia de corte que puede estar en un intervalo de 3 kHz a 10 kHz. Por otra parte, el bloque 102 de ampliación de ancho de banda, incluye además un analizador de banda alta para calcular los parámetros de ampliación de ancho de banda tales como una información sobre los parámetros de envolvente espectral, una información sobre los parámetros de margen de ruido, una información sobre los parámetros de filtración inversa, otra información paramétrica relacionada con determinadas líneas armónicas en la banda alta y parámetros adicionales, tal como se analizó en detalle en la norma MPEG-4 en el capítulo relacionado con la replicación de banda espectral.

35 En el lado del decodificador, el bloque 701 de ampliación de ancho de banda incluye un parcheador 701a, un ajustador 701b y un combinador 701c. El combinador 701c combina la señal de banda baja decodificada y la señal de banda alta reconstruida y ajustada emitida por el ajustador 701b. La entrada en el ajustador 701b se proporciona por un parcheador que se hace funcionar para derivar la señal de banda alta de la señal de banda baja, tal como mediante replicación de banda espectral, o en términos generales, mediante una ampliación de ancho de banda. El parcheado llevado a cabo por el parcheador 701a puede ser un parcheado llevado a cabo de una manera armónica o de una manera no armónica. La señal generada por el parcheador 701a se ajusta posteriormente por el ajustador 701b mediante el uso de la información paramétrica transmitida sobre la ampliación del ancho de banda.

45 Como se indica en las figuras 8 y 9, en una forma de realización preferida los bloques descritos pueden tener una entrada de control de modo. Esta entrada de control de modo se deriva de la señal de salida de la etapa 300 de decisión. En una forma de realización preferida de este tipo, una característica de un bloque correspondiente puede adaptarse a la salida de la etapa de decisión, es decir, si en una forma de realización preferida, se toma una decisión acerca del habla o una decisión acerca de la música para una determinada parte de tiempo de la señal de audio. Se prefiere que el control de modo se refiera solamente a una o más de las funcionalidades de estos bloques pero no a la totalidad de las funcionalidades de los bloques. Por ejemplo, la decisión puede influir solamente sobre el parcheador 701a, pero es posible que no influya sobre los otros bloques en la figura 9, o puede por ejemplo influir solamente sobre el analizador 101b conjunto de los parámetros de multicanal, en la figura 8 pero no sobre los otros bloques en la figura 8. Se prefiere que esta implementación sea tal que se obtenga una señal de salida de mayor flexibilidad, mayor calidad y menor tasa de bits, al proporcionar flexibilidad en la etapa de procesamiento previo común. Sin embargo, por otra parte el uso de algoritmos en la etapa de procesamiento previo común para ambos tipos de señales, permite implementar un esquema de codificación/decodificación eficiente.

60 Las figuras 10a y 10b ilustran dos implementaciones diferentes de la etapa 300 de decisión. En la figura 10a, se indica una decisión de bucle abierto. En este caso, el analizador 300a de señales en la etapa de decisión tiene determinadas reglas para decidir si la determinada parte de tiempo o una determinada parte de frecuencia de la señal de entrada tiene una característica que requiere que esta parte de señal esté codificada por el primer ramal 400 de codificación o por el segundo ramal 500 de codificación. Para este fin, el analizador 300a de señales puede

analizar la señal de audio de entrada en la etapa de procesamiento previo común, o puede analizar la señal de audio emitida por la etapa de procesamiento previo común, es decir, la señal intermedia de audio, o puede analizar una señal intermedia dentro de la etapa de procesamiento previo común tal como la salida de la señal de mezclado descendente que puede ser una señal mono o que puede ser una señal que tiene k canales indicadas en la figura 8.

5 En el lado de salida, el analizador 300a de señales genera la decisión de conmutación para controlar el conmutador 200 en el lado del codificador o el correspondiente conmutador 600 o el combinador 600 en el lado del decodificador.

10 Como alternativa, la etapa 300 de decisión puede llevar a cabo una decisión de bucle cerrado, lo que significa que ambos ramales de codificación llevan a cabo sus tareas sobre la misma parte de la señal de audio y ambas señales codificadas se decodifican por los correspondientes ramales 300c, 300d de decodificación. La salida de los dispositivos 300c y 300d se introduce en un comparador 300b que compara la salida de los dispositivos de decodificación para colocar la correspondiente parte de, por ejemplo, la señal intermedia de audio. A continuación, y en función de una función de coste tal como una relación de señal a ruido para cada ramal, se lleva a cabo una decisión de conmutación. Esta decisión de bucle cerrado tiene una complejidad incrementada en comparación con la
15 decisión de bucle abierto, pero esta complejidad existe solamente en el lado del codificador, y un decodificador no tiene ninguna desventaja debido a este procesamiento ya que el decodificador puede utilizar de manera ventajosa la salida de esta decisión de codificación. Por ello se prefiere el modo de bucle cerrado debido a consideraciones de complejidad y calidad en aquellas aplicaciones en las que la complejidad del decodificador no es un problema como en aplicaciones de difusión donde existe solamente una pequeña cantidad de codificadores pero existe una gran
20 cantidad de decodificadores, que además han de ser inteligentes y baratos.

La función de coste aplicada por el comparador 300d puede ser una función de coste basada en aspectos de calidad, o puede ser una función de coste basada en aspectos de ruido, o puede ser una función de coste basada en aspectos de tasa de bits, o puede ser una función de coste combinada basada en cualquier combinación de tasa de bits, calidad, ruido (introducido por los artefactos de codificación, específicamente, por la cuantificación), etc.
25

Se prefiere que el primer ramal de codificación o el segundo ramal de codificación incluyan una funcionalidad de distorsión del tiempo en el lado del codificador y correspondientemente en el lado del decodificador. En una forma de realización, el primer ramal de codificación comprende un módulo de distorsión del tiempo para calcular una característica de distorsión variable que depende de una parte de la señal de audio, un remuestreador para muestrear nuevamente según la característica de distorsión determinada, un conversor de dominio del tiempo/dominio de la frecuencia, y un codificador de entropía para convertir un resultado de la conversión de dominio del tiempo/dominio de la frecuencia en una representación codificada. La característica variable de distorsión está incluida en la señal de audio codificada. Esta información se lee en un ramal de decodificación reforzada en distorsión de tiempo, y procesada para tener finalmente una señal de salida en una escala de tiempo no distorsionada. Por ejemplo, el ramal de decodificación lleva a cabo una decodificación de entropía, descuantificación y una conversión a partir del dominio de la frecuencia de regreso al dominio del tiempo. En el dominio del tiempo, puede aplicarse una operación inversa a la distorsión, y la misma puede ser seguida por una correspondiente operación de remuestreo para obtener finalmente una señal de audio discreta con una escala de tiempo no distorsionada.
30
35
40

En función de determinados requisitos para la implementación de los métodos inventivos, los métodos inventivos pueden implementarse en hardware o en software. La implementación puede llevarse a cabo mediante el uso de un medio de almacenamiento digital, en particular, un disco, un DVD o un CD que tenga señales de control electrónicamente legibles almacenadas en ellos, que cooperen con sistemas informáticos programables de manera que los métodos inventivos se lleven a cabo. En términos generales, la presente invención es por lo tanto un producto de programa informático con un código de programa almacenado en un portador legible por máquina, operándose el código de programa para llevar a cabo los métodos inventivos cuando el producto de programa informático se ejecuta en un ordenador. En otras palabras, los métodos inventivos son por ello un programa informático que tiene un código de programa para llevar a cabo al menos uno de los métodos inventivos cuando el programa informático se ejecuta en un ordenador.
45
50

La señal de audio codificada inventiva puede almacenarse en un medio de almacenamiento digital, o puede transmitirse en un medio de transmisión tal como un medio de transmisión inalámbrica o un medio de transmisión por cable tal como Internet.
55

Las formas de realización anteriormente descritas son meramente ilustrativas de los principios de la presente invención. Se entiende modificaciones y variaciones de las disposiciones y de los detalles descritos en el presente documento serán evidentes para los expertos en la técnica. Por ello la intención es que la invención se limite solamente por el alcance de las reivindicaciones de patente adjuntas, y no por los detalles específicos presentados en modo descriptivo y explicativo de las formas de realización en el presente documento.
60

REIVINDICACIONES

1. Aparato para codificar una señal de audio para obtener una señal de audio codificada, estando la señal de audio en un primer dominio, que comprende:

- 5 un primer conversor (510) de dominio para convertir la señal de audio desde el primer dominio en un segundo dominio;
- una derivación (50) conmutable para puentear el primer conversor (510) de dominio o para provocar una conversión de la señal de audio por el primer conversor (510) de dominio en respuesta a una señal (51) de control de conmutación de la derivación;
- 10 un segundo conversor (410) de dominio para convertir una señal de audio recibida desde la derivación (50) conmutable o el primer conversor (510) de dominio en un tercer dominio, siendo el tercer dominio diferente del segundo dominio;
- un primer procesador (420) para codificar la señal de audio del tercer dominio según un primer algoritmo de codificación para obtener una primera señal procesada; y
- 15 un segundo procesador (520) para codificar la señal de audio recibida desde el primer conversor (510) de dominio según un segundo algoritmo de codificación que es diferente del primer algoritmo de codificación para obtener una segunda señal procesada,
- 20 en el que la señal codificada para una parte de la señal de audio incluye la primera señal procesada o la segunda señal procesada.

2. Aparato según la reivindicación 1, en el que el primer conversor (510) de dominio comprende un filtro de análisis de LPC para la filtración de LPC de la señal de audio para obtener una señal residual de LPC y datos de parámetros de LPC.

25 3. Aparato según la reivindicación 1 o 2, en el que el segundo conversor (410) de dominio comprende un conversor de tiempo-frecuencia para convertir una señal de entrada en una representación espectral de la misma.

30 4. Aparato según una de las reivindicaciones anteriores, en el que el segundo procesador (520) está operativo para generar una señal de salida codificada de manera que la señal de salida codificada esté en el mismo dominio que una señal de entrada en el segundo procesador (520).

35 5. Aparato según una de las reivindicaciones anteriores, en el que el primer procesador (420) comprende un cuantificador y un codificador de entropía y en el que el segundo procesador (520) comprende un codificador de fuente basado en un libro de código.

6. Aparato según una de las reivindicaciones anteriores, en el que el primer procesador (420) se basa en un modelo de sumidero de información y el segundo procesador (520) se basa en un modelo de fuente de información.

40 7. Aparato según una de las reivindicaciones anteriores, que además comprende una etapa (200) de conmutación conectada entre una salida del primer conversor (510) de dominio y una entrada del segundo conversor (410) de dominio y una entrada del segundo procesador (520),

45 en el que la etapa (200) de conmutación está adaptada para conmutar entre la entrada del segundo conversor (410) de dominio y la entrada del segundo procesador (520) en respuesta a una señal de control de la etapa de conmutación.

50 8. Aparato según una de las reivindicaciones anteriores, en el que una salida de la derivación (50) conmutable está conectada a una salida del primer conversor (510) de dominio y una entrada de la derivación (50) conmutable está conectada a una entrada en el primer conversor (510) de dominio.

9. Aparato según una de las reivindicaciones anteriores, que además comprende un clasificador de señales para controlar la derivación (50) conmutable para una parte de la señal de audio en función de un resultado de análisis para la parte de la señal de audio.

55 10. Aparato según una de las reivindicaciones anteriores, en el que el segundo conversor (410) de dominio está operativo para convertir una señal de entrada de una manera basada en bloques y en el que el segundo conversor de dominio está operativo para llevar a cabo una conmutación basada en bloques en respuesta a un análisis de señal de audio de manera que se controla el segundo conversor (410) de dominio porque los bloques de diferentes longitudes se convierten en función del contenido de la señal de audio.

60 11. Método de codificación de una señal de audio para obtener una señal de audio codificada, estando la señal de audio en un primer dominio, que comprende:

convertir (510) la señal de audio del primer dominio en un segundo dominio;

- puentear (50) la etapa de la conversión (510) de la señal de audio del primer dominio en un segundo dominio u ocasionar una conversión de la señal de audio del primer dominio en un segundo dominio en respuesta a una señal (51) de control de conmutación de la derivación;
- 5 convertir (410) una señal (50) de audio derivada o una señal de audio en el segundo dominio en un tercer dominio, siendo el tercer dominio diferente del segundo dominio;
- codificar (420) la señal de audio del tercer dominio generada por la etapa de convertir (410) la señal (50) de audio derivada o la señal de audio en el segundo dominio según un primer algoritmo de codificación para obtener una primera señal procesada; y
- 10 codificar (520) la señal de audio en el segundo dominio según un segundo algoritmo de codificación que es diferente de primer algoritmo de codificación para obtener una segunda señal procesada, en el que la señal codificada para una parte de la señal de audio incluye la primera señal procesada o la segunda señal procesada.
12. Aparato para decodificar una señal de audio codificada, comprendiendo la señal de audio codificada una primera señal procesada que está en un tercer dominio y una segunda señal procesada que está en un segundo dominio, en el que el segundo dominio y el tercer dominio son diferentes entre sí, que comprende:
- 15 un primer procesador (430) inverso para el procesamiento inverso de la primera señal procesada para obtener una primera señal procesada inversa;
- 20 un segundo procesador (530) inverso para el procesamiento inverso de la segunda señal procesada para obtener una segunda señal procesada inversa;
- un segundo conversor (440) para la conversión de dominio de la primera señal procesada inversa del tercer dominio en un dominio diferente;
- 25 un primer conversor (540) para convertir la segunda señal procesada inversa en un primer dominio o para convertir la primera señal procesada inversa, que se convirtió en un dominio diferente, en el primer dominio cuando el dominio diferente no es el primer dominio; y
- una derivación (52) para puentear el primer conversor (540) cuando el dominio diferente es el primer dominio.
13. Aparato según la reivindicación 12, que además comprende un combinador (600) para combinar una salida del primer conversor (540) y una salida de la derivación (52) para obtener una señal (699) de audio decodificada combinada.
- 30
14. Aparato para decodificar según una cualquiera de las reivindicaciones 12 o 13, que además comprende una interfaz (900) de entrada para extraer, de una señal de audio codificada, la primera señal procesada, la segunda señal procesada y la señal de control que indica si para una determinada primera señal procesada inversa, el primer conversor (540) ha de puentearse o no por la derivación.
- 35
15. Aparato para decodificar según una cualquiera de las reivindicaciones 12 a 14, en el que el primer conversor (540) comprende una etapa de síntesis de codificación de predicción lineal (LPC), y
- 40 en el que el segundo conversor (440) comprende un conversor de tiempo espectral para convertir una representación espectral de una señal de audio en una representación de tiempo de la señal de audio.
16. Aparato para decodificar según una cualquiera de las reivindicaciones 12 a 15, en el que el primer procesador (430) inverso comprende un decodificador de entropía y un descuantificador y en el que el segundo procesador (530) inverso comprende el decodificador de fuente basado en un libro de código.
- 45
17. Aparato para decodificar según una cualquiera de las reivindicaciones 12 a 16, en el que el segundo conversor (440) está operativo para llevar a cabo una operación de filtración de síntesis tal como una operación inversa de filtración de transformada de coseno discreta modificada distorsionada en tiempo controlable mediante información (434) adicional incluida en la señal de audio codificada.
- 50
18. Método de decodificación de una señal de audio codificada, comprendiendo la señal de audio codificada una primera señal procesada que está en un tercer dominio y una segunda señal procesada que está en un segundo dominio, en el que el segundo dominio y el tercer dominio son diferentes entre sí, que comprende:
- 55 el procesamiento (430) inverso de la primera señal procesada para obtener una primera señal procesada inversa;
- el procesamiento (530) inverso de la segunda señal procesada para obtener una segunda señal procesada inversa;
- 60 la conversión (440) del segundo dominio de la primera señal procesada inversa del tercer dominio en un dominio diferente;
- la conversión (540) del primer dominio de la segunda señal procesada inversa en un primer dominio o la conversión de la primera señal procesada inversa en el primer dominio cuando el dominio diferente no es el primer dominio; y

puentear (52) la etapa de la conversión (540) del primer dominio cuando el dominio diferente es el primer dominio.

- 5 19. Programa informático para llevar a cabo, cuando se ejecuta en un ordenador, un método de codificación de una señal de audio según la reivindicación 11 o un método de decodificación de una señal de audio codificada según la reivindicación 18.

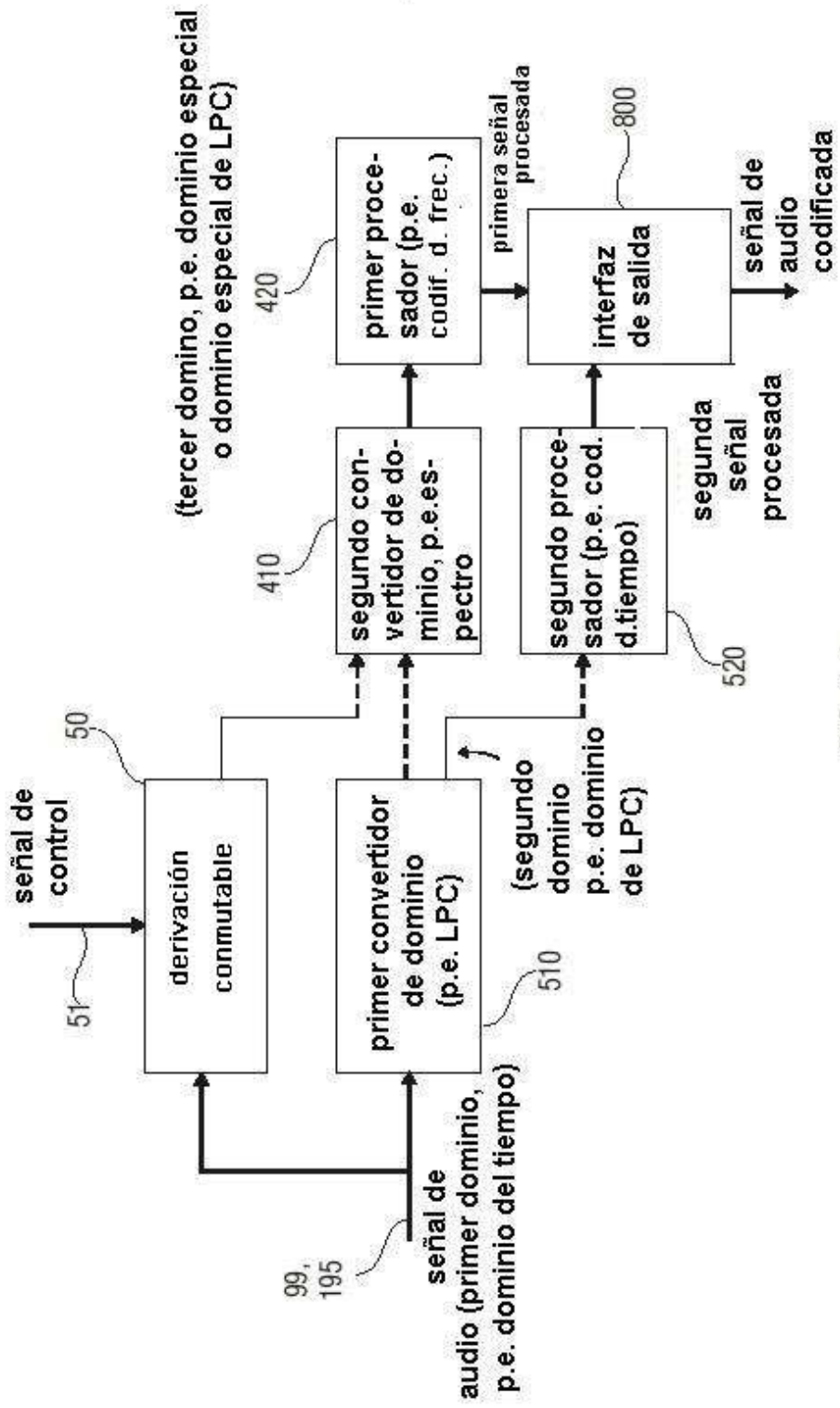


FIG 1A

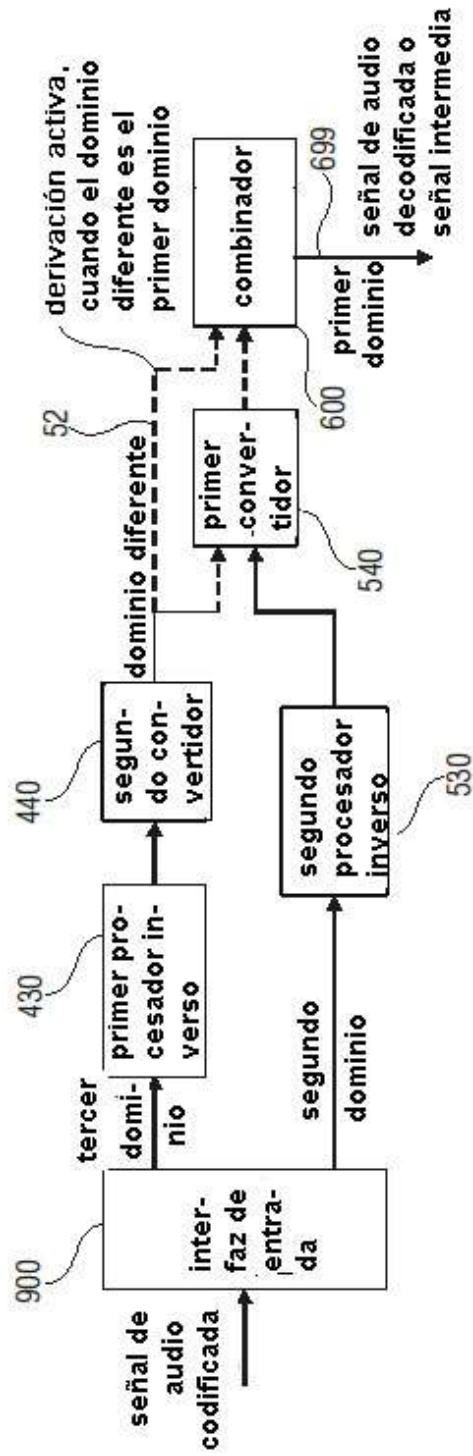


FIG 1B

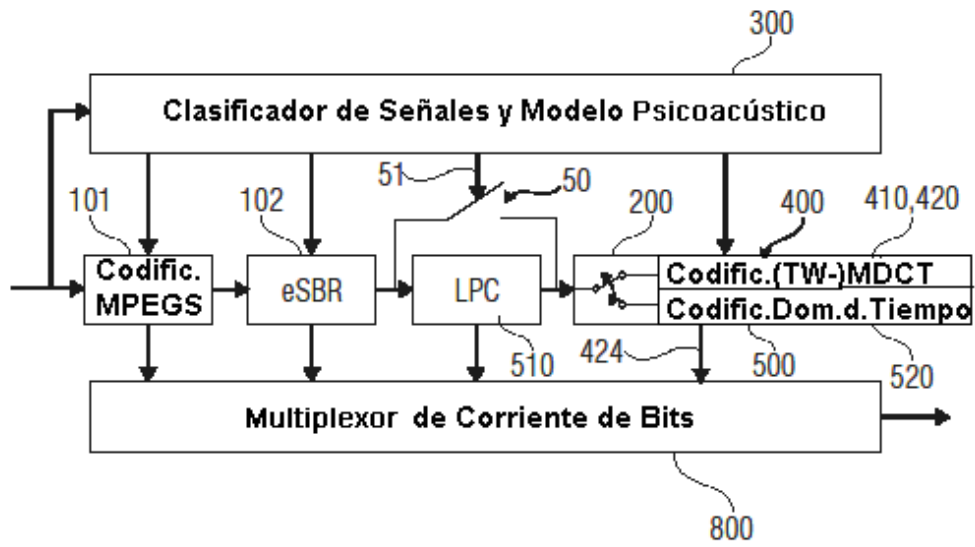


FIG 1C

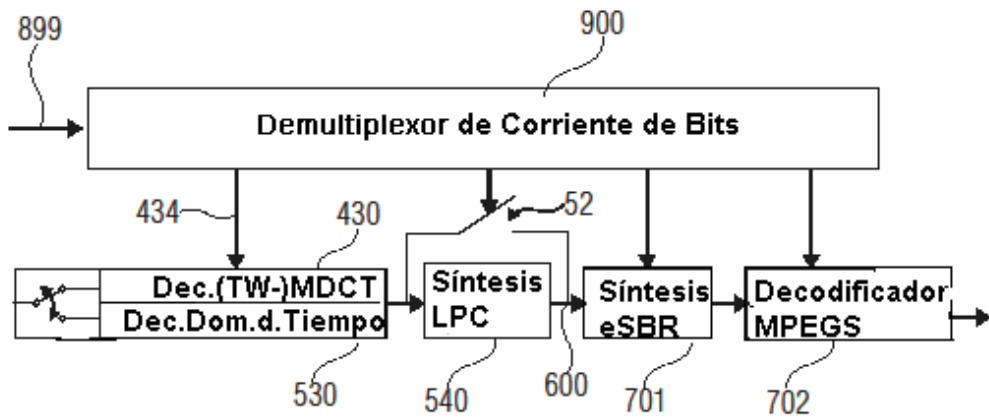


FIG 1D

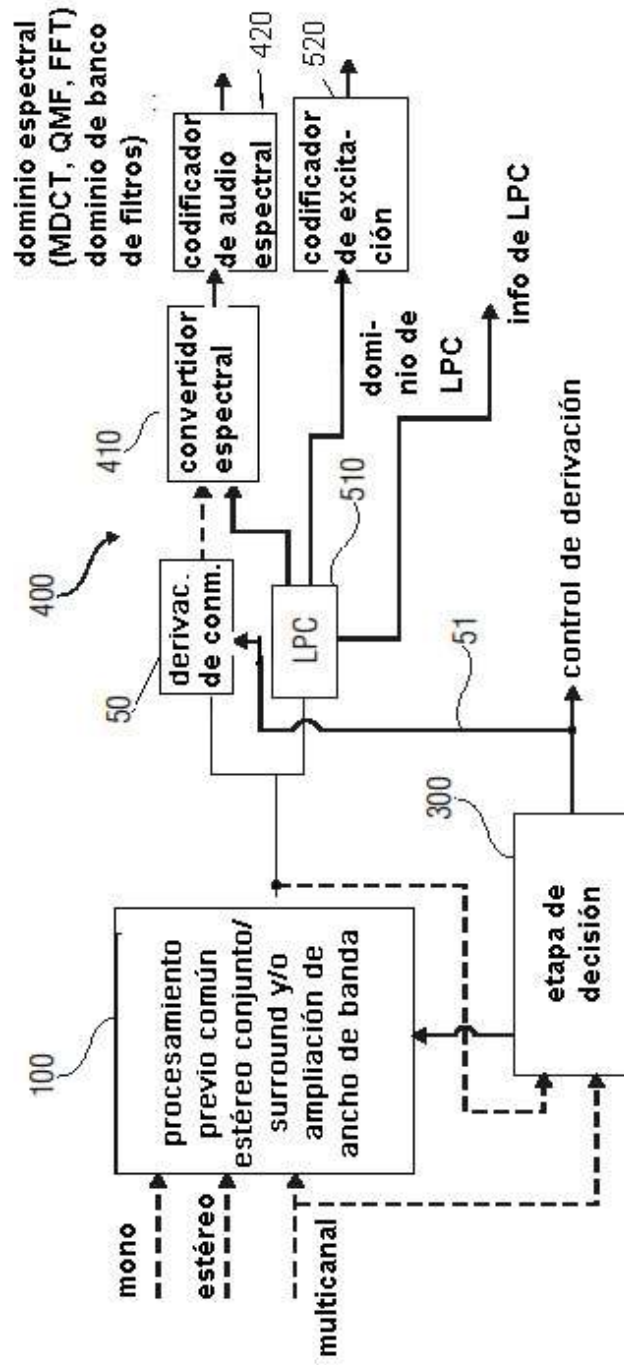


FIG 2A
(Codificador)

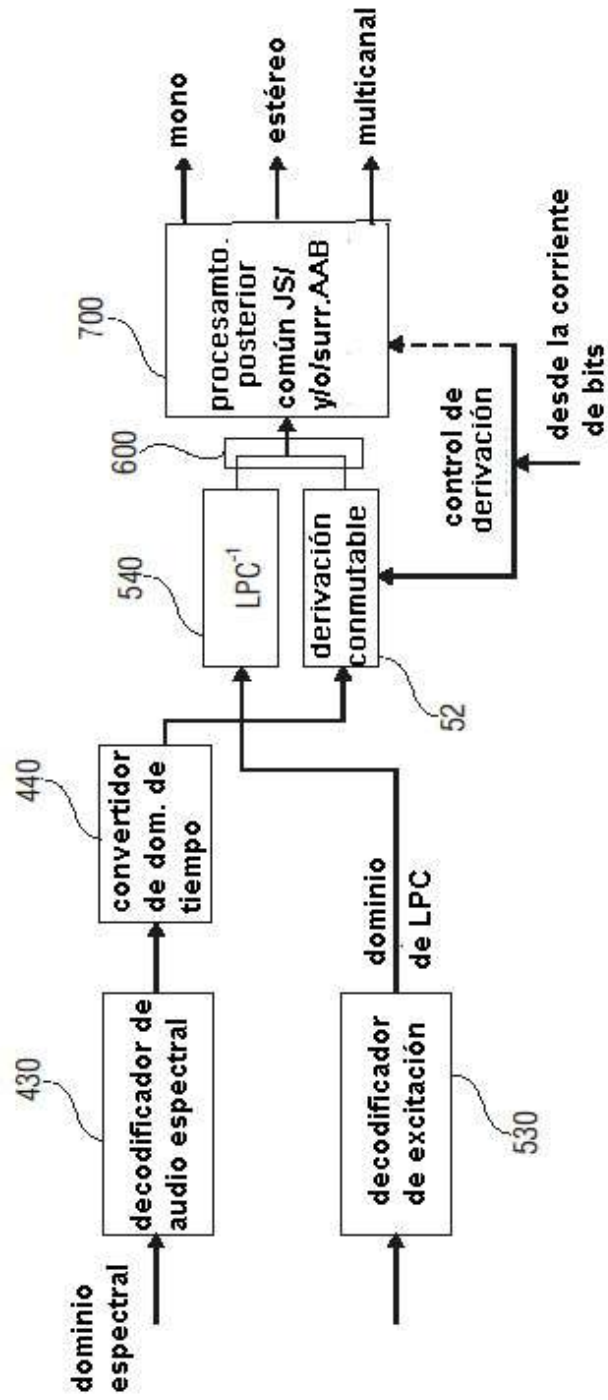


FIG 2B
(Decodificador)

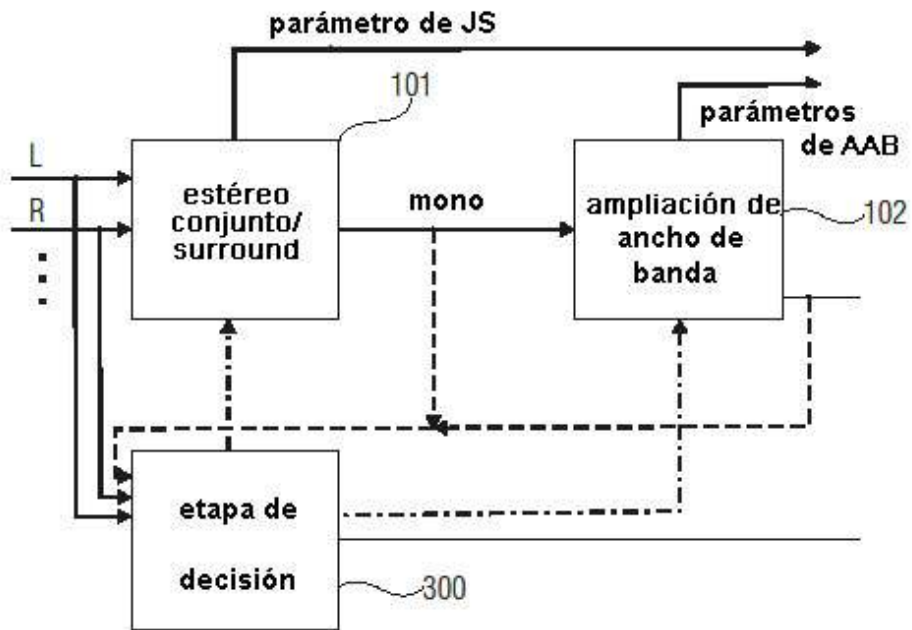


FIG 2C
(Codificador)

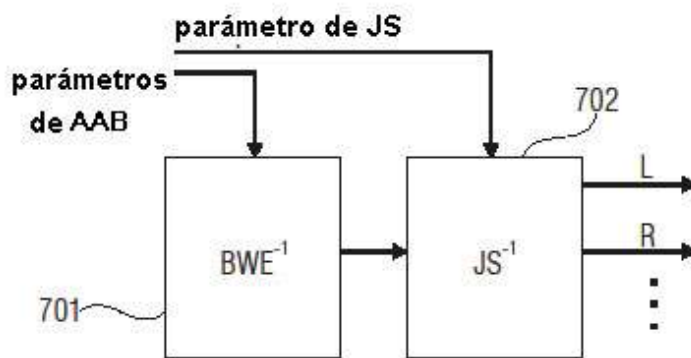


FIG 2D
(Decodificador)

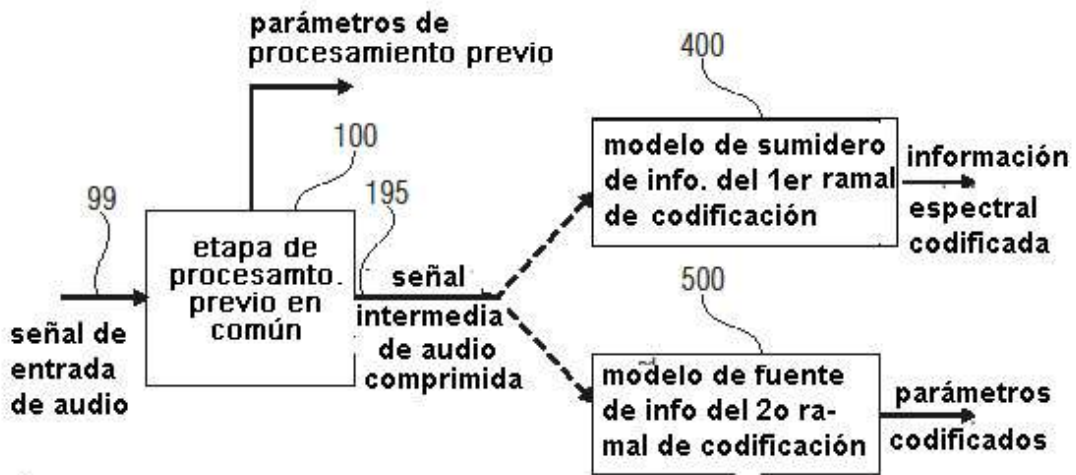


FIG 3A

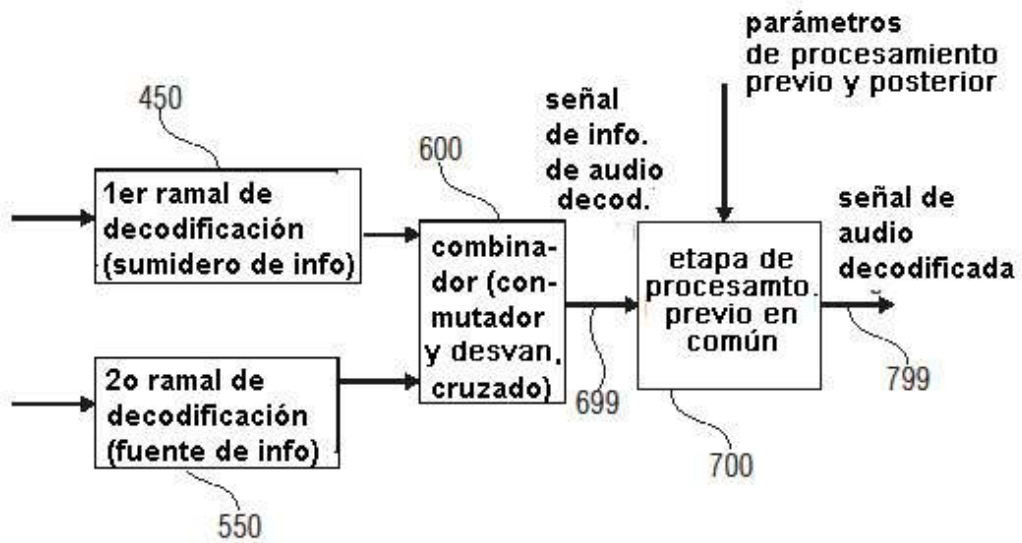
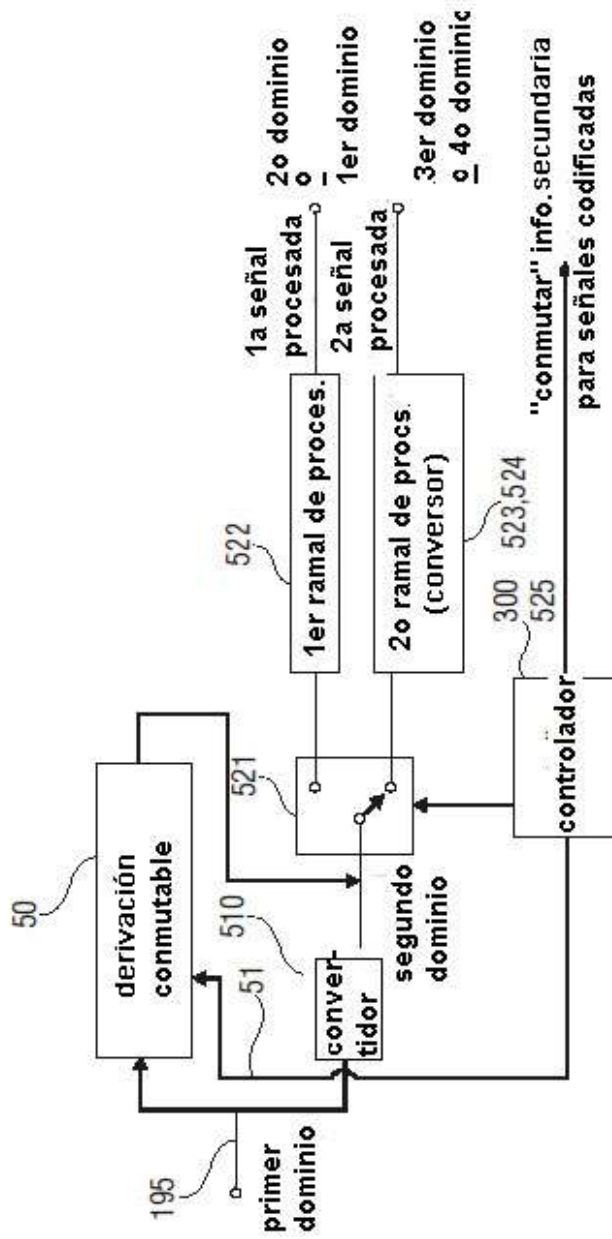


FIG 3B



* cada bloque de la señal de audio del 1er dominio está representado por una señal codificada de un 1er dominio o de un 2o dominio, de un 3er dominio o de un 4o dominio, aparte de una región de cruce opcional.

FIG 3C

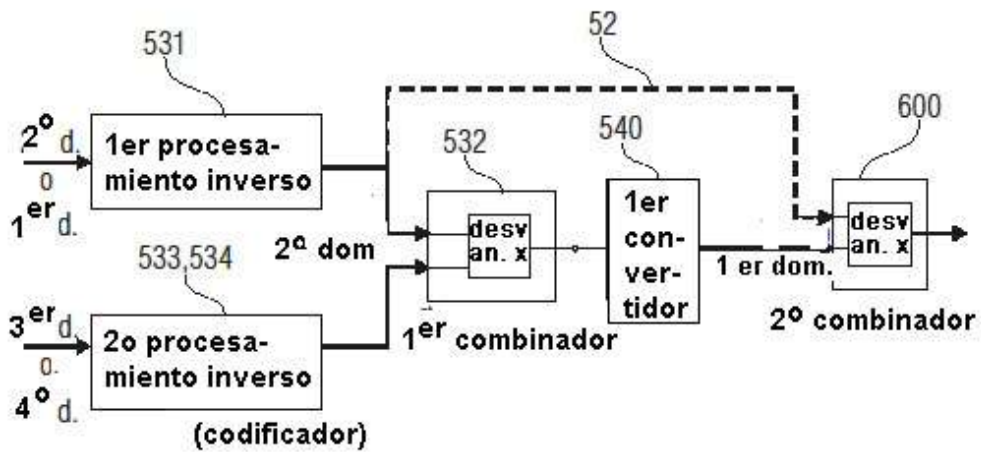


FIG 3D

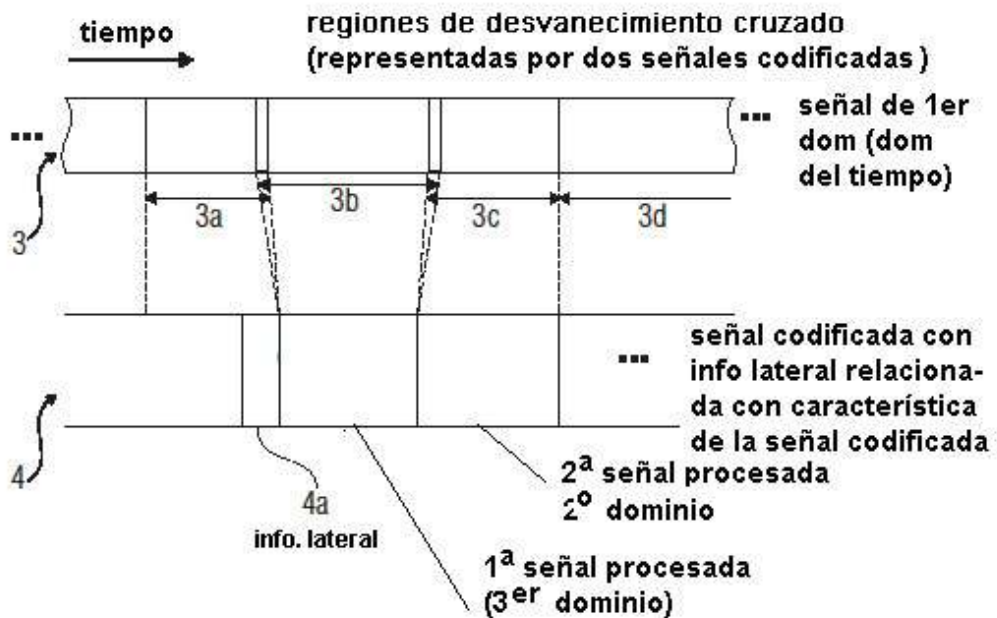


FIG 3E

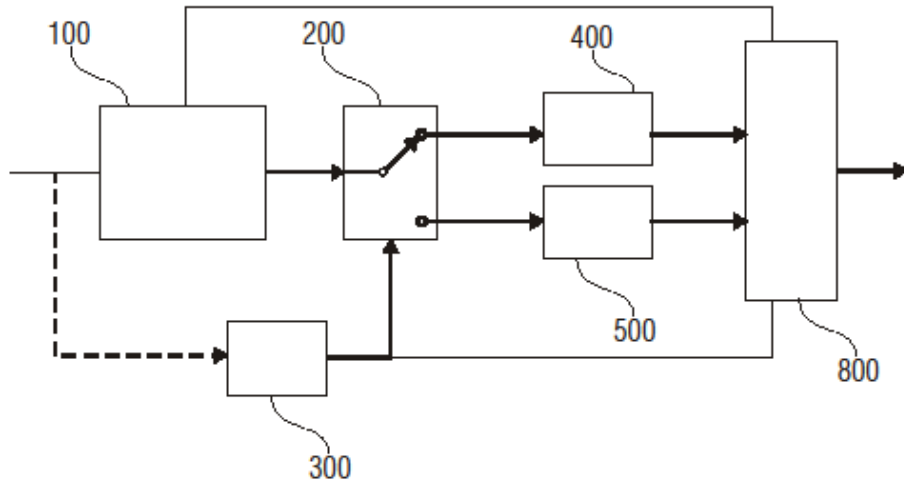


FIG 4A

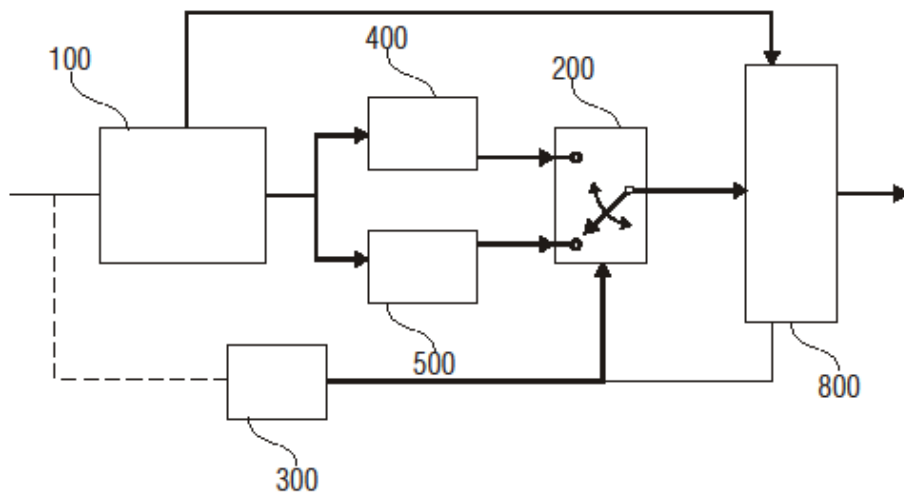


FIG 4B

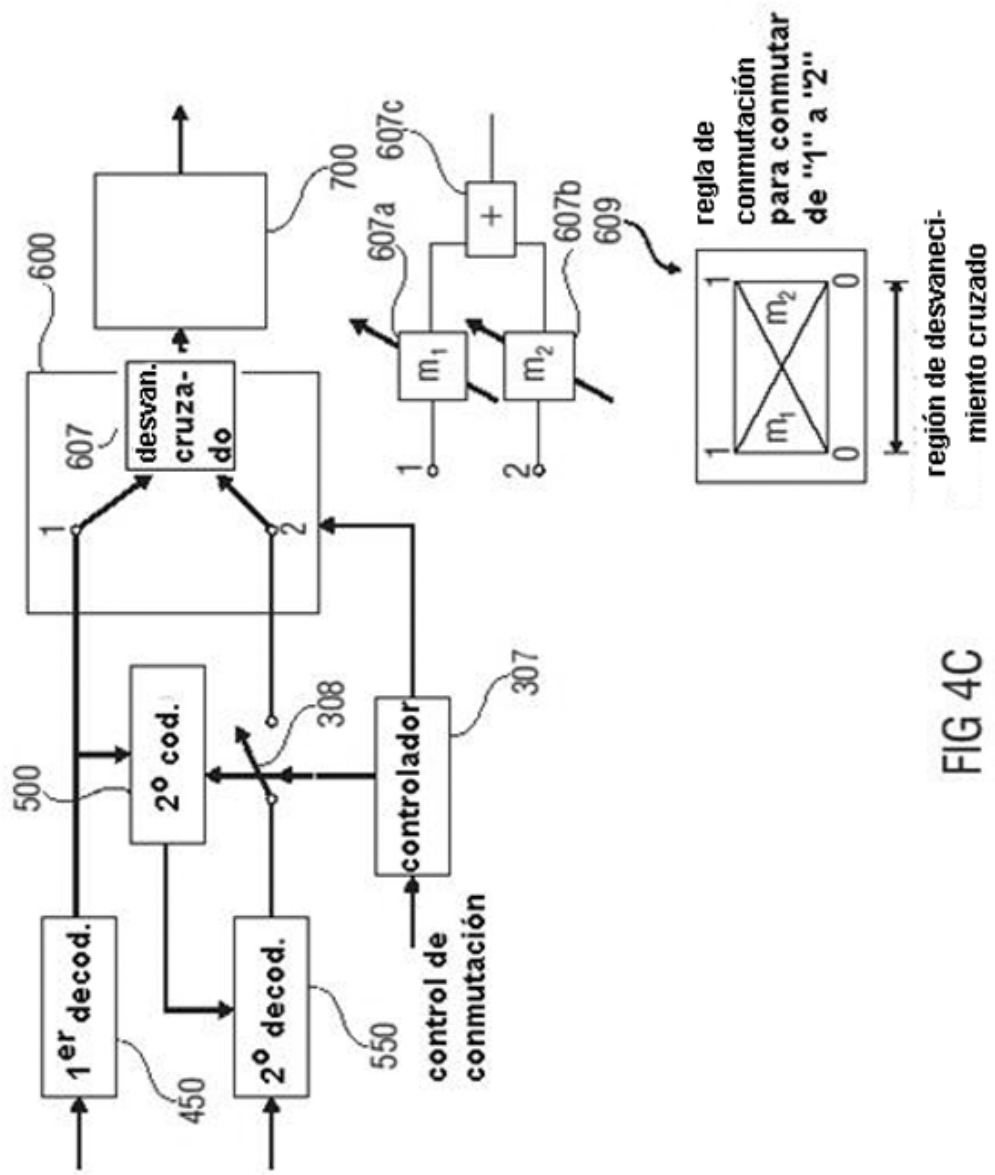


FIG 4C

segmento de señal similar a impulso (p.e., habla vocalizada)

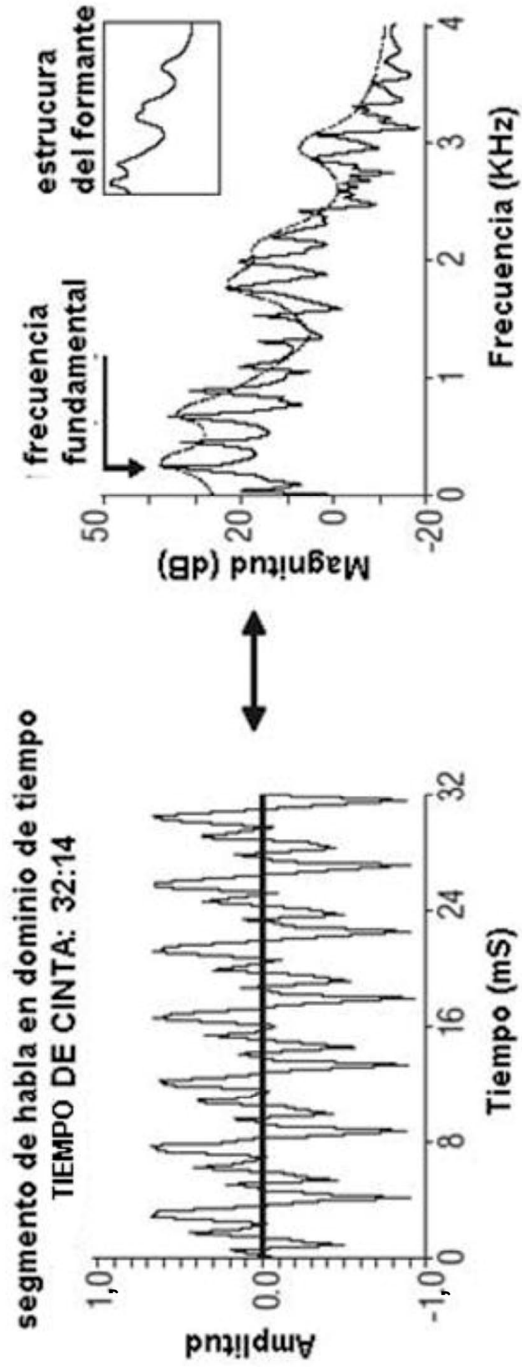


FIG 5A

FIG 5B

Segmento estacionario (por ejemplo, habla no vocalizada)

Segmento de habla en dominio de tiempo
TIEMPO DE CINTA: 30.00

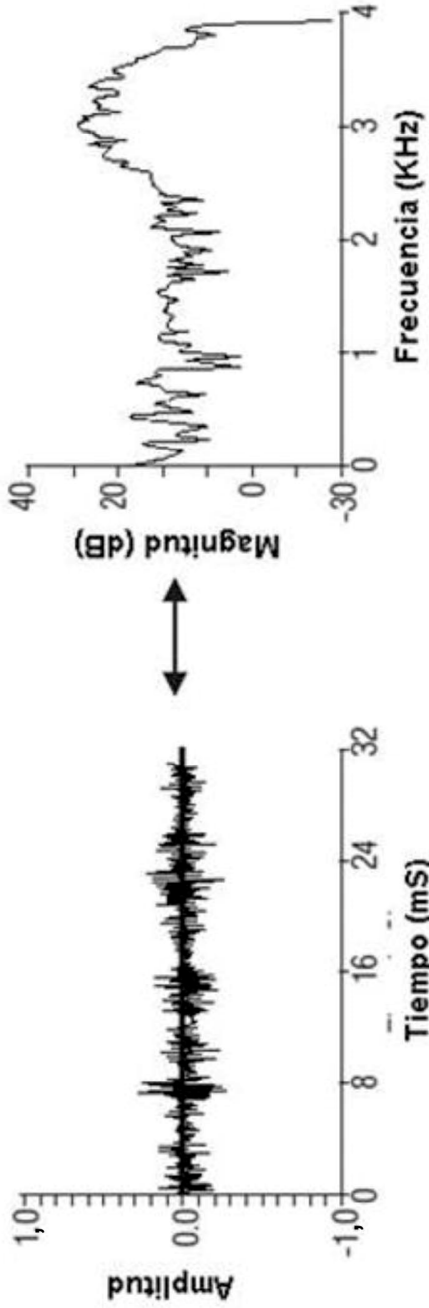
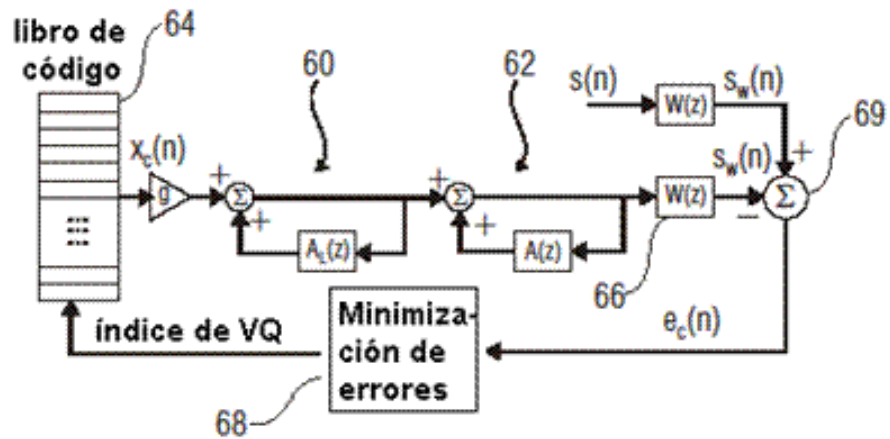


FIG 5D

FIG 5C

CELP de análisis por síntesis



$A_l(z)$: Predicción a largo plazo
 $\hat{=}$ estructura (fina) de paso

$A(z)$: Predicción a corto plazo
 $\hat{=}$ forma y estructura /
 envoltura especial

FIG 6

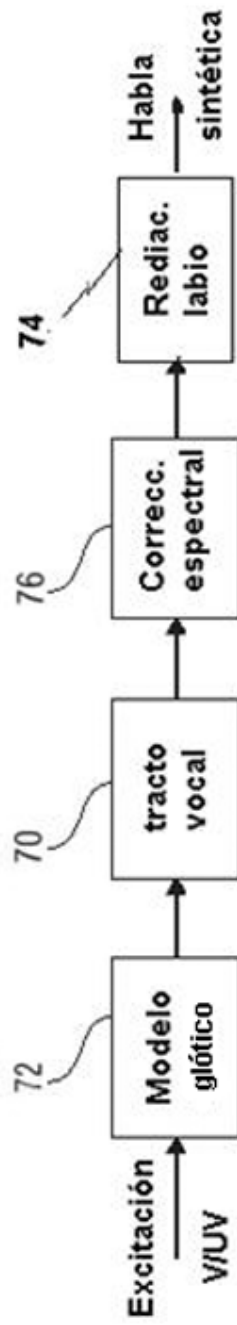


FIG 7A

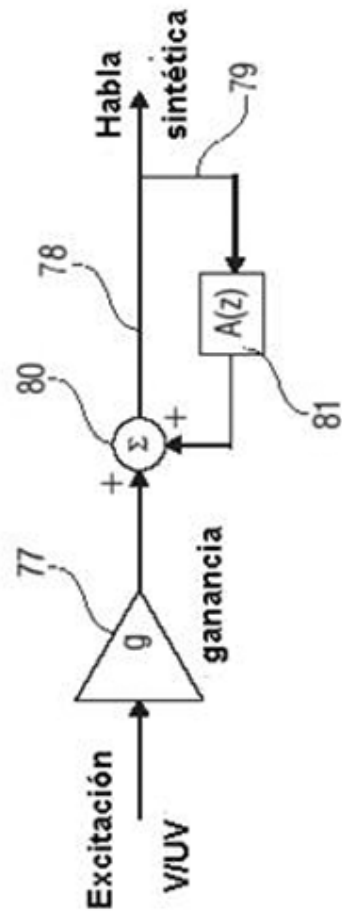


FIG 7B



FIG 7C

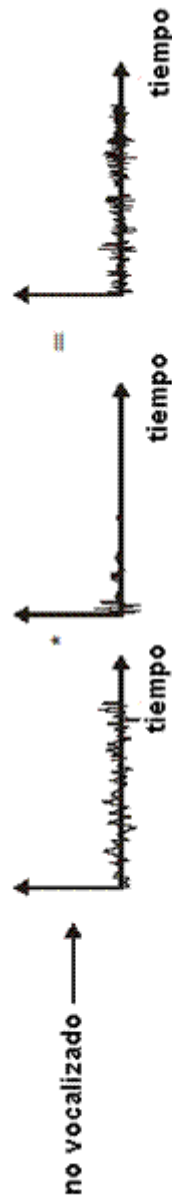


FIG 7D

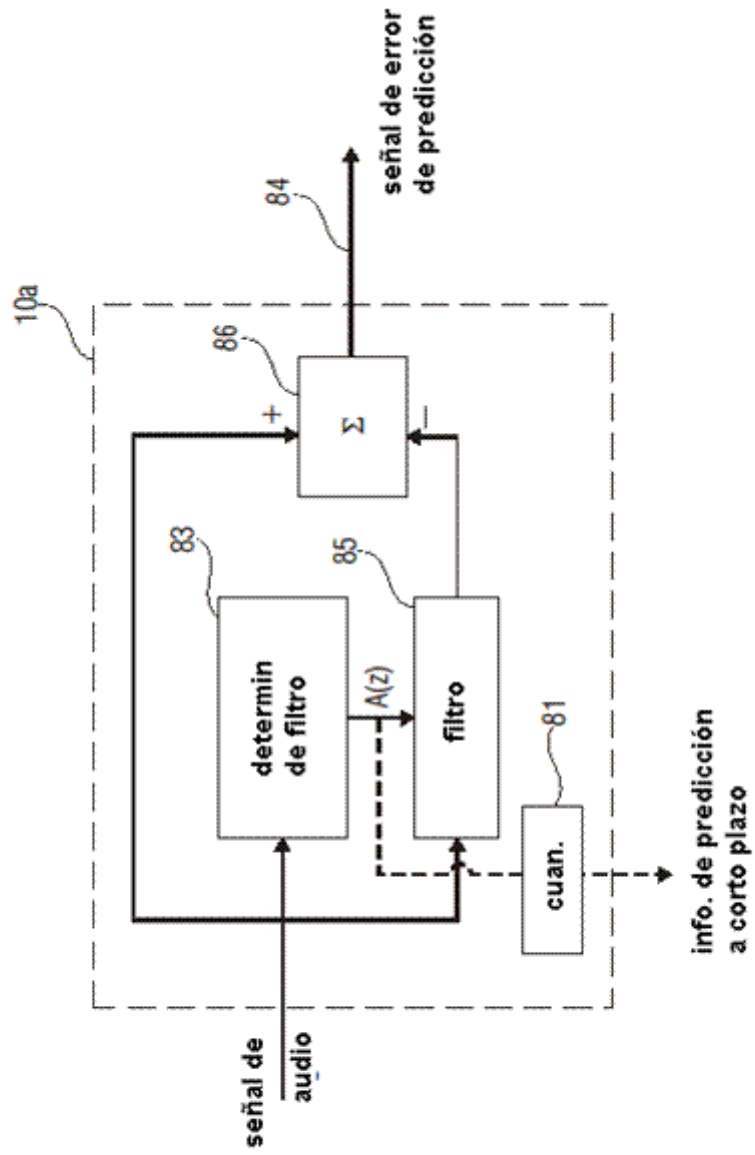


FIG 7E

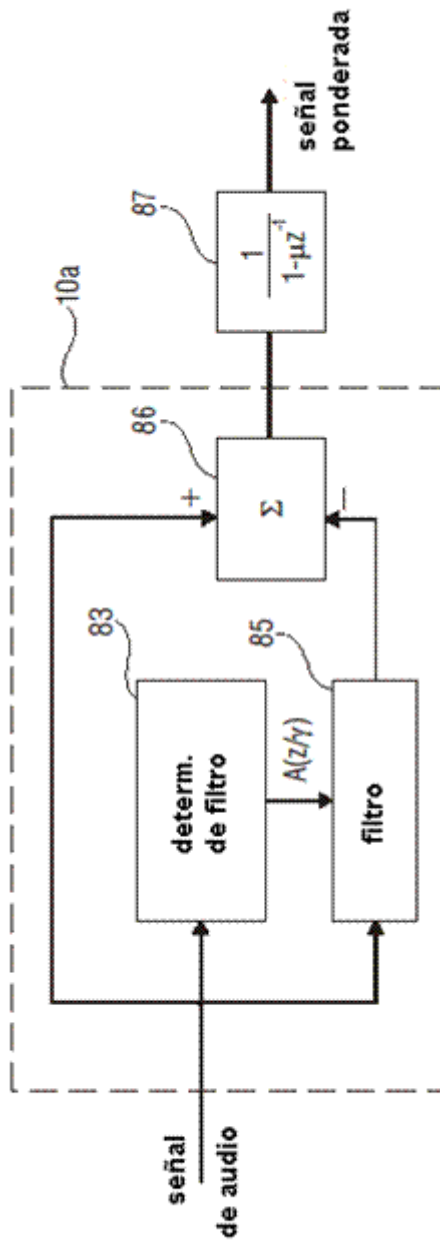


FIG 7F

(lado del codificador)

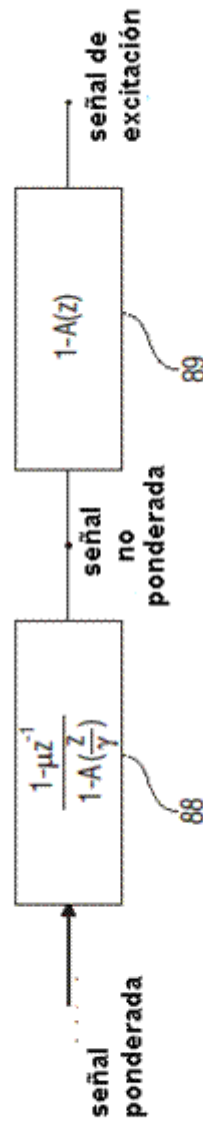


FIG 7G

(lado del decodificador)

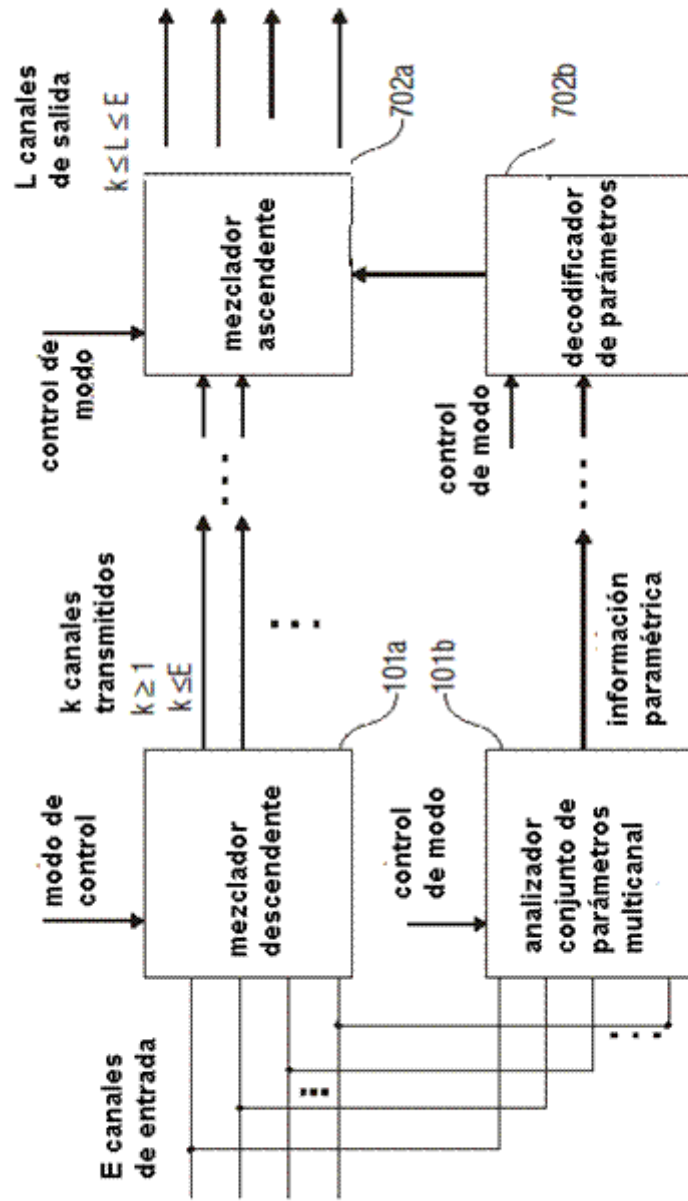


FIG 8

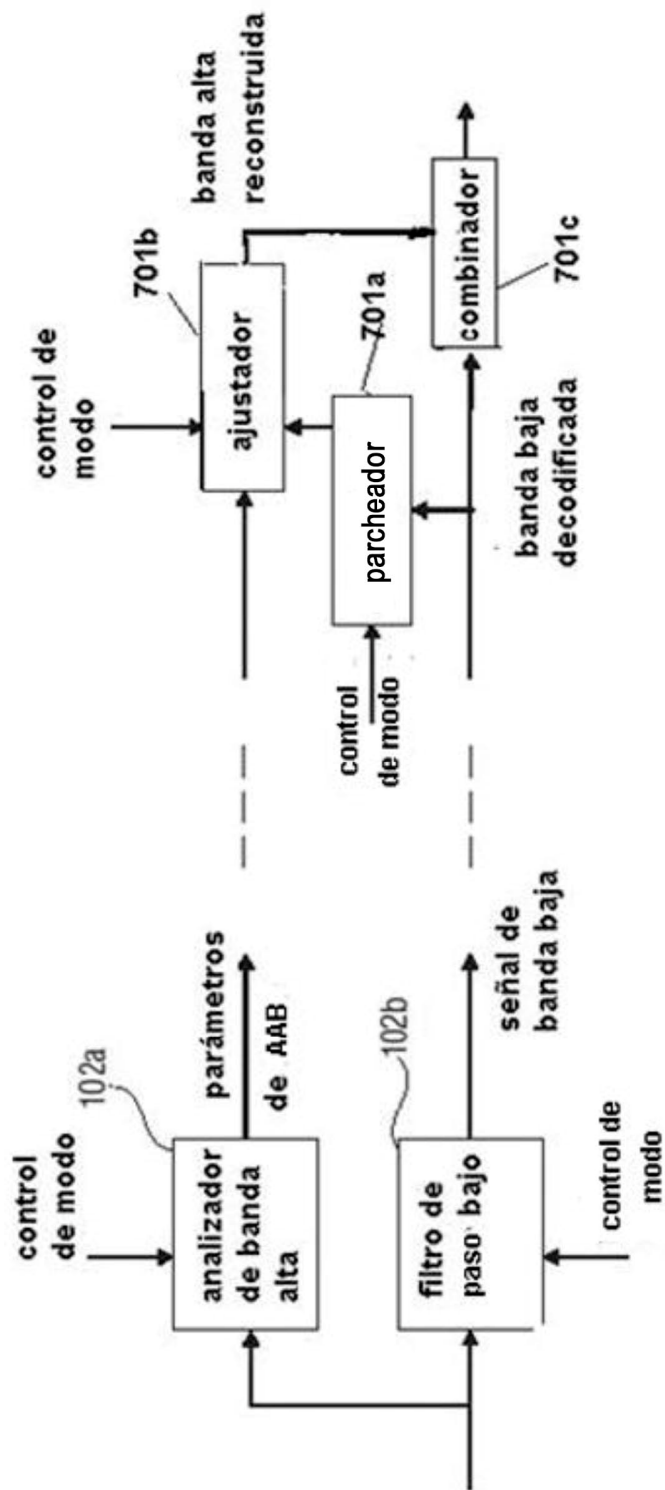


FIG 9



FIG 10A

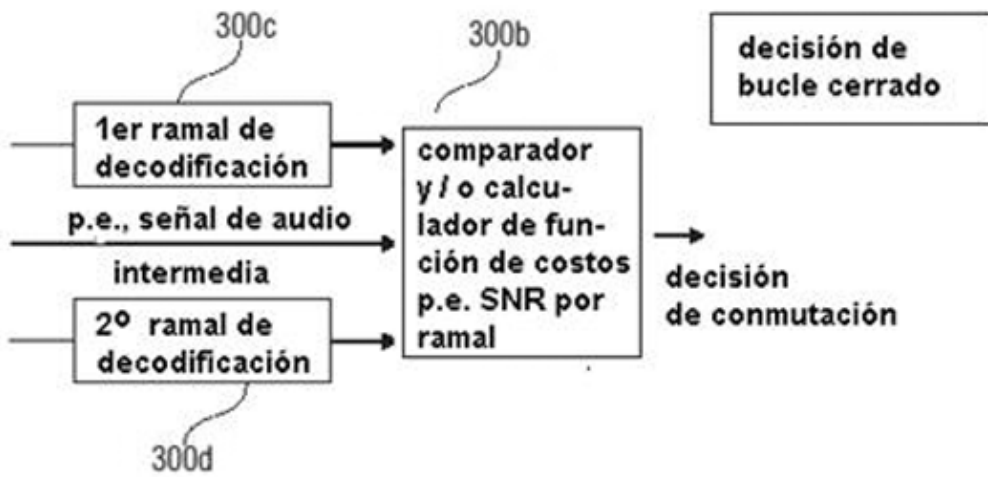


FIG 10B