



19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 380 307**

51 Int. Cl.:
G10L 19/14 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 96 Número de solicitud europea: **09793882 .3**
96 Fecha de presentación : **06.07.2009**
97 Número de publicación de la solicitud: **2311035**
97 Fecha de publicación de la solicitud: **20.04.2011**

54 Título: **Esquema de codificación/decodificación de audio de tasa de transmisión de bits baja con preprocesamiento común.**

30 Prioridad: **11.07.2008 US 79861**
08.10.2008 EP 08017662
18.02.2009 EP 09002272

45 Fecha de publicación de la mención BOPI:
10.05.2012

45 Fecha de la publicación del folleto de la patente:
10.05.2012

73 Titular/es: **Fraunhofer-Gesellschaft zur Förderung der Angewandten Forschung E.V.**
Hansastraße 27C
80686 München, DE

72 Inventor/es: **Grill, Bernhard;**
Bayer, Stefan;
Fuchs, Guillaume;
Geyersberger, Stefan;
Geiger, Ralf;
Hilpert, Johannes;
Kraemer, Ulrich;
Lecomte, Jeremie;
Multrus, Markus;
Neuendorf, Max;
Popp, Harald;
Rettelbach, Nikolaus;
Nagel, Frederik;
Disch, Sascha;
Herre, Juergen;
Yokotani, Yoshikazu;
Wabnik, Stefan;
Schuller, Gerald y
Hirschfeld, Jens

74 Agente/Representante:
Arizti Acha, Mónica

ES 2 380 307 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

DESCRIPCIÓN

Esquema de codificación/decodificación de audio de tasa de transmisión de bits baja con preprocesamiento común.

5 Campo de la invención

La presente invención se refiere a codificación de audio y, particularmente, a esquemas de codificación de audio de tasa de transmisión de bits baja.

10 Antecedentes de la invención y Técnica anterior

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

Por otro lado existen codificadores que son muy adecuados para el procesamiento de habla tal como AMR-WB+ tal como se describe en 3GPP TS 26.290. Tales esquemas de codificación de habla realizan un filtrado de predicción lineal de una señal de dominio de tiempo. Tal filtrado de LP se deriva de un análisis de predicción lineal de la señal de dominio de tiempo de entrada. Los coeficientes de filtro de LP resultantes se codifican y transmiten entonces como información secundaria. El proceso se conoce como codificación de predicción lineal (LPC). En la salida del filtro, la señal residual de predicción o señal de error de predicción que también se conoce como señal de excitación se codifica usando las fases de análisis por síntesis del codificador ACELP o, alternativamente, se codifica usando un codificador de transformada, que usa una transformada de Fourier con una superposición. La decisión entre la codificación ACELP y la codificación de excitación codificada por transformada que también se denomina codificación TCX se realiza usando un algoritmo de lazo abierto o de lazo cerrado.

Esquemas de codificación de audio de dominio de frecuencia tales como el esquema de codificación de AAC de alta eficacia, que combina un esquema de codificación de AAC y una técnica de replicación de ancho de banda espectral también puede combinarse con una herramienta de codificación multicanal o estéreo conjunto que se conoce con el término "MPEG envolvente".

Por otro lado, los codificadores de habla tales como la AMR-WB+ también tienen una fase de mejora de alta frecuencia y una funcionalidad estéreo.

Los esquemas de codificación de dominio de frecuencia son ventajosos porque presentan una alta calidad a tasas de transmisión de bits bajas para señales de música. Sin embargo, es problemática la calidad de señales de habla a tasas de transmisión de bits bajas.

Los esquemas de codificación de habla presentan una alta calidad para señales de habla incluso a tasas de transmisión de bits bajas, pero presentan una mala calidad para señales de música a tasas de transmisión de bits bajas.

45 Sumario de la invención

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

Este objeto se logra mediante el codificador de audio según la reivindicación 1, el método de codificación de audio según la reivindicación 13, el decodificador de audio según la reivindicación 14, el método de decodificación de audio según la reivindicación 24 o el programa informático según la reivindicación 25.

En un aspecto de la presente invención, una fase de decisión que controla un conmutador se usa para alimentar la salida de una fase de preprocesamiento común a cualquiera de las dos ramas. Una está motivada principalmente por un modelo de fuente y/o por mediciones objetivas tales como SNR, la otra por un modelo de sumidero y/o un modelo psicoacústico, es decir, por enmascaramiento auditivo.

A modo de ejemplo, una rama tiene un codificador de dominio de frecuencia y la otra rama tiene un codificador de dominio de LPC tal como un codificador de habla. El modelo de fuente es normalmente el procesamiento de habla y por tanto habitualmente se usa LPC. Por tanto, las fases de preprocesamiento típicas tales como una fase de codificación multicanal o estéreo conjunto y/o una fase de extensión de ancho de banda se usan habitualmente para ambos algoritmos de codificación, lo que ahorra una cantidad considerable almacenamiento, área de chip, consumo de energía, etc. en comparación con la situación, en la que se usa un codificador de audio completo y un codificador de habla completo para el mismo fin.

En una realización preferida, un codificador de audio comprende una fase de preprocesamiento común para dos ramas, en el que una primera rama está motivada principalmente por un modelo de sumidero y/o un modelo psicoacústico, es decir, por enmascaramiento auditivo, y en el que una segunda rama está motivada principalmente por un modelo de fuente y por cálculos de SNR segmentales. El codificador de audio preferiblemente tiene uno o más con-

mutadores para conmutar entre estas ramas en las entradas a estas ramas o salidas de estas ramas de manera controlada por una fase de decisión. En el codificador de audio la primera rama preferiblemente incluye un codificador de audio basado en psicoacústica, y en el que la segunda rama incluye una LPC y un analizador de SNR.

5 En una realización preferida, un decodificador de audio comprende rama de decodificación basada en sumidero de información tal como una rama de decodificación de dominio espectral, una rama de decodificación basada en fuente de información tal como una rama de decodificación de dominio de LPC, un conmutador para conmutar entre las ramas y una fase de posprocesamiento común para posprocesar una señal de audio de dominio de tiempo para obtener una señal de audio posprocesada.

10

Breve descripción de los dibujos

Realizaciones preferidas de la presente invención se describen a continuación con respecto a los dibujos adjuntos, en los que:

15

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

20

la figura 1b es un diagrama de bloques de un esquema de decodificación según el primer 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;

25

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

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

30

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

35

la figura 4a ilustra un diagrama de bloques con un conmutador situado antes de las ramas de codificación;

la figura 4b ilustra un diagrama de bloques de un esquema de codificación con el conmutador situado después de la codificación de las ramas;

40

la figura 4c ilustra un diagrama de bloques para una realización de combinador preferida;

la figura 5a ilustra una forma de onda de un segmento de habla de dominio de tiempo como un segmento de señal a modo de impulso o casi periódico;

45

la figura 5b ilustra un espectro del segmento de la figura 5a;

la figura 5c ilustra un segmento de habla de dominio de tiempo de habla sin voz como un ejemplo para un segmento a modo de ruido o estacionario;

50

la figura 5d ilustra un espectro de la forma de onda de dominio de tiempo de la figura 5c;

la figura 6 ilustra un diagrama de bloques de un análisis mediante codificador de CELP de síntesis;

las figuras 7a a 7d ilustran señales de excitación con voz/sin voz como un ejemplo para señales a modo de ruido/estacionarias y a modo de impulso;

55

la figura 7e ilustra una fase de LPC de lado de codificador que proporciona información de predicción a corto plazo y la señal de error de predicción;

60

la figura 8 ilustra un diagrama de bloques de un algoritmo multicanal conjunto según una realización de la presente invención;

la figura 9 ilustra una realización preferida de un algoritmo de extensión de ancho de banda;

65

la figura 10a ilustra una descripción detallada del conmutador cuando realiza una decisión de lazo abierto; y

la figura 10b ilustra una realización del conmutador cuando opera en un modo de decisión de lazo cerrado.

Descripción detallada o Realizaciones preferidas

Una señal monofónica, una señal estéreo o una señal multicanal se introducen en una fase 100 de preprocesamiento común en la figura 1a. El esquema de preprocesamiento común puede tener una funcionalidad de estéreo conjunto, una funcionalidad envolvente, y/o una funcionalidad de extensión de ancho de banda. En la salida del bloque 100 hay un canal monofónico, un canal estéreo o múltiples canales que se introducen en un conmutador 200 o múltiples conmutadores de tipo 200.

El conmutador 200 puede existir para cada salida de la fase 100, cuando la fase 100 tiene dos o más salidas, es decir, cuando la fase 100 emite una señal estéreo o una señal multicanal. A modo de ejemplo, el primer canal de una señal estéreo podría ser un canal de habla y el segundo canal de la señal estéreo podría ser un canal de música. En esta situación, la decisión en la fase de decisión puede ser diferente entre los dos canales para el mismo instante de tiempo.

El conmutador 200 se controla por una fase 300 de decisión. La fase de decisión recibe, como entrada, una señal introducida en el bloque 100 o una señal emitida por el bloque 100. Alternativamente, la fase 300 de decisión también puede recibir una información secundaria que se incluye en la señal monofónica, la señal estéreo o la señal multicanal o se asocia al menos con una señal de este tipo, en la que existe información, que, por ejemplo, se generó cuando se produjo originalmente la señal monofónica, la señal estéreo o la señal multicanal.

En una realización, la fase de decisión no controla la fase 100 de preprocesamiento, y la flecha entre el bloque 300 y el 100 no existe. En una realización adicional, el procesamiento en el bloque 100 se controla en cierto grado por la fase 300 de decisión con el fin 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 funcionalidad principal en el bloque 100 está activa independientemente de la decisión en la fase 300.

La fase 300 de decisión acciona el conmutador 200 con el fin de alimentar la salida de la fase de preprocesamiento común o bien en una parte 400 de codificación de frecuencia ilustrada en una rama superior de la figura 1a o bien una parte 500 de codificación de dominio de LPC ilustrada en una rama interior en la figura 1a.

En una realización, el conmutador 200 conmuta entre las dos ramas 400, 500 de codificación. En una realización adicional, puede haber ramas de codificación adicionales tales como una tercera rama de codificación o incluso una cuarta rama de codificación o incluso más ramas de codificación. En una realización con tres ramas de codificación, la tercera rama de codificación podría ser similar a la segunda rama de codificación, pero podría incluir un codificador de excitación diferente del codificador 520 de excitación en la segunda rama 500. En esta realización, la segunda rama comprende la fase 510 de LPC y un codificador de excitación basado en libro de códigos tal como en ACELP, y la tercera rama comprende una fase de LPC y un codificador de excitación que opera en una representación espectral de la señal de salida de fase de LPC.

Un elemento clave de la rama de codificación de dominio de frecuencia es un bloque 410 de conversión espectral que es operativo para convertir la señal de salida de fase de preprocesamiento común a un dominio espectral. El bloque de conversión espectral puede incluir un algoritmo de MDCT, un QMF, un algoritmo de FFT, análisis de ondículas o un banco de filtros tal como un banco de filtros muestreado de manera crítica que tiene un determinado número de canales de banco de filtros, en el que las señales de subbanda en este banco de filtros pueden ser señales de valor real o señales de valor complejo. La salida del bloque 410 de conversión espectral se codifica usando un codificador 420 de audio espectral, que puede incluir bloques de procesamiento tal como se conoce a partir del esquema de codificación de AAC.

En la rama 500 de codificación inferior, un elemento clave es un analizador de modelo de fuente tal como LPC 510, que emite dos clases de señales. Una señal es una señal de información de LPC que se usa 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 fase 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 de CELP, un codificador de ACELP o cualquier otro codificador que procesa una señal de dominio de LPC.

Otra implementación de codificador de excitación preferida es una codificación de transformada de la señal de excitación. En esta realización, la señal de excitación no se codifica usando un mecanismo de libro de códigos de ACELP, sino que la señal de excitación se convierte en una representación espectral y los valores de representación espectral tales como señales de subbanda en caso de un banco de filtros o coeficientes de frecuencia en caso de una transformada tal como una FFT se codifican para obtener una compresión de datos. Una implementación de esta clase de codificador de excitación es el modo de codificación de TCX conocido a partir de AMR-WB+.

La decisión en la fase de decisión puede ser adaptativa a la señal de modo que la fase de decisión realiza una discriminación de habla/música y controla el conmutador 200 de tal manera que las señales de música se introducen en la rama 400 superior, y las señales de habla se introducen en la rama 500 inferior. En una realización, la fase de decisión alimenta su información de decisión a un flujo de bits de salida, de modo que un decodificador pueda usar esta información de decisión con el fin de realizar las operaciones de decodificación correctas.

ES 2 380 307 T3

Un decodificador de este tipo se ilustra en la figura 1b. 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 convertidor 440 de dominio de tiempo. De manera análoga, la salida del codificador 520 de excitación de la figura 1a se introduce en un decodificador 530 de excitación que emite una señal de dominio de LPC. La señal de dominio de LPC se introduce en una fase 540 de síntesis de LPC, que recibe, como una entrada adicional, la información de LPC generada por la fase 510 de análisis de LPC correspondiente. La salida del convertidor 440 de dominio de tiempo y/o la salida de la fase 540 de síntesis de LPC se introducen en un conmutador 600. El conmutador 600 se controla a través de una señal de control de conmutador que se generó, por ejemplo, por la fase 300 de decisión, o que se proporcionó externamente tal como por un creador de la señal monofónica, señal estéreo o señal multicanal original.

La salida del conmutador 600 es una señal monofónica completa que se introduce, a continuación, en una fase 700 de posprocesamiento común, que puede realizar un procesamiento de estéreo conjunto o un procesamiento de extensión de ancho de banda, etc. Alternativamente, la salida del conmutador también podría ser una señal estéreo o incluso una señal multicanal. Es una señal estéreo, cuando el preprocesamiento incluye una reducción de canal a dos canales. Incluso puede ser una señal multicanal, cuando se realiza una reducción de canal a tres canales o no se realiza ninguna reducción de canal sino sólo una replicación de banda espectral.

Dependiendo de la funcionalidad específica de la fase de posprocesamiento común, una señal monofónica, se emite una señal estéreo o una señal multicanal que tiene, cuando la fase 700 de posprocesamiento común realiza una operación de extensión de ancho de banda, un ancho de banda más grande que la señal introducida en el bloque 700.

En una realización, el conmutador 600 conmuta entre las dos ramas 430, 440 y 530, 540 de decodificación. En una realización adicional, puede haber ramas de decodificación adicionales tales como una tercera rama de decodificación o incluso una cuarta rama de decodificación o incluso más ramas de decodificación. En una realización con tres ramas de decodificación, la tercera rama de decodificación podría ser similar a la segunda rama de decodificación, pero podría incluir un decodificador de excitación diferente del decodificador 530 de excitación en la segunda rama 530, 540. En esta realización, la segunda rama comprende la fase 540 de LPC y un decodificador de excitación basado en libro de códigos tal como en ACELP, y la tercera rama comprende una fase de LPC y un decodificador de excitación que opera en una representación espectral de la señal de salida de la fase 540 de LPC.

Tal como se indicó anteriormente, la figura 2a ilustra un esquema de codificación preferido según un segundo aspecto de la invención. El esquema de preprocesamiento común en 100 de la figura 1a comprende ahora un bloque 101 estéreo conjunto/envolvente que genera, como una salida, parámetros de estéreo conjunto y una señal de salida monofónica, que se genera mezclando de manera descendente la señal de entrada que es una señal que tiene dos o más canales. Generalmente, la señal en la salida del bloque 101 también puede ser una señal que tiene más canales, pero debido a la funcionalidad de 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 del bloque 101 se introduce en un bloque 102 de extensión de ancho de banda que, en el codificador de la figura 2a, emite una señal de banda limitada tal como la señal de banda baja o la señal de paso bajo en su salida. Además, para la banda alta de la señal introducida en el bloque 102, se generan y se envían a un multiplexor 800 de flujo de bits, parámetros de extensión de ancho de banda tales como parámetros de envolvente espectral, parámetros de filtrado inverso, parámetros de umbral mínimo de ruido, etc. tal como se conoce a partir del perfil de HE-AAC de MPEG-4.

Preferiblemente, la fase 300 de decisión recibe la señal introducida en el bloque 101 o introducida en el bloque 102 con el fin de decidir entre, por ejemplo, un modo de música o un modo de habla. En el modo de música, se selecciona la rama 400 de codificación superior, mientras que, en el modo de habla, se selecciona la rama 500 de codificación inferior. Preferiblemente, la fase de decisión controla adicionalmente el bloque 101 de estéreo conjunto y/o el bloque 102 de extensión de ancho de banda para adaptar la funcionalidad de estos bloques a la señal específica. Por tanto, cuando la fase 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 la fase 300 de decisión puede controlar características específicas del bloque 101 y/o del bloque 102. Alternativamente, cuando la fase 300 de decisión determina que la señal está en un modo de habla o, generalmente, en un modo de codificación de dominio de LPC, entonces pueden controlarse características específicas de los bloques 101 y 102 según la salida de la fase de decisión.

Dependiendo de la decisión del conmutador, que puede derivarse de la señal de entrada del conmutador 200 o de cualquier fuente externa tal como un productor de la señal de audio original que subyace a la señal introducida en la fase 200, el conmutador conmuta entre la rama 400 de codificación de frecuencia y la rama 500 de codificación de LPC.

La rama 400 de codificación de frecuencia comprende una fase 410 de conversión espectral y una fase 421 de cuantificación/codificación posteriormente conectada (tal como se muestra en la figura 2a). La fase de cuantificación/codificación puede incluir cualquiera de las funcionalidades tal como se conocen de los codificadores de dominio de frecuencia modernos tal como el codificador de AAC. Además, la operación de cuantificación en la fase 421 de cuantificación/codificación puede controlarse a través de un módulo psicoacústico que genera información psicoacústica tal como un umbral de enmascaramiento psicoacústico sobre la frecuencia, en el que esta información se introduce en la fase 421.

ES 2 380 307 T3

Preferiblemente, la conversión espectral se realiza usando una operación de MDCT que, incluso más preferiblemente, es la operación de MDCT alineada en tiempo, en la que puede controlarse la intensidad o, generalmente, la intensidad de alineamiento entre cero y una intensidad de alineamiento alta. En una intensidad de alineamiento cero, la operación de MDCT en el bloque 411 es una operación de MDCT sencilla conocida en la técnica. La intensidad de
5 alineamiento de tiempo junto con la información secundaria de alineamiento de tiempo puede transmitirse/introducirse al multiplexor 800 de flujo de bits como información secundaria. Por tanto, si se usa TW-MDCT, la información secundaria de alineamiento de tiempo debe enviarse al flujo de bits tal como se ilustra por 424 en la figura 2a, y, en el lado de decodificador, la información secundaria de alineamiento de tiempo debe recibirse del flujo de bits tal como se ilustra por el elemento 434 en la figura 2b.

10 En la rama de codificación de LPC, el codificador de dominio de LPC puede incluir un núcleo de ACELP que calcula una ganancia de altura tonal, un retraso de altura tonal y/o información de libro de códigos tal como un índice de libro de códigos y una ganancia de códigos.

15 En la primera rama 400 de codificación, un convertidor espectral preferiblemente comprende una operación de MDCT específicamente adaptada que tiene determinadas funciones ventana seguida por una fase de codificación de cuantificación/entropía que puede ser una fase de cuantificación de vector, pero preferiblemente es un cuantificador/codificador tal como se indica para el cuantificador/codificador en la rama de codificación de dominio de frecuencia, es decir, en el elemento 421 de la figura 2a.

20 La figura 2b ilustra un esquema de decodificación que corresponde al esquema de codificación de la figura 2a. El flujo de bits generado por el multiplexor 800 de flujo de bits de la figura 2a se introduce en un demultiplexor 900 de flujo de bits. Dependiendo de una información derivada, por ejemplo, del flujo de bits a través de un bloque 601 de detección de modo, se controla un conmutador 600 de lado de decodificador para enviar o bien señales desde la rama superior o bien señales desde la rama inferior al bloque 701 de extensión de ancho de banda. El bloque 701 de extensión de ancho de banda recibe, del demultiplexor 900 de flujo de bits, información secundaria y, basándose en esta información secundaria y en la salida de la detección 601 de modo, reconstruye la banda alta basándose en la salida de banda baja por el conmutador 600.

30 La señal de banda completa generada por el bloque 701 se introduce en la fase 702 de procesamiento de estéreo conjunto/envolvente, que reconstruye dos canales estéreo o varios multicanales. Generalmente, el bloque 702 emitirá más canales de los que se introdujeron en este bloque. Dependiendo de la aplicación, la entrada al bloque 702 puede incluir dos canales tal como en un modo estéreo e incluso puede incluir más canales siempre que la salida de este bloque tenga más canales que la entrada a este bloque.

35 Generalmente, existe un decodificador 530 de excitación. El algoritmo implementado en el bloque 530 se adapta al algoritmo correspondiente usado en el bloque 520 en el lado de codificador. Mientras que la fase 431 emite un espectro derivado de una señal de dominio de tiempo que se convierte al dominio de tiempo usando el convertidor 440 de frecuencia/tiempo, la fase 530 emite una señal de dominio de LPC. Los datos de salida de la fase 530 se transforman de vuelta al dominio de tiempo usando una fase 540 de síntesis de LPC, que se controla a través de una información de LPC transmitida y generada de lado de codificador. Entonces, después del bloque 540, ambas ramas tienen información de dominio de tiempo que se conmuta según una señal de control de conmutador con el fin de obtener finalmente una señal de audio tal como una señal monofónica, una señal estéreo o una señal multicanal.

45 Se ha mostrado que el conmutador 200 conmuta entre ambas ramas de modo que sólo una rama recibe una señal para procesar y la otra rama no recibe ninguna señal para procesar. En una realización alternativa, sin embargo, el conmutador también puede disponerse después, por ejemplo, del codificador 420 de audio y el codificador 520 de excitación, lo que significa que ambas ramas 400, 500 procesan la misma señal en paralelo. Con el fin de no duplicar la tasa de transmisión de bits, sin embargo, se selecciona sólo la señal emitida por una de esas ramas 400 ó 500 de codificación para escribirse en el flujo de bits de salida. La fase de decisión entonces operará de modo que la señal escrita en el flujo de bits minimice una determinada función de coste, en la que la función de coste puede ser la tasa de transmisión de bits generada o la distorsión de percepción generada o una función de coste de tasa/distorsión combinada. Por tanto, o bien en este modo o bien en el modo ilustrado en las figuras, la fase de decisión también puede operar en un modo de lazo cerrado con el fin de garantizar que, finalmente, se escribe sólo la salida de rama de
55 codificación en el flujo de bits que tiene para una distorsión de percepción dada la tasa de transmisión de bits más baja o, para una tasa de transmisión de bits dada, tiene la distorsión de percepción más baja.

60 Generalmente, el procesamiento en la rama 400 es un procesamiento en un modelo basado en percepción o modelo de sumidero de información. Por tanto, esta rama modela el sistema auditivo humano que recibe sonido. En cambio, el procesamiento en la rama 500 va a generar una señal en el dominio de excitación, residual o de LPC. Generalmente, el procesamiento en la rama 500 es un procesamiento en un modelo de habla o un modelo de generación de información. Para señales de habla, este modelo es un modelo del sistema de generación de habla/sonido humano que genera sonido. Sin embargo, si va a codificarse un sonido de una fuente diferente que requiere un modelo de generación de sonido diferente, entonces el procesamiento en la rama 500 puede ser diferente.

65 Aunque las figuras 1a a 2b se ilustran como diagramas de bloques de un aparato, estas figuras simultáneamente son una ilustración de un método, en las que las funcionalidades de bloque corresponden a las etapas del método.

ES 2 380 307 T3

La figura 3a ilustra un codificador de audio para generar una señal de audio codificada en una salida de la primera rama 400 de codificación y una segunda rama 500 de codificación. Además, la señal de audio codificada incluye preferiblemente información secundaria tal como parámetros de preprocesamiento a partir de la fase de preprocesamiento común o, tal como se comentó en relación con las figuras anteriores, información de control de conmutador.

Preferiblemente, la primera rama de codificación es operativa con el fin de codificar una señal 195 intermedia de audio según un primer algoritmo de codificación, en la que el primer algoritmo de codificación tiene un modelo de sumidero de información. La primera rama 400 de codificación genera la primera señal de salida de codificador que es una representación información espectral codificada de la señal 195 intermedia de audio.

Además, la segunda rama 500 de codificación está adaptada para codificar la señal 195 intermedia de audio según un segundo algoritmo de codificación, teniendo el segundo algoritmo de codificación un modelo de fuente de información y generando, en una primera señal de salida de codificador, parámetros codificados para el modelo de fuente de información que representan la señal de audio intermedia.

El codificador de audio comprende además la fase de preprocesamiento común para preprocesar una señal 99 de entrada de audio para obtener la señal 195 intermedia de audio. Específicamente, la fase de preprocesamiento común es operativa para procesar la señal 99 de entrada de audio de modo que la señal 195 intermedia de audio, es decir, la salida del algoritmo de preprocesamiento común es una versión comprimida de la señal de entrada de audio.

Un método preferido de codificación de audio para generar una señal de audio codificada, comprende una etapa de codificar 400 una señal 195 intermedia de audio según un primer algoritmo de codificación, teniendo el primer algoritmo de codificación un modelo de sumidero de información y generando, en una primera señal de salida, información espectral codificada que representa la señal de audio; una etapa de codificar 500 una señal 195 intermedia de audio según un segundo algoritmo de codificación, teniendo el segundo algoritmo de codificación un modelo de fuente de información y generando, en una segunda señal de salida, parámetros codificados para el modelo de fuente de información que representan la señal 195 intermedia, y una etapa de preprocesar de manera común una señal 99 de entrada de audio para obtener la señal 195 intermedia de audio, en la que, en la etapa de preprocesar de manera común se procesa la señal 99 de entrada de audio de modo que la señal 195 intermedia de audio es una versión comprimida de la señal 99 de entrada de audio, en el que la señal de audio codificada incluye, para una determinada parte de la señal de audio o bien la primera señal de salida o bien la segunda señal de salida. El método incluye preferiblemente la etapa adicional que codifica una determinada parte de la señal intermedia de audio o bien usando el primer algoritmo de codificación o bien usando el segundo algoritmo de codificación o bien codificando la señal usando ambos algoritmos y emitiendo en una señal codificada o bien el resultado del primer algoritmo de codificación o bien el resultado del segundo algoritmo de codificación.

Generalmente, el algoritmo de codificación de audio usado en la primera rama 400 de codificación refleja y modela la situación en un sumidero de audio. El sumidero de una información de audio es normalmente el oído humano. El oído humano puede modelarse como un analizador de frecuencia. Por tanto, la primera rama de codificación emite información espectral codificada. Preferiblemente, la primera rama de codificación incluye además un modelo psicoacústico para aplicar adicionalmente un umbral de enmascaramiento psicoacústico. Este umbral de enmascaramiento psicoacústico se usa cuando se cuantifican valores espectrales de audio en los que, preferiblemente, se realiza la cuantificación de manera que se introduce un ruido de cuantificación al cuantificar los valores de audio espectrales, que se ocultan debajo del umbral de enmascaramiento psicoacústico.

La segunda rama de codificación representa un modelo de fuente de información, que refleja la generación de sonido de audio. Por tanto, los modelos de fuente de información pueden incluir un modelo de habla que se refleja mediante una fase de LPC, es decir, transformando una señal de dominio de tiempo en una de dominio de LPC y procesando a continuación la señal residual de LPC, es decir, la señal de excitación. Sin embargo, modelos de fuente de sonido alternativos, son modelos de fuente de sonido para representar un determinado instrumento o cualquier otro generador de sonido tal como una fuente de sonido específica que exista en el mundo real. Puede realizarse una selección entre diferentes modelos de fuente de sonido cuando varios modelos de fuente de sonido están disponibles, basándose en un cálculo de SNR, es decir, basándose en un cálculo, cuál de los modelos de fuente es el más adecuado para codificar una parte de tiempo y/o una parte de frecuencia determinada de una señal de audio. Preferiblemente, sin embargo, la conmutación entre ramas de codificación se realiza en el dominio de tiempo, es decir, que se codifica una determinada parte de tiempo usando un modelo y se codifica una determinada parte de tiempo diferente de la señal intermedia usando la otra rama de codificación.

Los modelos de fuente de información se representan mediante determinados parámetros. Con respecto al modelo de habla, los parámetros son parámetros de LPC y parámetros de excitación codificados, cuando se considera un codificador de habla moderno tal como AMR-WB+. El AMR-WB+ comprende un codificador de ACELP y un codificador de TCX. En este caso, los parámetros de excitación codificados pueden ser códigos de ganancia global, de umbral mínimo de ruido y de longitud variable.

Generalmente, todos los modelos de fuente de información permitirán el establecimiento de un conjunto de parámetros que refleja la señal de audio original de manera muy eficaz. Por tanto, la salida de la segunda rama de codificación serán parámetros codificados para el modelo de fuente de información que representan la señal intermedia de audio.

ES 2 380 307 T3

La figura 3b ilustra un decodificador que corresponde al codificador ilustrado en la figura 3a. En general, la figura 3b ilustra un decodificador de audio para decodificar una señal de audio codificada para obtener una señal 799 de audio decodificada. El decodificador incluye la primera rama 450 de decodificación para decodificar una señal codificada, codificada según un primer algoritmo de codificación que tiene un modelo de sumidero de información. El decodificador de audio incluye además una segunda rama 550 de decodificación para decodificar una señal de información codificada, codificada según un segundo algoritmo de codificación que tiene un modelo de fuente de información. El decodificador de audio incluye además un combinador para combinar señales de salida de la primera rama 450 de decodificación y la segunda rama 550 de decodificación para obtener una señal combinada. La señal combinada que se ilustra en la figura 3b como la señal 699 intermedia de audio decodificada se introduce en una fase de posprocesamiento común para posprocesar la señal 699 intermedia de audio decodificada, que es la señal combinada emitida por el combinador 600 de modo que una señal de salida de la fase de preprocesamiento común sea una versión expandida de la señal combinada. Por tanto, la señal 799 de audio decodificada tiene un contenido de información mejorado en comparación con la señal 699 intermedia de audio decodificada. Esta expansión de información se proporciona por la fase de posprocesamiento común con la ayuda de parámetros de pre/posprocesamiento que pueden transmitirse de un codificador a un decodificador, o que pueden derivarse de la propia señal intermedia de audio decodificada. Preferiblemente, sin embargo, los parámetros de pre/posprocesamiento se transmiten de un codificador a un decodificador, puesto que este procedimiento permite una calidad mejorada de la señal de audio decodificada.

La figura 4a y 4b ilustra dos realizaciones diferentes, que difieren en el posicionamiento del conmutador 200. En la figura 4a, el conmutador 200 se posiciona entre una salida de la fase 100 de preprocesamiento común y una entrada de las dos ramas 400, 500 codificadas. La realización de la figura 4a garantiza que la señal de audio se introduce sólo en una única rama de codificación, y la otra rama de codificación, que no está conectada a la salida de la fase de preprocesamiento común no opera y, por tanto, está apagada o permanece en un modo de suspensión. Esta realización es preferible porque la rama de codificación no activa no consume energía ni recursos computacionales que son útiles para aplicaciones móviles en particular, que funcionan con batería y, por tanto, tienen la limitación general de consumo de energía.

Por otro lado, sin embargo, puede ser preferible la realización de la figura 4b cuando el consumo de energía no es un problema. En esta realización, ambas ramas 400, 500 de codificación están activas todo el tiempo, y sólo se envía la salida de la rama de codificación seleccionada durante una determinada parte de tiempo y/o para una determinada parte de frecuencia al formateador de flujo de bits que puede implementarse como multiplexor 800 de flujo de bits. Por tanto, en la realización de la figura 4b, ambas ramas de codificación están activas todo el tiempo, y la salida de una rama de codificación que se selecciona por la fase 300 de decisión se introduce en el flujo de bits de salida, mientras que se descarta la salida de la otra rama 400 de codificación no seleccionada, es decir, no se introduce en el flujo de bits de salida, es decir, la señal de audio codificada.

La figura 4c ilustra un aspecto adicional de una implementación de decodificador preferida. Con el fin de evitar artefactos audibles específicamente en la situación, en la que el primer decodificador es un decodificador de generación de solapamiento de tiempo o, en general, un decodificador de dominio de frecuencia y el segundo decodificador es un dispositivo de dominio de tiempo, los límites entre los bloques o tramas emitidas por el primer decodificador 450 y el segundo decodificador 550 no deben ser completamente continuos, específicamente en una situación de conmutación. Por tanto, cuando se emite el primer bloque del primer decodificador 450 y, cuando durante una parte de tiempo posterior, se emite un bloque del segundo decodificador, se prefiere realizar una operación de atenuación cruzada tal como se ilustra por un bloque 607 de atenuación cruzada. Para ello, el bloque 607 de atenuación cruzada podría implementarse tal como se ilustra en la figura 4c en 607a, 607b y 607c. Cada rama podría tener un ponderador que tiene un factor de ponderación m_1 entre 0 y 1 en la escala normalizada, en la que el factor de ponderación puede variar tal como se indica en el esquema 609, una regla de atenuación cruzada de este tipo garantiza que tiene lugar una atenuación cruzada continua y suave lo que garantiza adicionalmente que un usuario no percibirá ninguna variación de volumen.

En determinados casos, se generó el último bloque del primer decodificador usando una ventana en la que la ventana realizó de hecho una atenuación de salida de este bloque. En este caso, el factor de ponderación m_1 en el bloque 607a es igual a 1 y, de hecho, no se requiere ponderación alguna para esta rama.

Cuando tiene lugar una conmutación del segundo decodificador al primer decodificador, y cuando el segundo decodificador incluye una ventana que de hecho atenúa la salida al final del bloque, entonces no se requerirá el ponderador indicado con " m_2 " o el parámetro de ponderación puede establecerse a 1 a través de toda la región de atenuación cruzada.

Cuando el primer bloque después de generar una conmutación usando una operación de aplicación de ventana, y cuando esta ventana de hecho realizó una operación de atenuación de entrada, entonces el factor de ponderación correspondiente puede establecerse a 1 de modo que no es realmente necesario un ponderador. Por tanto, cuando se aplica una función ventana al último bloque con el fin de que el decodificador lo atenúe en salida y cuando se aplica una función ventana al primer bloque después del conmutador usando el decodificador con el fin de proporcionar una atenuación de entrada, entonces no se requieren los ponderadores 607a, 607b en absoluto y es suficiente una operación de adición por el sumador 607c.

ES 2 380 307 T3

En este caso, la parte de atenuación de salida de la última trama y la parte de atenuación de entrada de la trama siguiente definen la región de atenuación cruzada indicada en el bloque 609. Además, 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.

5

Si no se requiere o no es posible o no se desea una operación de atenuación cruzada, y si sólo se produce una conmutación brusca de un decodificador al otro decodificador, se prefiere realizar una conmutación de este tipo en pasos silenciosos de la señal de audio o al menos en pasos de la señal de audio en los que hay baja energía, es decir, que se perciben como silenciosos o casi silenciosos. Preferiblemente, la fase 300 de decisión garantiza en una realización de este tipo que el conmutador 200 sólo se activa cuando la parte de tiempo correspondiente que sigue el evento de conmutación tiene una energía que es, por ejemplo, inferior a la energía media de la señal de audio y es, preferiblemente, inferior al 50% de la energía media de la señal de audio en relación con, por ejemplo, dos o incluso más partes/tramas de tiempo de la señal de audio.

10

Preferiblemente, la segunda regla de codificación/regla de decodificación es un algoritmo de codificación basado en LPC. En la codificación de habla basada en LPC, se realiza una diferenciación entre partes de señal o segmentos de señal de excitación de a modo de impulso casi periódicos, y partes de señal o segmentos de señal de excitación a modo de ruido.

15

Los segmentos de señal de excitación a modo de impulso casi periódicos, es decir, los segmentos de señal que tienen una altura tonal específica se codifican con diferentes mecanismos que las señales de excitación a modo de ruido.

20

Mientras que las señales de excitación a modo de impulso casi periódicas están conectadas con habla con voz, las señales a modo de ruido están relacionadas con habla sin voz.

25

A modo de ejemplo, se hace referencia a las figuras 5a a 5d. En este caso, las partes de señal o segmentos a modo de impulso casi periódicos y las partes de señal o segmentos de señal a modo de ruido se comentan a modo de ejemplo. Específicamente, un habla con voz se ilustra en la figura 5a en el dominio de tiempo y en la figura 5b en el dominio de frecuencia se comenta como ejemplo para una parte de señal a modo de impulso casi periódica, y un segmento de habla sin voz como ejemplo para una parte de señal a modo de ruido se comenta en relación con las figuras 5c y 5d. El habla puede clasificarse generalmente como con voz, sin voz, o mezclado. Los esquemas de dominio de tiempo y frecuencia para segmentos con voz y sin voz muestreados se muestran en la figura 5a a 5d. El habla con voz es casi periódico en el dominio de tiempo y está estructurado de manera armónica en el dominio de frecuencia, mientras que el habla sin voz es a modo aleatorio y de banda ancha. Además, la energía de los segmentos con voz es generalmente mayor que la energía de los segmentos sin voz. El espectro de corto tiempo de habla con voz se caracteriza por su estructura fina y de formante. La estructura armónica fina es una consecuencia de la casi periodicidad del habla y puede atribuirse a la vibración de las cuerdas vocales. La estructura de formante (envolvente espectral) se debe a la interacción de la fuente y los tractos vocales. Los tractos vocales consisten en la faringe y la cavidad bucal. La forma de la envolvente espectral que "se adapta" al espectro de corto tiempo de habla con voz se asocia con las características de transferencia del tracto vocal y 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 modos resonantes del tracto vocal. Para el tracto vocal promedio hay de tres a cinco formantes por debajo de 5 kHz. Las amplitudes y ubicaciones de los primeros tres formantes, que se producen habitualmente por debajo de 3 kHz, son muy importantes tanto en síntesis de habla como en percepción. Los formantes superiores también son importantes para representaciones de banda amplia y habla sin voz. Las propiedades de habla se relacionan con el sistema de producción de habla física de la manera siguiente. El habla con voz se produce excitando el tracto vocal con pulsos glóticos de aire casi periódicos generados por la vibración de las cuerdas vocales. La frecuencia de los pulsos periódicos se denomina frecuencia fundamental o altura tonal. El habla sin voz se produce forzando el aire a través de una constricción en el tracto vocal. Los sonidos nasales se deben al acoplamiento acústico del tracto nasal al tracto vocal, y los sonidos explosivos se producen liberando abruptamente la presión de aire que se acumuló detrás del cierre en el tracto.

30

35

40

45

50

Por tanto, una parte a modo de ruido de la señal de audio no presenta una estructura de dominio de tiempo a modo de impulso ni una estructura de dominio de frecuencia armónica tal como se ilustra en la figura 5c y en la figura 5d, que es diferente de la parte a modo de impulso casi periódica tal como se ilustra por ejemplo en la figura 5a y en la figura 5b. Tal como se señalará más adelante, sin embargo, la diferenciación entre las partes a modo de ruido y las partes a modo de impulso casi periódicas puede observarse también después de una LPC para la señal de excitación. La LPC es un método que modela el tracto vocal y extrae de la señal la excitación de los tractos vocales.

55

Además, las partes a modo de impulso casi periódicas y las partes a modo de ruido pueden producirse de 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. Alternativamente, o adicionalmente, la característica de una señal puede ser diferente en diferentes bandas de frecuencia. Por tanto, la determinación de si la señal de audio es ruidosa o tonal, también puede realizarse de manera selectiva en frecuencia de modo que se considera(n) una banda de frecuencia determinada o varias bandas de frecuencia determinadas 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.

60

65

ES 2 380 307 T3

La figura 7a ilustra un modelo lineal de un sistema de producción de habla. Este sistema adopta una excitación de dos fases, es decir, un tren de impulsos para habla con voz tal como se indica en la figura 7c, y un ruido aleatorio para habla sin voz tal como se indica en la figura 7d. El tracto vocal se modela como un filtro 70 de todos los polos que procesa pulsos o ruido de la figura 7c o la figura 7d, generados por el modelo 72 glótico. La función de transferencia de todos los polos se forma mediante una cascada de un pequeño número de resonadores de dos polos que representan los formantes. El modelo glótico se representa como un filtro de paso bajo de dos polos, y el modelo 74 de radiación labial se representa por $L(z)=1-z^{-1}$. Finalmente, se incluye un factor 76 de corrección espectral para compensar los efectos de baja frecuencia de los polos superiores. En representaciones de habla individual se omite la corrección espectral y el 0 de la función de transferencia de radiación labial se cancela esencialmente por uno de los polos glóticos. Por eso, el sistema de la figura 7a puede reducirse a un modelo de filtro de todos los polos de la figura 7b que tiene una fase 77 de ganancia, un trayecto 78 hacia delante, un trayecto 79 de realimentación, y una fase 80 de adición. En el trayecto 79 de realimentación, hay un filtro 81 de predicción, y todo el sistema de síntesis de fuente-modelo ilustrado en la figura 7b puede representarse usando funciones de dominio z de la manera siguiente:

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

donde g representa la ganancia, $A(z)$ es el filtro de predicción según se determina por un análisis de LPC, $X(z)$ es la señal de excitación, y $S(z)$ es la salida de habla de síntesis.

Las figuras 7c y 7d proporcionan una descripción de dominio de tiempo gráfica de síntesis de habla con voz y sin voz usando el modelo de sistema de fuente lineal. Este sistema y los parámetros de excitación en la ecuación anterior no se conocen y deben determinarse a partir de un conjunto finito de muestras de habla. Los coeficientes de $A(z)$ se obtienen usando un análisis de predicción lineal de la señal de entrada y una cuantificación de los coeficientes de filtro. En un predictor lineal delantero de orden p , se predice la muestra actual de la secuencia de habla desde una combinación lineal de p muestras pasadas. Los coeficientes de predictor pueden determinarse por algoritmos bien conocidos tales como el algoritmo de Levinson-Durbin, o generalmente un método de autocorrelación o un método de reflexión. La cuantificación de los coeficientes de filtro obtenidos se realiza habitualmente por una cuantificación de vectores multifase en la LSF o en el dominio de ISP.

La figura 7e ilustra una implementación más detallada de un bloque de análisis de LPC, tal como 510 de la figura 1a. 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. En la realización de la figura 4a, es decir, la información de predicción a corto plazo podría requerirse para la señal de salida de codificador de impulso. Sin embargo, cuando sólo se requiere la señal de error de predicción en la línea 84, no tiene que emitirse la información de predicción a corto plazo. No obstante, se requiere la información de predicción a corto plazo por el filtro 85 de predicción real. En un sustractor 86, se introduce una muestra actual de la señal de audio y se sustrae un valor predicho para la muestra actual de modo que para esta muestra, se genera la señal de error de predicción en la línea 84. Una secuencia de tales muestras de señal de error de predicción se ilustra de manera esquemática en las figuras 7c o 7d, en las que, por motivos de claridad, no se ha ilustrado ningún problema con respecto a componentes de CC/CA, etc. Por tanto, la figura 7c puede considerarse como una clase de señal a modo de impulso rectificadas.

Posteriormente, se comentará un codificador de CELP de análisis por síntesis en relación con la figura 6 con el fin de ilustrar las modificaciones aplicadas a este algoritmo, tal como se ilustra en las figuras 10 a 13. Este codificador de CELP se comenta en detalle en "Speech Coding: A Tutorial Review", Andreas Spaniels, Proceedings of the IEEE, Vol. 82, n.º 10, octubre de 1994, páginas 1541-1582. El codificador de CELP tal 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 usa un libro de códigos que se indica en 64. Se implementa un filtro de ponderación de percepción $W(z)$ en 66, y se proporciona un controlador de minimización de errores en 68. $s(n)$ es la señal de entrada de dominio de tiempo. Después de haberse ponderado perceptivamente, la señal ponderada se introduce a un sustractor 69, que calcula el error entre la señal de síntesis ponderada en la salida del bloque 66 y la señal ponderada original $s_w(n)$. Generalmente, se calcula la predicción a corto plazo $A(z)$ y sus coeficientes se cuantifican por una fase de análisis de LPC tal como se indica en la figura 7e. La información de predicción a largo plazo $A_L(z)$ que incluye una ganancia de predicción a largo plazo g y el índice de cuantificación de vectores, es decir, referencias de libro de códigos, se calculan en la señal de error de predicción en la salida de la fase de análisis de LPC denominada 10a en la figura 7e. El algoritmo de CELP codifica entonces la señal residual obtenida después de las predicciones a corto plazo y a largo plazo usando un libro de códigos de, por ejemplo, secuencias gaussianas. El algoritmo de ACELP, en el que "A" significa "Algebraico" tiene un libro de códigos específico diseñado algebraicamente.

Un libro de códigos puede contener más o menos vectores de los que cada vector tiene algunas muestras de longitud.

Un factor de ganancia g ajusta a escala el vector de código y el código ganado se filtra por 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 el error cuadrático medio perceptivamente ponderado en la salida del sustractor 69 se minimiza. El proceso de búsqueda en CELP se realiza mediante una optimización de análisis por síntesis tal como se ilustra en la figura 6.

ES 2 380 307 T3

Para casos específicos, cuando una trama es una mezcla de habla sin voz y con voz o cuando se produce un habla sobre música, una codificación de TCX puede ser más apropiada para codificar la excitación en el dominio de LPC. La codificación de TCX procesa directamente la excitación en el dominio de frecuencia sin realizar ninguna suposición de producción de excitación. El TCX es entonces más genérico que la codificación de CELP y no se limita a un modelo de fuente con voz o sin voz de la excitación. El TCX es aún una codificación de modelo de fuente-filtro que usa un filtro de predicción lineal para modelar los formantes de las señales similares al habla.

En la codificación similar a AMR-WB+, tiene lugar una selección entre diferentes modos de TCX y ACELP tal como se conoce a partir de la descripción de AMR-WB+. Los modos de TCX son diferentes porque la longitud de la transformada rápida de Fourier por bloques es diferente para modos diferentes y el mejor modo puede seleccionarse mediante un enfoque de análisis por síntesis o mediante un modo de “alimentación” directa.

Tal como se comentó en relación con las figuras 2a y 2b, la fase 100 de preprocesamiento común preferiblemente incluye un multicanal 101 conjunto (dispositivo estéreo conjunto/envolvente) y, adicionalmente, una fase 102 de extensión de ancho de banda. De manera correspondiente, el decodificador incluye una fase 701 de extensión de ancho de banda y una fase 702 multicanal conjunta posteriormente conectada. Preferiblemente, la fase 101 multicanal conjunta se conecta, con respecto al codificador, antes de la fase 102 de extensión de ancho de banda, y, en el lado de decodificador, la fase 701 de extensión de ancho de banda se conecta antes de la fase 702 multicanal conjunta con respecto a la dirección de procesamiento de señal. Alternativamente, sin embargo, la fase de preprocesamiento común puede incluir una fase multicanal conjunta sin una fase de extensión de ancho de banda posteriormente conectada o una fase de extensión de ancho de banda sin una fase multicanal conjunta conectada.

Un ejemplo preferido para una fase multicanal conjunta en el lado 101a, 101b de codificador y en el lado 702a y 702b de decodificador se ilustra en el contexto de la figura 8. Un número de E canales de entrada originales se introduce al mezclador 101a descendente de modo que el mezclador descendente genera un número de K canales transmitidos, donde el número K es mayor que o igual a uno y es más pequeño que E.

Preferiblemente, los E canales de entrada se introducen en un analizador 101b de parámetros multicanal conjunto que genera información paramétrica. Esta información paramétrica se codifica preferiblemente por entropía tal como por una codificación diferente y una codificación de Huffman posterior o, alternativamente, una codificación aritmética posterior. La información paramétrica codificada emitida por el bloque 101b se transmite a un decodificador 702b de parámetros que puede formar parte del elemento 702 en la figura 2b. El decodificador 702b de parámetros decodifica la información paramétrica transmitida y envía la información paramétrica decodificada en el mezclador 702a ascendente. El mezclador 702a ascendente recibe los K canales transmitidos y genera un número de L canales emitidos, donde el número de L es mayor que K y menor que o igual a E.

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 tal como se conoce de la técnica de BCC o tal como se conoce y se describe en detalle en la norma envolvente de MPEG. El número de canales transmitidos puede ser un único canal monofónico para aplicaciones de tasa de transmisión de bits ultrabaja 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 cinco o quizás incluso mayor. Alternativamente, el número de E canales de entrada también puede ser E objetos de audio tal como se conoce en el contexto de codificación de objeto de audio espacial (SAOC).

En una implementación, el mezclador descendente realiza una adición ponderada o no ponderada de los E canales de entrada originales o una adición de los E objetos de audio de entrada. En el caso de objetos de audio como canales de entrada, el analizador 101b de parámetros multicanal conjunto calculará los parámetros de objeto de audio tales como una matriz de correlación entre los objetos de audio preferiblemente para cada parte de tiempo e incluso más preferiblemente para cada banda de frecuencia. Con este fin, todo el intervalo de frecuencia puede dividirse en al menos 10 y preferiblemente 32 ó 64 bandas de frecuencia.

La figura 9 ilustra una realización preferida para la implementación de la fase 102 de extensión de ancho de banda en la figura 2a y la fase 701 de extensión de ancho de banda correspondiente en la figura 2b. En el lado de codificador, el bloque 102 de extensión de ancho de banda preferiblemente incluye un bloque 102b de filtrado de paso bajo y analizador 102a de banda alta. La señal de audio original introducida en el bloque 102 de extensión de ancho de banda se filtra en paso bajo para generar la señal de banda baja que luego se introduce en las ramas de codificación y/o el conmutador. El filtro de paso bajo tiene una frecuencia de corte que está normalmente en un intervalo de 3 kHz a 10 kHz. Usando SBR, puede superarse este intervalo. Además, el bloque 102 de extensión de ancho de banda incluye además un analizador de banda alta para calcular los parámetros de extensión de ancho de banda tales como una información de parámetros de envolvente espectral, una información de parámetros de umbral mínimo de ruido, una información de parámetros de filtrado inverso, información paramétrica adicional en relación con determinadas líneas armónicas en la banda alta y parámetros adicionales según se comenta en detalle en la norma de MPEG-4 en el capítulo relacionado con la replicación de banda espectral (ISO/IEC 14496-3:2005, Parte 3, Capítulo 4.6.18).

En el lado de decodificador, el bloque 701 de extensión 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 ajustada y reconstruida emitida por el ajustador 701b. Se proporciona la entrada en el ajustador 701b por un

parcheador que se opera para derivar la señal de banda alta a partir de la señal de banda baja tal como mediante replicación de banda espectral o, generalmente, mediante extensión de ancho de banda. El parcheado realizado por el parcheador 701a puede ser un parcheado realizado de manera armónica o de manera no armónica. La señal generada por el parcheador 701a se ajusta, posteriormente, por el ajustador 701b usando la información de extensión de ancho de banda paramétrica transmitida.

Tal como se indica en la figura 8 y la figura 9, los bloques descritos pueden tener una entrada de control de modo en una realización preferida. Esta entrada de control de modo se deriva de la señal de salida de la fase 300 de decisión. En tal realización preferida, puede adaptarse una característica de un bloque correspondiente a la salida de la fase de decisión, es decir, si, en una realización preferida, se realiza una decisión para habla o una decisión para música durante una determinada parte de tiempo de la señal de audio. Preferiblemente, el control de modo sólo se refiere a una o más de las funcionalidades de estos bloques pero no a todas las funcionalidades de bloques. Por ejemplo, la decisión puede influir sólo al parcheador 701a pero puede no influir a los otros bloques en la figura 9, o, por ejemplo, puede influir sólo al analizador 101b de parámetros multicanal conjuntos en la figura 8 pero no a los otros bloques en la figura 8. Esta implementación es preferiblemente de manera que se obtiene una mayor flexibilidad y mayor calidad y menor señal de salida de tasa de transmisión de bits proporcionando flexibilidad en la fase de preprocesamiento común. Por otro lado, sin embargo, el uso de algoritmos en la fase de preprocesamiento común para ambas clases de señales permite implementar un esquema de codificación/decodificación eficaz.

La figura 10a y la figura 10b ilustran dos implementaciones diferentes de la fase 300 de decisión. En la figura 10a, se indica una decisión de lazo abierto. En este caso, el analizador 300a de señales en la fase de decisión tiene determinadas reglas con el fin de 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 se codifique por la primera rama 400 de codificación o por la segunda rama 500 de codificación. Con este fin, el analizador 300a de señales puede analizar la señal de entrada de audio en la fase de preprocesamiento común o puede analizar la señal de audio emitida por la fase de preprocesamiento común, es decir, la señal intermedia de audio o puede analizar una señal intermedia dentro de la fase de preprocesamiento común tal como la salida de la señal de mezcla descendente que puede ser una señal monofónica o que puede ser una señal que tiene k canales indicados en la figura 8. En el lado de salida, el analizador 300a de señales genera la decisión de conmutación para controlar el conmutador 200 en el lado de codificador y el correspondiente conmutador 600 o el combinador 600 en el lado de decodificador.

Alternativamente, la fase 300 de decisión puede realizar una decisión de lazo cerrado, que significa que ambas ramas de codificación realizan sus tareas en la misma parte de la señal de audio y ambas señales codificadas se decodifican mediante ramas 300c, 300d de decodificación correspondientes. La salida de los dispositivos 300c y 300d se introduce en un comparador 300b que compara la salida de los dispositivos de decodificación con la parte correspondiente de, por ejemplo, la señal intermedia de audio. Entonces, dependiendo de una función de coste tal como una relación señal a ruido por rama, se realiza una decisión de conmutación. Esta decisión de lazo cerrado tiene una complejidad aumentada en comparación con la decisión de lazo abierto, pero esta complejidad sólo existe en el lado de codificador, y un decodificador no tiene ninguna desventaja de este proceso, puesto que el decodificador puede usar ventajosamente la salida de esta decisión de codificación. Por tanto, se prefiere el modo de lazo cerrado debido a consideraciones de complejidad y calidad en aplicaciones en las que la complejidad del decodificador no es un problema tal como en aplicaciones de difusión en las que sólo hay un pequeño número de codificadores pero un gran número de decodificadores que, además, deben ser inteligentes y baratos.

La función de coste aplicada por el comparador 300b puede ser una función de coste guiada por aspectos de calidad o puede ser una función de coste guiada por aspectos de ruido o puede ser una función de coste guiada por aspectos de tasa de transmisión de bits o puede ser una función de coste combinada guiada por cualquier combinación de tasa de transmisión de bits, calidad, ruido (introducido por artefactos de codificación, específicamente, mediante cuantificación), etc.

Preferiblemente, la primera rama de codificación y/o la segunda rama de codificación incluye una funcionalidad de alineamiento de tiempo en el lado de codificador y de manera correspondiente en el lado de decodificador. En una realización, la primera rama de codificación comprende un módulo de alineamiento de tiempo para calcular una característica de alineamiento variable dependiente de una parte de la señal de audio, un remuestreador para volver a muestrear según la característica de alineamiento determinada, un convertidor de dominio de tiempo/dominio de frecuencia, y un codificador de entropía para convertir un resultado de la conversión de dominio de tiempo/dominio de frecuencia en una representación codificada. La característica de alineamiento variable se incluye en la señal de audio codificada. Esta información se lee mediante una rama de decodificación mejorada de alineamiento de tiempo y se procesa para tener finalmente una señal de salida en una escala de tiempo no alineada. Por ejemplo, la rama de decodificación realiza una decodificación de entropía, descuantificación y una conversión del dominio de frecuencia de vuelta al dominio de tiempo. En el dominio de tiempo, puede aplicarse el desalineamiento y puede seguirse por una operación de remuestreo correspondiente para obtener finalmente una señal de audio discreta con una escala de tiempo no alineada.

Dependiendo de determinados requisitos de implementación de los métodos de la invención, los métodos de la invención pueden implementarse en hardware o en software. La implementación puede realizarse usando un medio de almacenamiento digital, en particular, un disco, un DVD o un CD que tiene señales de control electrónicamente legibles almacenadas en el mismo, que actúan conjuntamente con sistemas informáticos programables de manera que

ES 2 380 307 T3

se realizan los métodos de la invención. Generalmente, la presente invención es por tanto un producto de programa informático con un código de programa almacenado un soporte legible por ordenador, pudiendo operarse el código de programa para realizar los métodos de la invención cuando el producto de programa informático se ejecuta en un ordenador. En otras palabras, los métodos de la invención son, por tanto, un programa informático que tiene un código de programa para realizar al menos uno de los métodos de la invención cuando el programa informático se ejecuta en un ordenador.

La señal de audio codificada 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.

Las realizaciones descritas anteriormente son meramente ilustrativas de los principios de la presente invención. Se entiende que las modificaciones y variaciones de las disposiciones y los detalles descritos en el presente documento serán evidentes para otros expertos en la técnica. Por tanto, se pretende que se limite sólo mediante el alcance de las reivindicaciones de patente adjuntas y no por los detalles específicos presentados a modo de descripción y explicación de las realizaciones en el presente documento.

20

25

30

35

40

45

50

55

60

65

REIVINDICACIONES

1. Codificador de audio para generar una señal de audio codificada, que comprende:

5 una primera rama (400) de codificación para codificar una señal (195) intermedia de audio según un primer algoritmo de codificación, teniendo el primer algoritmo de codificación un modelo de sumidero de información y generando, en una señal de salida de primera rama de codificación, información espectral codificada que representa la señal intermedia de audio, comprendiendo la primera rama de codificación un bloque (410) de conversión espectral para convertir la señal intermedia de audio en un dominio espectral y un codificador (420) de audio espectral para codificar una señal de salida del bloque (410) de conversión espectral para obtener la información espectral codificada;

15 una segunda rama (500) de codificación para codificar la señal (195) intermedia de audio según un segundo algoritmo de codificación, teniendo el segundo algoritmo de codificación un modelo de fuente de información y generando, en una señal de salida de segunda rama de codificación, parámetros codificados para el modelo de fuente de información que representan la señal (195) intermedia de audio, comprendiendo la segunda rama de codificación un analizador (510) de LPC para analizar la señal intermedia de audio y para emitir una señal de información de LPC que puede usarse para controlar un filtro de síntesis de LPC y una señal de excitación, y un codificador (520) de excitación para codificar la señal de excitación para obtener los parámetros codificados; y

20 una fase (100) de preprocesamiento común para preprocesar una señal (99) de entrada de audio para obtener la señal (195) intermedia de audio, en el que la fase (100) de preprocesamiento común es operativa para procesar la señal (99) de entrada de audio de modo que la señal (195) intermedia de audio es una versión comprimida de la señal (99) de entrada de audio.

25 2. Codificador de audio según la reivindicación 1, que comprende además una etapa (200) de conmutación conectada entre la primera rama (400) de codificación y la segunda rama (500) de codificación en entradas a las ramas o salidas de las ramas, controlándose la etapa de conmutación por una señal de control de conmutación.

30 3. Codificador de audio según la reivindicación 2, que comprende además una fase (300, 300a, 300b) de decisión para analizar la señal (99) de entrada de audio o la señal (195) intermedia de audio o una señal intermedia en la fase (100) de preprocesamiento común en tiempo o frecuencia con el fin de hallar una parte de tiempo o frecuencia de una señal que va a transmitirse en una señal de salida de codificador o bien como la señal de salida codificada generada por la primera rama de codificación o bien como la señal de salida codificada generada por la segunda rama de codificación.

40 4. Codificador de audio según una de las reivindicaciones anteriores, en el que la fase (100) de preprocesamiento común es operativa para calcular parámetros de preprocesamiento comunes para una parte de la señal de entrada de audio no incluida en una primera y una segunda parte diferente de la señal (195) intermedia de audio y para introducir una representación codificada de los parámetros de preprocesamiento en la señal de salida codificada, en el que la señal de salida codificada comprende adicionalmente una señal de salida de primera rama de codificación para representar una primera parte de la señal intermedia de audio y una señal de salida de segunda rama de codificación para representar la segunda parte de la señal intermedia de audio.

45 5. Codificador de audio según una de las reivindicaciones anteriores, en el que la fase (100) de preprocesamiento común comprende un módulo (101) multicanal conjunto, comprendiendo el módulo multicanal conjunto:

50 un mezclador (101a) descendente para generar un número de canales mezclados de manera descendente que son superiores a o igual a 1 y que son más pequeños que un número de canales introducidos en el mezclador (101a) descendente; y

55 un calculador (101b) de parámetros multicanal para calcular parámetros multicanal de modo que, usando los parámetros multicanal y el número de canales mezclados de manera descendente, pueda realizarse una representación del canal original.

60 6. Aparato según la reivindicación 5, en el que los parámetros multicanal son parámetros de diferencia de nivel entre canales, correlación entre canales o parámetros de coherencia, parámetros de diferencia de fase entre canales, parámetros de diferencia de tiempo entre canales, parámetros de objeto de audio o parámetros de difusión o dirección.

7. Codificador de audio según una de las reivindicaciones anteriores, en el que la fase (100) de preprocesamiento común comprende una fase (102) de análisis de extensión de ancho de banda, que comprende:

65 un dispositivo (102b) de limitación de banda para rechazar una banda alta en una señal de entrada y para generar una señal de banda baja; y

ES 2 380 307 T3

un calculador (102a) de parámetros para calcular parámetros de extensión de ancho de banda para la banda alta rechazada por el dispositivo de limitación de banda, en el que el calculador (102a) de parámetros es tal que usando los parámetros calculados y la señal de banda baja, puede realizarse una reconstrucción de una señal de entrada extendida en ancho de banda.

5

8. Codificador de audio según una de las reivindicaciones anteriores, en el que la fase (100) de preprocesamiento común incluye un módulo (101) multicanal conjunto, una fase (102) de extensión de ancho de banda, y un conmutador (200) para conmutar entre la primera rama (400) de codificación y la segunda rama (500) de codificación, en el que una salida de la fase (101) multicanal conjunta está conectada a una entrada de la fase (102) de extensión de ancho de banda, y una salida de la fase de extensión de ancho de banda está conectada a una entrada del conmutador (200), una primera salida del conmutador está conectada a una entrada de la primera rama de codificación y una segunda salida del conmutador está conectada a una entrada de la segunda rama (500) de codificación, y las salidas de las ramas de codificación están conectadas a un formador (800) de flujo de bits.

15

9. Codificador de audio según la reivindicación 3, en el que la fase (300) de decisión es operativa para analizar una señal de entrada de fase de decisión para buscar partes que van a codificarse por la primera rama (400) de codificación con una mejor relación señal a ruido a una determinada tasa de transmisión de bits en comparación con la segunda rama (500) de codificación, en el que la fase (300) de decisión es operativa para analizar basándose en un algoritmo de lazo abierto sin una señal codificada y de nuevo decodificada o basándose en un algoritmo de lazo cerrado usando una señal codificada y de nuevo decodificada.

20

10. Codificador de audio según la reivindicación 3,

25

en el que la fase de preprocesamiento común tiene un número específico de funcionalidades (101a, 101b, 102a, 102b) y en el que al menos una funcionalidad es adaptable por una señal de salida de fase (300) de decisión y en el que al menos una funcionalidad es no adaptable.

30

11. Codificador de audio según una de las reivindicaciones anteriores,

en el que la primera rama de codificación comprende un módulo de alineamiento de tiempo para calcular una característica de alineamiento variable dependiente de una parte de la señal de audio, en el que la primera rama de codificación comprende un remuestreador para volver a muestrear según una característica de alineamiento determinada,

35

y en el que la primera rama de codificación comprende un convertidor de dominio de tiempo/dominio de frecuencia y un codificador de entropía para convertir un resultado de la conversión de dominio de tiempo/dominio de frecuencia en una representación codificada,

40

en el que la característica de alineamiento variable se incluye en la señal de audio codificada.

45

12. Codificador de audio según una de las reivindicaciones anteriores, en el que la fase de preprocesamiento común es operativa para emitir al menos dos señales intermedias, y en el que, para cada señal intermedia de audio, se proporciona la rama de codificación primera y segunda y un conmutador para conmutar entre las dos ramas.

13. Método de codificación de audio para generar una señal de audio codificada, que comprende:

50

codificar (400) una señal (195) intermedia de audio según un primer algoritmo de codificación, teniendo el primer algoritmo de codificación un modelo de sumidero de información y generando, en una primera señal de salida, información espectral codificada que representa la señal de audio, comprendiendo el primer algoritmo de codificación una etapa (410) de conversión espectral de conversión de la señal intermedia de audio en un dominio espectral y una etapa (420) de codificación de audio espectral de codificación de una señal de salida de la etapa (410) de conversión espectral para obtener la información espectral codificada;

55

codificar (500) la señal (195) intermedia de audio según un segundo algoritmo de codificación, teniendo el segundo algoritmo de codificación un modelo de fuente de información y generando, en una segunda señal de salida, parámetros codificados para el modelo de fuente de información que representan la señal intermedia (195), comprendiendo la segunda rama de codificación una etapa de análisis (510) de LPC la señal intermedia de audio y emitir una señal de información de LPC que puede usarse para controlar un filtro de síntesis de LPC, y una señal de excitación, y una etapa de codificación (520) de excitación de la señal de excitación para obtener los parámetros codificados; y

60

preprocesar (100) de manera común una señal (99) de entrada de audio para obtener la señal (195) intermedia de audio, en el que, en la etapa de preprocesar de manera común la señal (99) de entrada de audio se procesa de modo que la señal (195) intermedia de audio es una versión comprimida de la señal (99) de entrada de audio,

65

ES 2 380 307 T3

en el que la señal de audio codificada incluye, para una parte determinada de la señal de audio o bien la primera señal de salida o bien la segunda señal de salida.

5 14. Decodificador de audio para decodificar una señal de audio codificada, que comprende:

una primera rama (430, 440) de decodificación para decodificar una señal codificada, codificada según un primer algoritmo de codificación que tiene un modelo de sumidero de información, comprendiendo la primera rama de decodificación un decodificador (430) de audio espectral para la decodificación de audio espectral de la señal codificada,
10 codificada según un primer algoritmo de codificación que tiene un modelo de sumidero de información, y un convertidor (440) de dominio de tiempo para convertir una señal de salida del decodificador (430) de audio espectral en el dominio de tiempo;

una segunda rama (530, 540) de decodificación para decodificar una señal de audio codificada, codificada según un segundo algoritmo de codificación que tiene un modelo de fuente de información, comprendiendo la segunda rama de decodificación un decodificador (530) de excitación para decodificar la señal de audio codificada, codificada según un segundo algoritmo de codificación para obtener una señal de dominio de LPC, y una fase (540) de síntesis de LPC para recibir una señal de información de LPC generada por una fase de análisis de LPC y para convertir la señal de dominio de LPC en el dominio de tiempo;

20 un combinador (600) para combinar señales de salida de dominio de tiempo del convertidor (440) de dominio de tiempo de la primera rama (430, 440) de decodificación y la fase (540) de síntesis de LPC de la segunda rama (530, 540) de decodificación para obtener una señal (699) combinada; y

25 una fase (700) de posprocesamiento común para procesar la señal (699) combinada de modo que una señal (799) de salida decodificada de la fase de posprocesamiento común sea una versión expandida de la señal (699) combinada.

15. Decodificador de audio según la reivindicación 14, en el que el combinador (600) comprende un conmutador para conmutar señales decodificadas de la primera rama (450) de decodificación y la segunda rama (550) de decodificación según una indicación de modo incluida explícita o implícitamente en la señal de audio codificada de modo que la señal (699) de audio combinada es una señal de dominio de tiempo discreta continua.

16. Decodificador de audio según la reivindicación 14 ó 15, en el que el combinador (600) comprende un atenuador (607) cruzado para atenuación cruzada, en caso de un evento de conmutación, entre una salida de una rama (450, 550) de decodificación y una salida de la otra rama (450, 550) de decodificación dentro de una región de atenuación cruzada de dominio de tiempo.

17. Decodificador de audio según la reivindicación 16, en el que el atenuador (607) cruzado está operativo para ponderar al menos una de las señales de salida de rama de decodificación dentro de la región de atenuación cruzada y para añadir al menos una señal ponderada a una señal ponderada o no ponderada de la otra rama (607c) de codificación, en el que los pesos usados para ponderar la al menos una señal (607a, 607b) son variables en la región de atenuación cruzada.

45 18. Decodificador de audio según una de las reivindicaciones 14 a 17, en el que la fase de preprocesamiento común comprende al menos una de un decodificador (101) multicanal conjunto o un procesador (102) de extensión de ancho de banda.

19. Decodificador de audio según la reivindicación 18,

50 en el que el decodificador (702) multicanal conjunto comprende un decodificador (702b) de parámetros y un mezclador (702a) ascendente controlado por una salida de decodificador (702b) de parámetros.

55 20. Decodificador de audio según la reivindicación 19,

en el que el procesador (702) de extensión de ancho de banda comprende un parcheador (701a) para crear una señal de banda alta, un ajustador (701b) para ajustar la señal de banda alta, y un combinador (701c) para combinar la señal de banda alta ajustada y una señal de banda baja para obtener una señal extendida en ancho de banda.

21. Decodificador de audio según una de las reivindicaciones 14 a 20, en el que la primera rama (450) de decodificación incluye un decodificador de audio de dominio de frecuencia, y la segunda rama (550) de decodificación incluye un decodificador de habla de dominio de tiempo.

65 22. Decodificador de audio según una de las reivindicaciones 14 a 20, en el que la primera rama (450) de decodificación incluye un decodificador de audio de dominio de frecuencia, y la segunda rama (550) de decodificación incluye un decodificador basado en LPC.

ES 2 380 307 T3

23. Decodificador de audio según una de las reivindicaciones 14 a 22,

5 en el que la fase de posprocesamiento común tiene un número específico de funcionalidades (700, 701, 702) y en el que al menos una funcionalidad es adaptable por una función (601) de detección de modo y en el que al menos una funcionalidad es no adaptable.

24. Método de decodificación de audio de una señal de audio codificada, que comprende:

10 decodificar (450) una señal codificada, codificada según un primer algoritmo de codificación que tiene un modelo de sumidero de información, que comprende la decodificación (430) de audio espectral de la señal codificada, codificada según un primer algoritmo de codificación que tiene un modelo de sumidero de información, y una conversión (440) de dominio de tiempo de una señal de salida de la etapa (430) de decodificación de audio espectral en el dominio de tiempo;

15 decodificar (550) una señal de audio codificada, codificada según un segundo algoritmo de codificación que tiene un modelo de fuente de información, que comprende la decodificación (530) de excitación de la señal de audio codificada, codificada según un segundo algoritmo de codificación para obtener una señal de dominio de LPC, y para recibir una señal de información de LPC generada por una fase de análisis de LPC y síntesis (540) de LPC para convertir la señal de dominio de LPC en el dominio de tiempo;

20 combinar (600) las señales de salida de dominio de tiempo a partir de la etapa de conversión (440) de dominio de tiempo y la etapa de síntesis (540) de LPC para obtener una señal (699) combinada; y

25 procesar (700) de manera común la señal (699) combinada de modo que una señal (799) de salida decodificada de la fase de posprocesamiento común sea una versión expandida de la señal (799) combinada.

30 25. Programa informático adaptado para realizar, cuando se ejecuta en un ordenador, el método según la reivindicación 14 o la reivindicación 24.

35

40

45

50

55

60

65

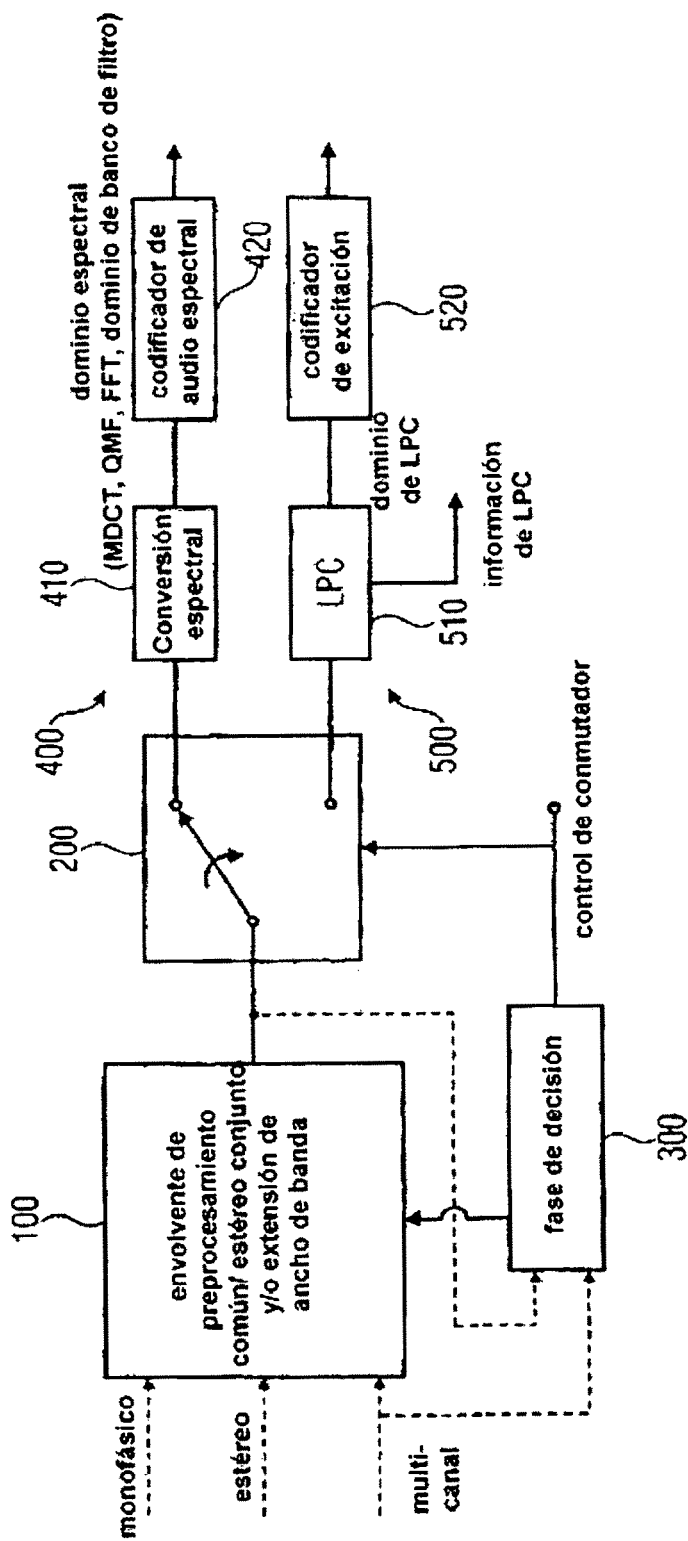


FIG 1A
(Codificador)

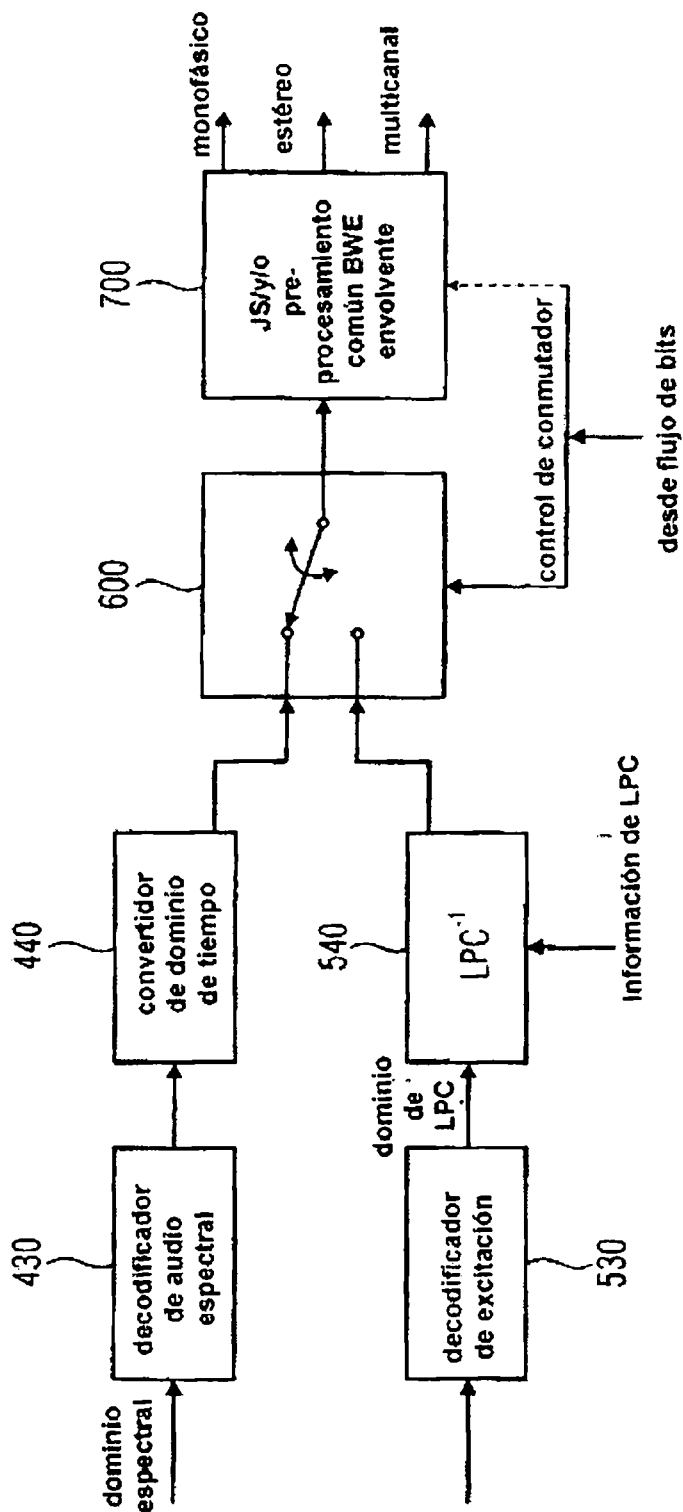


FIG 1B
(Decodificador)

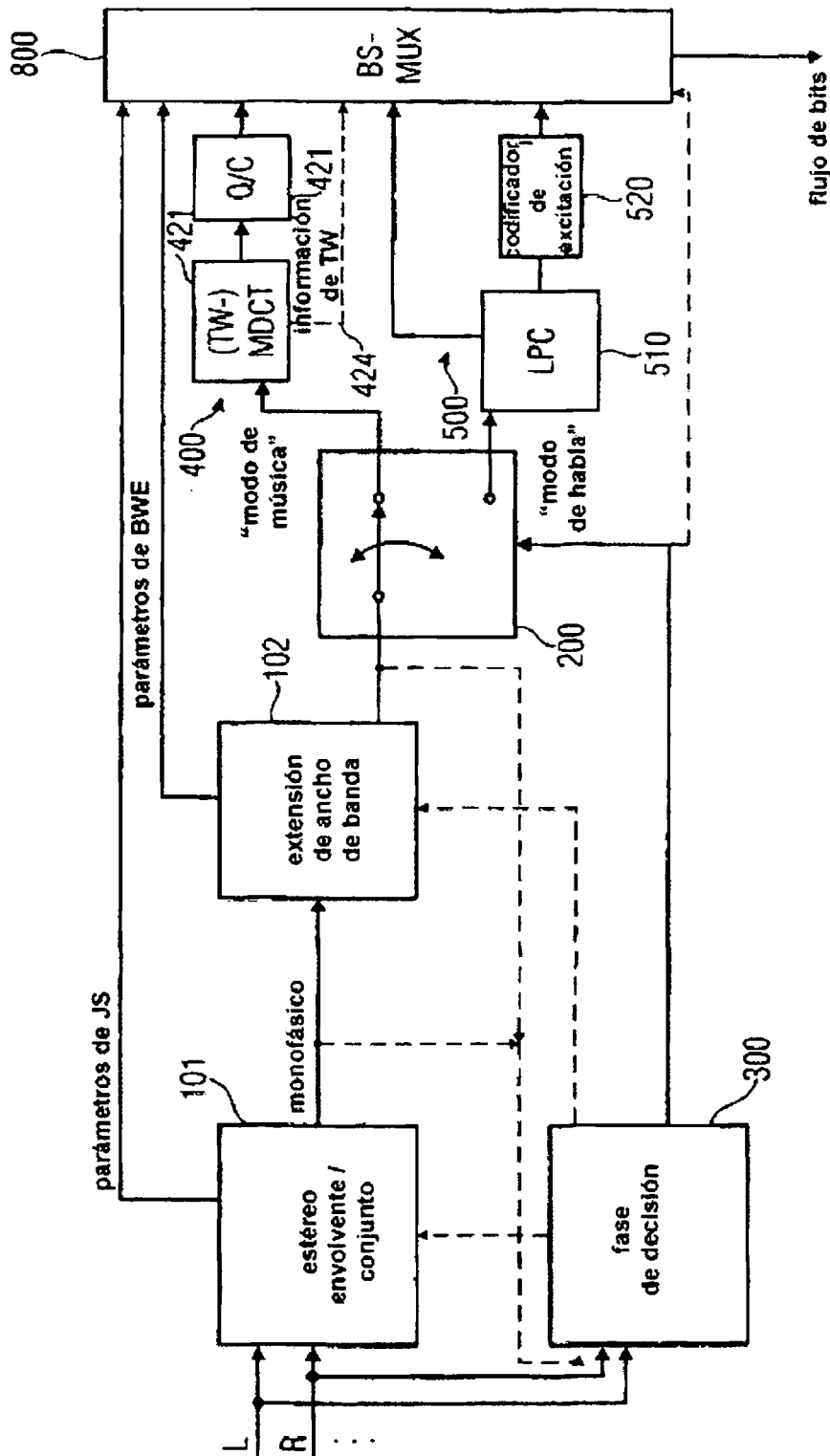


FIG 2A
(Codificador)

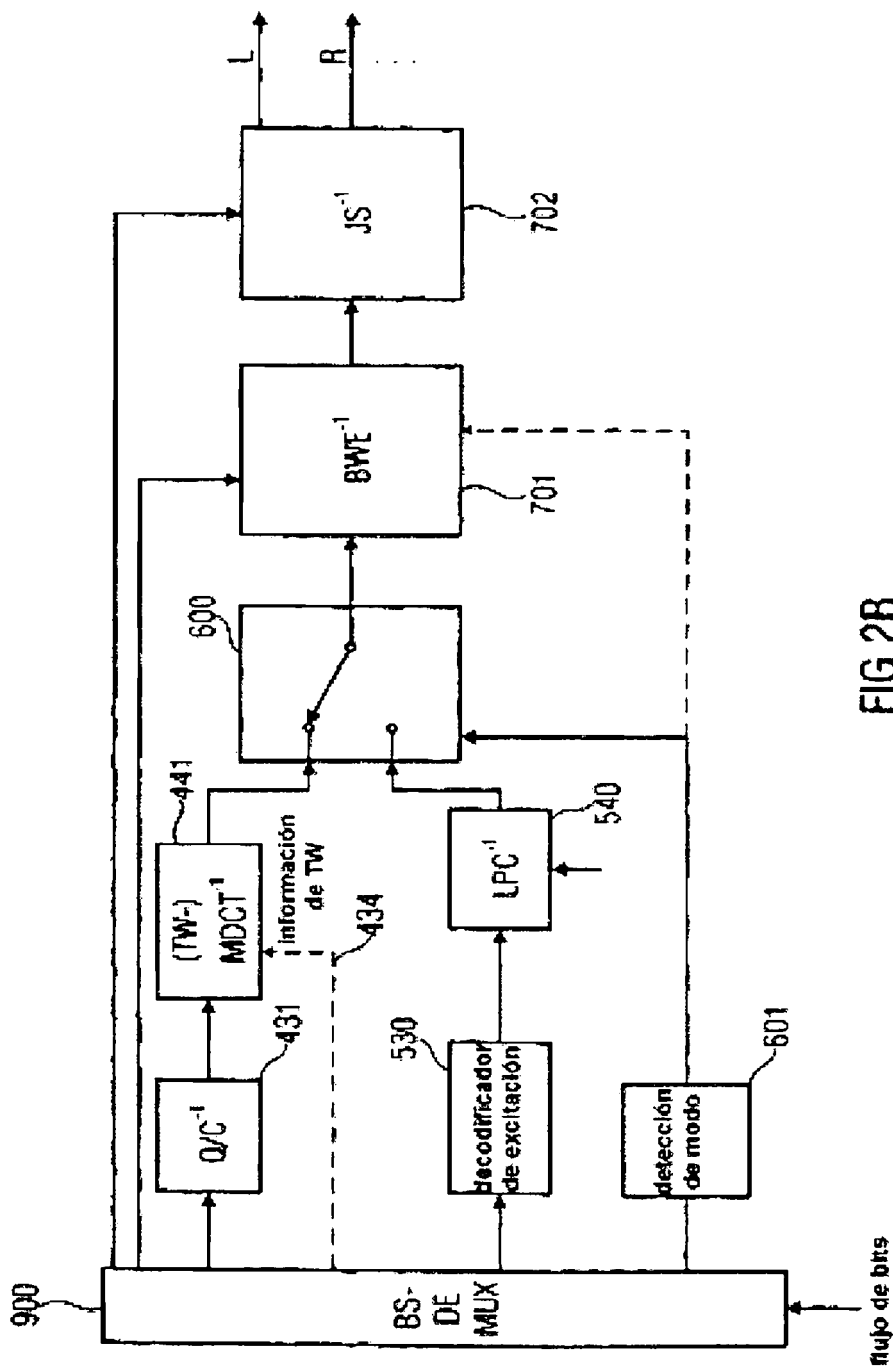


FIG 2B
(Decodificador)

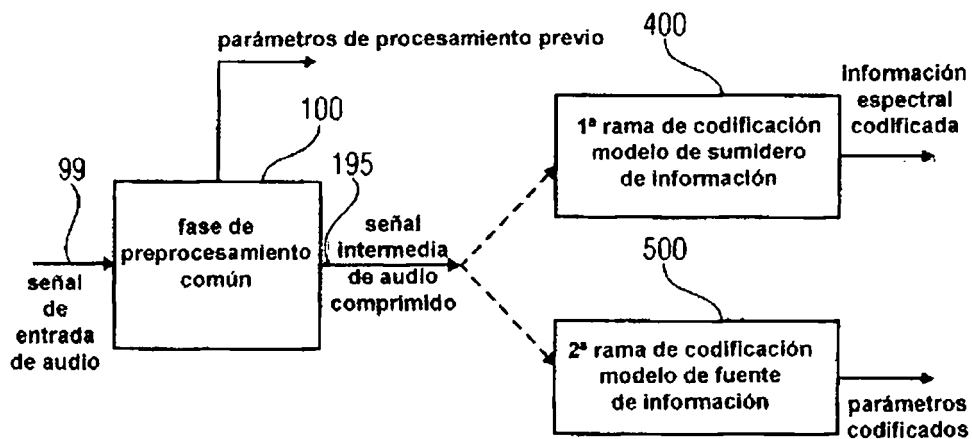


FIG 3A

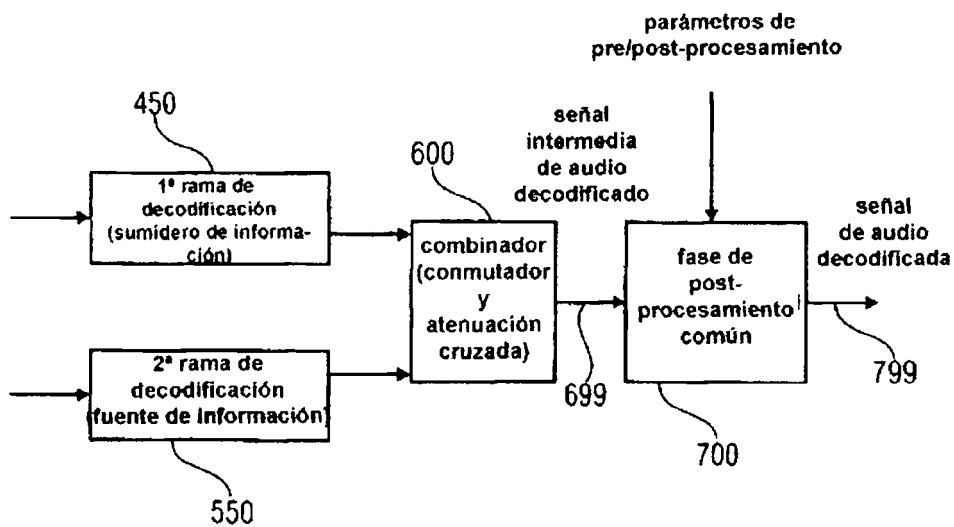


FIG 3B

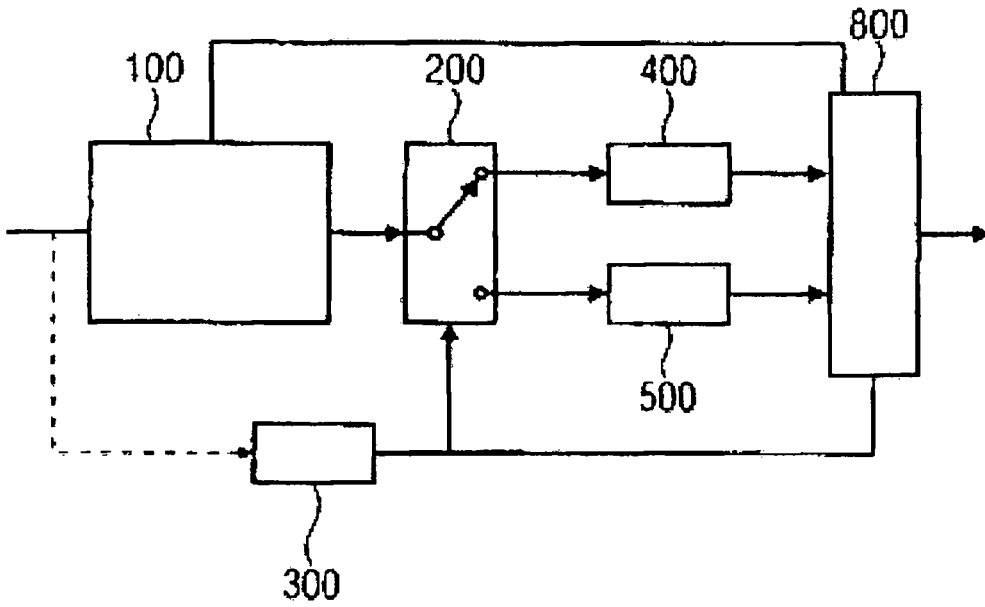


FIG 4A

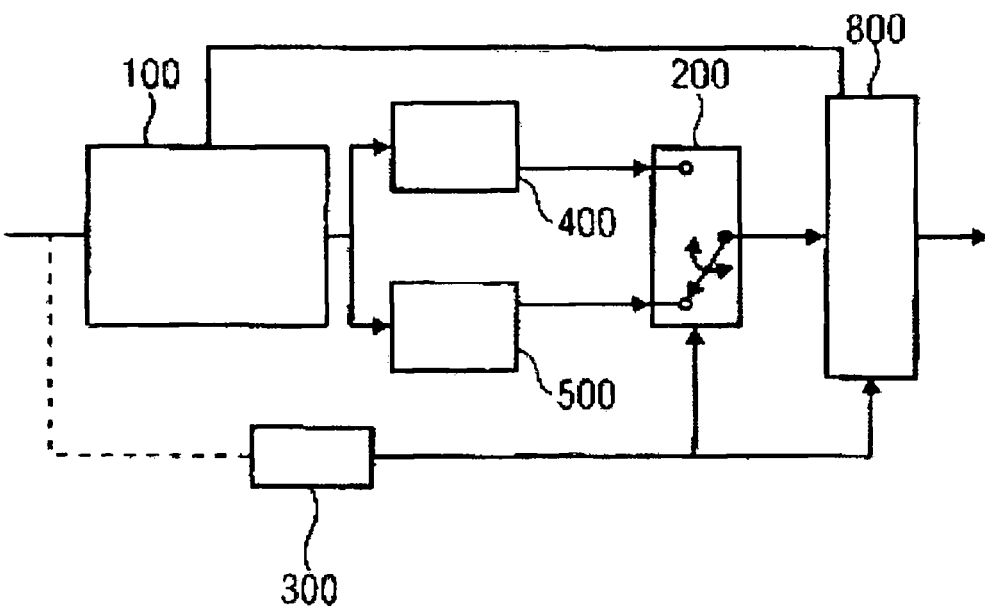


FIG 4B

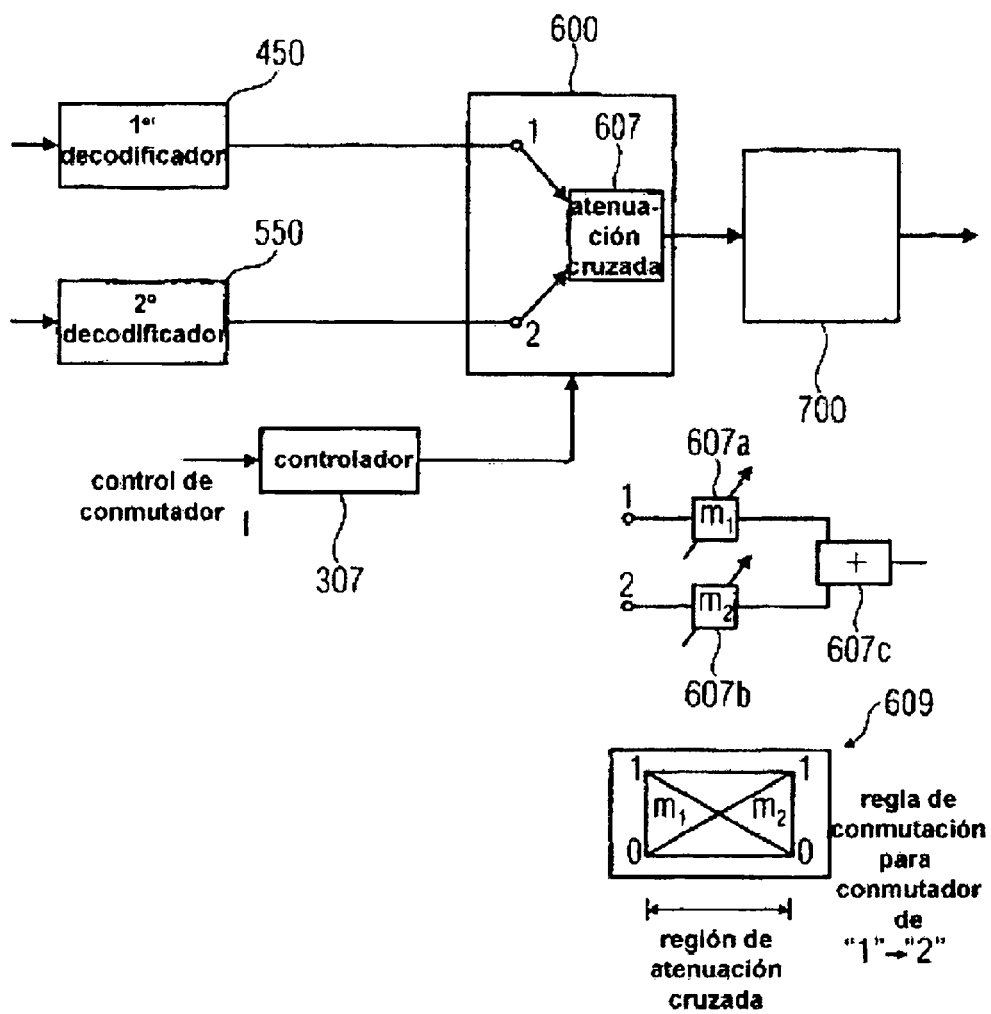


FIG 4C

segmento de señal similar a impulso (por ejemplo, habla con voz)

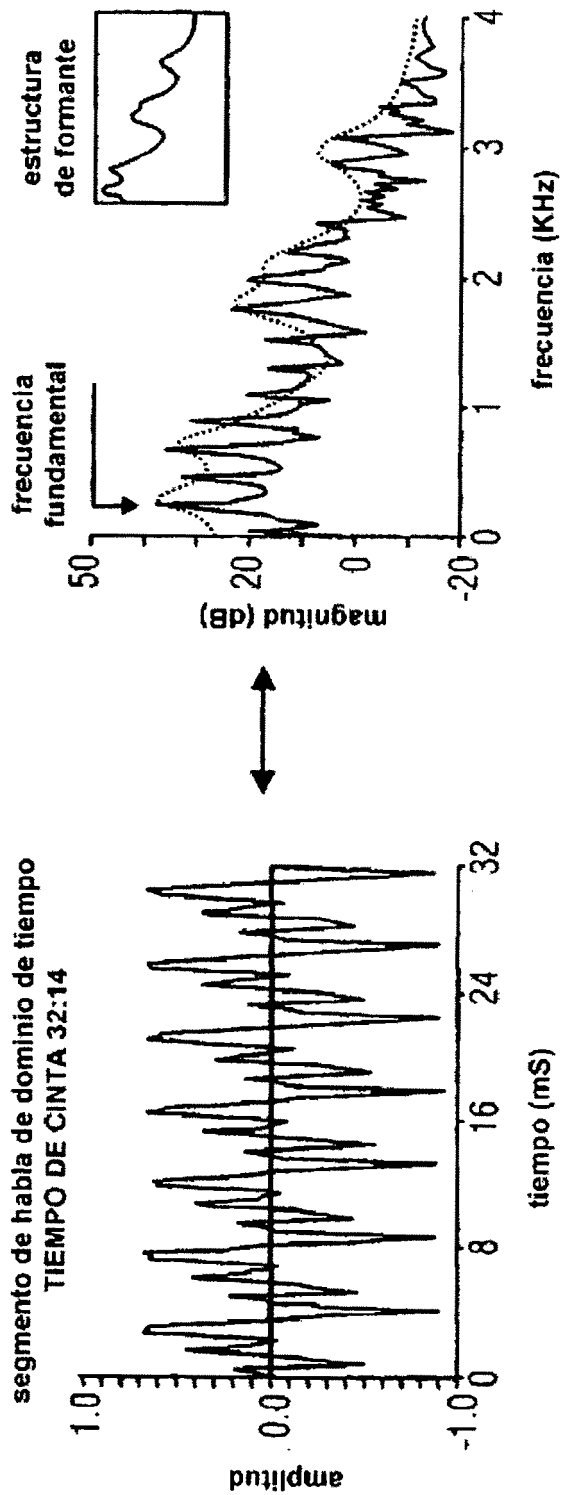


FIG 5A

FIG 5B

segmento estacionario (por ejemplo, habla sin voz)

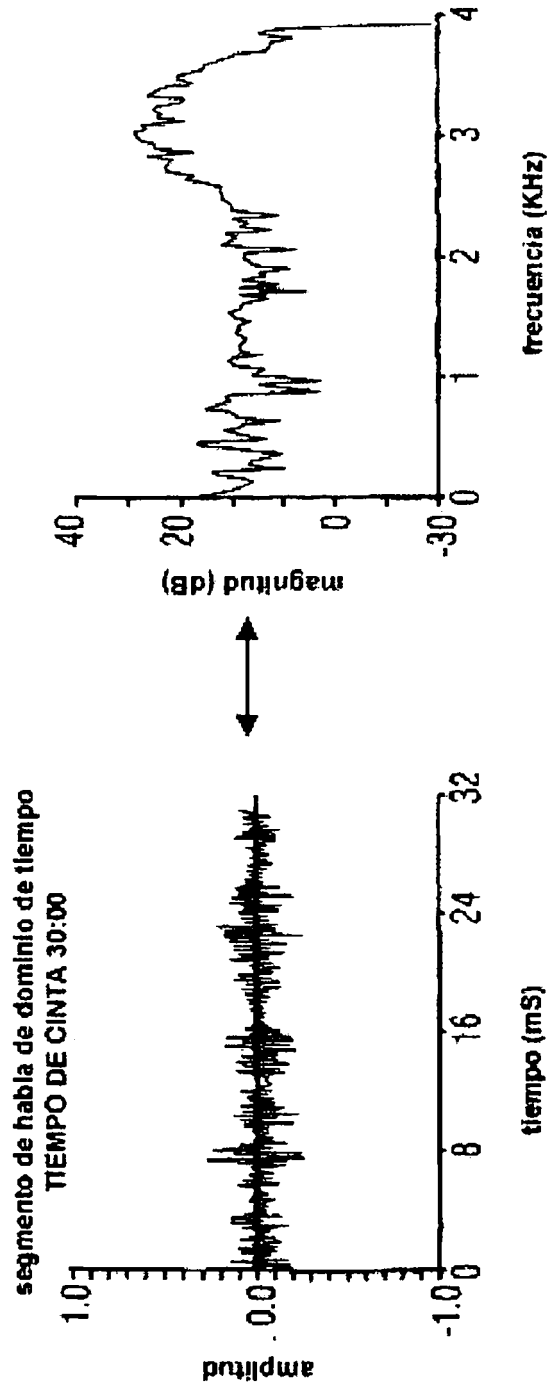
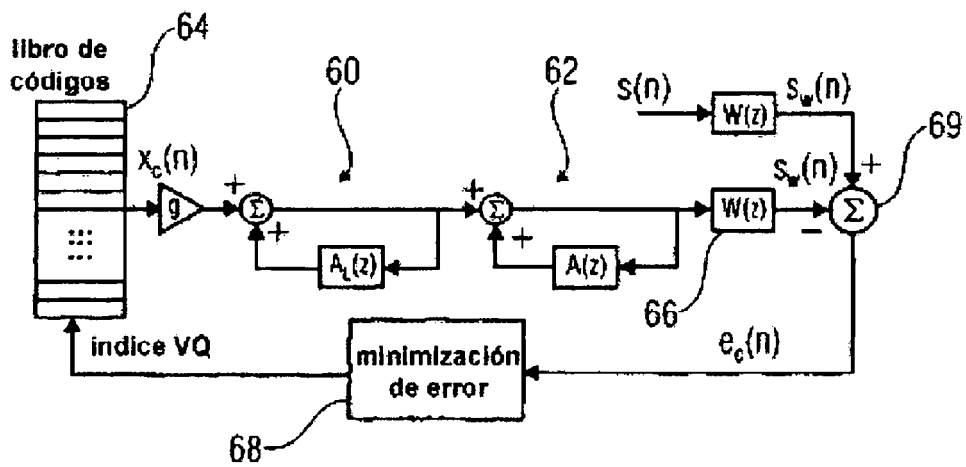


FIG 5C

FIG 5D

CELP de análisis por síntesis



$A_l(z)$: predicción a largo plazo
 $\hat{=}$ estructura (fina) de altura tonal

$A(z)$: predicción a corto plazo
 $\hat{=}$ estructura de formante /
 envolvente espectral

FIG 6

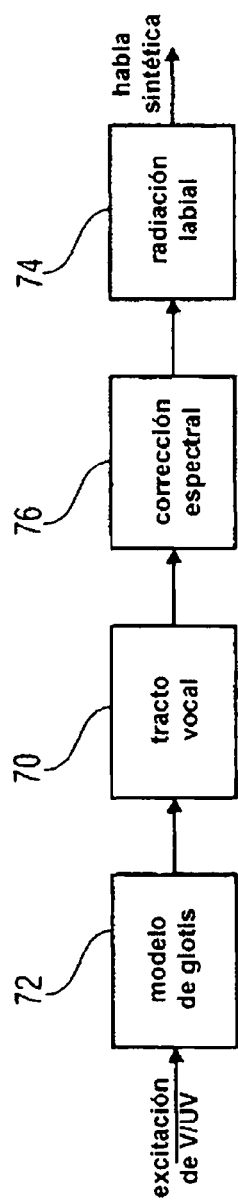


FIG 7A

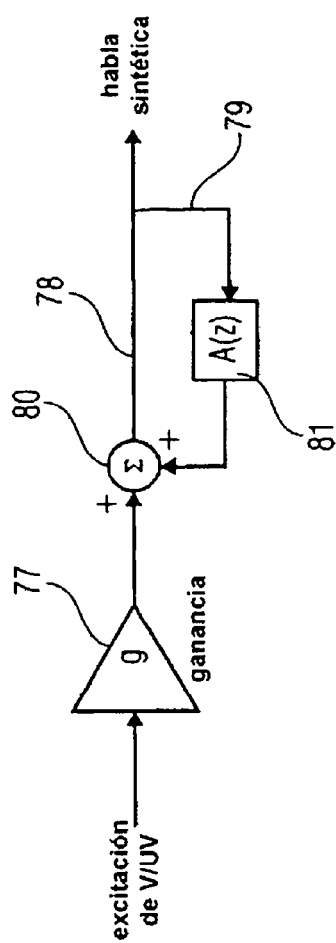


FIG 7B

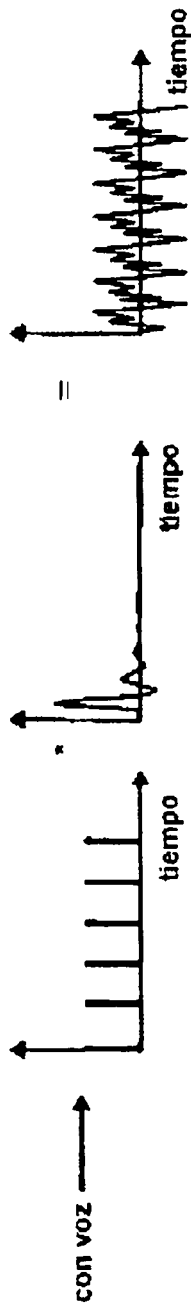


FIG 7C

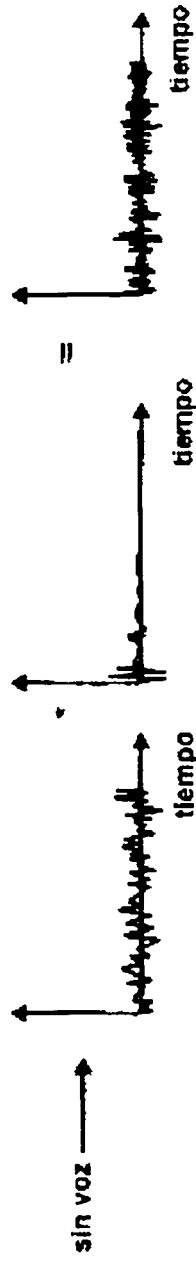


FIG 7D

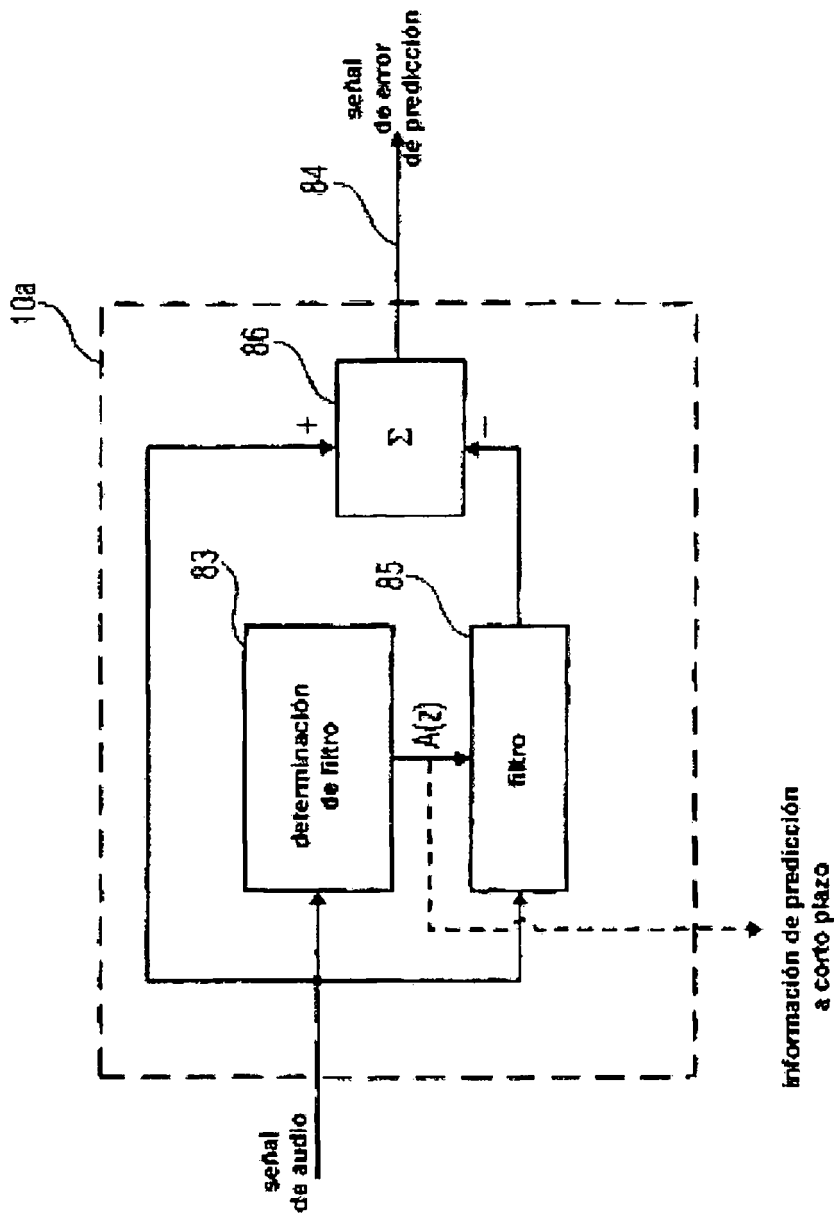


FIG 7E

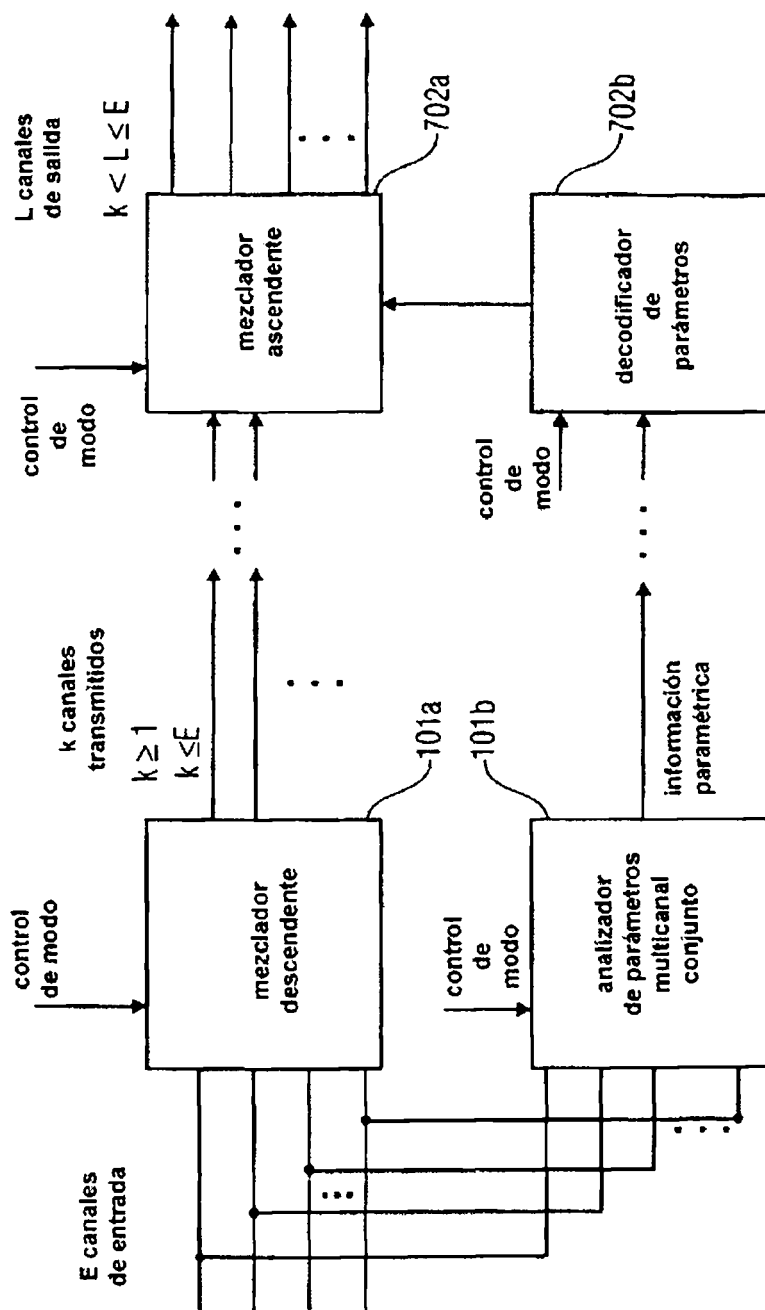


FIG 8

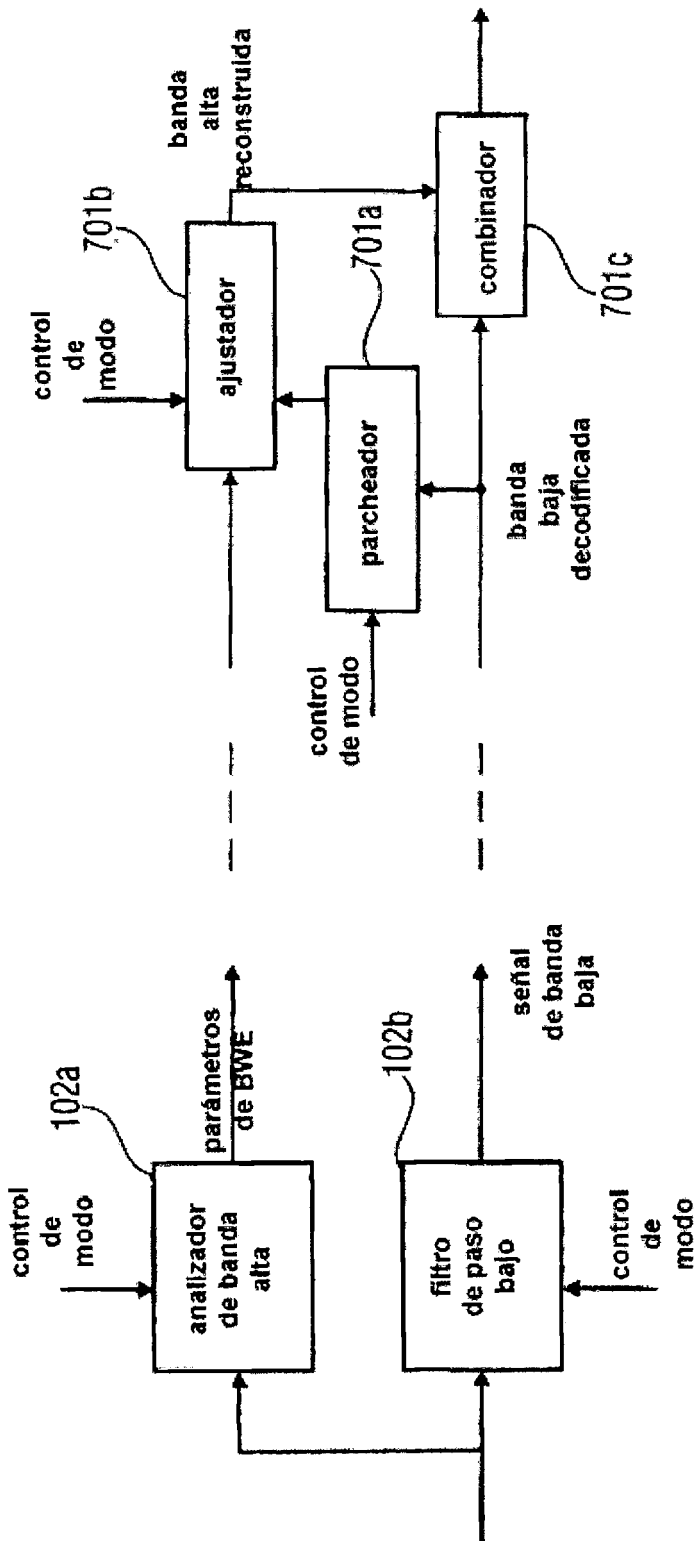


FIG 9

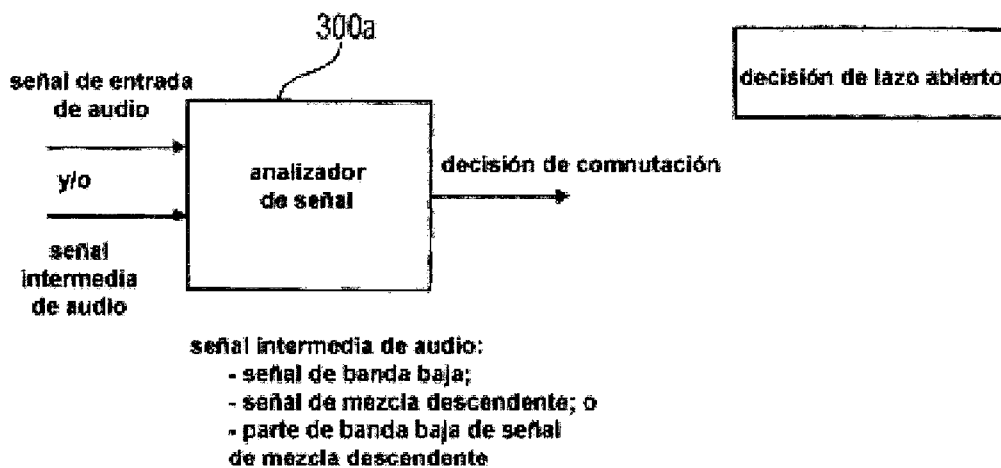


FIG 10A

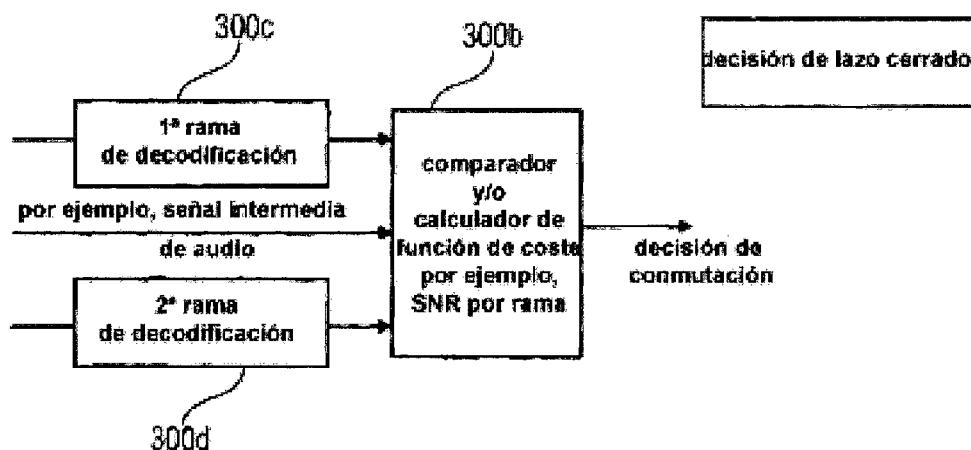


FIG 10B