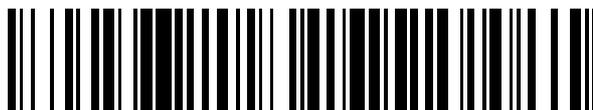


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 666 150**

51 Int. Cl.:

G10L 19/26 (2013.01)

G10L 19/107 (2013.01)

G10L 19/02 (2013.01)

G10L 19/20 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **23.06.2011 E 14164773 (5)**

97 Fecha y número de publicación de la concesión europea: **21.02.2018 EP 2757560**

54 Título: **Descodificación de audio con pos-filtro selectivo**

30 Prioridad:

02.07.2010 US 361237 P

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

03.05.2018

73 Titular/es:

**DOLBY INTERNATIONAL AB (100.0%)
Apollo Building, 3E, Herikerbergweg 1-35
1101 CN Amsterdam, NL**

72 Inventor/es:

**RESCH, BARBARA;
KJOERLING, KRISTOFER y
VILLEMOES, LARS**

74 Agente/Representante:

LEHMANN NOVO, María Isabel

ES 2 666 150 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

Descodificación de audio con pos-filtro selectivo

5 Campo técnico

La presente invención se refiere, en general, a la codificación de audio digital y, más precisamente, a técnicas de codificación para señales de audio que contienen componentes de distintos caracteres.

10 Antecedentes

Una extendida clase de métodos de codificación para señales de audio que contienen habla o canto incluye la predicción lineal excitada por código (CELP) aplicada en la alternación en el tiempo con distintos métodos de codificación, que incluyen métodos de codificación del dominio de la frecuencia, especialmente adaptados para la música, o métodos de naturaleza general, para admitir variaciones en el carácter entre sucesivos periodos de tiempo de la señal de audio. Por ejemplo, un descodificador simplificado de Codificación Unificada de Habla y Audio (USAC; véase la norma ISO / IEC 23003-3) del Grupo de Expertos en Imágenes en Movimiento (MPEG) es operable en al menos tres modalidades de descodificación, la Codificación Avanzada de Audio (AAC; véase la norma ISO / IEC 13818-7), la CELP algebraica (ACELP) y la excitación codificada por transformación (TCX), según se muestra en la parte superior de la figura 2 adjunta.

Las diversas realizaciones de la CELP están adaptadas a las propiedades de los órganos humanos del habla y, posiblemente, al sentido humano del oído. Según se usa en esta solicitud, la CELP se referirá a todas las posibles realizaciones y variantes, incluyendo, pero sin limitarse a, la ACELP, la CELP de banda ancha y estrecha, la SB-CELP (CELP de sub-banda), la CELP de baja y alta velocidad, la RCELP (CELP relajada), la LD-CELP (CELP de bajo retardo), la CS-CELP (CELP de estructura conjugada), la CS-ACELP (ACELP de estructura conjugada), la PSI-CELP (CELP de innovación sincrónica en tono) y la VSELP (predicción lineal excitada por suma vectorial). Los principios de la CELP son expuestos por R. Schroeder y S. Atal en los Anales de la Conferencia Internacional del IEEE sobre Acústica, Habla y Procesamiento de Señales (ICASSP), vol. 10, págs. 937 a 940, 1985, y algunas de sus aplicaciones están descritas en las referencias 25 a 29 citadas en la publicación de Chen y Gersho en las Transacciones del IEEE sobre Habla y Procesamiento de Audio, vol. 3, nº 1, de 1995. Según lo adicionalmente detallado en el artículo anterior, un descodificador de CELP (o, análogamente, un sintetizador de habla de CELP) puede incluir un predictor de tono, que restaura el componente periódico de una señal de habla codificada, y un libro de códigos de pulsos, desde el cual se añade una secuencia de innovación. El predictor de tono puede incluir, a su vez, un predictor de retardo largo para restaurar el tono y un predictor de retardo corto para restaurar las regiones formantes por medio de la modelación del envolvente espectral. En este contexto, el tono se concibe, en general, como la frecuencia fundamental del componente sonoro tonal producido por las cuerdas vocales, y adicionalmente realizado por las partes resonantes del conducto vocal. Esta frecuencia, junto con sus armónicos, dominará el habla o el canto. En términos generales, los métodos de CELP son los más adecuados para el procesamiento del canto solo o unipersonal, para el cual la frecuencia del tono está bien definida y es relativamente fácil de determinar.

Para mejorar la calidad percibida del habla codificada por CELP, es práctica común combinarla con el pos-filtrado (o realce tonal, en otras palabras). La Patente Estadounidense Nº 4.969.192 y la sección II del artículo de Chen y Gersho revelan propiedades deseables de tales pos-filtros, esto es, su capacidad de suprimir los componentes de ruido situados entre los armónicos del tono vocal detectado (parte a largo plazo, véase la sección IV). Se cree que una parte importante de este ruido emana del modelado del envolvente espectral. La parte a largo plazo de un pos-filtro sencillo puede ser diseñada para que tenga la siguiente función de transferencia:

$$H_E(z) = 1 + \alpha \left(\frac{z^T + z^{-T}}{2} - 1 \right),$$

donde T es un periodo tonal estimado en términos del número de muestras y α es una ganancia del pos-filtro, según se muestra en las figuras 1 y 2. De manera similar a un filtro peine, un filtro de ese tipo atenúa las frecuencias $1/(2T)$, $3/(2T)$, $5/(2T)$, ..., que están situadas a mitad de camino entre los armónicos de la frecuencia tonal y las frecuencias adyacentes. La atenuación depende del valor de la ganancia α . Los pos-filtros levemente más sofisticados aplican esta atenuación solamente a las frecuencias bajas – de aquí, el término habitualmente usado de pos-filtro bajo – donde el ruido es más perceptible. Esto puede ser expresado aplicando en cascada la función de transferencia H_E descrita anteriormente y un filtro H_{LP} de paso bajo. Así, el valor descodificado pos-procesado de S_E proporcionado por el pos-filtro estará dado, en el dominio de transformación, por

$$S_E(z) = S(z) - \alpha S(z) P_{LT}(z) H_{LP}(z),$$

donde

$$P_{LT}(z) = 1 - \frac{z^T + z^{-T}}{2}$$

5 y S es la señal descodificada que es suministrada como entrada al pos-filtro. La Figura 3 muestra una realización de un pos-filtro con estas características, que es adicionalmente expuesto en la sección 6.1.3 de la Especificación Técnica ETSI TS 126 290, versión 6.3.0, edición 6. Como sugiere esta figura, la información tonal está codificada como un parámetro en la señal del flujo de bits y es extraída por un módulo de rastreo tonal, conectado comunicativamente con el filtro de predicción a largo plazo que lleva a cabo las operaciones expresadas por P_{LT} .

10 La parte a largo plazo descrita en el párrafo anterior puede ser usada sola. Alternativamente, se dispone en serie con un filtro modelador del ruido que preserva los componentes en los intervalos de frecuencia correspondientes a las regiones formantes y que atenúa el ruido en otras regiones espectrales (parte a corto plazo; véase la sección III), es decir, en los 'valles espectrales' de la envolvente de región formante. Como otra posible variación, este grupo de filtros está adicionalmente suplementado por un filtro de tipo de paso alto, para reducir un deterioro percibido, debido al declive espectral de la parte a corto plazo.

15 El documento M. Neuendorf (ed.): WD7 de USAC, 92^{ésima} Reunión del MPEG, Dresde, N11299, y la Solicitud Internacional publicada como WO 99/38144 A1 revelan sistemas de procesamiento de audio con la aplicación selectiva del pos-filtrado.

20 Las señales de audio que contienen una mezcla de componentes de distintos orígenes – por ejemplo, tonal, no tonal, vocal, instrumental, no musical – no siempre son reproducidas por las tecnologías disponibles de codificación digital de manera satisfactoria. Más precisamente, ha sido observado que las tecnologías disponibles son deficientes en la manipulación de tal material de audio no homogéneo, favoreciendo generalmente a uno de los componentes en perjuicio del otro. En particular, la música que contiene canto acompañado por uno o más instrumentos o partes corales, que ha sido codificada por métodos de la naturaleza descrita anteriormente, será a menudo descodificada con distorsiones perceptibles que arruinan parte de la experiencia auditiva.

30 **Resumen de la invención**

A fin de mitigar al menos algunos de los inconvenientes esbozados en la sección anterior, es un objeto proporcionar métodos y dispositivos adaptados para la codificación y descodificación de audio de las señales que contienen una mezcla de componentes de distintos orígenes. Como objetos específicos, la invención busca proporcionar aquellos métodos y dispositivos que sean adecuados desde el punto de vista de la eficacia de codificación o la fidelidad (percibida) de reproducción, o ambas.

La invención logra al menos uno de estos objetos proporcionando un sistema descodificador, un método de descodificación y un producto de programa de ordenador para llevar a cabo el método, según lo definido en las reivindicaciones independientes. Las reivindicaciones dependientes definen realizaciones de la invención.

40 Los inventores han percibido que algunas distorsiones percibidas en las señales de audio descodificadas de origen no homogéneo provienen de una conmutación inadecuada entre varias modalidades de codificación, de las cuales al menos una incluye el pos-filtrado en el descodificador y al menos una no lo hace. Más precisamente, los pos-filtros disponibles eliminan no solamente el ruido inter-armónico (y, donde corresponda, el ruido en los valles espectrales), sino también los componentes de señales que representan acompañamiento instrumental o vocal, y otro material de naturaleza 'deseable'. El hecho de que la diferencia apenas perceptible en los valles espectrales puede ser tan grande como de 10 dB (según lo observado por Ghitza y Goldstein, Trans. del IEEE sobre Acústica, Habla y Procesamiento de Señales, vol. ASSP-4, págs. 697 a 708, de 1986) puede haber sido tomado como justificación por muchos diseñadores para filtrar con severidad estas bandas de frecuencia. La degradación de la calidad por la misma atenuación inter-armónica (y de valle espectral) puede, sin embargo, ser menos importante que la de las ocasiones de conmutación. Cuando el pos-filtro está activado, el fondo de una voz cantante suena repentinamente atenuado y, cuando el filtro es desactivado, el fondo se hace instantáneamente más sonoro. Si la conmutación tiene lugar frecuentemente, debido a la naturaleza de la señal de audio o a la configuración del dispositivo de codificación, habrá una distorsión de conmutación. Como ejemplo, un descodificador de USAC puede ser operable tanto en una modalidad de ACELP combinada con el pos-filtrado como en una modalidad de TCX sin pos-filtrado. La modalidad de ACELP se usa en episodios donde está presente un componente vocal dominante. De tal modo, la conmutación a la modalidad de ACELP puede ser activada por el inicio del canto, tal como al principio de una nueva frase musical, al comienzo de un nuevo verso, o simplemente después de un episodio donde se considera que el acompañamiento sofoca la voz cantante, en el sentido de que el componente vocal ya no es prominente. Los experimentos han confirmado que una solución alternativa, o más bien una elusión del problema, por la cual se usa la codificación TCX en toda su extensión (y la modalidad ACELP es inhabilitada), no remedia el problema, ya que aparecen distorsiones como de reverberación.

Se hace notar que los métodos y aparatos revelados en esta sección pueden ser aplicados, después de modificaciones adecuadas dentro de las capacidades del experto, que incluyen la experimentación rutinaria, a la codificación de señales con varios componentes, posiblemente correspondientes a distintos canales, tales como canales estéreo. A lo largo de la presente solicitud, el realce tonal y el pos-filtrado son usados como sinónimos. Se hace notar además que la AAC está expuesta como un ejemplo representativo de métodos de codificación del dominio de la frecuencia. En efecto, la aplicación de la invención a un descodificador o codificador operable en una modalidad de codificación del dominio de frecuencia, distinta a la AAC, solamente requerirá pequeñas modificaciones, si acaso, dentro de las capacidades del experto. De manera similar, TCX se menciona como un ejemplo de codificación por transformación de predicción lineal ponderada y de codificación por transformación en general.

Las características de dos o más realizaciones descritas anteriormente en la presente memoria pueden ser combinadas, a menos que sean claramente complementarias, en realizaciones adicionales. El hecho de que dos características sean mencionadas en distintas reivindicaciones no excluye que puedan ser combinadas con ventaja. Análogamente, también pueden ser proporcionadas realizaciones adicionales por la omisión de ciertas características que no sean necesarias o no sean esenciales para el propósito deseado.

Breve descripción de los dibujos

Se describirán ahora realizaciones de la presente invención con referencia a los dibujos adjuntos, en los cuales:

la figura 1 es un diagrama de bloques que muestra un descodificador convencional con pos-filtro;

la figura 2 es un diagrama de bloques esquemático de un descodificador convencional operable en modalidad de AAC, ACELP y TCX, y que incluye un pos-filtro permanentemente conectado aguas abajo del módulo de ACELP;

la figura 3 es un diagrama de bloques que ilustra la estructura de un pos-filtro;

las figuras 4 y 5 son diagramas de bloques de dos descodificadores de acuerdo a la invención;

las figuras 6 y 7 son diagramas de bloques que ilustran diferencias entre un descodificador convencional (figura 6) y un descodificador (figura 7) de acuerdo a la invención;

la figura 8 es un diagrama de bloques de un codificador;

las figuras 9 y 10 son diagramas de bloques que ilustran diferencias entre un descodificador convencional (figura 9) y un descodificador (figura 10) de acuerdo a la invención; y

la figura 11 es un diagrama de bloques de un pos-filtro autónomo que puede ser activado y desactivado selectivamente.

Descripción detallada de realizaciones

La figura 4 es un dibujo esquemático de un sistema descodificador 400 de acuerdo a una realización de la invención, que tiene como su entrada una señal de flujo de bits y como su salida una señal de audio. Como en los descodificadores convencionales mostrados en la figura 1, un pos-filtro 440 está dispuesto aguas abajo de un módulo descodificador 410, pero puede ser incorporado a, o apartado de, el trayecto de descodificación, operando un conmutador 442. El pos-filtro está habilitado en la posición de conmutación mostrada en la figura. Sería inhabilitado si el conmutador se fijara en la posición opuesta, por lo cual la señal del módulo descodificador 410, en cambio, sería conducida por la línea 444 de omisión. Como una contribución inventiva, el conmutador 442 es controlable por la información de pos-filtrado contenida en la señal del flujo de bits, de modo que el pos-filtrado pueda ser aplicado y eliminado, independientemente del estado actual del módulo descodificador 410. Debido a que un pos-filtro 440 funciona con algún retardo – por ejemplo, el pos-filtro mostrado en la figura 3 introducirá un retardo equivalente al menos al periodo tonal T – un módulo 443 de retardo de compensación está dispuesto en la línea 444 de omisión para mantener los módulos en una condición sincronizada en la conmutación. El módulo 443 de retardo retarda la señal en el mismo periodo que lo haría el pos-filtro 440, pero no procesa de otro modo la señal. Para minimizar el tiempo de traspaso, el módulo 443 de retardo de compensación recibe la misma señal que el pos-filtro 440 en todo momento. En una realización alternativa, donde el pos-filtro 440 es reemplazado por un pos-filtro de retardo cero (por ejemplo, un filtro causal, tal como un filtro con dos tomas, independiente de futuros valores de señal), el módulo 443 de retardo de compensación puede ser omitido.

La figura 5 ilustra un desarrollo adicional de acuerdo a las revelaciones de la invención del sistema descodificador 500 de triple modalidad de la figura 2. Un módulo descodificador 511 de ACELP está dispuesto en paralelo con un módulo descodificador 512 de TCX y un módulo descodificador 513 de AAC. En serie con el módulo descodificador

511 de ACELP está dispuesto un pos-filtro 540 para atenuar el ruido, en particular, el ruido situado entre armónicos de una frecuencia tonal, directamente o indirectamente obtenible de la señal del flujo de bits para el cual está adaptado el sistema descodificador 500. La señal del flujo de bits también codifica información de pos-filtrado que gobierna las posiciones del conmutador superior 541 operable para conmutar el pos-filtro 540 fuera del trayecto de procesamiento y reemplazarlo por un retardo 543 de compensación como en la figura 4. Un conmutador inferior 542 se usa para conmutar entre distintas modalidades de descodificación. Con esta estructura, la posición del conmutador superior 541 es irrelevante cuando se usa uno de los módulos 512, 513 de TCX o AAC; por tanto, la información de pos-filtrado no necesariamente indica esta posición, excepto en la modalidad de ACELP. Cualquiera que sea la modalidad de descodificación que se está usando actualmente, la señal es suministrada desde el punto de conexión aguas abajo del conmutador inferior 542 hasta un módulo 550 de copia de banda espectral (SBR), que emite una señal de audio. El experto se dará cuenta de que el dibujo es de naturaleza conceptual, como queda notablemente claro a partir de los conmutadores, que se muestran esquemáticamente como entidades físicas distintas con medios móviles de contacto. En una posible implementación realista del sistema descodificador, los conmutadores, así como los otros módulos, estarán realizados por instrucciones legibles por ordenador.

Las figuras 6 y 7 también son diagramas de bloques de dos sistemas descodificadores de triple modalidad, operables en una modalidad de descodificación ACELP, TCX o del dominio de frecuencia. Con referencia a la última figura, que muestra una realización de la invención, una señal del flujo de bits es suministrada a un punto 701 de entrada, que está a su vez permanentemente conectado, mediante los ramales respectivos, con los tres módulos descodificadores 711, 712, 713. El punto 701 de entrada también tiene un ramal conector 702 (no presente en el sistema descodificador convencional de la figura 6) con un módulo 740 de realce tonal, que actúa como un pos-filtro del tipo general descrito anteriormente. Como es práctica común en la técnica, un primer módulo 703 de ventanas de transición está dispuesto aguas abajo de los módulos 711, 712 de ACELP y TCX, para llevar a cabo las transiciones entre los módulos descodificadores. Un segundo módulo 704 de transición está dispuesto aguas abajo del módulo descodificador 713 del dominio de frecuencia y del primer módulo 703 de ventanas de transición, para llevar a cabo la transición entre las dos super-modalidades. Además, un módulo 750 de SBR está proporcionado inmediatamente aguas arriba del punto 705 de salida. Es claro que la señal de flujo de bits es suministrada directamente (o después del demultiplexado, según corresponda) a los tres módulos descodificadores 711, 712, 713 y al módulo 740 de realce tonal. La información contenida en el flujo de bits controla qué módulo de descodificación ha de estar activo. Por parte de la invención, sin embargo, el módulo 740 de realce tonal realiza una análoga auto-activación que, en respuesta a información de pos-filtrado en el flujo de bits, puede actuar como un pos-filtro o, simplemente, como un dispositivo de traspasamiento. Esto, por ejemplo, puede ser realizado mediante la provisión de una sección de control (no mostrada) en el módulo 740 de realce tonal, por medio del cual la acción de pos-filtrado puede ser activada o desactivada. El módulo 740 de realce tonal está siempre en su modalidad de traspasamiento cuando el sistema descodificador funciona en el dominio de frecuencia o la modalidad de descodificación TCX, en donde, en rigor, no es necesaria ninguna información de pos-filtrado. Se entiende que los módulos que no forman parte de la contribución inventiva, y cuya presencia es obvia para el experto, por ejemplo, un demultiplexador, han sido omitidos de la figura 7 y de otros dibujos similares, para aumentar la claridad.

Como variación, el sistema descodificador de la figura 7 puede estar equipado con un módulo de control (no mostrado) para decidir si ha de aplicarse o no el pos-filtrado, usando un enfoque de análisis por síntesis. Tal módulo de control está comunicativamente conectado con el módulo 740 de realce tonal y con el módulo 711 de ACELP, desde el cual extrae una señal descodificada intermedia $s_{i_DEC}(n)$ que representa una etapa intermedia en el proceso de descodificación, preferiblemente, una correspondiente a la excitación de la señal. El módulo de detección tiene la información necesaria para simular la acción del módulo 740 de realce tonal, según lo definido por las funciones de transferencia $P_{LT}(z)$ y $H_{LP}(z)$ (véase la sección Antecedentes y la figura 3) o, equivalentemente, sus respuestas de impulso de filtro $p_{LT}(z)$ y $h_{LP}(n)$. Como se desprende de la exposición en la sección Antecedentes, el componente a restar en el pos-filtrado puede ser estimado por una señal de diferencia aproximada $s_{AD}(n)$ que es proporcional a $[(s_{i_DEC} * P_{LT}) * h_{LP}](n)$, donde * indica la convolución discreta. Esto es una aproximación de la verdadera diferencia entre la señal de audio original y la señal descodificada pos-filtrada, a saber,

$$s_{ORIG}(n) - s_E(n) = s_{ORIG}(n) - (s_{DEC}(n) - \alpha [s_{DEC} * p_{LT} * h_{LP}](n)),$$

donde α es la ganancia del pos-filtro. Estudiando la energía total, la energía de banda baja, la tonalidad, el espectro efectivo de magnitudes o los espectros de magnitudes pasadas de la señal, según lo revelado en la sección Resumen y las reivindicaciones, la sección de control puede hallar una base para la decisión de activar o desactivar el módulo 740 de realce tonal.

La figura 8 muestra un sistema codificador 800. El sistema codificador 800 está adaptado para procesar señales de audio digital, que son generalmente obtenidas capturando una onda sonora por parte de un micrófono, y transduciendo la onda hacia una señal eléctrica analógica. La señal eléctrica es luego muestreada en una señal digital susceptible de ser suministrada, en un formato adecuado, al sistema codificador 800. El sistema consiste, en general, en un módulo codificador 810, un módulo 820 de decisión y un multiplexador 830. En virtud de los conmutadores 814, 815 (simbólicamente representados), el módulo codificador 810 es operable en una modalidad CELP, TCX o bien AAC, activando selectivamente los módulos 811, 812, 813. El módulo 820 de decisión aplica uno

o más criterios predefinidos para decidir si se inhabilita o no el pos-filtrado durante la descodificación de una señal de flujo de bits producida por el sistema codificador 800 para codificar una señal de audio. Con este fin, el módulo 820 de decisión puede examinar la señal de audio directamente o puede recibir datos desde el módulo codificador 810 mediante una línea 816 de conexión. Una señal indicativa de la decisión adoptada por el módulo 820 de
5 decisión es suministrada, junto con la señal de audio codificada proveniente del módulo codificador 810, a un multiplexador 830, que concatena las señales en un flujo de bits que constituye la salida del sistema codificador 800.

Preferiblemente, el módulo 820 de decisión basa su decisión sobre una señal de diferencia aproximada calculada a partir de una señal descodificada intermedia s_{i_DEC} , que puede ser restada del módulo codificador 810. La señal
10 descodificada intermedia representa una etapa intermedia en el proceso de descodificación, según lo expuesto en los párrafos precedentes, pero puede ser extraída desde una etapa correspondiente del proceso de codificación. Sin embargo, en el sistema codificador 800 está disponible la señal de audio original s_{ORIG} , de modo que, ventajosamente, la señal de diferencia aproximada se forme como:

$$s_{ORIG}(n) - (s_{i_DEC}(n) - \alpha [(s_{i_DEC} * p_{LT}) * h_{LP}](n)).$$

La aproximación reside en el hecho de que la señal descodificada intermedia se usa en lugar de la señal descodificada final. Esto permite una valoración de la naturaleza del componente que un pos-filtro eliminaría en la
20 descodificación y, aplicando uno de los criterios expuestos en la sección Resumen, el módulo 820 de decisión podrá adoptar una decisión en cuanto a inhabilitar o no el pos-filtrado.

Como una variación de esto, el módulo 820 de decisión puede usar la señal original en lugar de una señal descodificada intermedia, de modo que la señal de diferencia aproximada sea $[(s_{i_DEC} * p_{LT}) * h_{LP}](n)$. Es probable
25 que esto sea una aproximación menos fiel pero, por otra parte, hace optativa la presencia de una línea 816 de conexión entre el módulo 820 de decisión y el módulo codificador 810.

En otras variaciones de ese tipo de esta realización, donde el módulo 820 de decisión estudia directamente la señal de audio, pueden aplicarse uno o más de los siguientes criterios:

- 30 • ¿Contiene la señal de audio tanto un componente con frecuencia fundamental dominante como un componente situado por debajo de la frecuencia fundamental? (La frecuencia fundamental puede ser suministrada como un sub-producto del módulo codificador 810).
- 35 • ¿Contiene la señal de audio tanto un componente con frecuencia fundamental dominante como un componente situado entre los armónicos de la frecuencia fundamental?
- ¿Contiene la señal de audio energía significativa de señal por debajo de la frecuencia fundamental?
- 40 • ¿Es (probable que sea) preferible la descodificación pos-filtrada a la descodificación no filtrada con respecto a la optimalidad de velocidad-distorsión?

En todas las variaciones descritas de la estructura codificadora mostrada en la figura 8 – es decir, independientemente de la base del criterio de detección – la sección 820 de decisión puede ser habilitada para
45 decidir sobre un avance gradual o una eliminación gradual del pos-filtrado, a fin de lograr transiciones suaves. El avance y la eliminación graduales pueden ser controlados ajustando la ganancia de pos-filtro.

La figura 9 muestra un descodificador convencional operable en una modalidad de descodificación de frecuencia y en una modalidad de descodificación de CELP, según la señal de flujo de bits suministrada al descodificador. El pos-filtrado se aplica toda vez que se selecciona la modalidad de descodificación de CELP. Una mejora de este
50 descodificador se ilustra en la figura 10, que muestra un descodificador 1000 de acuerdo a una realización de la invención. Este descodificador es operable no solamente en una modalidad de descodificación basada en el dominio de frecuencia, en la cual está activo el módulo descodificador 1013 del dominio de frecuencia, y en una modalidad de descodificación de CELP filtrada, en la cual están activos el módulo descodificador 1011 de CELP y el pos-filtro 1040, sino también en una modalidad de CELP no filtrada, en la cual el módulo 1011 de CELP suministra su señal a
55 un módulo 1043 de retardo de compensación, mediante una línea 1044 de omisión. Un conmutador 1042 controla qué modalidad de descodificación se usa actualmente, en respuesta a la información de pos-filtrado contenida en la señal del flujo de bits proporcionada al descodificador 1000. En este descodificador y en el de la figura 9, la última etapa de procesamiento es efectuada por un módulo 1050 de SBR, desde el cual es emitida la señal de audio final.

La figura 11 muestra un pos-filtro 1100 adecuado para ser dispuesto aguas abajo de un descodificador 1199. El filtro 1100 incluye un módulo 1140 de pos-filtrado, que es habilitado o inhabilitado por un módulo de control (no
60 mostrado), notablemente, un controlador de ganancia binario o no binario, en respuesta a una señal de pos-filtrado recibida desde un módulo 1120 de decisión dentro del pos-filtro 1100. El módulo de decisión realiza una o más pruebas sobre la señal obtenida desde el descodificador, para llegar a una decisión en cuanto a si el módulo 1140

- de pos-filtrado ha de estar activo o inactivo. La decisión puede ser tomada según la funcionalidad del módulo 820 de decisión en la figura 8, que usa la señal original y/o una señal descodificada intermedia para predecir la acción del pos-filtro. La decisión del módulo 1120 de decisión también puede basarse en información similar a la que los módulos de decisión usan en esas realizaciones donde se forma una señal descodificada intermedia. Como ejemplo,
- 5 el módulo 1120 de decisión puede estimar una frecuencia de tono (a menos que esto sea inmediatamente extraíble desde la señal del flujo de bits) y calcular el contenido de energía en la señal por debajo de la frecuencia de tono y entre sus armónicos. Si este contenido de energía es significativo, probablemente representa un componente de señal relevante, en lugar de ruido, lo que motiva una decisión para inhabilitar el módulo 1140 de pos-filtrado.
- 10 Una prueba de escucha de 6 personas ha sido llevada a cabo, durante la cual muestras musicales codificadas y descodificadas de acuerdo a la invención fueron comparadas con muestras de referencia que contenían la misma música codificada, aplicando el pos-filtrado de la manera convencional, pero manteniendo sin cambios todos los otros parámetros. Los resultados confirman una mejora percibida de la calidad.
- 15 Realizaciones adicionales de la presente invención devendrán evidentes para una persona experta en la técnica, después de leer la descripción anterior. Incluso aunque la presente descripción y los dibujos revelan realizaciones y ejemplos, la invención no está restringida a estos ejemplos específicos. Pueden hacerse numerosas modificaciones y variaciones sin apartarse del alcance de la presente invención, que está definido por las reivindicaciones adjuntas.
- 20 Los sistemas y métodos revelados anteriormente en la presente memoria pueden ser implementados como software, firmware, hardware o una combinación de los mismos. Ciertos componentes, o todos los componentes, pueden ser implementados como software ejecutado por un procesador de señales digitales o un micro-procesador, o ser implementados como hardware, o como un circuito integrado específico de la aplicación. Tal software puede ser distribuido en medios legibles por ordenador, que pueden comprender medios de almacenamiento de
- 25 ordenadores (o medios no transitorios) y medios de comunicación (o medios transitorios). Como es bien sabido para una persona experta en la técnica, los medios de almacenamiento de ordenadores incluyen medios tanto volátiles como no volátiles, extraíbles como no extraíbles, implementados en cualquier método o tecnología para el almacenamiento de información, tal como instrucciones legibles por ordenador, estructuras de datos, módulos de programa u otros datos. Los medios de almacenamiento de ordenadores incluyen, pero no se limitan a, las
- 30 memorias RAM, ROM, EE-PROM, la memoria flash u otra tecnología de memoria, CD-ROM, discos versátiles digitales (DVD) u otro almacenamiento de disco óptico, casetes magnéticos, cinta magnética, almacenamiento de disco magnético u otros dispositivos de almacenamiento magnético, o cualquier otro medio que pueda ser usado para almacenar la información deseada y que pueda ser objeto de acceso por parte de un ordenador. Además, es bien sabido para el experto que los medios de comunicación realizan habitualmente instrucciones legibles por
- 35 ordenador, estructuras de datos, módulos de programa u otros datos en una señal de datos modulados, tal como una onda portadora u otro mecanismo de transporte, e incluyen cualquier medio de suministro de información.

REIVINDICACIONES

1. Un sistema descodificador (400; 500; 700; 1000) para descodificar una señal de flujo de bits como una señal de tiempo de audio, que incluye:
- 5 una sección (410; 511, 512, 513; 711, 712, 713; 1011, 1013) de descodificación para descodificar una señal de flujo de bits como una señal preliminar de tiempo de audio; y
- 10 un pos-filtro de atenuación de ruido inter-armónico (440; 540; 740; 1040) para filtrar la señal preliminar de tiempo de audio con el fin de obtener una señal de tiempo de audio, en el cual la atenuación de ruido inter-armónico está controlada por una ganancia del pos-filtro, que comprende una sección de control adaptada para inhabilitar de manera selectiva el pos-filtro, fijando la ganancia del pos-filtro a cero, en respuesta únicamente a la información de pos-filtrado codificada en la señal de flujo de bits, siendo la información de pos-filtrado indicativa de una decisión por parte del codificador en cuanto a inhabilitar o no el pos-filtrado, en el cual la señal preliminar de tiempo de audio es emitida como la señal de tiempo de audio.
- 15
2. El sistema descodificador de la reivindicación 1, en el cual el pos-filtrado está adaptado para atenuar solamente aquellos componentes espectrales que estén situados por debajo de una frecuencia predeterminada de corte.
- 20
3. El sistema descodificador de la reivindicación 1, en el cual la sección de descodificación comprende adicionalmente:
- un módulo (511; 711; 1011) descodificador de predicción lineal excitada por código, CELP; y
- un módulo (512; 712) descodificador de excitación codificada por transformación, TCX, para descodificar una señal de flujo de bits como una señal de tiempo de audio,
- 25 estando la sección de control adaptada para operar el sistema descodificador al menos en las siguientes modalidades:
- 30 a) el módulo de TCX está habilitado y el pos-filtro está inhabilitado;
- b) el módulo de CELP y el pos-filtro están habilitados; y
- c) el módulo de CELP está habilitado y el pos-filtro está inhabilitado, en donde la señal preliminar de tiempo de audio y la señal de tiempo de audio coinciden.
- 35
4. El sistema descodificador de la reivindicación 1, incluyendo dicha sección de descodificación un módulo de descodificación del habla, en el cual una frecuencia de tono estimada por una sección de predicción a largo plazo en el codificador se codifica en la señal de flujo de bits, y en el cual el pos-filtro está adaptado para atenuar los componentes espectrales situados entre los armónicos de la frecuencia de tono.
- 40
5. El sistema descodificador de la reivindicación 1, en el cual la señal de flujo de bits contiene una representación de una frecuencia de tono y el pos-filtro está adaptado para atenuar los componentes espectrales situados entre los armónicos de la frecuencia de tono.
6. El sistema descodificador de la reivindicación 4 o la reivindicación 5,
- 45 en el cual el pos-filtro está adaptado para atenuar únicamente dichos componentes espectrales que están situados por debajo de una frecuencia predeterminada de corte, y
- en el cual la sección descodificadora comprende adicionalmente un módulo (513; 713) descodificador de Codificación Avanzada de Audio, AAC, para descodificar una señal de flujo de bits como una señal de tiempo de audio,
- 50 estando la sección de control adaptada para operar el sistema descodificador también en la siguiente modalidad:
- d) el módulo de AAC está habilitado y el pos-filtro está inhabilitado.
- 55
7. El sistema descodificador de la reivindicación 1, en el cual la señal del flujo de bits está segmentada en tramas de tiempo, y la sección de control está adaptada para inhabilitar una trama de tiempo entera o una secuencia de tramas de tiempo enteras.
- 60
8. El sistema descodificador de la reivindicación 7, en el cual y la sección de control está adicionalmente adaptada para recibir, para cada trama de tiempo en un flujo de bits del Grupo de Expertos de Imágenes en Movimiento, MPEG, un campo de datos asociado a esta trama de tiempo, y es operable, en respuesta al valor del campo de datos, para inhabilitar el pos-filtro.

9. El sistema descodificador de la reivindicación 1, en el cual la sección de control está adaptada para disminuir y/o aumentar gradualmente la ganancia del pos-filtro.

5 10. Un método de descodificación de una señal de flujo de bits como una señal de tiempo de audio, que incluye las etapas de:

descodificar una señal de flujo de bits como una señal preliminar de tiempo de audio; y

10 pos-filtrar la señal preliminar de tiempo de audio atenuando el ruido inter-armónico, obteniendo por ello una señal de tiempo de audio, en el cual la atenuación del ruido inter-armónico está controlada por una ganancia del pos-filtro,

en el cual el paso de pos-filtrado se omite de manera selectiva, fijando la ganancia del pos-filtro a cero, en respuesta únicamente a la información de pos-filtrado codificada en la señal de flujo de bits, siendo la información de pos-filtrado indicativa de una decisión por parte del codificador en cuanto a inhabilitar o no el pos-filtrado.

15 11. Un producto de programa de ordenador que incluye un portador de datos que almacena instrucciones adaptado para realizar el método de la reivindicación 10.

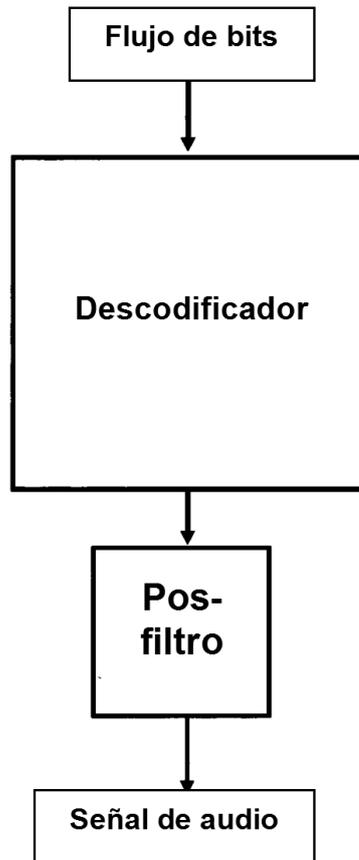


Fig. 1
(técnica anterior)

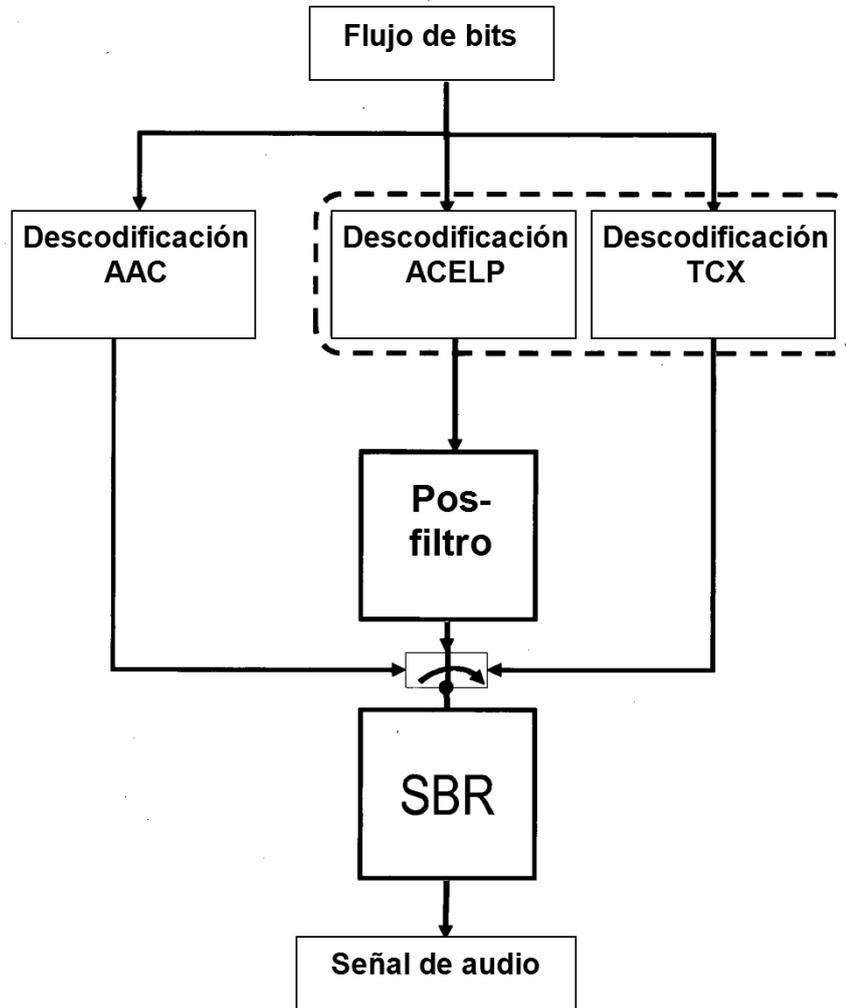


Fig. 2
(técnica anterior)

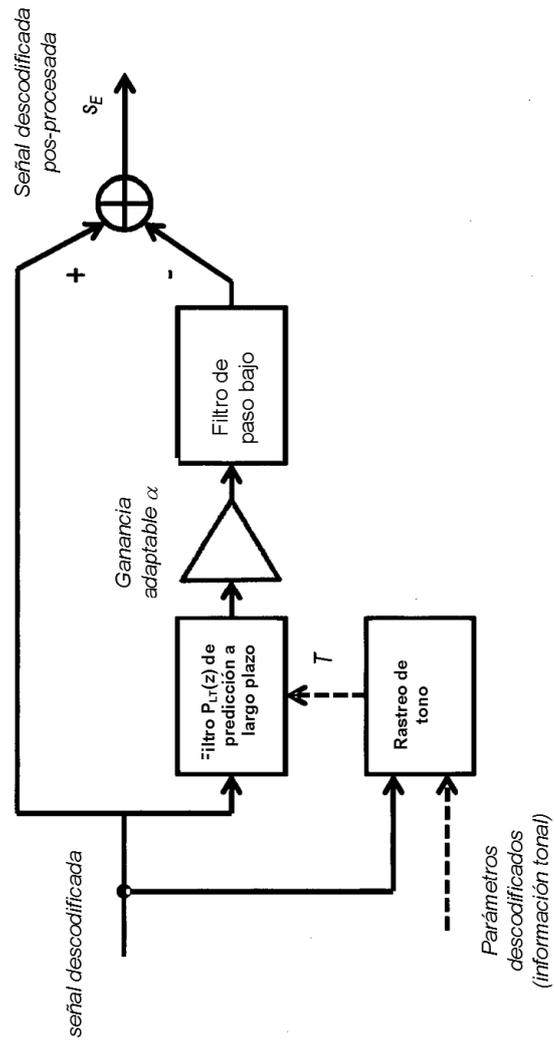


Fig. 3

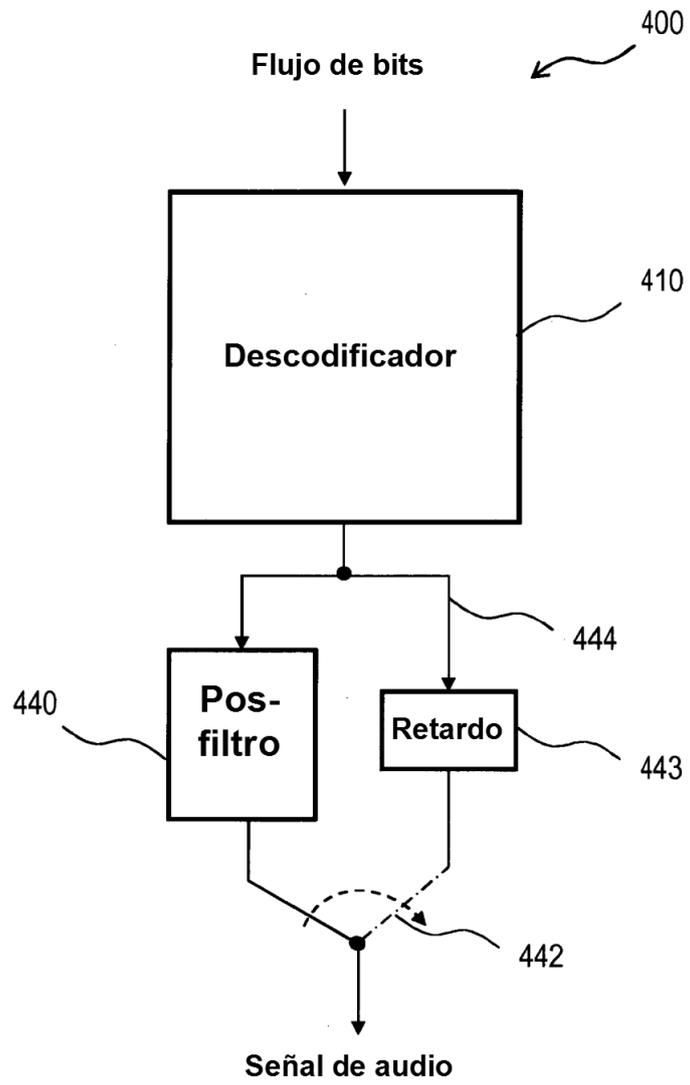


Fig. 4

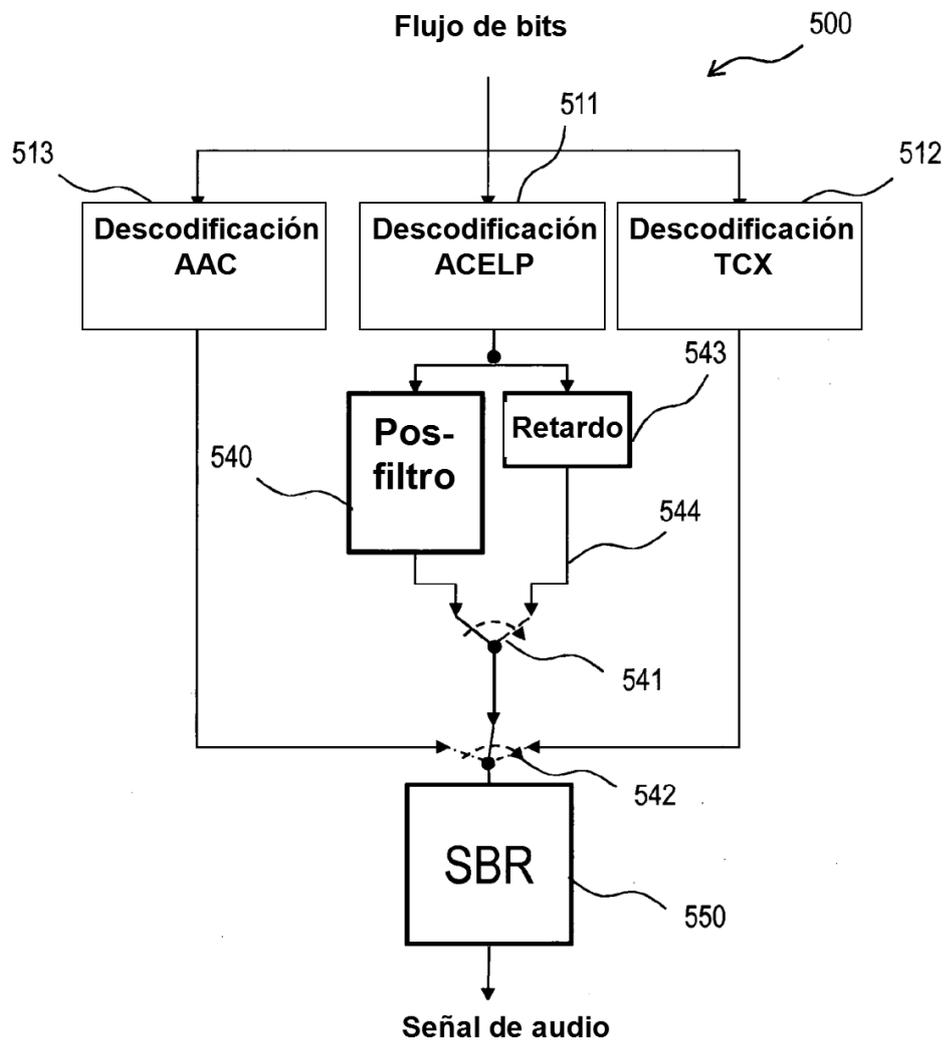


Fig. 5

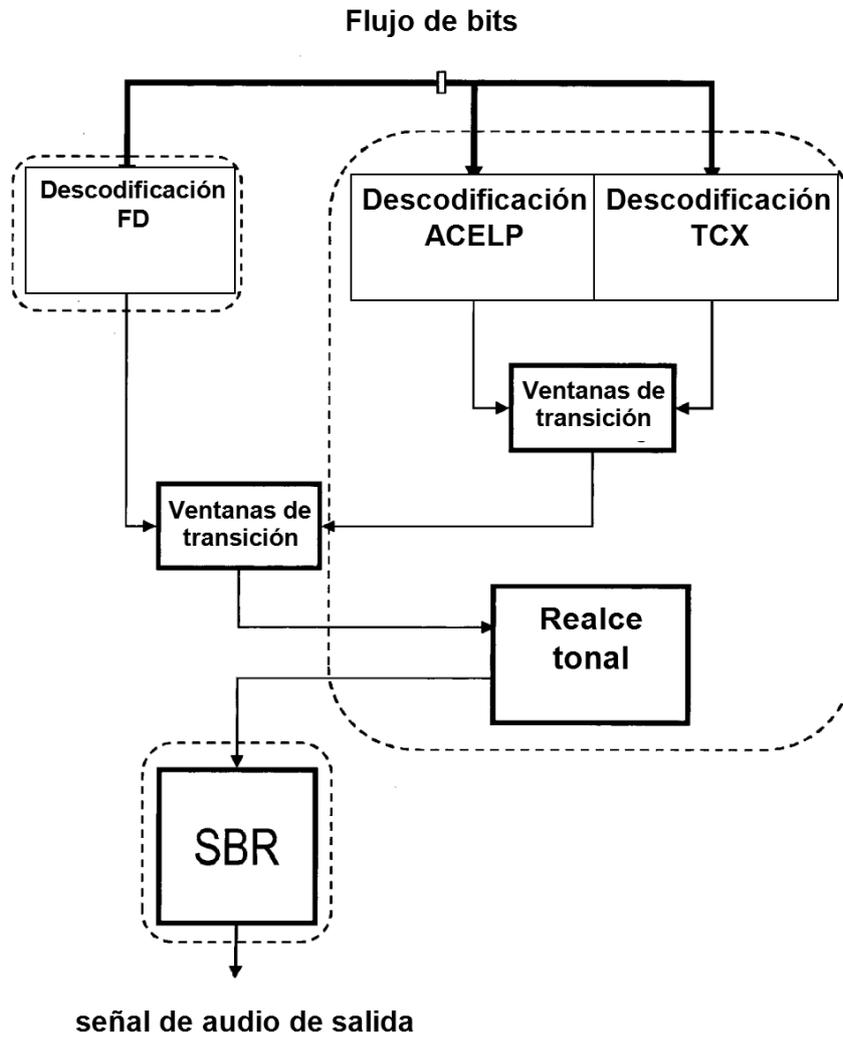


Fig. 6
(técnica anterior)

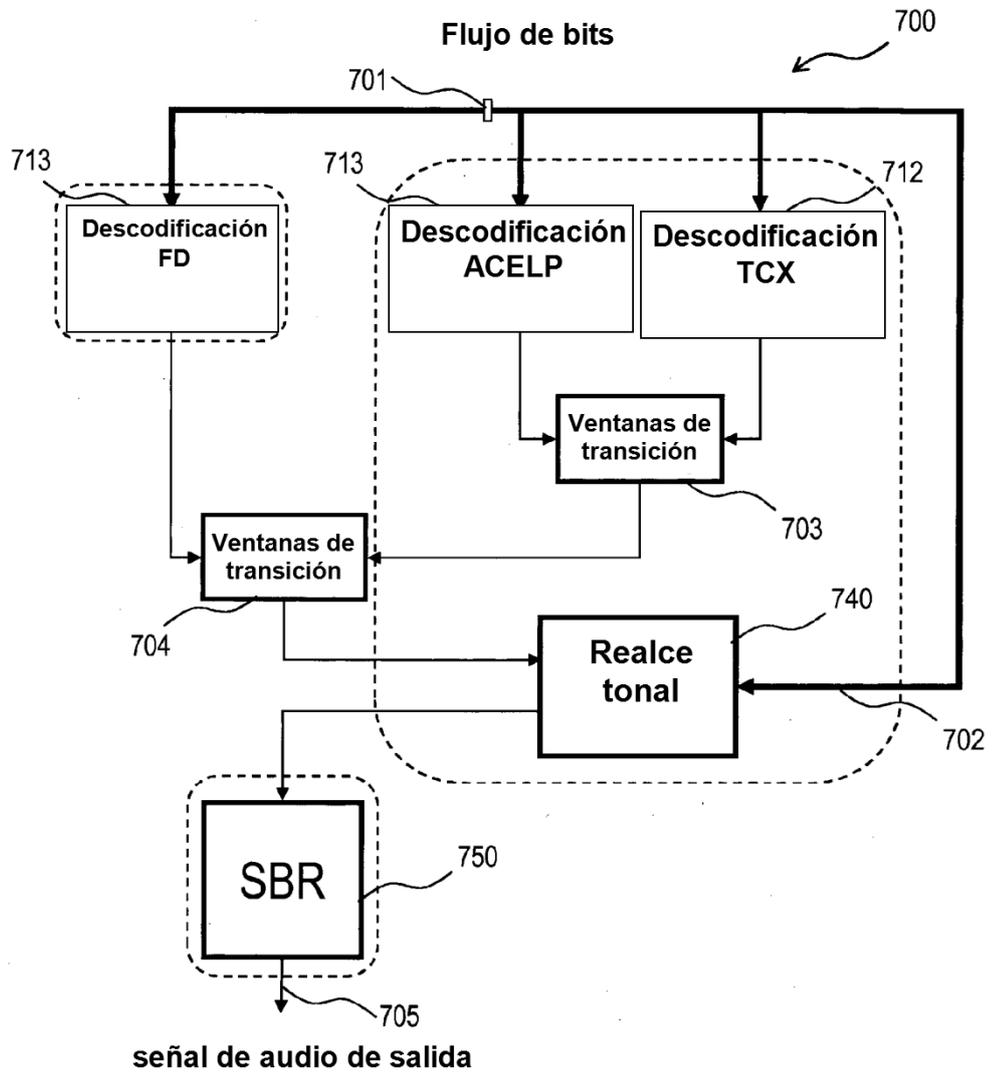


Fig. 7

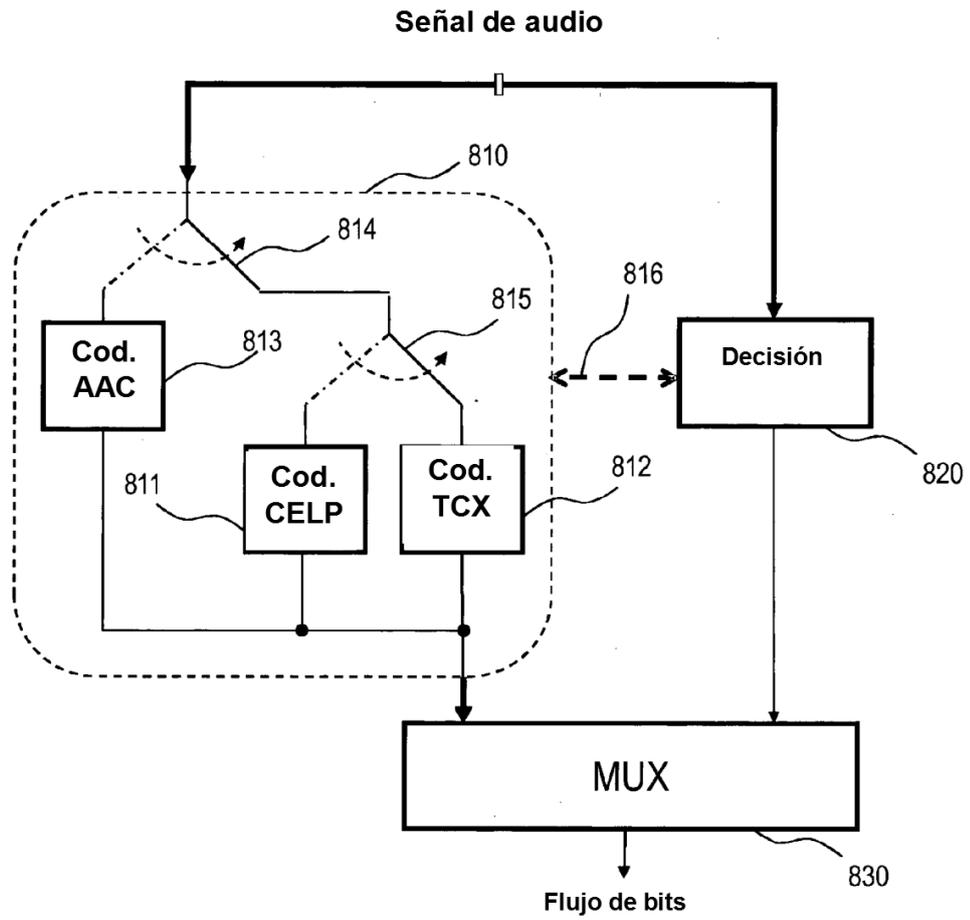


Fig. 8

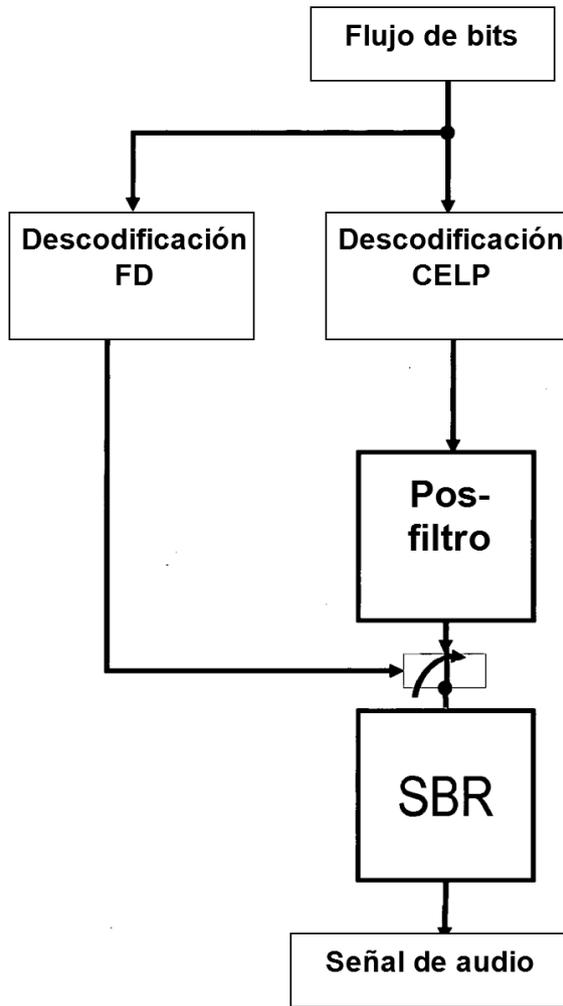


Fig. 9
(técnica anterior)

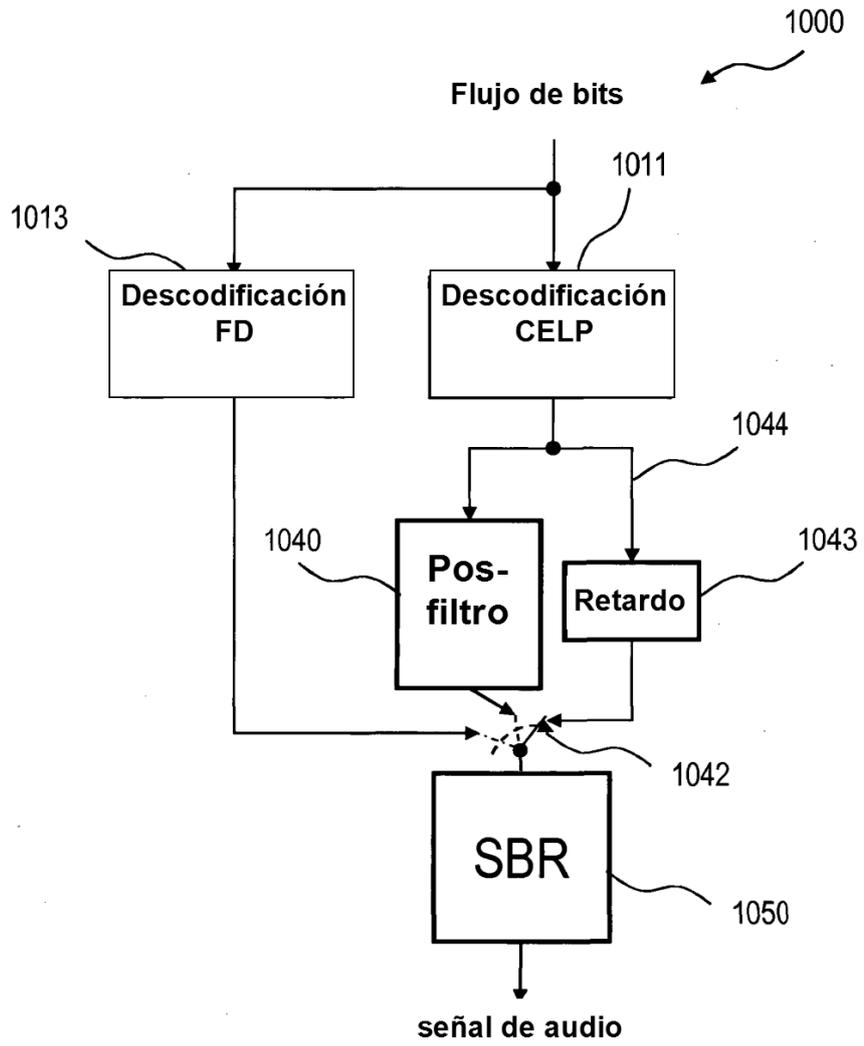


Fig. 10

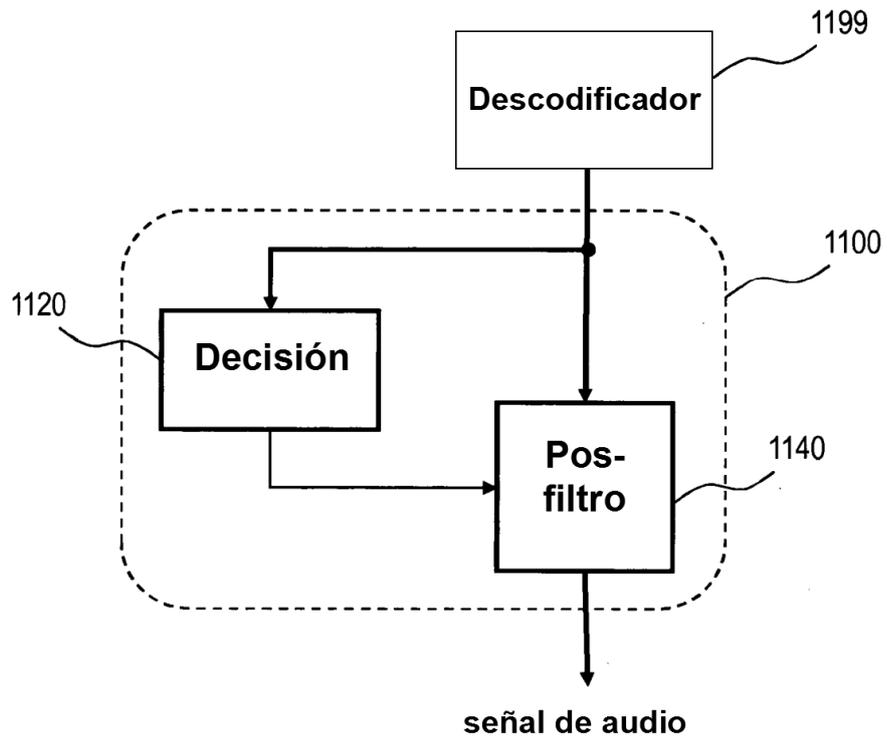


Fig. 11