

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 809 677**

51 Int. Cl.:

G10L 19/008	(2013.01)	G10L 19/032	(2013.01)
G10L 19/02	(2013.01)		
H04S 1/00	(2006.01)		
G10L 19/002	(2013.01)		
G10L 25/03	(2013.01)		
G10L 25/51	(2013.01)		
G10L 25/21	(2013.01)		
G10L 19/06	(2013.01)		
G10L 19/09	(2013.01)		
G10L 19/24	(2013.01)		

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 86 Fecha de presentación y número de la solicitud internacional: **22.09.2016 PCT/CA2016/051107**
- 87 Fecha y número de publicación internacional: **30.03.2017 WO17049398**
- 96 Fecha de presentación y número de la solicitud europea: **22.09.2016 E 16847685 (1)**
- 97 Fecha y número de publicación de la concesión europea: **24.06.2020 EP 3353779**

54 Título: **Método y sistema para codificar una señal de sonido estéreo utilizando parámetros de codificación de un canal primario para codificar un canal secundario**

30 Prioridad:

25.09.2015 US 201562232589 P
14.07.2016 US 201662362360 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
05.03.2021

73 Titular/es:

VOICEAGE CORPORATION (100.0%)
750 Lucerne Road Suite 250
Town of Mount Royal, Quebec H3R 2H6, CA

72 Inventor/es:

VAILLANCOURT, TOMMY y
JELINEK, MILAN

74 Agente/Representante:

SÁEZ MAESO, Ana

ES 2 809 677 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

Método y sistema para codificar una señal de sonido estéreo utilizando parámetros de codificación de un canal primario para codificar un canal secundario

Campo técnico

5 La presente divulgación se refiere a la codificación de sonido estéreo, en particular, pero no exclusivamente, a la codificación de voz y/o audio estéreo capaz de producir una buena calidad estéreo en una escena de audio compleja a baja velocidad de bits y bajo retardo.

Antecedentes

10 Históricamente, la telefonía conversacional se ha implementado con teléfonos portátiles que tienen un solo transductor para emitir sonido solamente a uno de los oídos del usuario. En la última década, los usuarios comenzaron a utilizar su teléfono portátil junto con un auricular para recibir el sonido en sus dos oídos principalmente para escuchar música, pero también, a veces, para escuchar el habla. Sin embargo, cuando se utiliza un teléfono portátil para transmitir y recibir voz conversacional, el contenido sigue siendo monofónico, pero se presenta a los dos oídos del usuario cuando se utilizan auriculares.

15 Con el nuevo estándar de codificación de voz 3GPP como se describe en la Referencia [1], la calidad del sonido codificado, por ejemplo, voz y/o audio que se transmite y recibe a través de un teléfono portátil se ha mejorado significativamente. El siguiente paso natural es transmitir información estéreo de modo que el receptor se acerque lo más posible a una escena de audio de la vida real que se captura en el otro extremo del enlace de comunicación.

20 En los códecs de audio, por ejemplo, como, se describe en la Referencia [2], normalmente se utiliza la transmisión de información estéreo.

25 Para los códecs de habla conversacional, la señal monofónica es la norma. Cuando se transmite una señal estereofónica, la velocidad de bits a menudo necesita duplicarse ya que los canales izquierdo y derecho se codifican utilizando un códec monofónico. Esto funciona bien en la mayoría de los escenarios, pero presenta los inconvenientes de duplicar la velocidad de bits y no aprovechar cualquier redundancia potencial entre los dos canales (canales izquierdo y derecho). Además, para mantener la velocidad de bits general a un nivel razonable, se utiliza una velocidad de bits muy baja para cada canal, lo que afecta la calidad de sonido general. Por lo tanto, para reducir la redundancia en la transmisión de señales estereofónicas, a menudo se utiliza un análisis de canales de entrada basado en la coherencia, como se describe en la Referencia [9].

30 Una posible alternativa es utilizar el llamado estéreo paramétrico como se describe en la Referencia [6]. El estéreo paramétrico envía información como, por ejemplo, la diferencia de tiempo interaural (ITD) o las diferencias de intensidad interaural (IID). La última información se envía por banda de frecuencia y, a baja velocidad de bits, el presupuesto de bits asociado a la transmisión estéreo no es lo suficientemente alto como para permitir que estos parámetros funcionen eficientemente.

35 La transmisión de un factor de desplazamiento podría ayudar a crear un efecto estéreo básico a baja velocidad de bits, pero dicha técnica no hace nada para preservar el ambiente y presenta limitaciones inherentes. Una adaptación demasiado rápida del factor de desplazamiento se vuelve perturbadora para el oyente, mientras que una adaptación demasiado lenta del factor de desplazamiento no refleja la posición real de los altavoces, lo que dificulta la obtención de una buena calidad en el caso de conversadores que interfieren o cuando la fluctuación del ruido de fondo es importante. Actualmente, la codificación de voz estéreo conversacional con una calidad decente para todas las escenas de audio posibles requiere una velocidad de bits mínima de alrededor de 24 kb/s para señales de banda ancha (WB); por debajo de esa velocidad de bits, la calidad del habla comienza a sufrir.

40 Con la creciente globalización de la fuerza laboral y la división de los equipos de trabajo en todo el mundo, existe la necesidad de mejorar las comunicaciones. Por ejemplo, los participantes de una teleconferencia pueden estar en lugares diferentes y distantes. Algunos participantes podrían estar en sus automóviles, otros podrían estar en una gran cámara anecoica o incluso en su sala de estar. De hecho, todos los participantes desean sentir que tienen una discusión cara a cara. Implementando voz estéreo, más generalmente sonido estéreo en dispositivos portátiles sería un gran paso en esta dirección.

Resumen

50 De acuerdo con un primer aspecto, la presente divulgación se refiere a un método de codificación de sonido estéreo para codificar canales izquierdo y derecho de una señal de sonido estéreo, que comprende la mezcla reductora de los canales izquierdo y derecho de la señal de sonido estéreo para producir canales primarios y secundarios, que codifican el canal primario y codifican el canal secundario. La codificación del canal secundario comprende analizar la coherencia entre los parámetros de codificación calculados durante la codificación del canal secundario y los parámetros de codificación calculados durante la codificación del canal primario para decidir si los parámetros de
55 codificación calculados durante la codificación del canal primario están suficientemente cerca de los parámetros de

codificación calculados durante la codificación del canal secundario a ser reutilizado durante la codificación del canal secundario.

De acuerdo con un segundo aspecto, se proporciona un sistema de codificación de sonido estéreo para codificar canales izquierdo y derecho de una señal de sonido estéreo, que comprende un mezclador reductor de los canales izquierdo y derecho de la señal de sonido estéreo para producir canales primarios y secundarios, un codificador del canal primario y un codificador del canal secundario; en el que el codificador de canal secundario comprende un analizador de coherencia entre los parámetros de codificación de canal secundario calculados durante la codificación de canal secundario y los parámetros de codificación de canal primario calculados durante la codificación de canal primario para decidir si los parámetros de codificación de canal primario están suficientemente cerca de los parámetros de codificación de canal secundario a estar reutilizado durante la codificación del canal secundario.

De acuerdo con un tercer aspecto, se proporciona un sistema de codificación de sonido estéreo para codificar canales izquierdo y derecho de una señal de sonido estéreo, que comprende: al menos un procesador; y una memoria acoplada al procesador y que comprende instrucciones no transitorias que, cuando se ejecutan, hacen que el procesador implemente: un mezclador reductor de los canales izquierdo y derecho de la señal de sonido estéreo para producir canales primarios y secundarios; un codificador del canal primario y un codificador del canal secundario; en el que el codificador de canal secundario comprende un analizador de coherencia entre los parámetros de codificación de canal secundario calculados durante la codificación de canal secundario y los parámetros de codificación de canal primario calculados durante la codificación de canal primario para decidir si los parámetros de codificación de canal primario están suficientemente cerca de los parámetros de codificación de canal secundario a ser reutilizado durante la codificación del canal secundario.

Un aspecto adicional se refiere a un sistema de codificación de sonido estéreo para codificar canales izquierdo y derecho de una señal de sonido estéreo, que comprende: al menos un procesador; y una memoria acoplada al procesador y que comprende instrucciones no transitorias que cuando se ejecutan hacen que el procesador: mezcle hacia abajo los canales izquierdo y derecho de la señal de sonido estéreo para producir canales primarios y secundarios; codificar el canal primario utilizando un codificador de canal primario y codificar el canal secundario utilizando un codificador de canal secundario; y analizar, en el codificador del canal secundario, la coherencia entre los parámetros de codificación del canal secundario calculados durante la codificación del canal secundario y los parámetros de codificación del canal primario calculados durante la codificación del canal primario para decidir si los parámetros de codificación del canal primario están suficientemente cerca de los parámetros de codificación del canal secundario a ser reutilizado durante la codificación del canal secundario.

La presente divulgación se refiere además a una memoria legible por procesador que comprende instrucciones no transitorias que, cuando se ejecutan, hacen que un procesador implemente las operaciones del método descrito anteriormente.

Los anteriores, y otros objetos, ventajas y características del método y sistema de codificación de sonido estéreo para codificar canales izquierdo y derecho de una señal de sonido estéreo se harán más evidentes al leer la siguiente descripción no restrictiva de realizaciones ilustrativas del mismo, dada a modo de ejemplo solo con referencia a los dibujos adjuntos.

Breve descripción de los dibujos

En los dibujos adjuntos:

La Figura 1 es un diagrama de bloques esquemático de un sistema de procesamiento y comunicación de sonido estéreo que representa un posible contexto de implementación del método y sistema de codificación de sonido estéreo como se divulga en la siguiente descripción;

La Figura 2 es un diagrama de bloques que ilustra simultáneamente un método y un sistema de codificación de sonido estéreo de acuerdo con un primer modelo, presentado como un diseño estéreo integrado;

La Figura 3 es un diagrama de bloques que ilustra simultáneamente un método y un sistema de codificación de sonido estéreo de acuerdo con un segundo modelo, presentado como un modelo integrado;

La Figura 4 es un diagrama de bloques que muestra simultáneamente suboperaciones de una operación de mezcla reductora de dominio del tiempo del método de codificación de sonido estéreo de las Figuras 2 y 3, y módulos de un mezclador de canales del sistema de codificación de sonido estéreo de las Figuras 2 y 3;

La Figura 5 es un gráfico que muestra cómo una diferencia de correlación linealizada a largo plazo se asigna a un factor β y a un factor de normalización de energía ϵ ;

La Figura 6 es un gráfico de curvas múltiples que muestra una diferencia entre utilizar un esquema *pca/klf* sobre una trama completa y utilizar una función de mapeo "coseno";

La Figura 7 es un gráfico de múltiples curvas que muestra un canal primario, un canal secundario y los espectros de estos canales primarios y secundarios que resultan de la aplicación de mezcla reductora de dominio en el tiempo para una muestra estéreo que se ha grabado en una pequeña cámara ecoica utilizando una configuración de micrófonos binaurales con ruido de oficina en el fondo;

- 5 La Figura 8 es un diagrama de bloques que ilustra simultáneamente un método y sistema de codificación de sonido estéreo, con una posible implementación de optimización de la codificación de los canales Y primario y X secundario de la señal de sonido estéreo;

La Figura 9 es un diagrama de bloques que ilustra una operación de análisis de coherencia de filtro LP y el analizador de coherencia de filtro LP correspondiente del método y sistema de codificación de sonido estéreo de la Figura 8;

- 10 La Figura 10 es un diagrama de bloques que ilustra simultáneamente un método de decodificación de sonido estéreo y un sistema de decodificación de sonido estéreo;

La Figura 11 es un diagrama de bloques que ilustra características adicionales del método y sistema de decodificación de sonido estéreo de la Figura 10;

- 15 La Figura 12 es un diagrama de bloques simplificado de una configuración de ejemplo de componentes de hardware que forman el sistema de codificación de sonido estéreo y el decodificador de sonido estéreo de la presente divulgación;

- 20 La Figura 13 es un diagrama de bloques que ilustra concurrentemente otras realizaciones de suboperaciones de la operación de mezcla reductora en el dominio del tiempo del método de codificación de sonido estéreo de las figuras 2 y 3, y módulos del mezclador de canales del sistema de codificación de sonido estéreo de las figuras 2 y 3, utilizando un factor de preadaptación para mejorar la estabilidad de la imagen estéreo;

La Figura 14 es un diagrama de bloques que ilustra simultáneamente operaciones de una corrección de retardo temporal y módulos de un corrector de retardo temporal;

La Figura 15 es un diagrama de bloques que ilustra al mismo tiempo un método y sistema alternativos de codificación de sonido estéreo;

- 25 La Figura 16 es un diagrama de bloques que ilustra simultáneamente las suboperaciones de un análisis de coherencia de tono y los módulos de un analizador de coherencia de tono;

La Figura 17 es un diagrama de bloques que ilustra el método y el sistema de codificación estéreo simultánea que utiliza una mezcla reductora en el dominio del tiempo con una capacidad de operar en el dominio del tiempo y en el dominio de la frecuencia; y

- 30 la Figura 18 es un diagrama de bloques que ilustra al mismo tiempo otro método y sistema de codificación estéreo que utiliza una mezcla reductora en el dominio del tiempo con una capacidad de operar en el dominio del tiempo y en el dominio de la frecuencia.

Descripción detallada

- 35 La presente divulgación se refiere a la producción y transmisión, con una baja velocidad de bits y bajo retardo, de una representación realista del contenido de sonido estéreo, por ejemplo, contenido de voz y/o audio, en particular, pero no exclusivamente, una escena de audio compleja. Una escena de audio compleja incluye situaciones en las que (a) la correlación entre las señales de sonido que graban los micrófonos es baja, (b) hay una fluctuación importante del ruido de fondo, y/o (c) está presente un conversador que interfiere. Los ejemplos de escenas de audio complejas comprenden una gran cámara de conferencias anecoica con una configuración de micrófonos A/B, una pequeña
- 40 cámara de eco con micrófonos binaurales y una pequeña cámara de eco con una configuración de micrófonos mono/laterales. Todas estas configuraciones de cámara podrían incluir ruido de fondo fluctuante y/o interferencias.

Los códecs de sonido estéreo conocidos, tales como 3GPP AMR-WB+ como se describe en la Referencia [7], son ineficientes para codificar sonido que no está cerca del modelo monofónico, especialmente a baja velocidad de bits. Ciertos casos son particularmente difíciles de codificar utilizando técnicas estéreo existentes. Tales casos incluyen:

- 45 - LAAB (Gran cámara anecoica con configuración de micrófonos A/B);
- SEBI (pequeña cámara de eco con configuración de micrófonos binaurales); y
- SEMS (pequeña cámara de eco con configuración de micrófonos mono/laterales).

- 50 Agregar un ruido de fondo fluctuante y/o conversadores que interfieren hace que estas señales de sonido sean aún más difíciles de codificar a baja velocidad de bits utilizando técnicas estéreo dedicadas, tales como estéreo paramétrico. Una alternativa para codificar tales señales es utilizar dos canales monofónicos, duplicando así la velocidad de bits y el ancho de banda de la red que se está utilizando.

El último estándar de voz conversacional 3GPP EVS proporciona un rango de velocidad de bits de 7.2 kb/s a 96 kb/s para operación de banda ancha (WB) y 9.6 kb/s a 96 kb/s para operación de banda súper ancha (SWB). Esto significa que las tres velocidades de bit mono dual más bajas que utilizan EVS son 14.4, 16.0 y 19.2 kb/s para la operación WB y 19.2, 26.3 y 32.8 kb/s para la operación SWB. Aunque la calidad de voz del 3GPP AMR-WB desplegado, como se describe en la Referencia [3], mejora con respecto a su códec predecesor, la calidad de la voz codificada a 7.2 kb/s en un entorno ruidoso está lejos de ser transparente y, por lo tanto, se puede anticipar que la calidad de voz de mono dual a 14.4 kb/s también sería limitada. A velocidades de bits tan bajas, el uso de la velocidad de bits se maximiza de tal manera que se obtiene la mejor calidad de voz posible tan a menudo como sea posible. Con el método y el sistema de codificación de sonido estéreo como se divulga en la siguiente descripción, la velocidad de bits total mínima para el contenido de voz estéreo conversacional, incluso en el caso de escenas de audio complejas, debe ser de alrededor de 13 kb/s para WB y 15.0 kb/s para SWB. A velocidades de bits que son más bajas que las velocidades de bits utilizadas en un enfoque mono dual, la calidad y la inteligibilidad de la voz estéreo se mejora en gran medida para escenas de audio complejas.

La Figura 1 es un diagrama de bloques esquemático de un sistema 100 de procesamiento y comunicación de sonido estéreo que representa un posible contexto de implementación del método y sistema de codificación de sonido estéreo como se divulga en la siguiente descripción.

El sistema 100 de procesamiento y comunicación de sonido estéreo de la Figura 1 admite la transmisión de una señal de sonido estéreo a través de un enlace 101 de comunicación. El enlace 101 de comunicación puede comprender, por ejemplo, un cable o un enlace de fibra óptica. Alternativamente, el enlace 101 de comunicación puede comprender al menos en parte un enlace de radiofrecuencia. El enlace de radiofrecuencia a menudo admite múltiples comunicaciones simultáneas que requieren recursos de ancho de banda compartido, como los que se pueden encontrar con la telefonía celular. Aunque no se muestra, el enlace 101 de comunicación puede ser reemplazado por un dispositivo de almacenamiento en una implementación de dispositivo único del sistema 100 de procesamiento y comunicación que graba y almacena la señal de sonido estéreo codificada para su posterior reproducción.

Aún con referencia a la Figura 1, por ejemplo, un par de micrófonos 102 y 122 produce los canales izquierdo 103 y derecho 123 de una señal de sonido estéreo analógica original detectada, por ejemplo, en una escena de audio compleja. Como se indica en la descripción anterior, la señal de sonido puede comprender, en particular, pero no exclusivamente, voz y/o audio. Los micrófonos 102 y 122 pueden estar dispuestos de acuerdo con una configuración A/B, binaural o mono/lateral.

Los canales izquierdo 103 y derecho 123 de la señal de sonido analógica original se suministran a un convertidor 104 de analógico a digital (A/D) para convertirlos en los canales izquierdo 105 y derecho 125 de una señal de sonido estéreo digital original. Los canales izquierdo 105 y derecho 125 de la señal de sonido estéreo digital original también pueden grabarse y suministrarse desde un dispositivo de almacenamiento (no mostrado).

Un codificador 106 de sonido estéreo codifica los canales izquierdo 105 y derecho 125 de la señal de sonido estéreo digital, produciendo así un conjunto de parámetros de codificación que se multiplexan bajo la forma de un 107 flujo de bits entregado a un codificador 108 opcional de corrección de errores. El codificador 108 opcional de corrección de errores, cuando está presente, agrega redundancia a la representación binaria de los parámetros de codificación en el flujo de bits 107 antes de transmitir el flujo 111 de bits resultante a través del enlace 101 de comunicación.

En el lado del receptor, un decodificador 109 de corrección de errores opcional utiliza la información redundante mencionada anteriormente en el flujo 111 de bits digital recibido para detectar y corregir errores que pueden haber ocurrido durante la transmisión a través del enlace 101 de comunicación, produciendo un flujo 112 de bits con los parámetros de codificación recibidos. Un decodificador 110 de sonido estéreo convierte los parámetros de codificación recibidos en el flujo 112 de bits para crear canales izquierdo 113 y derecho 133 sintetizados de la señal de sonido estéreo digital. Los canales izquierdo 113 y derecho 133 de la señal de sonido estéreo digital reconstruidos en el decodificador 110 de sonido estéreo se convierten en canales izquierdo 114 y derecho 134 sintetizados de la señal de sonido estéreo analógica en un convertidor 115 de digital a analógico (D/A).

Los canales izquierdo 114 y derecho 134 sintetizados de la señal de sonido estéreo analógica se reproducen respectivamente de nuevo en un par de unidades 116 y 136 de altavoces. Alternativamente, los canales izquierdo 113 y derecho 133 de la señal de sonido estéreo digital del decodificador 110 de sonido estéreo también se puede suministrar y grabar en un dispositivo de almacenamiento (no mostrado).

Los canales izquierdo 105 y derecho 125 de la señal de sonido estéreo digital original de la Figura 1 corresponden a los canales izquierdo L y derecho R de las Figuras 2, 3, 4, 8, 9, 13, 14, 15, 17 y 18. Además, el codificador 106 de sonido estéreo de la Figura 1 corresponde al sistema de codificación de sonido estéreo de las Figuras 2, 3, 8, 15, 17 y 18.

El método y el sistema de codificación de sonido estéreo de acuerdo con la presente divulgación son dobles; se proporcionan los modelos primero y segundo.

La Figura 2 es un diagrama de bloques que ilustra simultáneamente el método y el sistema de codificación de sonido estéreo de acuerdo con el primer modelo, presentado como un diseño estéreo integrado basado en el núcleo EVS.

Con referencia a la Figura 2, el método de codificación de sonido estéreo de acuerdo con el primer modelo comprende una operación 201 de mezcla reductora en el dominio del tiempo, una operación 202 de codificación de canal primario, una operación 203 de codificación de canal secundario y una operación 204 de multiplexación.

5 Para realizar la operación 201 de mezcla reductora en el dominio del tiempo, un mezclador de canales 251 mezcla los dos canales estéreo de entrada (canal derecho R y canal izquierdo L) para producir un canal primario Y y un canal secundario X.

10 Para llevar a cabo la operación 203 de codificación de canal secundario, un codificador 253 de canal secundario selecciona y utiliza un número mínimo de bits (velocidad de bits mínima) para codificar el canal secundario X utilizando uno de los modos de codificación como se define en la siguiente descripción y produce un flujo 206 de bits codificado de canal secundario correspondiente. El presupuesto de bits asociado puede cambiar cada trama dependiendo del contenido de la trama.

15 Para implementar la operación 202 de codificación de canal primario, se utiliza un codificador 252 de canal primario. El codificador 253 de canal secundario indica al codificador 252 de canal primario el número de bits 208 utilizados en la trama actual para codificar el canal secundario X. Cualquier tipo adecuado de codificador puede usarse como codificador 252 de canal primario. Como ejemplo no limitativo, el codificador 252 de canal primario puede ser un codificador de tipo CELP. En esta realización ilustrativa, el codificador de tipo CELP del canal primario es una versión modificada del codificador EVS heredado, donde el codificador EVS se modifica para presentar una mayor escalabilidad de la velocidad de bits para permitir una asignación flexible de la velocidad de bits entre los canales primario y secundario. De esta manera, el codificador EVS modificado podrá utilizar todos los bits que no se utilizan para codificar el canal secundario X para codificar, con una velocidad de bits correspondiente, el canal primario Y y producir un flujo 205 de bits codificado del canal primario correspondiente.

20 Un multiplexor 254 concatena el flujo 205 de bits del canal primario y el flujo 206 de bits del canal secundario para formar un flujo 207 de bits multiplexado, para completar la operación 204 de multiplexación.

25 En el primer modelo, el número de bits y la velocidad de bits correspondiente (en el flujo 206 de bits) usado para codificar el canal secundario X es menor que el número de bits y la velocidad de bits correspondiente (en el flujo 205 de bits) usado para codificar el canal primario Y. Esto puede verse como dos (2) canales de velocidad de bits variables en la que la suma de las velocidades de bits de los dos canales X e Y representa una velocidad de bits total constante. Este enfoque puede tener diferentes sabores con más o menos énfasis en el canal primario Y. De acuerdo con un primer ejemplo, cuando se pone un énfasis máximo en el canal primario Y, el presupuesto de bits del canal secundario X se fuerza agresivamente al mínimo. De acuerdo con un segundo ejemplo, si se pone menos énfasis en el canal primario Y, entonces el presupuesto de bits para el canal secundario X puede hacerse más constante, lo que significa que la velocidad de bits promedio del canal secundario X es ligeramente mayor en comparación con el primer ejemplo.

30 Se recuerda que los canales derecho R e izquierdo L de la señal de sonido estéreo digital de entrada son procesados por tramas sucesivas de una duración dada que puede corresponder a la duración de las tramas utilizadas en el procesamiento EVS. Cada trama comprende una cantidad de muestras de los canales derecho R e izquierdo L, dependiendo de la duración dada de la trama y la frecuencia de muestreo utilizada.

35 La Figura 3 es un diagrama de bloques que ilustra simultáneamente el método y el sistema de codificación de sonido estéreo de acuerdo con el segundo modelo, presentado como un modelo integrado.

40 En referencia a la Figura 3, el método de codificación de sonido estéreo de acuerdo con el segundo modelo comprende una operación 301 de mezcla reductora en el dominio del tiempo, una operación 302 de codificación de canal primario, una operación 303 de codificación de canal secundario y una operación 304 de multiplexación.

Para completar la operación 301 de mezcla reductora en el dominio del tiempo, un mezclador 351 de canales mezcla los dos canales de entrada derecho R e izquierdo L para formar un canal primario Y y un canal secundario X.

45 En la operación 302 de codificación del canal primario, un codificador 352 de canal primario codifica el canal primario Y para producir un flujo 305 de bits codificado del canal primario. Nuevamente, cualquier tipo de codificador adecuado puede usarse como el codificador 352 del canal primario. Como ejemplo no limitativo, el codificador 352 del canal primario puede ser un codificador tipo CELP. En esta realización ilustrativa, el codificador 352 de canal primario utiliza un estándar de codificación de voz tal como el modo de codificación mono EVS heredado o el modo de codificación AMR-WB-IO, por ejemplo, lo que significa que la porción monofónica del flujo 305 de bits sería interoperable con el EVS heredado, AMR-WB-IO o el decodificador AMR-WB heredado cuando la velocidad de bits es compatible con dicho decodificador. Dependiendo del modo de codificación que se seleccione, puede ser necesario algún ajuste del canal primario Y para procesar a través del codificador 352 de canal primario.

50 En la operación 303 de codificación de canal secundario, un codificador 353 de canal secundario codifica el canal X secundario en la velocidad de bits inferior utilizando uno de los modos de codificación definidos en la siguiente descripción. El codificador 353 de canal secundario produce un flujo 306 de bits codificado de canal secundario.

55

Para realizar la operación 304 de multiplexación, un multiplexor 354 concatena el flujo 305 de bits codificado de canal primario con el flujo 306 de bits codificado de canal secundario para formar un flujo 307 de bits multiplexado. Esto se llama un modelo embebido, porque el flujo 306 de bits codificado de canal secundario asociado a estéreo se agrega encima de un flujo 305 de bits interoperable. El flujo de bits de canal 306 secundario se puede quitar del flujo 307 de bits estéreo multiplexado (flujos 305 y 306 de bits concatenados) en cualquier momento que resulta en un flujo de bits decodificable por un códec heredado como se describió anteriormente en el presente documento, mientras que un usuario de una versión más reciente del códec aún podría disfrutar de la decodificación estéreo completa.

Los primeros y segundos modelos descritos anteriormente están, de hecho, cerca uno del otro. La principal diferencia entre los dos modelos es la posibilidad de utilizar una asignación dinámica de bits entre los dos canales Y y X en el primer modelo, mientras que la asignación de bits es más limitada en el segundo modelo debido a consideraciones de interoperabilidad.

En la siguiente descripción se dan ejemplos de implementación y enfoques utilizados para lograr los modelos primero y segundo descritos anteriormente.

1) Mezcla reductora en el dominio del tiempo

Como se expresa en la descripción anterior, los modelos estéreo conocidos que funcionan a baja velocidad de bits tienen dificultades para codificar el habla que no está cerca del modelo monofónico. Los enfoques tradicionales realizan una mezcla reductora en el dominio de la frecuencia, por banda de frecuencia, utilizando, por ejemplo, una correlación por banda de frecuencia asociada con un análisis de componentes principales (*pca*) utilizando, por ejemplo, una transformación de Karhunen-Loève (*klt*), para obtener dos vectores, como se describe en las referencias [4] y [5]. Uno de estos dos vectores incorpora todo el contenido altamente correlacionado, mientras que el otro vector define todo el contenido que no está muy correlacionado. El método más conocido para codificar voz a velocidades de bits bajas utiliza un códec de dominio de tiempo, como un códec CELP (predicción lineal excitada por código), en el que las soluciones conocidas de dominio de frecuencia no son directamente aplicables. Por esa razón, si bien la idea detrás de la *pca/klt* por banda de frecuencia es interesante, cuando el contenido es voz, el canal primario Y necesita convertirse de nuevo al dominio del tiempo y, después de tal conversión, su contenido ya no se parece a la voz tradicional, especialmente en el caso de las configuraciones descritas anteriormente que utilizan un modelo específico de voz como CELP. Esto tiene el efecto de reducir el rendimiento del códec de voz. Además, a una velocidad de bits baja, la entrada de un códec de voz debe estar lo más cerca posible de las expectativas del modelo interno del códec.

A partir de la idea de que una entrada de un códec de voz de baja velocidad de bits debe estar lo más cerca posible de la señal de voz esperada, se ha desarrollado una primera técnica. La primera técnica se basa en una evolución del esquema tradicional *pca/klt*. Mientras que el esquema tradicional calcula el *pca/klt* por banda de frecuencia, la primera técnica lo calcula en todo el marco, directamente en el dominio del tiempo. Esto funciona adecuadamente durante los segmentos activos del habla, siempre que no haya ruido de fondo o conversador que interfiere. El esquema *pca/klt* determina qué canal (canal izquierdo L o derecho R) contiene la información más útil, este canal se envía al codificador del canal primario. Desafortunadamente, el esquema *pca/klt* sobre una base de trama no es confiable en presencia de ruido de fondo o cuando dos o más personas están hablando entre sí. El principio del esquema *pca/klt* implica la selección de un canal de entrada (R o L) u otro, lo que a menudo conduce a cambios drásticos en el contenido del canal primario a codificar. Al menos por las razones anteriores, la primera técnica no es suficientemente confiable y, de acuerdo con lo anterior, se presenta aquí una segunda técnica para superar las deficiencias de la primera técnica y permitir una transición más suave entre los canales de entrada. Esta segunda técnica se describirá más adelante con referencia a las Figuras 4-9.

En referencia a la Figura 4, la operación 201/301 de mezcla reductora en el dominio del tiempo (Figuras 2 y 3) comprende las siguientes sub-operaciones: una suboperación 401 de análisis de energía, una suboperación 402 de análisis de tendencia de energía, una suboperación 403 de análisis de correlación normalizada del canal R y L, una suboperación 404 que calcula la diferencia de correlación a largo plazo (LT), una suboperación 405 de cuantificación y conversión del factor β de diferencia de correlación a largo plazo y una suboperación 406 de mezcla reductora en el dominio del tiempo.

Teniendo en cuenta la idea de que la entrada de un códec de sonido de baja velocidad de bits (como voz y/o audio) debe ser lo más homogénea posible, la suboperación 401 de análisis de energía se lleva a cabo en el mezclador 252/351 de canal por un analizador 451 de energía para determinar primero, por trama, la energía *rms* (Raíz Cuadrada de la Media) de cada canal de entrada R y L utilizando las relaciones (1):

$$rms_L(t) = \sqrt{\frac{\sum_{i=0}^{N-1} L(i)^2}{N}}; \quad rms_R(t) = \sqrt{\frac{\sum_{i=0}^{N-1} R(i)^2}{N}}; \quad (1)$$

donde los subíndices *L* y *R* representan los canales izquierdo y derecho respectivamente, *L* (*i*) representa la muestra *i* del canal *L*, *R* (*i*) representa la muestra *i* del canal *R*, *N* corresponde al número de muestras por trama, y *t* representa una trama actual.

El analizador 451 de energía luego utiliza los valores *rms* de las relaciones (1) para determinar los valores *rms a largo plazo* \overline{rms} para cada canal utilizando las relaciones (2):

$$\overline{rms}_{-L}(t) = 0.6 \cdot \overline{rms}_{-L}(t_{-1}) + 0.4 \cdot rms_L; \quad \overline{rms}_{-R}(t) = 0.6 \cdot \overline{rms}_{-R}(t_{-1}) + 0.4 \cdot rms_R; \quad (2)$$

donde *t* representa la trama actual y *t-1* la trama anterior.

- 5 Para realizar la suboperación 402 de análisis de tendencias de energía, un analizador 452 de tendencias de energía del mezclador 251/351 de canales utiliza los valores *rms a largo plazo* \overline{rms} para determinar la tendencia de la energía en cada canal \overline{rms}_{dt} L y R utilizando relaciones (3):

$$\overline{rms}_{dt_L} = \overline{rms}_{-L}(t) - \overline{rms}_{-L}(t_{-1}); \quad \overline{rms}_{dt_R} = \overline{rms}_{-R}(t) - \overline{rms}_{-R}(t_{-1}). \quad (3)$$

- 10 La tendencia de los valores *rms a largo plazo* se utiliza como información que muestra si los eventos temporales capturados por los micrófonos se están desvaneciendo o si están cambiando canales. Los valores *rms a largo plazo* y su tendencia también se utilizan para determinar una velocidad de convergencia α de una diferencia de correlación a largo plazo como se describirá aquí más adelante.

- 15 Para realizar la suboperación 403 de análisis de correlación normalizada de canales L y R, un analizador 453 de correlación normalizado L y R calcula una correlación G_{LR} para cada uno de los canales izquierdo L y derecho R normalizados contra una versión de señal monofónica *m(i)* del sonido, como voz y/o audio, en la trama *t* utilizando las relaciones (4):

$$G_L(t) = \frac{\sum_{i=0}^{N-1} (L(i) \cdot m(i))}{\sum_{i=0}^{N-1} m(i)^2}, \quad G_R(t) = \frac{\sum_{i=0}^{N-1} (R(i) \cdot m(i))}{\sum_{i=0}^{N-1} m(i)^2}, \quad m(i) = \left(\frac{L(i) + R(i)}{2} \right), \quad (4)$$

- 20 donde *N*, como ya se mencionó, corresponde al número de muestras en una trama, y *t* representa la trama actual. En la realización actual, todas las correlaciones normalizadas y los valores *rms* determinados por las relaciones 1 a 4 se calculan en el dominio de tiempo, para toda la trama. En otra configuración posible, estos valores se pueden calcular en el dominio de frecuencia. Por ejemplo, las técnicas descritas en el presente documento, que están adaptadas a señales de sonido que tienen características de voz, pueden ser parte de un marco más amplio que puede cambiar entre un método genérico de codificación de audio estéreo de dominio de frecuencia y el método descrito en la presente divulgación. En este caso, calcular las correlaciones normalizadas y los valores *rms* en el dominio de la frecuencia puede presentar alguna ventaja en términos de complejidad o reutilización del código.

- 25 Para calcular la diferencia de correlación a largo plazo (LT) en la suboperación 404, un calculador 454 calcula para cada canal L y R en las correlaciones normalizadas suavizadas de trama actual utilizando las relaciones (5):

$$\overline{G}_L(t) = \alpha \cdot \overline{G}_L(t_{-1}) + (1-\alpha) \cdot G_L(t) \quad \text{y} \quad \overline{G}_R(t) = \alpha \cdot \overline{G}_R(t_{-1}) + (1-\alpha) \cdot G_R(t), \quad (5)$$

- 30 donde α es la velocidad de convergencia mencionada anteriormente. Finalmente, el calculador 454 determina la diferencia de correlación a largo plazo \overline{G}_{LR} utilizando la relación (6):

$$\overline{G}_{LR}(t) = \overline{G}_L(t) - \overline{G}_R(t). \quad (6)$$

- 35 en una realización de ejemplo, la velocidad de convergencia α puede tener un valor de 0,8 o 0,5 dependiendo de las energías a largo plazo calculado en las relaciones (2) y la tendencia de las energías a largo plazo tal como se calcula en las relaciones (3). Por ejemplo, la velocidad de convergencia α puede tener un valor de 0.8 cuando las energías a largo plazo de los canales izquierdo L y derecho R evolucionan en la misma dirección, una diferencia entre la diferencia de correlación a largo plazo \overline{G}_{LR} en la trama *t* el largo de diferencia de correlación a largo plazo \overline{G}_{LR} en la trama *t-1* es baja (inferior a 0,31 para esta realización de ejemplo), y al menos uno de los valores *rms a largo plazo* de los canales izquierdo y derecho R está por encima de un cierto umbral (2000 en este ejemplo realización). Tales casos significan que ambos canales L y R están evolucionando sin problemas, que no hay un cambio rápido de energía de un canal a otro, y que al menos un canal contiene un nivel significativo de energía. De lo contrario, cuando las energías a largo plazo de los canales derecho R e izquierdo L evolucionan en diferentes direcciones, cuando la diferencia entre las diferencias de correlación a largo plazo es alta, o cuando los dos canales derecho R e izquierdo L tienen energías bajas, entonces α se establecerá en 0.5 para aumentar la velocidad de adaptación de la diferencia de correlación a largo plazo \overline{G}_{LR} .

- 45 Para llevar a cabo la suboperación 405 de conversión y cuantificación, una vez que la diferencia de correlación a largo plazo \overline{G}_{LR} se ha estimado adecuadamente en el calculador 454, el convertidor y cuantificador 455 convierte esta diferencia en un factor β que se cuantifica y se suministra a (a) el codificador 252 de canal primario (Figura 2), (b) el codificador 253/353 de canal secundario (Figuras 2 y 3), y (c) el multiplexor 254/354 (Figuras 2 y 3) para la transmisión a un decodificador dentro del flujo 207/307 de bits multiplexado a través de un enlace de comunicación tal como 101 de la Figura 1.

El factor β representa dos aspectos de la entrada estéreo combinados en un parámetro. Primero, el factor β representa una proporción o contribución de cada uno de los canales derecho R e izquierdo L que se combinan para crear el

canal primario Y, segundo, también puede representar un factor de escala de energía para aplicar al canal primario Y a obtener un canal primario que esté cerca en el dominio de la energía a cómo se vería una versión de señal monofónica del sonido. Por lo tanto, en el caso de una estructura embebida, permite que el canal primario Y se decodifique solo sin la necesidad de recibir el flujo 306 de bits secundario que transporta los parámetros estéreo. Este parámetro de energía también se puede utilizar para reescalar la energía del canal secundario X antes de codificarlo, de modo que la energía global del canal secundario X esté más cerca del rango de energía óptimo del codificador del canal secundario. Como se muestra en la Figura 2, la información de energía intrínsecamente presente en el factor β también se puede utilizar para mejorar la asignación de bits entre los canales primario y secundario.

El factor β cuantificado puede transmitirse al decodificador utilizando un índice. Dado que el factor β puede representar (a) contribuciones respectivas de los canales izquierdo y derecho al canal primario y (b) un factor de escala de energía para aplicar al canal primario para obtener una versión de señal monofónica del sonido o una información de energía/correlación que ayuda a asignar más eficientemente los bits entre el canal primario Y y el canal secundario X, el índice transmitido al decodificador transmite dos elementos de información distintos con el mismo número de bits.

Para obtener un mapeo entre la diferencia de correlación a largo plazo $\overline{G_{LR}}(t)$ y el factor β , en este ejemplo de realización, el convertidor y el cuantificador 455 limitan primero la diferencia de correlación a largo plazo $\overline{G_{LR}}(t)$ entre -1.5 a 1.5 y luego linealiza esta diferencia de correlación a largo plazo entre 0 y 2 para obtener una diferencia de correlación linealizada a largo plazo temporal $G'_{LR}(t)$ como se muestra en la relación (7):

$$G'_{LR}(t) = \begin{cases} 0, & \overline{G_{LR}}(t) \leq -1.5 \\ \frac{2}{3} \cdot G_{LR}(t) + 1.0, & -1.5 < \overline{G_{LR}}(t) < 1.5 \\ 2, & \overline{G_{LR}}(t) \geq 1.5 \end{cases} \quad (7)$$

En una implementación alternativa, se puede decidir utilizar solo una parte del espacio lleno con la diferencia de correlación linealizada a largo plazo $G'_{LR}(t)$ al limitar aún más sus valores entre, por ejemplo, 0.4 y 0.6. Esta limitación adicional tendría el efecto de reducir la localización de la imagen estéreo, pero también de guardar algunos bits de cuantificación. Dependiendo de la elección del diseño, esta opción puede ser considerada.

Después de la linealización, el convertidor y el cuantificador 455 realizan un mapeo de la diferencia de correlación linealizada a largo plazo $G'_{LR}(t)$ en el dominio "coseno" utilizando la relación (8):

$$\beta(t) = \frac{1}{2} \cdot \left(1 - \cos \left(\pi \cdot \frac{G'_{LR}(t)}{2} \right) \right) \quad (8)$$

Para realizar la suboperación 406 de mezcla reductora en el dominio del tiempo, un mezclador 456 reductor de dominio en el tiempo produce el canal primario Y y el canal secundario X como una mezcla de los canales derecho R e izquierdo L utilizando las relaciones (9) y (10):

$$Y(i) = R(i) \cdot (1 - \beta(t)) + L(i) \cdot \beta(t) \quad (9)$$

$$X(i) = L(i) \cdot (1 - \beta(t)) + R(i) \cdot \beta(t) \quad (10)$$

donde $i = 0, \dots, N-1$ es el índice de muestra en la trama y t es el índice de la trama.

La Figura 13 es un diagrama de bloques que muestra simultáneamente otras realizaciones de suboperaciones de la operación 201/301 de mezcla reductora en el dominio del tiempo del método de codificación de sonido estéreo de las Figuras 2 y 3, y módulos del mezclador 251/351 de canales del Sistema de codificación de sonido estéreo de las Figuras 2 y 3, que utiliza un factor de preadaptación para mejorar la estabilidad de la imagen estéreo. En una implementación alternativa como se representa en la Figura 13, la operación 201/301 de mezcla reductora en el dominio del tiempo comprende las siguientes sub-operaciones: una suboperación 1301 de análisis de energía, una suboperación 1302 de análisis de tendencia de energía, una suboperación 1303 de análisis de correlación normalizada de canal L y R, una suboperación 1304 de cálculo del factor de preadaptación, una operación 1305 de aplicar el factor de preadaptación a correlaciones normalizadas, una suboperación 1306 de cálculo de diferencia de correlación a largo plazo (LT), una ganancia para la suboperación 1307 de conversión y cuantificación del factor β , y una suboperación 1308 de mezcla reductora en el dominio del tiempo.

Las suboperaciones 1301, 1302 y 1303 son realizadas respectivamente por un analizador 1351 de energía, un analizador 1352 de tendencias de energía y un analizador 1353 de correlación normalizado L y R, sustancialmente de la misma manera que se explica en la descripción anterior en relación con las suboperaciones 401, 402 y 403, y los analizadores 451, 452 y 453 de la Figura 4.

Para realizar la suboperación 1305, el mezclador 251/351 de canales comprende un calculador 1355 para aplicar el factor de preadaptación a_r directamente a las correlaciones ($G_{L|R}$)($G_L(t)$ y $G_R(t)$) de las relaciones (4) de modo que su evolución sea suavizada dependiendo de la energía y las características de ambos canales. Si la energía de la señal es baja o si tiene algunas características son voz, la evolución de la ganancia de correlación puede ser más lenta.

Para llevar a cabo la suboperación 1304 de cálculo del factor de preadaptación, el mezclador 251/351 de canales comprende un calculador 1354 de factor de preadaptación, suministrado con (a) los valores de energía de las relaciones del canal izquierdo y derecho a largo plazo (2) del analizador 1351 de energía, (b) clasificación de trama de tramas anteriores y (c) información de actividad de voz de las tramas anteriores. El calculador 1354 del factor de preadaptación calcula el factor de preadaptación a_r , que puede linealizarse entre 0.1 y 1 dependiendo de los valores mínimos rms a largo plazo $\overline{rms}_{L|R}$ de los canales izquierdo y derecho del analizador 1351, utilizando la relación (6a):

$$a_r = \max(\min(M_a \cdot \min(\overline{rms}_{-L}(t), \overline{rms}_{-R}(t)) + B_a, 1), 0.1), \quad (11a)$$

En una realización, el coeficiente M_a puede tener el valor de 0.0009 y el coeficiente B_a el valor de 0.16. En una variante, el factor de preadaptación a_r puede ser forzado a 0.15, por ejemplo, si una clasificación previa de los dos canales R y L es indicativa de características sin voz y de una señal activa. También se puede utilizar un indicador de tiempo de retención de detección de actividad de voz (VAD) para determinar que una parte anterior del contenido de una trama era un segmento activo.

La operación 1305 de aplicar el factor de preadaptación a_r a las correlaciones de $G_{L|R}(G_L(t)$ y $G_R(t)$ normalizadas de las relaciones (4)) de los canales izquierdo L y derecho R es distinta de la operación 404 de la Figura 4. En lugar de calcular correlaciones normalizadas suavizadas a largo plazo (LT) al aplicar a las correlaciones de $G_{L|R}(G_L(t)$ y $G_R(t)$ normalizadas, un factor $(1-\alpha)$, siendo α la velocidad de convergencia definida anteriormente (Relaciones (5)), el calculador 1355 aplica el factor de preadaptación a_r directamente a las correlaciones de $G_{L|R}(G_L(t)$ y $G_R(t)$ normalizadas de los canales izquierdo L y derecho R utilizando la relación (11b):

$$\tau_L(t) = a_r \cdot G_L(t) + (1 - a_r) \cdot \overline{G}_L(t) \text{ y } \tau_R(t) = a_r \cdot G_R(t) + (1 - a_r) \cdot \overline{G}_R(t). \quad (11b)$$

El calculador 1355 genera ganancias de correlación adaptadas $\tau_{L|R}$ que se proporcionan a un calculador de diferencias 1356 de correlación a largo plazo (LT). La operación de la mezcla 201/301 reductora de dominio en el tiempo (Figuras 2 y 3) comprende, en la implementación de la Figura 13, una diferencia de correlación a largo plazo (LT) que calcula la suboperación 1306, una suboperación 1307 de cuantificación y conversión del factor β de diferencia de correlación a largo plazo y una suboperación 1358 de mezcla de reducción de dominio en el tiempo similar a las suboperaciones 404, 405 y 406, respectivamente, de la Figura 4.

La operación de la mezcla 201/301 de reducción de dominio en el tiempo (Figuras 2 y 3) comprende, en la implementación de la Figura 13, una suboperación 1306 de cálculo de diferencia de correlación a largo plazo (LT), una suboperación 1307 de cuantificación y conversión del factor β de diferencia de correlación a largo plazo y una suboperación 1358 de mezcla reductora de dominio en el tiempo similar a las suboperaciones 404, 405 y 406, respectivamente, de la Figura 4.

Las suboperaciones 1306, 1307 y 1308 se realizan respectivamente mediante un calculador 1356, un convertidor y cuantificador 1357 y un mezclador 1358 de reducción de dominio en el tiempo, sustancialmente de la misma manera que se explica en la descripción anterior en relación con las suboperaciones 404, 405 y 406, y el calculador 454, el convertidor y el cuantificador 455 y el mezclador 456 de reducción de dominio en el tiempo.

La Figura 5 muestra cómo se mapea una diferencia $G'_{LR}(t)$ de correlación a largo plazo linealizada al factor β y al escalado de energía. Se puede observar que para una diferencia $G'_{LR}(t)$ de correlación a largo plazo linealizada al factor β de 1.0, lo que significa que las energías/correlaciones del canal derecho R e izquierdo L son casi iguales, el factor β es igual a 0.5 y un factor ϵ de normalización de energía (reescalado) es 1.0. En esta situación, el contenido del canal primario Y es básicamente una mezcla mono y el canal secundario X forma un canal lateral. A continuación, se describe el cálculo del factor ϵ de normalización de energía (reescalado).

Por otro lado, si la diferencia $G'_{LR}(t)$ de correlación linealizada a largo plazo es igual a 2, lo que significa que la mayor parte de la energía está en el canal izquierdo L, entonces el factor β es 1 y el factor de normalización de energía (reescalado) es 0.5, lo que indica que el canal primario Y contiene básicamente el canal izquierdo L en una implementación de diseño integrada o una representación reducida del canal izquierdo L en una implementación de diseño integrada. En este caso, el canal secundario X contiene el canal derecho R. En las realizaciones de ejemplo, el convertidor y cuantificador 455 o 1357 cuantifica el factor β utilizando 31 posibles entradas de cuantificación. La versión cuantificada del factor β se representa utilizando un índice de 5 bits y, como se describió anteriormente, se suministra al multiplexor para su integración en el flujo 207/307 de bits multiplexado, y se transmite al decodificador a través del enlace de comunicación.

En una realización, el factor β también puede usarse como un indicador tanto para el codificador 252/352 de canal primario como para el codificador 253/353 de canal secundario para determinar la asignación de velocidad de bits. Por ejemplo, si el factor β es cercano a 0.5, lo que significa que las dos (2) energías del canal de entrada/correlación con el mono están cercanas entre sí, se asignarían más bits al canal secundario X y menos bits al canal primario Y, excepto si el contenido de ambos canales es bastante cercano, entonces el contenido del canal secundario será realmente de baja energía y probablemente se considerará inactivo, permitiendo así que muy pocos bits lo codifiquen. Por otro lado, si el factor β está más cerca de 0 o 1, entonces la asignación de velocidad de bits favorecerá el canal primario Y.

La Figura 6 muestra la diferencia entre utilizar el esquema *pca/klt* mencionado anteriormente en toda la trama (dos curvas superiores de la Figura 6) versus el uso de la función “coseno” como se desarrolló en la relación (8) para calcular el factor β (curva inferior de la Figura 6). Por naturaleza, el esquema *pca/klt* tiende a buscar un mínimo o un máximo. Esto funciona bien en el caso del habla activa como se muestra en la curva central de la Figura 6, pero esto no funciona realmente bien para el habla con ruido de fondo, ya que tiende a cambiar continuamente de 0 a 1 como se muestra en la curva central de la Figura 6. El cambio demasiado frecuente a las extremidades, 0 y 1, causa muchos artefactos al codificar a baja velocidad de bits. Una posible solución habría sido suavizar las decisiones del esquema *pca/klt*, pero esto habría afectado negativamente la detección de ráfagas de voz y sus ubicaciones correctas, mientras que la función “coseno” de la relación (8) es más eficiente a este respecto.

La Figura 7 muestra el canal primario Y, el canal secundario X y los espectros de estos canales primarios Y y secundarios X resultantes de la aplicación de la mezcla de dominio en el tiempo a una muestra estéreo que se ha grabado en una pequeña cámara ecoica utilizando configuración de micrófonos binaurales con ruido de oficina en segundo plano. Después de la operación de mezcla reductora en el dominio del tiempo, se puede observar que los dos canales todavía tienen formas de espectro similares y el canal X secundario todavía tiene una voz como contenido temporal, permitiendo así utilizar el modelo basado en voz para codificar el canal X secundario.

La mezcla reductora en el dominio del tiempo presentada en la descripción anterior puede mostrar algunos problemas en el caso especial de los canales derecho R e izquierdo L que están invertidos en fase. Sumar los canales derecho R e izquierdo L para obtener una señal monofónica daría como resultado que los canales derecho R e izquierdo se cancelen entre sí. Para resolver este posible problema, en una realización, el mezclador 251/351 de canales compara la energía de la señal monofónica con la energía de los canales derecho R e izquierdo L. La energía de la señal monofónica debe ser al menos mayor que la energía de uno de los canales derecho R e izquierdo L. De lo contrario, en esta realización, el modelo de mezcla reductora en el dominio del tiempo entra en el caso especial de fase invertida. En presencia de este caso especial, el factor β se fuerza a 1 y el canal secundario X se codifica de manera forzada utilizando el modo genérico o sin voz, evitando así el modo de codificación inactivo y asegurando la codificación adecuada del canal secundario X. Este caso especial, donde no se aplica reescalado de energía, se señala al decodificador utilizando la última combinación de bits (valor de índice) disponible para la transmisión del factor β (Básicamente, dado que β se cuantifica utilizando 5 bits y 31 entradas (niveles de cuantificación) se utilizan para la cuantificación como se describió anteriormente, la 32ª combinación de bits posible (valor de índice o entrada) se utiliza para señalar este caso especial).

En una implementación alternativa, se puede poner más énfasis en la detección de señales que son subóptimas para las técnicas de codificación y mezcla reductora descritas anteriormente, como en el caso de señales desfasadas o casi desfasadas. Una vez que se detectan estas señales, las técnicas de codificación subyacentes pueden adaptarse si es necesario.

Por lo general, para la mezcla reductora en el dominio del tiempo como se describe en el presente documento, cuando los canales izquierdo L y derecho R de una señal estéreo de entrada están desfasados, puede producirse una cancelación durante el proceso de mezcla reductora, lo que podría conducir a una calidad subóptima. En los ejemplos anteriores, la detección de estas señales es simple y la estrategia de codificación comprende codificar ambos canales por separado. Pero a veces, con señales especiales, como señales que están fuera de fase, puede ser más eficiente realizar una mezcla reductora similar a mono/lateral ($\beta = 0.5$), donde se pone un mayor énfasis en el canal lateral. Dado que un tratamiento especial de estas señales puede ser beneficioso, la detección de tales señales debe realizarse con cuidado. Además, la transición del modelo de mezcla reductora en el dominio del tiempo normal como se describió en la descripción anterior y el modelo de mezcla reductora en el dominio del tiempo que se ocupa de estas señales especiales puede activarse en una región de muy baja energía o en regiones donde el tono de ambos canales no es estable, de modo que el cambio entre los dos modelos tiene un efecto subjetivo mínimo.

La corrección de retardo temporal (TDC) (véase el corrector 1750 de retardo temporal en las Figuras 17 y 18) entre los canales L y R, o una técnica similar a la descrita en la referencia [8], se puede realizar antes de entrar en el módulo 201/301, 251/351 de mezcla reductora. En tal realización, el factor β puede terminar teniendo un significado diferente del que se ha descrito anteriormente. Para este tipo de implementación, a condición de que la corrección de retardo temporal funcione como se esperaba, el factor β puede acercarse a 0.5, lo que significa que la configuración de la mezcla reductora en el dominio del tiempo está cerca de una configuración mono/lateral. Con el correcto funcionamiento de la corrección de retardo temporal (TDC), el lado puede contener una señal que incluye una cantidad menor de información importante. En ese caso, la velocidad de bits del canal secundario X puede ser mínima cuando el factor β está cerca de 0.5. Por otro lado, si el factor β es cercano a 0 o 1, esto significa que la corrección de retardo temporal (TDC) puede no superar adecuadamente la situación de alineación errónea de retardo y es probable que el contenido del canal secundario X sea más complejo, por lo que necesita una velocidad de bits más alta. Para ambos tipos de implementación, el factor β y, por asociación, el factor ϵ de normalización de energía (reescalado), puede usarse para mejorar la asignación de bits entre el canal primario Y y el canal secundario X.

La Figura 14 es un diagrama de bloques que muestra operaciones simultáneas de una detección de señal desfasada y módulos de un detector de señal 1450 desfasada que forman parte de la operación 201/301 de mezcla reductora y mezclador 251/351 de canales. Las operaciones de detección de señal desfasada incluyen, como se muestra en la Figura 14, una operación 1401 de detección de señal desfasada, una operación 1402 de detección de posición de

5 conmutación y la operación 1403 de selección de mezclador de canales, para elegir entre la operación 201/301 de mezcla reductora de dominio en el tiempo y una operación 1404 de mezcla reductora de dominio en el tiempo específica fuera de fase. Estas operaciones se realizan respectivamente mediante un detector 1451 de señal desfasado, un detector 1452 de posición de conmutación, un selector 1453 de mezclador de canales, el mezclador 251/351 de canal reductor en el dominio del tiempo descrito anteriormente, y un mezclador 1454 de canal reductor de dominio en el tiempo específico fuera de fase.

La detección 1401 de señal fuera de fase se basa en una correlación de bucle abierto entre los canales primario y secundario en tramas anteriores. Con este fin, el detector 1451 calcula en las tramas anteriores una diferencia de energía $S_m(t)$ entre una señal lateral $s(i)$ y una señal mono $m(i)$ utilizando las relaciones (12a) y (12b):

$$10 \quad S_m(t) = 10 \cdot \left(\log_{10} \left(\frac{\sqrt{\sum_{i=0}^{N-1} s(i)^2}}{N} \right) - \log_{10} \left(\frac{\sqrt{\sum_{i=0}^{N-1} m(i)^2}}{N} \right) \right), \quad (12a)$$

$$m(i) = \left(\frac{L(i)+R(i)}{2} \right) \text{ y } s(i) = \left(\frac{L(i)-R(i)}{2} \right), \quad (12b)$$

Entonces, el detector 1451 calcula la diferencia $\overline{S}_m(t)$ de energía mono a lado a largo plazo utilizando la relación (12c):

$$\overline{S}_m(t) = \begin{cases} 0.9 \cdot \overline{S}_m(t-1), & \text{para contenido inactivo,} \\ 0.9 \cdot \overline{S}_m(t-1) + 0.1 \cdot S_m(t), & \text{de otra forma} \end{cases} \quad (12c)$$

15 donde t indica la trama actual, $t-1$ la trama anterior, y donde el contenido inactivo puede derivarse del indicador tiempo de retención del Detector de Actividad de Voz (VAD) o de un contador de tiempo de retención VAD.

Además de la diferencia $\overline{S}_m(t)$ de energía del lado a largo plazo a mono, también se toma en cuenta la última correlación C_{FIL} máxima de bucle abierto de tono de cada canal Y y X, como se define en la cláusula 5.1.10 de la Referencia [1], para decidir cuándo el modelo actual se considera subóptimo. $C_{P(t-1)}$ representa la correlación máxima del bucle abierto de tono del canal primario Y en una trama anterior y $C_{S(t-1)}$, la correlación máxima de bucle de tono abierto del canal secundario X en la trama anterior. Se calcula el indicador F_{sub} de suboptimalidad mediante el detector 1452 de posición de conmutación de acuerdo con los siguientes criterios:

25 Si la diferencia $\overline{S}_m(t)$ de energía de largo plazo a mono está por encima de un cierto umbral, por ejemplo, cuando $\overline{S}_m(t) > 2.0$, si tanto las correlaciones $C_{P(t-1)}$ y $C_{S(t-1)}$, máximas de bucle abierto de tono están entre 0.85 y 0.92, significa que las señales tienen una buena correlación, pero no están tan correlacionadas como lo estaría una señal de voz, el indicador de suboptimalidad F_{sub} se establece en 1, lo que indica una condición desfasada entre los canales izquierdo y derecho R.

De lo contrario, el indicador de suboptimalidad F_{sub} se establece en 0, lo que indica que no hay una condición desfasada entre los canales izquierdo L y derecho R.

30 Para agregar cierta estabilidad en la decisión del indicador de suboptimalidad, el detector 1452 de posición de conmutación implementa un criterio con respecto al contorno de tono de cada canal Y y X. El detector 1452 de posición de conmutación determina que el mezclador 1454 de canales se usará para codificar las señales subóptimas cuando, en la realización de ejemplo, al menos tres (3) instancias consecutivas del indicador de suboptimalidad F_{sub} se establecen en 1 y la estabilidad de tono del último cuadro de uno de los canales primarios, $p_{pc(t-1)}$, o del canal secundario, $p_{sc(t-1)}$, es mayor que 64. La estabilidad del tono consiste en la suma de las diferencias absolutas de los tres tonos de bucle abierto $p_{0|1|2}$ como se define en 5.1.10 de Referencia [1], calculada por el detector 1452 de posición de conmutación utilizando la relación (12d):

$$p_{pc} = |p_1 - p_0| + |p_2 - p_1| \text{ y } p_{sc} = |p_1 - p_0| + |p_2 - p_1| \quad (12d)$$

40 El detector 1452 de posición de conmutación proporciona la decisión al selector 1453 de mezclador de canales que, a su vez, selecciona el mezclador 251/351 de canales o el mezclador 1454 de canal en consecuencia. El selector 1453 de mezclador de canales implementa una histéresis de tal manera que, cuando se selecciona el mezclador 1454 de canales, esta decisión se mantiene hasta que se cumplan las siguientes condiciones: una cantidad de tramas consecutivas, por ejemplo 20 tramas, se consideran óptimas, la estabilidad de tono de la última trama de uno de los $p_{pc(t-1)}$ primarios o el canal $p_{pc(t-1)}$, secundario es mayor que un número predeterminado, por ejemplo 64, y la diferencia $\overline{S}_m(t)$ de energía del lado a largo plazo a mono está por debajo o igual a 0.

45 2) Codificación dinámica entre canales primarios y secundarios

La Figura 8 es un diagrama de bloques que ilustra al mismo tiempo el método y sistema de codificación de sonido estéreo, con una posible implementación de la optimización de la codificación tanto de los canales primario Y como secundario X de la señal de sonido estéreo, como voz o audio.

En referencia a la Figura 8, el método de codificación de sonido estéreo comprende una operación 801 de preprocesamiento de baja complejidad implementada por un preprocesador 851 de baja complejidad, una operación 802 de clasificación de señal implementada por un clasificador 852 de señal, una operación 803 de decisión implementada por un módulo 853 de decisión, un modelo de cuatro (4) subtramas genérico que codifica solamente la operación 804 de codificación implementada por un modelo de cuatro (4) subtramas genérico que codifica solamente el módulo 854, un modelo de dos (2) subtramas que codifica la operación 805 implementada por un modelo de dos (2) subtramas que codifica el módulo 855, y una operación 806 de análisis de coherencia de filtro LP implementada por un analizador 856 de coherencia de filtro LP.

Después de que el mezclador 351 de canales haya realizado la mezcla 301 reductora de dominio en el tiempo, en el caso del modelo incorporado, el canal primario Y se codifica (operación 302 de codificación de canal primario) (a) utilizando como el codificador 352 de canal primario un codificador heredado tal como el codificador EVS heredado o cualquier otro codificador de sonido heredado adecuado (se debe tener en cuenta que, como se mencionó en la descripción anterior, se puede utilizar cualquier tipo de codificador adecuado como el codificador 352 de canal primario). En el caso de una estructura integrada, se utiliza un códec de voz dedicado como el codificador 252 de canal primario. El codificador 252 de voz dedicado puede ser un codificador basado en velocidad de bits variable (VBR), por ejemplo, una versión modificada del codificador EVS heredado, que se ha modificado para tener una mayor escalabilidad de velocidad de bits que permite el manejo de una velocidad de bits variable en un nivel por trama (una vez más, se debe tener en cuenta que, como se mencionó en la descripción anterior, se puede utilizar cualquier tipo de codificador adecuado como el codificador 252 primario de canal). Esto permite que la cantidad mínima de bits utilizados para codificar el canal secundario X varíe en cada trama y se adapte a las características de la señal de sonido a codificar. Al final, la firma del canal secundario X será lo más homogénea posible.

La codificación del canal secundario X, es decir, la correlación de entrada de baja energía a mono, está optimizada para utilizar una velocidad de bits mínima, en particular, pero no exclusivamente, para contenido similar a voz. Para ese fin, la codificación del canal secundario puede aprovechar los parámetros que ya están codificados en el canal primario Y, como los coeficientes de filtro LP (LPC) y/o el desfase 807. Específicamente, se decidirá, como se describe más adelante, si los parámetros calculados durante la codificación del canal primario están suficientemente cerca de los parámetros correspondientes calculados durante la codificación del canal secundario para ser reutilizados durante la codificación del canal secundario.

Primero, la operación 801 de preprocesamiento de baja complejidad se aplica al canal secundario X utilizando el preprocesador 851 de baja complejidad, en el que se calcula un filtro LP, una detección de actividad de voz (VAD) y un tono de bucle abierto en respuesta al canal secundario X. Estos últimos cálculos pueden ser implementados, por ejemplo, por aquellos realizados en el codificador heredado EVS y descritos respectivamente en las cláusulas 5.1.9, 5.1.12 y 5.1.10 de la Referencia [1] de las cuales, como se indica anteriormente, el contenido completo se incorpora aquí mediante referencia. Dado que, como se mencionó en la descripción anterior, se puede utilizar cualquier tipo de codificador adecuado como codificador 252/352 de canal primario, los cálculos anteriores pueden implementarse por aquellos realizados en dicho codificador de canal primario.

Luego, las características de la señal del canal secundario X son analizadas por el clasificador 852 de señal para clasificar el canal secundario X como sin voz, genérico o inactivo utilizando técnicas similares a las de la función de clasificación de señal EVS, cláusula 5.1.13 de la misma referencia [1]. Estas operaciones son conocidas por los expertos en la materia y se pueden extraer del estándar 3GPP TS 26.445, v.12.0.0 por simplicidad, pero también se pueden utilizar implementaciones alternativas.

a. Reutilización de los coeficientes del filtro LP de canal primario

Una parte importante del consumo de velocidad de bits reside en la cuantificación de los coeficientes del filtro LP (LPC). A una velocidad de bits baja, la cuantificación completa de los coeficientes de filtro LP puede tomar hasta casi el 25% del presupuesto de bits. Dado que el canal secundario X a menudo tiene un contenido de frecuencia cercano al canal primario Y, pero con el nivel de energía más bajo, vale la pena verificar si sería posible reutilizar los coeficientes de filtro LP del canal primario Y. Para hacerlo, como se muestra en la Figura 8, se ha desarrollado una operación 806 de análisis de coherencia de filtro LP implementada por un analizador 856 de coherencia de filtro LP, en el que se calculan pocos parámetros y se comparan para validar la posibilidad de reutilizar o no los coeficientes 807 de filtro LP (LPC) del canal primario Y.

La Figura 9 es un diagrama de bloques que ilustra la operación 806 de análisis de coherencia del filtro LP y el analizador 856 de coherencia del filtro LP correspondiente del método y sistema de codificación de sonido estéreo de la Figura 8.

La operación 806 de análisis de coherencia del filtro LP y el analizador 856 de coherencia de filtro LP correspondiente del método y sistema de codificación de sonido estéreo de la Figura 8 comprenden, como se ilustra en la Figura 9, una suboperación 903 de análisis de filtro LP (predicción lineal) de canal primario implementada por un analizador 953 de filtro LP, una suboperación 904 de pesaje implementada por un filtro de pesaje 954, una suboperación 912 de análisis de filtro LP de canal secundario implementada por un analizador 962 de filtro LP, una suboperación 901 de pesaje implementada por un filtro 951 de pesaje, una suboperación 902 de análisis de distancia euclidiana

implementada por un analizador 952 de distancia euclidiana, una suboperación 913 de filtrado residual implementada por un filtro 963 residual, una suboperación 914 de cálculo de energía residual implementada por un calculador 964 de energía residual, una suboperación 915 de sustracción implementada por un sustractor 965, una suboperación 910 de cálculo de energía de sonido (como voz y/o audio) implementada por un calculador 960 de energía, una operación 906 de filtrado residual de canal secundario implementado por un filtro 956 residual de canal secundario, una suboperación 907 de cálculo de energía residual implementada por un calculador 957 de energía residual, una suboperación 908 de sustracción implementada por un sustractor 958, una suboperación 911 de cálculo de relación de ganancia implementada por un calculador de relación de ganancia, una suboperación 916 de comparación implementada por un comparador 966, una suboperación 917 de comparación implementada por un comparador 967, una suboperación 918 de decisión de uso del filtro LP del canal secundario implementada por un módulo 968 de decisión, y una suboperación 919 de decisión de reutilización del filtro LP del canal primario implementada por un módulo 969 de decisión.

Con referencia a la Figura 9, el analizador 953 de filtro LP realiza un análisis de filtro LP en el canal primario Y mientras que el analizador 962 de filtro LP realiza un análisis de filtro LP en el canal secundario X. El análisis de filtro LP realizado en cada uno de los canales Y primario y X secundario es similar al análisis descrito en cláusula 5.1.9 de la referencia [1].

Luego, los coeficientes de filtro LP A_y del analizador 953 de filtro LP se suministran al filtro 956 residual para un primer filtrado residual, r_y , del canal secundario X. De la misma manera, los coeficientes óptimos de filtro LP A_x desde el analizador 962 de filtro LP se suministra al filtro 963 residual para un segundo filtrado residual, r_x , del canal secundario X. El filtrado residual con coeficientes de filtro, A_y o A_x , se realiza utilizando la relación (11):

$$r_{y|x}(n) = s_x(n) + \sum_{i=0}^{16} (A_{y|x}(i) \cdot s_x(n-i)), n = 0, \dots, N-1 \quad (13)$$

donde, en este ejemplo, s_x representa el canal secundario, el orden del filtro LP es 16 y N es el número de muestras en la trama (tamaño de trama) que suele ser 256 correspondiente a una duración de trama de 20 ms a una frecuencia de muestreo de 12.8 kHz.

El calculador 910 calcula la energía E_x de la señal de sonido en el canal secundario X utilizando la relación (14):

$$E_x = 10 \cdot \log_{10}(\sum_{i=0}^{N-1} s_x(i)^2), \quad (14)$$

y el calculador 957 calcula la energía E_{ry} del residual del filtro 956 residual utilizando la relación (15):

$$E_{ry} = 10 \cdot \log_{10}(\sum_{i=0}^{N-1} r_y(i)^2). \quad (15)$$

El sustractor 958 resta la energía residual del calculador 957 de la energía del sonido del calculador 960 para producir una ganancia de predicción G_y .

De la misma manera, el calculador 964 calcula la energía E_{rx} del residuo del filtro 963 residual utilizando la relación (16):

$$E_{rx} = 10 \cdot \log_{10}(\sum_{i=0}^{N-1} r_x(i)^2), \quad (16)$$

y el sustractor 965 resta esta energía residual de la energía del sonido del calculador 960 para producir una ganancia de predicción G_x .

El calculador 961 calcula la relación de ganancia G_y/G_x . El comparador 966 compara la relación de ganancia G_y/G_x con un umbral τ , que es 0.92 en la realización de ejemplo. Si la relación G_y/G_x es menor que el umbral τ , el resultado de la comparación se transmite al módulo de decisión 968 que fuerza el uso de los coeficientes de filtro LP del canal secundario para codificar el canal secundario X.

El analizador 952 de distancia euclidiana realiza una medida de similitud de filtro LP, como la distancia euclidiana entre los pares espectrales de línea lsp_y calculada por el analizador 953 de filtro LP en respuesta al canal primario Y y los pares espectrales de línea lsp_x calculada por el analizador 962 de filtro LP en respuesta al canal secundario X. Como saben los expertos en la técnica, los pares espectrales de línea lsp_y y lsp_x representan los coeficientes de filtro LP en un dominio de cuantificación. El analizador 952 utiliza la relación (17) para determinar la distancia euclidiana $dist$:

$$dist = \sum_{i=0}^{M-1} (lsp_y(i) - lsp_x(i))^2 \quad (17)$$

donde M representa el orden del filtro, y lsp_y y lsp_x representan respectivamente los pares espectrales de línea calculados para los canales Y primario y X secundario.

Antes de calcular la distancia euclidiana en el analizador 952, es posible ponderar ambos conjuntos de pares espectrales de línea lsp_y e lsp_x a través de factores de pesaje respectivos de modo que se pone más o menos énfasis

en ciertas porciones del espectro. También se pueden utilizar otras representaciones de filtro LP para calcular la medida de similitud del filtro LP.

Una vez que se conoce la distancia euclidiana $dist$, se compara con un umbral σ en el comparador 967. En la realización de ejemplo, el umbral σ tiene un valor de 0.08. Cuando el comparador 966 determina que la relación G_V/G_X es igual o mayor que el umbral τ y el comparador 967 determina que la distancia euclidiana $dist$ es igual o mayor que el umbral σ , el resultado de las comparaciones se transmite al módulo 968 de decisión que fuerza el uso de los coeficientes de filtro LP del canal secundario para codificar el canal secundario X. Cuando el comparador 966 determina que la relación G_V/G_X es igual o mayor que el umbral τ y el comparador 967 determina que la distancia euclidiana $dist$ es menor que el umbral σ , el resultado de estas comparaciones se transmite al módulo 969 de decisión que fuerza la reutilización de los coeficientes del filtro LP del canal primario para codificar el canal secundario X. En este último caso, los coeficientes del filtro LP del canal primario se reutilizan como parte de la codificación del canal secundario.

Se pueden realizar algunas pruebas adicionales para limitar la reutilización de los coeficientes de filtro LP del canal primario para codificar el canal secundario X en casos particulares, por ejemplo, en el caso del modo de codificación son voz, donde la señal es lo suficientemente fácil de codificar que todavía hay una velocidad de bits disponible para codificar los coeficientes del filtro LP. También es posible forzar la reutilización de los coeficientes del filtro LP del canal primario cuando ya se obtiene una ganancia residual muy baja con los coeficientes del filtro LP del canal secundario o cuando el canal X secundario tiene un nivel de energía muy bajo. Finalmente, las variables τ , σ , el nivel de ganancia residual o el nivel de energía muy bajo en el que se puede forzar la reutilización de los coeficientes del filtro LP pueden adaptarse en función del presupuesto de bits disponible y/o en función de tipo de contenido. Por ejemplo, si el contenido del canal secundario se considera inactivo, incluso si la energía es alta, se puede decidir reutilizar los coeficientes de filtro LP del canal primario.

b. Codificación de baja velocidad de bits del canal secundario

Dado que los canales primarios Y y secundario X pueden ser una mezcla de los canales de entrada derecho R e izquierdo L, esto implica que, incluso si el contenido de energía del canal secundario X es bajo comparado con el contenido de energía del canal primario Y, se puede percibir un artefacto de codificación una vez que se realiza la mezcla de amplificación de los canales. Para limitar tal posible artefacto, la firma de codificación del canal secundario X se mantiene tan constante como sea posible para limitar cualquier variación de energía no deseada. Como se muestra en la Figura 7, el contenido del canal secundario X tiene características similares al contenido del canal primario Y y por esa razón se ha desarrollado un modelo de codificación de voz de muy baja velocidad de bits.

Con referencia de nuevo a la Figura 8, el analizador 856 de coherencia de filtro LP envía al módulo 853 de decisión la decisión de reutilizar los coeficientes de filtro LP del canal primario del módulo 969 de decisión o la decisión de utilizar los coeficientes de filtro LP del canal secundario del módulo 968 de decisión. El módulo 803 de decisión decide entonces no cuantificar los coeficientes del filtro LP del canal secundario cuando se reutilizan los coeficientes del filtro LP del canal primario y cuantificar los coeficientes del filtro LP del canal secundario cuando la decisión es utilizar los coeficientes del filtro LP del canal secundario. En el último caso, los coeficientes cuantificados del filtro LP del canal secundario se envían al multiplexor 254/354 para su inclusión en el flujo 207/307 de bits multiplexado.

En las cuatro (4) subtramas de modelo genérico que codifican solamente la operación 804 y las cuatro (4) subtramas correspondientes del modelo genérico que codifican el módulo 854, para mantener la velocidad de bits lo más baja posible, una búsqueda ACELP como se describe en la cláusula 5.2.3.1 de la Referencia [1] se utiliza solo cuando los coeficientes del filtro LP del canal primario Y pueden reutilizarse, cuando el canal secundario X se clasifica como genérico por el clasificador 852 de señal, y cuando la energía de los canales derecho R e izquierdo L están cerca uno del otro, lo que significa que las energías de los canales derecho R e izquierdo L están cerca unas de otras. Los parámetros de codificación encontrados durante la búsqueda de ACELP en los cuatro (4) modelos de subtramas genéricos que codifican el módulo 854 se utilizan luego para construir el flujo 206/306 de bits de canal secundario y enviado al multiplexor 254/354 para su inclusión en el flujo 207/307 de bits multiplexado.

De lo contrario, en el modelo de dos (2) subtramas que codifican la operación 805 y el modelo de dos (2) subtramas correspondientes que codifican el módulo 855, se utiliza un modelo de media banda para codificar el canal secundario X con contenido genérico cuando los coeficientes del filtro LP del canal primario Y no se pueden reutilizar. Para el contenido inactivo y sin voz, solo se codifica la forma del espectro.

En el módulo 855 de codificación, la codificación de contenido inactivo comprende (a) codificación de ganancia de banda espectral en el dominio de frecuencia más relleno de ruido y (b) codificación de los coeficientes de filtro LP del canal secundario cuando sea necesario como se describe respectivamente en (a) cláusulas 5.2.3.5.7 y 5.2.3.5.11 y (b) cláusula 5.2.2.1 de la Referencia [1]. El contenido inactivo se puede codificar a una velocidad de bits tan baja como 1.5 kb/s.

En el módulo 855 de codificación, la codificación sin voz del canal secundario X es similar a la codificación inactiva del canal secundario X, con la excepción de que la codificación sin voz utiliza un número adicional de bits para la

cuantificación de los coeficientes de filtro LP del canal secundario que están codificados para canal secundario sin voz.

5 El modelo de codificación genérico de media banda se construye de manera similar a ACELP como se describe en la cláusula 5.2.3.1 de la Referencia [1], pero se utiliza con solo dos (2) subtramas por trama. Por lo tanto, para hacerlo, el residual como se describe en la cláusula 5.2.3.1.1 de la Referencia [1], la memoria del libro de códigos adaptativo como se describe en la cláusula 5.2.3.1.4 de la Referencia [1] y el canal secundario de entrada se muestrean hacia abajo por un factor de 2. Los coeficientes del filtro LP también se modifican para representar el dominio de muestreo hacia abajo en lugar de la frecuencia de muestreo de 12.8 kHz utilizando una técnica como se describe en la cláusula 5.4.4.2 de la Referencia [1].

10 Después de la búsqueda ACELP, se realiza una extensión de ancho de banda en el dominio de frecuencia de la excitación. La extensión del ancho de banda primero replica las energías de la banda espectral inferior en la banda superior. Para replicar las energías de banda espectral, la energía de las primeras nueve (9) bandas espectrales, $G_{bd}(i)$, se encuentran como se describe en la cláusula 5.2.3.5.7 de la Referencia [1] y las últimas bandas se llenan como se muestra en relación (18):

$$15 \quad G_{bd}(i) = G_{bd}(16 - i - 1), \quad \text{para } i = 8, \dots, 15 \quad (18)$$

Entonces, el contenido de alta frecuencia del vector de excitación representado en el dominio de frecuencia $f_d(k)$ como se describe en la cláusula 5.2.3.5.9 de la Referencia [1] se rellena utilizando el contenido de frecuencia de banda inferior utilizando la relación (19):

$$f_d(k) = f_d(k - P_b), \quad \text{para } k = 128, \dots, 255 \quad (19)$$

20 donde el desplazamiento de tono, P_b , se basa en un múltiplo de la información de tono como se describe en la cláusula 5.2.3.1.4.1 de la Referencia [1] y se convierte en un desplazamiento de intervalos de frecuencia como se muestra en relación (20):

$$P_b = \begin{cases} \frac{8 \cdot \left(\frac{F_s}{\bar{T}}\right)}{F_r}, & \bar{T} > 64 \\ \frac{4 \cdot \left(\frac{F_s}{\bar{T}}\right)}{F_r}, & \bar{T} \leq 64 \end{cases}, \quad (20)$$

25 donde \bar{T} representa un promedio de la información de tono decodificada por subtrama, F_s es la frecuencia de muestreo interna, 12.8 kHz en esta realización de ejemplo, y F_r es la resolución de frecuencia.

Los parámetros de codificación encontrados durante la codificación inactiva de baja velocidad, la codificación sin voz de baja velocidad o la codificación genérica de media banda realizada en el módulo de codificación del modelo de dos (2) subtramas que codifican el módulo 855 se utilizan luego para construir el flujo 206/306 de bits del canal secundario enviado al multiplexor 254/354 para su inclusión en el flujo 207/307 de bits multiplexado.

30 C. Implementación alternativa de la codificación de baja velocidad de bits del canal secundario

La codificación del canal secundario X se puede lograr de manera diferente, con el mismo objetivo de utilizar un número mínimo de bits mientras se logra la mejor calidad posible y manteniendo una firma constante. La codificación del canal secundario X puede estar impulsada en parte por el presupuesto de bits disponible, independientemente de la posible reutilización de los coeficientes de filtro LP y la información de tono. Además, la codificación del modelo de dos (2) subtramas (operación 805) puede ser media banda o banda completa. En esta implementación alternativa de la codificación de baja velocidad de bits del canal secundario, los coeficientes de filtro LP y/o la información de tono del canal primario se pueden reutilizar y se pueden elegir las dos (2) codificaciones del modelo de subtramas en función del presupuesto de bits disponible para codificar el canal secundario X. Además, la codificación del modelo de 2 subtramas presentada a continuación se ha creado duplicando la longitud de la subtrama en lugar de muestrear hacia abajo/hacia arriba sus parámetros de entrada/salida.

La Figura 15 es un diagrama de bloques que ilustra simultáneamente un método de codificación de sonido estéreo alternativo y un sistema de codificación de sonido estéreo alternativo. El método y el sistema de codificación de sonido estéreo de la Figura 15 incluyen varias de las operaciones y módulos del método y sistema de la Figura 8, identificados utilizando los mismos números de referencia y cuya descripción no se repite aquí por brevedad. Además, el método de codificación de sonido estéreo de la Figura 15 comprende una operación 1501 de preprocesamiento aplicada al canal primario Y antes de su codificación en la operación 202/302, una operación 1502 de análisis de coherencia de tono, una operación 1504 de decisión sin voz/inactiva, una operación 1505 de decisión de codificación sin voz/inactiva, y una operación 1506 de decisión de modelo de 2/4 subtramas.

Las suboperaciones 1501, 1502, 1503, 1504, 1505 y 1506 son realizadas respectivamente por un preprocesador 1551 similar al preprocesador 851 de baja complejidad, un analizador 1552 de coherencia de tono, un estimador 1553 de asignación de bits, un módulo 1554 de decisión sin voz/inactivo, un módulo 1555 de decisión de codificación sin voz/inactivo y un módulo 1556 de decisión de modelo de 2/4 subtramas.

Para realizar la operación 1502 de análisis de coherencia de tono, el analizador 1552 de coherencia de tono es suministrado por los preprocesadores 851 y 1551 con tonos de bucle abierto de los canales Y primario y X secundario, respectivamente $OLpitch_{pri}$ y $OLpitch_{sec}$. El analizador 1552 de coherencia de tono de la Figura 15 se muestra con mayor detalle en la Figura 16, que es un diagrama de bloques que ilustra simultáneamente las suboperaciones de la operación 1502 de análisis de coherencia de tono y los módulos del analizador 1552 de coherencia de tono.

La operación 1502 de análisis de coherencia de tono realiza una evaluación de la similitud de los tonos de bucle abierto entre el canal primario Y y el canal secundario X para decidir en qué circunstancias el tono de bucle abierto primario se puede reutilizar en la codificación del canal secundario X. Para este fin, la operación 1502 de análisis de coherencia de tono comprende una suboperación 1601 de suma de tonos de bucle abierto de canal primario realizada por un sumador 1651 de tonos de bucle abierto de canal primario, y una suboperación 1602 de suma de tonos de bucle abierto de canal secundario realizada por un sumador 1652 de tonos de bucle abierto de canal secundario. La suma del sumador 1652 se resta (suboperación 1603) de la suma del sumador 1651 utilizando un sustractor 1653. El resultado de la resta de la suboperación 1603 proporciona una coherencia de tono estéreo. Como ejemplo no limitativo, las sumas en las suboperaciones 1601 y 1602 se basan en tres (3) tonos de bucle abierto anteriores consecutivos disponibles para cada canal Y y X. Los tonos de bucle abierto se pueden calcular, por ejemplo, como, se define en la cláusula 5.1.10 de la Referencia [1]. La coherencia S_{pc} de tono estéreo se calcula en las operaciones secundarias 1601, 1602 y 1603 utilizando la relación (21):

$$S_{pc} = \left| \sum_{i=0}^2 p_{p(i)} - \sum_{i=0}^2 p_{s(i)} \right| \quad (21)$$

donde $p_{p(s(i))}$ representa los tonos de bucle abierto de los canales primarios Y y secundarios X e i representa la posición de los tonos de bucle abierto.

Cuando la coherencia de tono estéreo está por debajo de un umbral predeterminado Δ , se puede permitir la reutilización de la información de tono del canal primario Y, dependiendo de un presupuesto de bit disponible para codificar el canal secundario X. Además, dependiendo del presupuesto de bits disponible, es posible limitar la reutilización de la información de tono para señales que tienen una característica de voz para los canales Y primario y X secundario.

Para este fin, la operación 1502 de análisis de coherencia de tono comprende una suboperación 1604 de decisión realizada por un módulo 1654 de decisión que considera el presupuesto de bits disponible y las características de la señal de sonido (indicada, por ejemplo, mediante los modos de codificación de canal primario y secundario). Cuando el módulo 1654 de decisión detecta que el presupuesto de bits disponible es suficiente o las señales de sonido para los canales X primario y secundario no tienen características de voz, la decisión es codificar la información de tono relacionada con el canal X secundario (1605).

Cuando el módulo 1654 de decisión detecta que el presupuesto de bits disponible es bajo con el fin de codificar la información de tono del canal secundario X o las señales de sonido para los canales X primario y secundario tienen una característica sonora, el módulo de decisión compara la coherencia de tono estéreo S_{pc} con el umbral Δ . Cuando el presupuesto de bits es bajo, el umbral Δ se establece en un valor mayor en comparación con el caso en el que el presupuesto de bits es más importante (suficiente para codificar la información de tono del canal secundario X). Cuando el valor absoluto de la coherencia de tono estéreo S_{pc} es menor o igual que el umbral Δ , el módulo 1654 decide reutilizar la información de tono del canal primario Y para codificar el canal secundario X (1607). Cuando el valor de la coherencia de tono estéreo S_{pc} es mayor que el umbral Δ , el módulo 1654 decide codificar la información de tono del canal secundario X (1605).

Asegurarse de que los canales tengan características sonoras aumenta la probabilidad de una evolución de tono suave, reduciendo así el riesgo de añadir artefactos reutilizando el tono del canal primario. Como ejemplo no limitativo, cuando el presupuesto de bit estéreo es inferior a 14 kb/s y la coherencia de tono estéreo S_{pc} es inferior o igual a 6 ($\Delta = 6$), la información de tono primario se puede reutilizar para codificar el canal secundario X. De acuerdo con otro ejemplo no limitativo, si el presupuesto de bits estéreo es superior a 14 kb/s e inferior a 26 kb/s, tanto el canal Y primario como el secundario X se consideran sonoros y se compara la coherencia de tono estéreo S_{pc} a un umbral más bajo $\Delta = 3$, lo que conduce a una tasa de reutilización más pequeña de la información de tono del canal primario Y a una velocidad de bits de 22 kb/s.

Con referencia de nuevo a la Figura 15, el estimador 1553 de asignación de bits recibe el factor β del mezclador 251/351 de canales, con la decisión de reutilizar los coeficientes de filtro LP del canal primario o de utilizar y codificar los coeficientes de filtro LP de canal secundario del analizador 856 de coherencia de filtro LP, y con la información de tono determinada por el analizador 1552 de coherencia de tono. Dependiendo de los requisitos de codificación de canal primario y secundario, el estimador 1553 de asignación de bits proporciona un presupuesto de bits para codificar el canal primario Y al codificador 252/352 de canal primario y un presupuesto de bits para codificar el canal secundario X al módulo 1556 de decisión. En una posible implementación, para todo el contenido que no es INACTIVO, se asigna una fracción de la velocidad de bits total al canal secundario. Entonces, la velocidad de bits del canal secundario aumentará en una cantidad relacionada con un factor ϵ de normalización de energía (reescalado) descrito anteriormente como:

$$B_x = B_M + (0.25 \cdot \varepsilon - 0.125) \cdot (B_t - 2 \cdot B_M) \quad (21a)$$

donde B_x representa la velocidad de bits asignada al canal secundario X, B_t representa la velocidad de bits estéreo total disponible, B_M representa la velocidad de bits mínima asignada al canal secundario y suele ser alrededor del 20% de la velocidad de bits estéreo total. Finalmente, ε representa el factor de normalización de energía descrito anteriormente. Por lo tanto, la velocidad de bits asignada al canal primario corresponde a la diferencia entre la velocidad de bits estéreo total y la velocidad de bits del canal secundario. En una implementación alternativa, la asignación de velocidad de bits del canal secundario puede describirse como:

$$B_x = \begin{cases} B_M + ((15 - \varepsilon_{idx}) \cdot (B_t - 2 \cdot B_M)) \cdot 0.05, & \text{si } \varepsilon_{idx} < 15 \\ B_M + ((\varepsilon_{idx} - 15) \cdot (B_t - 2 \cdot B_M)) \cdot 0.05, & \text{si } \varepsilon_{idx} \geq 15 \end{cases} \quad (21b)$$

donde nuevamente B_x representa la velocidad de bits asignada al canal secundario X, B_t representa la velocidad de bits estéreo total disponible y B_M representa la velocidad de bits mínima asignada al canal secundario. Finalmente, ε_{idx} representa un índice transmitido del factor de normalización de energía. Por lo tanto, la velocidad de bits asignada al canal primario corresponde a la diferencia entre la velocidad de bits estéreo total y la velocidad de bits del canal secundario. En todos los casos, para contenido INACTIVO, la velocidad de bits del canal secundario se establece en la velocidad de bits mínima necesaria para codificar la forma espectral del canal secundario, lo que da una velocidad de bits generalmente cercana a 2 kb/s.

Mientras tanto, el clasificador 852 de señal proporciona una clasificación de señal del canal secundario X al módulo 1554 de decisión. Si el módulo 1554 de decisión determina que la señal de sonido está inactiva o sin voz, el módulo 1555 de codificación voz/inactivo proporciona la forma espectral del canal secundario X al multiplexor 254/354. Alternativamente, el módulo 1554 de decisión informa al módulo 1556 de decisión cuando la señal de sonido no está inactiva ni sin voz. Para tales señales de sonido, utilizando el presupuesto de bits para codificar el canal secundario X, el módulo 1556 de decisión determina si hay un número suficiente de bits disponibles para codificar el canal secundario X utilizando el módulo 854 de codificación genérica solamente del modelo de cuatro (4) subtramas; de lo contrario, el módulo 1556 de decisión selecciona codificar el canal secundario X utilizando el módulo 855 que codifica el modelo de dos (2) subtramas. Para elegir el módulo de codificación solamente del modelo genérico de cuatro subtramas, el presupuesto de bits disponible para el canal secundario debe ser lo suficientemente alto como para asignar al menos 40 bits a los libros de códigos algebraicos, una vez que todo lo demás se cuantifica o reutiliza, incluido el coeficiente LP y la información y las ganancias de tono.

Como se entenderá de la descripción anterior, en la operación 804 de codificación solamente del modelo genérico de cuatro (4) subtramas y el módulo 854 de codificación solamente del modelo genérico de cuatro (4) subtramas correspondiente, para mantener la velocidad de bits lo más baja posible, se utiliza una búsqueda ACELP como se describe en la cláusula 5.2.3.1 de la Referencia [1]. En la codificación solamente del modelo genérico de cuatro (4) subtramas, la información de tono se puede reutilizar desde el canal primario o no. Los parámetros de codificación encontrados durante la búsqueda de ACELP en el módulo 854 de codificación solamente genérica del modelo de cuatro (4) subtramas se utilizan para construir el flujo 206/306 de bits del canal secundario y se envían al multiplexor 254/354 para su inclusión en el flujo 207/307 de bits multiplexado.

En la operación 805 de codificación del modelo de dos (2) subtramas alternativo y el correspondiente módulo 855 que codifica el modelo alternativo de dos (2) subtramas, el modelo de codificación genérico se construye de manera similar a ACELP como se describe en la cláusula 5.2.3.1 de la Referencia [1], pero se utiliza con solo dos (2) subtramas por trama. Por lo tanto, para hacerlo, la longitud de las subtramas se incrementa de 64 muestras a 128 muestras, manteniendo la frecuencia de muestreo interna a 12.8 kHz. Si el analizador 1552 de coherencia de tono ha decidido reutilizar la información de tono del canal primario Y para codificar el canal secundario X, entonces el promedio de los tonos de las dos primeras subtramas del canal primario Y se calcula y utiliza como estimación de tono para la primera mitad de trama del canal secundario X. Del mismo modo, el promedio de los tonos de las dos últimas subtramas del canal primario Y se calcula y utiliza para la segunda mitad de trama del canal secundario X. Cuando se reutiliza desde canal primario Y, los coeficientes del filtro LP se interpolan y la interpolación de los coeficientes del filtro LP como se describe en la cláusula 5.2.2.1 de la Referencia [1] se modifica para adaptarse a un esquema de dos (2) subtramas reemplazando los factores de interpolación primero y tercero con el segundo y cuarto factores de interpolación.

En la realización de la Figura 15, el proceso para decidir entre las cuatro (4) subtramas y el esquema de codificación de las dos (2) subtramas depende del presupuesto de bits disponible para codificar el canal secundario X. Como se mencionó anteriormente, el bit el presupuesto del canal secundario X se deriva de diferentes elementos, como el presupuesto total de bits disponible, el factor β o el factor ε de normalización de energía, la presencia o no de un módulo de corrección de retardo temporal (TDC), la posibilidad o no de reutilizar los coeficientes de filtro LP y/o la información de tono del canal primario Y.

La velocidad de bits mínima absoluta utilizada por los dos (2) modelos de codificación de subtramas del canal secundario X cuando tanto los coeficientes de filtro LP como la información de tono se reutilizan desde el canal primario Y es de alrededor de 2 kb/s para una señal genérica, mientras que es de alrededor de 3.6 kb/s para el esquema de codificación de cuatro (4) subtramas. Para un codificador similar a ACELP, que utiliza un modelo de codificación de

dos (2) o cuatro (4) subtramas, una gran parte de la calidad proviene del número de bits que se pueden asignar a la búsqueda de libro de códigos algebraico (ACB) como se define en cláusula 5.2.3.1.5 de la referencia [1].

Luego, para maximizar la calidad, la idea es comparar el presupuesto de bits disponible para las cuatro (4) subtramas de búsqueda de libro de códigos algebraico (ACB) y las dos (2) subtramas de búsqueda de libro de códigos algebraico (ACB) después de todo lo que será codificado se tiene en cuenta. Por ejemplo, si, para una trama específica, hay 4 kb/s (80 bits por trama de 20 ms) disponibles para codificar el canal secundario X y el coeficiente del filtro LP puede reutilizarse mientras la información de tono necesita ser transmitida. Luego se elimina de los 80 bits, la cantidad mínima de bits para codificar la señalización del canal secundario, la información de tono del canal secundario, las ganancias y el libro de códigos algebraico para las dos (2) subtramas y las cuatro (4) subtramas, para obtener el presupuesto de bits disponible para codificar el libro de códigos algebraico. Por ejemplo, el modelo de codificación de cuatro (4) subtramas se elige si hay al menos 40 bits disponibles para codificar el libro de códigos algebraico de cuatro (4) subtramas; de lo contrario, se utiliza el esquema de dos (2) subtramas.

3) Aproximación de la señal mono de un flujo de bits parcial

Como se describe en la descripción anterior, la mezcla reductora en el dominio del tiempo es mono amigable, lo que significa que en el caso de una estructura incorporada, el canal primario Y está codificado con un códec heredado (Debe tenerse en cuenta que, como se mencionó en la descripción anterior, se puede utilizar cualquier tipo de codificador adecuado como el codificador 252/352 de canal primario) y los bits estéreo se añaden al flujo de bits del canal primario, los bits estéreo podrían ser despojados y un decodificador heredado podría crear una síntesis subjetivamente cercana a una hipotética mono síntesis. Para hacerlo, se necesita una normalización de energía simple en el lado del codificador, antes de codificar el canal primario Y. Al reescalar la energía del canal primario Y a un valor suficientemente cercano a una energía de una versión de señal monofónica del sonido, decodificando el canal primario Y con un decodificador heredado puede ser similar a la decodificación del decodificador heredado de la versión de señal monofónica del sonido. La función de normalización de la energía está directamente relacionada con la diferencia $G'_{LR}(t)$ de correlación linealizada a largo plazo calculada utilizando la relación (7) y se calcula utilizando la relación (22):

$$\varepsilon = -0.485 \cdot G'_{LR}(t)^2 + 0.9765 \cdot G'_{LR}(t) + 0.5. \quad (22)$$

El nivel de normalización se muestra en la Figura 5. En la práctica, en lugar de utilizar la relación (22), se utiliza una tabla de búsqueda que relaciona los valores de normalización ε con cada valor posible del factor β (31 valores en esta realización de ejemplo). Incluso si este paso adicional no es necesario al codificar una señal de sonido estéreo, por ejemplo, voz y/o audio, con el modelo integrado, esto puede ser útil al decodificar solo la señal mono sin decodificar los bits estéreo.

4) Decodificación estéreo y mezcla de amplificación

La Figura 10 es un diagrama de bloques que ilustra simultáneamente un método de decodificación de sonido estéreo y un sistema de decodificación de sonido estéreo. La Figura 11 es un diagrama de bloques que ilustra características adicionales del método de decodificación de sonido estéreo y el sistema de decodificación de sonido estéreo de la Figura 10.

El método de decodificación de sonido estéreo de las figuras 10 y 11 comprende una operación 1007 de demultiplexación implementada por un demultiplexor 1057, una operación 1004 de decodificación de canal primario implementada por un decodificador 1054 de canal primario, una operación 1005 de decodificación de canal secundario implementada por un decodificador 1055 de canal secundario, y una operación 1006 de mezcla de amplificación de dominio en el tiempo implementada por un mezclador 1056 de amplificación de canal de dominio en el tiempo. La operación 1005 de decodificación de canal secundario comprende, como se muestra en la Figura 11, una operación 1101 de decisión implementada por un módulo 1151 de decisión, una operación 1102 de decodificación genérica de cuatro (4) subtramas implementada por un decodificador 1152 genérico de cuatro (4) subtramas y una operación 1103 de decodificación de dos (2) subtramas genérica/sin voz/inactivas implementadas por un decodificador 1153 genérico/sin voz/inactivo de dos (2) subtramas.

En la decodificación de sonido estéreo En un sistema, se recibe un flujo 1001 de bits desde un codificador. El demultiplexor 1057 recibe el flujo 1001 de bits y extrae de allí los parámetros de codificación del canal primario Y (flujo de 1002 bits), los parámetros de codificación del canal secundario X (flujo 1003 de bits) y el factor β suministrado al decodificador 1054 del canal primario, el decodificador 1055 del canal secundario y el mezclador 1056 de amplificación de canales. Como se mencionó anteriormente, el factor β se utiliza como un indicador tanto para el codificador 252/352 de canal primario como indicador para el codificador 253/353 de canal secundario para determinar la asignación de velocidad de bits, por lo tanto, el decodificador 1054 de canal primario y el decodificador 1055 de canal secundario están reutilizando el factor β para decodificar el flujo de bits correctamente.

Los parámetros de codificación del canal primario corresponden al modelo de codificación ACELP a la velocidad de bits recibida y podrían estar relacionados con un codificador EVS heredado o modificado (debe tenerse en cuenta aquí que, como se mencionó en la descripción anterior, cualquier tipo de codificador puede usarse como el codificador 252 de canal primario). El decodificador 1054 de canal primario se suministra con el flujo 1002 de bits para decodificar los parámetros de codificación del canal primario (modo de códec₁, β , LPC₁, Tono₁, índices de libro de códigos fijos₁ y

ganancias₁ como se muestra en la Figura 11) utilizando un método similar a la Referencia [1] a producir un canal primario decodificado Y'.

Los parámetros de codificación del canal secundario utilizados por el decodificador 1055 del canal secundario corresponden al modelo utilizado para codificar el segundo canal X y pueden comprender:

5 (a) El modelo de codificación genérico con reutilización de los coeficientes del filtro LP (LPC₁) y/u otros parámetros de codificación (como, por ejemplo, el desfase Tono₁) del canal primario Y. El decodificador 1152 genérico de cuatro (4) subtramas (Figura 11) del decodificador 1055 de canal secundario se suministra con los coeficientes de filtro LP (LPC₁) y/u otros parámetros de codificación (como, por ejemplo, el desfase Tono₁) del canal primario Y del decodificador 1054 y/o con el flujo 1003 de bits (β, Tono₂, índices fijos de libro de códigos₂ y ganancias₂ como se muestra en la Figura 11) y utiliza un método inverso al del módulo 854 de codificación (Figura 8) para producir el canal secundario decodificado X'.

15 (b) Otros modelos de codificación pueden o no reutilizar los coeficientes de filtro LP (LPC₁) y/u otros parámetros de codificación (como, por ejemplo, el desfase Tono₁) del canal primario Y, incluido el modelo de codificación genérico de media banda, el modelo de codificación sin voz de baja velocidad y el modelo de codificación inactivo de baja velocidad. Como ejemplo, el modelo de codificación inactivo puede reutilizar los coeficientes de filtro LP del canal primario LPC₁. El decodificador 1153 genérico/sin voz/inactivo de dos (2) subtramas (Figura 11) del decodificador 1055 de canal secundario se suministran con los coeficientes de filtro LP (LPC₁) y/u otros parámetros de codificación (como, por ejemplo, el desfase de Tono₁) del canal primario Y y/o con los parámetros de codificación del canal secundario del flujo 1003 de bits (modo códec₂, β, LPC₂, Tono₂, índices de libro de códigos fijos₂ y ganancias₂ como se muestra en la Figura 11) y utiliza métodos inversos a los del módulo 855 de codificación (Figura 8) para producir el canal secundario decodificado X'.

20 Los parámetros de codificación recibidos correspondientes al canal secundario X (flujo 1003 de bits) contienen información (modo códec₂) relacionada con el modelo de codificación que se está utilizando. El módulo 1151 de decisión utiliza esta información (modo códec₂) para determinar e indicar al decodificador 1152 genérico de cuatro (4) subtramas y el decodificador 1153 genérico/sin voz/inactivo de dos (2) subtramas cuyo modelo de codificación se va a utilizar.

25 En el caso de una estructura incorporada, el factor β se utiliza para recuperar el índice de escala de energía que se almacena en una tabla de búsqueda (no mostrada) en el lado del decodificador y se utiliza para reescalar el canal primario Y' antes de realizar la operación 1006 de mezcla de amplificación en el dominio del tiempo. Finalmente, el factor β se suministra al mezclador 1056 de amplificación de canal y se utiliza para mezclar por amplificación los canales Y' primarios decodificados y X' secundarios. La operación 1006 de mezcla de amplificación en el dominio del tiempo se realiza como la inversa de las relaciones (9) y (10) de mezcla reductora para obtener los canales decodificados derecho R' e izquierdo L', utilizando las relaciones (23) y (24):

$$L'(n) = \frac{\beta(t) \cdot Y'(n) - \beta(t) \cdot X'(n) + X'(n)}{2 \cdot \beta(t)^2 - 2 \cdot \beta(t) + 1}, \quad (23)$$

$$35 \quad R'(n) = \frac{-\beta(t) \cdot (Y'(n) + X'(n)) + Y'(n)}{2 \cdot \beta(t)^2 - 2 \cdot \beta(t) + 1}, \quad (24)$$

donde n = 0, ..., N-1 es el índice de la muestra en la trama y t es el índice de la trama.

5) Integración de la codificación en el dominio del tiempo y en el dominio de la frecuencia

40 Para aplicaciones de la presente técnica donde se utiliza un modo de codificación del dominio de la frecuencia, realizar la mezcla de reducción en el dominio el tiempo de la frecuencia para ahorrar algo de complejidad o para simplificar el flujo de datos también es contemplado. En tales casos, se aplica el mismo factor de mezcla a todos los coeficientes espectrales para mantener las ventajas de la mezcla reductora en el dominio del tiempo. Se puede observar que esto es una desviación de la aplicación de coeficientes espectrales por banda de frecuencia, como en el caso de la mayoría de las aplicaciones de mezcla reductora del dominio de frecuencia. El mezclador 456 reductor puede adaptarse para calcular relaciones (25.1) y (25.2):

$$45 \quad F_Y(k) = F_R(k) \cdot (1 - \beta(t)) + F_L(k) \cdot \beta(t) \quad (25.1)$$

$$F_X(k) = F_L(k) \cdot (1 - \beta(t)) - F_R(k) \cdot \beta(t) \quad (25.2)$$

50 donde F_R(k) representa un coeficiente de frecuencia k del canal derecho R y, de manera similar, F_L(k) representa un coeficiente de frecuencia k del canal izquierdo L. Los canales Y primario y X secundario se calculan aplicando una transformación de frecuencia inversa para obtener la representación en el tiempo de las señales mezcladas reductoras.

Las Figuras 17 y 18 muestran posibles implementaciones del método y sistema de codificación estéreo en el dominio del tiempo utilizando mezcla reductora en el dominio de la frecuencia capaz de cambiar entre la codificación en el dominio del tiempo y el dominio de la frecuencia de los canales Y primario y X secundario.

Una primera variante de dicho método y sistema se muestra en la Figura 17, que es un diagrama de bloques que ilustra el método y el sistema de codificación estéreo concurrente utilizando conmutación de reducción de dominio del tiempo con la capacidad de operar en el dominio del tiempo y en el dominio de frecuencia.

5 En la Figura 17, el método y sistema de codificación estéreo incluye muchas operaciones y módulos descritos previamente con referencia a figuras anteriores e identificados por los mismos números de referencia. Un módulo 1751 de decisión (operación 1701 de decisión) determina si los canales izquierdo L' y derecho R' del corrector 1750 de retardo temporal deben codificarse en el dominio del tiempo o en el dominio de la frecuencia. Si se selecciona la codificación en el dominio del tiempo, el método y el sistema de codificación estéreo de la Figura 17 funcionan sustancialmente de la misma manera que el método y el sistema de codificación estéreo de las figuras anteriores, por ejemplo y sin limitación como en la realización de la Figura 15.

10 Si el módulo 1751 de decisión selecciona la codificación de frecuencia, un convertidor 1752 de tiempo a frecuencia (operación 1702 de conversión de tiempo a frecuencia) convierte los canales izquierdo L' y derecho R' en dominio de frecuencia. Un mezclador 1753 reductor de dominio de frecuencia (operación 1703 de mezcla reductora de dominio de frecuencia) genera canales de dominio de frecuencia Y primario y X secundario. El canal primario del dominio de frecuencia se convierte de nuevo al dominio en el tiempo mediante un convertidor 1754 de frecuencia a tiempo (operación 1704 de conversión de frecuencia a tiempo) y el canal primario de dominio en el tiempo resultante Y se aplica al codificador 252/352 de canal primario. El canal secundario del dominio de frecuencia X del mezclador 1753 reductor del dominio de frecuencia se procesa a través de un codificador 1755 paramétrico y/o residual convencional (operación 1705 de codificación paramétrica y/o residual).

15 La Figura 18 es un diagrama de bloques que ilustra simultáneamente otro método y sistema de codificación estéreo que utiliza la mezcla reductora del dominio de frecuencia con la capacidad de operar en el dominio del tiempo y en el dominio de la frecuencia. En la Figura 18, el método y sistema de codificación estéreo son similares al método y sistema de codificación estéreo de la Figura 17 y solo se describirán las nuevas operaciones y módulos.

20 Un analizador 1851 de dominio en el tiempo (operación 1801 de análisis de dominio en el tiempo) reemplaza el mezclador 251/351 de canal de dominio en el tiempo descrito anteriormente (operación 201/301 de mezcla reductora de dominio en el tiempo). El analizador 1851 de dominio en el tiempo incluye la mayoría de los módulos de la Figura 4, pero sin el mezclador 456 reductor de dominio en el tiempo. Por lo tanto, su función es en gran parte proporcionar un cálculo del factor β . Este factor β se suministra al preprocesador 851 y a los convertidores 1852 y 1853 en el dominio de frecuencia a tiempo (operaciones 1802 y 1803 de conversión en el dominio de frecuencia a tiempo) que convierten respectivamente al dominio del tiempo los canales secundarios X y primarios Y del dominio de frecuencia recibidos del mezclador 1753 reductor en el dominio de frecuencia para la codificación en el dominio del tiempo para codificación de dominio de tiempo. La salida del convertidor 1852 es, por lo tanto, un canal secundario X en el dominio del tiempo que se proporciona al preprocesador 851, mientras que la salida del convertidor 1852 es un canal primario Y en el dominio del tiempo que se proporciona tanto al preprocesador 1551 como al codificador 252/352.

35 6) Ejemplo de configuración de hardware

La Figura 12 es un diagrama de bloques simplificado de una configuración de ejemplo de componentes de hardware que forman cada uno de los sistemas de codificación de sonido estéreo y sistema de decodificación de sonido estéreo descritos anteriormente.

40 Cada uno de los sistemas de codificación de sonido estéreo y sistema de decodificación de sonido estéreo puede implementarse como parte de un terminal móvil, como parte de un reproductor multimedia portátil o en cualquier dispositivo similar. Cada uno de los sistemas de codificación de sonido estéreo y sistema de decodificación de sonido estéreo (identificado como 1200 en la Figura 12) comprende una entrada 1202, una salida 1204, un procesador 1206 y una memoria 1208.

45 La entrada 1202 está configurada para recibir canales izquierdo L y derecho R de la señal de sonido estéreo de entrada en forma digital o analógica en el caso del sistema de codificación de sonido estéreo, o el flujo 1001 de bits en el caso del sistema de decodificación de sonido estéreo. La salida 1204 está configurada para suministrar el flujo 207/307 de bits multiplexado en el caso del sistema de codificación de sonido estéreo o el canal izquierdo L' y el canal derecho R' decodificados en el caso del sistema de decodificación de sonido estéreo. La entrada 1202 y la salida 1204 pueden implementarse en un módulo común, por ejemplo, un dispositivo de entrada/salida en serie.

50 El procesador 1206 está conectado operativamente a la entrada 1202, a la salida 1204 y a la memoria 1208. El procesador 1206 se realiza como uno o más procesadores para ejecutar instrucciones de código en apoyo de las funciones de los diversos módulos de cada uno de los sistemas de codificación de sonido estéreo como se muestra en las Figuras 2, 3, 4, 8, 9, 13, 14, 15, 16, 17 y 18 y el sistema de decodificación de sonido estéreo como se muestra en las Figuras 10 y 11.

55 La memoria 1208 puede comprender una memoria no transitoria para almacenar instrucciones de código ejecutables por el procesador 1206, específicamente, una memoria legible por procesador que comprende instrucciones no transitorias que, cuando se ejecutan, hacen que un procesador implemente las operaciones y módulos del método y sistema de codificación de sonido estéreo y el método y sistema de decodificación de sonido estéreo como se describe

en la presente divulgación. La memoria 1208 también puede comprender una memoria de acceso aleatorio o memoria intermedia (s) para almacenar datos de procesamiento intermedios de las diversas funciones realizadas por el procesador 1206.

5 Los expertos en la materia se darán cuenta de que la descripción del método y sistema de codificación de sonido estéreo y el método y sistema de decodificación de sonido estéreo son solo ilustrativos y no pretenden ser de ningún modo limitantes. Otras realizaciones se sugerirán fácilmente a tales personas con habilidades ordinarias en la técnica que tengan el beneficio de la presente divulgación. Además, el método y sistema de codificación de sonido estéreo divulgado y el método y sistema de decodificación de sonido estéreo pueden personalizarse para ofrecer soluciones valiosas a las necesidades y problemas existentes de codificación y decodificación de sonido estéreo.

10 En aras de la claridad, no se muestran y describen todas las características rutinarias de las implementaciones del método y sistema de codificación de sonido estéreo y el método y sistema de decodificación de sonido estéreo. Por supuesto, se apreciará que en el desarrollo de dicha implementación real del método y sistema de codificación de sonido estéreo y el método y sistema de decodificación de sonido estéreo, es posible que se deban tomar numerosas decisiones específicas de implementación para lograr los objetivos específicos del desarrollador, como el cumplimiento de las restricciones relacionadas con la aplicación, el sistema, la red y el negocio, y que estos objetivos específicos variarán de una implementación a otra y de un desarrollador a otro. Además, se apreciará que un esfuerzo de desarrollo puede ser complejo y lento, pero sin embargo sería una tarea rutinaria de ingeniería para los expertos en el campo del procesamiento de sonido que tiene el beneficio de la presente divulgación.

20 De acuerdo con la presente divulgación, los módulos, las operaciones de procesamiento y/o las estructuras de datos descritas en este documento pueden implementarse utilizando diversos tipos de sistemas operativos, plataformas informáticas, dispositivos de red, programas informáticos y/o máquinas de uso general. Además, los expertos en la materia reconocerán que también se pueden utilizar dispositivos de una naturaleza de propósito menos general, como dispositivos cableados, matrices de puertas programables en campo (FPGA), circuitos integrados específicos de aplicaciones (ASIC) o similares. Cuando un método que comprende una serie de operaciones y suboperaciones se implementa un procesador, ordenador o máquina y esas operaciones y suboperaciones pueden almacenarse como una serie de instrucciones de código no transitorias legibles por el procesador, ordenador o máquina, se pueden almacenar en un medio tangible y/o no transitorio.

25 Los módulos del método y sistema de codificación de sonido estéreo y el método y decodificador de decodificación de sonido estéreo como se describe en el presente documento pueden comprender software, firmware, hardware o cualquier combinación de software, firmware o hardware adecuado para los fines descritos en este documento.

30 En el método de codificación de sonido estéreo y el método de decodificación de sonido estéreo como se describe en el presente documento, las diversas operaciones y suboperaciones pueden realizarse en varios órdenes y algunas de las operaciones y suboperaciones pueden ser opcionales.

35 Aunque la presente divulgación se ha descrito anteriormente a modo de realización ilustrativa no restrictiva de la misma, estas realizaciones pueden modificarse a voluntad dentro del alcance de las reivindicaciones adjuntas sin apartarse del espíritu y la naturaleza de la presente divulgación.

Referencias

Las siguientes referencias se mencionan en la presente especificación.

40 [1] 3GPP TS 26.445, v.12.0.0, "Codec for Enhanced Voice Services (EVS); Detailed Algorithmic Description", Sep 2014.

[2] M. Neuendorf, M. Multrus, N. Rettelbach, G. Fuchs, J. Robillard, J. Lecompte, S. Wilde, S. Bayer, S. Disch, C. Helmrich, R. Lefebvre, P. Gournay, et al., "The ISO/MPEG Unified Speech and Audio Coding Standard - Consistent High Quality for All Content Types and at All Bit Rates", J. Audio Eng. Soc., vol. 61, no. 12, pp. 956-977, Dec. 2013.

45 [3] B. Bessette, R. Salami, R. Lefebvre, M. Jelinek, J. Rotola-Pukkila, J. Vainio, H. Mikkola, and K. Jarvinen, "The Adaptive Multi-Rate Wideband Speech Codec (AMR-WB)," Special Issue of IEEE Trans. Speech and Audio Proc., Vol. 10, pp.620-636, November 2002.

[4] R.G. van der Waal & R.N.J. Veldhuis, "Subband coding of stereophonic digital audio signals", Proc. IEEE ICASSP, Vol. 5, pp. 3601-3604, April 1991.

50 [5] Dai Yang, Hongmei Ai, Chris Kyriakakis and C.-C. Jay Kuo, "High-Fidelity Multichannel Audio Coding With Karhunen-Loève Transform", IEEE Trans. Speech and Audio Proc., Vol. 11, No.4, pp.365-379, July 2003.

[6] J. Breebaart, S. van de Par, A. Kohlrausch and E. Schuijers, "Parametric Coding of Stereo Audio", EURASIP Journal on Applied Signal Processing, Issue 9, pp. 1305-1322, 2005.

[7] 3GPP TS 26.290 V9.0.0, "Extended Adaptive Multi-Rate - Wideband (AMR-WB+) codec; Transcoding functions (Release 9)", September 2009.

[8] Jonathan A. Gibbs, "Apparatus and method for encoding a multi-channel audio signal", US 8577045 B2.

[9] Michael M. Goodwin et al: "Primary-ambient decomposition of stereo audio signals using a complex similarity index, US2009/0198356A1".

La siguiente es una descripción adicional que muestra otros posibles ejemplos útiles para la implementación.

5 Un método de codificación de sonido estéreo para codificar canales izquierdo y derecho de una señal de sonido estéreo, comprende: mezclar por reducción en el dominio del tiempo los canales izquierdo y derecho de la señal de sonido estéreo para producir canales primarios y secundarios; codificar el canal primario y codificar el canal secundario, en el que codificar el canal primario y codificar el canal secundario comprende seleccionar una primera velocidad de bits para codificar el canal primario y una segunda velocidad de bits para codificar el canal secundario, en el que la primera y segunda velocidades de bits se seleccionan en función de un nivel de énfasis que se dará a los canales primario y secundario; codificar el canal secundario comprende calcular los coeficientes del filtro LP en respuesta al canal secundario y analizar la coherencia entre los coeficientes del filtro LP calculados durante la codificación del canal secundario y los coeficientes del filtro LP calculados durante la codificación del canal primario para decidir si los coeficientes del filtro LP calculados durante el primario la codificación del canal está suficientemente cerca de los coeficientes del filtro LP calculados durante la codificación del canal secundario para ser reutilizados durante la codificación del canal secundario.

El método de codificación de sonido estéreo como se describe en el párrafo anterior puede comprender, en combinación, al menos una de las siguientes características (a) a (l).

20 (a) Decidir si los parámetros distintos de los coeficientes del filtro LP y calculados durante la codificación del canal primario están suficientemente cerca de los parámetros correspondientes calculados durante la codificación del canal secundario para ser reutilizados durante la codificación del canal secundario.

(b) Codificar el canal secundario comprende utilizar un número mínimo de bits para codificar el canal secundario; y codificar el canal primario comprende usar, para codificar el canal primario, todos los bits restantes que no se han usado para codificar el canal secundario.

25 (c) Codificar el canal secundario comprende utilizar una primera velocidad de bits fija para codificar el canal primario; y codificar el canal primario comprende utilizar una segunda velocidad de bits fija, más baja que la primera velocidad de bits, para codificar el canal secundario.

(d) Una suma de las velocidades de bits primera y segunda es igual a una velocidad de bits total constante.

30 (e) Analizar la coherencia entre los coeficientes del filtro LP calculados durante la codificación del canal secundario y los coeficientes del filtro LP calculados durante la codificación del canal primario comprende: determinar una distancia euclidiana entre los primeros parámetros representativos de los coeficientes del filtro LP calculados durante la codificación del canal primario y los segundos parámetros representativos de los coeficientes de filtro LP calculados durante la codificación del canal secundario; y comparando la distancia euclidiana a un primer umbral.

35 (f) El análisis de la coherencia entre los coeficientes del filtro LP calculados durante la codificación del canal secundario y los coeficientes del filtro LP calculados durante la codificación del canal primario comprende además: producir un primer residuo del canal secundario utilizando los coeficientes del filtro LP calculados durante la codificación del canal primario, y producir un segundo residuo del canal secundario utilizando los coeficientes de filtro LP calculados durante la codificación del canal secundario; producir una primera ganancia de predicción utilizando el primer residuo y producir una segunda ganancia de predicción utilizando el segundo residuo; calcular una relación entre las ganancias de predicción primera y segunda; comparar la relación con un segundo umbral.

40 (g) Analizar la coherencia entre los coeficientes del filtro LP calculados durante la codificación del canal secundario y los coeficientes del filtro LP calculados durante la codificación del canal primario comprende además: decidir, en respuesta a dichas comparaciones, si los coeficientes del filtro LP calculados durante la codificación del canal primario son suficientemente cerca de los coeficientes del filtro LP calculados durante la codificación del canal secundario para ser reutilizados durante la codificación del canal secundario.

(h) El primer y el segundo parámetro son pares espectrales de línea.

45 (i) Producir la primera ganancia de predicción comprende calcular una energía del primer residual, calcular una energía del sonido en el canal secundario y restar la energía del primer residual de la energía del sonido en el canal secundario; y producir la segunda ganancia de predicción comprende calcular una energía del segundo residual, calcular la energía del sonido en el canal secundario y restar la energía del segundo residual de la energía del sonido en el canal secundario.

(j) La codificación del canal secundario comprende clasificar el canal secundario y utilizar un modelo de codificación CELP de cuatro subtramas cuando el canal secundario se clasifica como genérico y la decisión es reutilizar los coeficientes del filtro LP calculados durante la codificación del canal primario para codificar el canal secundario.

(k) La codificación del canal secundario comprende la clasificación del canal secundario y el uso de un modelo de codificación de baja velocidad de dos subtramas cuando el canal secundario se clasifica como inactivo, sin voz o genérico y la decisión es no reutilizar los coeficientes del filtro LP calculados durante la codificación del canal primario para codificar el canal secundario.

- 5 (l) Una energía del canal primario se reescala a un valor suficientemente cercano a una energía de una versión de señal monofónica del sonido, de modo que la decodificación del canal primario con un decodificador heredado es similar a la decodificación del decodificador heredado de la versión de señal monofónica del sonido.

10 Un sistema de codificación de sonido estéreo para codificar canales izquierdo y derecho de una señal de sonido estéreo, comprende: un mezclador de reducción en el dominio del tiempo de los canales izquierdo y derecho de la señal de sonido estéreo para producir canales primarios y secundarios; un codificador del canal primario y un codificador del canal secundario, en el que el codificador del canal primario y el codificador del canal secundario seleccionan una primera velocidad de bits para codificar el canal primario y una segunda velocidad de bits para codificar el canal secundario, en el que la primera y segunda velocidades de bits dependen del nivel de énfasis que se otorgue a los canales primario y secundario; el codificador del canal secundario comprende un analizador de filtro LP para calcular los coeficientes del filtro LP en respuesta al canal secundario y un analizador de la coherencia entre los coeficientes del filtro LP del canal secundario y los coeficientes del filtro LP calculados en el codificador del canal primario para decidir si los coeficientes de filtro LP de canal primario están lo suficientemente cerca del coeficiente de filtro LP del canal secundario para ser reutilizados por el codificador del canal secundario.

20 El sistema de codificación de sonido estéreo como se describió en el párrafo anterior puede comprender, en combinación, al menos una de las siguientes características (1) a (12).

(1) El codificador de canal secundario decide además si los parámetros distintos de los coeficientes de filtro LP y calculados en el codificador de canal primario están suficientemente cerca de los parámetros correspondientes calculados en el codificador de canal secundario para ser reutilizados por el codificador de canal secundario.

25 (2) El codificador del canal secundario utiliza un número mínimo de bits para codificar el canal secundario, y el codificador del canal primario usa, para codificar el canal primario, todos los bits restantes que no han sido utilizados por el codificador del canal secundario para codificar el canal secundario.

(3) El codificador del canal secundario utiliza una primera velocidad de bits fija para codificar el canal primario, y el codificador del canal primario utiliza una segunda velocidad de bits fija, más baja que la primera velocidad de bits, para codificar el canal secundario.

30 (4) Una suma de las velocidades de bits primera y segunda es igual a una velocidad de bits total constante.

35 (5) El analizador de la coherencia entre los coeficientes del filtro LP del canal secundario y los coeficientes del filtro LP del canal primario comprende: un analizador de distancia euclidiana para determinar una distancia euclidiana entre los primeros parámetros representativos de los coeficientes del filtro LP del canal primario y los segundos parámetros representativos de coeficientes de filtro LP de canal secundario; y un comparador de la distancia euclidiana a un primer umbral.

40 (6) El analizador de la coherencia entre los coeficientes del filtro LP del canal secundario y los coeficientes del filtro LP del canal primario comprende: un primer filtro e residuo para producir un primer residuo del canal secundario utilizando los coeficientes del filtro LP del canal primario, y un segundo filtro de residuo para producir un segundo residuo del canal secundario utilizando los coeficientes de filtro LP del canal secundario; medios para producir una primera ganancia de predicción utilizando el primer residuo y medios para producir una segunda ganancia de predicción utilizando el segundo residuo; un calculador de una relación entre las ganancias de predicción primera y segunda; y un comparador de la relación a un segundo umbral.

45 (7) El analizador de la coherencia entre los coeficientes del filtro LP del canal secundario y los coeficientes del filtro LP del canal primario comprende, además: un módulo de decisión para decidir, en respuesta a las comparaciones, si los coeficientes del filtro LP del canal primario están suficientemente cerca del coeficiente de filtro LP de canal secundario para ser reutilizados por el codificador de canal secundario.

(8) El primer y el segundo parámetros son pares espectrales de línea.

50 (9) Los medios para producir la primera ganancia de predicción comprenden un calculador de una energía del primer residuo, un calculador de una energía del sonido en el canal secundario y un sustractor de la energía del primer residuo de la energía del sonido en el canal secundario; y los medios para producir la segunda ganancia de predicción comprenden un calculador de una energía del segundo residuo, el calculador de la energía del sonido en el canal secundario y un sustractor de la energía del segundo residuo de la energía del sonido en el canal secundario.

55 (10) El codificador del canal secundario comprende un clasificador del canal secundario y un módulo de codificación que utiliza un modelo de codificación CELP de cuatro subtramas cuando el canal secundario se clasifica como genérico y la decisión es reutilizar los coeficientes del filtro LP del canal primario para codificar el canal secundario.

(11) El codificador del canal secundario comprende un clasificador del canal secundario y un módulo de codificación que utiliza un modelo de codificación de dos subtramas cuando el canal secundario se clasifica como inactivo, sin voz o genérico y la decisión es no reutilizar los coeficientes de filtro LP de canal primario para codificar el canal secundario.

5 (12) Se proporcionan medios para reescalar una energía del canal primario a un valor suficientemente cercano a una energía de una versión de señal monofónica del sonido, de modo que la decodificación del canal primario con un decodificador heredado sea similar a la decodificación del decodificador heredado de la versión de señal monofónica del sonido.

10 Un sistema de codificación de sonido estéreo para codificar canales izquierdo y derecho de una señal de sonido estéreo, comprende: al menos un procesador; y una memoria acoplada al procesador y que comprende instrucciones no transitorias que, cuando se ejecutan, hacen que el procesador implemente: un mezclador de reducción en el dominio del tiempo de los canales izquierdo y derecho de la señal de sonido estéreo para producir canales primarios y secundarios; un codificador del canal primario y un codificador del canal secundario, en el que el codificador del canal primario y el codificador del canal secundario seleccionan una primera velocidad de bits para codificar el canal primario y una segunda velocidad de bits para codificar el canal secundario, en el que el primer y las segundas
15 velocidades de bits dependen del nivel de énfasis que se otorgue a los canales primario y secundario; el codificador del canal secundario comprende un analizador de filtro LP para calcular los coeficientes del filtro LP en respuesta al canal secundario y un analizador de la coherencia entre los coeficientes del filtro LP del canal secundario y los coeficientes del filtro LP calculados en el codificador del canal primario para decidir si los coeficientes de filtro LP de canal primario están lo suficientemente cerca del coeficiente de filtro LP del canal secundario para ser reutilizados por
20 el codificador del canal secundario.

REIVINDICACIONES

1. Un método de codificación de sonido estéreo para codificar canales izquierdo y derecho de una señal de sonido estéreo, que comprende:
- 5 mezclar por reducción los canales izquierdo y derecho de la señal de sonido estéreo para producir canales primarios y secundarios; y
- codificar el canal primario y codificar el canal secundario;
- en el que la codificación del canal secundario comprende analizar la coherencia entre los parámetros de codificación calculados durante la codificación del canal secundario y los parámetros de codificación calculados durante la codificación del canal primario para decidir si los parámetros de codificación calculados durante la codificación del canal primario están suficientemente cerca de los parámetros de codificación calculados durante la codificación de canal secundario para ser reutilizada durante la codificación del canal secundario.
- 10
2. Un método de codificación de sonido estéreo como se define en la reivindicación 1, en el que la mezcla reductora de los canales izquierdo y derecho de la señal de sonido estéreo comprende la mezcla reductora en el dominio de tiempo de los canales izquierdo y derecho de la señal de sonido estéreo para producir los canales primario y secundario.
- 15
3. Un método de codificación de sonido estéreo como se define en la reivindicación 1 o 2, en el que codificar el canal primario y codificar el canal secundario comprende seleccionar una primera velocidad de bits para codificar el canal primario y una segunda velocidad de bits para codificar el canal secundario, en el que la primera y la segunda velocidad de bits se seleccionan de acuerdo con el nivel de énfasis que se otorgue a los canales primario y secundario.
- 20
4. Un método de codificación de sonido estéreo como se define en una cualquiera de las reivindicaciones 1 a 3, en el que:
- codificar el canal secundario comprende utilizar un número mínimo de bits para codificar el canal secundario, y
- codificar el canal primario comprende usar, para codificar el canal primario, todos los bits restantes que no se han utilizado para codificar el canal secundario.
- 25
5. Un método de codificación de sonido estéreo como se define en una cualquiera de las reivindicaciones 1 a 3, en el que:
- codificar el canal primario comprende utilizar una primera velocidad de bits fija para codificar el canal primario, y
- codificar el canal secundario comprende utilizar una segunda velocidad de bits fija, inferior a la primera velocidad de bits, para codificar el canal secundario.
- 30
6. Un método de codificación de sonido estéreo como se define en una cualquiera de las reivindicaciones 3 a 5, en el que una suma de la primera y la segunda velocidad de bits es igual a una velocidad de bits total constante.
7. Un método de codificación de sonido estéreo como se define en una cualquiera de las reivindicaciones 3 a 6, en el que los parámetros de codificación comprenden coeficientes de filtro LP y en el que se analiza la coherencia entre los coeficientes de filtro LP calculados durante la codificación del canal secundario y los coeficientes de filtro LP calculados durante la codificación del canal primario comprende:
- 35
- determinar una distancia euclidiana entre los primeros parámetros representativos de los coeficientes del filtro LP calculados durante la codificación del canal primario y los segundos parámetros representativos de los coeficientes del filtro LP calculados durante la codificación del canal secundario; y
- comparar la distancia euclidiana a un primer umbral.
- 40
8. Un método de codificación de sonido estéreo como se define en la reivindicación 7, en el que analizar la coherencia entre los coeficientes del filtro LP calculados durante la codificación del canal secundario y los coeficientes del filtro LP calculados durante la codificación del canal primario comprende:
- producir un primer residuo del canal secundario utilizando los coeficientes de filtro LP calculados durante la codificación del canal primario, y producir un segundo residuo del canal secundario utilizando los coeficientes de filtro LP calculados durante la codificación del canal secundario;
- 45
- producir una primera ganancia de predicción utilizando el primer residuo y producir una segunda ganancia de predicción utilizando el segundo residual;
- calcular una relación entre las ganancias de predicción primera y segunda;
- comparar la relación con un segundo umbral.

9. Un método de codificación de sonido estéreo como se define en la reivindicación 8, en el que analizar la coherencia entre los coeficientes del filtro LP calculados durante la codificación del canal secundario y los coeficientes del filtro LP calculados durante la codificación del canal primario comprende:
- 5 decidir, en respuesta a dichas comparaciones, si los coeficientes de filtro LP calculados durante la codificación del canal primario están suficientemente cerca de los coeficientes del filtro LP calculados durante la codificación del canal secundario para ser reutilizados durante la codificación del canal secundario.
10. Un método de codificación de sonido estéreo como se define en la reivindicación 8 o 9, en el que:
- 10 producir la primera ganancia de predicción comprende calcular una energía del primer residuo, calcular una energía del sonido en el canal secundario y restar la energía del primer residuo de la energía del sonido en el canal secundario; y
- 10 producir la segunda ganancia de predicción comprende calcular una energía del segundo residuo, calcular la energía del sonido en el canal secundario y restar la energía del segundo residuo de la energía del sonido en el canal secundario.
- 15 11. Un método de codificación de sonido estéreo como se define en una cualquiera de las reivindicaciones 3 a 10, en el que los parámetros de codificación comprenden coeficientes de filtro LP y en el que la codificación del canal secundario comprende clasificar el canal secundario y utilizar un modelo de codificación CELP de cuatro subtramas cuando el canal secundario se clasifica como genérico y la decisión es reutilizar los coeficientes del filtro LP calculados durante la codificación del canal primario para codificar el canal secundario.
- 20 12. Un método de codificación de sonido estéreo como se define en una cualquiera de las reivindicaciones 3 a 10, en el que los parámetros de codificación comprenden coeficientes de filtro LP y en el que la codificación del canal secundario comprende clasificar el canal secundario y utilizar un modelo de codificación de baja velocidad de dos subtramas cuando el canal secundario se clasifica como inactivo, sin voz o genérico y la decisión es no reutilizar los coeficientes de filtro LP calculados durante la codificación del canal primario para codificar el canal secundario.
- 25 13. Un método de codificación de sonido estéreo como se define en una cualquiera de las reivindicaciones 1 a 12, que comprende reescalar una energía del canal primario a un valor suficientemente cercano a una energía de una versión de señal monofónica del sonido, de modo que la decodificación del canal primario con un decodificador heredado es similar a la decodificación del decodificador heredado de la versión de señal monofónica del sonido.
- 30 14. El método de codificación de sonido estéreo como se define en una cualquiera de las reivindicaciones 3 a 13, en el que:
- 30 los parámetros de codificación comprenden información de tono;
- 30 analizar la coherencia entre la información de tono calculada durante la codificación del canal secundario y la información de tono calculada durante la codificación del canal primario comprende calcular una coherencia de tonos de bucle abierto de los canales primario y secundario; y
- 35 codificar el canal secundario comprende (a) reutilizar la información de tono del canal primario para codificar el canal secundario cuando la coherencia de tono es inferior o igual a un umbral; y (b) codificar la información de tono del canal secundario cuando la coherencia de tono es mayor que el umbral.
- 40 15. El método de codificación de sonido estéreo como se define en la reivindicación 14, en el que el cálculo de la coherencia de los tonos de bucle abierto de los canales primario y secundario comprende (a) sumar los tonos de bucle abierto del canal primario, (b) sumar los tonos de bucle abierto del canal secundario, y (c) restar la suma de los tonos de bucle abierto del canal secundario de la suma de los tonos de bucle abierto del canal primario para obtener la coherencia de tono.
- 45 16. El método de codificación de sonido estéreo como se define en la reivindicación 14 o 15, que comprende:
- 45 detectar un presupuesto de bits disponible para codificar la información de tono del canal secundario;
- 45 detectar una característica de voz de los canales primario y secundario; y
- 45 reutilizar la información de tono del canal primario para codificar el canal secundario cuando el presupuesto de bits disponible es bajo con el fin de codificar la información de tono del canal secundario, cuando se detecta una característica sonora de los canales primario y secundario, y cuando la coherencia de tono es inferior o igual al umbral.
- 50 17. El método de codificación de sonido estéreo como se define en la reivindicación 16, que comprende establecer el umbral a un valor mayor cuando el presupuesto de bits disponible es bajo con el fin de codificar la información de tono del canal secundario y/o cuando una característica de voz del primario y se detectan canales secundarios.

18. El método como se define en una cualquiera de las reivindicaciones 1 a 17, en el que, cuando el canal secundario se clasifica como inactivo o sin voz, proporciona una forma espectral del canal secundario solo para codificar el canal secundario.

19. El método como se define en una cualquiera de las reivindicaciones 1 a 18, que comprende:

5 convertir los canales izquierdo y derecho desde el dominio en el tiempo hasta el dominio de frecuencia; y

mezclar por reducción el dominio de frecuencia los canales izquierdo y derecho del dominio de frecuencia para producir canales primarios y secundarios del dominio de frecuencia.

20. Un sistema de codificación de sonido estéreo para codificar canales izquierdo y derecho de una señal de sonido estéreo, que comprende:

10 un mezclador reductor de los canales izquierdo y derecho de la señal de sonido estéreo para producir canales primarios y secundarios; y

un codificador del canal primario y un codificador del canal secundario;

15 en el que el codificador de canal secundario comprende un analizador de coherencia entre los parámetros de codificación de canal secundario calculados durante la codificación de canal secundario y los parámetros de codificación de canal primario calculados durante la codificación de canal primario para decidir si los parámetros de codificación de canal primario están adecuadamente cerca de los parámetros de codificación del canal secundario a ser reutilizado durante la codificación del canal secundario.

20 21. Un sistema de codificación de sonido estéreo como se define en la reivindicación 20, en el que el mezclador reductor es un mezclador reductor en el dominio del tiempo de los canales izquierdo y derecho de la señal de sonido estéreo.

22. Un sistema de codificación de sonido estéreo como se define en la reivindicación 20 o 21, en el que el codificador de canal primario y el codificador de canal secundario seleccionan una primera velocidad de bits para codificar el canal primario y una segunda velocidad de bits para codificar el canal secundario, en el que las velocidades de bits primera y segunda se seleccionan dependiendo del nivel de énfasis que se dará a los canales primario y secundario.

25 23. Un sistema de codificación de sonido estéreo como se define en una cualquiera de las reivindicaciones 20 a 22, en el que:

el codificador del canal secundario utiliza un número mínimo de bits para codificar el canal secundario, y

el codificador del canal primario usa, para codificar el canal primario, todos los bits restantes que no han sido utilizados por el codificador del canal secundario para codificar el canal secundario.

30 24. Un sistema de codificación de sonido estéreo como se define en una cualquiera de las reivindicaciones 20 a 23, en el que:

el codificador de canal primario utiliza una primera velocidad de bits fija para codificar el canal primario; y

el codificador del canal secundario utiliza una segunda velocidad de bits fija, más baja que la primera velocidad de bits, para codificar el canal secundario.

35 25. Un sistema de codificación de sonido estéreo como se define en una cualquiera de las reivindicaciones 22 a 24, en el que una suma de la primera y la segunda velocidad de bits es igual a una velocidad de bits total constante.

26. Un sistema de codificación de sonido estéreo como se define en una cualquiera de las reivindicaciones 22 a 25, en el que los parámetros de codificación comprenden coeficientes de filtro LP y en el que el analizador de la coherencia entre los coeficientes de filtro LP del canal secundario y los coeficientes de filtro LP del canal primario comprende:

40 un analizador de distancia euclidiana para determinar una distancia euclidiana entre los primeros parámetros representativos de los coeficientes del filtro LP del canal primario y los segundos parámetros representativos de los coeficientes del filtro LP del canal secundario; y

un comparador de la distancia euclidiana a un primer umbral.

45 27. Un sistema de codificación de sonido estéreo como se define en la reivindicación 26, en el que el analizador de la coherencia entre los coeficientes del filtro LP del canal secundario y los coeficientes del filtro LP del canal primario comprende:

un primer filtro residuo para producir un primer residuo del canal secundario utilizando los coeficientes de filtro LP del canal primario, y un segundo filtro residuo para producir un segundo residuo del canal secundario utilizando los coeficientes de filtro LP del canal secundario;

un calculador de una primera ganancia de predicción utilizando el primer residuo; y
 un calculador de una segunda ganancia de predicción utilizando el segundo residuo;
 un calculador de una relación entre las ganancias de predicción primera y segunda; y
 un comparador de la relación a un segundo umbral.

- 5 28. Un sistema de codificación de sonido estéreo como se define en la reivindicación 27, en el que el analizador de la coherencia entre los coeficientes del filtro LP del canal secundario y los coeficientes del filtro LP del canal primario comprende, además:

un módulo de decisión para decidir, en respuesta a las comparaciones, si los coeficientes del filtro LP del canal primario están suficientemente cerca de los coeficientes del filtro LP del canal secundario para ser reutilizados por el codificador del canal secundario.

10

29. Un sistema de codificación de sonido estéreo como se define en la reivindicación 27 o 28, en el que:

el calculador de la primera ganancia de predicción comprende un calculador de una energía del primer residuo, un calculador de una energía del sonido en el canal secundario y un sustractor de la energía del primer residuo de la energía del sonido en el canal secundario; y

- 15 el calculador de la segunda ganancia de predicción comprende un calculador de una energía del segundo residuo, el calculador de la energía del sonido en el canal secundario y un sustractor de la energía del segundo residuo de la energía del sonido en el canal secundario.

- 20 30. Un sistema de codificación de sonido estéreo como se define en una cualquiera de las reivindicaciones 20 a 29, en el que los parámetros de codificación comprenden coeficientes de filtro LP y en el que el codificador de canal secundario comprende un clasificador del canal secundario y un módulo de codificación que utiliza un modelo de codificación CELP de cuatro subtramas cuando el canal secundario se clasifica como genérico y la decisión es reutilizar los coeficientes de filtro LP del canal primario para codificar el canal secundario.

- 25 31. Un sistema de codificación de sonido estéreo como se define en una cualquiera de las reivindicaciones 20 a 29, en el que los parámetros de codificación comprenden coeficientes de filtro LP y en el que el codificador de canal secundario comprende un clasificador del canal secundario y un módulo de codificación que utiliza un modelo de codificación de dos subtramas cuando el canal secundario se clasifica como inactivo, sin voz o genérico y la decisión es no reutilizar los coeficientes de filtro LP del canal primario para codificar el canal secundario.

- 30 32. Un sistema de codificación de sonido estéreo como se define en una cualquiera de las reivindicaciones 20 a 31, que comprende medios para reescalar una energía del canal primario a un valor suficientemente cercano a una energía de una versión de señal monofónica del sonido, de modo que la decodificación del canal primario con un decodificador heredado es similar a la decodificación del decodificador heredado de la versión de señal monofónica del sonido.

33. El sistema de codificación de sonido estéreo como se define en una cualquiera de las reivindicaciones 22 a 32, en el que:

los parámetros de codificación comprenden información de tono;

- 35 el analizador de coherencia de tono calcula una coherencia de tonos de bucle abierto de los canales primario y secundario; y

el codificador de canal secundario (a) reutiliza la información de tono del canal primario para codificar el canal secundario cuando la coherencia de tono es inferior o igual a un umbral; y (b) codifica la información de tono del canal secundario cuando la coherencia de tono es mayor que el umbral.

- 40 34. El sistema de codificación de sonido estéreo como se define en la reivindicación 33, en el que, para calcular la coherencia de los tonos de bucle abierto de los canales primario y secundario, el analizador de coherencia de tono comprende (a) un sumador de tonos de bucle abierto del canal primario, (b) un sumador de tonos de bucle abierto del canal secundario, y (c) un sustractor de la suma de los tonos de bucle abierto del canal secundario de la suma de los tonos de bucle abierto del canal primario para obtener la coherencia de tono.

- 45 35. El sistema de codificación de sonido estéreo como se define en la reivindicación 33 o 34, en el que:

el analizador de coherencia de tono detecta un presupuesto de bits disponible para codificar la información de tono del canal secundario y detecta una característica sonora de los canales primario y secundario; y

el codificador de canal secundario reutiliza la información de tono del canal primario para codificar el canal secundario cuando el presupuesto de bits disponible es bajo con el fin de codificar la información de tono del canal secundario,

cuando una característica de voz de los canales primario y secundario se detecta, y cuando la coherencia de tono es menor o igual al umbral.

5 36. El sistema de codificación de sonido estéreo como se define en la reivindicación 35, que comprende medios para establecer el umbral a un valor mayor cuando el presupuesto de bits disponible es bajo con el fin de codificar la información de tono del canal secundario y/o cuando una característica de voz de los canales primario y secundario se detecta.

37. El sistema como se define en una cualquiera de las reivindicaciones 20 a 36, en el que, cuando el canal secundario se clasifica como inactivo o sin voz, el codificador del canal secundario proporciona una forma espectral del canal secundario solo para codificar el canal secundario.

10 38. El sistema como se define en una cualquiera de las reivindicaciones 20 a 36, que comprende:

un convertidor de los canales izquierdo y derecho del dominio en el tiempo a dominio de frecuencia;

en el que el mezclador de canal reductor mezcla los canales izquierdo y derecho del dominio de frecuencia para producir canales primarios y secundarios del dominio de frecuencia.

15 39. Una memoria legible por procesador que comprende instrucciones no transitorias que, cuando se ejecutan, hacen que un procesador implemente las operaciones del método como se menciona una cualquiera de las reivindicaciones 1 a 19.

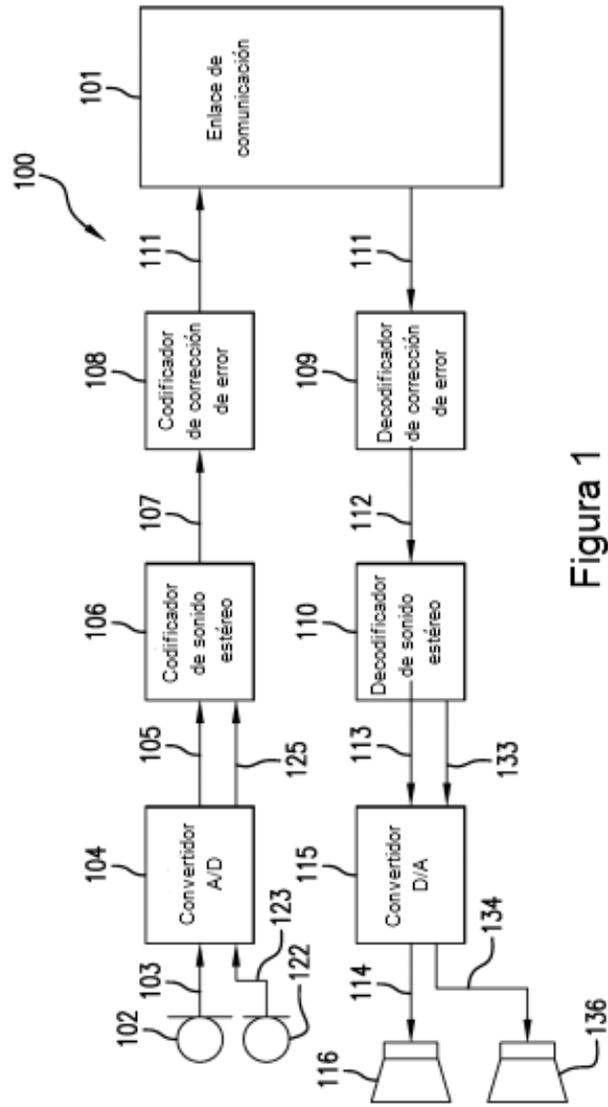


Figura 1

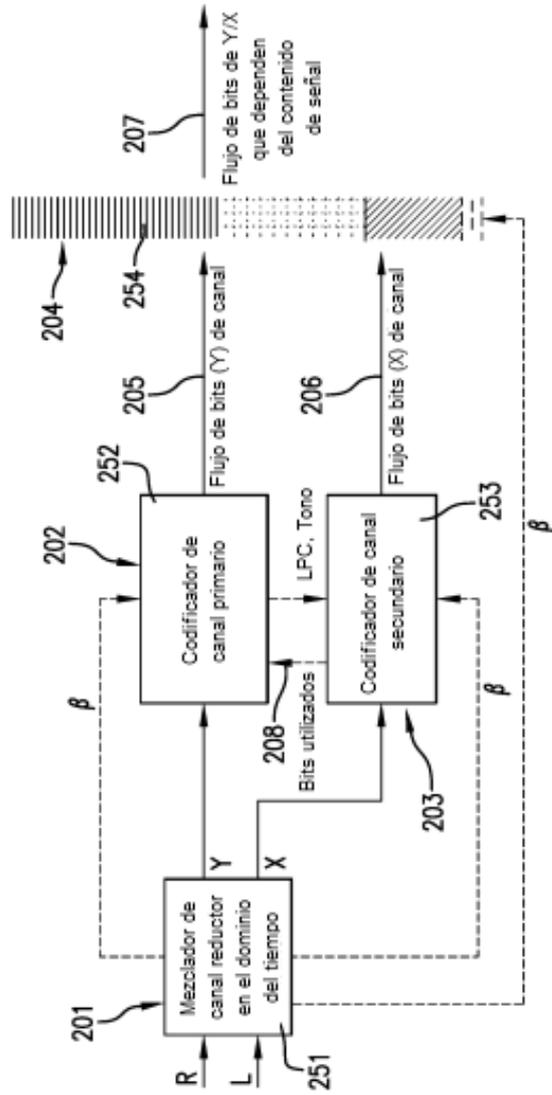


Figura 2

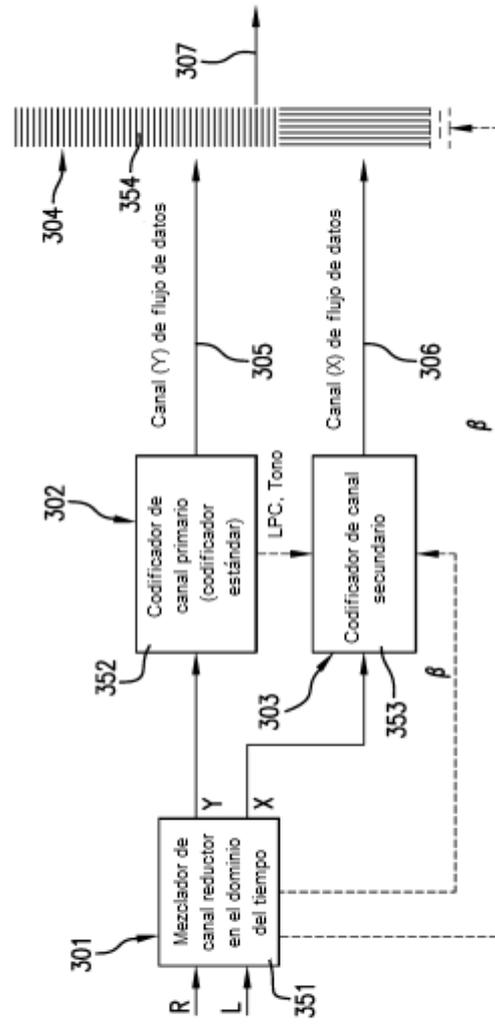


Figura 3

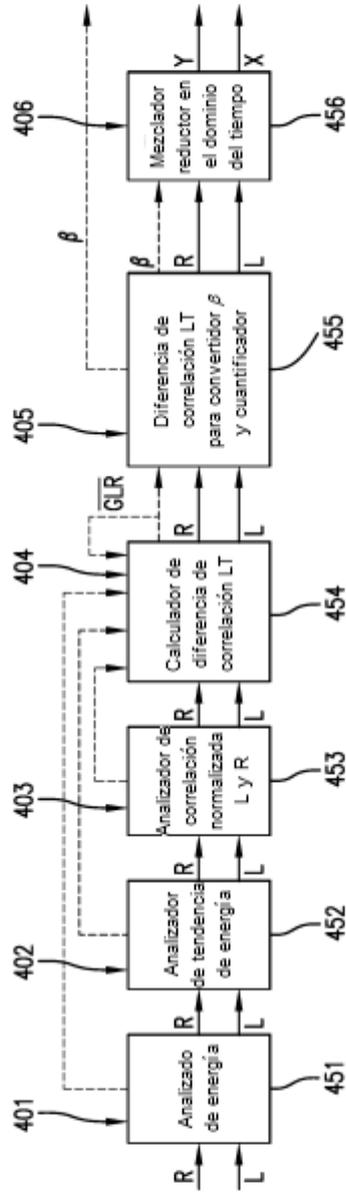


Figura 4

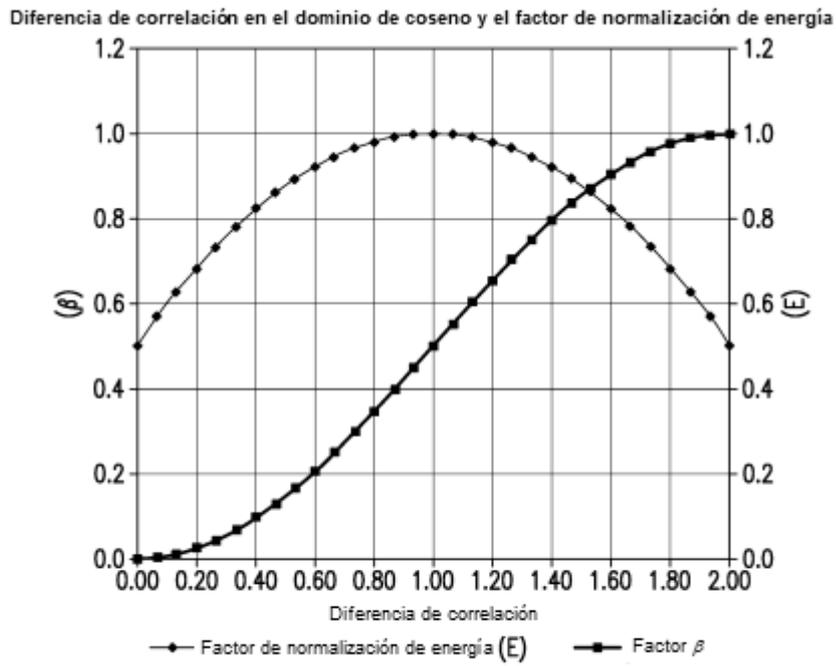


Figura 5

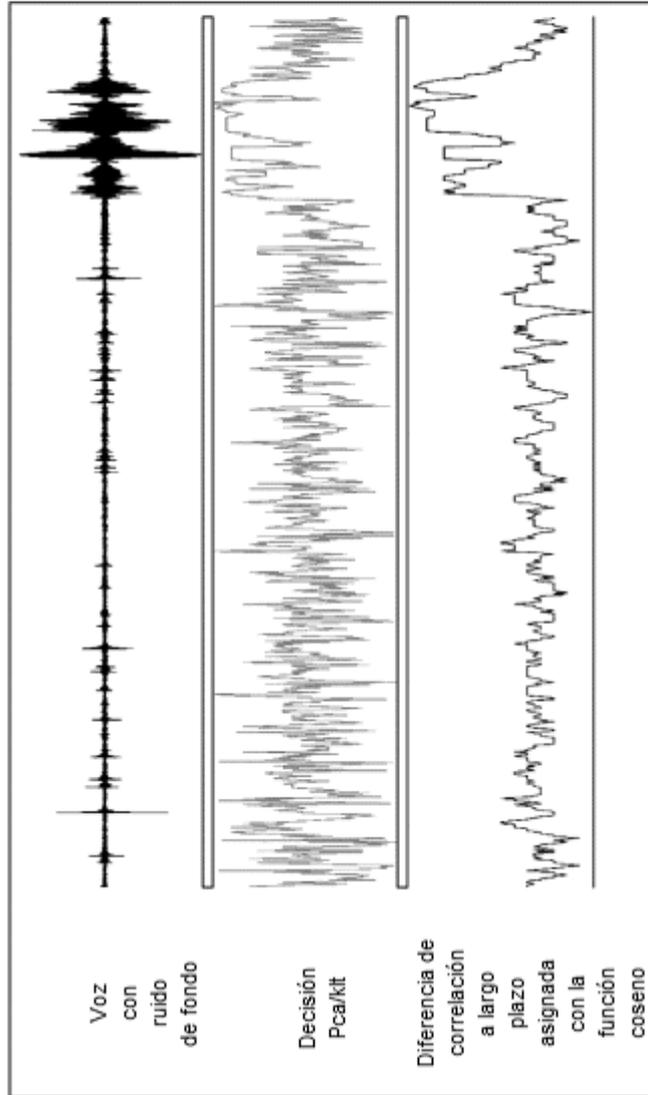


Figura 6

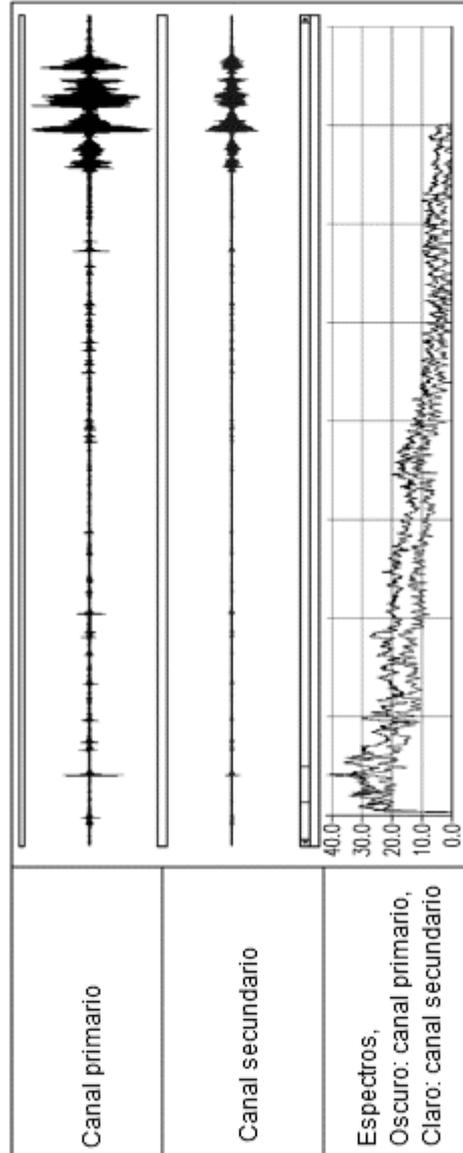


Figura 7

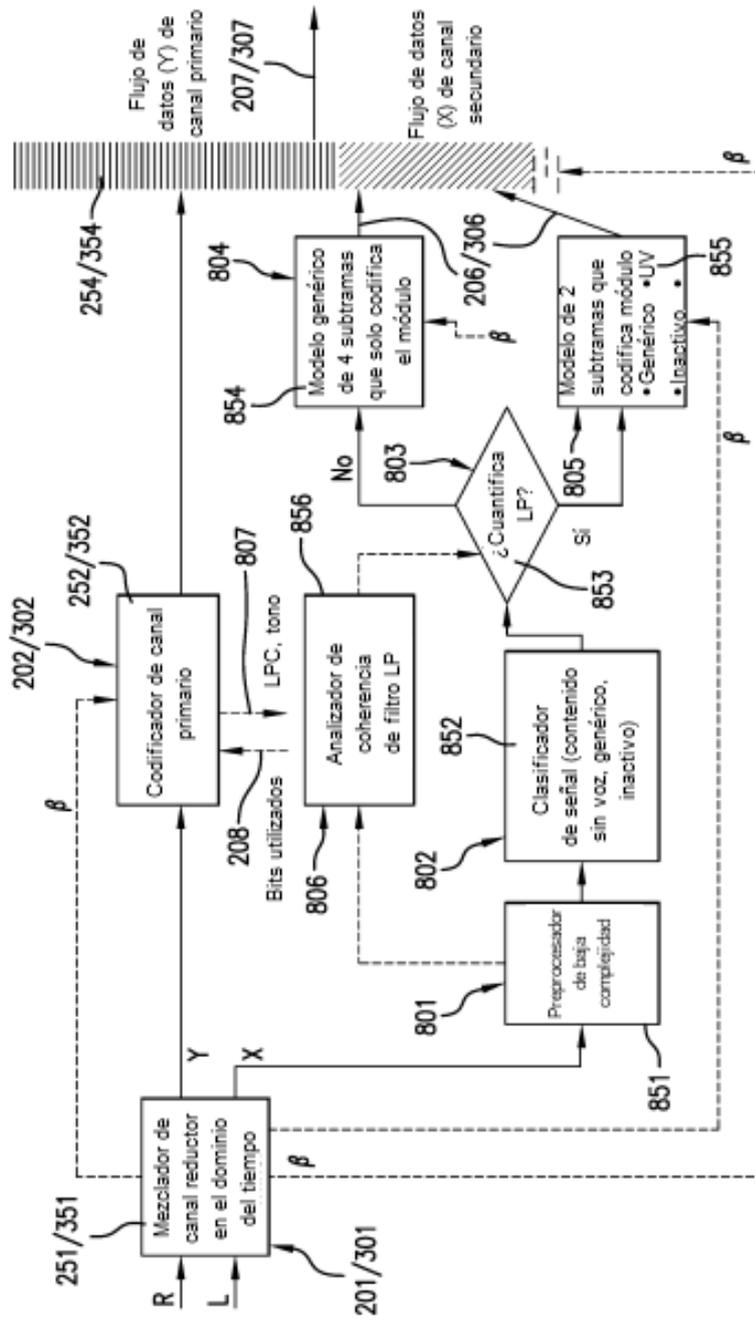
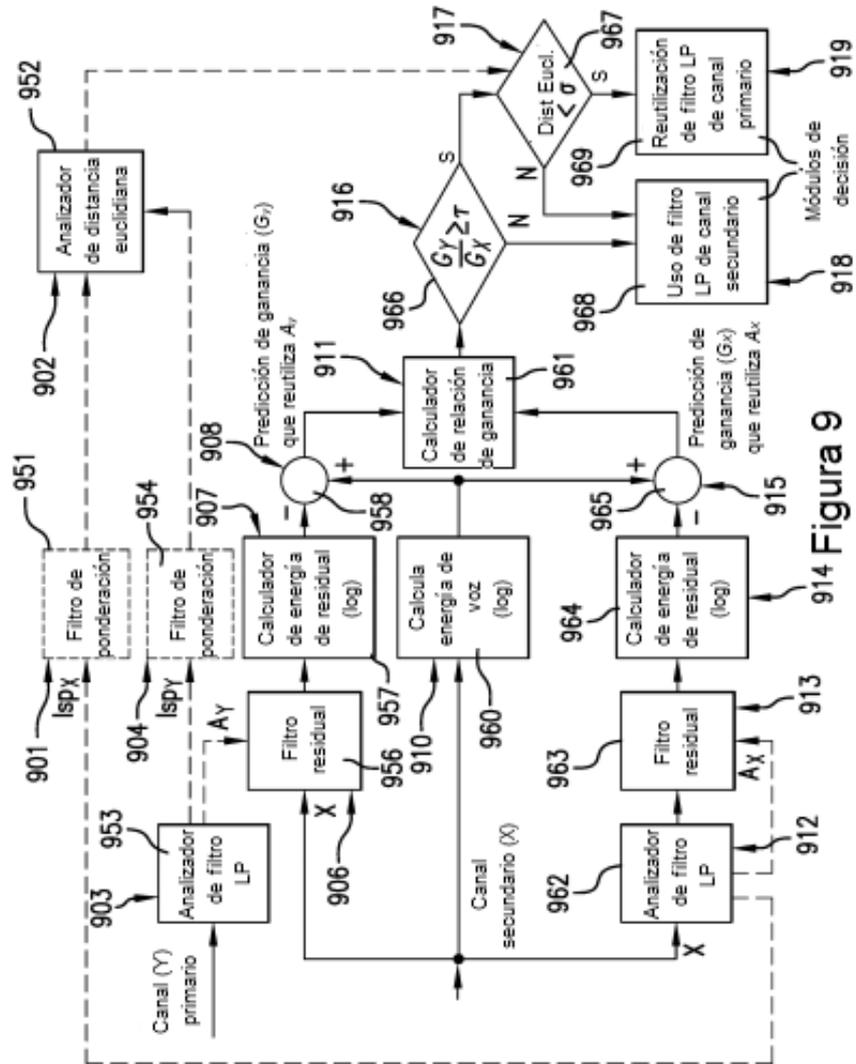


Figura 8



914 Figura 9

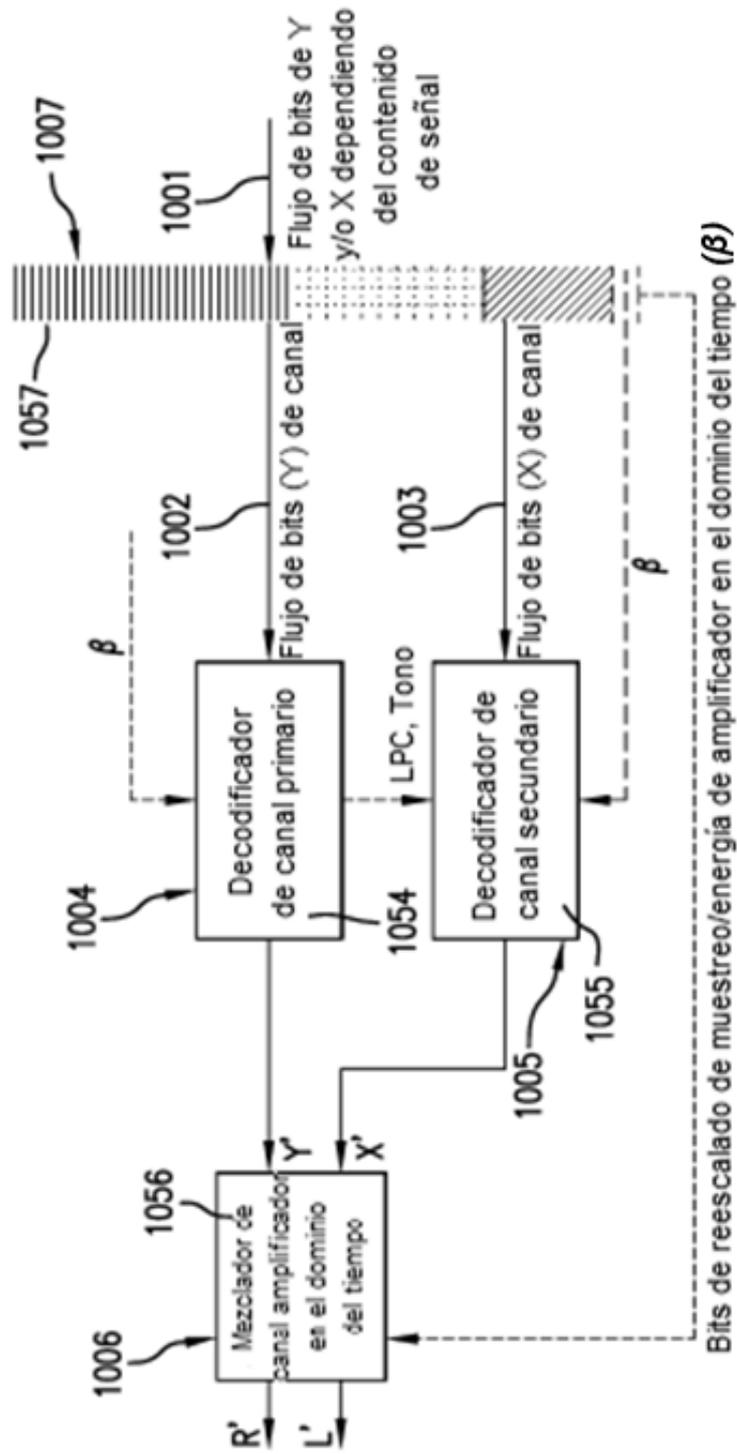


Figura 10

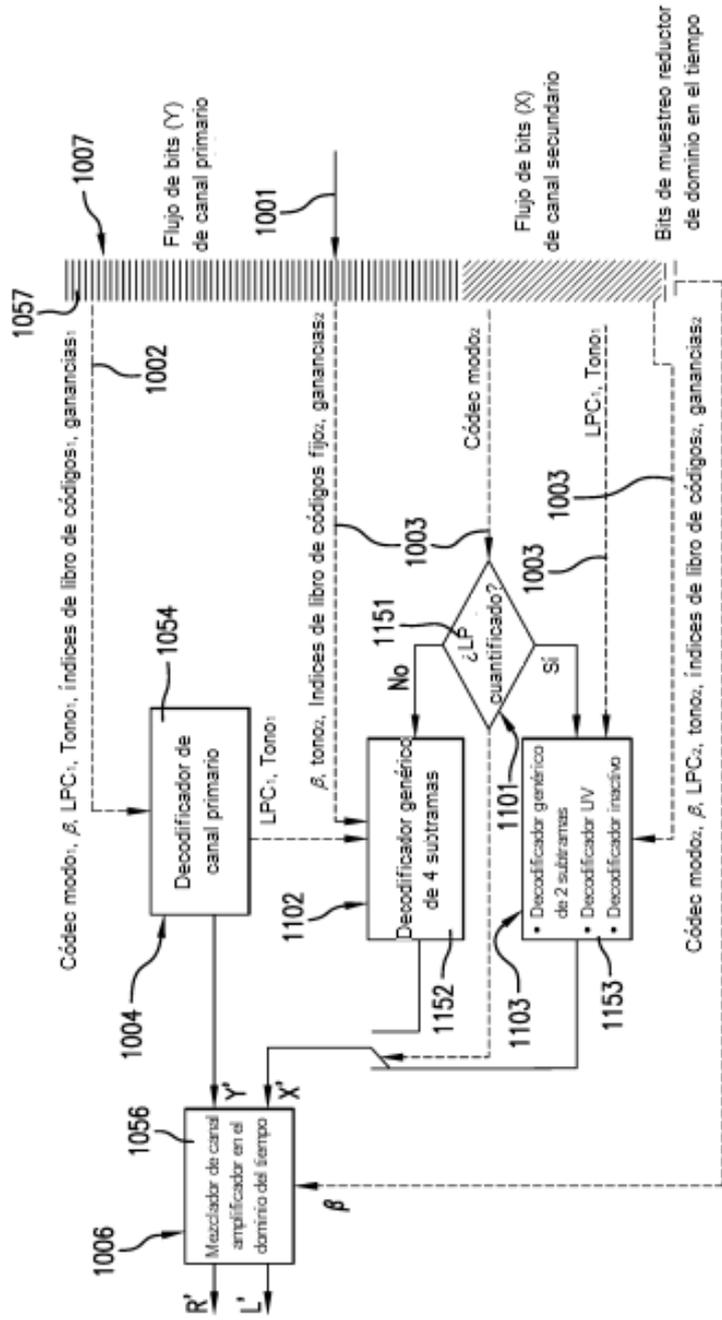


Figura 11

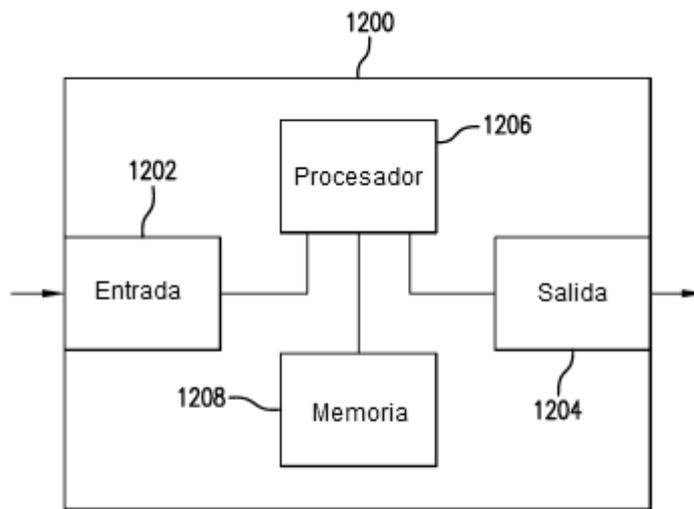


Figura 12

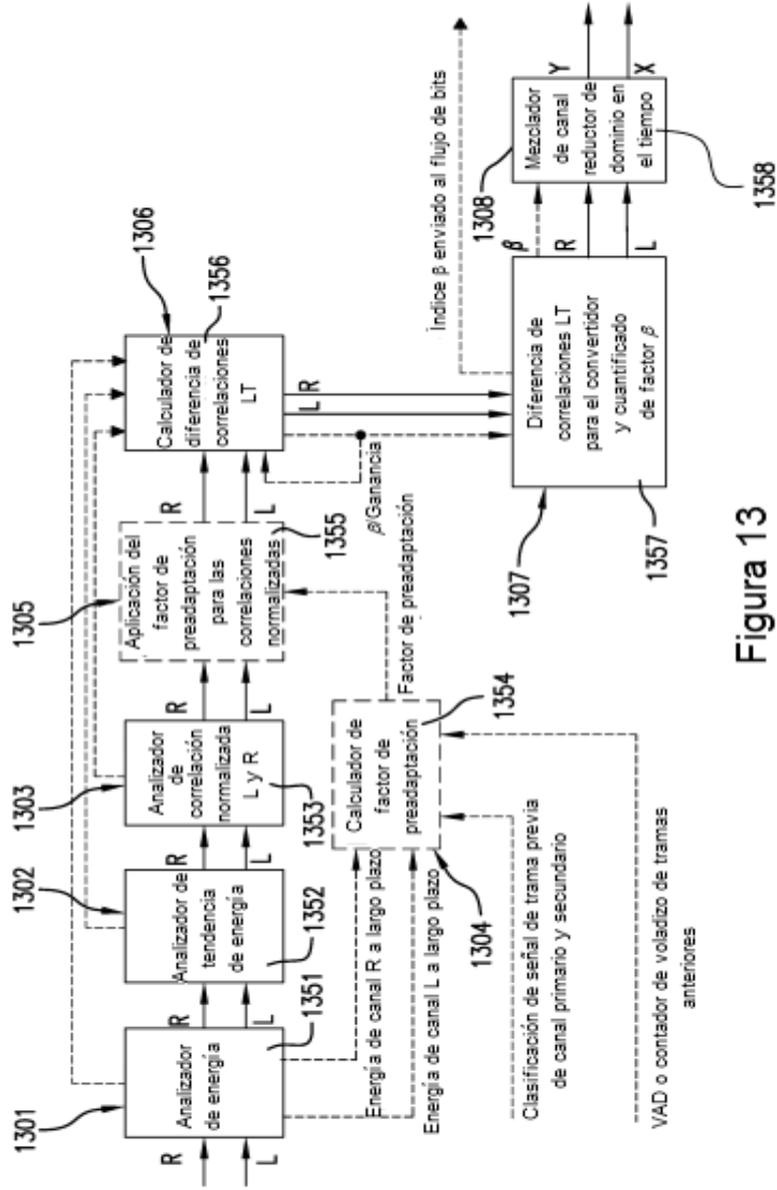


Figura 13

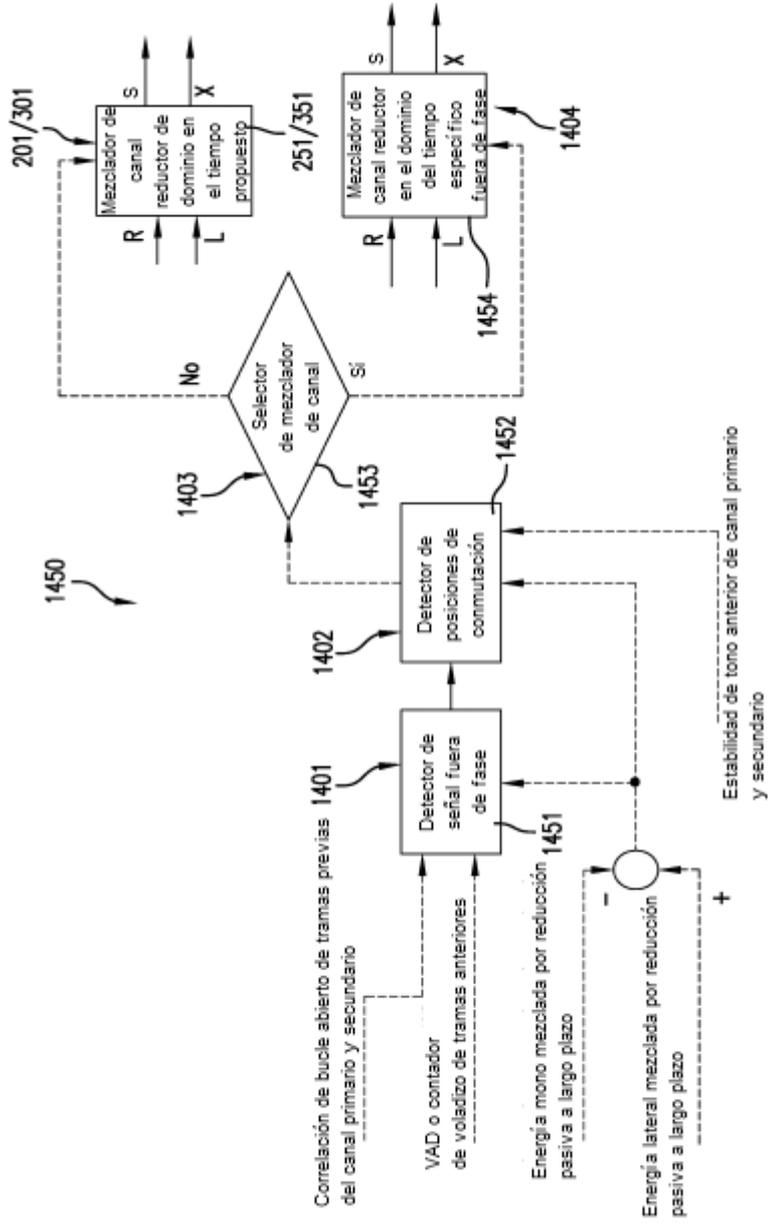


Figura 14

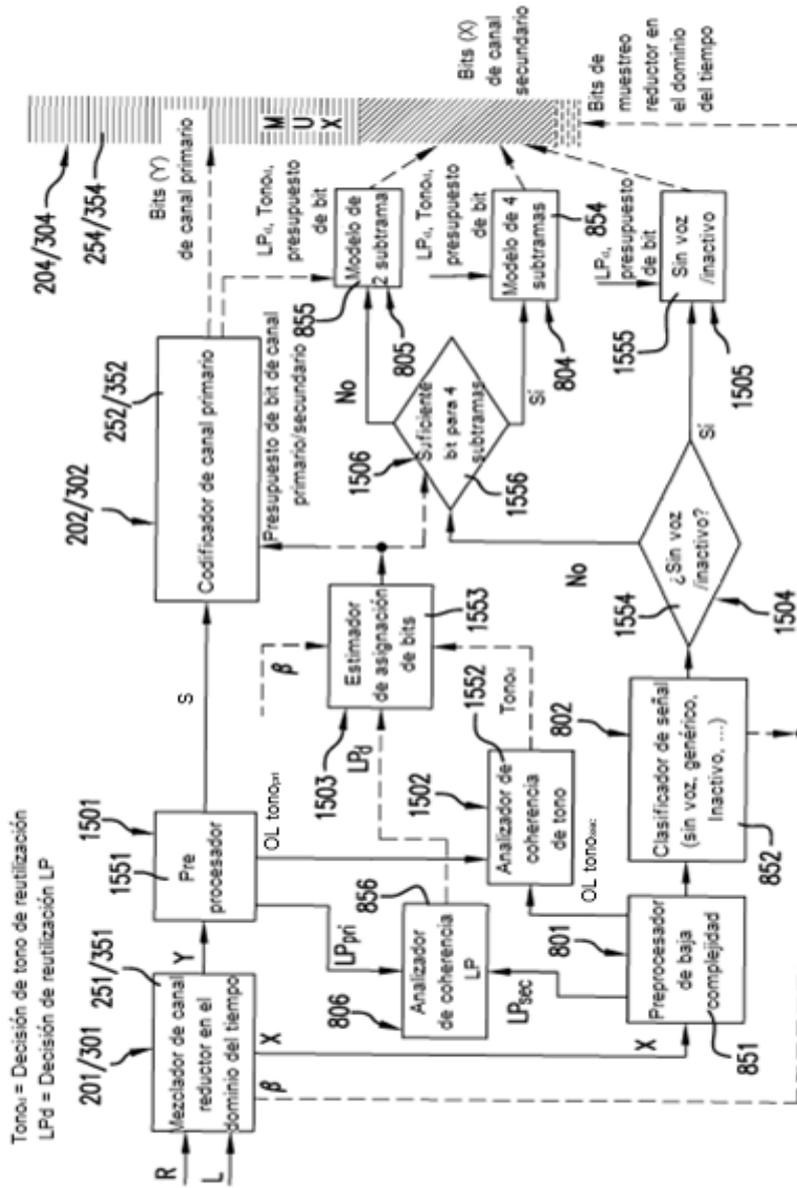


Figura 15

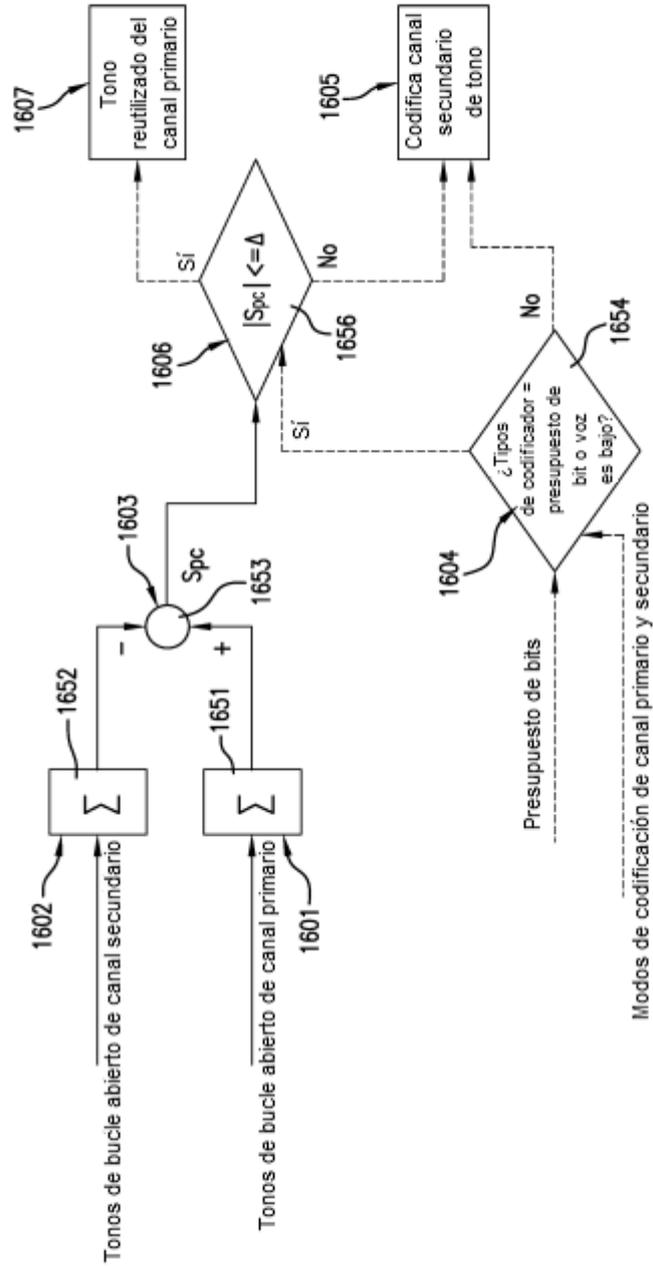


Figura 16

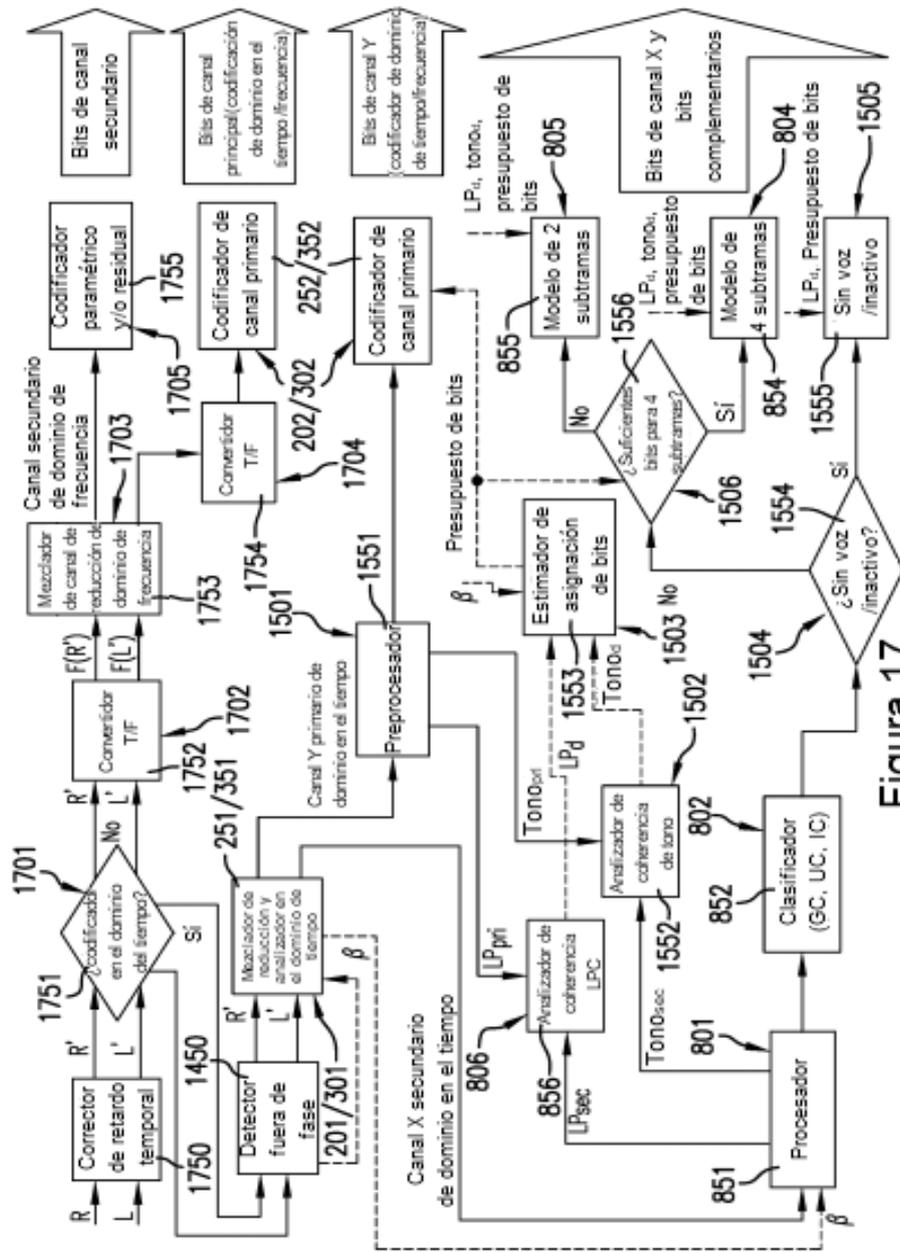


Figura 17

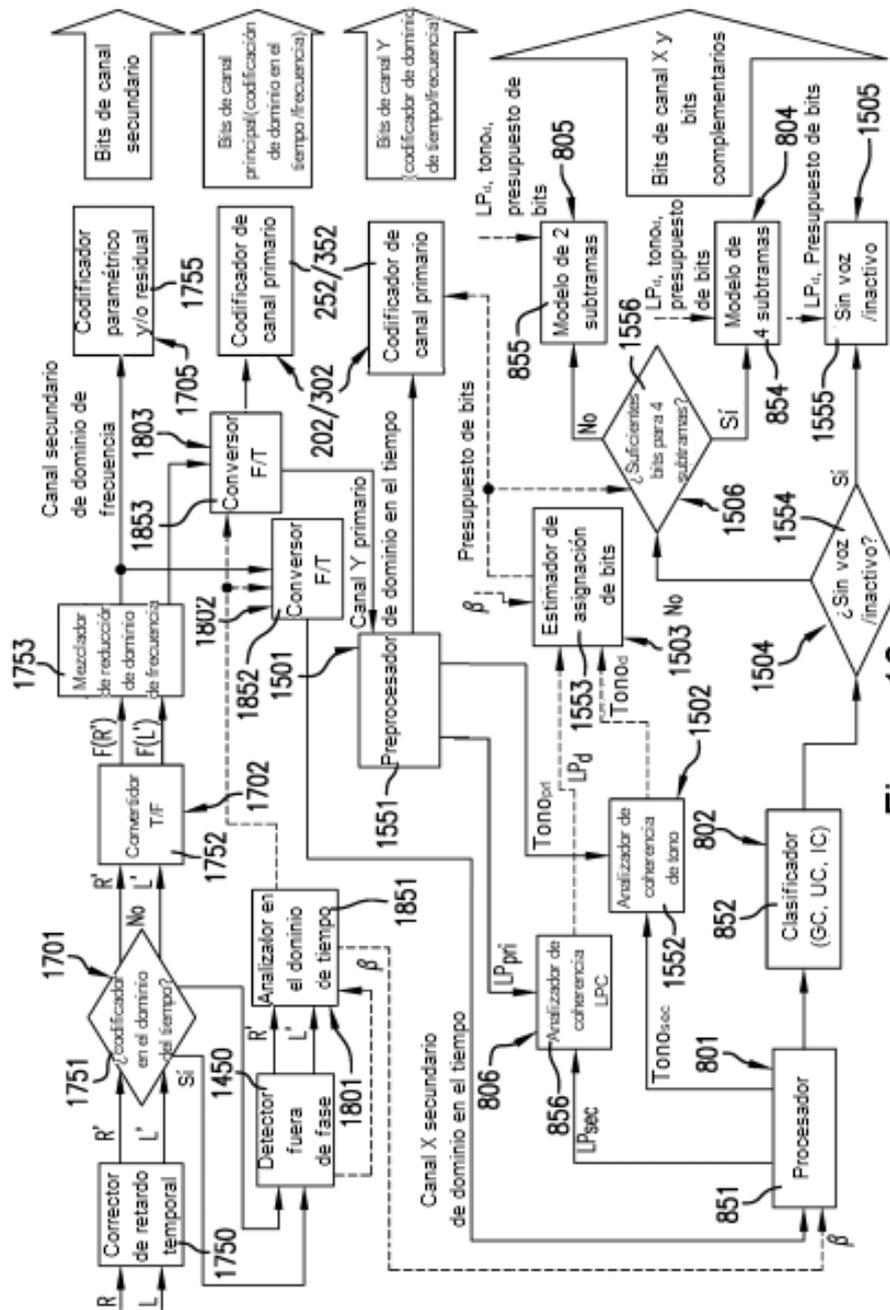


Figura 18