

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 984 875**

51 Int. Cl.:

**G10L 19/22**

(2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **13.11.2013** **E 21192621 (7)**

97 Fecha y número de publicación de la concesión europea: **31.07.2024** **EP 3933836**

54 Título: **Procedimiento y aparato para determinar un modo de codificación, procedimiento y aparato para codificar señales de audio, y procedimiento y aparato para decodificar señales de audio**

30 Prioridad:

**13.11.2012 US 201261725694 P**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**31.10.2024**

73 Titular/es:

**SAMSUNG ELECTRONICS CO., LTD. (100.0%)  
129, Samsung-ro, Yeongtong-gu  
Suwon-si, Gyeonggi-do 16677, KR**

72 Inventor/es:

**CHOO, KI-HYUN;  
POROV, ANTON VICTOROVICH;  
OSIPOV, KONSTANTIN SERGEEVICH y  
LEE, NAM-SUK**

74 Agente/Representante:

**ELZABURU, S.L.P**

ES 2 984 875 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

Procedimiento y aparato para determinar un modo de codificación, procedimiento y aparato para codificar señales de audio, y procedimiento y aparato para decodificar señales de audio

### Campo técnico

- 5 Los aparatos y procedimientos consistentes con las realizaciones ejemplares se relacionan con la codificación y decodificación de audio, y más particularmente, con un procedimiento y un aparato para determinar un modo de codificación para mejorar la calidad de una señal de audio reconstruida, determinando un modo de codificación apropiado para las características de una señal de audio y evitando la conmutación frecuente del modo de codificación, un procedimiento y un aparato para codificar una señal de audio, y un procedimiento y un aparato para decodificar una señal de audio.

### Técnica anterior

- 15 Es ampliamente conocido que es eficiente codificar una señal de música en el dominio de la frecuencia y es eficiente codificar una señal de voz en el dominio de tiempo. Por lo tanto, se han sugerido varias técnicas para determinar la clase de una señal de audio, en la cual se mezclan la señal de música y la señal de voz, y determinar un modo de codificación en correspondencia con la clase determinada.

Sin embargo, debido a la conmutación del modo de codificación de la frecuencia, no sólo se producen retrasos, sino que también se deteriora la calidad del sonido decodificado. Además, dado que no existe una técnica para corregir un modo de codificación determinado principalmente, es decir, una clase, si se produce un error durante la determinación de un modo de codificación, la calidad de una señal de audio reconstruida se deteriora.

- 20 El documento EP2144230A1 divulga un procedimiento para codificar señales de voz/música, en el que las señales de música se codifican utilizando códecs de audio MDCT de dominio de la frecuencia y las señales de voz se codifican utilizando codificación CELP/TCX. El procedimiento puede alternar el códec de voz entre CELP y TCX.

### Divulgación

#### Problema técnico

- 25 Los aspectos de una o más realizaciones ejemplares proporcionan un procedimiento para determinar un modo de codificación de acuerdo con las características de la reivindicación independiente.

#### Solución técnica

De acuerdo con un aspecto de una o más realizaciones ejemplares, existe un procedimiento para determinar un modo de codificación de acuerdo con las características de la reivindicación independiente.

- 30 Efectos ventajosos

De acuerdo con las realizaciones ejemplares, al determinar el modo de codificación final de una trama actual en base a la corrección del modo de codificación inicial y de los modos de codificación de las tramas correspondientes a una longitud de retención, puede seleccionarse un modo de codificación adaptativo a las características de una señal de audio a la vez que se evita la conmutación del modo de codificación de frecuencia entre tramas.

- 35 **Breve descripción de los dibujos**

La Figura 1 es un diagrama de bloques que ilustra una configuración de un aparato de codificación de audio de acuerdo con una realización ejemplar.

La Figura 2 es un diagrama de bloques que ilustra una configuración de un aparato de codificación de audio de acuerdo con otra realización ejemplar.

- 40 La Figura 3 es un diagrama de bloques que ilustra una configuración de una unidad de determinación del modo de codificación de acuerdo con una realización ejemplar;

La Figura 4 es un diagrama de bloques que ilustra una configuración de una unidad de determinación del modo de codificación inicial de acuerdo con una realización ejemplar;

- 45 La Figura 5 es un diagrama de bloques que ilustra una configuración de una unidad de extracción de parámetros de características de acuerdo con una realización ejemplar;

La Figura 6 es un diagrama que ilustra un procedimiento de conmutación adaptativo entre una codificación de dominio de predicción lineal y un dominio de espectro de acuerdo con una realización ejemplar;

La Figura 7 es un diagrama que ilustra el funcionamiento de una unidad de corrección del modo de codificación de acuerdo con una realización ejemplar;

La Figura 8 es un diagrama de bloques que ilustra una configuración de un aparato de decodificación de audio de acuerdo con una realización ejemplar.

- 5 La Figura 9 es un diagrama de bloques que ilustra una configuración de un aparato de decodificación de audio de acuerdo con otra realización ejemplar.

### Modo de la invención

A continuación, se hará referencia en detalle a las realizaciones, cuyos ejemplos se ilustran en los dibujos adjuntos, en los que los números de referencia similares se refieren a elementos similares en todo el conjunto. En este sentido, las presentes realizaciones pueden tener diferentes formas y no deben interpretarse como limitadas a las descripciones expuestas en la presente memoria. En consecuencia, las realizaciones se describen simplemente a continuación, haciendo referencia a las figuras, para explicar aspectos de la presente descripción.

Se pueden utilizar términos tales como “conectado” y “vinculado” para indicar un estado directamente conectado o vinculado, pero se entenderá que otro componente puede interponerse entre ellos.

- 15 Se pueden utilizar términos tales como “primero” y “segundo” para describir varios componentes, pero los componentes no se limitarán a los términos. Los términos pueden utilizarse únicamente para distinguir un componente de otro.

Las unidades descritas en las realizaciones ejemplares se ilustran de manera independiente para indicar diferentes funciones características, y no significa que cada unidad esté formada por un componente de hardware o software independiente. Cada unidad se ilustra para facilitar la explicación, y una pluralidad de unidades puede formar una unidad, y una unidad puede dividirse en una pluralidad de unidades.

La Figura 1 es un diagrama de bloques que ilustra una configuración de un aparato 100 de codificación de audio de acuerdo con una realización ejemplar.

El aparato 100 de codificación de audio que se muestra en la Figura 1 puede incluir una unidad 110 de determinación del modo de codificación, una unidad 120 de conmutación, una unidad 130 de codificación de dominio de espectro, una unidad 140 de codificación de dominio de predicción lineal, y una unidad 150 de generación de flujo de bits. La unidad 140 de codificación de dominio de predicción lineal puede incluir una unidad 141 de codificación de excitación de dominio de tiempo y una unidad 143 de codificación de excitación de dominio de frecuencia, donde la unidad 140 de codificación de dominio de predicción lineal puede estar incorporada como al menos una de las dos unidades 141 y 143 de codificación de excitación. A menos que sea necesario que se incorporen como un hardware separado, los componentes mencionados anteriormente pueden integrarse en al menos un módulo y pueden implementarse como al menos un procesador (no se muestra). En este caso, el término de una señal de audio puede referirse a una señal de música, una señal de voz, o una señal mixta de las mismas.

Con referencia a la Figura 1, la unidad 110 de determinación del modo de codificación puede analizar las características de una señal de audio para determinar la clase de la señal de audio, y determinar un modo de codificación en correspondencia con un resultado de la clasificación. La determinación del modo de codificación puede realizarse en unidades de supertramas, tramas, o bandas. De manera alternativa, la determinación del modo de codificación puede realizarse en unidades de una pluralidad de grupos de supertramas, una pluralidad de grupos de tramas, o una pluralidad de grupos de bandas. En este caso, los ejemplos de los modos de codificación pueden incluir un dominio de espectro y un dominio de tiempo o un dominio de predicción lineal, pero no están limitados a los mismos. Si el rendimiento y la velocidad de procesamiento de un procesador son suficientes y se pueden resolver los retrasos debidos a la conmutación del modo de codificación, los modos de codificación pueden subdividirse, y los esquemas de codificación también pueden subdividirse en correspondencia con el modo de codificación. De acuerdo con una realización ejemplar, la unidad 110 de determinación del modo de codificación puede determinar un modo de codificación inicial de una señal de audio como uno de un modo de codificación de dominio de espectro y un modo de codificación de dominio de tiempo. De acuerdo con otra realización ejemplar, la unidad 110 de determinación del modo de codificación puede determinar un modo de codificación inicial de una señal de audio como uno de un modo de codificación de dominio de espectro, un modo de codificación de excitación de dominio de tiempo y un modo de codificación de excitación de dominio de frecuencia. Si el modo de codificación de dominio de espectro se determina como el modo de codificación inicial, la unidad 110 de determinación del modo de codificación puede corregir el modo de codificación inicial a uno del modo de codificación de dominio de espectro y el modo de codificación de excitación de dominio de frecuencia. Si el modo de codificación de dominio de tiempo, es decir, el modo de codificación de excitación de dominio de tiempo se determina como el modo de codificación inicial, la unidad 110 de determinación del modo de codificación puede corregir el modo de codificación inicial a uno del modo de codificación de excitación de dominio de tiempo y el modo de codificación de excitación de dominio de frecuencia. Si el modo de codificación de excitación de dominio de tiempo se determina como el modo de codificación inicial, la determinación del modo de codificación final puede realizarse de manera selectiva. En otras palabras, se puede mantener el modo de codificación inicial, es decir, el modo de codificación de excitación de dominio de tiempo. La unidad 110 de determinación del modo

de codificación puede determinar los modos de codificación de una pluralidad de tramas correspondientes a una longitud de retención, y puede determinar el modo de codificación final para una trama actual. De acuerdo con una realización ejemplar, si el modo de codificación inicial o un modo de codificación corregido de una trama actual es idéntico a los modos de codificación de una pluralidad de tramas anteriores, por ejemplo, 7 tramas anteriores, el correspondiente modo de codificación inicial o modo de codificación corregido puede determinarse como el modo de codificación final de la trama actual. Mientras tanto, si el modo de codificación inicial o un modo de codificación corregido de una trama actual no es idéntico a los modos de codificación de una pluralidad de tramas anteriores, por ejemplo, 7 tramas anteriores, la unidad 110 de determinación del modo de codificación puede determinar el modo de codificación de la trama justo antes de la trama actual como el modo de codificación final de la trama actual.

Como se ha descrito anteriormente, al determinar el modo de codificación final de una trama actual en base a la corrección del modo de codificación inicial y de los modos de codificación de las tramas correspondientes a una longitud de retención, puede seleccionarse un modo de codificación adaptativo a las características de una señal de audio, a la vez que se evita la conmutación del modo de codificación de frecuencia entre tramas.

En general, la codificación de dominio de tiempo, es decir, la codificación de excitación de dominio de tiempo puede ser eficiente para una señal de voz, la codificación de dominio de espectro puede ser eficiente para una señal de música, y la codificación de excitación de dominio de frecuencia puede ser eficiente para una señal vocal y/o armónica.

En correspondencia con un modo de codificación determinado por la unidad 110 de determinación del modo de codificación, la unidad 120 de conmutación puede proporcionar una señal de audio a la unidad 130 de codificación de dominio de espectro o a la unidad 140 de codificación de dominio de predicción lineal. Si la unidad 140 de codificación de dominio de predicción lineal se incorpora como la unidad 141 de codificación de excitación de dominio de tiempo, la unidad 120 de conmutación puede incluir un total de dos ramas. Si la unidad 140 de codificación de dominio de predicción lineal se incorpora como la unidad 141 de codificación de excitación de dominio de tiempo y la unidad 143 de codificación de excitación de dominio de frecuencia, la unidad 120 de conmutación puede tener un total de 3 ramas.

La unidad 130 de codificación de dominio de espectro puede codificar una señal de audio en el dominio de espectro. El dominio de espectro puede referirse al dominio de frecuencia o a un dominio de transformación. Los ejemplos de procedimientos de codificación aplicables a la unidad 130 de codificación de dominio de espectro pueden incluir una codificación de audio avanzada (AAC), o una combinación de una transformada discreta de coseno modificada (MDCT) y una codificación de pulso factorial (FPC), pero no se limitan a ello. En detalle, se pueden utilizar otras técnicas de cuantificación y de codificación de entropía en lugar del FPC. Puede ser eficiente codificar una señal de música en la unidad 130 de codificación de dominio de espectro.

La unidad 140 de codificación de dominio de predicción lineal puede codificar una señal de audio en un dominio de predicción lineal. El dominio de predicción lineal puede referirse a un dominio de excitación o a un dominio de tiempo. La unidad 140 de codificación de dominio de predicción lineal puede estar incorporada como la unidad 141 de codificación de excitación de dominio de tiempo o puede estar incorporada para incluir la unidad 141 de codificación de excitación de dominio de tiempo y la unidad 143 de codificación de excitación de dominio de frecuencia. Los ejemplos de procedimientos de codificación aplicables a la unidad 141 de codificación de excitación de dominio de tiempo pueden incluir la predicción lineal excitada por código (CELP) o una CELP algebraica (ACELP), pero no se limitan a ello. Los ejemplos de procedimientos de codificación aplicables a la unidad 143 de codificación de excitación de dominio de frecuencia pueden incluir la codificación genérica de la señal (GSC) o la excitación codificada por transformación (TCX), sin limitarse a ello. Puede ser eficiente codificar una señal de voz en la unidad 141 de codificación de excitación de dominio de tiempo, mientras que puede ser eficiente codificar una señal vocal y/o armónica en la unidad 143 de codificación de excitación de dominio de frecuencia.

La unidad 150 de generación de flujo de bits puede generar un flujo de bits que incluya el modo de codificación proporcionado por la unidad 110 de determinación del modo de codificación, un resultado de codificación proporcionado por la unidad 130 de codificación de dominio de espectro, y un resultado de codificación proporcionado por la unidad 140 de codificación de dominio de predicción lineal.

La Figura 2 es un diagrama de bloques que ilustra una configuración de un aparato 200 de codificación de audio de acuerdo con otra realización ejemplar.

El aparato 200 de codificación de audio que se muestra en la Figura 2 puede incluir un módulo 205 de preprocesamiento común, una unidad 210 de determinación del modo de codificación, una unidad 220 de conmutación, una unidad 230 de codificación de dominio de espectro, una unidad 240 de codificación de dominio de la predicción lineal, y una unidad 250 de generación de flujo de bits. En este caso, la unidad 240 de codificación de dominio de predicción lineal puede incluir una unidad 241 de codificación de excitación de dominio de tiempo y una unidad 243 de codificación de excitación de dominio de frecuencia, y la unidad 240 de codificación de dominio de predicción lineal puede ser incorporada como la unidad 241 de codificación de excitación de dominio de tiempo o la unidad 243 de codificación de excitación de dominio de frecuencia. En comparación con el aparato 100 de codificación de audio que se muestra en la Figura 1, el aparato 200 de codificación de audio puede incluir además el módulo 205 de preprocesamiento común, y por lo tanto se omitirán las descripciones de los componentes idénticos a los del aparato 100 de codificación de audio.

Con referencia a la Figura 2, el módulo 205 de preprocesamiento común puede realizar el procesamiento estéreo conjunto, el procesamiento de sonido envolvente, y/o el procesamiento de extensión del ancho de banda. El procesamiento estéreo conjunto, el procesamiento de sonido envolvente, y el procesamiento de extensión del ancho de banda pueden ser idénticos a los empleados por un estándar específico, por ejemplo, el estándar MPEG, pero no están limitados a ello. La salida del módulo 205 de preprocesamiento común puede ser en un canal mono, en un canal estéreo, o en múltiples canales. De acuerdo con el número de canales de una señal emitida por el módulo 205 de preprocesamiento común, la unidad 220 de conmutación puede incluir al menos un interruptor. Por ejemplo, si el módulo 205 de preprocesamiento común emite una señal de dos o más canales, es decir, un canal estéreo o un multicanal, pueden disponerse interruptores correspondientes a los respectivos canales. Por ejemplo, el primer canal de una señal estéreo puede ser un canal de voz, y el segundo canal de la señal estéreo puede ser un canal de música. En este caso, se puede proporcionar de manera simultánea una señal de audio a los dos interruptores. La información adicional generada por el módulo 205 de preprocesamiento común puede proporcionarse a la unidad 250 de generación de flujo de bits e incluirse en un flujo de bits. La información adicional puede ser necesaria para realizar el procesamiento estéreo conjunto, el procesamiento de sonido envolvente, y/o el procesamiento de extensión del ancho de banda en un extremo de decodificación y puede incluir parámetros espaciales, información de envolvente, información de energía, etc. Sin embargo, puede haber varias informaciones adicionales en base a las técnicas de procesamiento aplicadas.

De acuerdo con una realización ejemplar, en el módulo 205 de preprocesamiento común, el procesamiento de extensión del ancho de banda puede realizarse de manera diferente en base a los dominios de codificación. La señal de audio en una banda central puede ser procesada utilizando el modo de codificación de excitación de dominio de tiempo o el modo de codificación de excitación de dominio de frecuencia, mientras que una señal de audio en una banda extendida de ancho de banda puede ser procesada en el dominio de tiempo. El procesamiento de extensión del ancho de banda en el dominio de tiempo puede incluir una pluralidad de modos que incluyen un modo con voz o un modo sin voz. De manera alternativa, una señal de audio en la banda central puede ser procesada utilizando el modo de codificación de dominio de espectro, mientras que una señal de audio en la banda extendida de ancho de banda puede ser procesada en el dominio de la frecuencia. El procesamiento de extensión del ancho de banda en el dominio de la frecuencia puede incluir una pluralidad de modos que incluyen un modo transitorio, un modo normal, o un modo armónico. Para realizar el procesamiento de extensión del ancho de banda en diferentes dominios, un modo de codificación determinado por la unidad 110 de determinación del modo de codificación puede proporcionarse al módulo 205 de preprocesamiento común como información de señalización. De acuerdo con una realización ejemplar, la última porción de la banda central y la porción inicial de la banda extendida de ancho de banda pueden superponerse entre sí hasta cierta medida. La ubicación y el tamaño de las porciones superpuestas pueden establecerse con anterioridad.

La Figura 3 es un diagrama de bloques que ilustra una configuración de una unidad 300 de determinación del modo de codificación de acuerdo con una realización ejemplar.

La unidad 300 de determinación del modo de codificación que se muestra en la Figura 3 puede incluir una unidad 310 de determinación del modo de codificación inicial y una unidad 330 de corrección del modo de codificación.

Con referencia a la Figura 3, la unidad 310 de determinación del modo de codificación inicial puede determinar si una señal de audio es una señal de música o una señal de voz utilizando parámetros de características extraídos a partir de la señal de audio. Si la señal de audio se determina como una señal de voz, la codificación de dominio de predicción lineal puede ser adecuada. Mientras tanto, si la señal de audio se determina como una señal de música, la codificación de dominio de espectro puede ser adecuada. La unidad 310 de determinación del modo de codificación inicial puede determinar la clase de la señal de audio indicando si la codificación de dominio de espectro, la codificación de excitación de dominio de tiempo, o la codificación de excitación de dominio de frecuencia es adecuada para la señal de audio utilizando parámetros de características extraídos a partir de la señal de audio. Se puede determinar un modo de codificación correspondiente en base a la clase de la señal de audio. Si una unidad de conmutación (120 de la Figura 1) tiene dos ramas, un modo de codificación puede expresarse en 1 bit. Si la unidad de conmutación (120 de la Figura 1) tiene tres ramas, un modo de codificación puede expresarse en 2 bits. La unidad 310 de determinación del modo de codificación inicial puede determinar si una señal de audio es una señal de música o una señal de voz utilizando cualquiera de las diversas técnicas conocidas en la técnica. Los ejemplos de estos pueden incluir la clasificación FD/LPD o la clasificación ACELP/TCX divulgada en una parte del codificador de la norma USAC y la clasificación ACELP/TCX utilizada en las normas AMR, pero no se limitan a ellas. En otras palabras, el modo de codificación inicial puede determinarse utilizando cualquiera de los diversos procedimientos distintos del procedimiento de acuerdo con las realizaciones descritas en la presente memoria.

La unidad 330 de corrección del modo de codificación puede determinar un modo de codificación corregido corrigiendo el modo de codificación inicial determinado por la unidad 310 de determinación del modo de codificación inicial utilizando parámetros de corrección. De acuerdo con una realización ejemplar, si el modo de codificación de dominio de espectro se determina como el modo de codificación inicial, el modo de codificación inicial puede corregirse al modo de codificación de excitación de dominio de frecuencia en base a los parámetros de corrección. Si el modo de codificación de dominio de tiempo se determina como el modo de codificación inicial, el modo de codificación inicial puede corregirse al modo de codificación de excitación de dominio de frecuencia en base a los parámetros de corrección. En otras palabras, se determina si hay un error en la determinación del modo de codificación inicial

utilizando parámetros de corrección. Si se determina que no existe error en la determinación del modo de codificación inicial, se puede mantener el modo de codificación inicial. Por el contrario, si se determina que existe un error en la determinación del modo de codificación inicial, éste puede corregirse. La corrección del modo de codificación inicial puede obtenerse a partir del modo de codificación de dominio de espectro al modo de codificación de excitación de dominio de frecuencia y a partir del modo de codificación de excitación de dominio de tiempo al modo de codificación de excitación de dominio de frecuencia.

Mientras tanto, el modo de codificación inicial o el modo de codificación corregido puede ser un modo de codificación temporal para una trama actual, donde el modo de codificación temporal para la trama actual puede ser comparado con los modos de codificación para las tramas anteriores dentro de una longitud de retención preestablecida y se puede determinar el modo de codificación final para la trama actual.

La Figura 4 es un diagrama de bloques que ilustra una configuración de una unidad 400 de determinación del modo de codificación inicial de acuerdo con una realización ejemplar.

La unidad 400 de determinación del modo de codificación inicial que se muestra en la Figura 4 puede incluir una unidad 410 de extracción de parámetros de características y una unidad 430 de determinación.

Con referencia a la Figura 4, la unidad 410 de extracción de parámetros de características puede extraer los parámetros de características necesarios para determinar un modo de codificación a partir de una señal de audio. Los ejemplos de los parámetros de características extraídos incluyen al menos uno o dos de entre un parámetro de tono, un parámetro de sonorización, un parámetro de correlación, y un error de predicción lineal, pero no se limitan a ellos. A continuación, se describen detalladamente los distintos parámetros.

En primer lugar, un primer parámetro  $F_1$  de característica se refiere a un parámetro de tono, donde un comportamiento de tono puede determinarse utilizando N valores de tono detectados en una trama actual y en al menos una trama anterior. Para evitar un efecto de una desviación aleatoria o de un valor de tono incorrecto, se pueden eliminar los M valores de tono significativamente diferentes del promedio de los N valores de tono. En este caso, N y M pueden ser valores obtenidos con anterioridad a través de experimentos o simulaciones. Además, N puede establecerse con anterioridad, y una diferencia entre un valor de tono por eliminar y el promedio de los N valores de tono puede determinarse con anterioridad a través de experimentos o simulaciones. El primer parámetro  $F_1$  de característica puede expresarse como se muestra a continuación en la Ecuación 1, utilizando el promedio  $m_p$  y la varianza  $\sigma_p$  con respecto a los valores de tono (N-M)

[Ecuación 1]

$$F_1 = \frac{\sigma_p}{m_p}$$

Un segundo parámetro  $F_2$  de característica también se refiere a un parámetro de tono y puede indicar la fiabilidad de un valor de tono detectado en una trama actual. El segundo parámetro  $F_2$  de característica puede expresarse como se muestra a continuación en la Ecuación 2, utilizando las varianzas  $\sigma_{SF1}$  y  $\sigma_{SF2}$  de los valores de tono detectados respectivamente en dos subtramas  $SF_1$  y  $SF_2$  de una trama actual

[Ecuación 2]

$$F_2 = \frac{cov(SF_1, SF_2)}{\sigma_{SF1} \sigma_{SF2}}$$

En este caso,  $cov(SF_1, SF_2)$  denota la covarianza entre las subtramas  $SF_1$  y  $SF_2$ . En otras palabras, el segundo parámetro  $F_2$  de característica indica la correlación entre dos subtramas como una distancia de tono. De acuerdo con una realización ejemplar, una trama actual puede incluir dos o más subtramas, y la Ecuación 2 puede modificarse en base al número de subtramas.

Un tercer parámetro  $F_3$  de característica puede expresarse como se muestra a continuación en la Ecuación 3, en base a un parámetro de sonorización y un parámetro Corr de correlación

[Ecuación 3]

$$F_3 = \sqrt{Q \frac{cSonorización - CorrC^2}{N}}$$

En este caso, el parámetro de Sonorización se refiere a las características vocales del sonido y puede obtenerse por cualquiera de los diversos procedimientos conocidos en la técnica, mientras que el parámetro Corr de correlación puede obtenerse sumando las correlaciones entre tramas para cada banda.

Un cuarto parámetro  $F_4$  de característica se relaciona con un error  $E_{LPC}$  de predicción lineal y puede expresarse como se muestra a continuación en la Ecuación 4.

[Ecuación 4]

$$F_4 = \sqrt{\frac{E_{LPCi} - M(E_{LPC})}{N}}$$

5 En este caso,  $M(E_{LPC})$  denota el promedio de  $N$  errores de predicción lineal.

La unidad 430 de determinación puede determinar la clase de una señal de audio utilizando al menos un parámetro de característica proporcionado por la unidad 410 de extracción de parámetros de características y puede determinar el modo de codificación inicial en base a la clase determinada. La unidad 430 de determinación puede emplear un mecanismo de decisión suave, en el que se puede formar al menos una mezcla por parámetro de característica. De acuerdo con una realización ejemplar, la clase de una señal de audio puede determinarse utilizando el modelo de mezcla de Gaussiano (GMM) en base a las probabilidades de mezcla. La probabilidad  $f(x)$  relativa a una mezcla puede calcularse de acuerdo con la Ecuación 5 siguiente.

[Ecuación 5]

$$f(x) = \frac{1}{\sqrt{(2\pi)^N \det(C^{-1})}} e^{-0.5(x-m)^T C^{-1}(x-m)}$$

15  $x = (x_1, \dots, x_N)$

$$m = (Cx_1, \dots, Cx_N C)$$

En este caso,  $x$  denota un vector de entrada de un parámetro de característica,  $m$  denota una mezcla, y  $c$  denota una matriz de covarianza.

La unidad 430 de determinación puede calcular una probabilidad de música  $P_m$  y una probabilidad de voz  $P_s$  utilizando la Ecuación 6 siguiente.

[Ecuación 6]

$$P_m = \sum_{i=1}^M P_i, P_s = \sum_{i=1}^S P_i$$

En este caso, la probabilidad de música  $P_m$  puede calcularse sumando las probabilidades  $P_i$  de  $M$  mezclas relacionadas con parámetros de características superiores para la determinación de la música, mientras que la probabilidad de voz  $P_s$  puede calcularse sumando las probabilidades  $P_i$  de  $S$  mezclas relacionadas con parámetros de características superiores para la determinación de voz.

Mientras tanto, para mejorar la precisión, la probabilidad de música  $P_m$  y la probabilidad de voz  $P_s$  pueden ser calculadas de acuerdo con la siguiente Ecuación 7.

[Ecuación 7]

$$P_m = \sum_{i=1}^M p_i(1 - p_i^{err}) + \sum_{i=1}^S p_i(p_i^{err})$$

$$P_s = \sum_{i=1}^S p_i(1 - p_i^{err}) + \sum_{i=1}^M p_i(p_i^{err})$$

En este caso,  $p_i^{err}$  denota la probabilidad de error de cada mezcla. La probabilidad de error puede obtenerse clasificando los datos de entrenamiento que incluyen señales de voz limpias y señales de música limpias utilizando cada una de las mezclas y contando el número de clasificaciones incorrectas.

A continuación, la probabilidad  $P^M$  de que todas las tramas incluyan sólo señales de música y la probabilidad  $P^S$  de que todas las tramas incluyan sólo señales de voz con respecto a una pluralidad de tramas tan numerosas como una longitud de retención constante pueden calcularse de acuerdo con la siguiente Ecuación 8. La longitud de retención puede establecerse en 8, pero no está limitada a ello. Ocho tramas pueden incluir una trama actual y 7 tramas anteriores

[Ecuación 8]

$$p^M = \frac{\sum_{i=0}^{-7} p_m^{(i)}}{\sum_{i=0}^{-7} p_m^{(i)} + \sum_{i=0}^{-7} p_s^{(i)}}$$

$$p^S = \frac{\sum_{i=0}^{-7} p_s^{(i)}}{\sum_{i=0}^{-7} p_m^{(i)} + \sum_{i=0}^{-7} p_s^{(i)}}$$

A continuación, se puede calcular una pluralidad de conjuntos de condiciones  $\{D_i^M\}$  y  $\{D_i^S\}$  utilizando la probabilidad Pm de música o la probabilidad Ps de voz obtenida utilizando la Ecuación 5 o la Ecuación 6. A continuación se describen detalladamente los mismos con referencia a la Figura 6. En este caso, se puede establecer de tal manera que cada condición tenga un valor 1 para la música y un valor 0 para la voz.

Con referencia a la Figura 6, en una operación 610 y en una operación 620, se puede obtener una suma de condiciones de música M y una suma de condiciones de voz S a partir de la pluralidad de conjuntos de condiciones  $\{D_i^M\}$  y  $\{D_i^S\}$  que se calculan utilizando la probabilidad de música Pm y la probabilidad de voz Ps. En otras palabras, la suma de las condiciones de música M y la suma de las condiciones de voz S pueden expresarse como se muestra a continuación en la Ecuación 9.

[Ecuación 9]

$$M = \sum_i Q_i D_i^M$$

$$S = \sum_i Q_i D_i^S$$

En una operación 630, la suma de las condiciones de música M se compara con un valor de umbral designado Tm. Si la suma de condiciones de música M es mayor que el valor de umbral Tm, un modo de codificación de una trama actual se conmuta a un modo de música, es decir, al modo de codificación de dominio de espectro. Si la suma de las condiciones de música M es menor que o igual al valor de umbral Tm, el modo de codificación de la trama actual no se cambia.

En una operación 640, la suma de las condiciones de voz S se compara con un valor de umbral designado Ts. Si la suma de las condiciones de voz S es mayor que el valor de umbral Ts, un modo de codificación de una trama actual se conmuta a un modo de voz, es decir, el modo de codificación de dominio de predicción lineal. Si la suma de las condiciones de voz S es menor que o igual al valor de umbral Ts, el modo de codificación de la trama actual no se cambia.

El valor de umbral Tm y el valor de umbral Ts pueden establecerse en valores obtenidos con anterioridad a través de experimentos o simulaciones.

La Figura 5 es un diagrama de bloques que ilustra una configuración de una unidad 500 de extracción de parámetros de características de acuerdo con una realización ejemplar.

Una unidad 500 de determinación del modo de codificación inicial que se muestra en la Figura 5 puede incluir una unidad 510 de transformación, una unidad 520 de extracción de parámetros de espectro, una unidad 530 de extracción de parámetros temporales, y una unidad 540 de determinación.

En la Figura 5, la unidad 510 de transformación puede transformar una señal de audio original a partir de dominio de tiempo al dominio de la frecuencia. En este caso, la unidad 510 de transformación puede aplicar cualquiera de las diversas técnicas de transformación para representar una señal de audio desde un dominio de tiempo hasta un dominio de espectro. Los ejemplos de las técnicas pueden incluir la transformada rápida de Fourier (FFT), la transformada discreta de coseno (DCT), o la transformada discreta de coseno modificada (MDCT), pero no se limitan a ellas.

La unidad 520 de extracción de parámetros de espectro puede extraer al menos un parámetro de espectro a partir de una señal de audio de dominio de frecuencia proporcionada por la unidad 510 de transformación. Los parámetros de espectro pueden clasificarse en parámetros de características a corto plazo y parámetros de características a largo plazo. Los parámetros de características a corto plazo pueden obtenerse a partir de una trama actual, mientras que los parámetros de características a largo plazo pueden obtenerse a partir de una pluralidad de tramas que incluyen la trama actual y al menos una trama anterior.



La unidad 530 de extracción de parámetros temporales puede extraer al menos un parámetro temporal a partir de una señal de audio de dominio de tiempo. Los parámetros temporales también pueden clasificarse en parámetros de características a corto plazo y parámetros de características a largo plazo. Los parámetros de características a corto plazo pueden obtenerse a partir de una trama actual, mientras que los parámetros de características a largo plazo pueden obtenerse a partir de una pluralidad de tramas que incluyen la trama actual y al menos una trama anterior.

Una unidad de determinación (430 de la Figura 4) puede determinar la clase de una señal de audio utilizando los parámetros de espectro proporcionados por la unidad 520 de extracción de parámetros de espectro y los parámetros temporales proporcionados por la unidad 530 de extracción de parámetros temporales y puede determinar el modo de codificación inicial en base a la clase determinada. La unidad de determinación (430 de la Figura 4) puede emplear un mecanismo de decisión suave.

La Figura 7 es un diagrama que ilustra una operación de una unidad 310 de corrección del modo de codificación de acuerdo con una realización ejemplar.

Con referencia a la Figura 7, en una operación 700, se recibe un modo de codificación inicial determinado por la unidad 310 de determinación del modo de codificación inicial y se puede determinar si el modo de codificación es el modo de dominio de tiempo, es decir, el modo de excitación de dominio de tiempo o el modo de dominio de espectro.

En una operación 701, si se determina en la operación 700 que el modo de codificación inicial es el modo de dominio de espectro ( $\text{estado}_{TS} == 1$ ), puede comprobarse un  $\text{estado}_{TSS}$  de índice que indica si la codificación de excitación de dominio de frecuencia es más apropiada. El  $\text{estado}_{TSS}$  de índice que indica si la codificación de la excitación de dominio de frecuencia (por ejemplo, GSC) es más apropiada puede obtenerse utilizando tonalidades de diferentes bandas de frecuencia. Las descripciones detalladas de los mismos se darán a continuación.

La tonalidad de una señal de banda baja puede obtenerse como una relación entre una suma de una pluralidad de coeficientes de espectro que tienen valores pequeños, incluyendo el valor más pequeño, y el coeficiente de espectro que tiene el valor más grande con respecto a una banda dada. Si las bandas dadas son 0~1 kHz, 1~2 kHz, y 2~4 kHz, las tonalidades  $t_{01}$ ,  $t_{12}$  y  $t_{24}$  de las bandas respectivas y la tonalidad  $t_L$  de una señal de banda baja, es decir, la banda central, pueden expresarse como se muestra en la siguiente Ecuación 10.

[Ecuación 10]

$$t_{01} = 0,2 \log_{10} \left( \frac{\max(x_i)}{\prod_{j=0}^{Q-1} \text{sort}(x_j)} \right), i, jH[0, \dots, 1kHz]$$

$$t_{12} = 0,2 \log_{10} \left( \frac{\max(x_i)}{\prod_{j=0}^{Q-1} \text{sort}(x_j)} \right), i, jH[2, \dots, 2kHz]$$

$$t_{24} = 0,2 \log_{10} \left( \frac{\max(x_i)}{\prod_{j=0}^{Q-1} \text{sort}(x_j)} \right), i, jH[2, \dots, 4kHz]$$

$$t_L = \max(t_{01}, t_{12}, t_{24})$$

Mientras tanto, el error de predicción lineal  $\text{err}$  puede obtenerse utilizando un filtro de codificación de predicción lineal (LPC) y puede utilizarse para eliminar los componentes tonales fuertes. En otras palabras, el modo de codificación de dominio de espectro puede ser más eficiente con respecto a los componentes tonales fuertes que el modo de codificación de excitación de dominio de frecuencia.

Una condición frontal  $\text{cond}_{\text{front}}$  para conmutar al modo de codificación de excitación de dominio de frecuencia utilizando las tonalidades y el error de predicción lineal obtenido como se ha descrito anteriormente puede expresarse como se muestra a continuación en la Ecuación 11.

[Ecuación 11]

$$\text{cond}_{\text{front}} = t_{12} > t_{12\text{front}} \text{ y } t_{24} > t_{24\text{front}} \text{ y } t_L > t_{L\text{front}} \text{ y } \text{err} > \text{err}_{\text{front}}$$

En este caso,  $t_{12front}$ ,  $t_{24front}$ ,  $t_{Lfront}$  y  $err_{front}$  son valores de umbral y pueden tener valores obtenidos con anterioridad a través de experimentos o simulaciones.

Mientras tanto, una condición de retroceso  $cond_{back}$  para finalizar el modo de codificación de excitación de dominio de frecuencia utilizando las tonalidades y el error de predicción lineal obtenido como se ha descrito anteriormente puede expresarse como se muestra a continuación en la Ecuación 12.

[Ecuación 12]

$$cond_{back} = t_{12} < t_{12back} \text{ y } t_{24} < t_{24back} \text{ y } t_L < t_{Lback}$$

En este caso,  $t_{12back}$ ,  $t_{24back}$ ,  $t_{Lback}$  son valores de umbral y pueden tener valores obtenidos con anterioridad a través de experimentos o simulaciones.

10 En otras palabras, se puede determinar si el estado<sub>TTSS</sub> de índice que indica si la codificación de excitación de dominio de frecuencia (por ejemplo, GSC) es más apropiada que la codificación de dominio de espectro sea 1, determinando si la condición frontal que se muestra en la Ecuación 11 se satisface o la condición de retroceso que se muestra en la Ecuación 12 no se satisface. En este caso, la determinación de la condición de retroceso que se muestra en la ecuación 12 puede ser opcional.

15 En una operación 702, si el estado<sub>TTSS</sub> de índice, es 1, el modo de codificación de excitación de dominio de frecuencia puede determinarse como el modo de codificación final. En este caso, el modo de codificación de dominio de espectro, el cual es el modo de codificación inicial, se corrige al modo de codificación de excitación de dominio de frecuencia, el cual es el modo de codificación final.

20 En una operación 705, si se determina en la operación 701 que el estado<sub>TTSS</sub> de índice es 0, se puede comprobar un estado<sub>SS</sub> de índice para determinar si una señal de audio incluye una característica de voz fuerte. Si existe un error en la determinación del modo de codificación de dominio de espectro, el modo de codificación de excitación de dominio de frecuencia puede ser más eficiente que el modo de codificación de dominio de espectro. El estado<sub>SS</sub> de índice para determinar si una señal de audio incluye una característica de voz fuerte puede obtenerse utilizando una diferencia  $vc$  entre un parámetro de sonorización y un parámetro de correlación.

25 Una condición frontal  $cond_{front}$  para conmutar a un modo de voz fuerte utilizando la diferencia  $vc$  entre un parámetro de sonorización y un parámetro de correlación puede ser expresado como se muestra a continuación en la Ecuación 13.

[Ecuación 13]

$$cond_{front} = vc > vc_{front}$$

30 En este caso,  $vc_{front}$  es un valor de umbral y puede tener un valor obtenido con anterioridad a través de experimentos o simulaciones.

Mientras tanto, una condición de retroceso  $cond_{back}$  para finalizar el modo de voz fuerte utilizando la diferencia  $vc$  entre un parámetro de sonorización y un parámetro de correlación puede ser expresado como se muestra a continuación en la Ecuación 14.

35 [Ecuación 14]

$$cond_{back} = vc < vc_{back}$$

En este caso,  $vc_{back}$  es un valor de umbral y puede tener un valor obtenido con anterioridad a través de experimentos o simulaciones.

40 En otras palabras, en una operación 705, se puede determinar si el estado<sub>SS</sub> de índice que indica si la codificación de excitación de dominio de frecuencia (por ejemplo, GSC) es más apropiada que la codificación de dominio de espectro sea 1, determinando si la condición frontal que se muestra en la Ecuación 13 se satisface o la condición de retroceso que se muestra en la Ecuación 14 no se satisface. En este caso, la determinación de la condición de retroceso que se muestra en la Ecuación 14 puede ser opcional.

45 En una operación 706, si se determina en la operación 705 que el estado<sub>SS</sub> de índice es 0, es decir, que la señal de audio no incluye una característica de voz fuerte, el modo de codificación de dominio de espectro puede determinarse como el modo de codificación final. En este caso, el modo de codificación de dominio de espectro, que es el modo de codificación inicial, se mantiene como modo de codificación final.

50 En una operación 707, si se determina en la operación 705 que el estado<sub>SS</sub> de índice, es 1, es decir, que la señal de audio incluye una característica de voz fuerte, el modo de codificación de excitación de dominio de frecuencia puede determinarse como el modo de codificación final. En este caso, el modo de codificación de dominio de espectro, el cual es el modo de codificación inicial, se corrige al modo de codificación de excitación de dominio de frecuencia, el

cual es el modo de codificación final.

Al realizar las operaciones 700, 701 y 705, se puede corregir un error en la determinación del modo de codificación de dominio de espectro como el modo de codificación inicial. En detalle, el modo de codificación de dominio de espectro, el cual es el modo de codificación inicial, puede mantenerse o conmutarse al modo de codificación de excitación de dominio de frecuencia como el modo de codificación final.

Mientras tanto, si se determina en la operación 700 que el modo de codificación inicial es el modo de codificación de dominio de predicción lineal ( $\text{estado}_{\text{TS}} == 0$ ), se puede comprobar un  $\text{estado}_{\text{SM}}$  de índice para determinar si una señal de audio incluye una característica de música fuerte. Si existe un error en la determinación del modo de codificación de dominio de predicción lineal, es decir, el modo de codificación de excitación de dominio de tiempo, el modo de codificación de excitación de dominio de frecuencia puede ser más eficiente que el modo de codificación de excitación de dominio de tiempo. El  $\text{estado}_{\text{SM}}$  para determinar si una señal de audio incluye una característica de música fuerte puede obtenerse utilizando un valor 1-vc obtenido restando de 1 la diferencia vc entre un parámetro de sonorización y un parámetro de correlación.

Una condición frontal  $\text{cond}_{\text{front}}$  para conmutar a un modo de música fuerte utilizando el valor 1-vc obtenido al restar la diferencia vc entre un parámetro de sonorización y un parámetro de correlación de 1 puede expresarse como se muestra a continuación en la Ecuación 15.

[Ecuación 15]

$$\text{cond}_{\text{front}} = 1 - \text{vc} > \text{vcm}_{\text{front}}$$

En este caso,  $\text{vcm}_{\text{front}}$  es un valor de umbral y puede tener un valor obtenido con anterioridad a través de experimentos o simulaciones.

Mientras tanto, una condición de retroceso  $\text{cond}_{\text{back}}$  para finalizar el modo de música fuerte utilizando el valor 1-vc obtenido al restar la diferencia vc entre un parámetro de sonorización y un parámetro de correlación de 1 puede expresarse como se muestra a continuación en la Ecuación 16.

[Ecuación 16]

$$\text{cond}_{\text{back}} = 1 - \text{vc} < \text{vcm}_{\text{back}}$$

En este caso,  $\text{vcm}_{\text{back}}$  es un valor umbral y puede tener un valor obtenido con anterioridad a través de experimentos o simulaciones.

En otras palabras, en una operación 709, se puede determinar si el  $\text{estado}_{\text{SM}}$  de índice que indica si la codificación de excitación de dominio de frecuencia (por ejemplo, GSC) es más apropiada que la codificación de excitación de dominio de tiempo sea 1, determinando si la condición frontal mostrada en la Ecuación 15 se satisface o la condición de retroceso mostrada en la Ecuación 16 no se satisface. En este caso, la determinación de la condición de retroceso que se muestra en la Ecuación 16 puede ser opcional.

En una operación 710, si se determina en la operación 709 que el  $\text{estado}_{\text{SM}}$  de índice es 0, es decir, que la señal de audio no incluye una característica de música fuerte, el modo de codificación de excitación de dominio de tiempo puede determinarse como el modo de codificación final. En este caso, el modo de codificación de dominio de predicción lineal, el cual es el modo de codificación inicial, se conmuta al modo de codificación de excitación de dominio de tiempo como el modo de codificación final. De acuerdo con una realización ejemplar, puede considerarse que el modo de codificación inicial se mantiene sin cambios, si el modo de codificación de dominio de predicción lineal corresponde al modo de codificación de excitación de dominio de tiempo.

En una operación 707, si se determina en la operación 709 que el  $\text{estado}_{\text{SM}}$  de índice es 1, es decir, que la señal de audio incluye una característica de música fuerte, el modo de codificación de excitación de dominio de frecuencia puede determinarse como el modo de codificación final. En este caso, el modo de codificación de dominio de predicción lineal, el cual es el modo de codificación inicial, se corrige al modo de codificación de excitación de dominio de frecuencia, el cual es el modo de codificación final.

Realizando las operaciones 700 y 709, se puede corregir un error en la determinación del modo de codificación inicial. En detalle, el modo de codificación de dominio de predicción lineal (por ejemplo, el modo de codificación de excitación de dominio de tiempo), el cual es el modo de codificación inicial, puede mantenerse o conmutarse al modo de codificación de excitación de dominio de frecuencia como el modo de codificación final.

De acuerdo con una realización ejemplar, la operación 709 para determinar si la señal de audio incluye una característica de música fuerte para corregir un error en la determinación del modo de codificación de dominio de predicción lineal puede ser opcional.

De acuerdo con otra realización ejemplar, se puede invertir una secuencia de realización de la operación 705 para determinar si la señal de audio incluye una característica de voz fuerte y la operación 701 para determinar si el modo de codificación de excitación de dominio de frecuencia es apropiado. En otras palabras, después de la operación 700, se puede realizar primero la operación 705, y luego la operación 701. En este caso, los parámetros utilizados para las determinaciones pueden modificarse de acuerdo a como lo exijan las ocasiones.

La Figura 8 es un diagrama de bloques que ilustra una configuración de un aparato 800 de decodificación de audio de acuerdo con una realización ejemplar.

El aparato 800 de decodificación de audio que se muestra en la Figura 8 puede incluir una unidad 810 de análisis del flujo de bits, una unidad 820 de decodificación de dominio de espectro, una unidad 830 de decodificación de dominio de predicción lineal y una unidad 840 de conmutación. La unidad 830 de decodificación de dominio de predicción lineal puede incluir una unidad 831 de decodificación de excitación de dominio de tiempo y una unidad 833 de decodificación de excitación de dominio de frecuencia, donde la unidad 830 de decodificación de dominio de predicción lineal puede estar incorporada como al menos una de la unidad 831 de decodificación de excitación de dominio de tiempo y la unidad 833 de decodificación de excitación de dominio de frecuencia. A menos que sea necesario que se incorporen como un hardware separado, los componentes mencionados anteriormente pueden integrarse en al menos un módulo y pueden implementarse como al menos un procesador (no se muestra).

Con referencia a la Figura 8, la unidad 810 de análisis del flujo de bits puede analizar un flujo de bits recibido y separar la información sobre un modo de codificación y los datos codificados. El modo de codificación puede corresponder a un modo de codificación inicial obtenido mediante la determinación de uno de entre una pluralidad de modos de codificación que incluye un primer modo de codificación y un segundo modo de codificación en correspondencia con las características de una señal de audio o un tercer modo de codificación corregido a partir del modo de codificación inicial si existe un error en la determinación del modo de codificación inicial.

La unidad 820 de decodificación de dominio de espectro puede decodificar datos codificados en el dominio de espectro a partir de los datos codificados separados.

La unidad 830 de decodificación de dominio de predicción lineal puede decodificar datos codificados en el dominio de predicción lineal a partir de los datos codificados separados. Si la unidad 830 de decodificación de dominio de predicción lineal incluye la unidad 831 de decodificación de excitación de dominio de tiempo y la unidad 833 de decodificación de excitación de dominio de frecuencia, la unidad 830 de decodificación de dominio de predicción lineal puede realizar la decodificación de excitación de dominio de tiempo o la decodificación de excitación de dominio de frecuencia con respecto a los datos codificados separados.

La unidad 840 de conmutación puede conmutar una señal reconstruida por la unidad 820 de decodificación de dominio de espectro o una señal reconstruida por la unidad 830 de decodificación de dominio de predicción lineal y puede proporcionar la señal conmutada como una señal reconstruida final.

La Figura 9 es un diagrama de bloques que ilustra una configuración de un aparato 900 de decodificación de audio de acuerdo con otra realización ejemplar.

El aparato 900 de decodificación de audio puede incluir una unidad 910 de análisis del flujo de bits, una unidad 920 de decodificación de dominio de espectro, una unidad 930 de decodificación de dominio de predicción lineal, una unidad 940 de conmutación y un módulo 950 de procesamiento posterior común. La unidad 930 de decodificación de dominio de predicción lineal puede incluir una unidad 931 de decodificación de excitación de dominio de tiempo y una unidad 933 de decodificación de excitación de dominio de frecuencia, donde la unidad 930 de decodificación de dominio de predicción lineal puede estar incorporada como al menos una de la unidad 931 de decodificación de excitación de dominio de tiempo y la unidad 933 de decodificación de excitación de dominio de frecuencia. A menos que sea necesario que se incorporen como un hardware separado, los componentes mencionados anteriormente pueden integrarse en al menos un módulo y pueden implementarse como al menos un procesador (no se muestra). En comparación con el aparato 800 de decodificación de audio que se muestra en la Figura 8, el aparato 900 de decodificación de audio puede incluir además el módulo 950 de procesamiento posterior común, y, por lo tanto, se omitirán las descripciones de los componentes idénticos a los del aparato 800 de decodificación de audio.

Con referencia a la Figura 9, el módulo 950 de procesamiento posterior común puede realizar un procesamiento estéreo conjunto, un procesamiento de sonido envolvente, y/o un procesamiento de extensión de ancho de banda, en correspondencia con un módulo de preprocesamiento común (205 de la Figura 2).

Los procedimientos de acuerdo con las realizaciones ejemplares pueden escribirse como programas ejecutables por ordenador y pueden implementarse en ordenadores digitales de uso general que ejecutan los programas utilizando un medio de grabación no transitorio legible por ordenador. Además, las estructuras de datos, las instrucciones de programa, o los archivos de datos, los cuales pueden utilizarse en las realizaciones, pueden grabarse en un medio de grabación no transitorio legible por ordenador de varias maneras. El medio de grabación no transitorio legible por ordenador es cualquier dispositivo de almacenamiento de datos que puede almacenar datos los cuales pueden ser leídos posteriormente por un sistema informático. Los ejemplos del medio de grabación no transitorio legible por ordenador incluyen los medios de almacenamiento magnético, tales como los discos duros, los disquetes, y las cintas

5 magnéticas, los medios de grabación óptica, tales como los CD-ROM y los DVDs, los medios ópticos magnéticos, tales como los discos ópticos, y los dispositivos de hardware, tales como la ROM, la RAM, y la memoria *flash*, especialmente configurados para almacenar y ejecutar instrucciones de programa. Además, el medio de grabación no transitorio legible por ordenador puede ser un medio de transmisión para transmitir la señal que designa las instrucciones del programa, las estructuras de datos, o similares. Los ejemplos de instrucciones de programa pueden incluir no sólo códigos de lenguaje mecánico creados por un compilador, sino también códigos de lenguaje de alto nivel ejecutables por un ordenador utilizando un intérprete o similar.

10 Si bien se han mostrado y descrito de manera particular las anteriores realizaciones ejemplares, se entenderá por aquellos expertos en la técnica que se pueden hacer varios cambios en la forma y los detalles de las mismas sin apartarse del ámbito del concepto inventivo como se define por las reivindicaciones adjuntas. Las realizaciones ejemplares deben considerarse únicamente en sentido descriptivo y no con fines de limitación. Por lo tanto, el ámbito del concepto inventivo se define no por la descripción detallada de las realizaciones ejemplares sino por las reivindicaciones adjuntas, y todas las diferencias dentro del ámbito se interpretarán como incluidas en el presente concepto inventivo.

15

**REIVINDICACIONES**

1. Un aparato para determinar un modo de codificación, comprendiendo el aparato:

- 5 una unidad de determinación del modo de codificación inicial (310) configurada para determinar un modo de codificación inicial de una trama actual como un modo de codificación de dominio de espectro o un modo de codificación de dominio de tiempo utilizando características de una señal de audio, en donde el módulo de codificación inicial de la trama actual se determina como el modo de codificación de dominio de espectro cuando la señal de audio es clasificada como una señal de música, y el modo de codificación inicial de la trama actual se determina como el modo de codificación de dominio de tiempo cuando la señal de audio es clasificada como una señal de voz; y
- 10 una unidad de corrección del modo de codificación (330) configurada para corregir el modo de codificación inicial a un modo de Codificación Genérica de la Señal, GSC, modo para la codificación de excitación basado en parámetros de corrección, y determinar un modo de codificación final de la trama actual de entre el modo de codificación inicial y el modo de GSC en base a los modos de codificación de tramas correspondientes a una longitud de retención, de manera que se evita una conmutación frecuente del modo de codificación.
- 15

FIG. 1

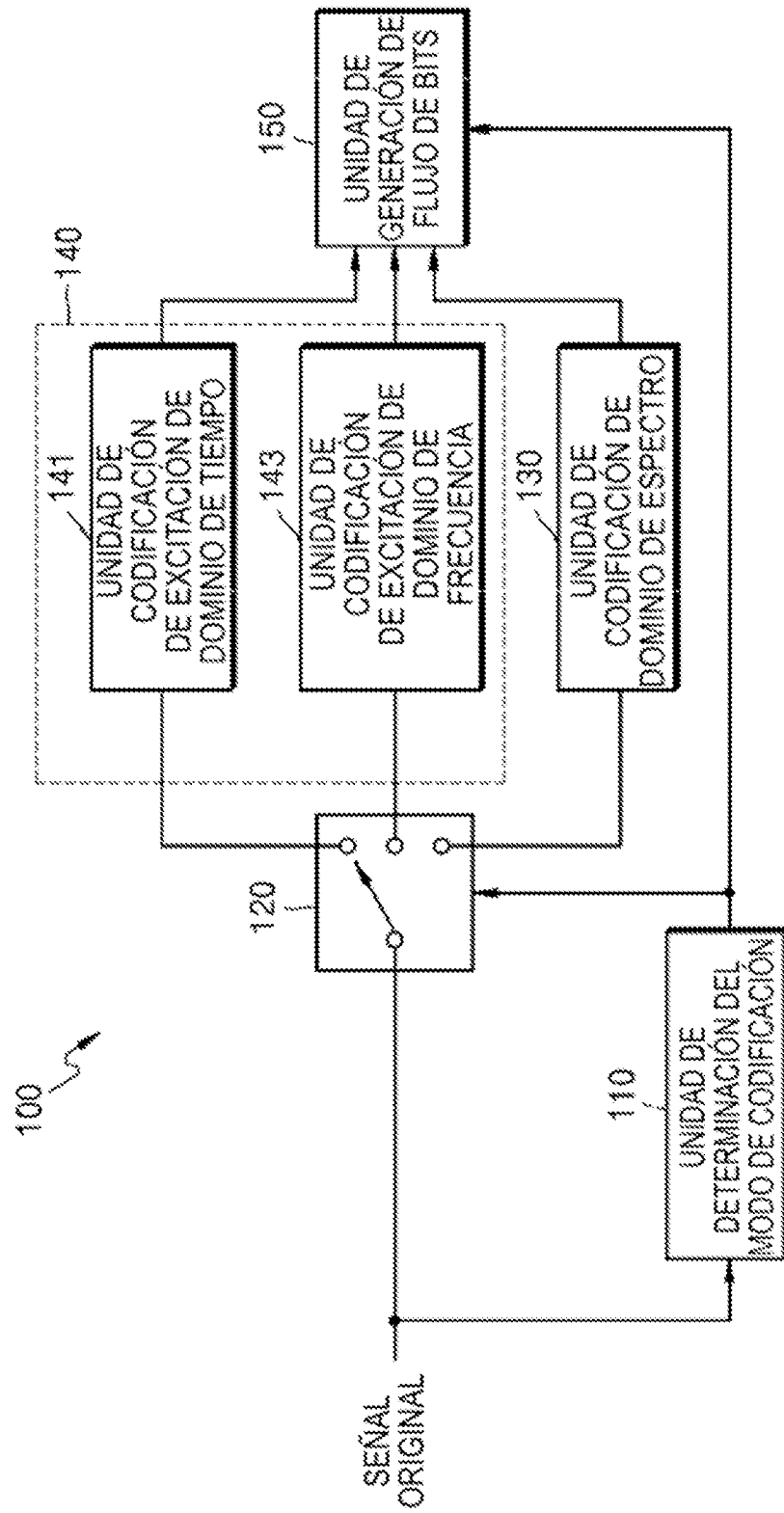


FIG. 2

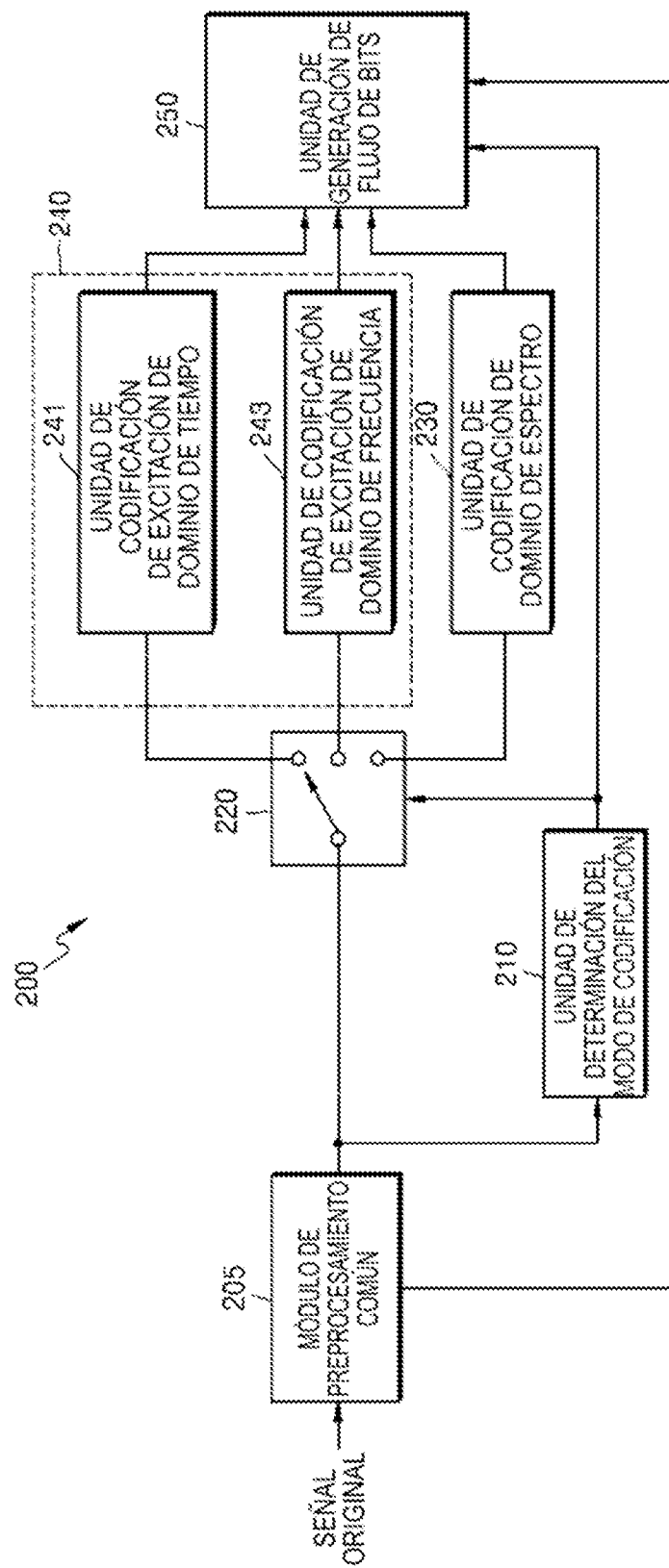




FIG. 3

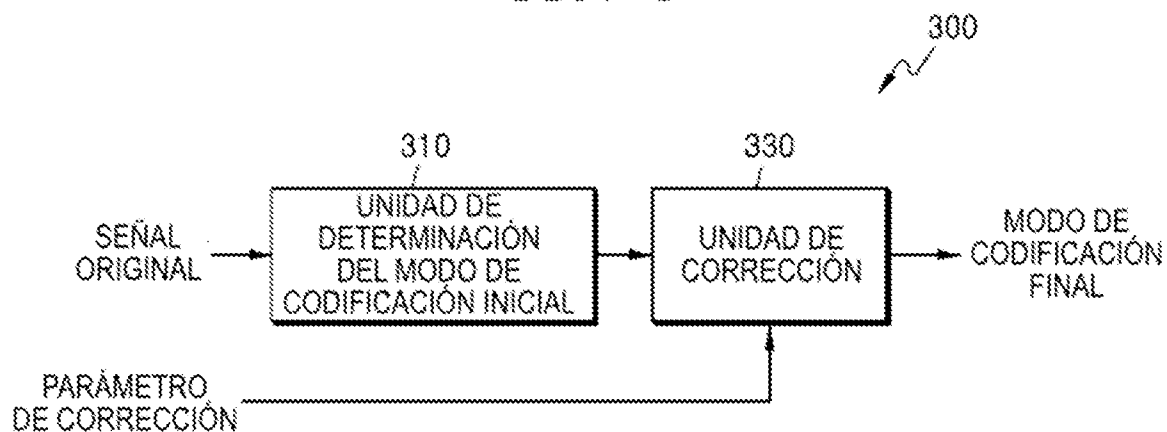


FIG. 4



FIG. 5

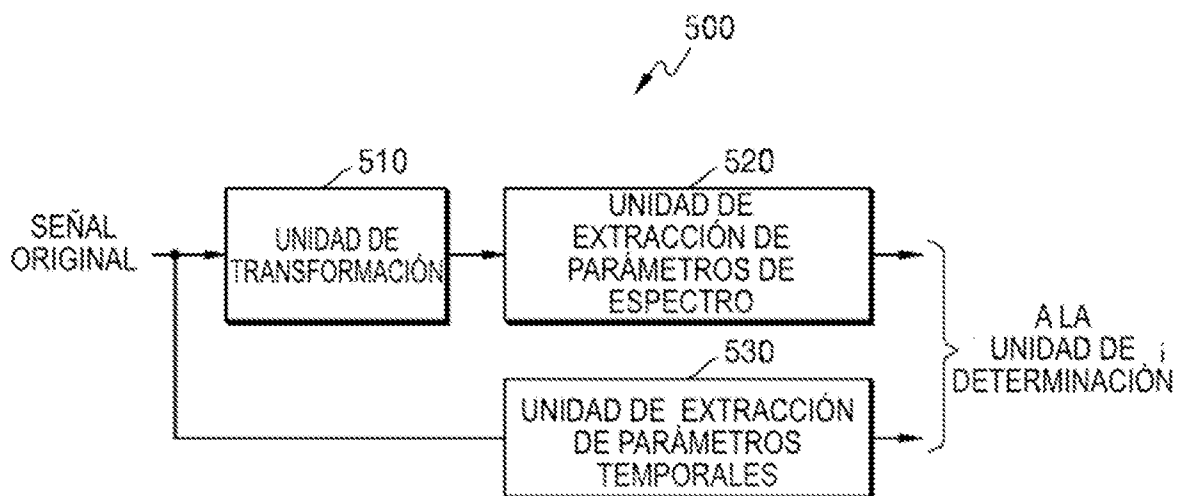


FIG. 6

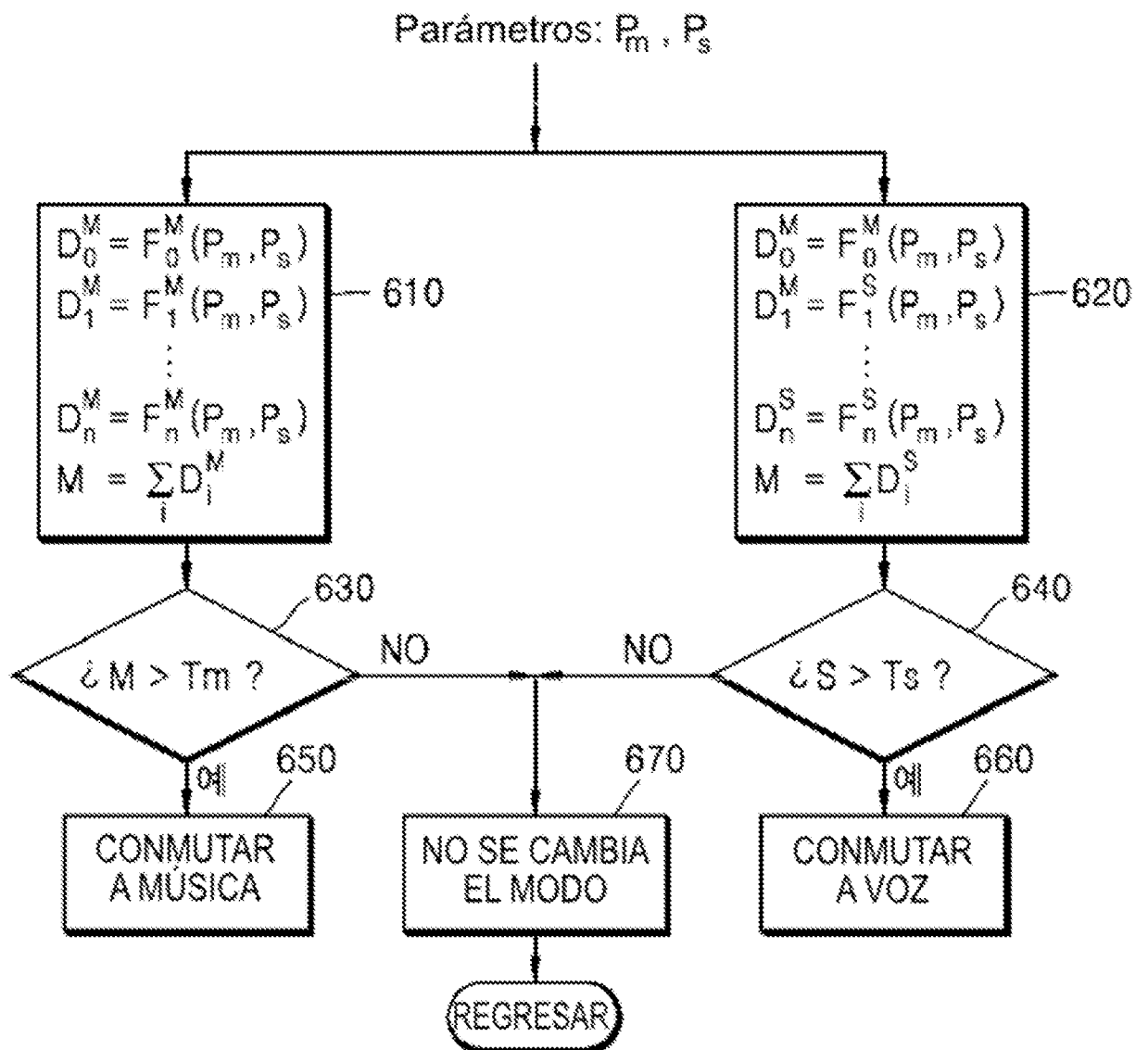


FIG. 7

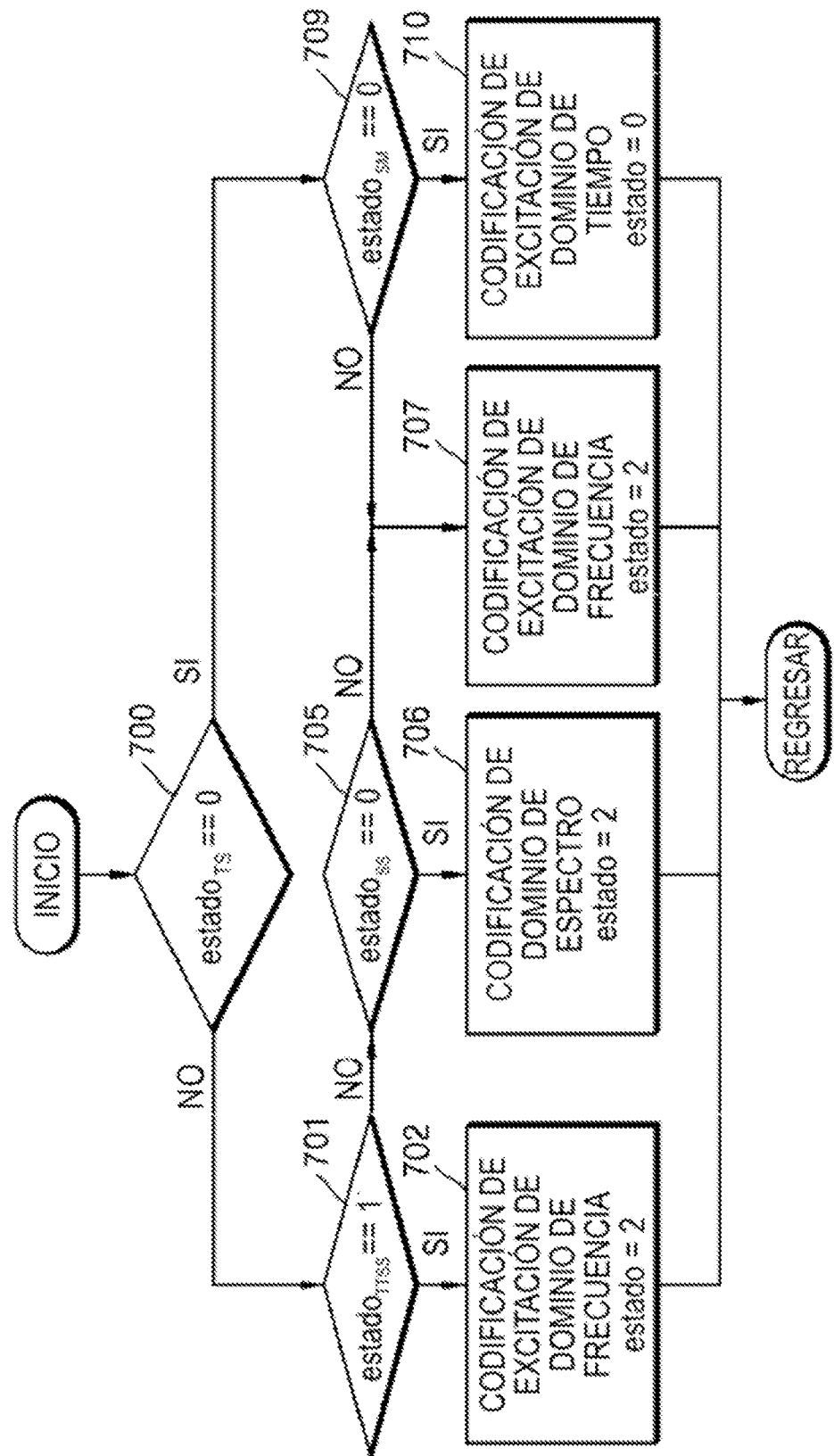


FIG. 8

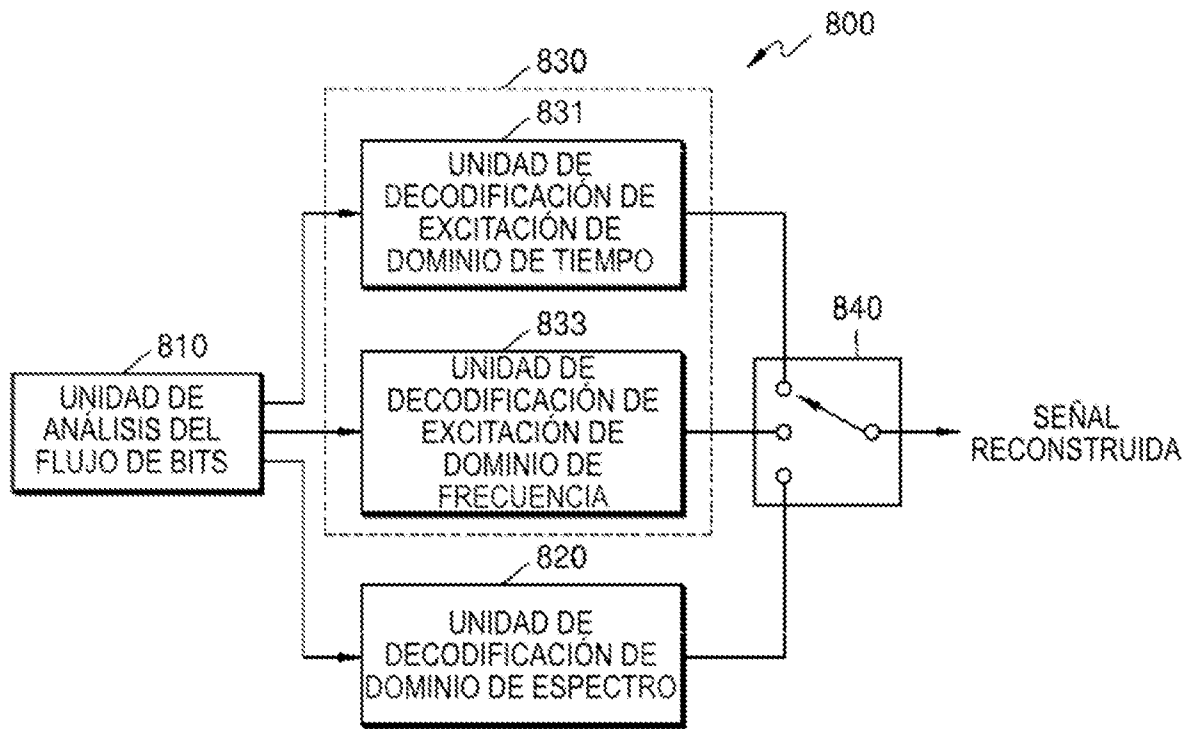


FIG. 9

