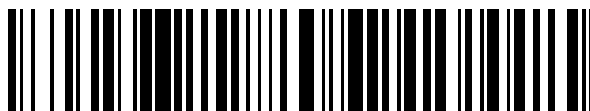


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 703 873**

51 Int. Cl.:

G10L 19/028 (2013.01)

G10L 19/038 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **30.10.2012 E 17164481 (8)**

97 Fecha y número de publicación de la concesión europea: **26.09.2018 EP 3220390**

54 Título: **Codificación/descodificación de la transformada de señales armónicas de audio**

30 Prioridad:

29.03.2012 US 201261617216 P

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

12.03.2019

73 Titular/es:

**TELEFONAKTIEBOLAGET LM ERICSSON (PUBL)
(100.0%)
164 83 Stockholm, SE**

72 Inventor/es:

**GRANCHAROV, VOLODYA;
NÄSLUND, SEBASTIAN;
JANSSON TOFTGÅRD, TOMAS y
POBLOTH, HARALD**

74 Agente/Representante:

ELZABURU, S.L.P

ES 2 703 873 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

Codificación/descodificación de la transformada de señales armónicas de audio

Campo técnico

5 La tecnología propuesta se refiere a la codificación/descodificación de la transformada de señales de audio, especialmente las señales armónicas de audio.

Antecedentes

10 La codificación de la transformada es la principal tecnología que se utiliza para comprimir y transmitir señales de audio. El concepto de codificar la transformada es convertir en primer lugar una señal al dominio en frecuencia, y a continuación cuantificar y transmitir los coeficientes de la transformada. El codificador utiliza los coeficientes recibidos de la transformada para reconstruir la forma de onda de la señal aplicando la transformada inversa de la frecuencia, véase la figura 1. En la figura 1 una señal de audio $X(n)$ se envía a un transformador de frecuencia 10. La transformada de frecuencia resultante $Y(k)$ se reenvía a un codificador de la transformada 12, y la transformada codificada se transmite al descodificador, en el cual se descodifica por medio de un descodificador de la transformada 14. La transformada descodificada $\hat{Y}(k)$ se reenvía a un transformador inverso de frecuencia 16 que la
15 transforma en una señal de audio descodificada $\hat{X}(k)$. La motivación existente tras este esquema es que los coeficientes en el dominio en frecuencia se pueden cuantificar más eficazmente, por las siguientes razones:

- 1) Los coeficientes de la transformada ($Y(k)$ en la figura 1) están menos correlacionados que las muestras de la señal de entrada ($X(n)$ en la figura 1).
- 2) La transformada en frecuencia proporciona compresión de la energía (más coeficientes ($Y(k)$ se aproximan a cero y se pueden desprestigiar), y
- 3) La motivación subjetiva que existe tras la transformación es que el sistema de audición humano opera en un dominio transformado, y es más sencillo seleccionar los componentes perceptualmente importantes de la señal en ese dominio.

25 En un códec típico de la transformada la forma de onda de la señal se transforma sobre la base de bloque a bloque (con un solape del 50%), utilizando la Transformada Discreta del Coseno Modificada (MDCT). En un códec de transformada tipo MDCT se transforma una forma de onda del bloque de la señal ($X(n)$ en un vector MDCT ($Y(k)$). La longitud de los bloques de la forma de onda corresponde a segmentos de audio de 20-40 milisegundos. Si la longitud viene definida por $2L$, la transformada MDCT se puede definir como:

$$Y(k) = \sqrt{\frac{2}{L}} \sum_{n=0}^{2L-1-k} \sin\left[\left(n + \frac{1}{2}\right)\frac{\pi}{L}\right] \cos\left[\left(n + \frac{1}{2} + \frac{L}{2}\right)\left(k + \frac{1}{2}\right)\frac{\pi}{L}\right] X(n) \quad (1)$$

30 para $k = 0, \dots, L-1$. A continuación el vector MDCT ($Y(k)$ se divide en múltiples bandas (subvectores), y la energía (o ganancia) $G(j)$ en cada banda se calcula como:

$$G(j) = \sqrt{\frac{1}{N_j} \sum_{k=m_j}^{m_j+N_j-1} Y^2(k)} \quad (2)$$

35 en donde m_j es el primer coeficiente en la banda J y N_j se refiere al número de coeficientes MDCT en las bandas correspondientes (un intervalo típico contiene 8-32 coeficientes). Como ejemplo de una estructura uniforme de banda, sea $N_j = 8$ para todos los j , entonces $G(0)$ sería la energía de los primeros 8 coeficientes, $G(1)$ sería la energía de los siguientes 8 coeficientes, etc.

40 Estos valores de energía o ganancias dan una aproximación de la envolvente del espectro, que está cuantificada y los índices de cuantificación se transmiten al descodificador. Los subvectores residuales o formas se obtienen por escalado de los vectores MDCT con las correspondientes ganancias de la envolvente, es decir, lo residual en cada banda se escala para tener la energía en valor eficaz (RMS) de la unidad. A continuación los subvectores residuales o formas se cuantifican con diferentes números de bits basándose en las correspondientes ganancias de la envolvente. Finalmente, en el descodificador, el vector MDCT se reconstruye escalando hacia arriba los subvectores residuales o formas con las correspondientes ganancias de la envolvente, y se utiliza una MDCT inversa para reconstruir la trama de audio en el dominio en el tiempo.

El concepto convencional de codificar la transformada no trabaja bien con señales de audio de gran contenido de armónicos, por ejemplo, instrumentos aislados. Un ejemplo de tal espectro de armónicos se ilustra en la figura 2 (para la comparación se muestra en la figura 3 un típico aspecto de audio sin excesivos armónicos). La razón es que la normalización con la envolvente del espectro no da lugar a un vector residual suficientemente "plano", y que el esquema residual de la codificación no puede producir una señal de audio de una calidad aceptable. La disparidad entre la señal y el modelo de codificación se puede resolver sólo a muy altas velocidades de bits, pero en la mayoría los casos esta solución no resulta apropiada.

La patente US 2012/0029923 describe un esquema para codificar un conjunto de coeficientes de la transformada que representan un intervalo de frecuencia de audio de una señal que utiliza un modelo armónico para parametrizar una relación entre puntos de regiones de energía significativa en el dominio en frecuencia.

Resumen

Un objeto de la tecnología propuesta es un esquema que es más apropiado para la codificación de una transformada para señales armónicas de audio.

La tecnología propuesta implica un aparato de codificar mediante la Transformada Discreta del Coseno Modificada, MDCT, los coeficientes ($Y(k)$) de una señal de armónicos de audio. El aparato comprende medios para localizar picos del espectro que tengan magnitudes que excedan un umbral predeterminado, en las que los picos del espectro se localizan mediante la comparación de los coeficientes con dicho umbral para formar un vector de candidatos de picos, y extraer los elementos del vector de candidatos de picos en orden decreciente, en los que dicho umbral se calcula como

$$\theta = \left(\frac{\bar{E}_p}{\bar{E}_{nf}} \right)^\gamma \bar{E}_{nf},$$

en donde \bar{E}_p es una energía media del pico, \bar{E}_{nf} es una energía media del suelo de ruido y γ tiene un valor fijo predeterminado, y en donde la energía del pico se calcula como $E_p(k) = \beta E_p(k) + (1-\beta)|Y(k)|$ y la energía del suelo de ruido se calcula como $E_{nf}(k) = \alpha E_{nf}(k) + (1-\alpha)|Y(k)|$, en donde la contribución de los coeficientes de alta energía se enfatiza en el cálculo de la energía del pico y la contribución de los coeficientes de baja energía se enfatiza en el cálculo de la energía del suelo de ruido. El aparato comprende además medios para codificar las regiones de picos incluyendo y rodeando los picos localizados, en donde los picos del espectro se cuantifican junto con los contenedores MDCT vecinos; medios para codificar, utilizando un número de bits reservados, un primer conjunto de coeficientes de baja frecuencia fuera de las regiones de picos y por debajo de una frecuencia de cruce que depende del número de bits utilizados para codificar las regiones de picos, y para codificar uno o más conjuntos adicionales de coeficientes de baja frecuencia fuera de las regiones de picos si existen bits no reservados disponibles después de codificar las regiones de picos; y medios para codificar, utilizando un número de bits reservados, una ganancia del suelo de ruido de al menos un conjunto de coeficientes de alta frecuencia aún no codificados fuera de las regiones de picos.

La tecnología propuesta también implica un equipo de usuario (UE) que comprende dicho aparato.

La tecnología propuesta también implica un código de programa de ordenador. El código de programa de ordenador, cuando se ejecuta por un procesador, provoca que un aparato localice picos del espectro que tienen magnitudes que exceden un umbral predeterminado, en el que los picos del espectro se localizan mediante la comparación de los coeficientes con dicho umbral para formar un vector de candidatos de picos, y extraen los elementos del vector de candidatos de picos en orden decreciente, en el que dicho umbral se calcula como

$$\theta = \left(\frac{\bar{E}_p}{\bar{E}_{nf}} \right)^\gamma \bar{E}_{nf},$$

en donde \bar{E}_p es una energía media de los picos, \bar{E}_{nf} es una energía media del suelo de ruido y γ tiene un valor fijo predeterminado, y en el que la energía de los picos se calcula como $E_p(k) = \beta E_p(k) + (1-\beta)|Y(k)|$ y la energía del suelo de ruido se calcula como $E_{nf}(k) = \alpha E_{nf}(k) + (1-\alpha)|Y(k)|$, en el que la contribución de los coeficientes de alta energía se enfatiza en el cálculo de la energía de los picos y la contribución de los coeficientes de baja energía se enfatiza en el cálculo de la energía del suelo de ruido. El código de programa de ordenador provoca además que un aparato codifique las regiones de picos incluyendo y rodeando los picos localizados, en donde los picos del espectro se cuantifican junto con los contenedores MDCT vecinos; codifique, utilizando un número de bits reservados, un

5 primer conjunto de coeficientes de baja frecuencia, LF, fuera de las regiones de picos y por debajo de una frecuencia de cruce que depende del número de bits utilizados para codificar las regiones de picos, y en el que la codificación comprende codificar uno o más conjuntos adicionales de coeficientes de baja frecuencia fuera de las regiones de picos si existen bits no reservados disponibles después de codificar las regiones de picos; y codificar, utilizando un número de bits reservados, una ganancia del suelo de ruido de al menos un conjunto de coeficientes de alta frecuencia aún no codificados fuera de las regiones de picos.

El esquema de codificación propuesto para codificar señales armónicas de audio proporciona mejor calidad perceptual que los esquemas de codificación convencionales para una amplia clase de señales armónicas de audio.

Breve descripción de los dibujos

10 La presente tecnología, junto con los objetivos y ventajas adicionales de la misma, se puede comprender mejor haciendo referencia a la siguiente descripción en conjunto con los dibujos que se acompañan, en los cuales:

La figura 1 ilustra el concepto de codificación de la transformada en frecuencia;

La figura 2 ilustra un espectro típico que una señal armónica de audio;

La figura 3 ilustra un espectro típico de una señal no armónica de audio;

15 La figura 4 ilustra una región de picos;

La figura 5 es un diagrama de flujo que ilustra el método de codificación propuesto;

La figura 6A-D ilustra una realización de ejemplo del método de codificación propuesto;

La figura 7 es un diagrama de bloques de una realización de ejemplo del codificador propuesto;

La figura 8 es un diagrama de flujo que ilustra el método de descodificación propuesto;

20 La figura 9A-C ilustra una realización de ejemplo del método de descodificación propuesto;

La figura 10 es un diagrama de bloques de una realización de ejemplo del descodificador propuesto;

La figura 11 es un diagrama de bloques de una realización de ejemplo del codificador propuesto;

La figura 12 es un diagrama de bloques de una realización de ejemplo del descodificador propuesto;

25 La figura 13 es un diagrama de bloques de una realización de ejemplo de un UE que incluye el codificador propuesto;

La figura 14 es un diagrama de bloques de una realización de ejemplo de un UE que incluye el descodificador propuesto;

La figura 15 es un diagrama de flujo de una realización de ejemplo de una parte del método de codificación propuesto;

30 La figura 16 es un diagrama de bloques de una realización de ejemplo de un codificador de la región de picos del codificador propuesto;

La figura 17 es un diagrama de flujo de una realización de ejemplo de una parte del método de descodificación propuesto;

35 La figura 16 es un diagrama de bloques de una realización de ejemplo de un descodificador de la región de picos en el descodificador propuesto.

Descripción Detallada

40 La figura 2 ilustra un espectro típico de una señal armónica de audio, y la figura 3 ilustra un espectro típico de una señal no armónica de audio. El espectro de la señal armónica está formado por fuertes picos del espectro separados por bandas de frecuencia mucho más débiles, mientras que el espectro de la señal no armónica de audio es mucho más liso.

45 La tecnología propuesta proporciona un modelo alternativo de codificación de audio que trata mejor las señales armónicas de audio. El principal concepto es que el vector de la transformada en frecuencia, por ejemplo un vector MDCT, no está dividido en parte envolvente y parte residual, sino que, al contrario, los picos del espectro se extraen directamente y se cuantifican, junto con los contenidos MDCT vecinos. A altas frecuencias, los coeficientes de baja energía fuera de los picos de la vecindad no se codifican, pero se rellenan de ruido en el descodificador. Aquí el modelo de señal utilizado en la codificación convencional, (envolvente del espectro + residuo) se reemplaza por un

nuevo modelo (picos del espectro + suelo de ruido). A bajas frecuencias, los coeficientes fuera de los picos de la vecindad aún se codifican, dado que ellos tienen un importante papel perceptual.

Codificador

Las principales etapas de la parte codificadora son:

- 5 • Localizar y codificar las regiones de picos del espectro
- Codificar los coeficientes de baja frecuencia (LF) del espectro. El tamaño de la región codificada depende del número de bits que resten tras la codificación de la región de los picos.
- Codificar las ganancias del suelo de ruido para los coeficientes del espectro fuera de las regiones de los picos.

10 En primer lugar se estima el suelo de ruido, a continuación se extraen los picos del espectro por medio de un algoritmo de extracción de picos (los correspondientes algoritmos se describen con más detalle en el APENDICE I-II. Cada pico y sus 4 vecinos que lo rodean se normalizan a la energía unidad en la posición del pico, véase figura 4. En otras palabras, la región entera se escala de tal forma que el pico tenga amplitud unidad. La posición del pico, la ganancia (representa la amplitud del pico, la magnitud) y el signo se cuantifican. Se aplica un Cuantificador de Vectores (VQ) a los contenedores MDCT que rodean el pico y se busca el índice I_{shape} del vector del libro de códigos

15 que proporcione la mejor coincidencia. La posición del pico, la ganancia y el signo, así como los vectores de forma que los rodean se cuantifican y los índices de cuantificación $\{ I_{position} I_{gain} I_{sign} I_{shape} \}$ se transmiten al descodificador. En adición a estos índices el descodificador también es informado de número total de picos.

En el ejemplo anterior cada región de picos incluye cuatro vecinos que rodean simétricamente el pico. Sin embargo es también factible tener tanto menos como más vecinos rodeando el pico de forma simétrica o asimétrica.

20 Después de cuantificar las regiones de picos, todos los bits restantes disponibles (excepto los bits reservados para la codificación del suelo de ruido, véase más adelante) se utilizan para cuantificar los coeficientes MDCT de baja frecuencia. Esto se hace agrupando los coeficientes MDCT restantes no cuantificados en, por ejemplo, 24 bandas dimensionales que comienzan a partir del primer contenedor. Por ello, estas bandas cubrirán las menores frecuencias hasta una cierta frecuencia de corte. Los coeficientes que ya han sido cuantificados en la codificación de

25 los picos no se incluyen, de modo que las bandas no están hechas necesariamente a partir de los 24 coeficientes consecutivos. Por esta razón las bandas serán referidas también a continuación como "conjuntos".

El número total de bandas de LF o conjuntos depende del número de bits disponibles, pero siempre hay bastantes bits reservados para crear al menos un conjunto. Cuando hay más bits disponibles, el primer conjunto adquiere más bits asignados hasta que se alcanza un umbral para el máximo número de bits por conjunto. Si hay más bits disponibles, se crea otro conjunto y se asignan los bits a este conjunto hasta que se alcanza el umbral. Este procedimiento se repite hasta que se utilizan todos los bits disponibles. Esto quiere decir que la frecuencia de cruce a la cual este proceso se detiene dependerá de las tramas, ya que el número de picos variará de trama en trama. La frecuencia de cruce estará determinada por el número de bits disponibles para la codificación de LF una vez que se han codificado las regiones de picos.

35 La cuantificación de los conjuntos de LF se puede hacer con cualquier esquema adecuado de cuantificación de vectores, pero normalmente se utiliza algún tipo de codificación de la forma de la señal. Por ejemplo, se puede utilizar la codificación factorial de impulsos para el vector de la forma, y se puede utilizar un cuantificador escalar para la ganancia.

40 Se reserva siempre un cierto número de bits para codificar una ganancia del suelo de ruido de al menos una banda de coeficientes de alta frecuencia fuera de las regiones de picos, y por encima de la frecuencia superior de las bandas de LF. Preferiblemente se utilizan dos ganancias para este fin. Estas ganancias se pueden obtener del algoritmo del suelo de ruido descrito en el APENDICE I. Si se utiliza la codificación factorial de impulsos para codificar las bandas de baja frecuencia algunos coeficientes de LF pueden no ser codificados. Estos coeficientes pueden en cambio incluirse en la codificación de la banda de alta frecuencia. Como en el caso de las bandas de LF,

45 las bandas de HF no están necesariamente constituidas por coeficientes consecutivos. Por esta razón, las bandas serán también referidas a continuación como "conjuntos".

Si fuera aplicable, la envolvente del espectro para una región de extensión de banda ancha (BWE) también se codifica y se transmite. El número de bandas (y la frecuencia de transición en la que se inicia la BWE) depende de la velocidad de bits, por ejemplo, 5,6 kHz a 24 kbps y 6,4 kHz a 32 kbps,

50 La figura 5 es un diagrama de flujo que ilustra el método propuesto de codificación desde una perspectiva general. La etapa S1 localiza los picos del espectro que tienen magnitudes que superan un umbral predeterminado dependiente de la frecuencia. La etapa S2 codifica las regiones de picos que incluyen y rodean los picos localizados. La etapa S3 codifica al menos un conjunto de coeficientes de baja frecuencia fuera de las regiones de picos y por debajo de una frecuencia de cruce que depende del número de bits utilizados para codificar las regiones de picos.

La etapa S4 codifica una ganancia del suelo de ruido de al menos un conjunto de coeficientes de alta frecuencia aún no codificados (aún no codificados o restantes) fuera de las regiones de picos.

5 La figura 6A-D ilustra una realización de ejemplo del método de codificación propuesto. La figura 6A ilustra la transformada MDCT de la trama de la señal a codificar. En la figura existen menos coeficientes que en una señal real. Sin embargo, se debe tener en cuenta que el propósito de la figura es sólo ilustrar el procedimiento de codificación. La figura 6B ilustra 4 regiones de picos identificadas listas para codificar la forma de la ganancia. El método descrito en el APENDICE II se puede utilizar para encontrarlas. A continuación se recogen los coeficientes de LF fuera de las regiones de picos en la figura 6C. Estos se concatenan en bloques que codifican la forma de la ganancia. Los coeficientes restantes de la señal original en la figura 6A son los coeficientes de alta frecuencia ilustrados en la figura 6D. Estos se dividen en 2 conjuntos y se codifican (como bloques concatenados) por medio de una ganancia del suelo de ruido para cada conjunto. Esta ganancia del suelo de ruido se puede obtener a partir de la energía de cada conjunto o por estimaciones obtenidas a partir del algoritmo de estimación del suelo de ruido descrito en el APENDICE I.

15 La figura 7 es un diagrama de bloques de una realización de ejemplo de un codificador propuesto 20. Un localizador de picos 22 está configurado para localizar los picos del espectro que tienen magnitudes que exceden un umbral predeterminado dependiente de la frecuencia. Un codificador de regiones de picos 24 está configurado para codificar regiones de picos que incluyen y rodean los picos extraídos. Un codificador de conjuntos de baja frecuencia 26 está configurado para codificar al menos un conjunto de coeficientes de baja frecuencia fuera de las regiones de picos y por debajo de una frecuencia de cruce que depende del número de bits utilizados para codificar las regiones de picos. Un codificador de la ganancia del suelo de ruido 28 está configurado para codificar una ganancia del suelo de ruido de al menos un conjunto de coeficientes de baja frecuencia aún no codificados fuera de las regiones de picos. En esta realización, los codificadores 24, 26, 28 utilizan la posición de los picos detectados para decidir qué coeficientes incluir en la codificación respectiva.

Descodificador

25 Las principales etapas en el descodificador son:

- Reconstruir las regiones de picos del espectro
- Reconstruir los coeficientes del espectro de LF
- Rellenar con ruido las regiones no codificadas, escaladas con las ganancias del suelo de ruido recibidas.

30 El descodificador de audio extrae, procedente del flujo de bits, el número de regiones de picos y los índices de cuantificación $\{ I_{position} I_{gain} I_{sign} I_{shape} \}$ con objeto de reconstruir las regiones de picos codificadas. Estos índices de cuantificación contienen información acerca de la posición de los picos del espectro, la ganancia y signo del pico, así como el índice para el vector del libro de códigos que proporcione la mejor coincidencia para la vecindad del pico.

Los coeficientes de baja frecuencia de MDCT fuera de las regiones de picos se reconstruyen a partir de los coeficientes de LF codificados.

35 Los coeficientes de alta frecuencia de MDCT fuera de las regiones de los picos se rellenan con ruido en el descodificador. El nivel del suelo de ruido se recibe por el descodificador, preferiblemente en forma de dos ganancias del suelo de ruido codificadas (una para la mitad inferior y otra para la mitad superior o parte del vector).

Si resulta aplicable, el descodificador de audio realiza una BWE a partir de una frecuencia de transición predefinida con las ganancias de la envolvente recibidas para los coeficientes de HF de MDCT.

40 La figura 8 es un diagrama de flujo que ilustra el método de descodificación propuesto desde una perspectiva general. La etapa S11 descodifica las regiones de picos del espectro de la señal armónica de audio codificada transformada en frecuencia. La etapa S12 descodifica al menos un conjunto de coeficientes de baja frecuencia. La etapa S13 distribuye los coeficientes de cada conjunto de baja frecuencia fuera de las regiones de picos. La etapa S14 descodifica una ganancia del suelo de ruido de al menos un conjunto de coeficientes de alta frecuencia fuera de las regiones de picos. La etapa S15 rellena cada conjunto de alta frecuencia con ruido que tiene la correspondiente ganancia del suelo de ruido.

En una realización de ejemplo, la descodificación de un conjunto de baja frecuencia se basa en un esquema de descodificación de la forma de la ganancia.

50 En una realización de ejemplo el esquema de descodificación de la forma de la ganancia se basa en la descodificación escalar de la ganancia y en la descodificación factorial de la forma del impulso.

Una realización de ejemplo incluye la etapa de descodificar una ganancia del suelo de ruido para cada uno de los dos conjuntos de alta frecuencia.

- La figura 9A-C ilustra una realización de ejemplo del método de descodificación propuesto. La reconstrucción de la transformada en frecuencia se inicia con la descodificación de la forma de la ganancia de las regiones de picos del espectro y sus posiciones, como se ilustra en la figura 9A. En la figura 9B el (los) conjunto(s) de LF descodifica(n) la forma de la ganancia y los coeficientes codificados de la transformada se distribuyen en bloques fuera de las regiones de picos. En la figura 9C la ganancia del suelo de ruido se descodifica y los coeficientes restantes de la transformada se rellenan con ruido que tiene las correspondientes ganancias del suelo de ruido. De este modo, se ha reconstruido aproximadamente la transformada de la figura 6A. Una comparación de la figura 9C con las figuras 6A y 6D muestra que las regiones rellenas de ruido tienen diferentes coeficientes individuales pero la misma energía, tal como se esperaba.
- La figura 10 es un diagrama de bloques de una realización de ejemplo de un descodificador propuesto 40. Un descodificador de la región de picos 42 está configurado para descodificar regiones de picos del espectro de la señal armónica de audio codificada transformada en frecuencia. Un descodificador de conjuntos de baja frecuencia 44 está configurado para descodificar al menos un conjunto de coeficientes de baja frecuencia. Un distribuidor de coeficientes 46 está configurado para distribuir coeficientes de cada conjunto de baja frecuencia fuera de las regiones de picos. Un descodificador de la ganancia del suelo de ruido 48 está configurado para descodificar un suelo de ruido de al menos un conjunto de coeficientes de alta frecuencia fuera de las regiones de picos. Un rellenedor de ruido 50 está configurado para rellenar cada conjunto de alta frecuencia con ruido que tiene la correspondiente ganancia del suelo de ruido. En esta realización las posiciones de los picos se reenvían al distribuidor de coeficientes 46 y al rellenedor de ruido 50 para evitar sobre escribir las regiones de picos.
- Las etapas, funciones, procedimientos y/o bloques descritos en este documento se pueden realizar mediante hardware utilizando cualquier tecnología convencional, tal como tecnología de circuitos discretos o de circuitos integrados, incluyendo tanto la circuitería electrónica de propósito general como la circuitería específica de la aplicación.
- Alternativamente, al menos alguna de las etapas, funciones, procedimientos y/o bloques descritos en este documento se pueden realizar mediante software para su ejecución por medio del equipo de tratamiento adecuado. Este equipamiento puede incluir, por ejemplo, uno o varios microprocesadores, uno o varios Procesadores Digitales de Señal (DSP), uno o varios Circuitos Integrados de Aplicación Específica (ASIC), hardware acelerado de vídeo o uno o varios dispositivos lógicos programables adecuados, tales como Conjuntos de Puertas Programables en Campo (FPGA). También es factible la combinación de tales elementos de tratamiento.
- También se debe entender que puede ser posible reutilizar las capacidades generales del tratamiento ya presentes en el codificador/descodificador. Esto puede, por ejemplo, ser hecho reprogramando el software existente o añadiendo nuevos componente de software
- La figura 11 es un diagrama de bloques de una realización de ejemplo del codificador propuesto 20. Esta realización se basa en un procesador 110, por ejemplo un microprocesador, que ejecuta el software 120 para localizar los picos, el software 130 para codificar las regiones de picos, el software 140 para codificar al menos un conjunto de baja frecuencia y el software 150 para codificar al menos una ganancia del suelo de ruido. El software está almacenado en la memoria 160. El procesador 110 se comunica con la memoria sobre un bus del sistema. La transformada en frecuencia entrante se recibe por medio de un controlador de entrada/salida (I/O) 170 que controla un bus de I/O, al cual están conectados el procesador 110 y la memoria 160. La transformada en frecuencia codificada obtenida mediante el software 150 se envía desde la memoria 160 por medio del controlador de I/O 170 sobre el bus de I/O.
- La figura 12 es un diagrama de bloques de una realización de ejemplo del descodificador propuesto 40. Esta realización se basa en un procesador 210, por ejemplo un microprocesador, que ejecuta el software 220 para descodificar las regiones de picos, el software 230 para descodificar al menos un conjunto de baja frecuencia, el software 240 para distribuir los coeficientes de LF, el software 250 para descodificar al menos una ganancia del suelo de ruido y el software 260 para el relleno de ruido. El software se almacena en la memoria 270. El procesador 210 se comunica con la memoria sobre un bus del sistema. La transformada en frecuencia codificada entrante se recibe por medio de un controlador de entrada/salida (I/O) 280 que controla un bus de I/O, al cual están conectados el procesador 210 y la memoria 280. La transformada en frecuencia reconstruida obtenida mediante el software 260 se envía desde la memoria 270 por medio del controlador de I/O 280 sobre el bus de I/O.
- La tecnología descrita anteriormente se pretende que se utilice en un codificador/descodificador de audio, el cual se puede utilizar en un dispositivo móvil (por ejemplo, teléfono móvil, ordenador portátil) o en un dispositivo estacionario, tal como un ordenador personal. En este documento el término equipo de usuario (UE) se utilizará como un nombre genérico para tales dispositivos.
- La figura 13 es un diagrama de bloques de una realización de ejemplo de un UE que incluye el codificador propuesto. Una señal de audio procedente de un micrófono 70 se reenvía a un convertidor A/D 72, cuya salida se reenvía a un codificador de audio 74. El codificador de audio 74 incluye un transformador de frecuencia 76 que transforma las muestras digitales de audio en el dominio en frecuencia. Un detector de armónicos de la señal 78 determina si la transformada representa audio con armónicos o sin armónicos. Si representa audio sin armónicos, se codifica de modo convencional (no mostrado). Si representa audio con armónicos, se reenvía a un codificador

transformador de frecuencia 20 de acuerdo con la tecnología propuesta. La señal codificada se reenvía a una unidad de radio 80 para su transmisión a un receptor.

La decisión del detector de armónicos de la señal 78 se basa en la energía del suelo de ruido \bar{E}_{nf} y la energía del pico \bar{E}_p , en los APENDICES I y II. La lógica es como sigue: IF \bar{E}_p / \bar{E}_{nf} está por encima de un umbral AND el número de picos detectados se encuentra en un intervalo predefinido, THEN la señal se clasifica como con armónicos. Si no, la señal se clasifica como con no armónicos. La clasificación y por ello el modo de codificación se señalará explícitamente al descodificador.

La figura 14 es un diagrama de bloques de una realización de ejemplo de un UE que incluye el descodificador propuesto. Una señal de audio recibida por una unidad de radio 82 se convierte a banda base, se descodifica en canales y se reenvía a un descodificador de audio 84. El descodificador de audio incluye un selector del modo de descodificación 86, que reenvía la señal a un descodificador de la transformada en frecuencia 40 de acuerdo con la tecnología propuesta si ha sido clasificada como con armónicos. Si ha sido clasificada como audio sin armónicos, se descodifica en un descodificador convencional (no mostrado). El descodificador de la transformada en frecuencia 40 reconstruye la transformada en frecuencia como se describió anteriormente. La transformada en frecuencia reconstruida se convierte en el dominio en el tiempo en un transformador inverso de frecuencia 88. Las muestras de audio resultantes se reenvían a una unidad de conversión D/A y de amplificación 90, que reenvía la señal final de audio a un altavoz 92

La figura 15 es un diagrama de flujo de una realización de ejemplo de una parte del método de codificación propuesto. En esta realización la etapa de codificar la región de picos S2 en la figura 5 se ha dividido en subetapas S2-A a S2-E. La etapa S2-A codifica la posición del espectro y el signo de un pico. La etapa S2-B cuantifica la ganancia del pico. La etapa S2-C codifica la ganancia cuantificada del pico. La etapa S2-D escala los contenidos predeterminados de frecuencia que rodean el pico por medio de la inversa de la ganancia cuantificada del pico. La etapa S2-E codifica la forma de los contenidos escalados de frecuencia.

La figura 16 es un diagrama de bloques de una realización de ejemplo de un codificador de la región de picos en el codificador propuesto. En esta realización el codificador de la región de picos 24 incluye los elementos 24-A a 24-D. El codificador de posición y signo 24-A está configurado para codificar la posición del espectro y el signo de un pico. El codificador de la ganancia de pico 24-B está configurado para cuantificar la ganancia del pico y para codificar la ganancia cuantificada del pico. La unidad de escalado 24-C está configurada para escalar los contenidos predeterminados de frecuencia que rodean el pico por medio de la inversa de la ganancia cuantificada del pico. El codificador de la forma 24-D está configurado para codificar la forma de los contenidos escalados de frecuencia.

La figura 17 es un diagrama de flujo de una realización de ejemplo de una parte del método propuesto de descodificación. En esta realización la etapa descodificadora de la región de picos S11 en la figura 8 se ha dividido en subetapas S11-A a S11-D. La etapa S11-A descodifica la posición del espectro y el signo de un pico. La etapa S11-B descodifica la ganancia del pico. La etapa S11-C descodifica una forma de los contenidos predeterminados de frecuencia que rodean al pico. La etapa S11-D escala la forma descodificada por medio de la ganancia descodificada del pico.

La figura 18 es un diagrama de bloques de una realización de ejemplo de un descodificador de la región de picos en el descodificador propuesto. En esta realización el descodificador de la región de picos 42 incluye los elementos 42-A a 42-D. Un descodificador de posición y signo 42-A está configurado para descodificar la posición del espectro y el signo de un pico. Un descodificador de la ganancia del pico 42-B está configurado para descodificar la ganancia del pico. Un descodificador de la forma 42-C está configurado para descodificar una forma de los contenidos predeterminados de frecuencia que rodean al pico. Una unidad de escalado 42-D está configurada para escalar la forma descodificada por medio de la ganancia descodificada del pico.

A continuación se dan los detalles específicos de realización para un modo de 24 kbps.

- El codec opera sobre tramas del 20 ms, lo cual a una velocidad de bits de 24 kbps da 480 bits por trama.
- La señal procesada de audio se muestrea a 32 kHz, y tiene un ancho de banda de audio de 16 kHz.
- La frecuencia de transición se fija en 5,6 kHz (todo los componentes de la frecuencia por encima de 5,6 kHz se amplían en el ancho de banda).
- Bits reservados para la señalización y la ampliación del ancho de banda de las frecuencias por encima de la frecuencia de transición: ≈30-40.
- Bits para codificar dos ganancias del suelo de ruido: 10.
- El número de regiones codificadas del espectro de picos es 7-17. El número de bits utilizados por cada región de picos es ≈20-22, que da un número total de ≈140-340 para codificar todas las posiciones de picos, ganancias, signos y formas.

- Bits para codificar las bandas de baja frecuencia: ≈100-300.
- Bandas codificadas de baja frecuencia: 1-4 (cada banda contiene ocho contenidos MDCT). Dado que cada contenido MDCT corresponde a 25 Hz, la región codificada de baja frecuencia corresponde a 200-800 Hz.
- Las ganancias utilizadas para la ampliación del ancho de banda y para las ganancias de picos se tratan con la codificación Huffman de modo que el número de bits utilizados para esas podría variar entre tramas incluso para un número constante de picos.
- La posición del pico y la codificación del signo utilizan una optimización que es más eficiente según aumenta el número de picos. Para 7 picos, la posición y el signo requieren alrededor de 6,9 bits por pico y para 17 picos el número es de alrededor de 5,7 bits por pico.
- Esta variabilidad de cuantos bits utilizar en diferentes etapas de la codificación no es un problema ya que la codificación de la banda de baja frecuencia se hace la última y sólo se usa siempre que queden bits. Sin embargo el sistema está diseñado para que queden siempre bastantes bits para codificar una banda de baja frecuencia.

La tabla siguiente presenta los resultados procedentes de una prueba de escucha realizada de acuerdo con el procedimiento descrito en ITU-R BS.1534-1 MUSHRA (Estímulos Múltiples con Referencia y Anclaje Ocultos). La escala en una prueba MUSHRA va de 0 a 100, en la que los valores bajos corresponden a baja calidad percibida y los valores altos corresponden a la calidad alta. Ambos codecs operan a 24 kbps. Los resultados de la prueba se promedian sobre 24 elementos musicales y los votos de ocho auditores.

Sistema Bajo Prueba	Puntuacion de MUSHRA
Señal de anclaje de paso bajo (banda ancha 7 kHz)	48,49
Esquema de codificación convencional	49,94
Esquema de codificación armónica propuesto	55,87
Señal de referencia (banda ancha 16 kHz)	100,00

Los expertos en la técnica comprenderán que se pueden hacer diversas modificaciones y cambios a la tecnología propuesta sin apartarse del ámbito de la misma, el cual se define por medio de las reivindicaciones adjuntas.

Apéndice I

El algoritmo de estimación del suelo de ruido opera sobre los valores absolutos de los coeficientes de la transformada $|Y(k)|$. Las energías instantáneas del suelo de ruido $E_{nf}(k)$ se estiman de acuerdo con la recursividad:

$$E_{nf}(k) = \alpha E_{nf}(k-1) + (1 - \alpha)|Y(k)| \quad (3)$$

en la que

$$\alpha = \begin{cases} 0.9578 & \text{if } |Y(k)| > E_{nf}(k-1) \\ 0.6472 & \text{if } |Y(k)| \leq E_{nf}(k-1) \end{cases} \quad (4)$$

El formato particular del factor de ponderación α minimiza el efecto de los coeficientes de alta energía de la transformada y enfatiza la contribución de los coeficientes de baja energía. Finalmente, el nivel del suelo de ruido \bar{E}_{nf} se estima por promediado simple de las energías instantáneas $E_{nf}(k)$.

Apéndice II

El algoritmo de pico escogido requiere el conocimiento del nivel del suelo de ruido y del nivel promediado de los picos del espectro. El algoritmo de estimación de la energía del pico es similar al algoritmo de estimación del suelo de ruido, pero en lugar de la baja energía, él hace el seguimiento de las energías de alto espectro:

$$E_p(k) = \beta E_p(k-1) + (1 - \beta)|Y(k)| \quad (5)$$

en donde

$$\beta = \begin{cases} 0.4223 & \text{if } |Y(k)| > E_p(k-1) \\ 0.8029 & \text{if } |Y(k)| \leq E_p(k-1) \end{cases} \quad (6)$$

En este caso el factor de ponderación β minimiza el efecto de los coeficientes de baja energía de la transformada y enfatiza la contribución de los coeficientes de alta energía. La energía total del pico \bar{E}_p se estima promediando simplemente las energías instantáneas.

Cuando se calculan los niveles del pico y del suelo de ruido, se forma un nivel de umbral θ como:

$$\theta = \left(\frac{\bar{E}_p}{\bar{E}_{nf}} \right)^\gamma \bar{E}_{nf} \quad (7)$$

5 con $\gamma = 0,88579$. Los coeficientes de la transformada se comparan con el umbral, y los que tienen una amplitud por encima de él, forman un vector de los candidatos de picos. Dado que las fuentes naturales no producen normalmente picos muy cercanos, por ejemplo, 80 hz, el vector con los candidatos de picos se perfecciona adicionalmente. Los elementos de los vectores se extraen en orden decreciente, y la vecindad de cada elemento se fija a valor cero. De esta manera sólo los elementos mayores permanecen en cierta región espectral, y el conjunto de esos elementos forma los picos del espectro para la trama actual.

Abreviaturas

- ASIC Circuitos Integrados de Aplicación Específica
- BWE Aumento del Ancho de Banda
- 15 DSP Procesadores Digitales de Señal
- FPGA Conjuntos de Puertas Programables en Campo
- HF Alta Frecuencia
- LF Baja Frecuencia
- MDCT Transformada Discreta del Coseno Modificada
- 20 RMS Valor Eficaz
- VQ Cuantificador del Vector

REIVINDICACIONES

1. Un aparato para codificar mediante la Transformada Discreta del Coseno Modificada, MDCT, los coeficientes $(Y(k))$ de una señal armónica de audio, comprendiendo el aparato:

5 medios (22) para localizar los picos del espectro que tienen magnitudes que exceden un umbral predeterminado, en el que los picos del espectro se localizan mediante la comparación de los coeficientes con dicho umbral para formar un vector de candidatos de picos, y para extraer los elementos procedentes del vector de los candidatos de picos en orden decreciente, en el que dicho umbral se calcula como

$$\theta = \left(\frac{\bar{E}_p}{\bar{E}_{nf}} \right)^\gamma \bar{E}_{nf},$$

10 en donde \bar{E}_p es una energía promedio del pico, \bar{E}_{nf} es una energía promedio del suelo de ruido y γ tiene un valor fijo predeterminado, y en donde se calcula una energía del pico como $E_p(k) = \beta E_p(k) + (1-\beta)|Y(k)|$ y se calcula una energía del suelo de ruido como $E_{nf}(k) = \alpha E_{nf}(k) + (1-\alpha)|Y(k)|$, en donde la contribución de los coeficientes de alta energía se enfatiza en el cálculo de la energía del pico y la contribución de los coeficientes de baja energía se enfatiza en el cálculo de la energía del suelo de ruido;

15 medios (24) para codificar las regiones de picos que incluyen y rodean los picos localizados, en donde los picos del espectro se cuantifican junto con los contenidos MDCT vecinos;

20 medios (26) para codificar utilizando un cierto número de bits reservados, una primer conjunto de coeficientes de baja frecuencia, fuera de la regiones de picos y por debajo de una frecuencia de cruce que depende del número de bits utilizados para codificar las regiones de picos, y para codificar uno o más conjuntos de coeficientes de baja frecuencia adicionales fuera de la regiones de picos si no hay bits reservados disponibles después de codificar las regiones de picos; y

medios (28) para codificar, utilizando un cierto número de bits reservados, una ganancia del suelo de ruido de al menos un conjunto de coeficientes de alta frecuencia aún no codificados fuera de las regiones de picos.

2. El aparato de acuerdo con la reivindicación 1, en el que un factor de ponderación α se define como

$$\alpha = \begin{cases} 0.9578 & \text{if } |Y(k)| > E_{nf}(k-1) \\ 0.6472 & \text{if } |Y(k)| \leq E_{nf}(k-1) \end{cases},$$

25 y un factor de promediado se define como

$$\beta = \begin{cases} 0.4223 & \text{if } |Y(k)| > E_p(k-1) \\ 0.8029 & \text{if } |Y(k)| \leq E_p(k-1) \end{cases}.$$

3. El aparato de acuerdo con la reivindicación 1 o 2, en el que los medios para codificar las regiones de picos comprende:

medios (24-A) para codificar la posición del espectro y el signo de un pico;

30 medios (24-B) para cuantificar la ganancia del pico;

medios (24-C) para escalar los contenidos predeterminados de la frecuencia que rodean el pico por medio de la inversa de la ganancia cuantificada del pico;

medios (24-D) para codificar la forma de los contenidos escalados de frecuencia.

35 4. El aparato de acuerdo con cualquiera de las reivindicaciones 1 a 3, en el que la región de picos comprende el pico y cuatro contenidos MDCT que rodean dicho pico.

5. El aparato de acuerdo con cualquiera de las reivindicaciones precedentes, en el que los medios (26) para codificar el conjunto de coeficientes de baja frecuencia comprende medios para agrupar los coeficientes MDCT restantes no cuantificados en 24 bandas dimensionales.

6. El aparato de acuerdo con cualquiera de las reivindicaciones precedentes, en el que la codificación de un conjunto de baja frecuencia se basa en un esquema de codificación de la forma de la ganancia, estando basado dicho esquema de codificación de la forma de la ganancia en la cuantificación escalar de la ganancia y en la codificación factorial de la forma del impulso.

5 7. Un equipo de usuario que comprende el aparato de acuerdo con la reivindicación 1.

8. El equipo de usuario de acuerdo con la reivindicación 7, en el que el equipo de usuario es un dispositivo móvil.

9. Un código de programa de ordenador, cuando se ejecuta por un procesador, el código de programa de ordenador provoca un aparato para:

10 localizar los picos del espectro que tienen magnitudes que exceden un umbral predeterminado, en el que los picos del espectro se localizan comparando los coeficientes con dicho umbral para formar un vector de candidatos de picos, y extraen los elementos del vector de los candidatos de picos en orden decreciente, en el que dicho umbral se calcula como

$$\theta = \left(\frac{\bar{E}_p}{\bar{E}_{nf}} \right)^\gamma \bar{E}_{nf},$$

15 en donde \bar{E}_p es una energía del pico promedio, \bar{E}_{nf} es una energía del suelo de ruido promedio y γ tiene un valor fijo predeterminado, y en el que una energía del pico se calcula como $E_p(k) = \beta E_p(k) + (1-\beta)|Y(k)|$ y una energía de suelo de ruido se calcula como $E_{nf}(k) = \alpha E_{nf}(k) + (1-\alpha)|Y(k)|$, en donde la contribución de los coeficientes de alta energía se enfatiza en el cálculo de la energía del pico y la contribución de los coeficientes de baja energía se enfatiza en el cálculo de la energía del suelo de ruido;

20 codificar las regiones de picos que incluyen y rodean los picos localizados, en el que los picos del espectro se cuantifican junto con los contenidos MDCT vecinos;

25 codificar, utilizando un número de bits reservados, un primer conjunto de coeficientes de baja frecuencia, LF, fuera de las regiones de picos y por debajo de una frecuencia de cruce que depende del número de bits utilizados para codificar las regiones de picos, en donde la codificación comprende codificar adicionalmente uno o más conjuntos de coeficientes de baja frecuencia fuera de las regiones de picos si existen bits no reservados disponibles después decodificar las regiones de picos; y

codificar, utilizando un número reservado de bits, una ganancia del suelo de ruido de al menos un conjunto de coeficientes de alta frecuencia no codificados aún fuera de las regiones de picos.

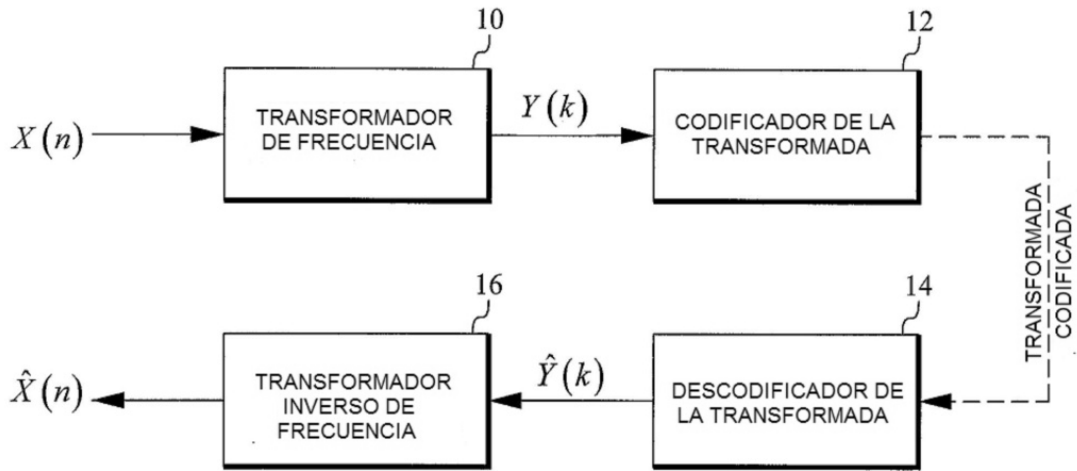


FIG. 1

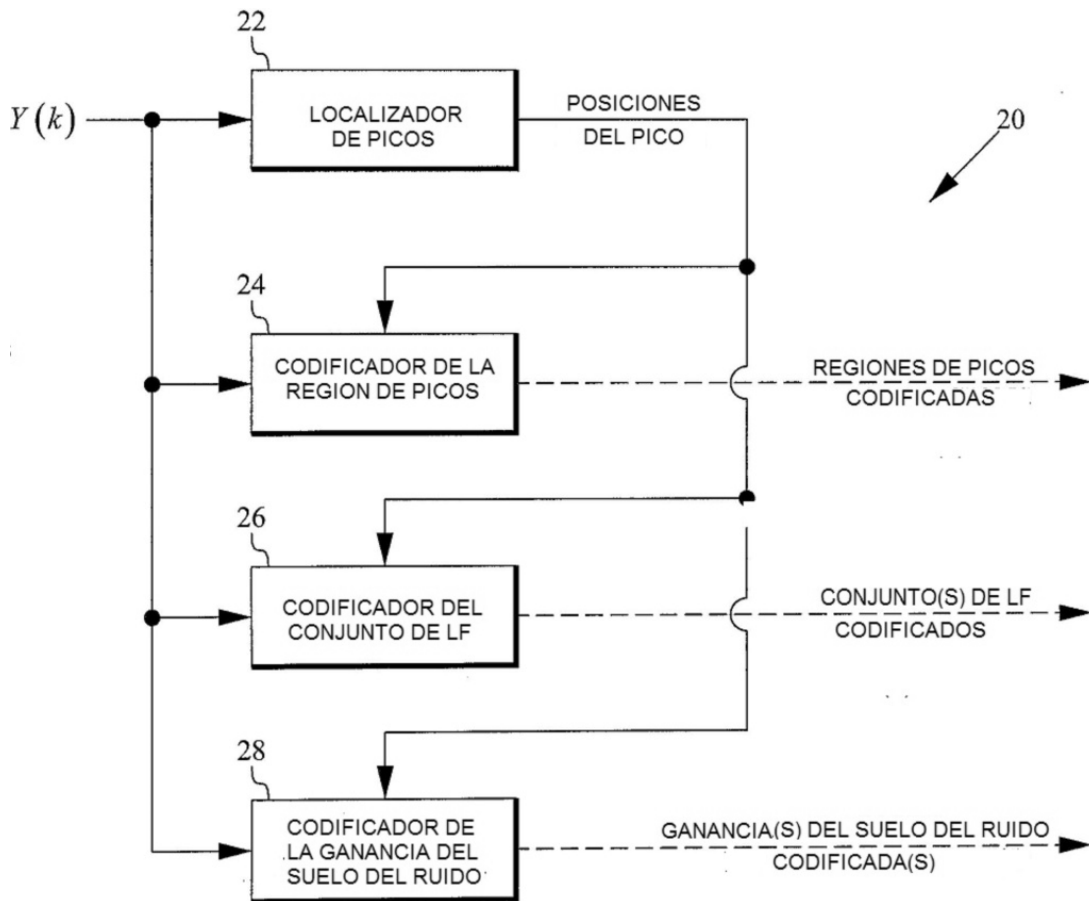


FIG. 7

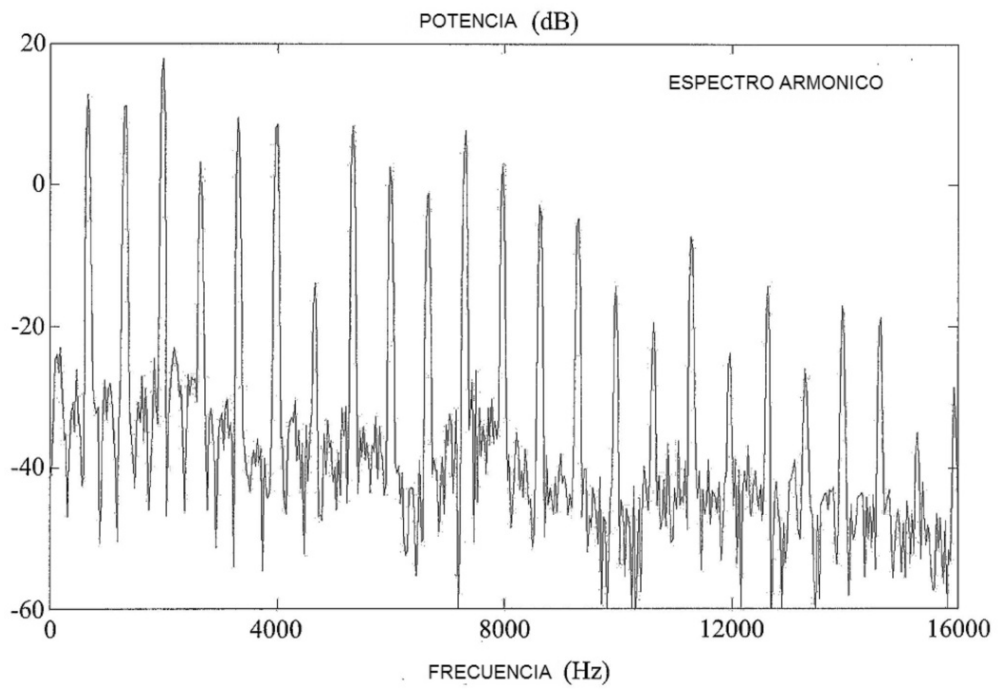


Fig. 2

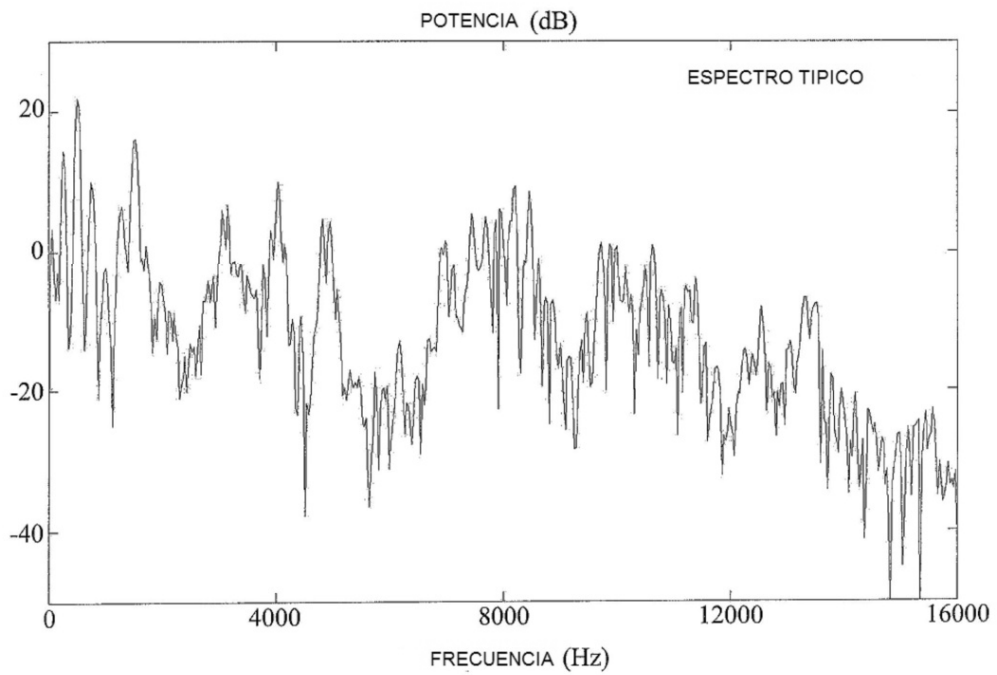


Fig. 3

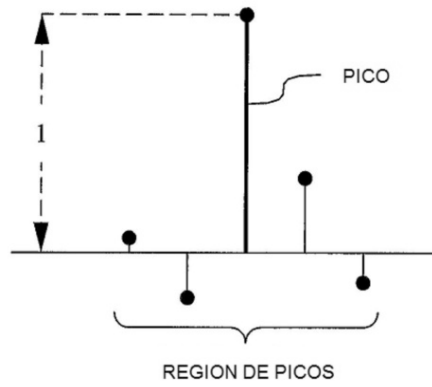


FIG. 4

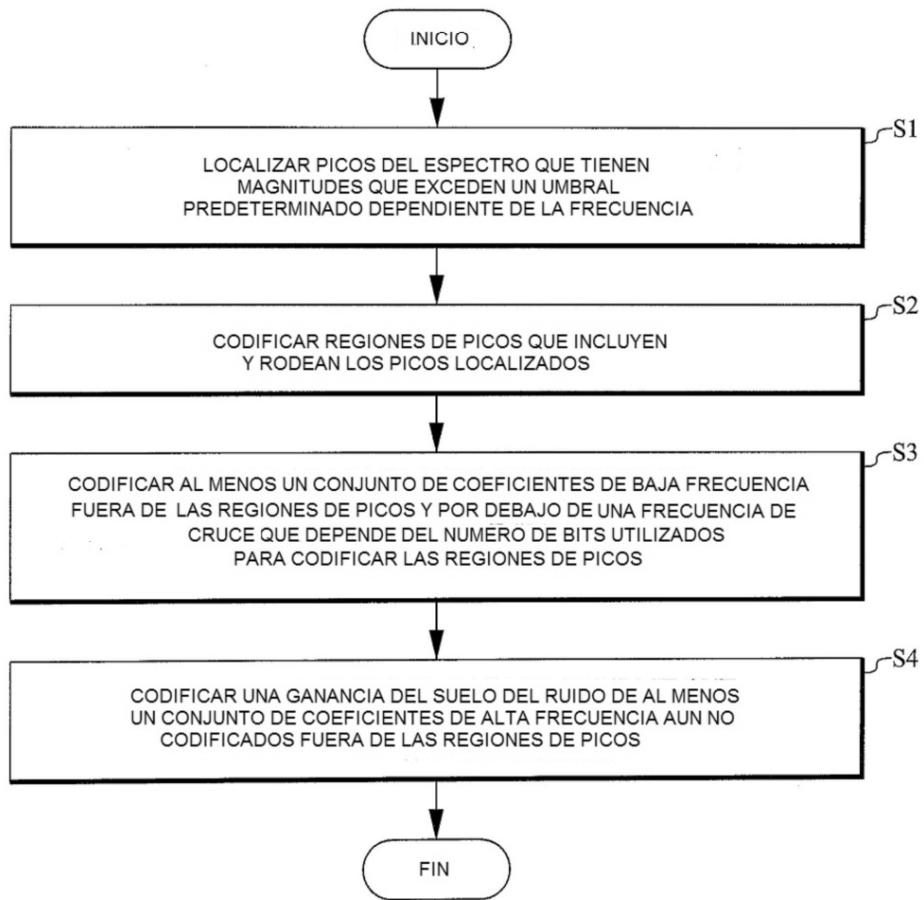


FIG. 5

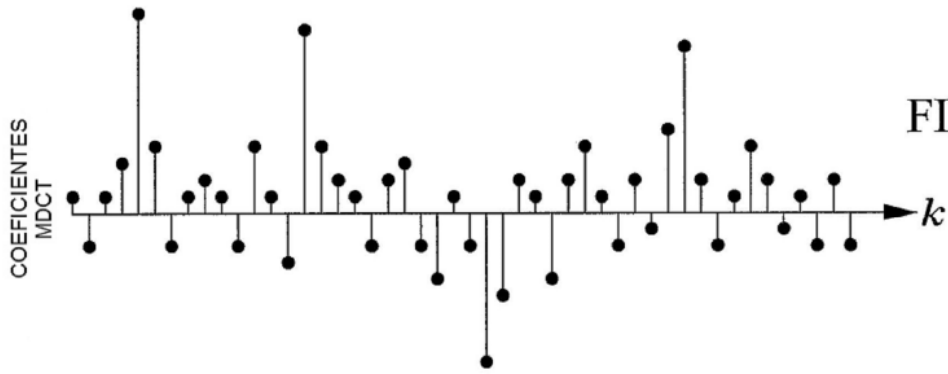


FIG. 6A

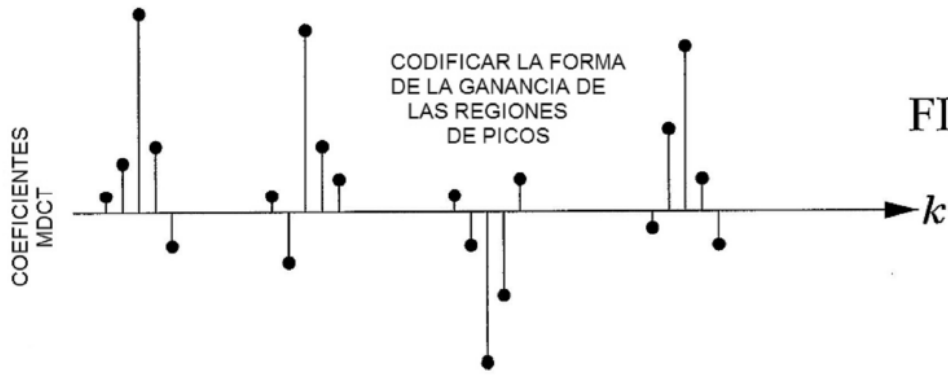


FIG. 6B

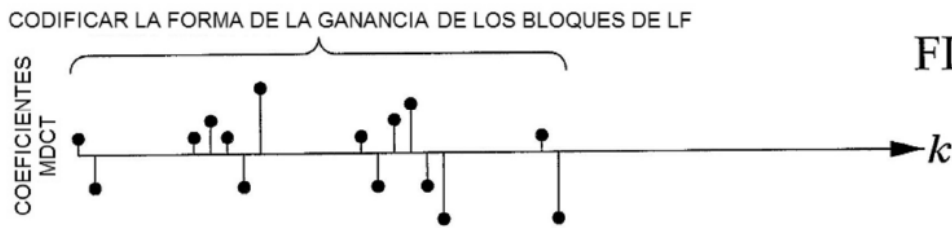


FIG. 6C

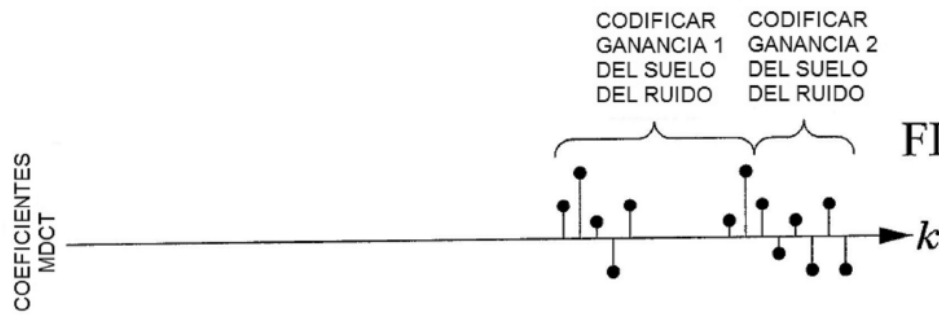


FIG. 6D

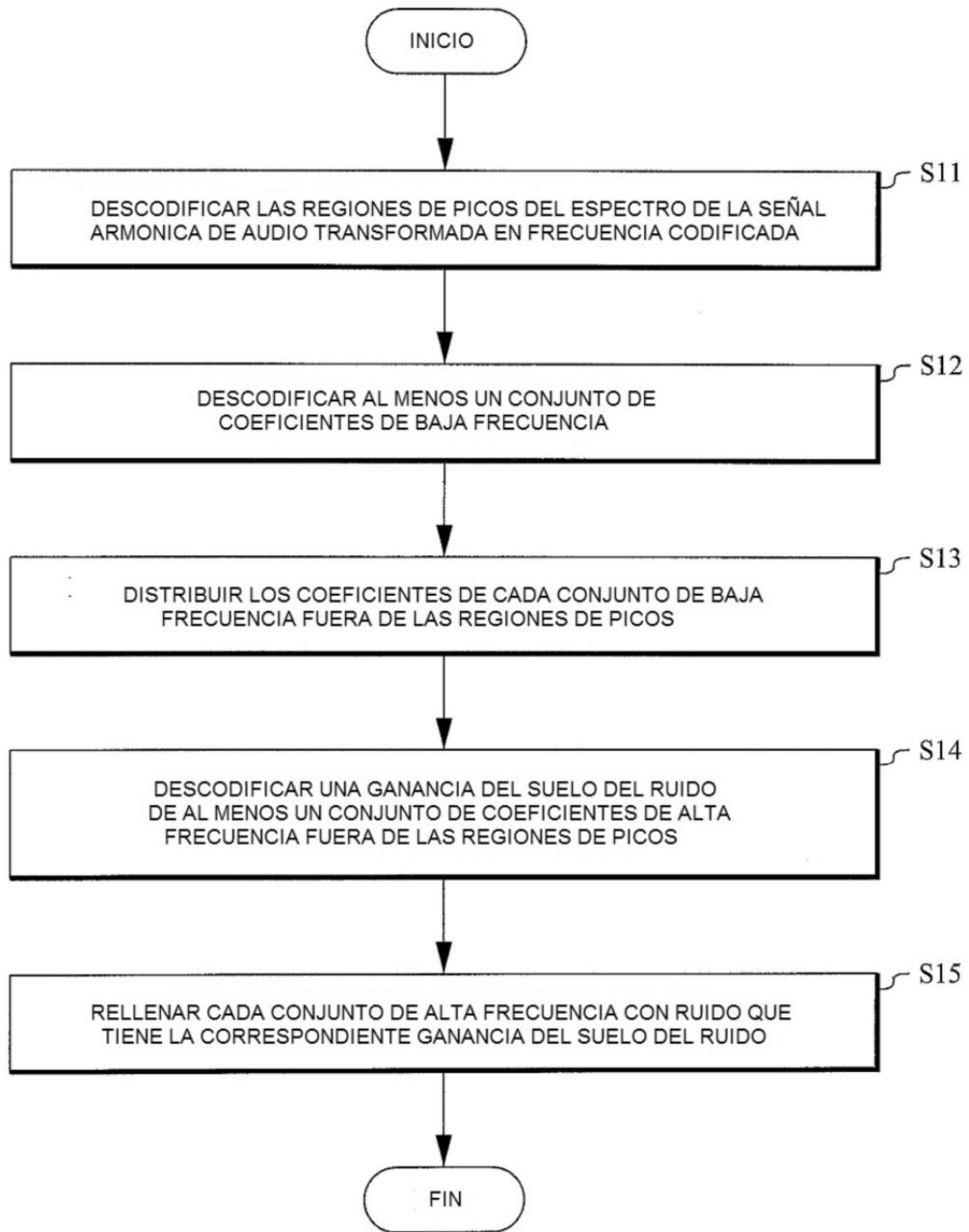


FIG. 8



FIG. 9A

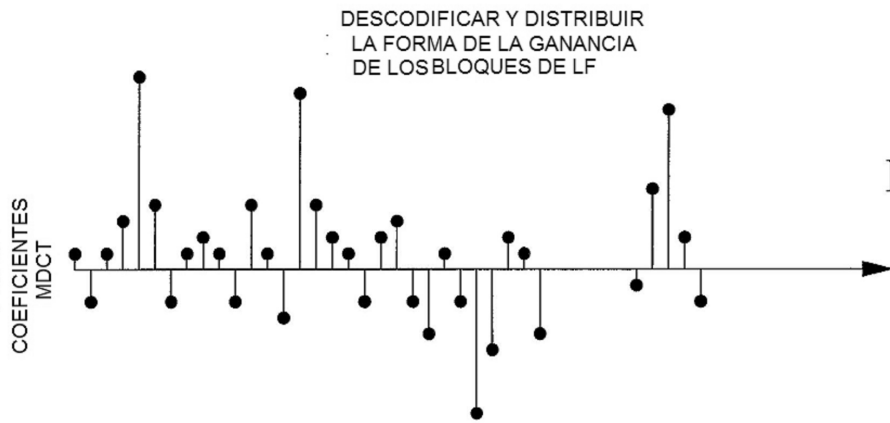


FIG. 9B

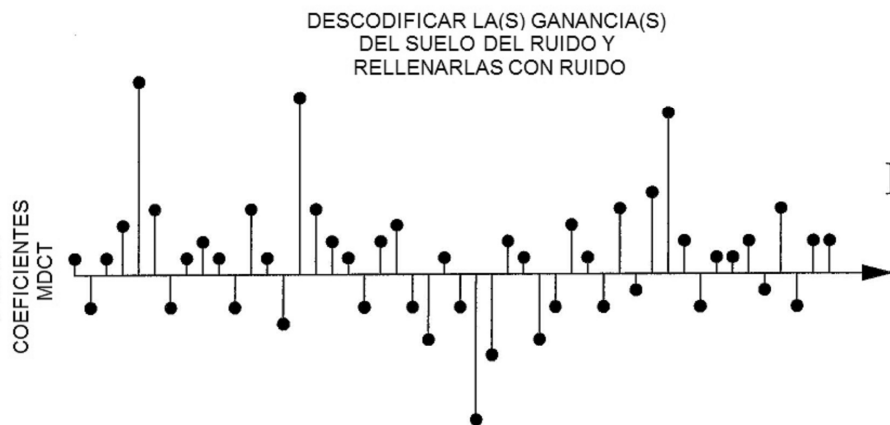


FIG. 9C

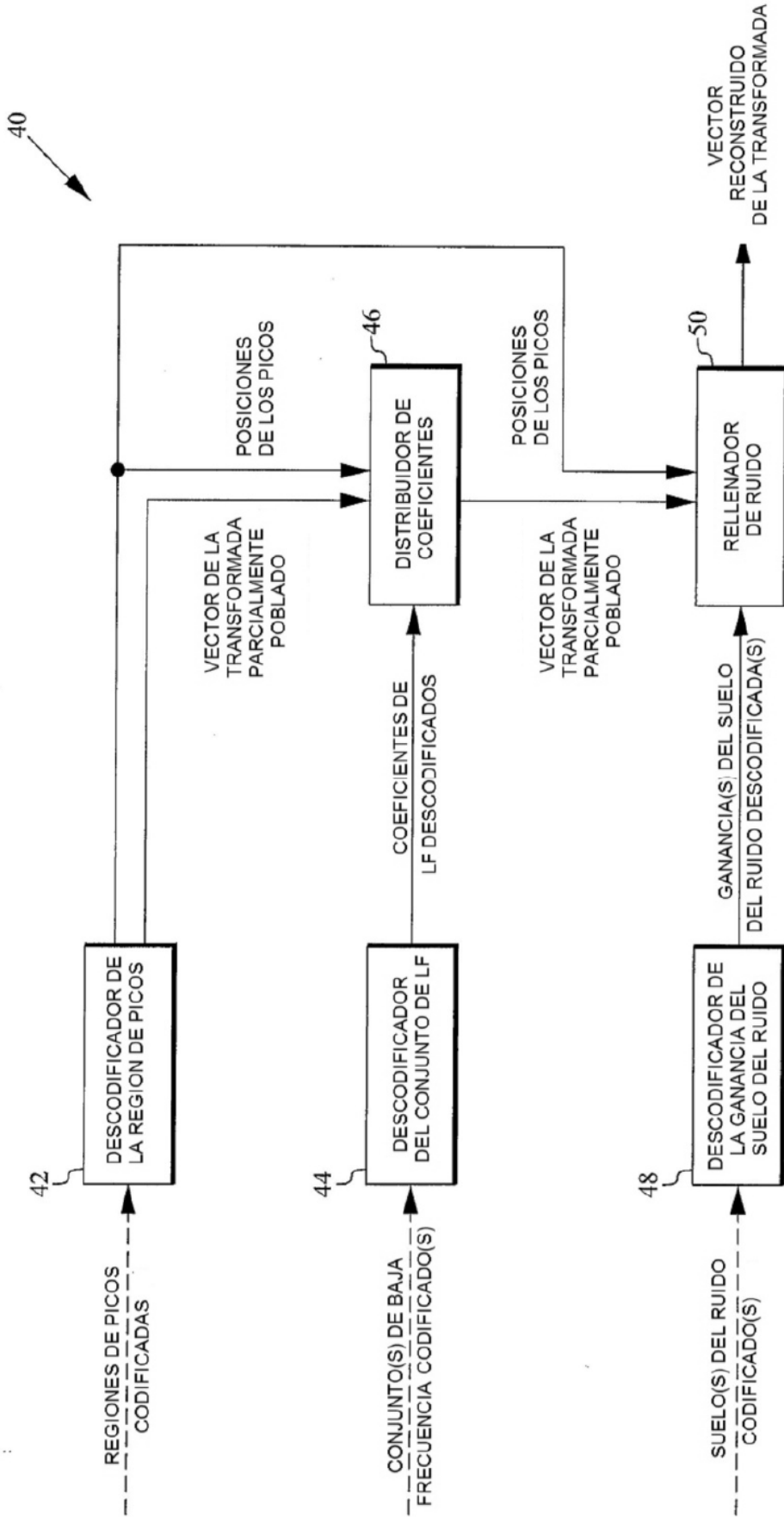


FIG. 10

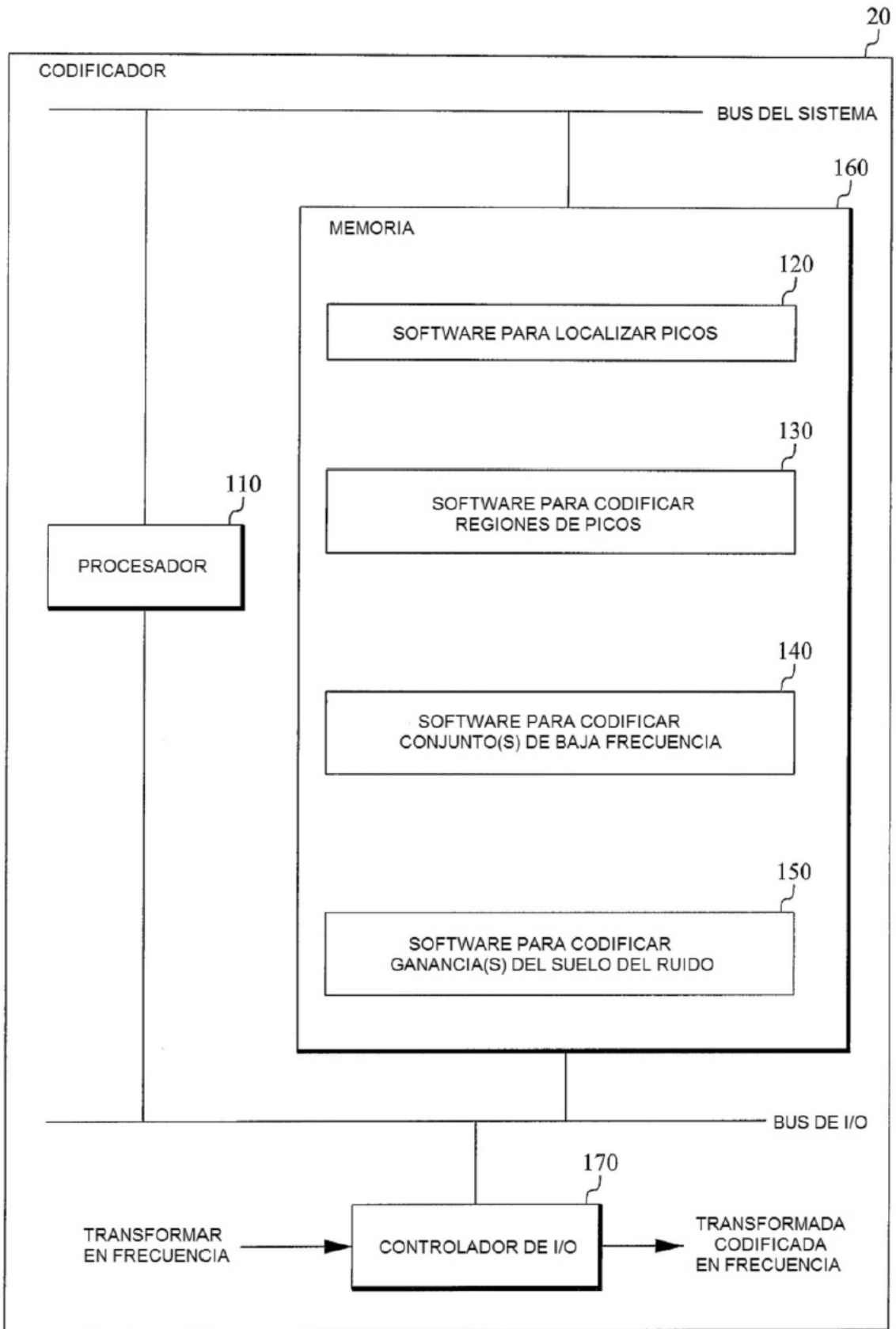


FIG. 11

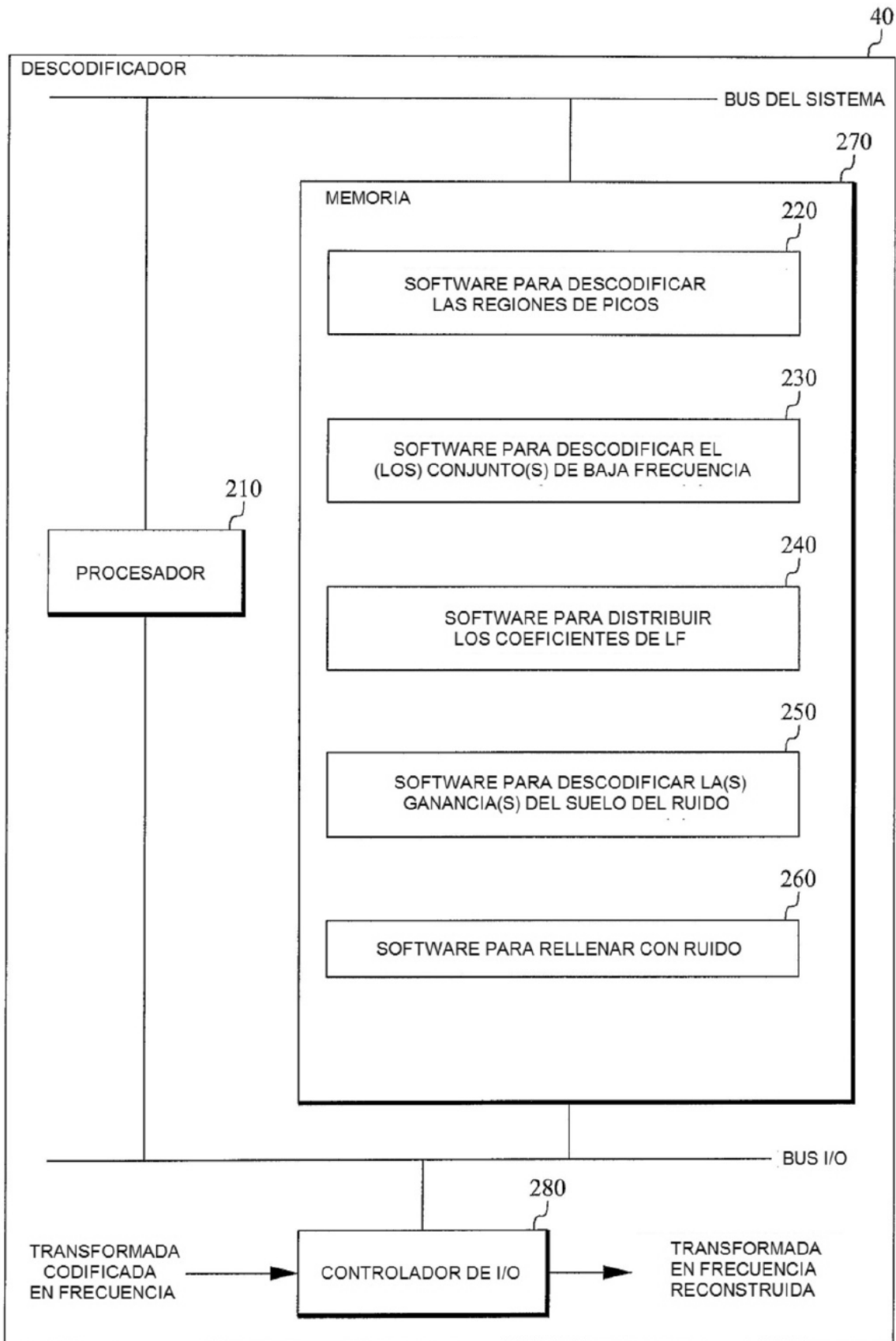


FIG. 12

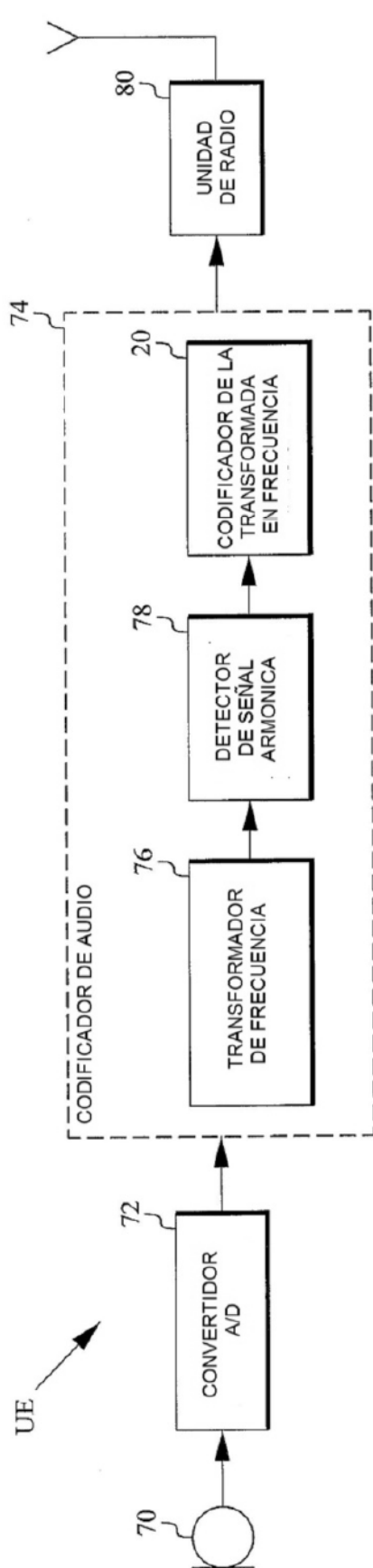


FIG. 13

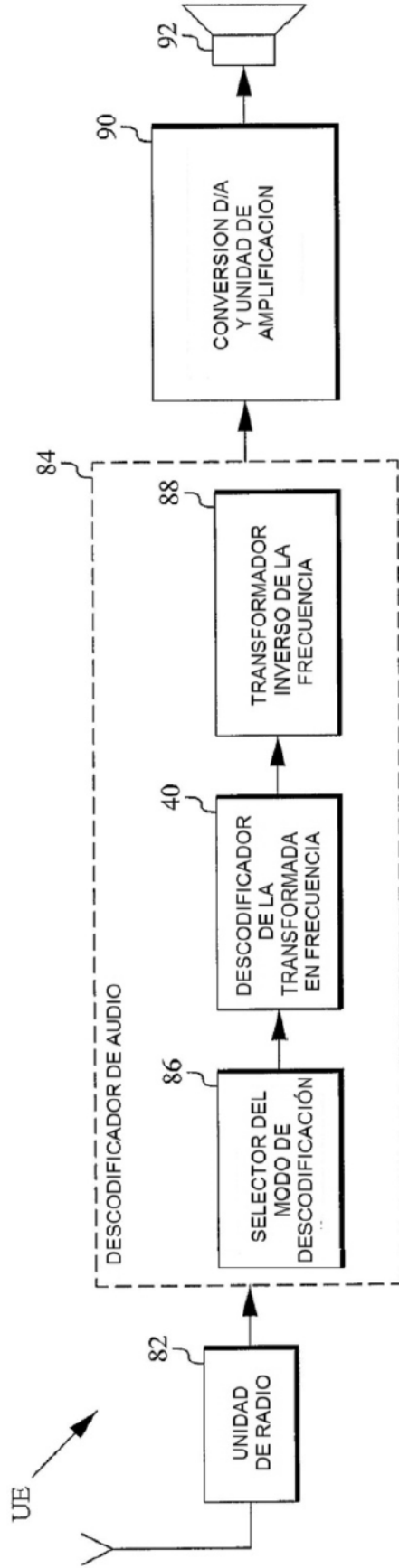


FIG. 14

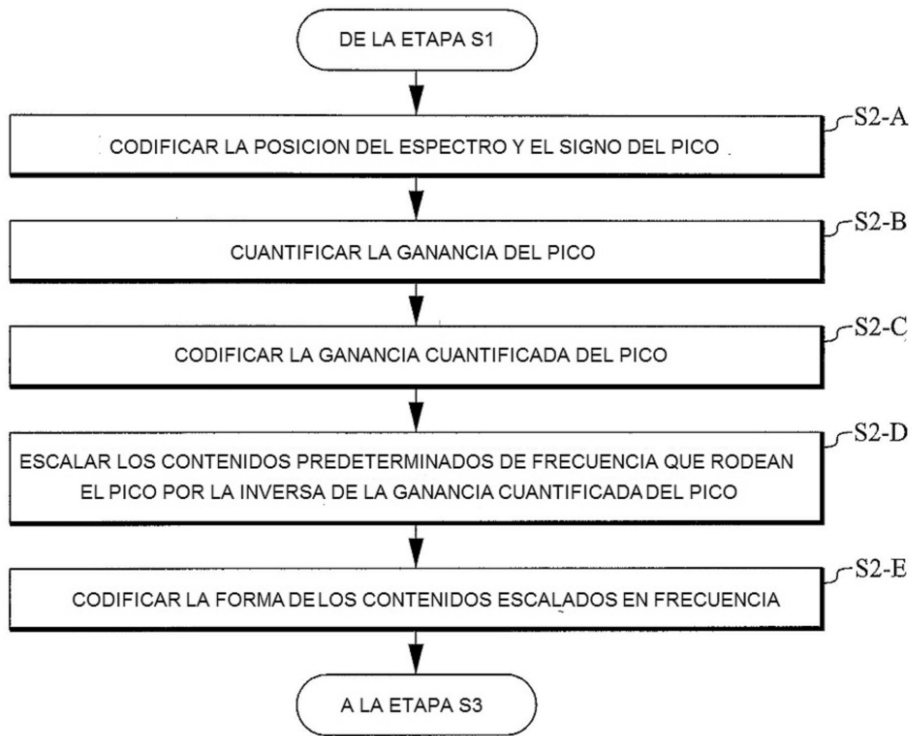


FIG. 15

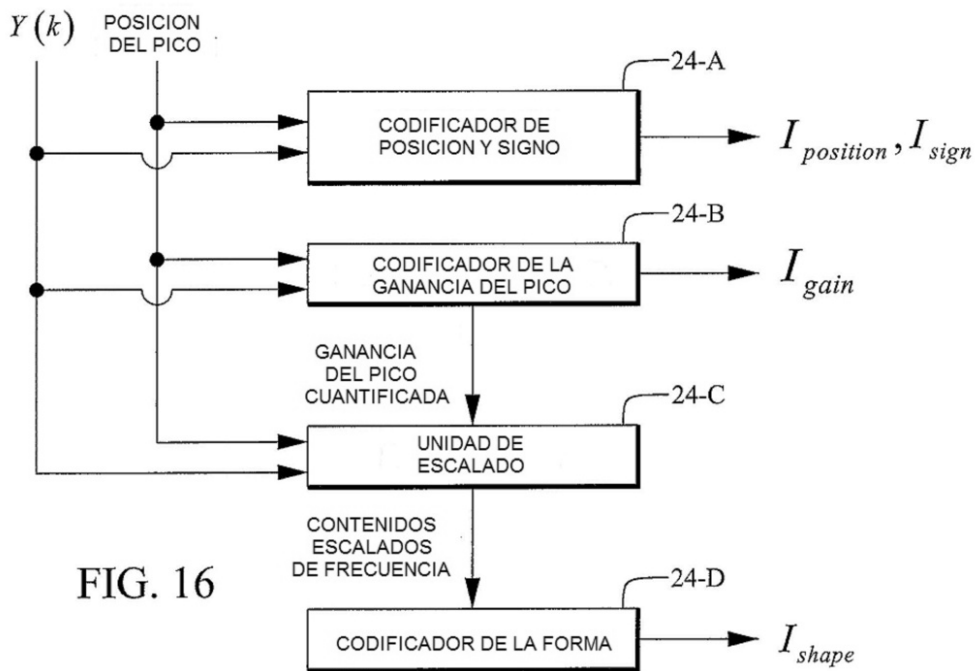


FIG. 16



FIG. 17

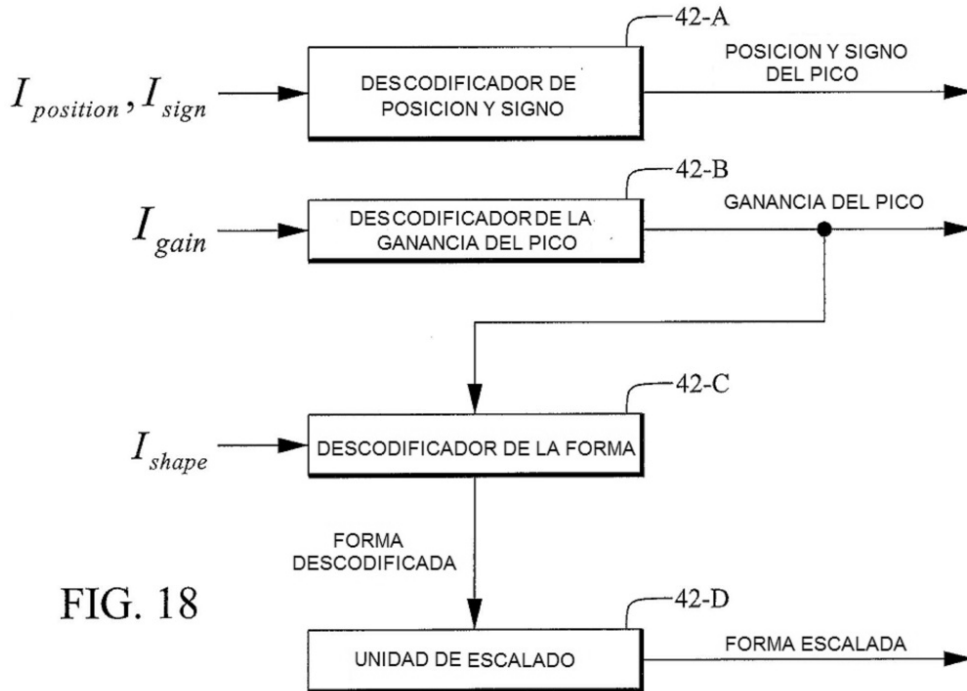


FIG. 18