

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 657 802**

51 Int. Cl.:

G10L 19/06 (2013.01)

G10L 19/032 (2013.01)

G10L 21/038 (2013.01)

G10L 19/02 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **15.05.2012 E 16156708 (6)**

97 Fecha y número de publicación de la concesión europea: **25.10.2017 EP 3040988**

54 Título: **Decodificación de audio basada en una representación eficiente de coeficientes autoregresivos**

30 Prioridad:

02.11.2011 US 201161554647 P

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

06.03.2018

73 Titular/es:

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

72 Inventor/es:

**GRANCHAROV, VOLODYA y
SVERRISSON, SIGURDUR**

74 Agente/Representante:

LINAGE GONZÁLEZ, Rafael

ES 2 657 802 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

Decodificación de audio basada en una representación eficiente de coeficientes autoregresivos

5 **Campo técnico**

La tecnología propuesta se refiere a la decodificación de audio basada en una representación eficiente de coeficientes autoregresivos (AR).

10 **Antecedentes**

El análisis AR se utiliza comúnmente tanto en la codificación de audio de dominio de la transformada [2] como del tiempo [1]. Las diferentes aplicaciones utilizan vectores AR de diferente longitud (el orden de modelo depende principalmente del ancho de banda de la señal codificada; de 10 coeficientes para señales con un ancho de banda de 4 kHz, a 24 coeficientes para señales con un ancho de banda de 16 kHz). Estos coeficientes AR se cuantifican con cuantificación de vector de múltiples etapas (VQ) dividida, lo que garantiza una reconstrucción casi transparente. Sin embargo, los esquemas de cuantificación convencionales no están diseñados para el caso en que los coeficientes AR modelan frecuencias de audio altas (por ejemplo, por encima de 6 kHz) y funcionan con presupuestos de bits muy limitados (que no permiten la codificación transparente de los coeficientes). Esto introduce grandes errores perceptuales en la señal reconstruida cuando estos esquemas de cuantificación convencionales se utilizan en intervalos de frecuencia no óptimos y tasas de bits no óptimas.

El documento EP 1818913 A1 divulga un aparato y método de codificación de banda ancha que codifica los LSP de banda ancha utilizando LSP de banda estrecha cuantificados de una señal de discurso, y un dispositivo de predicción de LSP de banda ancha y otros capaces de predecir un LSP de banda ancha a partir de un LSP de banda estrecha con una eficiencia de cuantificación alta y una precisión alta mientras se suprime el tamaño de una tabla de conversión que correlaciona el LSP de banda estrecha con el LSP de banda ancha.

30 **Sumario**

Un objeto de la tecnología propuesta es un esquema de cuantificación más eficiente para los coeficientes autoregresivos.

Este objeto se logra de acuerdo con las reivindicaciones adjuntas.

La tecnología propuesta proporciona un esquema de baja tasa de bits para la compresión o codificación de coeficientes autoregresivos. Además de las mejoras perceptuales, la tecnología propuesta también tiene la ventaja de reducir la complejidad computacional en comparación con los métodos de cuantificación de espectro completo.

40 **Breve descripción de los dibujos**

La tecnología propuesta, junto con otros objetos y ventajas de la misma, puede comprenderse mejor haciendo referencia a la siguiente descripción tomada junto con los dibujos que la acompañan, en los que:

45 la figura 1 es un diagrama de flujo del método de codificación de acuerdo con la tecnología propuesta;

la figura 2 ilustra ejemplo del método de lado del codificador de la tecnología propuesta;

50 la figura 3 ilustra el volteo de elementos LSF de baja frecuencia cuantificados (representados por puntos negros) a alta frecuencia al reflejarlos en el espacio previamente ocupado por la mitad superior del vector LSF;

la figura 4 ilustra el efecto del alisado de parrilla en un espectro de señal;

55 la figura 5 es un diagrama de bloques de un ejemplo del codificador de acuerdo con la tecnología propuesta;

la figura 6 es un diagrama de bloques de un ejemplo del codificador de acuerdo con la tecnología propuesta;

la figura 7 es un diagrama de flujo del método de decodificación de acuerdo con la tecnología propuesta;

60 la figura 8 ilustra una realización del método del lado del decodificador de la tecnología propuesta;

la figura 9 es un diagrama de bloques de una realización del decodificador de acuerdo con la tecnología propuesta;

65 la figura 10 es un diagrama de bloques de una realización del decodificador de acuerdo con la tecnología propuesta;

la figura 11 es un diagrama de bloques de un ejemplo del codificador de acuerdo con la tecnología propuesta;

la figura 12 es un diagrama de bloques de una realización del decodificador de acuerdo con la tecnología propuesta;

la figura 13 ilustra un ejemplo de un equipo de usuario que incluye un codificador de acuerdo con la tecnología propuesta; y

la figura 14 ilustra una realización de un equipo de usuario que incluye un decodificador de acuerdo con la tecnología propuesta.

Descripción detallada

La tecnología propuesta requiere como entrada un vector a de coeficientes AR (otro nombre comúnmente utilizado es coeficientes de predicción lineal (LP)). Estos se obtienen generalmente calculando primero las autocorrelaciones $r(j)$ de los segmentos de audio con ventana $s(n)$, $n = 1, \dots, N$, es decir:

$$r(j) = \sum s(n)s(n-j), \quad j = 0, \dots, M \quad (1)$$

donde M es una orden de modelo predefinida. Después, los coeficientes AR a se obtienen de la secuencia de autocorrelación $r(j)$ a través del algoritmo Levinson-Durbin [3].

En un sistema de comunicación de audio, los coeficientes AR deben transmitirse de manera eficiente desde el codificador a la parte decodificadora del sistema. En la tecnología propuesta, esto se logra cuantificando solo ciertos coeficientes, y representando los coeficientes restantes con solo un pequeño número de bits.

Codificador

La figura 1 es un diagrama de flujo del método de codificación de acuerdo con la tecnología propuesta. El paso S1 codifica una parte de baja frecuencia de la representación espectral paramétrica mediante elementos de cuantificación de la representación espectral paramétrica que corresponde a una parte de baja frecuencia de la señal de audio. El paso S2 codifica una parte de alta frecuencia de la representación espectral paramétrica mediante un promedio ponderado basado en los elementos cuantificados volteados alrededor de una frecuencia de reflexión cuantificada, que separa la parte de baja frecuencia de la parte de alta frecuencia y una parrilla de frecuencia determinada a partir de un libro de códigos de parrilla de frecuencia en un procedimiento de búsqueda de bucle cerrado.

La figura 2 ilustra los pasos realizados en el lado del codificador de un ejemplo de la tecnología propuesta. Primero, los coeficientes AR se convierten a una representación de frecuencias espectrales de línea (LSF) en el paso S3, por ejemplo, por el algoritmo descrito en [4]. Luego, el vector LSF se divide en dos partes, indicadas como partes de baja (L) y de alta frecuencia (H) en el paso S4. Por ejemplo, en un vector LSF dimensional 10, los primeros 5 coeficientes se pueden asignar al subvector L^L y los coeficientes restantes al subvector H^H .

Aunque la tecnología propuesta se describirá con referencia a una representación LSF, los conceptos generales también se pueden aplicar a una implementación alternativa en la que el vector AR se convierte a otra representación espectral paramétrica, como el par espectral de línea (LSP) o pares espectrales de imitación (ISP) en lugar de LSF.

Solo el subvector LSF de baja frecuencia L^L se cuantifica en el paso S5, y sus índices de cuantificación l_{iL} se transmiten al decodificador. Las LSF de alta frecuencia del subvector H^H no se cuantifican, sino que solo se utilizan en la cuantificación de una frecuencia de reflexión f_m (to \hat{f}_m), y la búsqueda de bucle cerrado para una parrilla de frecuencia óptima g^{opt} de un conjunto de parrillas de frecuencia g^i formando un libro de códigos de la parrilla de frecuencia, como se describe con referencia a las ecuaciones (2)-(13) a continuación. Los índices de cuantificación l_m e l para la frecuencia de reflexión y la parrilla de frecuencia óptima, respectivamente, representan el vector LSF de alta frecuencia codificado f^H y se transmiten al decodificador. La codificación del subvector de alta frecuencia f^H ocasionalmente se denominará "extrapolación" en la siguiente descripción.

En el ejemplo propuesto, la cuantificación se basa en un conjunto de cuantificadores escalares (SQ) optimizados individualmente sobre las propiedades estadísticas de los parámetros anteriores. En una implementación alternativa, los elementos LSF podrían enviarse a un cuantificador vectorial (VQ) o uno podría incluso entrenar un VQ para el conjunto combinado de parámetros (LSF, frecuencia de reflexión y parrilla óptima).

Las LSF de baja frecuencia del subvector L^L están en el paso S6 volteadas en el espacio abarcado por las LSF de alta frecuencia del subvector f^H . Esta operación se ilustra en la figura 3. Primero, la frecuencia de reflexión cuantificada \hat{f}_m se calcula de acuerdo con:

$$\hat{f}_m = Q(f(M/2) - \hat{f}(M/2-1)) + \hat{f}(M/2-1) \quad (2)$$

5 donde f denota todo el vector LSF, y $Q(\cdot)$ es la cuantificación de la diferencia entre el primer elemento en f^H (en concreto $f(M/2)$) y el último elemento cuantificado en f^H (en concreto $\hat{f}(M/2-1)$), y donde M denota el número total de elementos en la representación espectral paramétrica.

A continuación, las LSF volteadas $f_{flip}(k)$ se calculan de acuerdo con:

$$10 \quad f_{flip}(k) = 2\hat{f}_m - \hat{f}(M/2-1-k), \quad 0 \leq k \leq M/2-1 \quad (3)$$

Después, las LSF volteadas se vuelven a escalar para que estén dentro del intervalo $[0 \dots 0,5]$ (como alternativa, el intervalo se puede representar en radianes como $[0 \dots \pi]$) de acuerdo con:

$$15 \quad \tilde{f}_{flip}(k) = \begin{cases} (f_{flip}(k) - f_{flip}(0)) \cdot (f_{max} - \hat{f}_m) / (\hat{f}_m + f_{flip}(0)), & \hat{f}_m > 0,25; \\ f_{flip}(k), & \text{de otro modo} \end{cases} \quad (4)$$

Las parrillas de frecuencia g^i se reescalan para encajarlas al intervalo entre el último elemento LSF cuantificado $\hat{f}(M/2-1)$ y un valor máximo de punto de parrilla g_{max} , es decir:

$$20 \quad \tilde{g}^i(k) = g^i(k) \cdot (g_{max} - \hat{f}(M/2-1)) + \hat{f}(M/2-1) \quad (5)$$

Estos coeficientes volteados y reescalados $\hat{f}_{flip}(k)$ (denominados colectivamente \hat{f}^H en la figura 2) se procesan adicionalmente en el paso S7 mediante el alisado con las parrillas de frecuencia reescaladas $\hat{g}^i(k)$. El alisado tiene la forma de una suma ponderada entre las LSF volteadas y reescaladas $\hat{f}_{flip}(k)$ y las parrillas de frecuencia reescaladas $\hat{g}^i(k)$, de acuerdo con:

$$25 \quad f_{smooth}(k) = [1 - \lambda(k)] \tilde{f}_{flip}(k) + \lambda(k) \tilde{g}^i(k) \quad (6)$$

30 donde $\lambda(k)$ y $[1-\lambda(k)]$ son pesos predefinidos.

Como la ecuación (6) incluye un índice i libre, esto significa que se generará un vector $f_{smooth}(k)$ para cada $\hat{g}^i(k)$. Por lo tanto, la ecuación (6) puede expresarse como:

$$35 \quad f_{smooth}^i(k) = [1 - \lambda(k)] \tilde{f}_{flip}(k) + \lambda(k) \tilde{g}^i(k) \quad (7)$$

El alisado se realiza en el paso S7 en una búsqueda de bucle cerrado sobre todas las parrillas de frecuencia g^i , para encontrar la que minimice un criterio predefinido (descrito después de la ecuación (12) a continuación).

40 Para $M/2 = 5$ los pesos $\lambda(k)$ en la ecuación (7) se pueden elegir como:

$$\lambda = \{0,2, 0,35, 0,5, 0,75, 0,8\} \quad (8)$$

En un ejemplo, estas constantes se optimizan perceptivamente (se sugieren diferentes conjuntos de valores, y finalmente se selecciona el conjunto que maximiza la calidad, según lo informado por un panel de oyentes). En general, los valores de los elementos en λ aumentan a medida que aumenta el índice k . Dado que un índice más alto corresponde a una frecuencia más alta, las frecuencias más altas del espectro resultante están más influenciadas por $\hat{g}^i(k)$ que por \hat{f}_{flip} (véase la ecuación (7)). Este resultado de este promedio alisado o ponderado es un espectro más plano hacia las frecuencias altas (la estructura del espectro potencialmente introducida por los f_{flip} se elimina progresivamente hacia las frecuencias altas).

50 Aquí, g_{max} se selecciona cerca de 0,5 pero menor. En este ejemplo, g_{max} se selecciona igual a 0,49.

El método en este ejemplo utiliza 4 parrillas entrenadas g^i (es posible menos o más parrillas). Los vectores de parrilla de plantilla en un intervalo [0 ... 1], almacenados previamente en la memoria, son de la siguiente forma:

$$\begin{cases} g^1 = \{0,17274857, 0,35811835, 0,52369229, 0,71552804, 0,85539771\} \\ g^2 = \{0,16313042, 0,30782962, 0,43109281, 0,59395830, 0,81291897\} \\ g^3 = \{0,17172427, 0,33157177, 0,48528862, 0,66492442, 0,82952486\} \\ g^4 = \{0,16666667, 0,33333333, 0,50000000, 0,66666667, 0,83333333\} \end{cases} \quad (9)$$

- 5 Si asumimos que la posición del último coeficiente LSF cuantificado $\hat{f}(M/2-1)$ es 0,25, los vectores de parrilla reescalada toman la forma:

$$\begin{cases} \tilde{g}^1 = \{0,2915, 0,3359, 0,3757, 0,4217, 0,4553\} \\ \tilde{g}^2 = \{0,2892, 0,3239, 0,3535, 0,3925, 0,4451\} \\ \tilde{g}^3 = \{0,2912, 0,3296, 0,3665, 0,4096, 0,4491\} \\ \tilde{g}^4 = \{0,2900, 0,3300, 0,3700, 0,4100, 0,4500\} \end{cases} \quad (10)$$

- 10 En la figura 4 se ilustra un ejemplo del efecto de alisar los coeficientes LSF volteados y reescalados a los puntos de la parrilla. Con el aumento del número de vectores de parrilla utilizados en el procedimiento de bucle cerrado, el espectro resultante se acerca más y más al espectro de destino.

Si $g_{\max} = 0,5$ en vez de 0,49, el código de libros de la parrilla de frecuencia puede estar formado por:

$$\begin{cases} g^1 = \{0,15998503, 0,31215086, 0,47349756, 0,66540429, 0,84043882\} \\ g^2 = \{0,15614473, 0,30697672, 0,45619822, 0,62493785, 0,77798001\} \\ g^3 = \{0,14185823, 0,26648724, 0,39740108, 0,55685745, 0,74688616\} \\ g^4 = \{0,15416561, 0,27238427, 0,39376780, 0,59287916, 0,86613986\} \end{cases} \quad (11)$$

- 15 Si suponemos nuevamente que la posición del último coeficiente LSF cuantificado $\hat{f}(M/2-1)$ es 0,25, los vectores de parrilla reescalados toman la forma:

$$\begin{cases} \tilde{g}^1 = \{0,28999626, 0,32803772, 0,36837439, 0,41635107, 0,46010970\} \\ \tilde{g}^2 = \{0,28903618, 0,32674418, 0,36404956, 0,40623446, 0,44449500\} \\ \tilde{g}^3 = \{0,28546456, 0,31662181, 0,34935027, 0,38921436, 0,43672154\} \\ \tilde{g}^4 = \{0,28854140, 0,31809607, 0,34844195, 0,39821979, 0,46653496\} \end{cases} \quad (12)$$

- 20 Se observa que las parrillas reescaladas \tilde{g}^j pueden ser diferentes de trama a trama, ya que $\hat{f}(M/2-1)$ en la ecuación (5) de reescalado puede no ser constante sino variar con el tiempo. Sin embargo, el libro de códigos formado por las parrillas de plantilla g^j es constante. En este sentido, las parrillas reescaladas \tilde{g}^j pueden considerarse como un libro de códigos adaptativo formado a partir de un libro de códigos fijo de parrillas de plantilla g^j .

- 25 Los vectores LSF f_{smooth}^i creados por la suma ponderada en (7) se comparan con el vector LSF de destino f^H , y se selecciona la parrilla óptima como la que minimiza el error cuadrático medio (MSE) entre estos dos vectores. El índice opt de esta parrilla óptima puede expresarse matemáticamente como:

$$opt = \arg \min_i \left(\sum_{k=0}^{M/2-1} (f_{\text{smooth}}^i(k) - f^H(k))^2 \right) \quad (13)$$

- 30 donde $f^H(k)$ es un vector de destino formado por los elementos de la parte de alta frecuencia de la representación espectral paramétrica.

En una implementación alternativa, se pueden utilizar medidas de error más avanzadas que imitan la distorsión

espectral (SD), por ejemplo, la media armónica inversa u otra ponderación en el dominio LSF.

En un ejemplo, el libro de códigos de la parrilla de frecuencia se obtiene con un algoritmo de agrupación de medias K en un gran conjunto de vectores LSF, que se ha extraído de una base de datos de voz. Los vectores de parrilla en las ecuaciones (9) y (11) se seleccionan como los que, después de reescalarse de acuerdo con la ecuación (5) y el promedio ponderado con \hat{f}_{rip} de acuerdo con la ecuación (7), minimizan la distancia cuadrada a f^H . En otras palabras, estos vectores de parrilla, cuando se utilizan en la ecuación (7), proporcionan la mejor representación de los coeficientes LSF de alta frecuencia.

La figura 5 es un diagrama de bloques de un ejemplo del codificador de acuerdo con la tecnología propuesta. El codificador 40 incluye un codificador 10 de baja frecuencia configurado para codificar una parte de baja frecuencia de la representación espectral paramétrica f por la cuantificación de elementos de la representación espectral paramétrica que corresponde a una parte de baja frecuencia de la señal de audio. El codificador 40 también incluye un codificador 12 de alta frecuencia configurado para codificar una parte de alta frecuencia f^H de la representación espectral paramétrica mediante un promedio ponderado basado en los elementos cuantificados \hat{f}^L volteados alrededor de una frecuencia de reflexión cuantificada que separa la parte de baja frecuencia de la parte de alta frecuencia, y una parrilla de frecuencia determinada a partir de un libro 24 de códigos de parrilla de frecuencia en un procedimiento de búsqueda de bucle cerrado. Las entidades cuantificadas $\hat{f}^L, \hat{f}_m, g^{opt}$ están representadas por los correspondientes índices de cuantificación l_L, l_m, l_g , que se transmiten al decodificador.

La figura 6 es un diagrama de bloques de un ejemplo del codificador de acuerdo con la tecnología propuesta. El codificador 10 de baja frecuencia recibe el vector LSF completo f , que se divide en una parte de baja frecuencia o subvector f^L y una parte de alta frecuencia o subvector f^H mediante un divisor 14 de vector. La parte de baja frecuencia se reenvía a un cuantificador 16, que está configurado para codificar la parte de baja frecuencia f^L cuantificando sus elementos, ya sea mediante cuantificación escalar o vectorial, en una parte o subvector de baja frecuencia cuantificada \hat{f}^L . Al menos un índice de cuantificación de l_L , (dependiendo del método de cuantificación utilizado) se envía para su transmisión al decodificador.

El subvector f de baja frecuencia cuantificado \hat{f}^L y el subvector de frecuencia alta todavía no codificado f^H se envían al codificador 12 de alta frecuencia. Un calculador 18 de frecuencia de reflexión está configurado para calcular la frecuencia de reflexión cuantificada \hat{f}_m de acuerdo con la ecuación (2). Las líneas discontinuas indican que solo el último elemento cuantificado $\hat{f}^{(M/2-1)}$ en \hat{f}^L y el primer elemento $f^{(M/2)}$ en f^H son necesarios para esto. El índice de cuantificación l_m que representa la frecuencia de reflexión cuantificada \hat{f}_m se envía para su transmisión al decodificador.

La frecuencia de reflexión cuantificada \hat{f}_m se envía a una unidad 20 de volteado de subvector cuantificada de baja frecuencia configurada para voltear los elementos del subvector de baja frecuencia cuantificado \hat{f}^L alrededor de la frecuencia de reflexión cuantificada \hat{f}_m de acuerdo con la ecuación (3). Los elementos volteados $f_{rip}(k)$ y la frecuencia de reflexión cuantificada \hat{f}_m se envían a un reescalador 22 de elemento volteado configurado para reescalar los elementos volteados de acuerdo con la ecuación (4).

Las parrillas de frecuencia $g'(k)$ se envían desde el libro 24 de códigos de parrilla de frecuencia a un reescalador 26 de parrilla de frecuencia, que también recibe el último elemento cuantificado $\hat{f}^{(M/2-1)}$ en \hat{f}^L . El reescalador 26 está configurado para realizar un reescalado de acuerdo con la ecuación (5).

Las LSF volteadas y reescaladas $\hat{f}_{rip}(k)$ del reescalador 22 de elementos volteados y las parrillas de frecuencia reescaladas $\tilde{g}^j(k)$ del reescalador 26 de parrilla de frecuencia se envían a una unidad 28 de ponderación, que está configurada para realizar un promedio ponderado de acuerdo con la ecuación (7). Los elementos alisados resultantes $f_{smooth}^j(k)$ y el vector de destino de alta frecuencia f^H se envían a una unidad 30 de búsqueda de parrilla de frecuencia configurada para seleccionar una parrilla de frecuencia g^{opt} de acuerdo con la ecuación (13). El índice l_g correspondiente se transmite al decodificador.

Decodificador

La figura 7 es un diagrama de flujo del método de decodificación de acuerdo con la tecnología propuesta. El paso S11 reconstruye los elementos de una parte de baja frecuencia de la representación espectral paramétrica correspondiente a una parte de baja frecuencia de la señal de audio a partir de al menos un índice de cuantificación que codifica esa parte de la representación espectral paramétrica. El paso S12 reconstruye elementos de una parte de alta frecuencia de la representación espectral paramétrica por promedio ponderado basada en los elementos

decodificados volteados alrededor de una frecuencia de reflexión decodificada, que separa la parte de baja frecuencia de la parte de alta frecuencia y una parrilla de frecuencia decodificada.

5 Los pasos del método realizados en el decodificador se ilustran mediante la realización en la figura 8. Primero, los índices de cuantificación l_{fL} , l_m , l_g para las LSF de baja frecuencia, la frecuencia de reflexión óptima y la parrilla óptima, respectivamente, son recibidos.

10 En el paso S13, la parte de baja frecuencia cuantificada se reconstruye a partir de un libro de códigos de baja frecuencia utilizando el índice recibido l_{fL} .

Los pasos del método realizados en el decodificador para reconstruir la parte de alta frecuencia \hat{f}^H son muy similares a los pasos de procesamiento de codificador ya descritos en las ecuaciones (3)-(7).

15 Los pasos de volteado y reescalado realizados en el decodificador (en S14) son idénticos a las operaciones del codificador, y por lo tanto se describen exactamente mediante las ecuaciones (3)-(4).

20 Los pasos (en S15) de reescalado de la parrilla (ecuación (5)) y el alisado con ella (ecuación (6)), requieren solo una ligera modificación en el decodificador, porque la búsqueda de bucle cerrado no se realiza (búsqueda sobre i). Esto se debe a que el decodificador recibe el índice óptimo opt desde el flujo de bits. Estas ecuaciones toman la siguiente forma:

$$\tilde{g}^{opt}(k) = g^{opt}(k) \cdot (g_{max} - \hat{f}(M/2-1)) + \hat{f}(M/2-1) \quad (14)$$

y

$$f_{smooth}(k) = [1 - \lambda(k)] \tilde{f}_{flip}(k) + \lambda(k) \tilde{g}^{opt}(k) \quad (15)$$

respectivamente. El vector f_{smooth} representa la parte de alta frecuencia \hat{f}^H de la señal decodificada.

30 Finalmente, las partes de baja y alta frecuencia \hat{f}^L, \hat{f}^H del vector LSF se combinan en el paso S16, y el vector \hat{f} resultante se transforma en coeficientes AR \hat{a} en el paso S17.

35 La figura 9 es un diagrama de bloques de una realización del decodificador 50 de acuerdo con la tecnología propuesta. Un decodificador 60 de baja frecuencia se configura para reconstruir elementos \hat{f}^L de una parte de baja frecuencia f^L de la representación espectral paramétrica f que corresponde a una parte de baja frecuencia de la señal de audio desde al menos un índice de cuantificación l_{fL} , codificando esa parte de la representación espectral paramétrica. Un decodificador 62 de alta frecuencia está configurado para reconstruir elementos \hat{f}^H de una parte de alta frecuencia f^H de la representación espectral paramétrica mediante promedio ponderado basado en los elementos decodificados \hat{f}^L volteados alrededor de una frecuencia de reflexión decodificada \hat{f}_m , que separa la parte de baja frecuencia de la parte de alta frecuencia, y una parrilla de frecuencia decodificada g^{opt} . La parrilla de frecuencia g^{opt} se obtiene al recuperar la parrilla de frecuencia que corresponde a un índice recibido l_g de un libro 24 de códigos de parrilla de frecuencia (este es el mismo libro de códigos que en el codificador).

45 La figura 10 es un diagrama de bloques de una realización del decodificador de acuerdo con la tecnología propuesta. El decodificador de baja frecuencia recibe al menos un índice de cuantificación l_{fL} , dependiendo de si se utiliza cuantificación escalar o vectorial, y lo envía a un decodificador 66 de índice de cuantificación, que reconstruye los elementos \hat{f}^L de la parte de baja frecuencia de la representación espectral paramétrica. El decodificador 62 de alta frecuencia recibe un índice de cuantificación de frecuencia de reflexión l_m , que se envía a un decodificador 66 de frecuencia de reflexión para decodificar la frecuencia de reflexión \hat{f}_m . Los bloques restantes 20, 22, 24, 26 y 28 realizan las mismas funciones que los bloques numerados correspondientemente en el codificador ilustrado en la figura 6. Las diferencias esenciales entre el codificador y el decodificador son que la frecuencia de reflexión se decodifica desde el índice l_m en vez de calcularse a partir de la ecuación (2), y que la unidad 30 de búsqueda de parrilla de frecuencia en el codificador no es necesaria, ya que se obtiene la parrilla de frecuencia óptima directamente desde el libro 24 de códigos de la parrilla de frecuencia buscando la parrilla de frecuencia g^{opt} que corresponde al índice recibido l_g .

60 Los pasos, funciones, procedimientos y/o bloques descritos en el presente documento pueden implementarse en hardware utilizando cualquier tecnología convencional, tal como la tecnología de circuito discreto o de circuito integrado, que incluye tanto circuitería electrónica de propósito general como circuitería específica de aplicación.

Alternativamente, al menos algunos de los pasos, funciones, procedimientos y/o bloques descritos en el presente documento pueden implementarse en un software para su ejecución mediante un equipo de procesamiento adecuado. Este equipo puede incluir, por ejemplo, uno o varios microprocesadores, uno o varios procesadores de señal digital (DSP), uno o varios circuitos integrados específicos de aplicación (ASIC), hardware acelerado por video o uno o varios dispositivos lógicos programables adecuados, como matriz de puertas programables de campo (FPGA). Las combinaciones de tales elementos de procesamiento también son factibles.

También debe entenderse que puede ser posible reutilizar las capacidades de procesamiento generales ya presentes en un UE. Esto puede, por ejemplo, hacerse reprogramando el software existente o agregando nuevos componentes de software.

La figura 11 es un diagrama de bloques de un ejemplo del codificador 40 de acuerdo con la tecnología propuesta. Este ejemplo se basa en un procesador 110, por ejemplo un microprocesador, que ejecuta el software 120 para cuantificar la parte de baja frecuencia f^l de la representación espectral paramétrica, y el software 130 para buscar

una extrapolación óptima representada por la frecuencia de reflexión \hat{f}_m y el vector de parrilla de frecuencia óptima g^{opt} . El software está almacenado en la memoria 140. El procesador 110 se comunica con la memoria a través de un bus del sistema. La representación espectral paramétrica entrante f se recibe mediante un controlador 150 de entrada/salida (E/S) que controla un bus de E/S, al que están conectados el procesador 110 y la memoria 140. El software 120 puede implementar la funcionalidad del codificador 10 de baja frecuencia. El software 130 puede implementar la funcionalidad del codificador 12 de alta frecuencia. Los parámetros cuantificados $\hat{f}_l, \hat{f}_m, g^{opt}$ (o preferiblemente los índices correspondientes l_l, l_m, l_g) obtenidos del software 120 y 130 son enviados desde la memoria 140 por el controlador 150 de E/S a través del bus de E/S.

La figura 12 es un diagrama de bloques de una realización del decodificador 50 de acuerdo con la tecnología propuesta. Esta realización se basa en un procesador 210, por ejemplo un microprocesador, que ejecuta el software 220 para decodificar la parte de baja frecuencia f^l de la representación espectral paramétrica, y el software 230 para decodificar la parte de baja frecuencia f^h de la representación espectral paramétrica por extrapolación. El software se almacena en la memoria 240. El procesador 210 se comunica con la memoria a través de un bus del sistema. Los parámetros codificados entrantes $\hat{f}_l, \hat{f}_m, g^{opt}$ (representados por l_l, l_m, l_g) son recibidos por un controlador 250 de entrada/salida (E/S) que controla un bus de E/S, al que el procesador 210 y la memoria 240 están conectados. El software 220 puede implementar la funcionalidad del decodificador 60 de baja frecuencia. El software 230 puede implementar la funcionalidad del decodificador 62 de alta frecuencia. La representación paramétrica decodificada \hat{f} (\hat{f}_l combinada con \hat{f}^h) obtenida del software 220 y 230 se envía desde la memoria 240 por el controlador 250 de E/S a través del bus de E/S.

La figura 13 ilustra un ejemplo de un equipo de usuario UE que incluye un codificador de acuerdo con la tecnología propuesta. Un micrófono 70 envía una señal de audio a un convertidor A/D 72. La señal de audio digitalizada está codificada por un codificador 74 de audio. Solamente los componentes relevantes para ilustrar la tecnología propuesta se ilustran en el codificador 74 de audio. El codificador 74 de audio incluye un estimador 76 de coeficiente AR, un convertidor 78 de AR a representación espectral paramétrica y un codificador 40 de la representación espectral paramétrica. La representación espectral paramétrica codificada (junto con otros parámetros de audio codificados que no son necesarios para ilustrar la presente tecnología) se envía a una unidad 80 de radio para codificación de canal y conversión ascendente a radiofrecuencia y transmisión a un decodificador a través de una antena.

La figura 14 ilustra una realización de un equipo de usuario UE que incluye un decodificador de acuerdo con la tecnología propuesta. Una antena recibe una señal que incluye la representación espectral paramétrica codificada y la envía a la unidad 82 de radio para la conversión descendente desde la frecuencia de radio y la decodificación de canal. La señal digital resultante se envía a un decodificador 84 de audio. Solamente los componentes relevantes para ilustrar la tecnología propuesta se ilustran en el decodificador 84 de audio. El decodificador 84 de audio incluye un decodificador 50 de la representación espectral paramétrica y una representación espectral paramétrica al convertidor AR 86. Los coeficientes AR se utilizan (junto con otros parámetros de audio decodificados que no son necesarios para ilustrar la tecnología actual) para decodificar la señal de audio, y las muestras de audio resultantes se envían a una unidad 88 de conversión y amplificación D/A, que envía la señal de audio a un altavoz 90.

En una aplicación de ejemplo, el esquema de cuantificación-extrapolación AR propuesto se utiliza en un contexto BWE. En este caso, el análisis AR se realiza en una cierta banda de alta frecuencia, y los coeficientes AR se utilizan solo para el filtro de síntesis. En lugar de obtenerse con el filtro de análisis correspondiente, la señal de excitación para esta banda alta se extrapola a partir de una excitación de banda baja codificada independientemente.

En otra aplicación de ejemplo, el esquema de cuantificación-extrapolación AR propuesto se utiliza en un esquema de codificación de tipo ACELP. Los codificadores ACELP modelan el tracto vocal de un hablante con un modelo AR. Una señal de excitación $e(n)$ se genera pasando una forma de onda $s(n)$ a través de un filtro de blanqueamiento $e(n)$

= $A(z)s(n)$, donde $A(z) = 1 + a_1z^{-1} + a_2z^{-2} + \dots + a_Mz^{-M}$, es el modelo AR de orden M. Trama por trama se cuantifica un conjunto de coeficientes AR $a = [a_1 a_2 \dots a_M]^T$ y la señal de excitación, y los índices de cuantificación se transmiten a través de la red. En el decodificador, el habla sintetizada se genera trama por trama enviando la señal de excitación reconstruida a través del filtro de síntesis reconstruido $A(z)^{-1}$.

5 En una aplicación de ejemplo adicional, el esquema de cuantificación-extrapolación AR propuesto se utiliza como una manera eficiente de parametrizar una envolvente de espectro de un códec de audio de transformación. En un tiempo breve, la forma de onda se transforma en dominio de la frecuencia, y la respuesta de frecuencia de los coeficientes AR se utiliza para aproximar la envolvente del espectro y normalizar el vector transformado (para crear un vector residual). A continuación, los coeficientes AR y el vector residual se codifican y transmiten al decodificador.

Los expertos en la técnica entenderán que pueden realizarse diversas modificaciones y cambios en la tecnología propuesta sin apartarse del alcance de la misma, que se define mediante las reivindicaciones adjuntas.

15 **Abreviaturas**

ACELP: Predicción lineal excitada de código algebraico

ASIC: Circuitos integrados específicos de aplicación

AR: Autoregresión

20 BWE: Extensión de ancho de banda

DSP: Procesador de señal digital

FPGA: Matriz de puertas programables de campo

ISP: Pares espectrales de inmitancia

LP: Predicción lineal

25 LSF: Frecuencias espectrales de línea

LSP: Par espectral de línea

MSE: Error cuadrático medio

SD: Distorsión espectral

SQ: Cuantificador escalar

30 UE: Equipo de usuario

VQ: Cuantificación de vector

Referencias

35 [1] 3GPP TS 26.090, *Códec de voz adaptable de multivelocidad (AMR); funciones de transcodificación*, p.13, 2007

[2] N. Iwakami, et al., *Codificación de audio de alta calidad a menos de 64 kbit/s utilizando la cuantificación del vector de intercalación ponderado en el dominio de transformación (TWINVQ)*, IEEE ICASSP, vol. 5, pp. 3095-3098, 1995

40 [3] J. Makhoul, *Predicción lineal: una revisión tutorial*, Proc. IEEE, vol 63, p.566, 1975

[4] P. Kabal y R.P. Ramachandran, *El cálculo de las frecuencias espectrales de línea utilizando polinomios de Chebyshev*, IEEE Trans. en ASSP, vol. 34, no. 6, pp. 1419-1426, 1986

REIVINDICACIONES

1.- Un método de decodificación de una representación espectral paramétrica codificada \hat{f} de coeficientes autoregresivos (a) que representan parcialmente una señal de audio, incluyendo dicho método los pasos de:

5 reconstruir (S11) coeficientes \hat{f}^L de una parte de baja frecuencia (f^L) de la representación espectral paramétrica (f) correspondiente a una parte de baja frecuencia de la señal de audio a partir de al menos un índice de cuantificación (I_{fL}) que codifica esa parte de la representación espectral paramétrica;

10 reconstruir (S12) coeficientes \hat{f}^H de una parte de alta frecuencia (f^H) de la representación espectral paramétrica mediante un promedio ponderado basado en los coeficientes decodificados \hat{f}^L volteados alrededor de una frecuencia de reflexión decodificada \hat{f}_m , que separa la parte de baja frecuencia de la parte de alta frecuencia y una parrilla de frecuencia decodificada (g^{opt}).

15 2.- El método de decodificación de la reivindicación 1, que incluye el paso de voltear los coeficientes decodificados \hat{f}^L de la parte de baja frecuencia alrededor de la frecuencia de reflexión \hat{f}_m de acuerdo con:

$$f_{flip}(k) = 2\hat{f}_m - \hat{f}(M/2 - 1 - k), \quad 0 \leq k \leq M/2 - 1$$

20 donde M denota el número total de coeficientes en la representación espectral paramétrica, y $\hat{f}(M/2-1-k)$ denota el elemento decodificado M/2-1-k.

3.- El método de decodificación de la reivindicación 2, que incluye el paso de reescalar los coeficientes volteados $f_{flip}(k)$ de acuerdo con:

$$\tilde{f}_{flip}(k) = \begin{cases} (f_{flip}(k) - f_{flip}(0)) \cdot (f_{max} - \hat{f}_m) / \hat{f}_m + f_{flip}(0), & \hat{f}_m > 0,25; \\ f_{flip}(k), & \text{de otro modo} \end{cases}$$

4.- El método de decodificación de la reivindicación 3, que incluye el paso de reescalar la parrilla de frecuencia decodificada g^{opt} para encajar en el intervalo entre el último elemento cuantificado $\hat{f}(M/2-1)$ en la parte de baja frecuencia y un valor máximo de punto de parrilla g_{max} de acuerdo con:

$$\tilde{g}^{opt}(k) = g^{opt}(k) \cdot (g_{max} - \hat{f}(M/2 - 1)) + \hat{f}(M/2 - 1).$$

5.- El método de decodificación de la reivindicación 4, que incluye el paso del promedio ponderado de los coeficientes volteados y reescalados $\tilde{f}_{flip}(k)$ y la parrilla de frecuencia reescalada $\tilde{g}^{opt}(k)$ de acuerdo con:

$$f_{smooth}(k) = [1 - \lambda(k)] \tilde{f}_{flip}(k) + \lambda(k) \tilde{g}^{opt}(k)$$

40 donde $\lambda(k)$ y $[1-\lambda(k)]$ son pesos predefinidos.

6.- El método de decodificación de la reivindicación 5, en el que $M = 10$, $g_{max} = 0,5$ y los pesos $\lambda(k)$ se definen como $\lambda = \{0,2, 0,35, 0,5, 0,75, 0,8\}$.

7.- El método de cualquiera de las reivindicaciones anteriores 1-6, en el que la decodificación se realiza en una representación de frecuencias espectrales de línea de los coeficientes autoregresivos.

8.- Un decodificador (50) para decodificar una representación espectral paramétrica codificada \hat{f} de coeficientes autoregresivos (a) que representan parcialmente una señal de audio, incluyendo dicho decodificador:

50 un decodificador (60) de baja frecuencia configurado para reconstruir coeficientes \hat{f}^L de una parte de baja frecuencia (f^L) de la representación espectral paramétrica (f) correspondiente a una parte de baja frecuencia de la señal de audio de al menos un índice de cuantificación (I_{fL}) que codifica esa parte de la representación espectral paramétrica;

un decodificador (62) de alta frecuencia configurado para reconstruir coeficientes $\hat{f}^{(H)}$ de una parte de alta frecuencia $(f^{(H)})$ de la representación espectral paramétrica mediante un promedio ponderado basado en los coeficientes decodificados $\hat{f}^{(L)}$ volteados alrededor de una frecuencia de reflexión decodificada \hat{f}_m , que separa la parte de baja frecuencia de la parte de alta frecuencia, y una parrilla de frecuencia decodificada (g^{opt}) .

9.- El decodificador de la reivindicación 8, en el que el decodificador (62) de alta frecuencia incluye una unidad (20) de volteado de subvector de baja frecuencia cuantificada configurada para voltear los coeficientes decodificados $\hat{f}^{(L)}$ de la parte de baja frecuencia alrededor de la frecuencia de reflexión de acuerdo con:

$$f_{flip}(k) = 2\hat{f}_m - \hat{f}(M/2 - 1 - k), \quad 0 \leq k \leq M/2 - 1$$

donde M denota el número total de coeficientes en la representación espectral paramétrica, y $\hat{f}^{(M/2-1-k)}$ denota el elemento decodificado M/2 - 1 - k.

10.- El decodificador de la reivindicación 9, en el que el decodificador (62) de alta frecuencia incluye un reescalador (22) de elemento volteado configurado para reescalar los coeficientes volteados $f_{flip}(k)$ de acuerdo con:

$$\tilde{f}_{flip}(k) = \begin{cases} (f_{flip}(k) - f_{flip}(0)) \cdot (f_{max} - \hat{f}_m) / \hat{f}_m + f_{flip}(0), & \hat{f}_m > 0,25; \\ f_{flip}(k), & \text{de otro modo} \end{cases}$$

11.- El decodificador de la reivindicación 10, en el que el decodificador (62) de alta frecuencia incluye un reescalador (26) de parrilla de frecuencia configurado para reescalar la parrilla de frecuencia decodificada g^{opt} para encajar en el intervalo entre el último coeficiente cuantificado $\hat{f}^{(M/2-1)}$ en la parte de baja frecuencia y un valor máximo de punto de parrilla g_{max} de acuerdo con:

$$\tilde{g}^{opt}(k) = g^{opt}(k) \cdot (g_{max} - \hat{f}(M/2 - 1)) + \hat{f}(M/2 - 1).$$

12.- El decodificador de la reivindicación 11, en el que el decodificador (62) de alta frecuencia incluye una unidad (28) de ponderación configurada para realizar promedios ponderados de los coeficientes volteados y reescalados $\tilde{f}_{flip}(k)$ y la parrilla de frecuencia reescalada $\tilde{g}^{opt}(k)$ de acuerdo con:

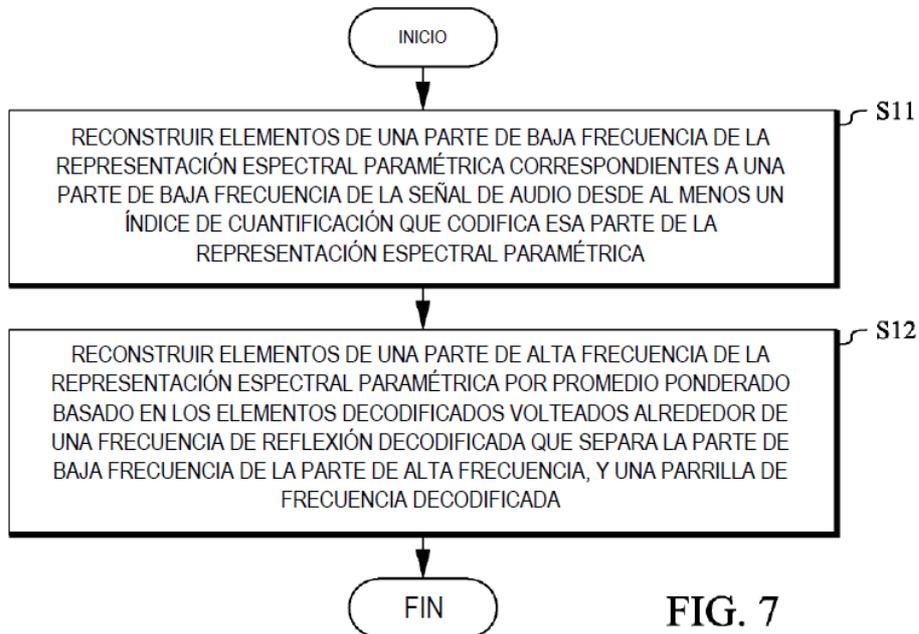
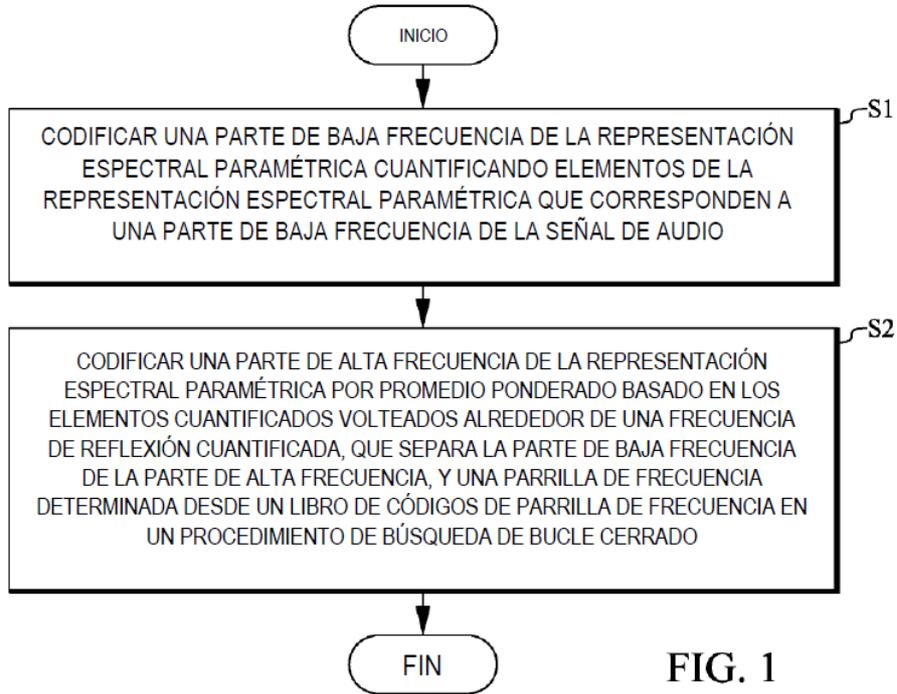
$$f_{smooth}(k) = [1 - \lambda(k)] \tilde{f}_{flip}(k) + \lambda(k) \tilde{g}^{opt}(k)$$

donde $\lambda(k)$ y $[1 - \lambda(k)]$ son pesos predefinidos.

13.- El decodificador de la reivindicación 12, en el que M = 10, $g_{max} = 0,5$ y los pesos $\lambda(k)$ se definen como $\lambda = \{0,2, 0,35, 0,5, 0,75, 0,8\}$.

14.- El decodificador de cualquiera de las reivindicaciones anteriores 8-13, en el que el decodificador está configurado para realizar la decodificación en una representación de frecuencias espectrales de línea de los coeficientes de autoregresivos.

15.- Un equipo de usuario que incluye un decodificador de acuerdo con cualquiera de las reivindicaciones precedentes 8-14.



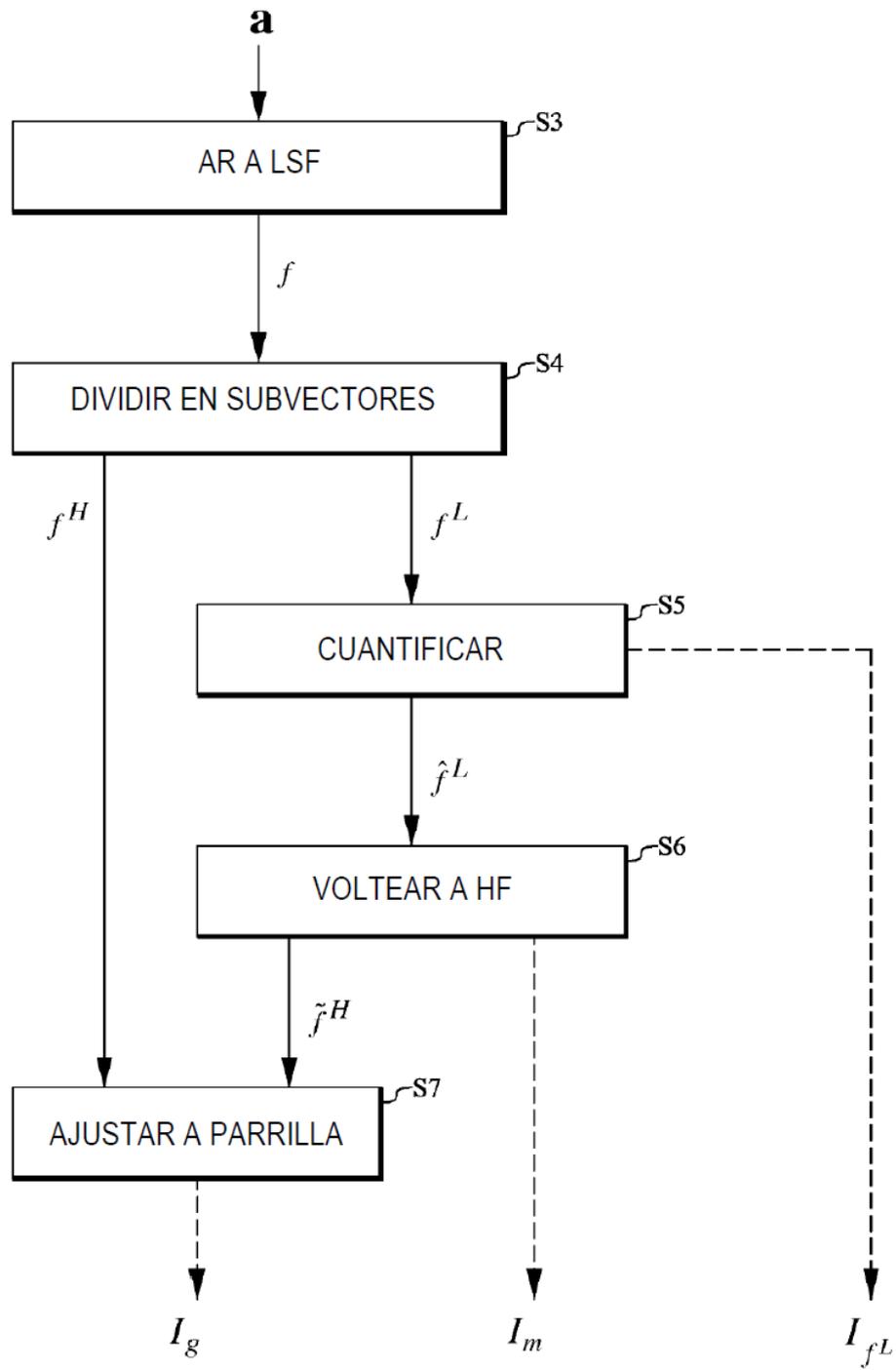


FIG. 2

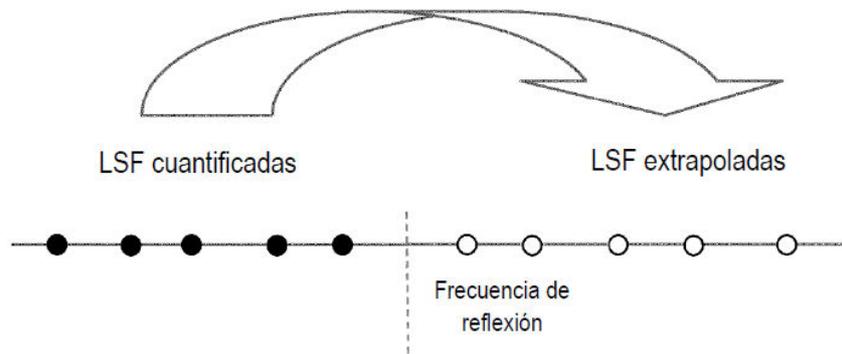


FIG. 3

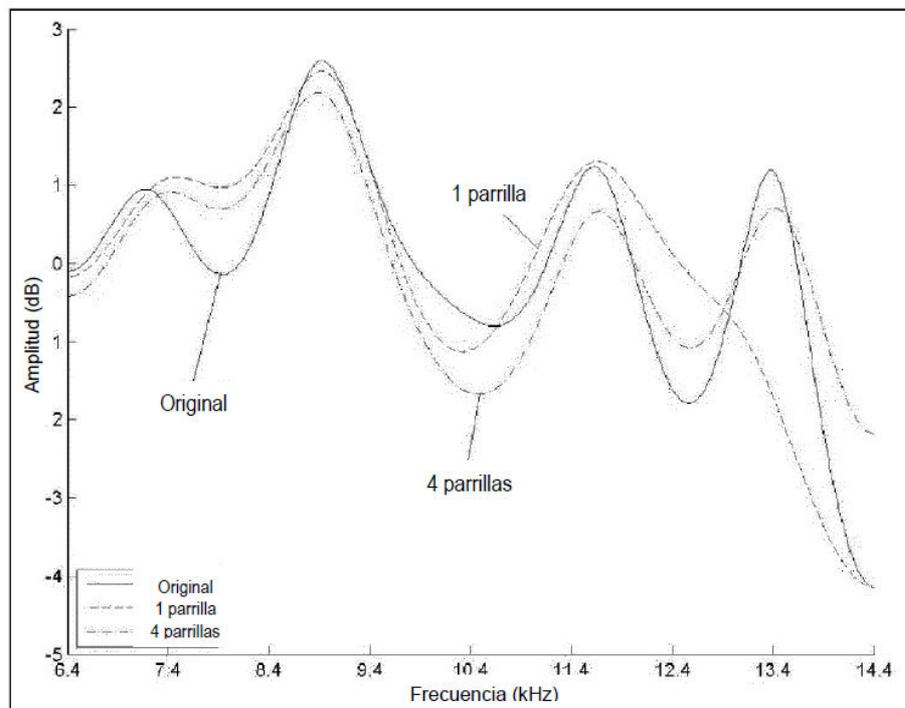


FIG. 4

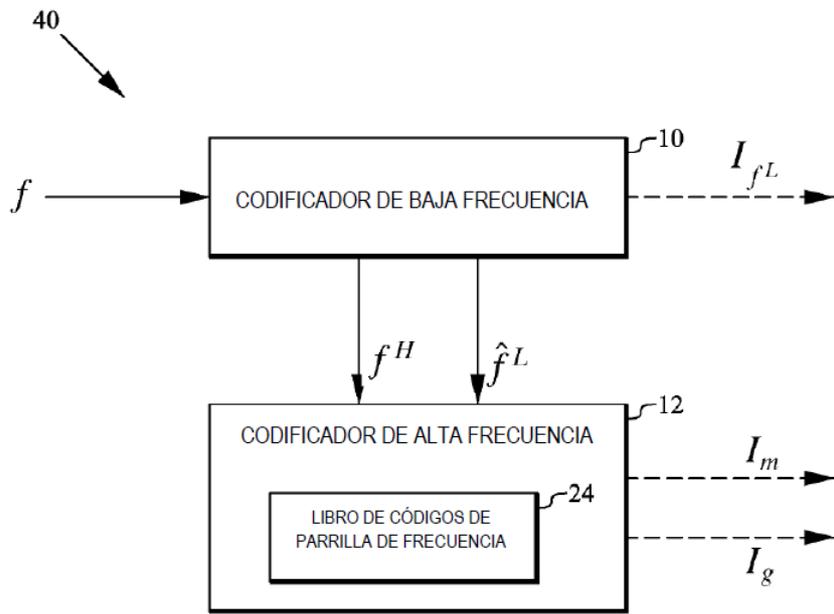


FIG. 5

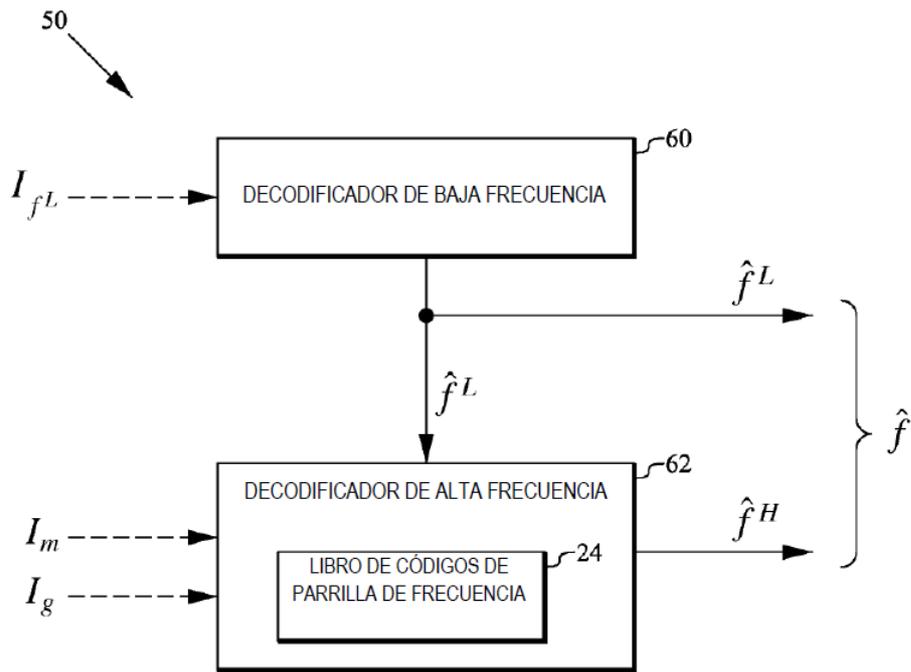


FIG. 9

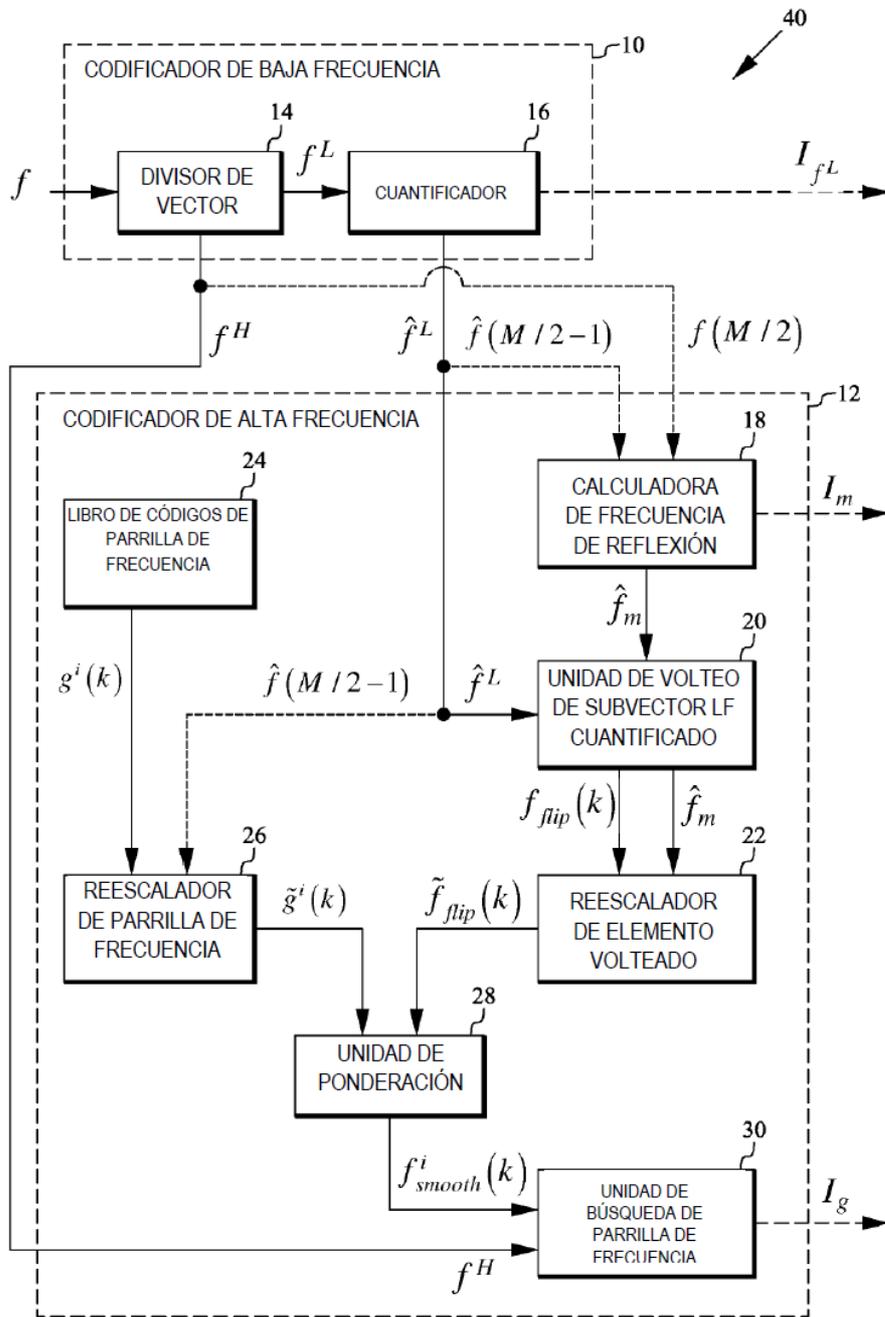


FIG. 6

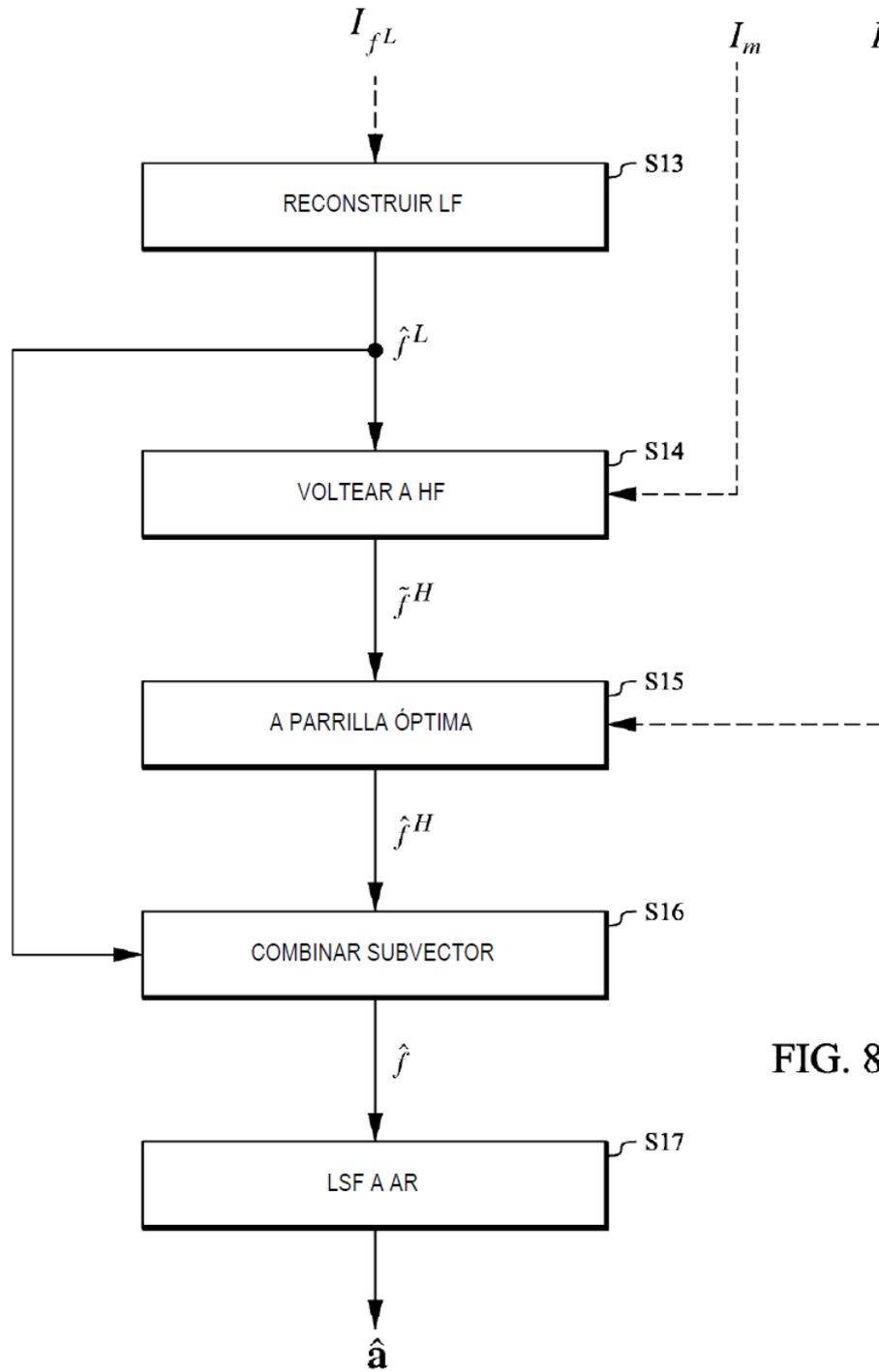


FIG. 8

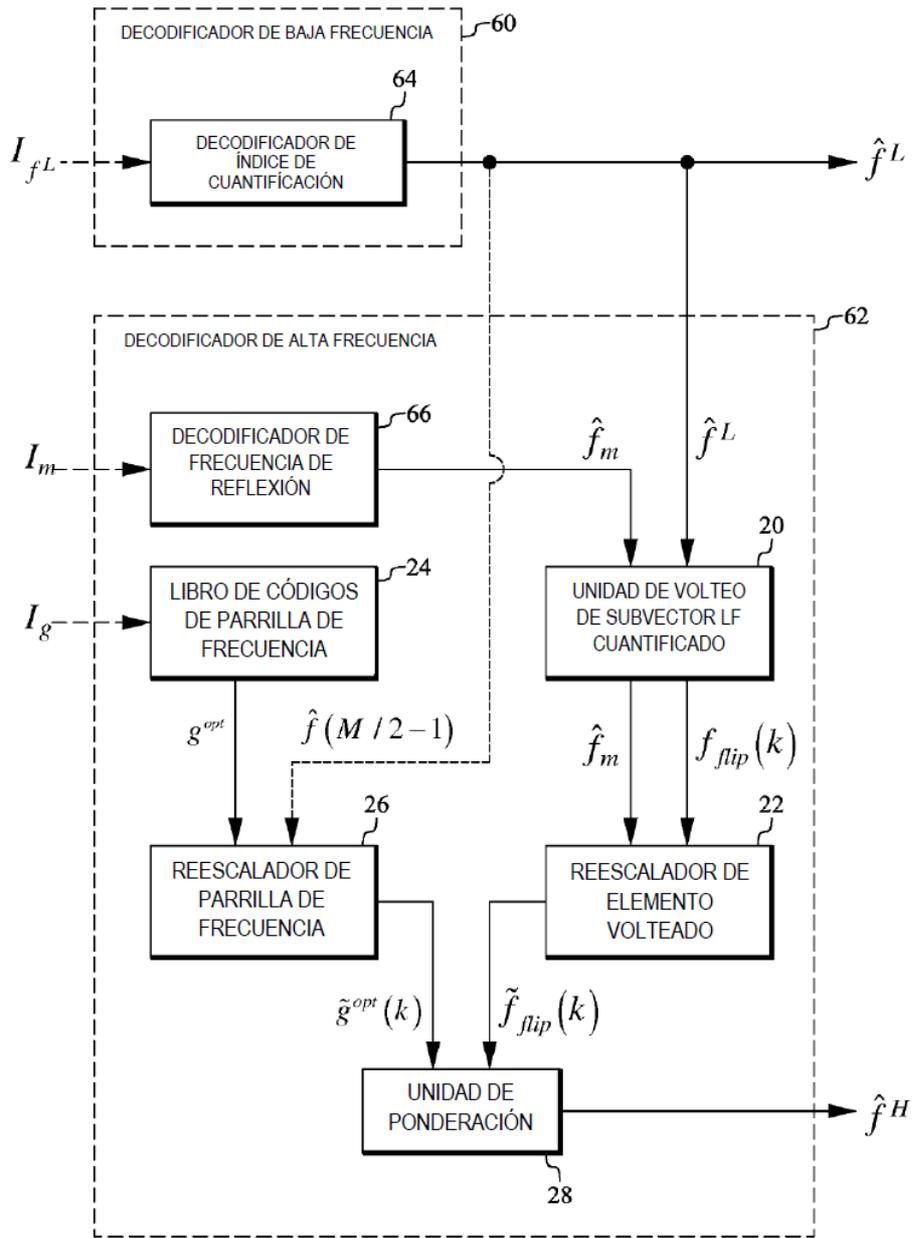


FIG. 10

50

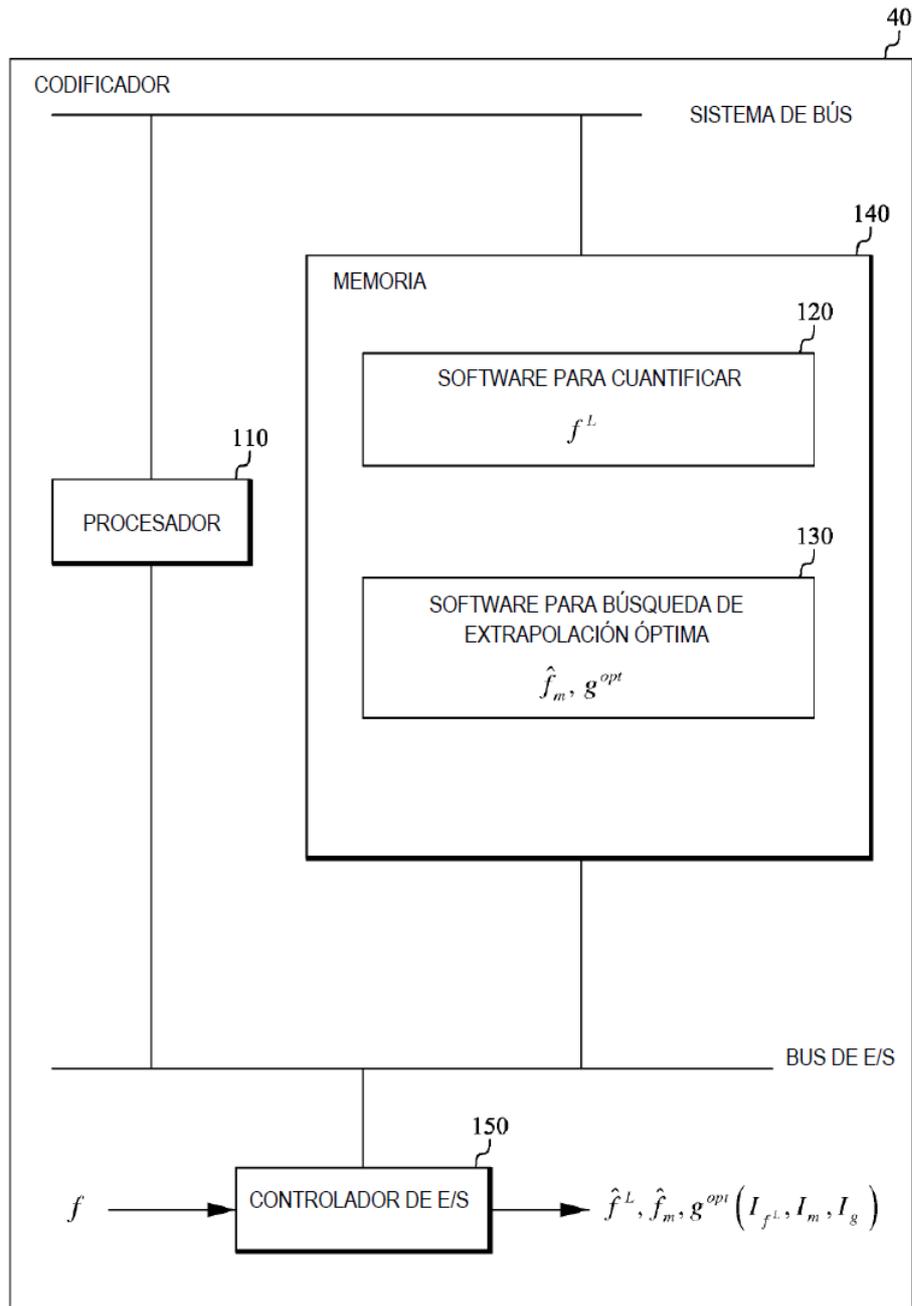


FIG. 11

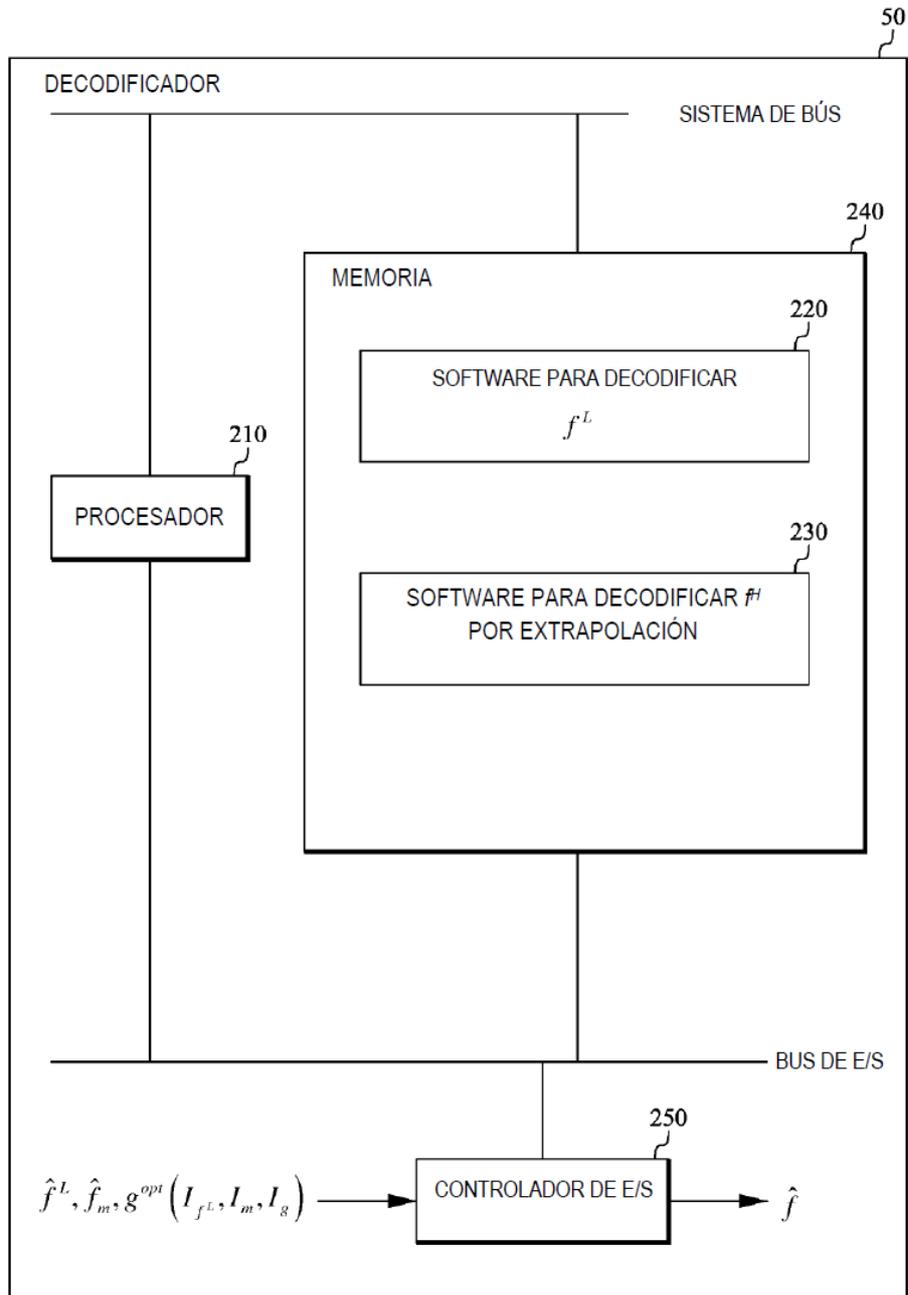


FIG. 12

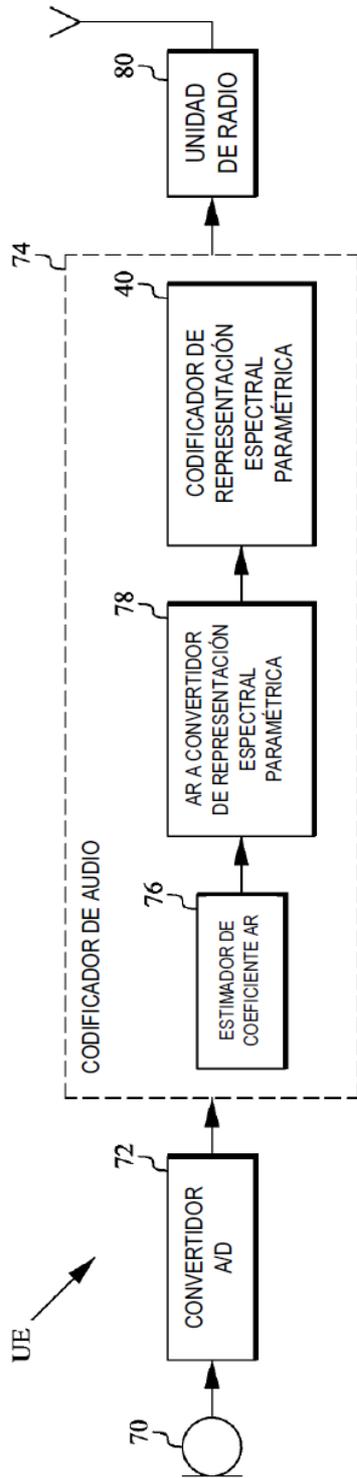


FIG. 13

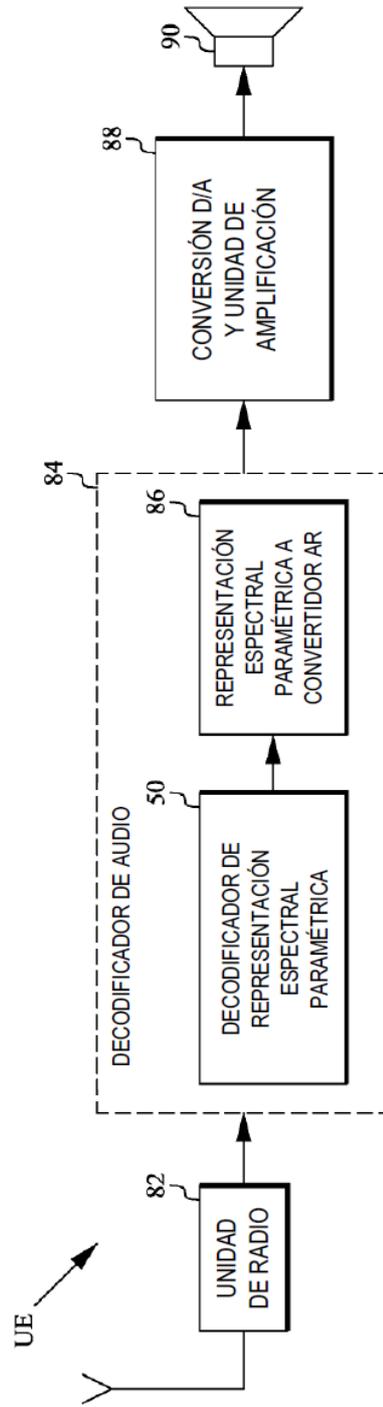


FIG. 14