

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 673 668**

51 Int. Cl.:

G10L 19/26 (2013.01)

G10L 25/18 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **05.03.2009 E 15151693 (7)**

97 Fecha y número de publicación de la concesión europea: **31.01.2018 EP 2863390**

54 Título: **Sistema y método para mejorar una señal sonora tonal decodificada**

30 Prioridad:

05.03.2008 US 64430

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

25.06.2018

73 Titular/es:

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

72 Inventor/es:

**VAILLANCOURT, TOMMY;
JELINEK, MILAN;
MALENOVSKY, VLADIMIR y
SALAMI, REDWAN**

74 Agente/Representante:

ELZABURU, S.L.P

ES 2 673 668 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

Sistema y método para mejorar una señal sonora tonal decodificada

Campo de la invención

5 La presente invención se refiere a un sistema y método para mejorar una señal sonora tonal decodificada, por ejemplo, una señal de audio, tal como una señal de música decodificada mediante el uso de un códec específico de voz. Con este fin, el sistema y método reducen un nivel de ruido de cuantificación en regiones del espectro que presentan baja energía.

Antecedentes de la invención

10 La demanda por técnicas eficientes de codificación de voz y audio digital con un equilibrio entre calidad subjetiva y tasa de bits está en aumento en diversas áreas de aplicación, tales como teleconferencias, multimedia y comunicaciones inalámbricas.

15 Un codificador de voz convierte una señal de voz en un flujo de bits digital que se transmite a través de un canal de comunicación o se almacena en un medio de almacenamiento. La señal de voz se digitaliza, es decir, se muestrea y cuantifica con usualmente 16 bits por muestra. El codificador de voz tiene el papel de representar las muestras digitales con una menor cantidad de bits a la vez que mantiene una calidad de voz con una buena calidad subjetiva. El decodificador o sintetizador de voz trabaja sobre el flujo de bits transmitido o almacenado y lo convierte de nuevo a una señal digital.

20 La codificación por *Predicción lineal con excitación por código* (CELP, por sus siglas en inglés) es una de las mejores técnicas de la técnica previa para lograr un buen equilibrio entre calidad subjetiva y tasa de bits. La técnica de codificación por CELP es la base de varias normas de codificación de voz, tanto en aplicaciones inalámbricas como por cable. En la codificación por CELP, la señal de voz muestreada se procesa en bloques sucesivos de L muestras, usualmente llamados *tramas*, donde L es una cantidad predeterminada de muestras que corresponde típicamente a 10-30 ms. Se computa y transmite un filtro de predicción (LP, por sus siglas en inglés) con cada trama. La computación del filtro de LP típicamente utiliza un parámetro *lookahead* [mirar adelante], por ejemplo, un segmento de voz de 5-15 ms de la trama subsiguiente. La trama de L muestras se divide en bloques menores llamados *subtramas*. Usualmente la cantidad de subtramas es de 3 (tres) o 4 (cuatro), lo cual da como resultado subtramas de 4-10 ms. En cada subtrama, se obtiene usualmente una señal de excitación de dos componentes, una excitación pasada y una excitación innovadora de libro de códigos fijo. El componente formado a partir de la excitación pasada se menciona a menudo como excitación de libro de códigos adaptativo o excitación de libro de códigos de tono. Los parámetros que caracterizan la señal de excitación se codifican y transmiten al decodificador, donde la señal de excitación se reconstruye y utiliza como entrada del filtro de LP.

25 En algunas aplicaciones, como la música de espera, se utilizan códigos específicos de voz de tasa de bits baja para trabajar sobre señales musicales. Esto usualmente da como resultado una mala calidad musical debido al uso de un modelo de producción de voz en un códec específico de voz con tasa de bits baja.

35 En algunas señales musicales, el espectro presenta una estructura tonal donde están presentes varios tonos (correspondientes a picos tonales) y no se relacionan armónicamente. Estas señales musicales son difíciles de codificar con un códec específico de voz con tasa de bits baja mediante el uso de un filtro de síntesis omnipolar y un filtro de tono. El filtro de tono es capaz de modelar segmentos de voz en los cuales el espectro presenta una estructura armónica que comprende una frecuencia fundamental y armónicos de esta frecuencia fundamental. Sin embargo, dicho filtro de tono no modela de forma adecuada los tonos que no estén armónicamente relacionados. Además, el filtro de síntesis omnipolar no modela los valles espectrales entre los tonos. Por lo tanto, cuando se utiliza un códec específico de voz con tasa de bits baja que utiliza un modelo de producción tal como CELP, las señales musicales presentan un ruido de cuantificación audible en las regiones de baja energía del espectro (regiones intertonales o valles espectrales). Un enfoque para reducir dicho ruido de cuantificación intertonal se describe, por ejemplo, en RAPPORTEUR Q9/16: "Updated draft new of new ITU-T Recommendation G.VBR-EV", ITU-T SG16 MEETING, 22-4-2008 - 2-5-2008, GINEBRA, n.º T05-SG16-080422-TD-WP3-0338, 24 de abril de 2008.

Compendio de la invención

50 Un objeto de la presente invención es mejorar una señal sonora tonal decodificada por un decodificador de un códec específico de voz en respuesta a un flujo de bits codificado recibido, por ejemplo, una señal de audio, tal como una señal musical, al reducir el ruido de cuantificación en las regiones de energía baja del específico (regiones intertonales o valles espectrales).

Más específicamente, según la presente invención, se proporciona un sistema para mejorar una señal sonora tonal decodificada según la reivindicación 2.

55 La presente invención se relaciona también con un método para mejorar una señal sonora tonal decodificada según la reivindicación 1.

Los objetivos, ventajas y características anteriores y otras de la presente invención serán evidentes al leer la descripción no restrictiva que sigue a continuación de realizaciones ilustrativas de esta, que se proporciona a modo de ejemplo solamente con referencia a los dibujos adjuntos.

Breve descripción de los dibujos

- 5 En los dibujos adjuntos:
 - la Figura 1 es un diagrama de bloques esquemático que muestra un resumen de un sistema y método para mejorar una señal sonora tonal decodificada;
 - la Figura 2 es una gráfica que ilustra la aplicación de una función de ventana en análisis espectral;
 - 10 la Figura 3 es un diagrama de bloques esquemático que muestra un resumen de un sistema y método para mejorar una señal sonora tonal decodificada;
 - la Figura 4 es un diagrama de bloques esquemático que ilustra la corrección de aumento de tono;
 - la Figura 5 es un diagrama de bloques esquemático de un ejemplo de clasificador de tipo de señal; y
 - 15 la Figura 6 es un diagrama de bloques esquemático de un decodificador de un códec específico de voz con tasa de bits baja que utiliza un modelo de producción que comprende el modelado con filtro de síntesis por LP de la forma del tracto vocal (envoltura espectral) y un modelado de filtro de tono de las cuerdas vocales (estructura fina armónica).

Descripción detallada

En la descripción detallada que sigue a continuación, se realiza una técnica de reducción de ruido intertonal en un códec específico de voz con tasa de bits baja para reducir el nivel de ruido de cuantificación intertonal, por ejemplo, en contenido musical. La técnica de reducción de ruido intertonal puede llevarse a cabo ya sea con señales sonoras de banda estrecha muestreadas a 8000 muestras/s o señales sonoras de banda ancha muestreadas a 16.000 muestras/s o a cualquier otra frecuencia de muestreo. La técnica de reducción de ruido intertonal se aplica a una señal sonora tonal decodificada para reducir el ruido de cuantificación en los valles espectrales (regiones de baja energía entre tonos). En algunas señales musicales, el espectro presenta una estructura tonal donde están presentes varios tonos (correspondientes a picos tonales) y no se relacionan armónicamente. Estas señales musicales son difíciles de codificar con un códec específico de voz con tasa de bits baja que utiliza un filtro de síntesis por LP omnipolar y un filtro de tono. El filtro de tono puede modelar los segmentos de voz sonoros que tengan un espectro que presente una estructura armónica con una frecuencia fundamental y armónicos de esta frecuencia fundamental. Sin embargo, el filtro de tono no modela de forma adecuada los tonos que no estén armónicamente relacionados. Además, el filtro de síntesis por LP omnipolar no modela los valles espectrales entre los tonos. Por lo tanto, mediante el uso de un códec específico de voz con tasa de bits baja con un modelo de producción de voz tal como CELP, las señales modeladas presentarán un ruido de cuantificación audible en las regiones de energía baja del espectro (regiones intertonales o valles espectrales). La técnica de reducción de ruido intertonal, por lo tanto, se refiere a la reducción del ruido de cuantificación en regiones espectrales de energía baja para mejorar una señal sonora tonal decodificada, más específicamente, para mejorar la calidad de la señal sonora tonal decodificada.

En una realización, el códec específico de voz con tasa de bits baja se basa en un modelo de producción de voz por CELP que trabaja ya sea sobre señales de banda estrecha o banda ancha (frecuencia de muestreo de 8 o 16 kHz). También puede utilizarse otra frecuencia de muestreo.

Se describirá brevemente un ejemplo 600 del decodificador de un códec específico de voz con tasa de bits baja que utiliza un modelo de producción de voz por CELP con referencia a la Figura 6. En respuesta a un índice de libro de códigos fijo del flujo de bits codificado recibido, un libro de códigos fijo 601 produce un vector de libro de códigos fijo 602 multiplicado por un aumento de libro de códigos fijo g para producir una excitación innovadora de libro de códigos fijo 603. De manera similar, un libro de códigos adaptativo 604 responde a un retraso de tono extraído del flujo de bits codificado recibido para producir un vector de libro de códigos adaptativo 607; también se le proporciona al libro de códigos adaptativo 604 (véase 605) la señal de excitación 610 a través de un bucle de retroalimentación que comprende un filtro de tono 606. El vector de libro de códigos adaptativo 607 se multiplica por un aumento G para producir una excitación de libro de códigos adaptativo 608. La excitación innovadora de libro de códigos fijo 603 y la excitación de libro de códigos adaptativo 608 se suman mediante un sumador 609 para formar la señal de excitación 610 que se proporciona a un filtro de síntesis por LP 611; el filtro de síntesis por LP 611 se controla mediante parámetros de filtro de LP extraídos del flujo de bits codificado recibido. El filtro de síntesis por LP 611 produce una señal sonora de síntesis 612, o una señal sonora tonal decodificada, que se puede someter a sobremuestreo/submuestreo en el módulo 613 antes de mejorarla mediante el uso del sistema 100 y el método para mejorar una señal sonora tonal decodificada.

Por ejemplo, se puede utilizar un códec basado en la estructura AMR-WB ([1] - 3GPP TS 26,190, "Adaptive Multi-Rate

- Wideband (AMR-WB) speech codec; Transcoding functions”). El códec de voz AMR-WB utiliza una frecuencia de muestreo interno de 12,8 kHz, y la señal se puede remuestrear ya sea a 8 o 16 kHz antes de llevar a cabo la reducción del ruido de cuantificación intertonal o, de manera alternativa, la reducción de ruido y mejora del audio se puede llevar a cabo a 12,8 kHz.

5 La Figura 1 es un diagrama de bloques esquemático que muestra un resumen de un sistema y método 100 para mejorar una señal sonora tonal decodificada.

Haciendo referencia a la Figura 1, un flujo de bits codificado 101 (señal sonora codificada) se recibe y procesa a través de un decodificador 102 (por ejemplo, el decodificador 600 de la Figura 6) de un códec específico de voz con tasa de bits baja para producir una señal sonora decodificada 103. Tal como se indica en la descripción que
10 antecede, el decodificador 102 puede ser, por ejemplo, un decodificador específico de voz que utiliza un modelo de producción por CELP, tal como un decodificador AMR-WB.

La señal sonora decodificada 103 en la salida del decodificador de señal sonora 102 se convierte (remuestrea) a una frecuencia de muestreo de 8 kHz. Sin embargo, debería recordarse que la técnica de reducción de ruido intertonal descrita en la presente puede aplicarse del mismo modo a señales sonoras tonales decodificadas a otras
15 frecuencias de muestreo, tales como 12,8 kHz o 16 kHz.

Se puede aplicar o no el procesamiento a la señal sonora decodificada 103. Cuando se aplica procesamiento, la señal sonora decodificada 103, por ejemplo, se preenfatisa mediante un preprocesador 104 antes de realizar el análisis espectral en el analizador espectral 105.

Para preenfatisar la señal sonora decodificada 103, el preprocesador 104 comprende un filtro de paso alto de primer orden (no se muestra). El filtro de paso alto de primer orden enfatiza frecuencias más altas de la señal sonora decodificada 103 y, con este fin, puede tener la siguiente función de transferencia:
20

$$H_{\text{pre-emph}}(z) = 1 - 0.68z^{-1} \quad (1)$$

donde z representa la variable transformada Z .

El preénfasis de las frecuencias más elevadas de la señal sonora decodificada 103 tiene la propiedad de aplanar el espectro de la señal sonora decodificada 103, que es útil para la reducción de ruido intertonal.
25

Después del preénfasis de las frecuencias más elevadas de la señal sonora decodificada 103 en el procesador 104:

- Se lleva a cabo el análisis espectral de la señal sonora decodificada preenfatisada 106 en el analizador espectral 105. Este análisis espectral utiliza la transformada discreta de Fourier (DFT, por sus siglas en inglés) y se describirá en mayor detalle en la descripción más adelante.

30 - La técnica de reducción de ruido intertonal se aplica en respuesta a los parámetros espectrales 107 del analizador espectral 107 y se implementa en un reductor 108 de ruido de cuantificación en las regiones espectrales de energía baja de la señal sonora tonal decodificada. El funcionamiento del reductor 108 del ruido de cuantificación se describirá en mayor detalle en la descripción más adelante.

35 - Un analizador inverso y un operador de solapar y sumar (*overlap-add*) 110 (a) aplica una DFT inversa (transformada discreta de Fourier) a los parámetros 109 espectrales con ruido intertonal reducido para convertir estos parámetros 109 de vuelta al dominio temporal y (b) utiliza una operación de solapar y sumar para reconstruir la señal sonora tonal decodificada mejorada 111. El funcionamiento del analizador inverso y operador de solapar y sumar 110 se describirá en mayor detalle en la descripción más adelante.

40 - Un postprocesador 112 postprocesa la señal sonora tonal decodificada mejorada 111 reconstruida a partir del analizador inverso y operador de solapar y sumar 110. El postprocesamiento es la inversa de la etapa de preprocesamiento (preprocesador 104) y, por lo tanto, puede consistir en atenuar las frecuencias más elevadas de la señal sonora tonal decodificada mejorada. Dicha atenuación se describirá en mayor detalle en la descripción más adelante.

45 - Finalmente, se puede proporcionar un sistema de reproducción de sonido 114 para convertir la señal sonora tonal decodificada mejorada postprocesada 113 del postprocesador 112 en un sonido audible.

Por ejemplo, el códec específico de voz en el cual se implementa la técnica de reducción de ruido intertonal trabaja sobre tramas de 20 ms que contienen 160 muestras a una frecuencia de muestreo de 8 kHz. Además, según este ejemplo, el decodificador de señal sonora 102 utiliza un parámetro *lookahead* de 10 ms de la trama futura para un mejor rendimiento de ocultación de borrado de tramas. Este parámetro *lookahead* se utiliza además en la técnica de
50 reducción de ruido intertonal para obtener una mejor resolución de frecuencia. La técnica de reducción de ruido

intertonal implementada en el reductor 108 de ruido de cuantificación sigue la misma estructura de tramas que el decodificador 102. Sin embargo, puede introducirse cierto desplazamiento entre la estructura de tramas del decodificador y la estructura de tramas de reducción de ruido intertonal para maximizar el uso del parámetro *lookahead*. En la descripción que sigue a continuación, los índices atribuidos a las muestras se reflejarán en la estructura de tramas de reducción de ruido intertonal.

Análisis espectral

Haciendo referencia a la Figura 3, se utiliza DFT (transformada discreta de Fourier) en el analizador espectral 105 para llevar a cabo un análisis espectral y estimación de la energía de espectro de la señal sonora tonal decodificada preenfatzada 106. En el analizador espectral 105, se realiza el análisis espectral en cada trama mediante el uso de ventanas de análisis de 30 ms con una superposición de 33 %. Más específicamente, el análisis espectral en el analizador 105 (Figura 3) se lleva a cabo una vez por trama mediante el uso de transformada rápida de Fourier (FFT, por sus siglas en inglés) de 256 puntos con la aplicación de una función de ventana de 33,3 % como se ilustra en la Figura 2. Las ventanas de análisis se colocan de manera de explotar la totalidad del parámetro *lookahead*. El comienzo de la primera ventana de análisis se desplaza 80 muestras después del comienzo de la trama actual del decodificador de señal sonora 102.

Las ventanas de análisis se utilizan para ponderar la señal sonora tonal decodificada preenfatzada 106 para el análisis de frecuencia. Las ventanas de análisis son planas en el centro con una función de seno en los bordes (Figura 2), lo cual se adapta bien para operaciones de solapar y sumar. Más específicamente, la ventana de análisis puede describirse de la siguiente manera:

$$w_{FFT}(n) = \begin{cases} \text{sen}\left(\frac{\pi n}{2L_{window}/3}\right), & n = 0, \dots, L_{window}/3 - 1 \\ 1, & n = L_{window}/3, \dots, 2L_{window}/3 - 1 \\ \text{sen}\left(\frac{\pi(n - L_{window}/3)}{2L_{window}/3}\right), & n = 2L_{window}/3, \dots, L_{window} - 1 \end{cases}$$

donde $L_{window} = 240$ muestras es el tamaño de la ventana de análisis. Dado que se utiliza una FFT de 256 puntos ($L_{FFT} = 256$), la señal sometida a función de ventana se rellena con 16 muestras cero.

Se podría utilizar una ventana de análisis alternativa en el caso de una señal de banda ancha con solamente un *lookahead* pequeño disponible. La ventana de análisis podría tener la siguiente forma:

$$w_{FFT_{WB}}(n) = \begin{cases} \text{sen}\left(\frac{\pi n}{2 \cdot L_{window_{WB}}/9}\right), & n = 0, \dots, L_{window_{WB}}/9 - 1 \\ 1, & n = L_{window_{WB}}/9, \dots, 8 \cdot L_{window_{WB}}/9 - 1 \\ \text{sen}\left(\frac{\pi(n - L_{window_{WB}}/9)}{2 \cdot L_{window_{WB}}/9}\right), & n = 8 \cdot L_{window_{WB}}/9, \dots, L_{window_{WB}} - 1 \end{cases}$$

donde $L_{window_{WB}} = 360$ es el tamaño de la ventana de análisis de banda ancha. En este caso, se utiliza una FFT de 512 puntos. Por lo tanto, la señal sometida a una función de ventana se rellena con 152 muestras cero. Posiblemente se pueden utilizar otras FFT de radix para reducir tanto como sea posible el relleno con ceros y para reducir la complejidad.

Mediante $s'(n)$ se indica la señal sonora tonal decodificada con un índice 0 correspondiente a la primera muestra en la trama de reducción de ruido intertonal (tal como se indicó anteriormente, en esta realización, este corresponde a 80 muestras después del comienzo de la trama decodificadora de señal sonora). La señal sonora tonal decodificada sometida a una función de ventana para el análisis espectral puede obtenerse mediante el uso de la siguiente

relación:

$$x_n^{(1)}(n) = \begin{cases} w_{FFT}(n)s'(n), & n = 0, \dots, L_{window} - 1 \\ 0, & n = L_{window}, \dots, L_{FFT} - 1 \end{cases} \quad (2)$$

donde $s'(0)$ es la primera muestra en la trama de reducción de ruido intertonal actual.

5 Se realiza una FFT a la señal sonora tonal decodificada sometida a una función de ventana para obtener un conjunto de parámetros espectrales por trama:

$$X^{(1)}(k) = \sum_{n=0}^{N-1} x_n^{(1)}(n) e^{-j2\pi \frac{kn}{N}}, \quad k = 0, \dots, L_{FFT} - 1 \quad (3)$$

donde $N = L_{FFT}$.

La salida de la FFT proporciona partes reales e imaginarias del espectro indicadas mediante $X_R(k)$, $k=0$ para $\frac{L_{FFT}}{2}$,
 y $X_I(k)$, $k=1$ para $\left(\frac{L_{FFT}}{2}-1\right)$. Nótese que $X_R(0)$ corresponde al espectro a 0 Hz (DC) y $X_R\left(\frac{L_{FFT}}{2}\right)$ corresponde al

10 espectro a $\frac{F_S}{2}$ Hz, donde F_S corresponde a la frecuencia de muestreo. El espectro en estos 2 (dos) puntos es solamente un valor real y usualmente se ignora en el análisis subsiguiente.

Después del análisis de FFT, el espectro resultante se divide en bandas críticas de frecuencia mediante el uso de intervalos que tienen los siguientes límites superiores; (17 bandas críticas en el rango de frecuencia de 0-4000 Hz y 21 bandas críticas de frecuencia en el rango de 0-8000 Hz) (véase [2]: J. D. Johnston, "Transform coding of audio signal using perceptual noise criteria," *IEEE J. Select. Areas Commun.*, vol. 6, pp. 314—323, feb. 1988).

15 En el caso de codificación de banda estrecha, las bandas críticas de frecuencia son = {100,0, 200,0, 300,0, 400,0, 510,0, 630,0, 770,0, 920,0, 1080,0, 1270,0, 1480,0, 1720,0, 2000,0, 2320,0, 2700,0, 3150,0, 3700,0, 3950,0} Hz.

20 En el caso de codificación de banda ancha, las bandas críticas de frecuencia son = {100,0, 200,0, 300,0, 400,0, 510,0, 630,0, 770,0, 920,0, 1080,0, 1270,0, 1480,0, 1720,0, 2000,0, 2320,0, 2700,0, 3150,0, 3700,0, 4400,0, 5300,0, 6700,0, 8000,0} Hz.

25 La FFT de 256 puntos o 512 puntos tiene como resultado una resolución de frecuencia de 31,25 Hz (4000/128=8000/256). Después de ignorar el componente DC del espectro, la cantidad de intervalos (*bins*) de frecuencia por banda crítica de frecuencia en el caso de codificación de banda estrecha es $M_{CB} = \{3, 3, 3, 3, 3, 4, 5, 4, 5, 6, 7, 7, 9, 10, 12, 14, 17, 12\}$, respectivamente, cuando la respectivamente se aproxima a 32Hz. En el caso de codificación de banda ancha $M_{CB} = \{3, 3, 3, 3, 3, 4, 5, 4, 5, 6, 7, 7, 9, 10, 12, 14, 17, 22, 28, 44, 41\}$.

La energía espectral promedio por banda crítica de frecuencia se computa de la siguiente manera:

$$E_{CB}(i) = \frac{1}{(L_{FFT}/2)^2 M_{CB}(i)} \sum_{k=0}^{M_{CB}(i)-1} (X_R^2(k+j_i) + X_I^2(k+j_i)), \quad i = 0, \dots, 17, \quad (4)$$

30 donde $X_R(k)$ y $X_I(k)$ son, respectivamente las partes real e imaginaria de la k^a cada de frecuencia, y j_i es el índice del primer intervalo en la i^a banda crítica proporcionado por $j_i = \{1, 4, 7, 10, 13, 16, 20, 25, 29, 34, 40, 47, 54, 63, 73, 85, 99, 116\}$ en el caso de codificación de banda estrecha y $j_i = \{1, 4, 7, 10, 13, 16, 20, 25, 29, 34, 40, 47, 54, 63, 73, 85, 99, 116, 138, 166, 210\}$ en el caso de codificación de banda ancha.

El analizador espectral 105 de la Figura 3 computa además la energía del espectro por intervalo de frecuencia, $E_{BIN}(k)$, para las primeras 17 bandas críticas (115 intervalos, sin incluir el componente DC) mediante el uso de la siguiente relación:

$$E_{BIN}(k) = X_R^2(k) + X_I^2(k), \quad k = 0, \dots, 114 \quad (5)$$

Finalmente, el analizador espectral 105 computa una energía espectral de trama total como un promedio de las energías espectrales de las primeras 17 bandas críticas de frecuencia mediante el analizador espectral 105 en una trama, mediante el uso de la siguiente relación:

$$E_{fr}^t = 10 \log \left(\sum_{i=0}^{i=16} \bar{E}_{CB}(i) \right), \quad \text{dB} \quad (6)$$

Los parámetros espectrales 107 del analizador espectral 105 de la Figura 3, más específicamente, la energía espectral promedio por banda crítica, la energía espectral por intervalo de frecuencia y la energía espectral de trama total calculadas anteriormente se utilizan en el reductor 108 para reducir el ruido de cuantificación y realizar una corrección de aumento.

Cabe destacar que, para una señal sonora tonal decodificada muestreada a 16.000 muestras/s, podrían utilizarse hasta 21 bandas críticas de frecuencia, pero la computación de la energía de trama total E_{fr}^t en el tiempo t se realizará sobre las primeras 17 bandas críticas.

Clasificador del tipo de señal:

La técnica de reducción de ruido intertonal llevada a cabo mediante el sistema y método 100 mejora una señal sonora tonal decodificada, tal como una señal musical, codificada por medio de un códec específico de voz. Usualmente, los sonidos no tonales, tales como la voz, se codifican bien mediante un código específico de voz y no necesitan este tipo de mejor basada en la frecuencia.

El sistema y método 100 para mejorar una señal sonora tonal decodificada comprende, además, tal como se ilustra en la Figura 3, un clasificador del tipo de señal 301 diseñado para maximizar además la eficacia del reductor 108 de ruido de cuantificación mediante la identificación de cuál sonido es adecuado para la reducción de ruido intertonal, como la música, y cuál sonido no lo es, tal como la voz.

El clasificador del tipo de señal 301 comprende la característica de no solamente separar la señal sonora decodificada en categorías de señal sonora, sino que también aporta una instrucción al reductor 108 de ruido de cuantificación para reducir al mínimo cualquier degradación posible de la voz.

Se ilustra un diagrama de bloques esquemático del clasificador del tipo de señal 301 en la Figura 5. En la realización presentada, el clasificador del tipo de señal 301 se mantuvo lo más simple posible. La entrada principal para el clasificador del tipo de señal 301 es la energía espectral de trama total E_t tal como se formula en la Ecuación (6).

Primero, el clasificador del tipo de señal 301 comprende un buscador 501 que determine un promedio de las pasadas 40 (cuarenta) variaciones de energía espectral de trama total (E_t) calculado mediante el uso de la siguiente relación:

$$\bar{E}_{diff} = \frac{\left(\sum_{t=40}^{t=-1} \Delta E^t \right)}{40}, \quad \text{donde } \Delta E^t = E_{fr}^t - E_{fr}^{(t-1)} \quad (7)$$

Entonces, el buscador 501 determina una desviación estadística del historial de variación de energía σ_E en las últimas 15 (quince) tramas mediante el uso de la siguiente relación:

$$\sigma_E = 0.7745967 \cdot \sqrt{\sum_{t=-15}^{t=-1} \frac{(\Delta E^t - \bar{E}_{diff})^2}{15}} \quad (8)$$

El clasificador del tipo de señal 301 comprende una memoria 502 actualizada con el promedio y la desviación de la variación de la energía espectral de trama total E_t según se calcula en las Ecuaciones (7) y (8).

La desviación resultante σ_E se compara con 4 (cuatro) umbrales flotantes en los comparadores 503-506 para determinar la eficacia del reductor 108 de ruido de cuantificación sobre la señal sonora decodificada actual. En el ejemplo de la Figura 5, la salida 302 (Figura 3) del clasificador del tipo de señal 301 se divide en 5 (cinco) categorías de señal sonora, llamadas categorías de señal sonora 0 a 4, donde cada categoría de señal sonora tiene su propio ajuste de reducción del ruido intertonal.

Las 5 (cinco) categorías de señal sonora 0-4 pueden determinarse tal como se indica en la siguiente Tabla:

Categoría	Banda mejorada (banda estrecha)	Banda mejorada (banda ancha)	Reducción permitida
	Hz	Hz	dB
0	NC	NC	0
1	[2000, 4000]	[2000, 8000]	6
2	[1270,4000]	[1270, 8000]	9
3	[700, 4000]	[700, 8000]	12
4	[400, 4000]	[400, 8000]	12

5 La categoría 0 de señal sonora es una categoría de señal sonora no tonal, como la voz, que no se modifica mediante la técnica de reducción del ruido intertonal. Esta categoría de señal sonora decodificada tiene una gran desviación estadística de historial de variación de energía espectral. Cuando la detección de las categorías 1-4 mediante los comparadores 503-506 sea negativa, un controlador 511 le indica al reductor 108 de ruido de cuantificación que no reduzca el ruido de cuantificación no tonal (Reducción = 0 dB).

10 Las tres categorías intermedias de señal sonora incluyen señales sonoras con diferentes tipos de desviación estadística del historial de variación de energía espectral.

La categoría 1 de señal sonora (la mayor variación después de la señal sonora decodificada de “tipo voz”) se detecta mediante el comparador 506 cuando la desviación estadística del historial de variación de energía espectral sea menor que un Umbral 1. Un controlador 510 responde a dicha detección mediante el comparador 506, cuando la última categoría de señal sonora detectada fue ≥ 0 , para indicarle al reductor 108 de ruido de cuantificación que

15 mejore la señal sonora tonal decodificada dentro de la banda de frecuencia de 2000 a $\frac{F_S}{2}$ Hz al reducir el ruido de cuantificación intertonal en una amplitud permitida máxima de 6 dB.

La categoría 2 de señal sonora se detecta mediante el comparador 505 cuando la desviación estadística del historial de variación de energía espectral sea menor que un Umbral 2. Un controlador 509 responde a dicha detección mediante el comparador 505, cuando la última categoría de señal sonora detectada fue ≥ 1 , para indicarle al reductor 108 de ruido de cuantificación que mejore la señal sonora tonal decodificada dentro de la banda de frecuencia de

20 $\frac{F_S}{2}$ 1270 a $\frac{F_S}{2}$ Hz al reducir el ruido de cuantificación intertonal en una amplitud permitida máxima de 9 dB.

La categoría 3 de señal sonora se detecta mediante el comparador 504 cuando la desviación estadística del historial de variación de energía espectral sea menor que un Umbral 3. Un controlador 508 responde a dicha detección mediante el comparador 504, cuando la última categoría de señal sonora detectada fue ≥ 2 , para indicarle al reductor 108 de ruido de cuantificación que mejore la señal sonora tonal decodificada dentro de la banda de frecuencia de

25 $\frac{F_S}{2}$ 700 a $\frac{F_S}{2}$ Hz al reducir el ruido de cuantificación intertonal en una amplitud permitida máxima de 12 dB.

La categoría 4 de señal sonora se detecta mediante el comparador 503 cuando la desviación estadística del historial de variación de energía espectral sea menor que un Umbral 4. Un controlador 507 responde a dicha detección mediante el comparador 503, cuando el último tipo de señal sonora detectado fue ≥ 3 , para indicarle al reductor 108 de ruido de cuantificación que mejore la señal sonora tonal decodificada dentro de la banda de frecuencia de 400 a

30 $\frac{F_S}{2}$ Hz al reducir el ruido de cuantificación intertonal en una amplitud permitida máxima de 12 dB.

En la realización de la Figura 5, el clasificador del tipo de señal 301 utiliza umbrales flotantes 1-4 para dividir la señal sonora decodificada en las diferentes categorías 0-4. Estos umbrales flotantes 1-4 son particularmente útiles para evitar la clasificación errónea del tipo de señal. Típicamente, una señal sonora tonal decodificada, como la música, tiene una desviación estadística mucho menor de su variación de energía espectral que una señal sonora no tonal, como la voz. Pero la música puede contener una desviación estadística mayor y la voz podría contener una variación estadística menor. Es improbable que el contenido de voz o música cambie de una trama a otra. Los

umbrales flotantes actúan como refuerzo para prevenir cualquier clasificación errónea que pudiera dar como resultado un rendimiento subóptimo del reductor 108 de ruido de cuantificación.

Se utilizan contadores de una serie de tramas de señal sonora de categoría 0 y de una serie de tramas de señal sonora de categoría 3 o 4, respectivamente, para disminuir o aumentar los umbrales.

5 Por ejemplo, si el contador 512 cuenta una serie de más de 30 tramas de señal de sonido de categoría 3 o 4, los umbrales flotantes 1-4 aumentarán en un controlador de umbral 514 para permitir que se consideren más tramas como señal sonora de categoría 4. Cada vez que el conteo del contador 512 aumenta, el contador 513 se reinicia a cero.

10 También se cumple la inversa con una señal sonora de categoría 0. Por ejemplo, si el contador 513 cuenta una serie de más de 30 tramas de señal de sonido de categoría 0, el controlador 514 disminuye los umbrales flotantes 1-4 para permitir que se consideren más tramas como señal sonora de categoría 0. Los umbrales flotantes 1-4 se limitan a valores máximos y mínimos absolutos para asegurar que el clasificador del tipo de señal 301 no se bloquee en una categoría fija.

El aumento y disminución de los umbrales 1-4 puede ilustrarse mediante las siguientes relaciones:

$$\begin{aligned}
 &IF (Nbr_cat4_frame > 30) \\
 &\quad Thres(i) = Thres(i) + TH_UP|_{i=1}^4 \\
 &ELSE IF (Nbr_cat0_frame > 30) \\
 &\quad Thres(i) = Thres(i) - TH_DWN|_{i=1}^4 \\
 &\quad Thres(i) = MIN(Thres(i), MAX_TH)|_{i=1}^4 \\
 &\quad Thres(i) = MAX(Thres(i), MIN_TH)|_{i=1}^4
 \end{aligned}$$

15 En el caso de eliminación de tramas, todos los umbrales 1-4 se reinician en sus valores mínimos y la salida del clasificador del tipo de señal 301 se considera como no tonal (categoría 0 de señal sonora) durante 3 (tres) tramas, incluso la trama perdida.

20 Si la información de un detector de actividad de voz (VAD, por sus siglas en inglés) (no se muestra) está disponible e indica que no hay actividad de voz (presencia de silencio), se fuerza la decisión del clasificador del tipo de señal 301 a la categoría 0 de señal sonora.

Según una alternativa del clasificador del tipo de señal 301, la banda de frecuencia de la mejora permitida y/o el nivel de reducción de ruido intertonal máximo podría ser completamente dinámicos (sin aumento súbito).

25 En el caso de un *lookahead* pequeño, podría ser necesario introducir un ajuste de reducción de aumento mínimo en las primeras bandas críticas para reducir adicionalmente cualquier posible distorsión introducida con la reducción de ruido intertonal. Este ajuste se puede realizar mediante el uso de la siguiente relación:

$$\begin{aligned}
 RedGain_i &= 1.0|_{i \in [0, FEhBand]} ; \\
 RedGain_i &= RedGain_{i-1} \frac{(1.0 - Allow_red)}{(10 - FEhBand)}|_{i \in [FEhBand, 10]} ; \\
 RedGain_i &= Allow_red|_{i \in [0, max_band]}
 \end{aligned}$$

30 donde $RedGain_i$ es un máximo de reducción de aumento por banda, $FEhBand$ es la primera banda donde se permite la reducción de ruido intertonal (que varía típicamente entre 400Hz y 2kHz o las bandas críticas de frecuencia 3 y 12), $Allow_red$ es el nivel de reducción de ruido permitido por categoría de señal sonora presentada en la tabla anterior y max_band es la banda máxima para la reducción de ruido intertonal (17 para la banda estrecha (NB) y 20 para la banda ancha (WB)).

Reducción de ruido intertonal:

35 Se aplica la reducción de ruido intertonal (ver el reductor 108 de ruido de cuantificación (Figura 3)) y la señal sonora decodificada mejorada se reconstruye mediante el uso de una operación de solapar y sumar (véase el operador de solapar y sumar 303 (Figura 3)). La reducción del ruido de cuantificación intertonal se realiza al modificar la escala

del espectro en cada banda crítica de frecuencia con un aumento de escala limitado entre g_{min} y 1 y se deriva de la proporción señal a ruido (SNR, por sus siglas en inglés) en esa banda crítica de frecuencia. Una característica de la técnica de reducción de ruido intertonal es que para las frecuencias menores que una determinada frecuencia, por ejemplo, relacionadas con la sonorización de señal, el procesamiento se realiza basándose en el intervalo de frecuencia y no en la banda crítica de frecuencia. Por lo tanto, se aplica un aumento de escala en cada intervalo de frecuencia derivada de la SNR en ese intervalo (la SNR se computa mediante el uso de la energía del intervalo dividida por la energía del ruido de la banda crítica que incluye ese intervalo). Esta característica tiene el efecto de preservar la energía a frecuencias cerca de los armónicos o tonos, lo cual previene la distorsión a la vez que reduce en gran medida el ruido de cuantificación entre los armónicos. En el caso de señales de banda estrecha, se puede utilizar el análisis por intervalo para el total de espectro. El análisis por intervalo se puede utilizar de forma alternativa en todas las bandas críticas de frecuencia excepto la última.

Haciendo referencia a la Figura 3, la reducción de ruido intertonal se realiza en el reductor 108 de ruido de cuantificación. Según una primera implementación posible, se puede realizar el procesamiento por intervalo en los 115 intervalos de frecuencia en la codificación de banda estrecha (250 intervalos de frecuencia en la codificación de banda ancha) en un atenuador de ruido 304.

En una implementación alternativa, el atenuador de ruido 304 realiza el procesamiento por intervalo para aplicar un aumento de escala a cada intervalo de frecuencia en las primeras bandas K sonoras y entonces el atenuador de ruido 305 realiza el procesamiento por banda para modificar la escala del espectro en cada una de las bandas críticas de frecuencia restantes con un aumento de escala. Si $K=0$, entonces el atenuador de ruido 305 realiza un procesamiento por banda en todas las bandas críticas de frecuencia.

El aumento de escala mínimo g_{min} se deriva de la reducción de ruido intertonal máxima permitida en dB, NR_{max} . Tal como se describió en la descripción que antecede (ver la tabla anterior), el clasificador del tipo de señal 301 realiza la reducción de ruido máxima permitida NR_{max} que varía entre 6 y 12 dB. Por lo tanto, el aumento de escala mínimo está dado por la relación:

$$g_{min} = 10^{-NR_{max} / 20} \quad (9)$$

En el caso de una trama tonal de banda estrecha, el aumento de escala puede computarse en relación con la SNR por intervalo de frecuencia, entonces se realiza la reducción de ruido por intervalo. El procesamiento por intervalo se aplica únicamente a las primeras 17 bandas críticas correspondientes a una frecuencia máxima de 3700 Hz. La cantidad máxima de intervalos de frecuencia en las cuales se puede utilizar el procesamiento por intervalo es de 115 (la cantidad de intervalos en las primeras 17 bandas a 4 kHz).

En el caso de una trama tonal de banda estrecha, el procesamiento por intervalo se aplica a las 21 bandas críticas de frecuencia correspondientes a frecuencia máxima de 8000 Hz. La cantidad máxima de intervalos de frecuencia para las cuales se puede utilizar el procesamiento por intervalo es de 250 (la cantidad de intervalos en las primeras 21 bandas a 8kHz).

En la técnica de reducción de ruido intertonal, la reducción de ruido comienza en la cuarta banda crítica de frecuencia (no se realiza reducción antes de 400 Hz). Para reducir cualquier impacto negativo de la técnica de reducción de ruido de cuantificación intertonal, el clasificador del tipo de señal 301 puede mover la banda crítica de frecuencia inicial hasta la 12ª. Esto significa que la primera banda crítica de fe en la cual se realiza la reducción de ruido intertonal se encuentra entre 400 Hz y 2 kHz y podría variar entre tramas.

El aumento de escala para una determinada banda crítica de frecuencia, o para una determinado intervalo de frecuencia, puede computarse en función de la SNR en esa banda o intervalo de frecuencia mediante el uso de la siguiente relación:

$$(g_s)^2 = k_s SNR + c_s, \text{ limitada por } g_{min} \leq g_s \leq 1 \quad (10)$$

Los valores de k_s y c_s se determinan de manera que $g_s = g_{min}$ para $SNR = 1$ dB, y $g_s = 1$ para $SNR = 45$ dB. Es decir, para SNR de 1 dB y menores, el aumento de escala se limita a g_s y para SNR de 45 dB y mayores, no se realiza reducción de ruido intertonal en la banda crítica de frecuencia dada ($g_s = 1$). Por lo tanto, dados los dos extremos, los valores de k_s y c_s en la Ecuación (10) pueden calcularse mediante el uso de las siguientes relaciones:

$$k_s = (1 - g_{min}^2) / 44 \quad \text{y} \quad c_s = (45g_{min}^2 - 1) / 44. \quad (11)$$

La variable SNR de la Ecuación (10) es ya sea la SNR por banda crítica de frecuencia, $SNR_{CB}(i)$, o la SNR por intervalo de frecuencia, $SNR_{BIN}(k)$, dependiendo del tipo de procesamiento por intervalo o por banda.

La SNR por banda crítica de frecuencia se computa de la siguiente manera:

$$SNR_{CB}(i) = \frac{0.3E_{CB}^{(1)}(i) + 0.7E_{CB}^{(2)}(i)}{N_{CB}(i)} \quad i = 0, \dots, 17 \quad (12)$$

5 donde $E_{CB}^{(1)}(i)$ y $E_{CB}^{(2)}(i)$ indican la energía por banda crítica de frecuencia para el análisis espectral de la trama pasada y actual, respectivamente (tal como se computa en la Ecuación (4)), y $N_{CB}(i)$ indica el estimado de energía de ruido por banda crítica de frecuencia.

La SNR por intervalo de frecuencia en una determinada banda crítica de frecuencia i se computa mediante el uso de la siguiente relación:

$$SNR_{BIN}(k) = \frac{0.3E_{BIN}^{(1)}(k) + 0.7E_{BIN}^{(2)}(k)}{N_{CB}(i)}, \quad k = j_i, \dots, j_i + M_{CB}(i) - 1 \quad (13)$$

10 donde $E_{BIN}^{(1)}(k)$ y $E_{BIN}^{(2)}(k)$ indican la energía por intervalo de frecuencia para el análisis espectral de la trama pasada⁽¹⁾ y actual⁽²⁾, respectivamente (tal como se computa en la Ecuación (5)), $N_{CB}(i)$ indica el estimado de energía de ruido por banda crítica de frecuencia, j_i es el índice del primer intervalo de frecuencia en la i .ª banda crítica de frecuencia, y $M_{CB}(i)$ es la cantidad de intervalos de frecuencia en la banda crítica de frecuencia i como se definió anteriormente en la presente.

15 De acuerdo con otra implementación alternativa, el aumento de escala podría computarse en relación con la SNR por banda crítica de frecuencia o por intervalo de frecuencia para las primeras bandas sonoras. Si $K_{VOIC} > 0$ entonces se puede realizar el procesamiento por intervalo en las primeras K_{VOIC} bandas. El procesamiento por banda entonces se puede utilizar para el resto de las bandas. En el caso donde $K_{VOIC} = 0$, se puede utilizar el procesamiento por banda en la totalidad del espectro.

20 En el caso del procesamiento por banda para una banda crítica de frecuencia con índice i , después de determinar el aumento de escala mediante el uso de la Ecuación (10) y la SNR, según se definió en la Ecuación (12) o (13), el cambio de escala real se realiza mediante el uso de un aumento de escala ajustado actualizado en cada análisis espectral por medio de la siguiente relación:

$$g_{CB,LP}(i) = \alpha_{gs} g_{CB,LP}(i) + (1 - \alpha_{gs}) g_s \quad (14)$$

25 De acuerdo con una característica, el factor de ajuste α_{gs} utilizado para ajustar el aumento de escala g_s puede hacerse adaptativo y relacionado de forma inversa con el aumento de escala g_s en sí. Por ejemplo, el factor de ajuste se puede definir mediante $\alpha_{gs} = 1 - g_s$. Por lo tanto, el ajuste es más potente para aumentos g_s menores. Este enfoque previene la distorsión en segmentos con SNR altas precedidos por tramas con SNR baja, tal como es el caso para comienzos sonoros. En el enfoque propuesto, el procedimiento de ajuste es capaz de adaptarse y utilizar rápidamente aumentos de escala ante la aparición, por ejemplo, de un comienzo sonoro.

30 El cambio de escala en una banda crítica de frecuencia se realiza de la siguiente manera:

$$\begin{aligned} X'_R(k + j_i) &= g_{CB,LP}(i) X_R(k + j_i), & \mathbf{y} \\ X'_I(k + j_i) &= g_{CB,LP}(i) X_I(k + j_i), & k = 0, \dots, M_{CB}(i) - 1 \end{aligned} \quad (15)$$

donde j_i es el índice del primer intervalo de frecuencia en la banda crítica de frecuencia i y $M_{CB}(i)$ es la cantidad de intervalos de frecuencia en esa banda crítica de frecuencia.

35 En el caso del procesamiento por intervalo en una banda crítica de frecuencia con índice i , después de determinar el aumento de escala mediante el uso de la Ecuación (10) y la SNR, según se definió en la Ecuación (12) o (13), el cambio de escala real se realiza mediante el uso de un aumento de escala ajustado actualizado en cada análisis espectral de la siguiente manera:

$$g_{BIN,LP}(k) = \alpha_{gs} g_{BIN,LP}(k) + (1 - \alpha_{gs}) g_s \quad (16)$$

donde el factor de ajuste $\alpha_{gs} = 1 - g_s$ es similar a la Ecuación (14).

El ajuste temporal de los aumentos de escala previene oscilaciones de energía audibles, a la vez que controlar el ajuste mediante el uso de α_{gs} previene la distorsión en segmentos de voz con SNR alta precedidos por tramas con SNR baja, como es el caso de comienzos sonoros, por ejemplo.

El cambio de escala en una banda crítica de frecuencia i se realiza entonces de la siguiente manera:

$$\begin{aligned} X'_R(k + j_i) &= g_{BIN,LP}(k + j_i) X_R(k + j_i), \quad Y \\ X'_I(k + j_i) &= g_{BIN,LP}(k + j_i) X_I(k + j_i), \quad k = 0, \dots, M_{CB}(i) - 1 \end{aligned} \quad (17)$$

donde j_i es el índice del primer intervalo de frecuencia en la banda crítica de frecuencia i y $M_{CB}(i)$ es la cantidad de intervalos de frecuencia en esa banda crítica de frecuencia.

Los aumentos de escala ajustados $g_{BIN,LP}(k)$ y $g_{CB,P}(i)$ se establecen inicialmente en 1,0. Cada vez que se procesa una trama sonora no tonal ($music_flag = 0$), el valor de los aumentos de escala ajustados se reinicia a 1,0 para reducir la posible reducción de estos aumentos de escala ajustados en la siguiente trama.

En cada análisis espectral realizado mediante el analizador espectral 105, los aumentos de escala ajustados $g_{CB,LP}(i)$ se actualizan para todas las bandas críticas de frecuencia (incluso para las bandas críticas de frecuencia sonoras procesadas a través de procesamiento por intervalo, en este caso $g_{CB,LP}(i)$ se actualiza con un promedio de $g_{BIN,LP}(k)$ que pertenece a la banda crítica de frecuencia i). De manera similar, los aumentos de escala ajustados $g_{BIN,LP}(k)$ se actualizan en las primeras 17 bandas críticas de frecuencia, es decir, hasta el intervalo de frecuencia 115 en el caso de codificación de banda estrecha (las primeras 21 bandas críticas de frecuencia, es decir, hasta el intervalo de frecuencia 250 en el caso de codificación de banda ancha). Para las bandas críticas de frecuencia procesadas con procesamiento por banda, los aumentos de escala se actualizan al establecer que son iguales a $g_{CB,LP}(i)$ en las primeras 17 (codificación de banda estrecha) o 21 (codificación de banda ancha) bandas críticas de frecuencia.

En el caso de una señal sonora tonal decodificada de baja energía, no se realiza la reducción de ruido intertonal. Se detecta una señal sonora de baja energía al hallar la energía de ruido máxima en todas las bandas críticas de frecuencia, $max(N_{CB}(i))$, $i = 0, \dots, 17$, (17 en el caso de codificación de banda estrecha y 21 en el caso de codificación de banda ancha) y si este valor es menor que un determinado valor, por ejemplo, 15 dB, entonces no se realiza la reducción de ruido intertonal.

En el caso de procesamiento de señales de banda estrecha, la reducción de ruido intertonal se realiza en las primeras 17 bandas críticas de frecuencia (hasta 3680 Hz). Para los restantes 11 intervalos de frecuencia entre 3680 Hz y 4000 Hz, se cambia la escala del espectro mediante el uso del último aumento de escala g_s del intervalo de frecuencia correspondiente a 3680 Hz.

Corrección del aumento espectral

La relación de Parseval muestra que la energía en el dominio temporal es igual a la energía en el dominio de frecuencia. La reducción de la energía del ruido intertonal da como resultado una reducción general de la energía en los dominios de frecuencia y temporal. Una característica adicional es que el reductor 108 de ruido de cuantificación comprende un corrector de aumento por banda 306 para modificar la escala de la energía por banda crítica de frecuencia de una manera tal que la energía en cada banda crítica de frecuencia al finalizar la modificación de escala será cercana a la energía antes de la reducción del ruido intertonal.

Para lograr dicha modificación de escala, no es necesario modificar la escala de todos los intervalos de frecuencia, sino modificar la escala solamente de los intervalos con mayor energía. El corrector de aumento por banda 306 comprende un analizador 401 (Figura 4) que identifica los intervalos con mayor energía antes de la reducción de ruido intertonal a medida que se modifica la escala de los intervalos al modificar un aumento de escala entre]0,8, 1,0] en la fase de reducción de ruido intertonal. Según una alternativa, el analizador 401 puede determinar también la energía por intervalo antes de la reducción del ruido intertonal mediante el uso de, por ejemplo, la Ecuación (5) para identificar los intervalos con mayor energía.

La energía que se elimina del ruido intertonal se moverá a los eventos con mayor energía (correspondientes a los intervalos con mayor energía) de la banda crítica de frecuencia. De esta manera, la muestra musical final tendrá un sonido más claro que al realizar solamente una reducción de ruido intertonal simple, porque la dinámica entre los eventos de mayor energía y el ruido de fondo aumentarán adicionalmente.

La energía espectral de una banda crítica de frecuencia después de la reducción de ruido intertonal se computa de la misma manera que la energía espectral antes de la reducción de ruido intertonal.

$$E_{CB}(i) = \frac{1}{(L_{FFT}/2)^2 M_{CB}(i)} \sum_{k=0}^{M_{CB}(i)-1} (X_R^2(k+j_i) + X_I^2(k+j_i)), \quad i = 0, \dots, 16 \quad (18)$$

5 A este respecto, el corrector de aumento por banda 306 comprende un analizador 402 para determinar la energía espectral por banda antes de la reducción de ruido intertonal mediante el uso de la Ecuación (18), y un analizador 403 para determinar la energía espectral por banda después de la reducción del ruido intertonal mediante el uso de la Ecuación (18).

10 El corrector de aumento por banda 306 comprende además un calculador 404 para determinar un aumento correctivo a medida que ha sido aplicada la proporción entra la energía espectral de una banda crítica de frecuencia antes de la reducción de ruido intertonal y la energía espectral de esta banda crítica de frecuencia después de la reducción de ruido intertonal.

$$G_{corr}(i) = \sqrt{\left(\frac{E_{CB}(i)}{E_{CB'}(i)} \right)}, \quad i = 0, \dots, 16 \quad (19)$$

15 donde E_{CB} es la energía espectral de banda crítica antes de la reducción de ruido intertonal $E_{CB'}$ es la energía espectral de banda crítica de frecuencia después de la reducción de ruido intertonal. La cantidad de bandas críticas de frecuencia cubre el espectro completo desde 17 bandas en codificación de banda estrecha hasta 21 bandas en codificación de banda ancha.

El cambio de escala a lo largo de la banda crítica de frecuencia i se puede llevar a cabo de la siguiente manera:

$$\begin{aligned} & IF (g_{BIN,LP}(k+j_i) > 0.8 \ \& \ i > 4) \\ & \quad X_R^*(k+j_i) = G_{corr}(k+j_i) X_R'(k+j_i), \quad Y \\ & \quad X_I^*(k+j_i) = G_{corr}(k+j_i) X_I'(k+j_i), \quad k = 0, \dots, M_{CB}(i) - 1, \quad (20) \\ & ELSE \\ & \quad X_R''(k+j_i) = X_R'(k+j_i), \quad Y \\ & \quad X_I''(k+j_i) = X_I'(k+j_i), \quad k = 0, \dots, M_{CB}(i) - 1 \end{aligned}$$

20 donde j_i es el índice del primer intervalo de frecuencia en la banda crítica de frecuencia i y $M_{CB}(i)$ es la cantidad de intervalos de frecuencia en esa banda crítica de frecuencia. No se aplica corrección de aumento por debajo de 600 Hz, dado que se asume que la energía espectral a muy baja frecuencia ha sido codificada de forma precisa mediante el códec específico de voz con tasa de bits baja y cualquier aumento de tono interarmónico será audible.

Potenciación del aumento espectral

25 Es posible aumentar de forma adicional la claridad de la muestra musical al aumentar adicionalmente el aumento G_{corr} en bandas críticas de frecuencia donde no se producen muchos eventos de alta energía. Un calculador 405 del corrector de aumento por banda 306 determina la proporción de eventos de alta energía (proporción entre cantidad de intervalos de alta energía sobre la cantidad total de intervalos de frecuencia) por banda crítica de frecuencia de la siguiente manera:

$$\begin{aligned} REv_{CB} &= \frac{NumBin_{max}}{NumBin_{total}} \quad k = 0, \dots, M_{CB}(i-1) \\ NumBin_{max} &= \sum (g_{BIN,LP} > 0.8) \\ NumBin_{total} &= \text{Total de intervalos en una banda crítica} \end{aligned}$$

El calculador 405 computa entonces un factor de corrección adicional para el aumento correctivo mediante el uso de la siguiente fórmula:

$$IF(NumBin_{max} > 0)$$

$$C_F = -0.2778 \cdot REV_{CB} + 1.2778$$

- 5 En un corrector de aumento por banda 406, este nuevo factor de corrección C_F multiplica el aumento correctivo G_{corr} por un valor situado entre [1,0, 1,2778]. Cuando se considera este factor de corrección C_F , el cambio de escala a lo largo de la banda crítica de frecuencia i se convierte en:

$$IF(g_{BIN,LP}(k + j_i) > 0.8 \ \& \ i > 4)$$

$$X_R^*(k + j_i) = G_{corr} \cdot C_F \cdot (k + j_i) X'_R(k + j_i), \quad Y$$

$$X_I^*(k + j_i) = G_{corr} \cdot C_F \cdot (k + j_i) X'_I(k + j_i), \quad k = 0, \dots, M_{CB}(i) - 1$$

ELSE

$$X_R^*(k + j_i) = X'_R(k + j_i), \quad Y$$

$$X_I^*(k + j_i) = X'_I(k + j_i), \quad k = 0, \dots, M_{CB}(i) - 1$$

- 10 En el caso particular de codificación de banda estrecha, el cambio de escala se realiza únicamente en los intervalos de frecuencia a los que se le modificó la escala mediante un aumento de escala entre]0,96, 1,0] en la fase de reducción de ruido intertonal. Usualmente, cuanto mayor sea la tasa de bits, más cerca estará la energía del espectro al nivel de energía deseado. Por este motivo, puede que no se utilice siempre la segunda parte de la corrección de aumento, el factor de corrección de aumento C_F . Finalmente, con una tasa de bits muy elevada, podría ser beneficioso realizar un cambio de escala de aumento solamente en los intervalos de frecuencia que no se modificaron anteriormente (que tengan un aumento de escala de 1,0).

- 15 Reconstrucción de señal sonora sin ruido mejorada

Después de determinar los componentes espectrales con escala modificada 308, $X'_R(k)$ o $X''_R(k)$ and $X'_I(k)$ o $X''_I(k)$, un calculador 307 del analizador inverso y operador de solapar y sumar 110 computa la FFT inversa. La FTT inversa calculada se aplica a los componentes espectrales con escala modificada 308 para obtener una señal sonora decodificada mejorada sometida a función de ventana en el dominio temporal mediante la siguiente relación:

$$x_{w,d}(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{j2\pi \frac{kn}{N}}, \quad n = 0, \dots, L_{FFT} - 1 \quad (21)$$

- 20 La señal se reconstruye entonces en el operador 303 mediante el uso de una operación de solapar y sumar para las partes superpuestas del análisis. Dado que se utiliza una ventana sinusoidal sobre la señal sonora tonal decodificada original 103 antes del análisis espectral en el analizador espectral 105, se aplica la misma función de ventana a la señal sonora tonal decodificada mejorada sometida a función de ventana 309 en la salida del calculador de FFT inversa antes de la operación de solapar y sumar. Por lo tanto, la señal sonora tonal decodificada mejorada sometida a una función de ventana doble está proporcionada por la relación:
- 25

$$x_{ww,d}^{(1)}(n) = w_{FFT}(n) x_{w,d}^{(1)}(n), \quad n = 0, \dots, L_{FFT} - 1$$

(22)

Para el primer tercio de la ventana de análisis de banda estrecha, la operación de solapar y sumar para construir la señal sonora mejorada se realiza mediante el uso de la relación:

$$s(n) = x_{ww,d}^{(0)}\left(n + 2 \cdot \frac{L_{window}}{3}\right) + x_{ww,d}^{(1)}(n), \quad n = 0, \dots, L_{window}/3 - 1 \quad (23)$$

y para el primer noveno de la ventana de análisis de banda ancha, la operación de solapar y sumar para construir la señal sonora tonal decodificada mejorada se realiza de la siguiente manera:

$$s(n) = x_{ww,d}^{(0)}\left(n + 2 \cdot \frac{L_{window_{wa}}}{9}\right) + x_{ww,d}^{(1)}(n), \quad n = 0, \dots, L_{window_{wa}}/9 - 1$$

5 donde $x_{ww,d}^{(0)}(n)$ es la señal sonora tonal decodificada mejorada sometida a función de ventana doble a partir del análisis de la trama anterior.

Mediante el uso de la operación de solapar y sumar, dado que hay un desplazamiento de 80 muestras (40 en el caso de codificación de banda estrecha) entre la trama decodificadora de señal sonora y la trama de reducción de ruido intertonal, la señal sonora decodificada mejorada puede reconstruirse hasta 80 muestras desde el *lookahead*, además de la presente trama de reducción de ruido intertonal.

Después de la operación de solapar y sumar para reconstruir la señal sonora tonal decodificada mejorada, se realiza la atenuación en el postprocesador 112 sobre la señal sonora decodificada mejorada mediante el uso de la inversa del filtro de preénfasis descrito anteriormente. EL postprocesador 112, por lo tanto, comprende un filtro de atenuación que, en esta realización, está proporcionado por la relación:

$$H_{de-emph}(z) = 1/(1 - 0.68z^{-1}) \quad (24)$$

Actualización de energía de ruido intertonal

Los estimados de energía de ruido intertonal por banda crítica de frecuencia para la reducción de ruido intertonal pueden calcularse para cada trama en un estimador de energía de ruido intertonal (no se muestra) mediante el uso de, por ejemplo, la siguiente fórmula:

$$N_{cr}^0(i) = \frac{(0.6 \cdot E_{cr}^0(i) + 0.2 \cdot E_{cr}^1(i) + 0.2 \cdot N_{cr}^1(i))}{16.0}, \quad i = 0, \dots, 16 \quad (25)$$

donde N_{cr}^0 y E_{cr}^0 representan las energías espectral y de ruido actuales para la banda crítica de frecuencia (i) especificada y N_{cr}^1 y E_{cr}^1 representan las energías espectral y de ruido para la trama pasada de la misma banda crítica de frecuencia.

Este método para calcular los estimados de energía de ruido intertonal por banda crítica de frecuencia es simple y podría introducir algunas distorsiones en la señal sonora tonal decodificada mejorada. Sin embargo, en codificación de banda estrecha de tasa de bits baja, estas distorsiones se compensan en gran medida por la mejora en la claridad la síntesis de las señales sonoras.

En la codificación de banda ancha, cuando está presente el ruido intertonal, pero no es tan molesto, el método para actualizar la energía de ruido intertonal tiene que ser más sofisticada para prevenir la introducción de distorsión molesta. Se podría utilizar una técnica diferente con más o menos complejidad computacional.

Actualización de la energía de ruido intertonal mediante el uso del promedio ponderado de la energía por banda:

Según esta técnica, los segundos valores de energía máximos y mínimos de cada banda crítica de frecuencia se utilizan para computar un umbral de energía por banda crítica de frecuencia de la siguiente manera:

$$thr_ener_{CB}(i) = 1.85 \cdot \left(\frac{\max_2(E_{CB}^0(i)) + \min(E_{CB}^0(i))}{2} \right), \quad i = 0, \dots, 20$$

donde \max_2 representa el intervalo de frecuencia que tiene el segundo valor de energía máximo y \min es el intervalo de frecuencia que tiene el valor de energía mínimo en la banda crítica de frecuencia de interés.

- 5 El umbral de energía (thr_ener_{CB}) se utiliza para computar una primera estimación del nivel de ruido intertonal por banda crítica (tmp_ener_{CB}) que corresponde al promedio de las energías (E_{BIN}) de todos los intervalos de energía posteriores al umbral de energía dentro de la banda crítica de frecuencia, mediante el uso de la siguiente relación:

```

mcnt = 0
tmp_ener_CB(i) = 0
for (k = 0 : M_CB(i))
    if (E_BIN(k) < thr_ener_CB)
        tmp_ener_CB(i) = tmp_ener_CB(i) + E_BIN(k)
        mcnt = mcnt + 1
    endif
endfor

tmp_ener_CB(i) = tmp_ener_CB(i) / mcnt

```

- 10 donde $mcnt$ es la cantidad de intervalos de frecuencia de los cuales se incluyen las energías (E_{BIN}) en la suma y $mcnt \leq M_{CB}(i)$. Además, la cantidad $mcnt$ de intervalos de frecuencia de los cuales la energía (E_{BIN}) se encuentra por debajo del umbral de energía se compara con la cantidad de intervalos de energía (M_{CB}) dentro de una banda crítica de frecuencia para evaluar la proporción de intervalos de frecuencia por debajo del umbral de energía. Esta proporción $accepted_ratio_{CB}$ se utiliza para ponderar la primera estimación del nivel de ruido intertonal hallada previamente (tmp_ener_{CB}).

$$accepted_ratio_{CB}(i) = \frac{mcnt}{M_{CB}(i)}, \quad i = 0, \dots, 20$$

- 15 Un factor de ponderación β_{CB} de la estimación del nivel de ruido intertonal es diferente entre la tasa de bits utilizada y la $accepted_ratio_{CB}$. Un valor elevado de $accepted_ratio_{CB}$ para una banda crítica de frecuencia significa que será difícil diferenciar entre la energía de ruido y la energía de señal. En este caso es deseable no reducir demasiado el nivel de ruido de esta banda crítica de frecuencia para no arriesgarse a producir ninguna alteración de la energía de señal. Pero un valor bajo de $accepted_ratio_{CB}$ indica una gran diferencia entre los niveles de energía de señal y ruido, entonces el nivel de ruido estimado podría ser mayor en esta banda crítica de frecuencia sin agregar distorsión. El factor β_{CB} se modifica de la siguiente manera:
- 20

$IF((accepted_ratio(i)<0.6|accepted_ratio(i-1)<0.5) \& i > 9)$
 $\beta_{CB}(i) = 1$
 $ELSE IF(accepted_ratio(i)<0.75 \& i > 15)$
 $\beta_{CB}(i) = 2$
 $ELSE IF\left(\left(\begin{matrix} accepted_ratio(i)>0.85 & \& \\ accepted_ratio(i-1)>0.85 & \& \\ accepted_ratio(i-2)>0.85 \end{matrix}\right) \& bitrate > 16000\right), \quad i = 0, \dots, 20$
 $\beta_{CB}(i) = 30$
 $ELSE IF(bitrate > 16000)$
 $\beta_{CB}(i) = 20$
 $ELSE$
 $\beta_{CB}(i) = 16$

Finalmente, la estimación de ruido intertonal por banda crítica de frecuencia puede ajustarse de forma diferente si aumenta o disminuye el ruido intertonal.

Disminuye el ruido:

Disminuye el ruido:

$$N_{CB}^0(i) = (1 - \alpha) \left(\frac{imp_ener_{CB}(i)}{\beta_{CB}(i)} \right) + \alpha \cdot N^1(i)$$

Aumenta el ruido: $i = 0, \dots, 20$

$$N_{CB}^0(i) = (1 - \alpha_2) \left(\frac{imp_ener_{CB}(i)}{\beta_{CB}(i)} \right) + \alpha_2 \cdot N^1(i)$$

Donde

$$\alpha = 0.1$$

$$\alpha_2 = \begin{cases} 0.98 & \text{para tasa de bits} > 16000 \text{ bps} \\ 0.95 & \text{de otro modo} \end{cases}$$

5

donde N_{CB}^0 representa la energía de ruido actual para la banda crítica de frecuencia (i) especificada y N_{CB}^1 representa la energía de ruido de la trama pasada de la misma banda crítica de frecuencia.

Aunque la presente invención se ha descrito en la descripción que antecede a modo de realizaciones ilustrativas no restrictivas de esta, son posibles muchas otras modificaciones y variaciones dentro del alcance de las reivindicaciones adjuntas.

10

REFERENCIAS

- [1] 3GPP TS 26.190, "Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; Transcoding functions".
- [2] J. D. Johnston, "Transform coding of audio signal using perceptual noise criteria," *IEEE J. Select. Areas Commun.*, vol. 6, pp. 314--323, Feb. 1988.

REIVINDICACIONES

1. Un método (100) para mejorar una señal sonora decodificada, que comprende:

analizar espectralmente (105) la señal sonora tonal decodificada para producir parámetros espectrales (107) representativos de la señal sonora tonal decodificada, donde analizar espectralmente (105) la señal sonora tonal decodificada comprende dividir un espectro resultante del análisis espectral en un conjunto de bandas críticas de frecuencia, donde cada una comprende una cantidad de intervalos de frecuencia;

reducir (108) un ruido de cuantificación en regiones espectrales de baja energía de la señal sonora tonal decodificada en respuesta a los parámetros espectrales (107) del análisis espectral, donde reducir (108) el ruido de cuantificación comprende cambiar la escala (108, 304, 305, 306) del espectro de la señal sonora tonal decodificada por banda crítica de frecuencia, por intervalo de frecuencia o tanto por banda crítica de frecuencia como por intervalo de frecuencia;

realizar una clasificación del tipo de señal que comprende:

determinar (501) (a) un promedio \bar{E}_{diff} de variaciones de una energía espectral por trama total para 40 tramas pasadas de la señal sonora decodificada mediante el uso de la relación

$$\bar{E}_{diff} = \frac{\left(\sum_{t=-40}^{t=-1} \Delta_E^t \right)}{40}, \text{ donde } \Delta_E^t = E_{fr}^t - E_{fr}^{(t-1)}$$

donde E_{fr}^t es la energía espectral de trama total para una trama actual t y $E_{fr}^{(t-1)}$ es la energía espectral de trama total para una trama anterior $(t-1)$, y (b) una desviación estadística σ_E de la variación de energía en las últimas 15 tramas de la señal sonora decodificada mediante el uso de la relación

$$\sigma_E = 0.7745967 \cdot \sqrt{\sum_{t=-15}^{t=-1} \frac{\left(\Delta_E^t - \bar{E}_{diff} \right)^2}{15}}$$

almacenar el promedio \bar{E}_{diff} y la desviación estadística σ_E en una memoria (50);

comparar (503-506), mediante un primer a cuarto comparador, la desviación estadística σ_E con cuatro umbrales flotantes que incluyen el umbral 1, umbral 2, umbral 3 y umbral 4, para clasificar la señal sonora decodificada en la categoría 0 de señal sonora, categoría 1 de señal sonora, categoría 2 de señal sonora, categoría 3 de señal sonora y categoría 4 de señal sonora;

contar (512), mediante un primer contador, las tramas de categoría 3 o 4 de señal sonora y aumentar (514) los umbrales flotantes 1 a 4 el valor TH_UP cuando el contador cuente una serie de más de 30 tramas de la categoría 3 o 4 de señal sonora; y

contar (513), mediante un segundo contador, las tramas de categoría 0 de señal sonora y disminuir (514) los umbrales flotantes 1 a 4 el valor TH_DOWN cuando el contador cuente una serie de más de 30 tramas de la categoría 0 de señal sonora,

donde los umbrales 1 a 4 se limitan a valores máximos y mínimos absolutos y donde cada vez que aumente el conteo del primer contador, el segundo contador se reinicia en cero;

caracterizado por que la clasificación del tipo de señal comprende:

- controlar (510), mediante un primer controlador, la reducción de ruido de cuantificación (108) para mejorar la señal sonora ton decodificada dentro de una banda de frecuencia de 2000 a $F_s/2$ Hz al reducir el ruido de cuantificación intertonal una amplitud permitida máxima de 6 dB, cuando (a) se detecte la categoría 1 de señal sonora mediante el primer comparador (506) lo cual muestra una desviación estadística σ_E menor que el umbral 1 y (b) la última categoría de señal sonora detectada era ≥ 0 , donde F_s es una frecuencia de muestreo de la señal sonora decodificada;

- controlar (509), mediante un segundo controlador, la reducción de ruido de cuantificación (108) para mejorar la señal sonora ton decodificada dentro de una banda de frecuencia de 1270 a $F_s/2$ Hz al reducir el ruido de cuantificación intertonal una amplitud permitida máxima de 9 dB, cuando (a) se detecte la categoría 2 de señal sonora mediante el segundo comparador (505) lo cual muestra una desviación estadística σ_E menor que el umbral 2 y (b) la última categoría de señal sonora detectada era ≥ 1 ;

5 - controlar (508), mediante un tercer controlador, la reducción de ruido de cuantificación (108) para mejorar la señal sonora ton decodificada dentro de una banda de frecuencia de 700 a $F_s/2$ Hz al reducir el ruido de cuantificación intertonal una amplitud permitida máxima de 12 dB, cuando (a) se detecte la categoría 3 de señal sonora mediante el tercer comparador (504) lo cual muestra una desviación estadística σ_E menor que el umbral 3 y (b) la última categoría de señal sonora detectada era ≥ 2 ;

10 - controlar (507), mediante un cuarto controlador, la reducción de ruido de cuantificación (108) para mejorar la señal sonora ton decodificada dentro de una banda de frecuencia de 400 a $F_s/2$ Hz al reducir el ruido de cuantificación intertonal una amplitud permitida máxima de 12 dB, cuando (a) se detecte la categoría 4 de señal sonora mediante el cuarto comparador (503) lo cual muestra una desviación estadística σ_E menor que el umbral 4 y (b) la última categoría de señal sonora detectada era ≥ 3 ; y

- controlar (511), mediante un quinto controlador, la reducción de ruido de cuantificación (108) para no reducir el ruido de cuantificación intertonal cuando se detecte la categoría 0 de señal sonora, cuando la detección de las categorías 1 a 4 de señal sonora mediante el primer al cuarto comparador sea negativa.

15 2. Un sistema (100) para mejorar una señal sonora tonal decodificada que comprende: un analizador espectral (105) de la señal sonora tonal decodificada adaptado para producir parámetros espectrales (107) representativos de la señal sonora tonal decodificada, donde el analizador espectral (105) está adaptado para dividir un espectro resultante del análisis espectral en un conjunto de bandas críticas de frecuencia, y donde cada banda crítica de frecuencia comprende una cantidad de intervalos de frecuencia;

20 un reductor (108) de ruido de cuantificación en regiones espectrales de baja energía de la señal sonora tonal decodificada mediante el uso de los parámetros espectrales (107) del analizador espectral (105), donde el reductor (108) de ruido de cuantificación comprende un atenuador de ruido (108, 304, 305, 306) que está adaptado para cambiar la escala del espectro de la señal sonora tonal decodificada por banda crítica de frecuencia, por intervalo de frecuencia o tanto por banda crítica de frecuencia como por intervalo de frecuencia; y

un clasificador del tipo de señal (301) que comprende:

25 - un buscador (501) para determinar (a) un promedio \bar{E}_{diff} de variaciones de una energía espectral por trama total para 40 tramas pasadas de la señal sonora decodificada mediante el uso de la relación

$$\bar{E}_{diff} = \frac{\left(\sum_{t=-40}^{t=-1} \Delta_E^t \right)}{40}, \text{ donde } \Delta_E^t = E_{fr}^t - E_{fr}^{(t-1)}$$

30 donde E_{fr}^t es la energía espectral de trama total para una trama actual t y $E_{fr}^{(t-1)}$ es la energía espectral de trama total para una trama anterior $(t-1)$, y (b) una desviación estadística σ_E de la variación de energía en las últimas 15 tramas de la señal sonora decodificada mediante el uso de la relación

$$\sigma_E = 0.7745967 \cdot \sqrt{\sum_{t=-15}^{t=-1} \frac{(\Delta_E^t - \bar{E}_{diff})^2}{15}}$$

- una memoria (502) adaptada para actualizarse con el promedio \bar{E}_{diff} y la desviación estadística σ_E ;

35 - primer, segundo y tercer comparadores (503-506) para comparar la desviación estadística σ_E con cuatro umbrales flotantes que incluyen el umbral 1, umbral 2, umbral 3 y umbral 4, para clasificar la señal sonora decodificada en la categoría 0 de señal sonora, categoría 1 de señal sonora, categoría 2 de señal sonora, categoría 3 de señal sonora y categoría 4 de señal sonora;

- un primer contador (512) de tramas de categoría 3 o 4 de señal sonora y un controlador (514) adaptado para aumentar los umbrales flotantes 1 a 4 el valor TH_UP cuando el primer contador cuente una serie de más de 30 tramas de la categoría 3 o 4 de señal sonora, y

40 - un segundo contador (513) de tramas de categoría 0 de señal sonora y un controlador (514) adaptado para disminuir los umbrales flotantes 1 a 4 el valor TH_DOWN cuando el segundo contador cuente una serie de más de 30 tramas de la categoría 0 de señal sonora,

donde los umbrales 1 a 4 se limitan a valores máximos y mínimos absolutos y donde cada vez que aumente el conteo del primer contador, el segundo contador se reinicia en cero;

45 caracterizado por que el clasificador del tipo de señal comprende:

- 5 - un primer controlador (510) para instruir al reductor de ruido de cuantificación (108) a que mejore la señal sonora ton decodificada dentro de una banda de frecuencia de 2000 a $F_s/2$ Hz al reducir el ruido de cuantificación intertonal una amplitud permitida máxima de 6 dB, cuando (a) el primer comparador (506) detecte una categoría 1 de señal sonora al detectar una desviación estadística σ_E menor que el umbral 1 y (b) la última categoría de señal sonora detectada era ≥ 0 , donde F_s es una frecuencia de muestreo de la señal sonora decodificada;
- 10 - un segundo controlador (509) para instruir al reductor de ruido de cuantificación (108) a que mejore la señal sonora ton decodificada dentro de una banda de frecuencia de 1270 a $F_s/2$ Hz al reducir el ruido de cuantificación intertonal una amplitud permitida máxima de 9 dB, cuando (a) el segundo comparador (505) detecte una categoría 2 de señal sonora al detectar una desviación estadística σ_E menor que el umbral 2 y (b) la última categoría de señal sonora detectada era ≥ 1 ;
- 15 - un tercer controlador (508) para instruir al reductor de ruido de cuantificación (108) a que mejore la señal sonora ton decodificada dentro de una banda de frecuencia de 700 a $F_s/2$ Hz al reducir el ruido de cuantificación intertonal una amplitud permitida máxima de 12 dB, cuando (a) el tercer comparador (504) detecte una categoría 3 de señal sonora al detectar una desviación estadística σ_E menor que el umbral 3 y (b) la última categoría de señal sonora detectada era ≥ 2 ;
- 20 - un cuarto controlador (507) para instruir al reductor de ruido de cuantificación (108) a que mejore la señal sonora ton decodificada dentro de una banda de frecuencia de 400 a $F_s/2$ Hz al reducir el ruido de cuantificación intertonal una amplitud permitida máxima de 12 dB, cuando (a) el cuarto comparador (503) detecte una categoría 4 de señal sonora al detectar una desviación estadística σ_E menor que el umbral 4 y (b) la última categoría de señal sonora detectada era ≥ 3 ; y
- un quinto controlador (511) para instruir al reductor de ruido de cuantificación (108) a que no reduzca el ruido de cuantificación intertonal cuando se detecte la categoría 0 de señal sonora, cuando la detección de las categorías 1 a 4 de señal sonora mediante el primer al cuarto comparador sea negativa.

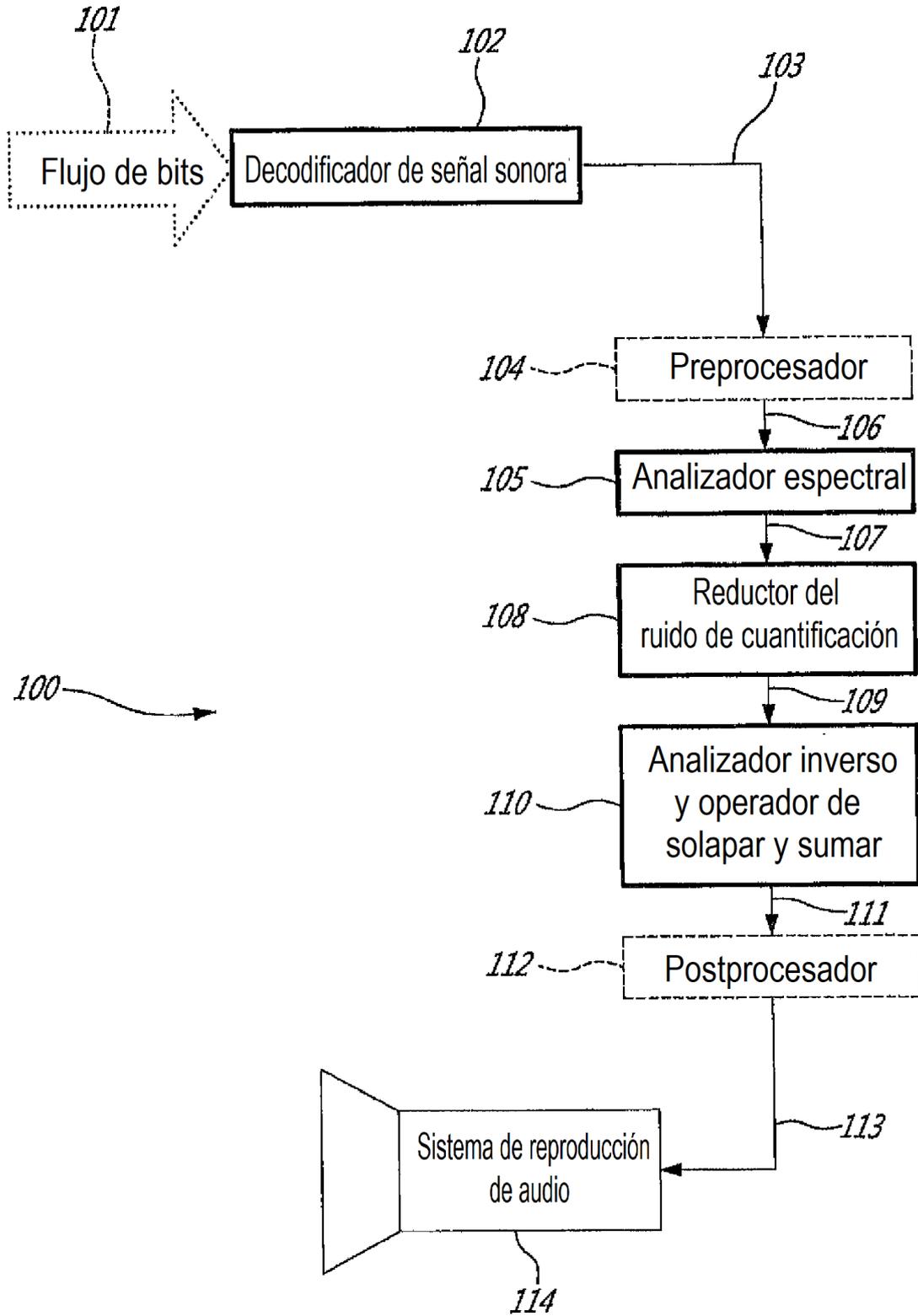


Fig. 1

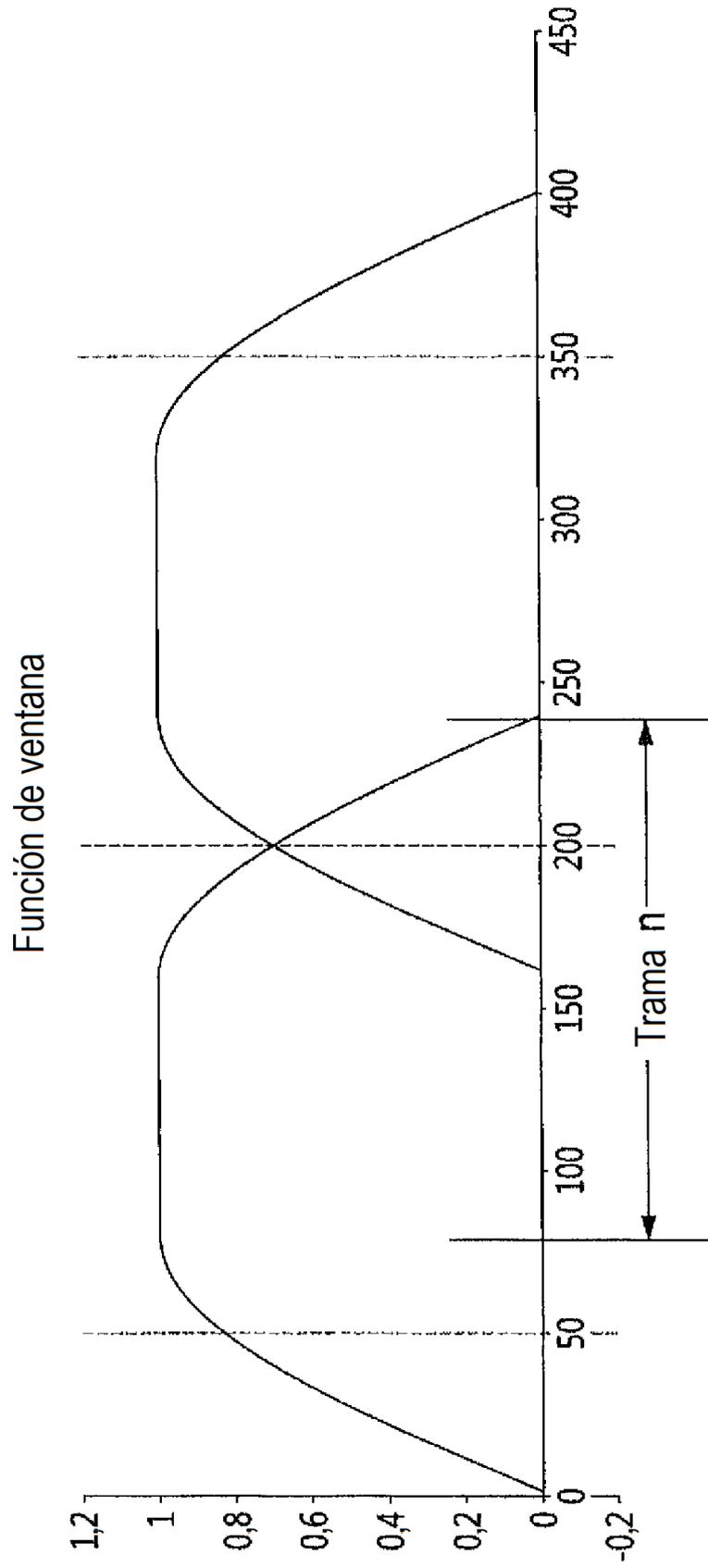


FIG. 2

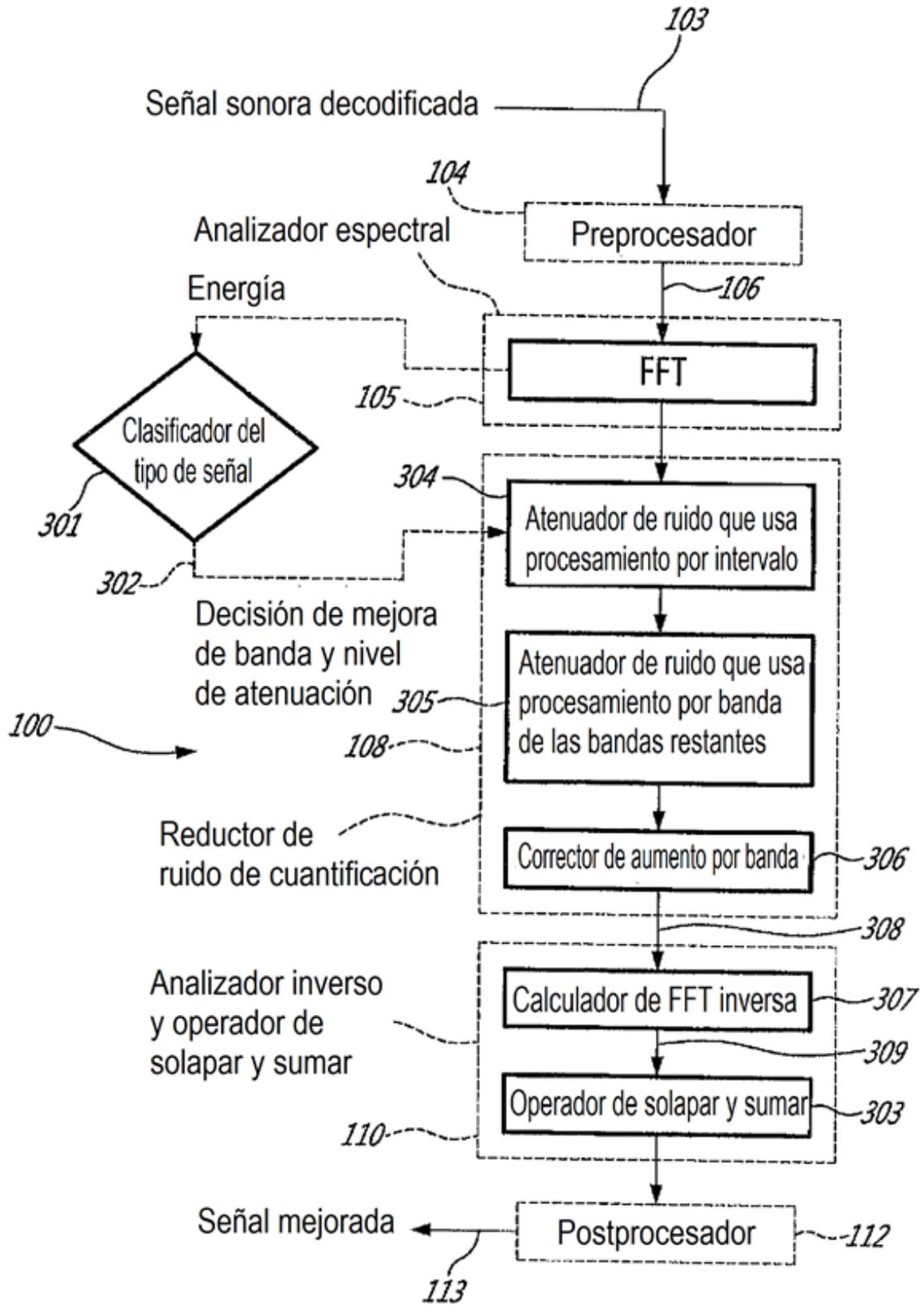


FIG. 3

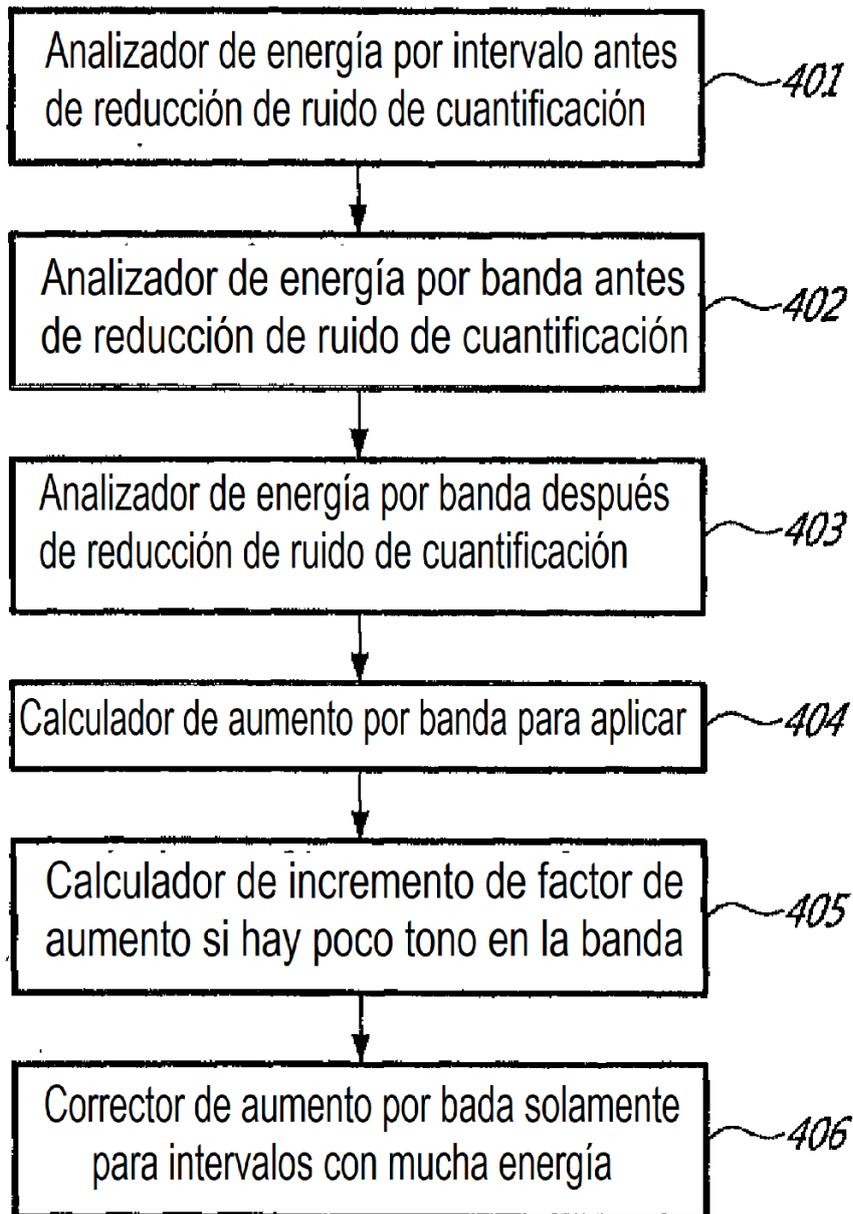
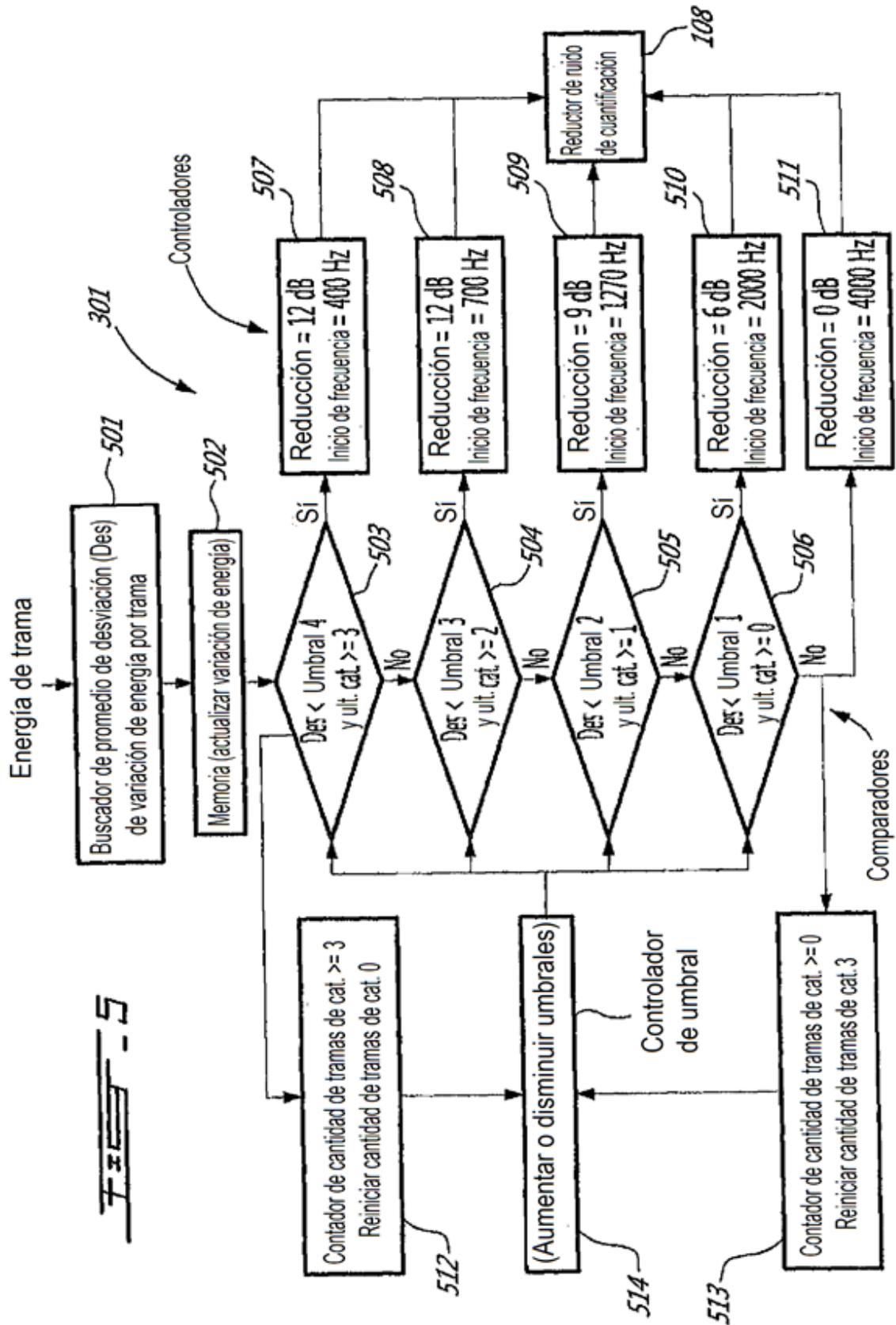


FIG. 4



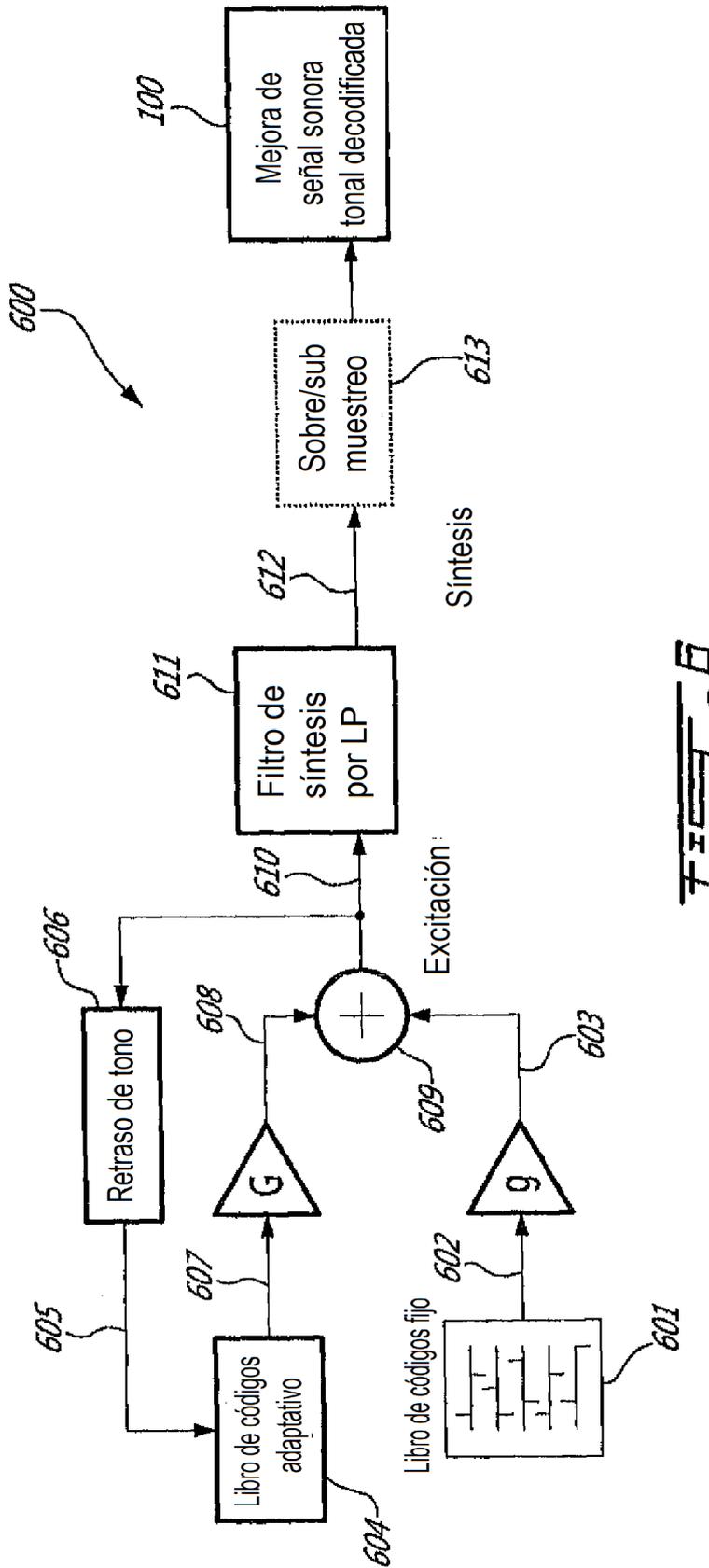


FIG. 6