(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 805 308**

(51) Int. CI.:

G10L 19/20 (2013.01)
G10L 19/26 (2013.01)
G10L 19/08 (2013.01)
G10L 19/22 (2013.01)
G10L 25/78 (2013.01)
G10L 25/81 (2013.01)
G10L 25/93 (2013.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **01.11.2012 PCT/CA2012/001011**

(87) Fecha y número de publicación internacional: **10.05.2013 WO13063688**

(96) Fecha de presentación y número de la solicitud europea: **01.11.2012 E 12844916 (2)**

(97) Fecha y número de publicación de la concesión europea: **17.06.2020 EP 2774145**

(54) Título: **Mejora del contenido insonoro para decodificador CELP de tasa baja**

(30) Prioridad:

03.11.2011 US 201161555246 P

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
11.02.2021

(73) Titular/es:

VOICEAGE EVS LLC (100.0%)
620 Newport Center Drive, Suite 1100
Newport Beach, CA 92660, US

(72) Inventor/es:

VAILLANCOURT, TOMMY y
JELINEK, MILAN

(74) Agente/Representante:

SÁEZ MAESO, Ana

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

Mejora del contenido insonoro para decodificador CELP de tasa baja

Campo

La presente divulgación se refiere a un dispositivo y método para modificar una síntesis de una excitación de dominio de tiempo decodificada por un decodificador de dominio de tiempo.

Antecedentes

Un códec conversacional de última generación puede representar con una muy buena calidad una señal sonora limpia con una tasa de bits de alrededor de 8 kbps y una transparencia de aproximación a una tasa de bits de 16 kbps. Para mantener esta alta calidad sonora incluso a tasa baja de bits, se puede utilizar un esquema de codificación multimodal. Usualmente (véase, por ejemplo, el documento US 6,240,386 B1) la señal de sonido de entrada se divide entre diferentes categorías que reflejan sus características. Por ejemplo, las diferentes categorías pueden incluir de voz, sorda y de aparición. El códec utiliza diferentes modos de codificación optimizados para todas estas categorías.

Sin embargo, algunos códecs sonoros desplegados no utilizan este enfoque multimodal, lo que da como resultado una calidad subóptima, especialmente a tasas de bits bajas para una señal de sonido diferente de sonoridad limpia. Cuando se despliega un códec, es difícil modificar el codificador debido al hecho de que el flujo de bits está estandarizado y cualquier modificación al flujo de bits rompería la interoperabilidad del códec. Sin embargo, se pueden implementar modificaciones en el decodificador para mejorar la calidad percibida en el lado del receptor.

Resumen

De acuerdo con un primer aspecto, la presente divulgación se refiere a un dispositivo para modificar, durante la decodificación de una señal de sonido, una síntesis de una excitación de dominio de tiempo decodificada por un decodificador de dominio de tiempo, que comprende: un clasificador de la síntesis de excitación de dominio del tiempo decodificado en uno de un número de categorías; un convertidor de la excitación de dominio de tiempo decodificada en una excitación de dominio de frecuencia; un modificador de la excitación de dominio de frecuencia en función de la categoría en la que el clasificador clasifica la síntesis de la excitación de dominio de tiempo decodificada; un convertidor de la excitación de dominio de frecuencia modificada en una excitación de dominio del tiempo modificado; y un filtro de síntesis suministrado con la excitación de dominio del tiempo modificado para producir una síntesis modificada de la excitación de dominio del tiempo decodificada. El modificador puede comprender una calculadora de una frecuencia de corte donde una contribución de excitación de dominio de tiempo deja de usarse; un cero de la excitación del dominio de frecuencia por encima de la frecuencia de corte; y un normalizador de la excitación de dominio de frecuencia por debajo de la frecuencia de corte para producir una excitación de dominio de frecuencia normalizada.

De acuerdo con un segundo aspecto, la presente divulgación se refiere a un método para modificar, durante la decodificación de una señal de sonido, una síntesis de una excitación de dominio de tiempo decodificada por un decodificador de dominio de tiempo, que comprende: clasificar la síntesis de la excitación de dominio de tiempo decodificada en uno de un número de categorías; convertir la excitación de dominio de tiempo decodificada en una excitación de dominio de frecuencia; modificar la excitación de dominio de frecuencia en función de la categoría en la que se clasifica la síntesis de la excitación de dominio de tiempo decodificada; convertir la excitación de dominio de frecuencia modificada en una excitación de dominio del tiempo modificada; y sintetizar la excitación de dominio de tiempo modificada para producir una síntesis modificada de la excitación de dominio de tiempo decodificada. La modificación de la excitación de dominio de frecuencia puede comprender el cálculo de una frecuencia de corte donde una contribución de excitación de dominio de tiempo deja de usarse; poner a cero la excitación del dominio de frecuencia por encima de la frecuencia de corte; y normalizar la excitación de dominio de frecuencia por debajo de la frecuencia de corte para producir una excitación de dominio de frecuencia normalizada.

Las características anteriores y otras del dispositivo y método para modificar la síntesis de una excitación de dominio de tiempo se harán más evidentes al leer la siguiente descripción no restrictiva, dada a modo de ejemplo no limitativo con referencia a los dibujos acompañantes.

Breve descripción de los dibujos

En los dibujos adjuntos:

La figura 1 es un diagrama esquemático simplificado que muestra la modificación de un decodificador CELP para la mejora de tramas sordas inactivas y activas;

La figura 2 es un diagrama esquemático detallado que muestra la modificación del decodificador CELP para la mejora de tramas sordas inactivas y activas;

La figura 3 es un diagrama esquemático simplificado que muestra la modificación de un decodificador CELP para mejorar las tramas de audio genéricas; y

La figura 4 es un diagrama esquemático detallado que muestra la modificación del decodificador CELP para la mejora de tramas de audio genéricas.

Descripción

La presente divulgación se refiere a un enfoque para implementar en el lado del decodificador una decodificación multimodal de modo que se mantenga la interoperabilidad y se incremente la calidad percibida. En la divulgación, aunque AMR-WB como se describe en la referencia [3GPP TS 26.190, "Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; Transcoding functions] se utiliza como ejemplo ilustrativo, debe tenerse en cuenta que este enfoque también se puede aplicar a otros tipos de decodificadores sonoros de tasa baja de bits.

En referencia a la figura 1, para lograr esta decodificación multimodal, un decodificador 102 de excitación de dominio de tiempo primero decodifica completamente el flujo 101 de bits recibido, por ejemplo el flujo de bits AMR-WB, para obtener una excitación decodificada de predicción lineal con excitación de código (CELP) completa de dominio de tiempo. La excitación de dominio de tiempo decodificada se procesa a través de un filtro 103 de síntesis de predicción lineal (LP) para obtener una síntesis de dominio de tiempo de la señal de sonido/sonora a la frecuencia de muestreo interna del decodificador. Para AMR-WB, esta frecuencia de muestreo interna es de 12.8 kHz, pero para otro códec podría ser diferente.

La síntesis de dominio de tiempo de la trama actual del filtro 103 de síntesis LP se procesa a través de un clasificador 104-105-106-301 (figuras 1, 2 y 3) suministrado con la información 109 de detección de actividad de voz (VAD) del flujo 101 de bits. El clasificador 104-105-106-301 analiza y categoriza la síntesis de dominio de tiempo bien sea como sonoridad inactiva, sonoridad de voz activa, sonoridad sorda activa, o audio genérico. La sonoridad inactiva (detectada en 1051) incluye todos los ruidos de fondo entre la ráfaga sonora, la sonoridad de voz activa (detectada en 1061) representa una trama durante una ráfaga sonora activa que tiene características de voz, la sonoridad sorda activa (detectada en 1062) representa una trama durante una ráfaga sonora que tiene características sordas, y el audio genérico (detectado en 3010) representa música o sonoridad reverberante. Se pueden agregar o derivar otras categorías de las categorías anteriores. El enfoque divulgado apunta a mejorar en particular, pero no exclusivamente, la calidad percibida de la sonoridad inactiva, la sonoridad sorda activa y el audio genérico.

Una vez que se determina la categoría de la síntesis de dominio de tiempo, un convertidor/modificador 107 convierte la excitación decodificada del decodificador 102 de excitación de dominio de tiempo en dominio de frecuencia usando una transformación de frecuencia sin superposición. También se puede usar una transformación de superposición, pero implica un aumento del retraso de extremo a extremo que no es deseable en la mayoría de los casos. La representación de frecuencia de la excitación se divide en diferentes bandas de frecuencia en el convertidor/modificador 107. Las bandas de frecuencia pueden tener un tamaño fijo, pueden basarse en bandas críticas [J. D. Johnston, "Transform coding of audio signal using perceptual noise criteria," IEEE J. Select. Areas Commun., vol. 6, págs. 314-323, febrero de 1988], o cualquier otra combinación. Luego, la energía por banda se calcula y se guarda en la memoria en el convertidor/modificador 107 para su uso después del proceso de remodelación para garantizar que la modificación no altera el nivel de energía global de la trama.

La modificación de la excitación de dominio de frecuencia realizada por el convertidor/modificador 107 puede diferir con la clasificación de la síntesis. Para la sonoridad inactiva y la sonoridad sorda activa, la remodelación puede consistir en una normalización de las bajas frecuencias con una adición de ruido y la sustitución del contenido de alta frecuencia con solo ruido. Una frecuencia de corte de la síntesis de dominio del tiempo decodificada, el límite entre baja y alta frecuencia, se puede fijar en un valor de alrededor de 1 a 1.2 kHz. Parte del contenido de baja frecuencia de la síntesis decodificada de dominio de tiempo se mantiene para evitar artefactos cuando se conmuta entre una trama no modificada y una trama modificada. También es posible hacer que la frecuencia de corte sea variable de trama a trama eligiendo un segmento de frecuencia en función del tono decodificado del decodificador 102 de excitación de dominio de tiempo. El proceso de modificación tiene el efecto de eliminar el tipo de ruido eléctrico asociado con el códec sonoro de tasa baja de bits. Después del proceso de modificación, se aplica una ganancia de emparejamiento por banda de frecuencia para recuperar el nivel de energía inicial por banda de frecuencia con un ligero aumento de la energía para las frecuencias superiores a 6 kHz para compensar una caída de ganancia del filtro LP en esas frecuencias.

Para una trama categorizada como audio genérico, el procesamiento en el convertidor/modificador 107 es diferente. Primero, la normalización se realiza por banda de frecuencia para todas las bandas. En la operación de normalización, todos los segmentos dentro de una banda de frecuencia que están por debajo de una fracción del valor de frecuencia máximo dentro de la banda se establecen en cero. Para bandas de frecuencia más altas, se ponen a cero más segmentos por banda. Esto simula un esquema de cuantificación de frecuencia con un presupuesto de bits alto, pero con más bits asignados a las frecuencias más bajas. Después del proceso de normalización, se puede aplicar un relleno de ruido para reemplazar los segmentos a cero con ruido aleatorio, pero, dependiendo de la tasa de bits, el relleno de ruido no siempre se usa. Después del proceso de modificación, se aplica una ganancia de emparejamiento por banda de frecuencia para recuperar el nivel de energía inicial por banda de frecuencia y se aplica una corrección de inclinación según la tasa de bits a lo largo de la banda de frecuencia para compensar la subestimación sistemática del filtro LP en caso de entrada de audio genérica. Otra diferenciación para el trayecto de audio genérico proviene del hecho de que la ganancia de emparejamiento no se aplica a todos los segmentos de frecuencia. Debido a que el

espectro del audio genérico es usualmente más intenso que el de la sonoridad, la calidad percibida mejora cuando es posible identificar pulsos espectrales y poner cierto énfasis al respecto. Para hacerlo, la ganancia de emparejamiento completa con la corrección de inclinación se aplica solo a los segmentos de energía más altos dentro de una banda de frecuencia. Para los segmentos de energía más bajos, solo se aplica una fracción de la ganancia de emparejamiento a esos segmentos. Esto da como resultado un aumento de la dinámica espectral.

Después de la remodelación de la frecuencia de excitación y la ganancia de emparejamiento, el convertidor/modificador 107 aplica una transformación de frecuencia inversa para obtener la excitación de dominio de tiempo modificada. Esta excitación modificada se procesa a través del filtro 108 de síntesis LP para obtener una síntesis de dominio del tiempo modificada. Un sobrescritor 110 simplemente sobrescribe la síntesis decodificada de dominio de tiempo del filtro 103 de síntesis LP con la síntesis modificada de dominio de tiempo del filtro 108 de síntesis LP dependiendo de la clasificación de la síntesis decodificada de dominio de tiempo antes de la última desacentuación y remuestreo a 16 kHz (para el ejemplo de AMR-WB) en un filtro y remuestreador 112 que enfatiza el énfasis.

En caso de sonoridad inactiva, la única diferencia en comparación con la modificación de sonoridad sorda activa es el uso de un suavizador 111 para suavizar el filtro 108 de síntesis LP para proporcionar una variación de ruido más suave. Las modificaciones restantes son las mismas que para el trayecto sordo activo. En el siguiente texto se describe un ejemplo más detallado de implementación del enfoque divulgado con referencia a la figura 2.

1) Clasificación de señal

Con referencia a la figura 2, el clasificador 104-105-106-301 realiza en el decodificador una clasificación de la síntesis 1021 de dominio de tiempo de la señal de sonido/sonora como se describe anteriormente para las tasas de bits donde se aplica la modificación. Para simplificar los dibujos, el filtro 103 de síntesis LP no se muestra en la figura 2. La clasificación en el decodificador es similar a la descrita en las referencias [Milan Jelinek y Philippe Gournay; solicitud de patente PCT WO03102921A1, "A method and device for efficient frame erasure concealment in linear predictive based speech codecs"] y [T.Vaillancourt et al., solicitud de patente PCT WO2007073604A1, "Method and device for efficient frame erasure concealment in speech codecs"], más algunas adaptaciones para la detección de audio genérico. Los siguientes parámetros se utilizan para la clasificación de las tramas en el decodificador: una correlación normalizada r_x, una medida de inclinación espectral e_t, un contador de estabilidad de tono pc, una energía de trama relativa de la señal de sonido al final de la trama actual E_s, y un contador de cruce por cero zc. El cálculo de estos parámetros que se utilizan para clasificar la señal se explica a continuación.

La correlación r_x normalizada se calcula al final de la trama con base en la síntesis de dominio del tiempo de señal de sonido/sonora $s_{salida}(n)$. Se utiliza el retardo de tono de la última subtrama del decodificador 102 de excitación de dominio del tiempo. Más específicamente, la correlación r_x normalizada calcula el tono sincrónicamente de la siguiente manera:

$$r_x = \frac{\sum_{i=0}^{T-1} x(t+i)x(t+i-T)}{\sqrt{\sum_{i=0}^{T-1} x^2(t+i)\sum_{i=0}^{T-1} x^2(t+i-T)}}, \qquad (1)$$

donde $x(n) = s_{salida}(n)$, T es el retardo de tono de la última subtrama, t=L-T, y L es el tamaño de la trama. Si el retardo de tono de la última subtrama es mayor que 3N/2 (N es el tamaño de la subtrama), T se establece en el retardo de tono promedio de las últimas dos subtramas.

Por lo tanto, la correlación r_x normalizada se calcula utilizando la síntesis de dominio del tiempo de señal de sonido/sonora $s_{salida}(n)$. Para retardos de tono más bajos que el tamaño de la subtrama (64 muestras), la correlación normalizada se calcula dos veces en instantes t=L-T y t=L-2T, y la correlación r_x normalizada se da como el promedio de estos dos cálculos.

El parámetro e_t de inclinación espectral contiene la información sobre la distribución de frecuencia de la energía. Como ejemplo no limitativo, la inclinación espectral en el decodificador se estima como el primer coeficiente de autocorrelación normalizado de la síntesis de dominio de tiempo. Se calcula con base en las últimas 3 subtramas como:

$$e_t = \frac{\sum_{i=N}^{L-1} x(i)x(i-1)}{\sum_{i=N}^{L-1} x^2(i)} \qquad (2)$$

donde x(n) = s_{salida}(n) es la señal de síntesis de dominio de tiempo, N es el tamaño de la subtrama y L es el tamaño de la trama (N = 64 y L = 256 en el ejemplo de AMR-WB).

El contador de estabilidad de tono pc evalúa la variación del período de tono. Se calcula en el decodificador de la siguiente manera:

$$pc = |p_3 + p_2 - p_1 - p_0| \quad (3)$$

Los valores p_0, p_1, p_2 y p_3 corresponden al retardo de tono de circuito cerrado de las 4 subtramas de la trama actual (en el ejemplo de AMR-WB).

La energía E_s de trama relativa se calcula como una diferencia entre la energía de trama actual E_f en dB y su promedio E_{lt} a largo plazo

$$E_s = E_f - E_{lt} \quad (4)$$

donde la energía de trama E_f actual es la energía de la síntesis de dominio de tiempo s_{salida}(n) en dB de tono calculado sincrónicamente al final de la trama como

$$E_f = 10\log_{10}\left(\frac{1}{T}\sum_{i=0}^{T-1} s^2_{salida}(i + L - T)\right) \quad (5)$$

donde L=256 (en el ejemplo de AMR-WB) es la longitud de la trama y T es el retardo de tono promedio de las últimas dos subtramas. Si T es menor que el tamaño de la subtrama, entonces T se establece en 2T (la energía calculada utilizando dos períodos de tono para retardos de tono cortos).

La energía promedio a largo plazo se actualiza en las tramas de sonoridad activas utilizando la siguiente relación:

$$E_{lt} = 0.99E_{lt} + 0.01E_f \quad (6)$$

El último parámetro es el contador de cruce por cero zc calculado en una trama de la síntesis de dominio del tiempo s_{salida}(n). Como ejemplo no limitativo, el contador de cruce por cero zc cuenta el número de veces que el signo de la síntesis de dominio de tiempo cambia de positivo a negativo durante ese intervalo.

Para hacer la clasificación más robusta, los parámetros de clasificación se consideran juntos formando una función de mérito fm. Para ese propósito, los parámetros de clasificación se escalan primero usando una función lineal. Consideremos un parámetro p_x, su versión escalada se obtiene usando:

$$p^s = k_p \cdot p_x + c_p \quad (7)$$

El contador de estabilidad de tono escalado pc se recorta entre 0 y 1. Los coeficientes de función k_p y c_p se han encontrado experimentalmente para cada uno de los parámetros. Los valores utilizados en este ejemplo de implementación se resumen en la tabla 1:

Tabla 1. Parámetros de clasificación de trama en el decodificador y los coeficientes de sus respectivas funciones de escala

Parámetro	Significado	k_p	c_p
r_x	Correlación Normalizada	0.8547	0.2479
e_t	Inclinación espectral	0.8333	0.2917
pc	Contador de estabilidad de tono	-0.0357	1.6074
E_s	Energía relativa de trama	0.04	0.56

Parámetro	Significado	k_p	c_p
zc	Contador de cruce por cero	-0.04	2.52

La función del mérito se define como:

$$f_m = \frac{1}{6}(2 \cdot r_x^s + e_t^s + pc^s + E_s^s + zc^s) \qquad (8)$$

donde el superíndice s indica la versión a escala de los parámetros.

La clasificación de las tramas se realiza utilizando la función de mérito f_m y siguiendo las reglas resumidas en la tabla 2:

Tabla 2: Reglas de clasificación de señales en el decodificador

Clase de trama anterior	Regla	Clase de trama actual
DE APARICIÓN	$f_m \geq 0.63$	DE VOZ
DE VOZ		
TRANSICIÓN DE VOZ		
DE APARICIÓN ARTIFICIAL		
SONIDO DE AUDIO GENÉRICO		
	$0.39 \leq f_m < 0.63$	TRANSICIÓN DE VOZ
	$f_m < 0.39$	SORDA
SORDA DE TRANSICIÓN SORDA	$f_m > 0.56$	DE APARICIÓN
	$0.56 \geq f_m > 0.45$	TRANSICIÓN SORDA
	$f_m \leq 0.45$	SORDA
Información de VAD de trama actual		
	VAD = 0	SORDA

Además de esta clasificación, la información 109 sobre la detección de actividad de voz (VAD) por el codificador puede transmitirse al flujo 101 de bits (figura 1) como es el caso con el ejemplo de AMR-WB. Por lo tanto, se envía un bit al flujo 101 de bits para especificar si el codificador considera o no la trama actual como contenido activo (VAD = 1) o contenido inactivo (ruido de fondo, VAD = 0). Cuando la información de VAD indica que el contenido es inactivo, la porción de clasificador 104, 105, 106 y 301 sobrescribe la clasificación como sorda.

El esquema de clasificación también incluye una detección de audio genérico (véase la porción 301 del clasificador de la figura 3). La categoría de audio genérico incluye música, sonoridad reverberante y también puede incluir música de fondo. Un segundo paso de clasificación permite al clasificador 104-105-106-301 determinar con buena confianza que la trama actual puede categorizarse como audio genérico. Se utilizan dos parámetros para realizar este segundo paso de clasificación. Uno de los parámetros es la energía de trama E_f total como se formula en la ecuación (5).

Primero, una media de las últimas cuarenta (40) variaciones totales de energía de trama E_{df} se calcula utilizando la siguiente relación:

$$\overline{E}_{df} = \frac{\sum_{t=-40}^{t=-1} \Delta_E^t}{40}; \qquad \text{donde} \quad \Delta_E^t = E_f^t - E_f^{(t-1)} \tag{9}$$

Luego, se determina una desviación estadística del historial de variación de energía σ_E en las últimas quince (15) tramas utilizando la siguiente relación:

$$\sigma_E = 0.7745967 \cdot \sqrt{\sum_{t=-15}^{t=-1} \frac{\left(\Delta_E^t - \overline{E}_{df}\right)^2}{15}} \tag{10}$$

La desviación σ_E resultante da una indicación sobre la estabilidad energética de la síntesis decodificada. Típicamente, la música tiene una mayor estabilidad energética (menor desviación estadística del historial de variación de energía) que la sonora.

Además, la clasificación del primer paso se usa para evaluar el intervalo entre dos tramas clasificadas como N_{UV} sordas cuando la energía de la trama E_f, como se formula en la ecuación (5) es superior a -12dB. Cuando una trama se clasifica como sorda y la energía de la trama E_f es mayor que -9dB, lo que significa que la señal es sorda pero no silenciosa, si la energía sonora activa a largo plazo E_{lt}, como se formula en la ecuación (6), está por debajo de 40dB el contador de intervalos sordos se establece en 16, de lo contrario, el N_{UV} del contador de intervalos sordos se reduce en 8. El N_{UV} del contador también está limitado entre 0 y 300 para la señal sonora activa y entre 0 y 125 para la señal sonora inactiva. Se recuerda que, en el ejemplo ilustrativo, la diferencia entre la señal sonora activa e inactiva puede deducirse de la información VAD de detección de actividad de voz incluida en el flujo 101 de bits.

Se deriva un promedio a largo plazo de este contador de trama sorda de la siguiente manera para la señal sonora activa:

$$N_{uv_{lt}} = 0.9 \cdot N_{uv_{lt}} + 0.1 \cdot N_{uv} \tag{11}$$

Y de la siguiente manera para la señal sonora inactiva:

$$N_{uv_{lt}} = 0.95 \cdot N_{uv_{lt}} \tag{12}$$

Además, cuando el promedio a largo plazo es muy alto y la desviación σ_E es alta, por ejemplo cuando $N_{UVlt} > 140$ y $\sigma_E > 5$ en el ejemplo actual de implementación, el promedio a largo plazo se modifica de la siguiente manera:

$$N_{uv_{lt}} = 0.2 \cdot N_{uv_{lt}} + 80 \tag{13}$$

El clasificador 104-105-106-301 utiliza este parámetro en promedio a largo plazo del número de tramas entre tramas clasificadas como sordas para determinar si la trama debe considerarse como audio genérico o no. Cuanto más se acerquen las tramas sordas en el tiempo, más probable es que la trama tenga características sonoras (menos probable audio genérico). En el ejemplo ilustrativo, el umbral para decidir si una trama se considera como audio genérico G_A se define de la siguiente manera:

$$\text{Una trama es } G_A \text{ si:} \quad N_{uv_{lt}} > 140 \quad y \quad \Delta_E^t < 12 \tag{14}$$

El parámetro Δ_E^t, definido en la ecuación (9), se agrega para no clasificar grandes variaciones de energía como audio genérico, sino para mantenerlo como sonora activa.

La modificación realizada en la excitación depende de la clasificación de la trama y para algún tipo de tramas no hay modificación en absoluto. La siguiente tabla 3 resume el caso en el que se puede realizar una modificación o no.

Tabla 3: Categoría de señal para modificación de excitación

Clasificación de trama	¿Actividad de voz detectada? Sí/No	Categoría	Modificación Sí/No
DE APARICIÓN	Sí (VAD=1)	Voz activa	No
DE VOZ			
TRANSICIÓN SORDA			
DE APARICIÓN ARTIFICIAL			
SONIDO DE AUDIO GENÉRICO	Sí	Audio genérico	Sí*
TRANSICIÓN DE VOZ SORDA	Sí	Activa sorda	Sí
DE APARICIÓN	No	Audio inactivo	Sí
DE VOZ			
TRANSICIÓN SORDA			
DE APARICIÓN ARTIFICIAL			
SONIDO DE AUDIO GENÉRICO			
TRANSICIÓN DE VOZ			
SORDA			

* La categoría de audio genérico puede modificarse o no dependiendo de la implementación. Por ejemplo, el audio genérico puede modificarse solo cuando está inactivo, o el audio genérico puede modificarse solo cuando está activo, todo el tiempo o nada.

2) Transformación de frecuencia

Durante la fase de modificación del dominio de frecuencia, la excitación necesita ser representada en el dominio de transformación. Por ejemplo, la conversión de tiempo a frecuencia se logra mediante un convertidor 201 de dominio del tiempo a frecuencia del convertidor/modificador 107 usando una DCT tipo II (transformación discreta de coseno) que proporciona una resolución de frecuencia de 25 Hz pero cualquier otra transformación adecuada puede ser usada. En caso de que se utilice otra transformación, la resolución de frecuencia (definida anteriormente), el número de bandas de frecuencia y el número de segmentos de frecuencia por bandas (definidos más adelante) pueden necesitar una revisión en consecuencia. La representación de frecuencia de la excitación CELP de dominio del tiempo f_e calculada en el convertidor 201 de dominio del tiempo a frecuencia se da a continuación:

$$f_e(k) = \begin{cases} \sqrt{\frac{1}{L}} \cdot \sum_{n=0}^{L-1} e_{id}(n), & k = 0 \\ \sqrt{\frac{2}{L}} \cdot \sum_{n=0}^{L-1} e_{id}(n) \cdot \cos\left(\frac{\pi}{L}\left(n+\frac{1}{2}\right)k\right), & 1 \le k \le L-1 \end{cases} \quad (15)$$

donde $e_{id}(n)$ es la excitación CELP de dominio de tiempo y L es la longitud de la trama. En el ejemplo de AMR-WB, la longitud de la trama es de 256 muestras para una frecuencia de muestreo interna correspondiente de 12.8 kHz.

En un decodificador CELP de dominio de tiempo tal como 102, la señal de excitación de dominio de tiempo viene dada por

$$e_{td}(n) = bv(n) + gc(n) \qquad (15)$$

donde v(n) es la contribución del libro de códigos adaptativo, b es la ganancia del libro de códigos adaptativo, c(n) es la contribución fija del libro de códigos, g es la ganancia fija del libro de códigos.

3) Análisis de energía por banda

Antes de cualquier modificación de la excitación de dominio de tiempo, el convertidor/modificador 107 comprende una calculadora 208-209-210 de ganancia que incluye una subcalculadora 209 para calcular la energía por banda E_b de la excitación de dominio de frecuencia y mantiene la energía calculada por banda E_b en memoria para el ajuste de energía después de la remodelación del espectro de excitación. Para una frecuencia de muestreo de 12.8 kHz, la energía puede ser calculada por la subcalculadora 209 de la siguiente manera:

$$E_b(i) = \sqrt{\sum_{j=C_{Bb}(i)}^{j=C_{Bb}(i)+B_b(i)} f_e(j)^2} \qquad (16)$$

donde C_{Bb} representa los segmentos de frecuencia acumulados por banda y B_b el número de segmentos por banda de frecuencia definidos como:

$$B_b = \{4, 4, 4, 4, 4, 5, 6, 6, 6, 8, 8, 10, 11, 13, 15, 18, 22, 16, 16, 20, 20, 20, 16\}$$

$$C_{Bb} = \left\{ \begin{matrix} 0,8,12,16,20,25,31,37,43,51,59,69,80,93, \\ 108,126,148,164,180,200,220,240 \end{matrix} \right\}$$

Las bandas de baja frecuencia pueden corresponder a las bandas de audio críticas como se describe en [Milan Jelinek y Philippe Gournay. La solicitud de patente PCT WO03102921A1, "A method and device for efficient frame erasure concealment in linear predictive based speech codecs"], pero las bandas de frecuencia por encima de 3700 Hz pueden ser un poco más cortas para igualar mejor la posible variación de energía espectral en esas bandas. Cualquier otra configuración de bandas espectrales también es posible.

4) Modificación de excitación para tramas sordas inactivas y activas

A) Frecuencia de corte de la contribución de dominio de tiempo frente a relleno de ruido

Para lograr una conmutación transparente entre la excitación no modificada y la excitación modificada para tramas inactivas y tramas sordas activas, se mantienen al menos las frecuencias más bajas de la contribución de excitación de dominio de tiempo. El convertidor/modificador 107 comprende una calculadora 203 de frecuencia de corte para determinar una frecuencia en la que la contribución de dominio de tiempo deja de usarse, la frecuencia de corte f_c, que tiene un valor mínimo de 1.2 kHz. Esto significa que los primeros 1.2 kHz de la excitación decodificada se mantienen siempre y, dependiendo del valor de tono decodificado del decodificador 102 de excitación de dominio de tiempo, esta frecuencia de corte puede ser mayor. El octavo armónico se calcula a partir del tono más bajo de todas las subtramas y la contribución de dominio de tiempo se mantiene hasta este octavo armónico. Una estimación del octavo armónico se calcula de la siguiente manera:

$$h_{octavo} = \frac{(8 \cdot F_s)}{\min_{0 \le i < N_{sub}}(T(i))} \qquad (17)$$

donde F_s = 12800 Hz, N_{sub} es el número de subtramas y T es el tono de subtrama decodificado. Para todo i <N_b donde N_b es la banda de frecuencia máxima incluida en el rango de frecuencia L_f, se realiza una verificación para encontrar la banda en la que se ubica el octavo armónico buscando la banda más alta para la cual aún se verifica la siguiente desigualdad:

$$\left(h_{octavo} \ge L_f(i)\right) \qquad (18)$$

donde L_f se define como:

$$L_f = \begin{Bmatrix} 175, 275, 375, 475, 600, 750, 900, 1050, 1250, 1450, 1700, 1975, \\ 2300, 2675, 3125, 3675, 4075, 4475, 4975, 5475, 5975, 6375 \end{Bmatrix}$$

El índice de esa banda de frecuencia en L_f se llamará i_{octavo} e indica la banda de frecuencia donde es probable que se ubique el octavo armónico. La calculadora 203 de frecuencia de corte calculadora calcula la frecuencia de corte final f_{tc} como la frecuencia más alta entre 1.2 kHz y la última frecuencia de la banda de frecuencias en la que es probable que se ubique el octavo armónico ($L_f(i_{octavo})$), usando la siguiente relación:

$$f_{tc} = \max\left(L_f\left(i_{octavo}\right), 1.2 \text{ kHz}\right) \qquad (19)$$

B) Normalización y relleno de ruido.

El convertidor/modificador 107 comprende además un cero 204 que pone a cero los segmentos de frecuencia de las bandas de frecuencia por encima de la frecuencia de corte f_c.

Para las tramas inactivas y las tramas sordas activas, un normalizador 205 del convertidor/modificador 107 normaliza los segmentos de frecuencia por debajo de f_c de las bandas de frecuencia de la representación de frecuencia de la excitación f_e CELP del dominio del tiempo entre [0, 4] usando la siguiente relación:

$$f_{eN}(j) = \begin{cases} \dfrac{4 \cdot f_e(j)}{\max\limits_{0 \le i < f_c}\left(\left|f_e(i)\right|\right)}, & \text{para} \quad 0 \le j < f_c \\ 0, & \text{para} \quad f_c \le j < 256 \end{cases} \qquad (20)$$

Entonces, el convertidor/modificador 107 comprende un generador 206 de ruido aleatorio para generar ruido aleatorio y se realiza un relleno de ruido simple a través de un sumador 207 para agregar ruido sobre todos los segmentos de frecuencia a un nivel constante. La función que describe la adición de ruido se define a continuación como:

$$\text{para } j = 0 : L-1$$
$$f'_{eN}(j) = f_{eN}(j) + 0.75 \cdot r_y \quad ( ) \qquad (21)$$

donde r_y es un generador de números aleatorios que está limitado entre -1 y 1.

c) Análisis de energía por banda del espectro de excitación modificado

La subcalculadora 208 de la calculadora 208-209-210 de ganancia determina la energía por banda después de la remodelación del espectro E_b' usando el mismo método que se describe en la sección 3 anterior.

D) Emparejamiento de energía

Para tramas inactivas y tramas sordas activas, el emparejamiento de energía consiste solo en ajustar la energía por banda después de la modificación del espectro de excitación a su valor inicial. Para cada banda i, la subcalculadora 210 de la calculadora 208-209-210 de ganancia determina una ganancia de emparejamiento G_b para aplicar a todos los segmentos en la banda de frecuencia para emparejar la energía de la siguiente manera:

$$G_b(i) = \frac{E_b(i)}{E_b'(i)} \qquad (22)$$

donde $E_b(i)$ es la energía por banda antes de la modificación del espectro de excitación como se determina en la subcalculadora 209 usando el método de la sección 3 anterior y $E'_b(i)$ es la energía por banda después de la modificación del espectro de excitación como se calcula en la subcalculadora 208. Para una banda i específica, la excitación modificada (desnormalizada) del dominio de frecuencia f'_{eN} como lo determinado en la subcalculadora 210 puede escribirse como:

$$\text{para } C_{Bb}(i) \le j < C_{Bb}(i) + B_b(i)$$
$$f'_{edN}(j) = G_b(i) \cdot f'_{eN}(j) \qquad (23)$$

donde C_{Bb} y B_b se definen en la sección 3 anterior.

5) Modificación de excitación para tramas de audio genéricas

A) Normalización y relleno de ruido.

Ahora se hará referencia a la figura 3. Para las tramas de audio genéricas según lo determinado por la porción 301 del clasificador, la normalización es ligeramente diferente y realizada por un normalizador 302. Primero, el factor de normalización N_f cambia de banda a banda, usando un valor más alto. para bandas de baja frecuencia y un valor inferior para bandas de alta frecuencia. La idea es permitir una mayor amplitud en las bandas de baja frecuencia donde la ubicación de los pulsos es más precisa y una amplitud más baja en las bandas de alta frecuencia donde la ubicación de los pulsos es menos precisa. En este ejemplo ilustrativo, el factor de normalización N_f variable por banda de frecuencia se define como:

$$N_f = \{16, 16, 16, 16, 16, 16, 16, 12, 12, 12, 12, 8, 8, 8, 8, 8, 4, 4, 2, 2, 1, 1, 1\}$$

Para una banda i de frecuencia específica, la normalización de la representación de frecuencia de la excitación de dominio de tiempo (excitación de dominio de frecuencia) de tramas de audio genéricas se puede describir de la siguiente manera:

$$f_{eN}(j) = \frac{N_f(i) \cdot f_e(j)}{\max_{k=C_{Bb}(i)}^{C_{Bb}(i)+B_b(i)} \left(|f_e(k)|\right)}, \quad \text{para} \quad C_{Bb}(i) \le j < C_{Bb}(i) + B_b(i) \qquad (24)$$

donde B_b es el número de segmentos por banda de frecuencia, los segmentos de frecuencia acumulada por bandas es C_{Bb} y $f_{eN}(j)$ es la excitación de dominio de frecuencia normalizada. B_b y C_{Bb} se describen en la sección 3 anterior.

Además, el normalizador 302 comprende una reducción a cero (no mostrado) para poner a cero todos los segmentos de frecuencia por debajo de una fracción Z_f del valor máximo de $f_{eN}(j)$ en cada banda de frecuencia para obtener $f'_{eN}(j)$:

$$f'_{eN}(j) = \begin{cases} 0 & si \left(f_{eN}(j) < Z_f(i)\right) \\ f_{eN}(j) & \text{de otra manera} \end{cases}\Bigg|_{\text{para } C_{Bb}(i) \le j < C_{Bb}(i) + B_b(i)} \qquad (25)$$

donde Z_f se puede representar como:

$$Z_f = \{1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0.5, 0.5, 0.5\}$$

Se puede realizar una reducción a cero más agresiva aumentando el valor del vector Z_f, si se desea aumentar el pico del espectro.

B) Análisis de energía por banda del espectro de excitación modificado

La porción 303 de calculadora de una calculadora 303-304-306 de ganancia determina la energía por banda después de remodelar el espectro E_b' usando el mismo método que se describe en la sección 3 anterior.

C) Emparejamiento de energía

La figura 3 muestra la calculadora 303-304-306 de ganancia y la figura 4 describe con más detalle la porción 306 de calculadora de esta calculadora de ganancia.

Para tramas de audio genéricas, el emparejamiento de energía es más complicado ya que también apunta a aumentar la dinámica espectral. Para cada banda i de frecuencia, una subcalculadora 413 de la porción 306 de calculadora de

la calculadora 303-304-306 de ganancia calcula una ganancia estimada G_e definida de manera similar a la ecuación (22):

$$G_e(i) = \frac{E_b(i)}{E_b'(i)} \qquad (26)$$

donde $E_b(i)$ es la energía por banda antes de la modificación del espectro de excitación según lo determinado en la porción 304 de la calculadora usando el método descrito en la sección 3 anterior, y $E'_b(i)$ es la energía por banda después de la modificación del espectro de excitación según lo calculado en la porción 303 de calculadora.

Una subcalculadora 414 de la porción 306 de calculadora aplica la ganancia G_e a los primeros 400 Hz (o las primeras 4 bandas) de la excitación de dominio de frecuencia normalizada f'_{eN} del normalizador 302 y el divisor 401-420 de espectro para proporcionar una excitación de dominio de frecuencia modificada (desnormalizada) f'_{edN} usando la siguiente relación:

$$f'_{edN}(j) = G_e(i) \cdot f'_{eN}(j), \quad para \ C_{Bb}(i) \le j < C_{Bb}(i) + B_b(i)|_{0 \le i < 4} \qquad (27)$$

Un buscador 404 determina el valor máximo $\max_{a \le j < b}(|f_{eN}(j)|)$ por banda i por encima de 400 Hz, donde a = $C_{Bb}(i)$ y b = $C_{Bb}(i) + B_b(i)$ son definidos en la sección 3 anterior.

Para las bandas de frecuencia comprendidas entre 400 Hz y 2 kHz (bandas 4 a 12) de la excitación del dominio de frecuencia normalizada (véanse los módulos 420 y 450), si la excitación del dominio de frecuencia normalizada en un segmento de frecuencia $f'_{eN}(j) \ge 0.86 \max_{a \le j < b}(|f_{eN}(j)|)$ (véase el módulo 451), un amplificador 402 amplifica la ganancia G_e de la subcalculadora 413 por un factor 1.1 como se muestra en la línea superior de la ecuación (28). Una subcalculadora 403 aplica la ganancia amplificada del amplificador 402 a la excitación espectral normalizada f'eN en el segmento de frecuencia de acuerdo con la primera línea de la ecuación (28) para obtener la excitación modificada (desnormalizada) del dominio de frecuencia f'_{edN}.

Una vez más, para las bandas de frecuencia comprendidas entre 400 Hz y 2 kHz (bandas 4 a 12) de la excitación de dominio de frecuencia normalizada (véanse los módulos 420 y 450), si la excitación de dominio de frecuencia normalizada en un segmento de frecuencia $f'_{eN}(j) < 0.86 \max_{a \le j < b}(|f_{eN}(j)|)$ (véase el módulo 451), un atenuador 405 atenúa la ganancia G_e de la subcalculadora 413 por un factor 0.86 como se muestra en la línea inferior de la ecuación (28). Una subcalculadora 406 aplica la ganancia atenuada del atenuador 405 a la excitación espectral normalizada f'_{eN} en el segmento de frecuencia de acuerdo con la línea inferior de la ecuación (28) para obtener la excitación modificada (desnormalizada) del dominio de frecuencia f'_{edN}.

Para resumir, la excitación espectral modificada (desnormalizada) f'_{edN} se da de la siguiente manera:

$$f'_{edN}(j) = \begin{cases} 1.1 \cdot G_e(i) \cdot f'_{eN}(j), & si \quad f'_{eN}(j) \ge 0.86 \cdot \max\limits_{a \le j < b}\left(\left|f_{eN}(j)\right|\right) \\ 0.86 \cdot G_e(i) \cdot f'_{eN}(j), & si \quad f'_{eN}(j) < 0.86 \cdot \max\limits_{a \le j < b}\left(\left|f_{eN}(j)\right|\right) \end{cases} \qquad (28)$$

Finalmente, para las partes más altas del espectro, en este ejemplo, las bandas de frecuencia por encima de 2 kHz (bandas > 12) de la excitación de dominio de frecuencia normalizada (véanse los módulos 420 y 450), si la excitación de dominio de frecuencia normalizada en una frecuencia segmento $f'_{eN}(j) \ge 0.86 \max_{a \le j < b}(|f_{eN}(j)|)$ (véase el módulo 452), una inclinación que es función de la banda i de frecuencia y que también puede ser función de la tasa de bits se agrega a la ganancia G_e para compensar la estimación de energía demasiado baja del filtro LPC. El valor de la inclinación por banda de frecuencia δ(i) se formula como:

$$\delta(i) = 1.5 \cdot G_e(i) \cdot \frac{(j - 12)}{32} \qquad (29)$$

La inclinación se calcula mediante la calculadora 407-408 de inclinación y se aplica a la excitación de dominio de frecuencia normalizada f'_{eN} por el segmento de frecuencia de acuerdo con la línea superior de la ecuación (30) mediante una subcalculadora 409 para obtener la excitación f'_{edN} de dominio de frecuencia modificada (desnormalizada).

Una vez más, para las partes más altas del espectro, en este ejemplo ilustrativo, las bandas de frecuencia por encima de 2 kHz (bandas > 12) de la excitación de dominio de frecuencia normalizada (véanse los módulos 420 y 450), si la

excitación de dominio de frecuencia normalizada en un frecuencia segmento $f'_{eN}(j)$ <0.86 $\max_{a\leq j<b}(|f_{eN}(j)|)$ (véase el módulo 452), un atenuador 410 calcula una ganancia de atenuación $[f'_{eN}(j)/\max_{a\leq j<b}(|f_{eN}(j)|)]^2$ aplicado a la excitación espectral normalizada fe_N por segmento de frecuencia de acuerdo con la línea inferior de la ecuación (30) por una subcalculadora 411 para obtener la excitación de dominio de frecuencia modificada (desnormalizada) f'_{edN}.

Para resumir, la excitación espectral desnormalizada f'_{edN} se determina de la siguiente manera:

$$f'_{edN}(j)=\begin{cases}\delta(i)\cdot f'_{eN}(j), & si \quad f'_{eN}(j)\geq 0.86\cdot\max\limits_{a\leq j<b}\left(\left|f_{eN}(j)\right|\right)\\ \left(\dfrac{f'_{eN}(j)}{\max\limits_{a\leq j<b}\left(\left|f_{eN}(j)\right|\right)}\right)^2\cdot f'_{eN}(j), & de\ otra\ manera\end{cases} \qquad (30)$$

donde a y b se describen en este documento anteriormente. También es posible aumentar aún más la ganancia aplicada a las últimas bandas, donde el emparejamiento de energía de LPC es la peor.

6) Transformación de frecuencia inversa

Un combinador 453 combina las contribuciones a la excitación de dominio de frecuencia modificada (desnormalizada) f'_{edN} de las subcalculadoras 414, 403, 406, 409 y 411 para formar la excitación de dominio de frecuencia modificada (desnormalizada) completa f'_{edN}.

Después de que se completa el procesamiento de dominio de frecuencia, se aplica una transformación 202 inversa de frecuencia-tiempo a la excitación de dominio de frecuencia modificada (desnormalizada) f'_{edN} del combinador 453 para encontrar la excitación modificada de dominio del tiempo. En esta realización ilustrativa, la conversión de frecuencia a tiempo se logra con la inversa de la misma DCT tipo II que se usó para la conversión de tiempo a frecuencia dando una resolución de 25 Hz. De nuevo, se pueden usar otras transformaciones. La excitación modificada de dominio de tiempo e'_{td} se obtiene de la siguiente manera:

$$e'_{id}(k)=\begin{cases}\sqrt{\dfrac{1}{L}}\cdot\sum\limits_{n=0}^{L-1}f'_{edN}(n), & k=0\\ \sqrt{\dfrac{2}{L}}\cdot\sum\limits_{n=0}^{L-1}f'_{edN}(n)\cdot\cos\left(\dfrac{\pi}{L}\left(n+\dfrac{1}{2}\right)k\right), & 1\leq k\leq L-1\end{cases} \qquad (31)$$

donde $f'_{edN}(n)$ es la representación de frecuencia de la excitación modificada, y L es la longitud de la trama. En este ejemplo ilustrativo, la longitud de la trama es de 256 muestras para una frecuencia de muestreo interna correspondiente de 12.8 kHz (AMR-WB).

7) Síntesis filtrando y sobrescribiendo la síntesis actual de CELP

Una vez que se completa la modificación de la excitación, la excitación modificada se procesa a través del filtro 108 de síntesis para obtener una síntesis modificada para la trama actual. El sobreescritor 110 usa esta síntesis modificada para sobrescribir la síntesis decodificada y así aumentar la calidad perceptiva.

La desacentuación final y el remuestreo a 16 kHz se pueden realizar luego en el filtro de desacentuación y el remuestreador 112.

REIVINDICACIONES

1. Un dispositivo para modificar, durante la decodificación de una señal de sonido, una síntesis de una excitación de dominio de tiempo decodificada por un decodificador (102) de dominio de tiempo, que comprende:

un clasificador (104, 105, 106, 301) de la síntesis de la excitación de dominio de tiempo decodificada en uno de un número de categorías;

un convertidor (107, 201) de la excitación de dominio de tiempo decodificada en una excitación de dominio de frecuencia;

un modificador (107, 203, 204, 205, 206, 207, 208, 209, 210) de la excitación de dominio de frecuencia en función de la categoría en la que el clasificador (104, 105, 106, 301) clasifica la síntesis de la excitación de dominio de tiempo decodificado;

un convertidor (107, 202) de la excitación modificada de dominio de frecuencia en una excitación modificada de dominio de tiempo;

un filtro (108) de síntesis suministrado con la excitación modificada de dominio de tiempo para producir una síntesis modificada de la excitación de dominio de tiempo decodificada.

2. Un dispositivo para modificar una síntesis de una excitación de dominio de tiempo de acuerdo con la reivindicación 1, en donde el modificador (107, 203, 204, 206, 207, 208, 209, 210) comprende:

una calculadora (203) de una frecuencia de corte donde se deja de utilizar una contribución de excitación de dominio de tiempo;

una reducción a cero (204) de la excitación del dominio de frecuencia por encima de la frecuencia de corte; y

un normalizador (205) de la excitación de dominio de frecuencia por debajo de la frecuencia de corte para producir una excitación de dominio de frecuencia normalizada.

3. Un dispositivo para modificar una síntesis de una excitación de dominio de tiempo de acuerdo con la reivindicación 2, en donde el modificador (107, 203, 204, 205, 206, 207, 208, 209, 210) comprende un generador (206) de ruido aleatorio y un sumador (207) del ruido aleatorio a la excitación del dominio de frecuencia puesto a cero por encima de la frecuencia de corte y normalizado por debajo de dicha frecuencia de corte.

4. Un dispositivo para modificar una síntesis de una excitación de dominio de tiempo de acuerdo con la reivindicación 3, en donde el modificador (107, 203, 204, 205, 206, 207, 208, 209, 210) comprende:

una calculadora (208, 209, 210) de una ganancia de emparejamiento que utiliza una energía de la excitación del dominio de frecuencia antes y después de la modificación, la calculadora (208, 209, 210) aplica la ganancia de emparejamiento a la excitación normalizada del dominio de frecuencia para producir excitación de dominio de frecuencia modificada.

5. Un dispositivo para modificar una síntesis de una excitación de dominio de tiempo de acuerdo con cualquiera de las reivindicaciones 1 a 4, en donde el clasificador (104, 105, 106, 301) clasifica la síntesis de la excitación de dominio de tiempo decodificada como sorda inactiva o activa.

6. Un dispositivo para modificar una síntesis de una excitación de dominio de tiempo de acuerdo con una cualquiera de las reivindicaciones 1 a 5, que comprende un suavizador (111) de filtro de síntesis cuando la síntesis de la excitación de dominio de tiempo decodificada se clasifica como una dada de las categorías por el clasificador.

7. Un dispositivo para decodificar una señal de sonido codificada por parámetros de codificación, que comprende:

un decodificador (102) de una excitación de dominio de tiempo en respuesta a los parámetros de codificación de la señal de sonido;

un filtro (103) de síntesis sensible a la excitación de dominio de tiempo decodificada para producir una síntesis de dicha excitación de dominio de tiempo; y

un dispositivo de acuerdo con una cualquiera de las reivindicaciones 1 a 6, para modificar la síntesis de la excitación de dominio de tiempo.

8. Un método para modificar, durante la decodificación de una señal de sonido, una síntesis de una excitación de dominio de tiempo decodificada por un decodificador (102) de dominio de tiempo, que comprende:

clasificar (104, 105, 106, 301) la síntesis de la excitación de dominio de tiempo decodificada en uno de un número de categorías;

convertir (107, 201) la excitación de dominio de tiempo decodificada en una excitación de dominio de frecuencia;

modificar (107, 203, 204, 205, 206, 207, 208, 209, 210) la excitación de dominio de frecuencia en función de la categoría en la que se clasifica la síntesis de la excitación de dominio de tiempo decodificado;

convertir (107, 202) la excitación de dominio de frecuencia modificada en una excitación de dominio del tiempo modificada;

sintetizar (108) la excitación modificada de dominio de tiempo para producir una síntesis modificada de la excitación de dominio de tiempo decodificada.

9. Un método para modificar una síntesis de una excitación de dominio de tiempo de acuerdo con la reivindicación 8, en donde la excitación de dominio de frecuencia se divide en bandas de frecuencia, cada una dividida en segmentos de frecuencia, y en donde la modificación de la excitación de dominio de frecuencia comprende:

normalizar (302) la excitación de dominio de frecuencia utilizando un factor de normalización que varía en la banda de frecuencia para producir una excitación de dominio de frecuencia normalizada.

10. Un método para modificar una síntesis de una excitación de dominio de tiempo de acuerdo con la reivindicación 9, en donde modificar la excitación de dominio de frecuencia comprende reducir a cero (302) los segmentos de frecuencia por debajo de una fracción de un valor máximo de la excitación de dominio de frecuencia normalizada en la banda de frecuencia que comprende los segmentos de frecuencia.

11. Un método para modificar una síntesis de una excitación de dominio de tiempo de acuerdo con la reivindicación 9 o 10, en donde modificar la excitación de dominio de frecuencia comprende:

calcular (303, 304, 306) una ganancia de emparejamiento por banda de frecuencia utilizando una energía de la excitación del dominio de frecuencia antes y después de la modificación.

12. Un método para modificar una síntesis de una excitación de dominio de tiempo de acuerdo con la reivindicación 11, en donde modificar la excitación de dominio de frecuencia comprende, para las bandas de frecuencia por debajo de una primera frecuencia, aplicar (306) la ganancia de emparejamiento a la excitación de dominio de frecuencia normalizada para producir la excitación de dominio de frecuencia modificada.

13. Un método para modificar una síntesis de una excitación de dominio de tiempo de acuerdo con la reivindicación 11, que comprende, para las bandas de frecuencia entre una primera frecuencia inferior y una segunda frecuencia superior:

encontrar (404) un valor máximo por banda de frecuencia de la excitación de dominio de frecuencia normalizada;

amplificar (402) la ganancia de emparejamiento por un factor de amplificación por segmento de frecuencia si la excitación normalizada del dominio de frecuencia en el segmento de frecuencia es igual a o mayor que un valor proporcional a dicho valor máximo de la banda de frecuencia;

atenuar (405) la ganancia de emparejamiento por un factor de atenuación por segmento de frecuencia de la banda de frecuencia si la excitación normalizada del dominio de frecuencia en el segmento de frecuencia es inferior al valor proporcional a dicho valor máximo de la banda de frecuencia; y

aplicar (403, 406) la ganancia de emparejamiento amplificada o atenuada a la excitación del dominio de frecuencia normalizada en dicho segmento de frecuencia para producir en dicho segmento de frecuencia la excitación del dominio de frecuencia modificada.

14. Un método para modificar una síntesis de una excitación de dominio de tiempo de acuerdo con la reivindicación 11, que comprende, para las bandas de frecuencia por encima de una frecuencia dada:

encontrar (404) un valor máximo por banda de frecuencia de la excitación de dominio de frecuencia normalizada;

calcular (407, 408) una corrección de inclinación para la ganancia de emparejamiento cuando la excitación del dominio de frecuencia normalizada en el segmento de frecuencia es mayor que un valor proporcional a dicho valor máximo de la banda de frecuencia, y aplicar la corrección de inclinación calculada a la ganancia de emparejamiento; y

aplicar (409) la ganancia de emparejamiento a la que se ha aplicado la corrección de inclinación calculada a la excitación del dominio de frecuencia normalizada en dicho segmento de frecuencia para producir en dicho segmento de frecuencia la excitación del dominio de frecuencia modificada.

15. Un método para modificar una síntesis de una excitación de dominio de tiempo de acuerdo con la reivindicación 11, que comprende, para las bandas de frecuencia por encima de una frecuencia dada:

encontrar (404) un valor máximo por banda de frecuencia de la excitación de dominio de frecuencia normalizada;

atenuar (410) la ganancia de emparejamiento por un factor de atenuación por segmento de frecuencia de la banda de frecuencia cuando la excitación normalizada del dominio de frecuencia en el segmento de frecuencia es inferior a un valor proporcional al valor máximo de la banda de frecuencia; y

aplicar (411) la ganancia de emparejamiento atenuada a la excitación del dominio de frecuencia normalizada en dicho segmento de frecuencia para producir en dicho segmento de frecuencia la excitación del dominio de frecuencia modificada.

16. Un método para decodificar una señal de sonido codificada por parámetros de codificación, que comprende:

decodificar (102) una excitación de dominio de tiempo en respuesta a los parámetros de codificación de la señal de sonido;

sintetizar (103) la excitación de dominio de tiempo decodificada para producir una síntesis de dicha excitación de dominio de tiempo; y

un método de acuerdo con una cualquiera de las reivindicaciones 8 a 15, para modificar la síntesis de la excitación de dominio de tiempo.

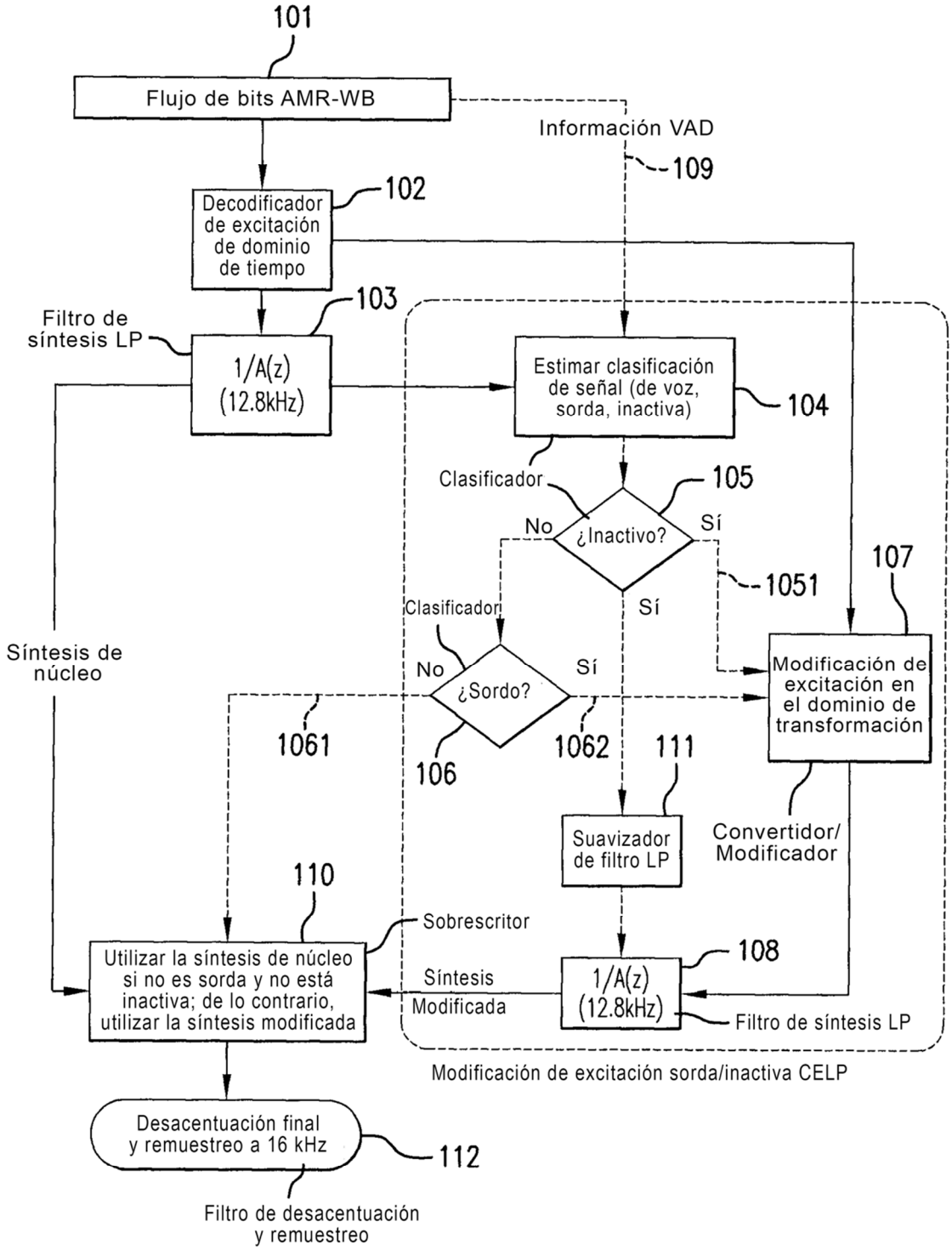
101
Flujo de bits AMR-WB
Información VAD
109
Decodificador de excitación de dominio de tiempo
102
Filtro de síntesis LP
103
1/A(z) (12.8kHz)
Estimar clasificación de señal (de voz, sorda, inactiva)
104
Clasificador
105
No
¿Inactivo?
Sí
1051
107
Clasificador
Sí
Síntesis de núcleo
No
¿Sordo?
Sí
Modificación de excitación en el dominio de transformación
1061
1062
111
106
Suavizador de filtro LP
Convertidor/ Modificador
110
Sobrescritor
108
Utilizar la síntesis de núcleo si no es sorda y no está inactiva; de lo contrario, utilizar la síntesis modificada
Síntesis Modificada
1/A(z) (12.8kHz)
Filtro de síntesis LP
Modificación de excitación sorda/inactiva CELP
Desacentuación final y remuestreo a 16 kHz
112
Filtro de desacentuación y remuestreo

FIG.1

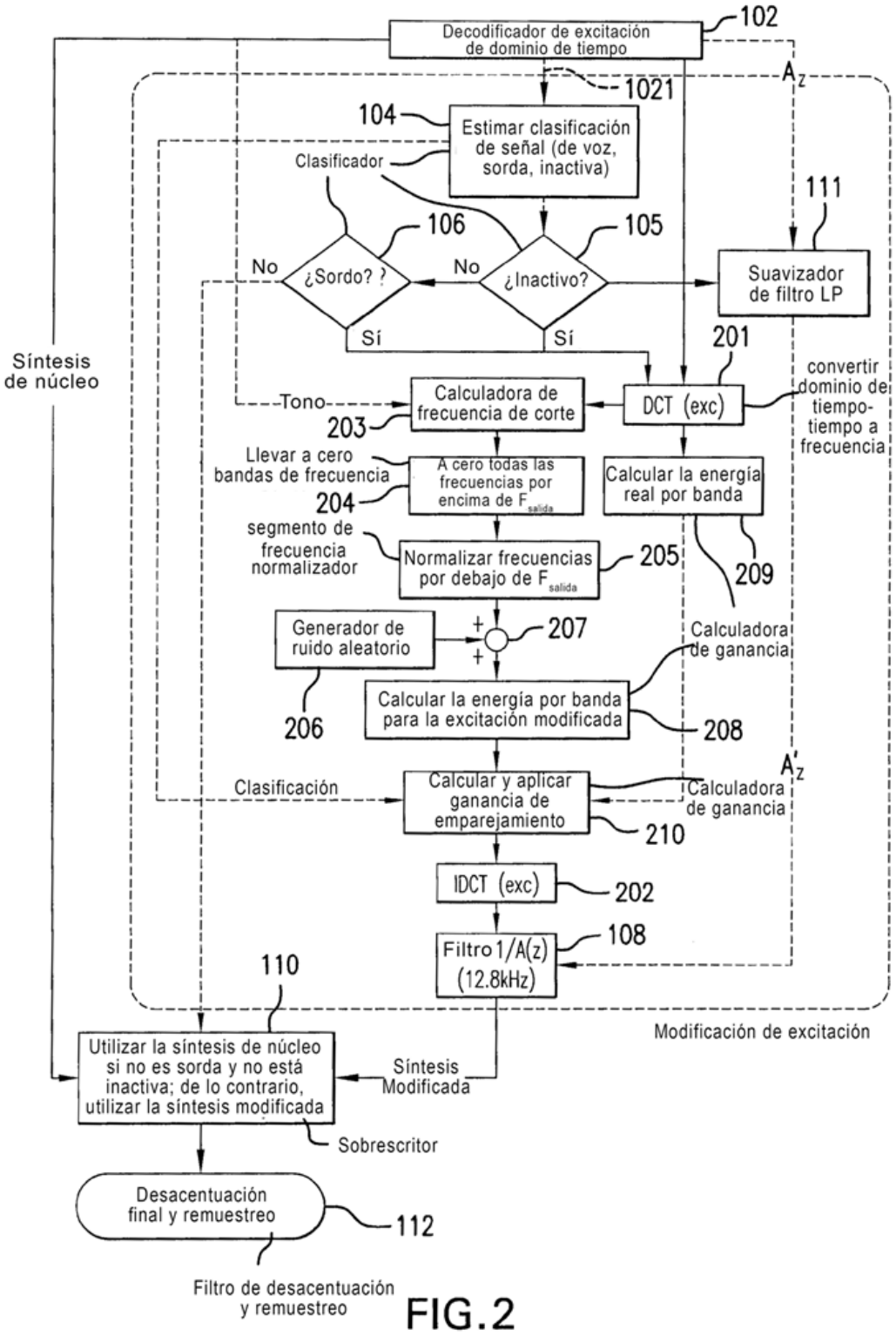
Decodificador de excitación de dominio de tiempo
102
1021
A_z
104
Estimar clasificación de señal (de voz, sorda, inactiva)
Clasificador
111
106
105
No
¿Sordo? ?
No
¿Inactivo?
Suavizador de filtro LP
Sí
Sí
201
Síntesis de núcleo
convertir dominio de tiempo-tiempo a frecuencia
Tono
Calculadora de frecuencia de corte
DCT (exc)
203
Llevar a cero bandas de frecuencia
A cero todas las frecuencias por encima de F_{salida}
Calcular la energía real por banda
204
segmento de frecuencia normalizador
Normalizar frecuencias por debajo de F_{salida}
205
209
Generador de ruido aleatorio
207
Calculadora de ganancia
206
Calcular la energía por banda para la excitación modificada
208
A'_z
Clasificación
Calcular y aplicar ganancia de emparejamiento
Calculadora de ganancia
210
IDCT (exc)
202
Filtro 1/A(z) (12.8kHz)
108
110
Modificación de excitación
Utilizar la síntesis de núcleo si no es sorda y no está inactiva; de lo contrario, utilizar la síntesis modificada
Síntesis Modificada
Sobrescritor
Desacentuación final y remuestreo
112
Filtro de desacentuación y remuestreo

FIG.2

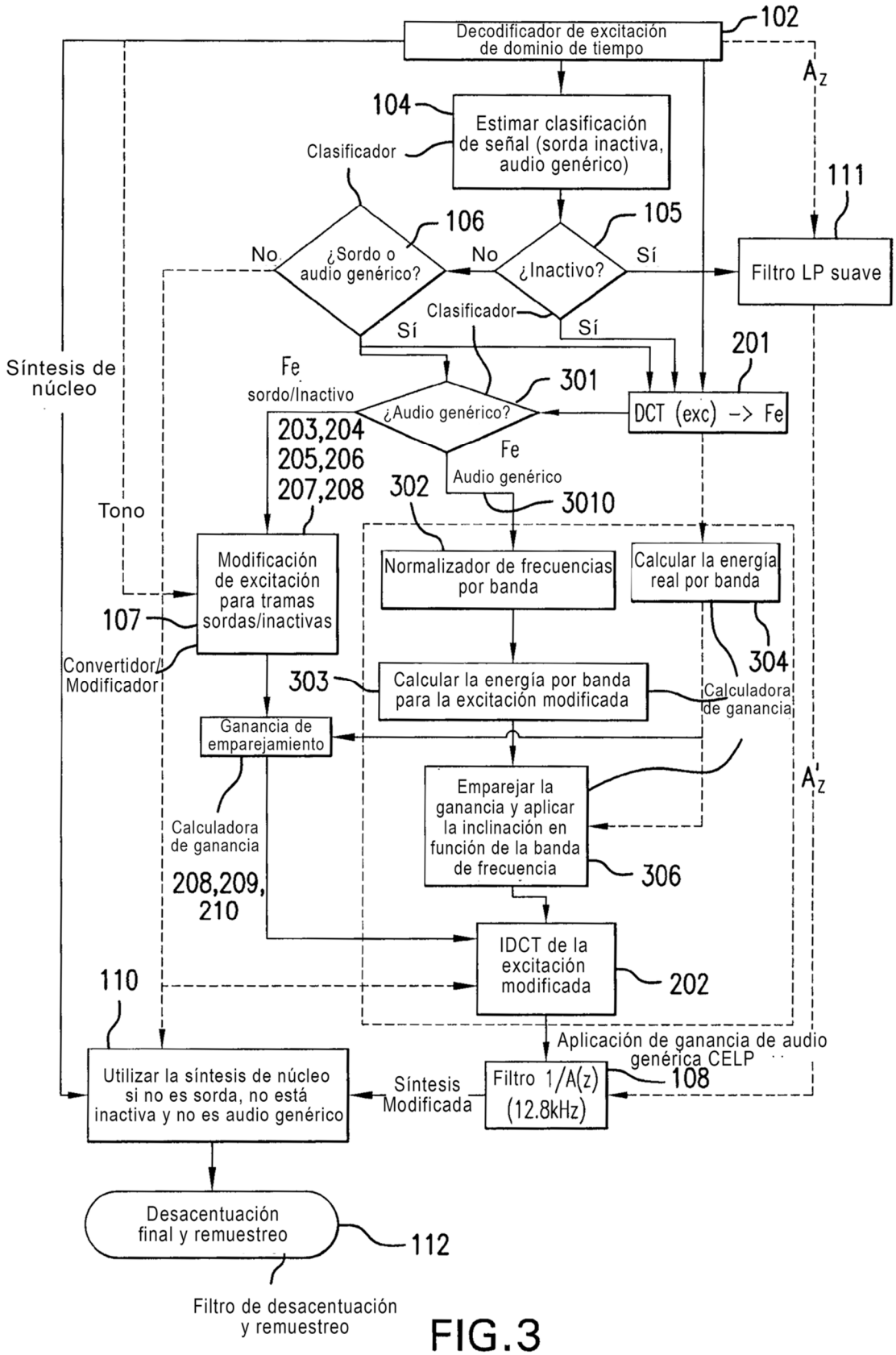
Decodificador de excitación de dominio de tiempo
102
A_z
104
Estimar clasificación de señal (sorda inactiva, audio genérico)
Clasificador
111
106
105
No
¿Sordo o audio genérico?
No
¿Inactivo?
Sí
Filtro LP suave
Clasificador
Sí
Sí
201
Síntesis de núcleo
Fe
sordo/Inactivo
301
¿Audio genérico?
DCT (exc) -> Fe
203,204
205,206
207,208
Fe
Audio genérico
302
3010
Tono
Modificación de excitación para tramas sordas/inactivas
Normalizador de frecuencias por banda
Calcular la energía real por banda
107
304
Convertidor/ Modificador
303
Calcular la energía por banda para la excitación modificada
Calculadora de ganancia
Ganancia de emparejamiento
A'_z
Emparejar la ganancia y aplicar la inclinación en función de la banda de frecuencia
Calculadora de ganancia
306
208,209, 210
IDCT de la excitación modificada
110
202
Aplicación de ganancia de audio genérica CELP
Utilizar la síntesis de núcleo si no es sorda, no está inactiva y no es audio genérico
Síntesis Modificada
Filtro 1/A(z) (12.8kHz)
108
Desacentuación final y remuestreo
112
Filtro de desacentuación y remuestreo

FIG.3

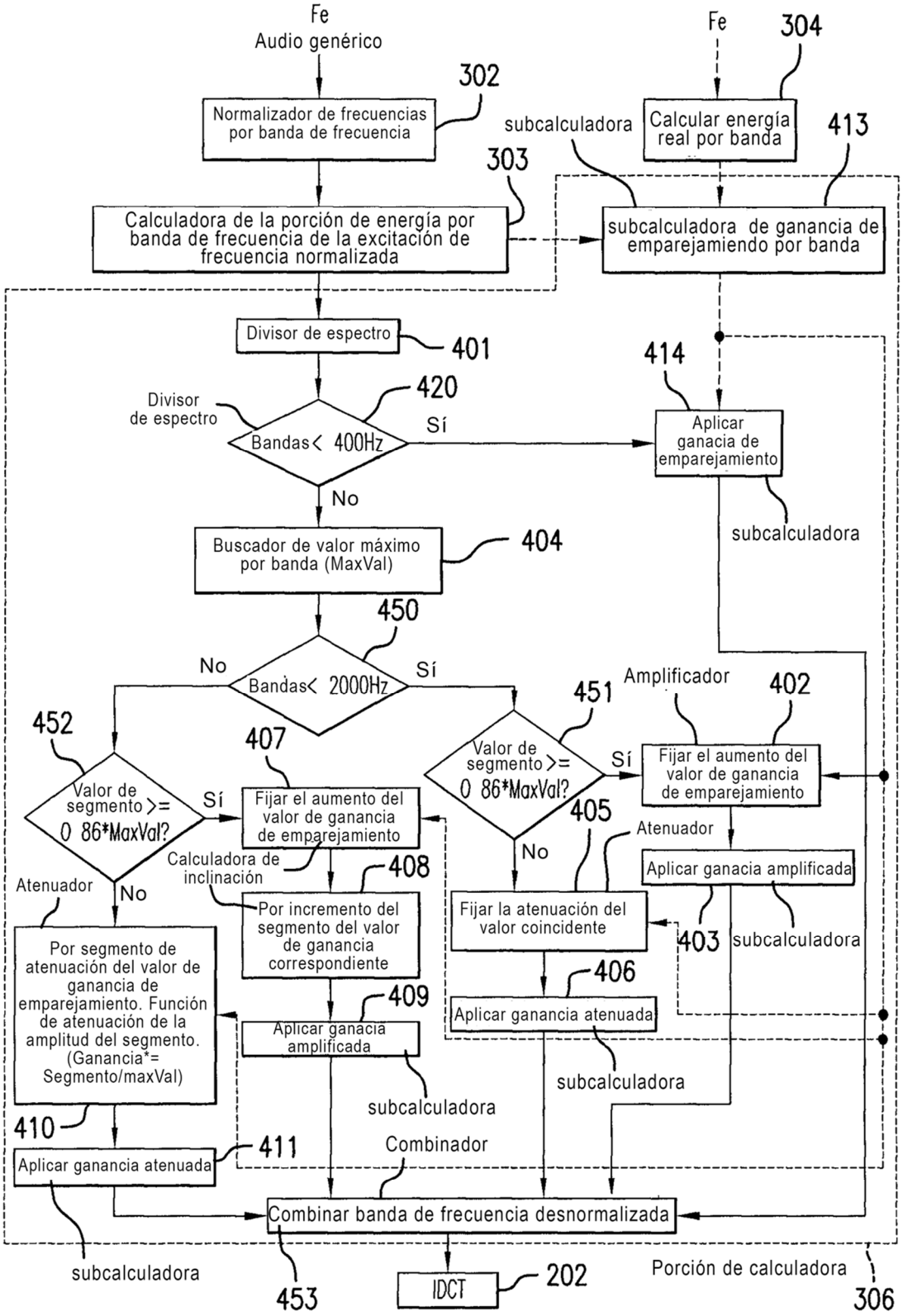
Fe
Audio genérico
302
Normalizador de frecuencias por banda de frecuencia
Fe
304
Calcular energía real por banda
413
subcalculadora
303
Calculadora de la porción de energía por banda de frecuencia de la excitación de frecuencia normalizada
subcalculadora de ganancia de emparejamiendo por banda
Divisor de espectro
401
414
420
Divisor de espectro
Bandas< 400Hz
Sí
Aplicar ganacia de emparejamiento
No
subcalculadora
Buscador de valor máximo por banda (MaxVal)
404
450
No
Bandas< 2000Hz
Sí
Amplificador
402
451
452
407
Valor de segmento >= 0 86*MaxVal?
Sí
Fijar el aumento del valor de ganancia de emparejamiento
Valor de segmento >= 0 86*MaxVal?
Sí
Fijar el aumento del valor de ganancia de emparejamiento
405
Atenuador
Calculadora de inclinación
408
No
Aplicar ganancia amplificada
Atenuador
No
Por incremento del segmento del valor de ganancia correspondiente
Fijar la atenuación del valor coincidente
403
subcalculadora
Por segmento de atenuación del valor de ganancia de emparejamiento. Función de atenuación de la amplitud del segmento. (Ganancia*= Segmento/maxVal)
406
409
Aplicar ganancia atenuada
Aplicar ganancia amplificada
subcalculadora
subcalculadora
410
411
Combinador
Aplicar ganancia atenuada
Combinar banda de frecuencia desnormalizada
subcalculadora
453
IDCT
202
Porción de calculadora
306

FIG.4