(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 757 700**

(51) Int. CI.:

G10L 25/90 (2013.01)
G10L 19/09 (2013.01)
G10L 25/21 (2013.01)
G10L 25/06 (2013.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Fecha de presentación y número de la solicitud europea: **21.12.2012 E 17193357 (5)**

(97) Fecha y número de publicación de la concesión europea: **28.08.2019 EP 3301677**

(54) Título: **Detección y codificación de altura tonal muy débil**

(30) Prioridad:

21.12.2011 US 201161578398 P

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
29.04.2020

(73) Titular/es:

HUAWEI TECHNOLOGIES CO., LTD. (100.0%)
Huawei Administration Building, Bantian, Longgang District
Shenzhen, Guangdong 518129, CN

(72) Inventor/es:

GAO, YANG y
QI, FENGYAN

(74) Agente/Representante:

LEHMANN NOVO, María Isabel

DESCRIPCIÓN

Detección y codificación de altura tonal muy débil

CAMPO TÉCNICO

La presente invención se refiere, en general, al campo de codificación de señales y, en formas de realización particulares, a un sistema y método para la detección y codificación de altura tonal muy débil.

ANTECEDENTES DE LA INVENCIÓN

Por lo general, los métodos paramétricos de codificación vocal hacen uso de la redundancia inherente en la señal vocal con el fin de reducir la cantidad de información que ha de enviarse, y estimar los parámetros de muestras vocales de una señal en intervalos cortos. Esta redundancia puede ser el resultado de la repetición de formas de onda vocales a una tasa casi periódica y la envolvente espectral, que cambia lentamente, de la señal vocal. La redundancia de las formas de onda vocales puede considerarse con respecto a tipos diferentes de señal vocal, tales como de voz y sin voz. Para la señal vocal con voz, dicha señal vocal es prácticamente periódica. Sin embargo, esta periodicidad puede variar a través de la duración de un segmento vocal, y la forma de la onda periódica puede cambiar, de forma gradual, de un segmento a otro. Una codificación vocal de baja tasa binaria podría beneficiarse, de forma sustancial, mediante la exploración de dicha periodicidad. El período vocal de voz se denomina también altura tonal, y la predicción de altura tonal se denomina, a menudo, Predicción a Largo Plazo (LTP). En cuanto a la señal vocal sin voz, la señal se asemeja más a un ruido aleatorio y tiene menor cantidad de predictibilidad.

La Solicitud de Patente de Estados Unidos 2010/070270A da a conocer un método para recibir una señal de audio decodificada que tiene un retardo de altura notal transmitido. El método incluye: la estimación de correlaciones de alturas tonales de posibles retardos de altura tonal débil que son inferiores a una limitación mínima de altura tonal, y tiene una relación aproximada múltiplo con el retardo de altura tonal transmitido, la comprobación de si una de las correlaciones de altura tonal de los posibles retardos de altura tonal débil, es lo suficientemente grande en comparación con una correlación de altura tonal estimada con el retardo de altura tonal transmitido, la selección de un retardo de altura tonal débil como un retardo de altura tonal correcto si una correlación de altura tonal correspondiente es suficientemente grande. El post-procesamiento se realiza utilizando el retardo de altura total corregido. En otra forma de realización, cuando se detecta la existencia de armónicos irregulares o retardo de altura tonal erróneo, un post-filtro de predicción lineal excitada por código (CELP) se hace más agresivo.

SUMARIO DE LA INVENCIÓN

De conformidad con una forma de realización, se divulga un método para la detección y codificación de altura tonal muy débil implementado mediante un aparato para la codificación vocal o audio según una cualquiera de las reivindicaciones 1-14.

De conformidad con otra forma de realización, se divulga un aparato que soporta la detección y codificación de altura tonal muy débil para la codificación de audio o vocal según la reivindicación 15.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

Para un entendimiento más completo de la presente invención, y de sus ventajas, se hace ahora referencia a las descripciones siguientes que se toman en conjunción con los dibujos adjuntos, en las que:

La Figura 1 es un diagrama de bloques de un codificador de la Técnica de Predicción Lineal Excitada por Código (CELP).

La Figura 2 es un diagrama de bloques de un decodificador que corresponde al codificador de CELP de la Figura 1.

La Figura 3 es un diagrama de bloques de otro codificador de CELP con un componente adaptativo.

La Figura 4 es un diagrama de bloques de otro decodificador que corresponde al codificador de CELP de la Figura 3.

La Figura 5 es un ejemplo de una señal vocal con voz, en donde un período de altura tonal es menor que un tamaño de sub-trama y un tamaño de mitad de trama.

La Figura 6 es un ejemplo de una señal vocal con voz, en donde un período de altura tonal es mayor que un tamaño de sub-trama y menor que un tamaño de mitad de trama.

La Figura 7 ilustra un ejemplo de un espectro de una señal vocal con voz.

La Figura 8 ilustra un ejemplo de un espectro de la misma señal ilustrada en la Figura 7 con codificación de retardo

de altura tonal doble.

La Figura 9 ilustra una forma de realización de un método para la detección y codificación de un retardo de altura tonal muy débil para una señal vocal o de voz.

La Figura 10 es un diagrama de bloques de un sistema de procesamiento que puede utilizarse para poner en práctica varias formas de realización.

DESCRIPCIÓN DETALLADA DE FORMAS DE REALIZACIÓN ILUSTRATIVAS

Todos los sucesos siguientes del término "formas de realización", si se refieren a combinaciones de características diferentes de las definidas por las reivindicaciones independientes, se refieren a ejemplos que se presentaron originalmente pero que no representan formas de realización de la invención actualmente reivindicada; estos ejemplos todavía se muestran solamente con fines ilustrativos.

La creación y utilización de las formas de realización actualmente preferidas se describen, en detalle, a continuación. Debe entenderse que, sin embargo, la presente invención da a conocer numerosos conceptos inventivos aplicables que pueden realizarse en una amplia diversidad de contextos específicos. Las formas de realización específicas, aquí descritas, son simplemente ilustrativas de modos específicos de la realización y utilización de la invención, y no limitan el alcance de la idea inventiva.

Para un caso de señal vocal con voz o sin voz, se puede utilizar la codificación paramétrica con el fin de reducir la redundancia de los segmentos vocales mediante la separación de la componente de excitación de la señal vocal del componente de envolvente espectral. La envolvente espectral que cambia lentamente puede representarse por una Codificación de Predicción Lineal (LPC), también denominada Predicción a Corto Plazo (STP). Una codificación vocal de baja tasa binaria podría beneficiarse, también, de una exploración tal como la Predicción a Corto Plazo. La ventaja de la codificación es el resultado de la baja tasa a la que cambian los parámetros. Además, los parámetros de señal de voz pueden no ser muy diferentes de los valores mantenidos en el espacio de unos pocos milisegundos. En la tasa de muestreo de 8 kilohercios (kHz), 12.8 kHz o 16 kHz, el algoritmo de codificación vocal es tal que la duración de la trama nominal está en el margen de diez a treinta milisegundos. Una duración de trama de veinte milisegundos puede ser una elección común. En las normas bien conocidas más recientes, tales como G.723.1, G.729, G.718, EFR, SMV, AMR, VMR-WB o AMR-WB, se ha adoptado una Técnica de Predicción Lineal Excitada por Código (CELP). CELP es una combinación técnica de Predicción a Largo Plazo y Predicción a Corto Plazo de Excitación por Código. La codificación vocal de CELP es un principio algorítmico muy popular en el área de compresión vocal, aunque los detalles de la técnica CELP para diferentes *códecs* podría ser bastante distinta.

La Figura 1 ilustra un ejemplo de un codificador de CELP 100, en donde puede minimizarse un error ponderado 109 entre una señal vocal sintetizada 102 y una señal vocal original 101 mediante la utilización de un método de análisis por síntesis. El codificador de CLP 100 realiza diferentes operaciones o funciones. La función W(z) correspondiente se consigue por un filtro de ponderación de error 110. La función 1/B(z) se consigue mediante un filtro de predicción lineal a largo plazo 105. La función 1/A(z) se consigue mediante un filtro de predicción lineal a corto plazo 103. Una excitación codificada 107, a partir de un bloque de excitación codificada 108, que se denomina también excitación de libro de código fijo, se pone a escala mediante una ganancia G_c 106 antes de pasar a través de los filtros posteriores. Un filtro de predicción lineal a corto plazo 103 se pone en práctica mediante el análisis de la señal original 101 y se representa por un conjunto de coeficientes:

$$A(z) = \sum_{i=1}^{P} 1 + a_i \cdot z^{-i} \quad , i = 1,2,...., P \qquad (1)$$

El filtro de ponderación de error 110 está relacionado con la función de filtro de predicción lineal a corto plazo anterior. Una forma típica de la función de filtro de ponderación podría ser

$$W(z) = \frac{A(z/\alpha)}{1 - \beta \cdot z^{-1}} \quad , \qquad (2)$$

en donde $\beta<\alpha$, $0<\beta<1$ y $0<\alpha \leq 1$. El filtro de predicción lineal a largo plazo 105 depende de la altura tonal de la señal y de su ganancia. Una altura tonal puede estimarse a partir de la señal original, la señal residual o la señal original ponderada. La función de filtro de predicción lineal a largo plazo puede expresarse como

$$W(z) = \frac{A(z/\alpha)}{1 - \beta \cdot z^{-1}} \quad , \qquad (3)$$

La excitación codificada 107, a partir del bloque de excitación codificada 108, puede estar constituida por señales

similares a pulsos o señales similares a ruido, que se construyen matemáticamente o se memorizan en un libro de códigos. Un índice de excitación codificada, un índice de ganancia cuantificada, un índice de parámetro de predicción a largo plazo cuantificado y un índice de parámetro de predicción a corto plazo cuantificado, pueden transmitirse desde el codificador 100 a un decodificador.

La Figura 2 ilustra un ejemplo de un decodificador 200, que puede recibir señales procedentes del codificador 100. El decodificador 200 incluye un bloque de post-procesamiento 207 que proporciona, a la salida, una señal vocal sintetizada 206. El decodificador 200 comprende una combinación de múltiples bloques, que incluyen un bloque de excitación codificada 201, un filtro de predicción lineal a largo plazo 203, un filtro de predicción lineal a corto plazo 205 y un bloque de post-procesamiento 207. Los bloques del decodificador 200 están configurados de forma similar a los bloques correspondientes del codificador 100. El bloque de post-procesamiento 207 puede incluir funciones de post-procesamiento a corto plazo y post-procesamiento a largo plazo.

La Figura 3 ilustra otro codificador de CELP 300 que pone en práctica la predicción lineal a largo plazo utilizando un bloque de libro de códigos adaptativo 307. El bloque de libro de códigos adaptativo 307 utiliza una excitación sintetizada anterior 304, o repite un ciclo de altura tonal de excitación anterior en un período de altura tonal. Los bloques restantes y los componentes del codificador 300 son similares a los bloques y componentes descritos con anterioridad. El codificador 300 puede codificar un retardo de altura tonal en un valor entero cuando el retardo de altura tonal es relativamente grande o largo. El retardo de altura tonal puede codificarse en un valor fraccional más preciso cuando la altura tonal es relativamente pequeña o corta. La información periódica de la altura tonal se utiliza con el fin de generar el componente adaptativo de la excitación (en el bloque de libro de códigos adaptativo 307). Este componente de excitación se pone luego a escala mediante una ganancia G_p 305 (también denominada ganancia de altura tonal). Los dos componentes de excitación puestos a escala, a partir del bloque de libro de códigos adaptativo 307, y del bloque de excitación codificada 308, se añaden juntos antes de pasar a través de un filtro de predicción lineal a corto plazo 303. Las dos ganancias (G_p y G_c) son objeto de cuantización y a continuación, se envían a un decodificador.

La Figura 4 ilustra un decodificador 400, que puede recibir señales a partir del codificador 300. El decodificador 400 incluye un bloque de post-procesamiento 408 que proporciona, a la salida, una señal vocal sintetizada 407. El decodificador 400 es similar al decodificador 200 y los componentes del decodificador 400 pueden ser similares a los correspondientes componentes del decodificador 200. Sin embargo, el decodificador 400 incluye un bloque de libro de códigos adaptativo 307, además de una combinación de otros bloques, que incluyen un bloque de excitación codificada 402, un libro de códigos adaptativo 401, un filtro de predicción lineal a corto plazo 406 y un bloque de post-procesamiento 408. El bloque de post-procesamiento 408 puede incluir funciones de post-procesamiento a corto plazo y post-procesamiento a largo plazo. Otros bloques son similares a los componentes correspondientes en el decodificador 200.

La predicción a largo plazo puede ser utilizada, de forma eficaz, en una codificación vocal de voz, debido a la naturaleza de periodicidad relativamente fuerte de la señal vocal con voz. Los ciclos de altura tonal adyacentes de la señal vocal con voz pueden ser similares entre sí, lo que significa, matemáticamente, que la ganancia de altura tonal G_p en la expresión de excitación siguiente es relativamente alta o próxima a 1,

$$e(n) = G_p \cdot e_p(n) + G_c \cdot e_c(n) \qquad (4)$$

en donde $e_p(n)$ es una sub-trama de series de muestras indexadas por n, y se envía desde el bloque de libro de códigos adaptativo 307 o 401, que utiliza la excitación sintetizada anterior 304 o 403. El parámetro $e_p(n)$ puede ser filtrado de modo adaptativo de paso bajo desde la zona de baja frecuencia que puede ser más periódica o más armónica que la zona de alta frecuencia. El parámetro $e_c(n)$ se envía desde el libro de códigos de excitación codificada 308 o 402 (también denominado libro de códigos fijo), que es una contribución de excitación actual. El parámetro $e_c(n)$ puede mejorarse, además, a modo de ejemplo, utilizando un filtrado de paso alto mejorado, una mejora de altura tonal, mejora de dispersión, mejora de los formantes, etc. Para la señal vocal con voz, la contribución del parámetro $e_p(n)$ procedente del bloque de libro de códigos adaptativo 307 o 401 puede ser dominante y la ganancia de altura tonal G_p 305 o 404 es aproximadamente un valor de 1. La excitación puede actualizarse para cada sub-trama. A modo de ejemplo, un tamaño de trama típico es de aproximadamente 20 milisegundos y un tamaño de sub-trama típico es de aproximadamente 5 milisegundos.

Para señales vocales con voz típicas, una trama puede incluir más de 2 ciclos de altura tonal. La Figura 5 ilustra un ejemplo de una señal vocal con voz 500, en donde un período de altura tonal 503 es menor que un tamaño de sub-trama 502 y un tamaño de mitad de trama 501. La Figura 6 ilustra otro ejemplo de una señal vocal con voz 600, en donde un período de altura tonal 603 es mayor que un tamaño de sub-trama 602 y menor que un tamaño de mitad de trama 601.

La técnica CELP se utiliza para codificar la señal vocal beneficiándose de las características de la voz humana o del modelo de generación de señal vocal humana. El algoritmo de CELP ha sido utilizado en varias normas como ITU-T, MPEG, 3GPP y 3GPP2. Para una codificación más eficiente de señales vocales, dichas señales vocales se pueden

clasificar en diferentes clases, en donde cada clase se codifica de un modo distinto. A modo de ejemplo, en algunas normas tales como G.718, VMR-WB o AMR-WB, las señales vocales se clasifican en clases de señal vocal de UNVOICED, TRANSITION, GENERIC, VOICED y NOISE. Para cada clase, se utiliza un filtro LPC o STP para representar una envolvente espectral, pero la excitación para el filtro LPC puede ser diferente. Las clases UNVOICED y NOISE pueden codificarse con una excitación por ruido y alguna excitación mejorada. La clase de TRANSITION puede codificarse con una excitación por pulsos y alguna excitación mejorada sin utilizar un libro de códigos adaptativo o LTP. La clase GENERIC puede codificarse con una técnica CELP tradicional, tal como una técnica CELP algebraica utilizada en las normas G.729 o AMR-WB, en la que una trama de 20 milisegundos (ms) contiene cuatro sub-tramas de 5 ms. El componente de excitación de libro de códigos adaptativo y el componente de excitación de libro de códigos fijo se generan, ambos, con alguna mejora de excitación para cada sub-trama. Retardos de altura tonal para el libro de códigos adaptativo en la primera y tercera sub-tramas se codifican en un margen completo a partir de un límite de altura tonal mínimo *PIT_MIN* a un límite de altura tonal máximo *PIT_MAX*, y retardos de altura tonal para el libro de códigos adaptativo, en la segunda y cuarta sub-tramas se codifican, de forma distinta del anterior retardo de altura tonal codificado. La clase VOICED se puede codificar, de una forma ligeramente distinta, de la clase GENERIC, en la que el retardo de altura tonal en la primera sub-trama se codifica en un margen completo a partir de un límite de altura tonal mínimo *PIT_MIN* a un límite de altura tonal máximo *PIT_MAX*, y retardos de altura tonal en las otras sub-tramas se codifican, de forma distinta del anterior retardo de altura tonal codificado. A modo de ejemplo, si se supone una tasa de muestreo de excitación de 12.8 kHz, el valor de *PIT_MIN* puede ser 34 y el valor de *PIT_MAX* puede ser 231.

Los *códecs* de CELP (codificadores/decodificadores) funcionan, de forma eficiente, para señales vocales normales, pero *códecs* CELP de baja tasa binaria pueden fallar para señales musicales y señales vocales de canto. Para señales vocales de voz estable, el método de codificación de altura tonal de la clase VOICED puede proporcionar un mejor rendimiento que el método de codificación de altura tonal de la clase GENERIC mediante la reducción de la tasa binaria para codificar retardos de altura tonal con codificación de altura tonal más diferencial. Sin embargo, el método de codificación de altura tonal de la clase VOICED o de la clase GENERIC pueden tener, todavía, un problema de que se degrada el rendimiento o no es suficientemente bueno cuando la altura tonal real es prácticamente o relativamente, muy débil, a modo de ejemplo, cuando el retardo de altura tonal real es menor que *PIT_MIN*. Un margen de altura tonal desde *PIT_MIN*= 34 a *PIT_MAX* =231 para frecuencias de muestreo $F_s = 12.8$ kHz, se puede adaptar para diversas voces humanas. Sin embargo, el retardo de altura tonal real de señales típicas de música o señales vocales de canto, puede ser sustancialmente más corto que la limitación mínima *PIT_MIN = 34* definida en el algoritmo de CELP. Cuando el retardo de altura tonal real es *P*, la frecuencia armónica fundamental correspondiente es $F0=F_s/P$, en donde F_s es la frecuencia de muestreo y *F0* es la localización del primer pico armónico en el espectro. De este modo, la limitación mínima de altura tonal *PIT_MIN* puede definir, realmente, la limitación de frecuencia armónica fundamental máxima $F_{MIN} = F_s/PIT\_MIN$ para el algoritmo de CELP.

La Figura 7 ilustra un ejemplo de un espectro 700 de una señal vocal con voz que comprende picos armónicos 701 y una envolvente espectral 702. La frecuencia armónica fundamental real (la localización del primer pico armónico) supera ya la limitación máxima de frecuencia armónica fundamental F_{MIN} de modo que el retardo de altura tonal transmitido para el algoritmo de CELP es igual a un doble o un múltiplo del retardo de altura tonal real. El retardo de altura tonal incorrecto que se transmite como un múltiplo del retardo de altura tonal real puede hacer que se degrade la calidad. Dicho de otro modo, cuando el retardo de altura tonal real para una señal armónica de música o una señal vocal de canto es menor que la limitación de retardo mínima *PIT_MIN* que se define en el algoritmo de CELP, el retardo transmitido puede ser el doble, el triple o un múltiplo del retardo de altura tonal real. La Figura 8 ilustra un ejemplo de un espectro 800 de la misma señal de codificación de retardo de altura tonal doble (el retardo de altura tonal transmitido y codificado es el doble del retardo de altura tonal real). El espectro 800 incluye picos armónicos 801, una envolvente espectral 802 y picos pequeños no deseados entre los picos armónicos reales. Los pequeños picos del espectro, en la Figura 8, pueden causar una distorsión perceptual no deseada.

Las formas de realización del sistema y método se dan a conocer en este documento con el fin de evitar el problema potencial anterior de la codificación de altura tonal para la clase VOICED o la clase GENERIC. Las formas de realización del sistema y método están configuradas para codificar un retardo de altura tonal en un margen que comienza desde un valor prácticamente corto *PIT_MIN0 (PIT_MIN0 < PIT_MIN)*, que puede estar definido con anterioridad. El sistema y método incluye la detección de si existe, o no, una altura tonal muy débil en una señal vocal o de audio (p.ej., de 4 sub-tramas) con la utilización de una combinación de procedimientos del dominio temporal y del dominio frecuencial, p.ej., utilizando una función de correlación de altura tonal y un análisis del espectro de energía. A la detección de que existe una altura tonal muy débil, se puede determinar, entonces, un valor de altura tonal muy débil en el margen desde *PIT_MIN0* a *PIT_MIN*.

En condiciones normales, las señales armónicas musicales o las señales vocales de canto son más estacionarias que las señales vocales normal. El retardo de altura tonal (o frecuencia fundamental) de una señal vocal normal puede seguir cambiando en el transcurso del tiempo. Sin embargo, el retardo de altura tonal (o frecuencia fundamental) de las señales musicales o señales vocales de canto, pueden cambiar relativamente despacio a través de una duración temporal considerablemente larga. Para un retardo de altura tonal sustancialmente corto, es deseable tener un retardo de altura tonal preciso para la finalidad de una codificación eficiente. El retardo de altura tonal relativamente corto puede cambiar muy lentamente desde una sub-trama a una sub-trama siguiente. Lo que

antecede significa que no se necesita un margen dinámico relativamente largo de codificación de altura tonal cuando el retardo de altura tonal real es sustancialmente corto. En consecuencia, un modo de codificación de altura tonal puede estar configurado para definir alta precisión con un margen dinámico relativamente menor. Este modo de codificación de altura tonal se utiliza para codificar señales de altura tonal, sustancial o relativamente cortas o señales de altura tonal prácticamente estables que tienen una diferencia de altura tonal relativamente pequeña entre una sub-trama anterior y una sub-trama actual.

El margen de altura tonal sustancialmente corto se define a partir de *PIT_MIN0* a *PIT_MIN*. A modo de ejemplo, en la frecuencia de muestreo *Fs* = 12.8 kHz, la definición del margen de altura tonal sustancialmente corto puede ser *PIT_MIN0 = 17* y *PIT_MIN = 34*. Cuando la altura tonal candidato es sustancialmente corta, puede no ser fiable la detección de altura tonal utilizando solamente un método de dominio temporal o de dominio frecuencial. Con el fin de detectar, de forma fiable, un valor de altura tonal débil, puede *ser necesaria la comprobación de tres condiciones: (1) en el dominio frecuencial, la energía desde 0 Hz a* $F_{MIN} = Fs/PIT\_MIN$ Hz es relativamente baja; (2) en el dominio temporal, la correlación de altura tonal máxima en el margen de *PIT_MIN0* a *PIT_MIN* es, relativamente, lo suficientemente alta en comparación con la correlación de altura tonal máxima en el margen de *PIT_MIN* a *PIT_MAX*; y (3) en el dominio temporal, la correlación de altura tonal normalizada máxima en el margen de *PIT_MIN0* a *PIT_MIN* es lo suficientemente alta con referencia en sentido hacia 1. Estas tres condiciones son más importantes que otras condiciones que pueden también añadirse, tales como Detección de Actividad de Voz y Clasificación por Voz.

Para una altura tonal candidato *P*, la correlación de altura tonal normalizada se puede definir en forma matemática como,

$$R(P) = \frac{\sum_n s_w(n) \cdot s_w(n-P)}{\sqrt{\sum_n \|s_w(n)\|^2 \cdot \sum_n \|s_w(n-P)\|^2}} . \qquad (5)$$

En la ecuación (5), $s_w(n)$ es una señal vocal ponderada, el numerador es la correlación, y el denominador es un factor de normalización de la energía. Suponiendo que *Voicing* sea el valor de correlación de altura tonal normalizada media de las cuatro sub-tramas, en la trama actual:

$$Voicing = [\, R_1(P_1) + R_2(P_2) + R_3(P_3) + R_4(P_4) \,] / 4 \qquad (6)$$

en donde $R_1(P_1)$, $R_2(P_2)$, $R_3(P_3)$ y $R_4(P_4)$, son las cuatro correlaciones de altura tonal normalizadas que se calculan para cada sub-trama y siendo P_1, P_2, P_3 y P_4, para cada sub-trama, las mejores candidatas de altura tonal encontradas en el margen de altura tonal desde *P = PIT_MIN* a *P = PIT_MAX*. La correlación de altura tonal de magnitud limitada desde la trama anterior a la trama actual puede ser

$$Voicing\_sm \Leftarrow (3 \cdot Voicing\_sm + Voicing)/4 . \qquad (7)$$

Utilizando un sistema de detección de altura tonal de bucle abierto, la altura tonal candidato puede ser una altura tonal múltiple. Si la altura tonal de bucle abierto es la correcta, existe un pico de espectro alrededor de la frecuencia de altura tonal correspondiente (la frecuencia fundamental o la primera frecuencia armónica) y la energía del espectro relacionada es relativamente grande. Además, la energía media entorno a la frecuencia de altura tonal correspondiente es relativamente grande. De no ser así, es posible que exista una altura tonal sustancialmente corta. Esta etapa puede combinarse con un sistema de detección de falta de energía de baja frecuencia, que se describe a continuación con el fin de detectar la posible altura tonal sustancialmente corta.

En el sistema para detectar la falta de energía de baja frecuencia, la energía máxima en la zona de frecuencia [*0,* F_{MIN}] (Hz) se define como *Energy0* (dB), la energía máxima en la zona de frecuencia [F_{MIN}, 900] (Hz) se define como *Energy1* (dB), y la relación de energía relativa entre *Energy0* y *Energy1* se define como

$$Ratio = Energy1 - Energy0 . \qquad (8)$$

Esta relación de energía puede ser ponderada multiplicando un valor de correlación de altura tonal normalizada media Voicing:

$$Ratio \Leftarrow Ratio \cdot Voicing. \qquad (9)$$

El motivo para realizar la ponderación en la ecuación (9) utilizando el factor *Voicing* es que la detección de altura tonal débil es significativa para la señal vocal de voz o la música armónica, pero puede no ser significativa para la señal vocal sin voz o la música no armónica. Antes de utilizar el parámetro *Ratio* para detectar la falta de energía de baja frecuencia, resulta ventajoso limitar la magnitud del parámetro *Ratio* con el fin de reducir la incertidumbre:

$$LF\_EnergyRatio\_sm \Leftarrow (15 \cdot LF\_EnergyRatio\_sm + Ratio)/16 . \quad (10)$$

Suponiendo que *LF_lack_flag=1* designa que se detecta la falta de energía de baja frecuencia (de no ser así *LF_lack_flag=0*), el valor *LF_lack_flag* puede determinarse mediante el siguiente procedimiento A:

Si (LF_EnergyRatio_sm>35 o Ratio>50) {
LF_lack_flag=1;
}
Si (LF_EnergyRatio_sm<16) {
LF_lack_flag=0;
}

Si las condiciones anteriores no se satisfacen, *LF_lack_flag* se mantiene invariable.

Se puede encontrar una altura tonal débil candidato inicial *Pitch_Tp* maximizando la ecuación (5) y buscando desde *P=PIT_MIN0* a *PIT_MIN,*

$$R(Pitch\_Tp) = MAX\{ R(P), P = PIT\_MIN0, ..., PIT\_MIN \}. \quad (11)$$

Si *Voicing0* representa la correlación de altura tonal débil actual,

$$Voicing0 = R(Pitch\_Tp). \quad (12)$$

entonces, la correlación de altura tonal débil, de magnitud limitada, desde la trama anterior a la trama actual puede ser

$$Voicing0\_sm \Leftarrow (3 \cdot Voicing0\_sm + Voicing0)/4 \quad (13)$$

Utilizando los parámetros disponibles con anterioridad, se puede decidir el retardo de altura tonal final sustancialmente corto con el procedimiento B siguiente:

Si ((coder_type no es UNVOICED o TRANSITION) y
(LF_lack_flag=1) y (VAD=1) y
(Voicing0_sm>0.7) y (Voicing0_sm>0.7 Voicing_sm))
{
Open_Loop_Pitch = Pitch_Tp;
stab_pit_flag = 1;
coder_type = VOICED;
}

En el procedimiento anterior, *VAD* significa *Detección de Actividad de Voz.*

La Figura 9 ilustra una forma de realización de un método 900 para la detección y codificación de retardo de altura tonal muy débil para una señal vocal o de audio. El método 900 puede ponerse en práctica por un codificador para la codificación vocal/audio tal como el codificador 300 (o 100). Un método similar puede ponerse en práctica también por un decodificador para la codificación de señal vocal/audio, tal como el decodificador 400 (o 200). En la etapa 901, se clasifica una señal vocal o de audio, o trama, que incluye 4 sub-tramas, a modo de ejemplo, para la clase VOICED o GENERIC. En la etapa 902, se calcula una correlación de altura tonal normalizada *R(P)* para una altura tonal candidato *P*, p.ej., utilizando la ecuación (5). En la etapa 903, se calcula una correlación de altura tonal normalizada media *Voicing*, p.ej., utilizando la ecuación (6). En la etapa 904, se calcula una correlación de altura tonal de magnitud limitada *Voicing_sm*, p.ej., utilizando la ecuación (7). En la etapa 905, se detecta una energía máxima *Energy0* en la zona de la frecuencia [*0, F_{MIN}*]. En la etapa 906, se detecta una energía máxima *Energy1* en la zona de la frecuencia [*F_{MIN},* 900], a modo de ejemplo. En la etapa 907, se calcula una relación de energía *Ratio* entre los valores *Energy1* y *Energy0*, p.ej., utilizando la ecuación (8). En la etapa 908, se ajusta la relación *Ratio* utilizando la correlación de altura tonal normalizada media *Voicing* p.ej., utilizando la ecuación (9). En la etapa 909, se calcula una relación de magnitud limitada *LF_EnergyRatio_sm* p.ej., utilizando la ecuación (10). En la etapa 910, se calcula una correlación *Voicing0* para una altura tonal inicial muy débil *Pitch_Tp*, p.ej., utilizando las ecuaciones (11) y (12). En la etapa 911, se calcula una correlación de altura tonal débil de magnitud limitada *Voicing0_sm* p.ej., utilizando la ecuación (13). En la etapa 912, se calcula una altura tonal final muy débil, p.ej., utilizando los procedimientos A y B.

La Relación de Señal a Ruido (SNR) es uno de los métodos de medición de prueba objetivo para la codificación vocal. La relación SNR Segmental Ponderada (WsegSNR) es otro método de medición de prueba objetivo, que puede ser ligeramente más próximo a la medición real de la calidad perceptual que la relación SNR. Puede no ser audible una diferencia relativamente pequeña en SNR o WsegSNR, mientras que las diferencias más grandes en SNR o WsegSNR pueden ser más o claramente audibles. Las tablas 1 y 2 ilustran el hecho de que la introducción

de una codificación de retardo de altura tonal muy débil puede mejorar, de forma significativa, la calidad de codificación de música o vocal cuando la señal contiene un retardo de altura tonal real muy débil. Los resultados de prueba adicional de audición ilustran que se mejora, de forma significativa, la calidad vocal o musical con un retardo de altura tonal real <= *PIT_MIN* después de la utilización de las etapas y métodos anteriores.

Tabla 1: Relación SNR para señal vocal limpia con retardo de altura tonal real <= *PIT_MIN*.

	6.8 kbps	7.6 kbps	9.2 kbps	12.8 kbps	16 kbps
Sin altura tonal débil	5.241	5.865	6.792	7.974	9.223
Con altura tonal débil	5.732	6.424	7.272	8.332	9.481
Diferencia	0.491	0.559	0.480	0.358	0.258

Tabla 2: Relación WsegSNR para señal vocal limpia con retardo de altura tonal real <= *PIT_MIN*.

	6.8 kbps	7.6 kbps	9.2 kbps	12.8 kbps	16 kbps
Sin altura tonal débil	6.073	6.593	7.719	9.032	10.257
Con altura tonal débil	6.591	7.303	8.184	9.407	10.511
Diferencia	0.528	0.710	0.465	0.365	0.254

La Figura 10 es un diagrama de bloques de un aparato o sistema de procesamiento 1000 que puede utilizarse para poner en práctica varias formas de realización. A modo de ejemplo, el sistema de procesamiento 1000 puede ser parte de, o acoplarse a, un componente de red, tal como un enrutador, un servidor, o cualquier otro componente de red o aparato. Dispositivos específicos pueden utilizar la totalidad de los componentes ilustrados, o solamente un subconjunto de los componentes, y los niveles de integración pueden variar de un dispositivo a otro. Además, un dispositivo puede incluir múltiples instancias operativas de un componente, tal como múltiples unidades de procesamiento, procesadores, memorias, transmisores, receptores, etc. El sistema de procesamiento 1000 puede incluir una unidad de procesamiento 1001 provista con uno o más dispositivos de entrada/salida, tal como un altavoz, micrófono, ratón, pantalla táctil, teclado numérico, teclado, impresora, pantalla, etc. La unidad de procesamiento 1001 puede incluir una unidad central de procesamiento (CPU) 1010, una memoria 1020, un dispositivo de almacenamiento masivo 1030, un adaptador de vídeo 1040, y una interfaz de I/O (entrada/salida) 1060 que se conecta a un bus. El bus puede ser uno o más de cualquier tipo de varias arquitecturas de bus, que incluyen un bus de memoria o un controlador de memoria, un bus periférico, un bus de vídeo, o similar.

La unidad CPU 1010 puede incluir cualquier tipo de procesador de datos electrónico. La memoria 1020 puede comprender cualquier tipo de memoria del sistema, tal como una memoria de acceso aleatorio estática (SRAM), una memoria de acceso aleatorio dinámica (DRAM), una memoria DRAM síncrona (SDRAM), una memoria de solamente lectura (ROM), una de sus combinaciones, etc. En una forma de realización, la memoria 1020 puede incluir una memoria ROM para su uso durante el arranque, y una memoria DRAM para memorizar programas y datos para uso mientras se ejecutan dichos programas. En formas de realización, la memoria 1020 es una memoria no transitoria. El dispositivo de almacenamiento masivo 1030 puede incluir cualquier tipo de dispositivo de almacenamiento configurado para memorizar datos, programas y otra información y para hacer que los datos, los programas y otra información sean accesibles a través de un bus. El dispositivo de almacenamiento masivo 1030 puede incluir, a modo de ejemplo, uno o más de entre una unidad de estado sólido, una unidad de disco duro, una unidad de disco magnético, una unidad de disco óptico, o similar.

El adaptador de vídeo 1040 y la interfaz de I/O (entrada/salida) 1060 proporcionan interfaces con el fin de acoplar, de forma externa, dispositivos de entrada y salida a la unidad de procesamiento. Tal como se ilustra, ejemplos de dispositivos de entrada y salida incluyen una pantalla de visualización 1090 acoplada al adaptador de vídeo 1040 y cualquier combinación de ratón/teclado/impresora 1070 que se acopla a la interfaz de entrada/salida (I/O) 1060. Otros dispositivos pueden acoplarse a la unidad de procesamiento 1001, y se pueden utilizar menos, o adicionales tarjetas de interfaz. A modo de ejemplo, una tarjeta de interfaz serie (no ilustrada) puede utilizarse para proporcionar una interfaz serie para una impresora.

La unidad de procesamiento 1001 incluye, además, una o más interfaces de red 1050, que puede incluir enlaces cableados, tal como un cable de Ethernet o similar, y/o enlaces inalámbricos para acceder a nodos o una o más redes 1080. La interfaz de red 1050 permite a la unidad de procesamiento 1001 su comunicación con unidades distantes a través de las redes 1080. A modo de ejemplo, la interfaz de red 1050 puede proporcionar comunicación inalámbrica, a través de uno o más transmisores/antenas de transmisión y uno o más receptores/antenas de recepción. En una forma de realización, la unidad de procesamiento 1001 está acoplada a una red de área local o una red de área amplia para el procesamiento de datos y comunicaciones con dispositivos distantes, tales como otras unidades de procesamiento, la red Internet, instalaciones de almacenamiento distantes, etc.

Aunque esta invención ha sido descrita haciendo referencia a las formas de realización ilustrativas, la presente descripción no está prevista para crearse en un sentido limitativo. Varias modificaciones y combinaciones de las formas de realización ilustrativas, así como otras formas de realización de la invención, serán evidentes para los expertos en esta técnica, con referencia a la descripción. Por lo tanto, está previsto que las reivindicaciones adjuntas abarquen cualesquiera de dichas modificaciones o formas de realización.

REIVINDICACIONES

1. Un método para la detección y codificación de altura tonal muy débil implementado mediante un aparato para una codificación vocal o audio, comprendiendo dicho método:

detectar en una señal vocal o de audio un retardo de altura tonal muy débil, que está en un margen desde una limitación de altura tonal muy débil mínima a una limitación de altura tonal mínima convencional PIT_MIN, que se define mediante un algoritmo predeterminado de Técnica de Predicción Lineal Excitada por Código (CELP), utilizando una combinación de técnicas de detección de altura tonal de dominio temporal y dominio frecuencial que incluyen el utilización de la correlación de altura tonal y la detección de una falta de energía de baja frecuencia, en donde, la limitación de altura tonal muy débil mínima es menor que la limitación de PIT_MIN;
el método está caracterizado por cuanto que comprende, además:
codificación del retardo de altura tonal muy débil;
en donde la detección de una falta de energía de baja frecuencia comprende:
calcular (907) una relación de energía como
Ratio = Energy1 – Energy0,

en donde *Ratio* es la relación de energía, *Energy0* es la energía máxima en decibelios (dB) en una primera zona de frecuencia [*0, F_{MIN}*] Hertz (Hz), *Energy1* es la energía máxima en dB en una segunda zona de frecuencia [*F_{MIN},* 900] Hz, y *F_{MIN}* es una frecuencia mínima predeterminada;

ponderar (908) la relación de energía usando la correlación de altura tonal normalizada media como

$$Ratio \quad = \quad Ratio \cdot Voicing \ ;$$

en donde *Ratio*, en el lado derecho de la ecuación, representa la relación de energía que ha de ajustarse; *Ratio,* en el lado izquierdo de la ecuación, representa la relación de energía ajustada; y *Voicing* representa la correlación de altura tonal normalizada media;

calcular (909) una relación de energía de magnitud limitada utilizando la relación de energía como:

$$LF\_EnergyRatio\_sm = (15 \cdot LF\_EnergyRatio\_sm + Ratio)/16$$

en donde *LF_EnergyRatio_sm*, en el lado izquierdo de la ecuación, representa la relación de energía de magnitud limitada y *Ratio* representa la relación de energía ajustada;

determinar que se detecta la falta de energía de baja frecuencia si la relación de energía ajustada es mayor que un primer valor umbral predeterminado o si la relación de energía de magnitud limitada es mayor que un secundo valor umbral predeterminado.

2. El método según la reivindicación 1, en donde la detección del retardo de altura tonal muy débil, utilizando la combinación de técnicas de detección de altura tonal de dominio temporal y de dominio frecuencial, comprende:

calcular (902) una correlación de altura tonal normalizada, utilizando una altura tonal candidato y un valor ponderado para la señal vocal o de audio;

calcular (903) la correlación de altura tonal normalizada media *Voicing* utilizando la correlación de altura tonal normalizada; y

calcular (904) una correlación de altura tonal de magnitud limitada de la correlación de altura tonal normalizada.

3. El método según la reivindicación 2, en donde el cálculo de la correlación de altura tonal normalizada utilizando una altura tonal candidato y el valor ponderado para la señal vocal o de audio, comprende:

calcular la correlación de altura tonal normalizada como

$$R(P) \quad = \quad \frac{\sum_n s_w(n) \cdot s_w(n-P)}{\sqrt{\sum_n \left\| s_w(n) \right\|^2 \cdot \sum_n \left\| s_w(n-P) \right\|^2}},$$

en donde *R(P)* es la correlación de altura tonal normalizada, *P* es la altura tonal candidato, y *$s_w(n)$* es un valor ponderado de la señal vocal.

4. El método según cualquiera de las reivindicaciones 2 o 3, en donde $R_1(P_1)$, $R_2(P_2)$, $R_3(P_3)$ y $R_4(P_4)$, son cuatro correlaciones de altura tonal normalizadas que se calculan para cuatro sub-tramas respectivas en una trama actual de la señal vocal o de audio, y P_1, P_2, P_3 y P_4, son cuatro alturas tonales candidatos que se encuentran en un margen de altura tonal desde *PIT_MIN* a una altura tonal limitada máxima *PIT_MAX* que se define por el algoritmo de CELP predeterminado ;
en donde el cálculo de la correlación de altura tonal normalizada media, utilizando la correlación de altura tonal normalizada, comprende:

calcular la correlación de altura tonal normalizada media como

$$Voicing = [\ R_1(P_1) + R_2(P_2) + R_3(P_3) + R_4(P_4)\ ] / 4\ ,$$

en donde *Voicing* es la correlación de altura tonal normalizada media.

5. El método según cualquiera de las reivindicaciones 1, 2 a 4 en donde la detección del retardo de altura tonal muy débil, utilizando la combinación de técnicas de detección de altura tonal de dominio temporal y de dominio frecuencial comprende, además:

calcular una correlación de altura tonal de magnitud limitada como :
Voicing_sm = (3•Voicing_sm + Voicing)/4;

en donde *Voicing_sm*, en el lado izquierdo de la ecuación, es la correlación de altura tonal de magnitud limitada de la trama actual, *Voicing_sm* en el lado derecho de la ecuación es la correlación de altura tonal de magnitud limitada de la trama anterior.

6. El método según cualquiera de las reivindicaciones 2 a 5, en donde la detección del retardo de altura tonal muy débil, utilizando la combinación de técnicas de detección de altura tonal de dominio temporal y de dominio frecuencial, comprende, además:

calcular (910) una correlación para un retardo de altura tonal inicial muy débil; y

calcular (911) una correlación de altura tonal débil de magnitud limitada utilizando la correlación para el retardo de altura tonal inicial muy débil.

7. El método según la reivindicación 6, en donde el retardo de altura tonal muy débil inicial se encuentra como

$$R(Pitch\_Tp) = MAX\{\ R(P),\ P = PIT\_MIN0, \ldots., PIT\_MIN\ \},$$

en donde *Pitch_Tp* es el retardo de altura tonal inicial muy débil, *PIT_MIN0* es la limitación mínima predeterminada de altura tonal muy débil; y

la correlación para el retardo de altura tonal inicial muy débil se representa como:

$$Voicing0 = R(Pitch\_Tp),$$

en donde *Voicing0* es la correlación para el retardo de altura tonal inicial muy débil.

8. El método según la reivindicación 7, en donde el cálculo de una correlación de altura tonal débil de magnitud limitada, utilizando la correlación para el retardo de altura tonal inicial muy débil, comprende:

calcular una correlación de altura tonal débil de magnitud limitada usando la correlación para el retardo de altura tonal inicial muy débil como:

$$Voicing\ 0\_sm = (3 \cdot Voicing\ 0\_sm + Voicing\ 0) / 4\ ;$$

en donde *Voicing0_sm*, en el lado izquierdo de la ecuación, es la correlación de altura tonal débil de magnitud limitada de una trama actual, *Voicing0_sm*, en el lado derecho de la ecuación, es la correlación de altura tonal débil de magnitud limitada de una trama anterior.

9. El método según las reivindicaciones 6 a 8, en donde la detección del retardo de altura tonal muy débil, utilizando la combinación de técnicas de dominio temporal y de dominio frecuencial comprende, además:

decidir (912) el retardo de altura tonal muy débil de conformidad con las condiciones que comprenden:

se detecta la falta de energía de baja frecuencia;

la correlación de altura tonal débil de magnitud limitada es mayor que un tercer umbral predeterminado; y

la correlación de altura tonal débil de magnitud limitada mayor que una multiplicación de un producto de un cuarto umbral predeterminado y la correlación de altura tonal de magnitud limitada.

10. El método según cualquiera de las reivindicaciones 1 a 9, en donde la limitación convencional de altura tonal mínima *PIT_MIN* es igual a 34 para una frecuencia de muestreo de 12.8 kilohercios (kHz).

11. El método según cualquiera de las reivindicaciones 1 a 9, en donde la limitación de altura tonal muy débil mínima es igual a 17 para una frecuencia de muestreo de 12.8 kilohercios (kHz).

12. El método según cualquiera de las reivindicaciones 1 a 9, en donde el primer valor umbral predeterminado es 50 y el segundo valor umbral predeterminado es 35.

13. El método según la reivindicación 9, en donde el cuarto valor umbral predeterminado es 0,7.

14. El método según la reivindicación 1, en donde la limitación convencional de altura tonal mínima *PIT_MIN* define la limitación de frecuencia armónica fundamental máxima $F_{MIN} = F_s/PIT\_MIN$ para el algoritmo de CELP.

15. Un aparato que soporta la detección y codificación de altura tonal muy débil para una codificación vocal o de audio, que comprende:

un procesador; y

un soporte de memorización legible por ordenador que memoriza la programación para su ejecución por el procesador, de los programas que incluyen instrucciones para poner en práctica el método de conformidad con cualquiera de las reivindicaciones 1 a 14.

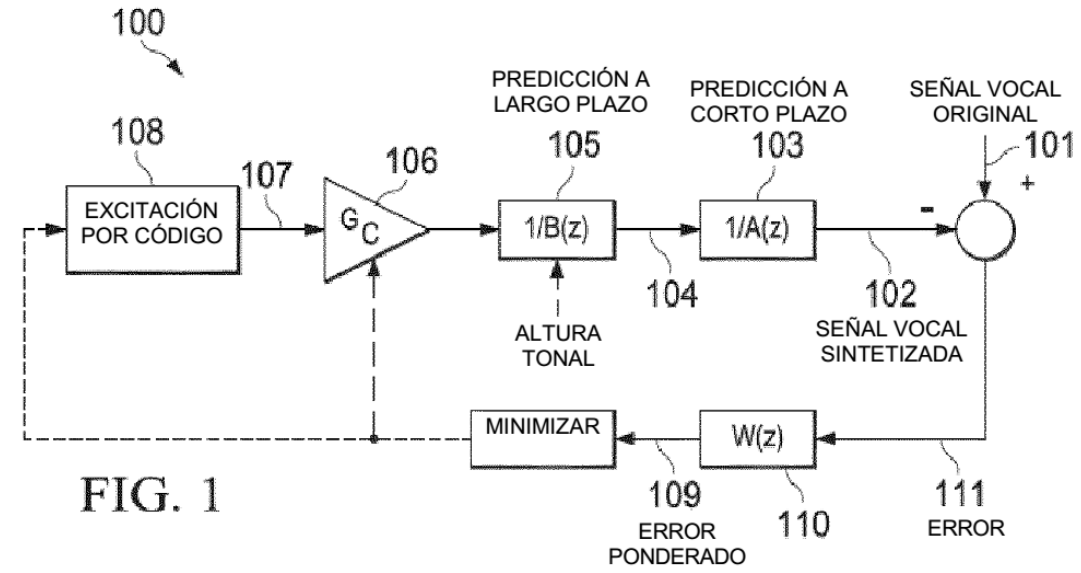

100
108
EXCITACIÓN POR CÓDIGO
107
106
G_C
PREDICCIÓN A LARGO PLAZO
105
1/B(z)
ALTURA TONAL
104
PREDICCIÓN A CORTO PLAZO
103
1/A(z)
102
SEÑAL VOCAL SINTETIZADA
SEÑAL VOCAL ORIGINAL
101
+
-
MINIMIZAR
W(z)
109
ERROR PONDERADO
110
111
ERROR

FIG. 1

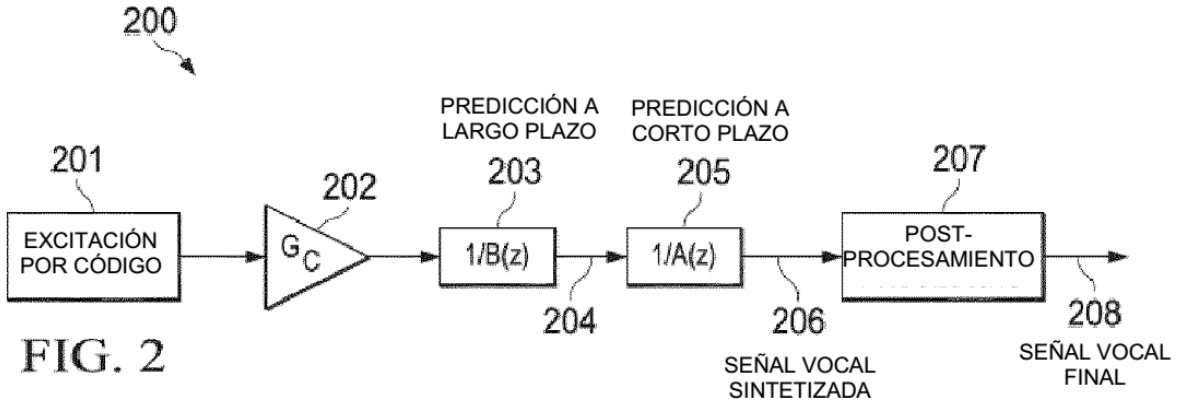

200
201
EXCITACIÓN POR CÓDIGO
202
G_C
PREDICCIÓN A LARGO PLAZO
203
1/B(z)
204
PREDICCIÓN A CORTO PLAZO
205
1/A(z)
206
SEÑAL VOCAL SINTETIZADA
207
POST-PROCESAMIENTO
208
SEÑAL VOCAL FINAL

FIG. 2

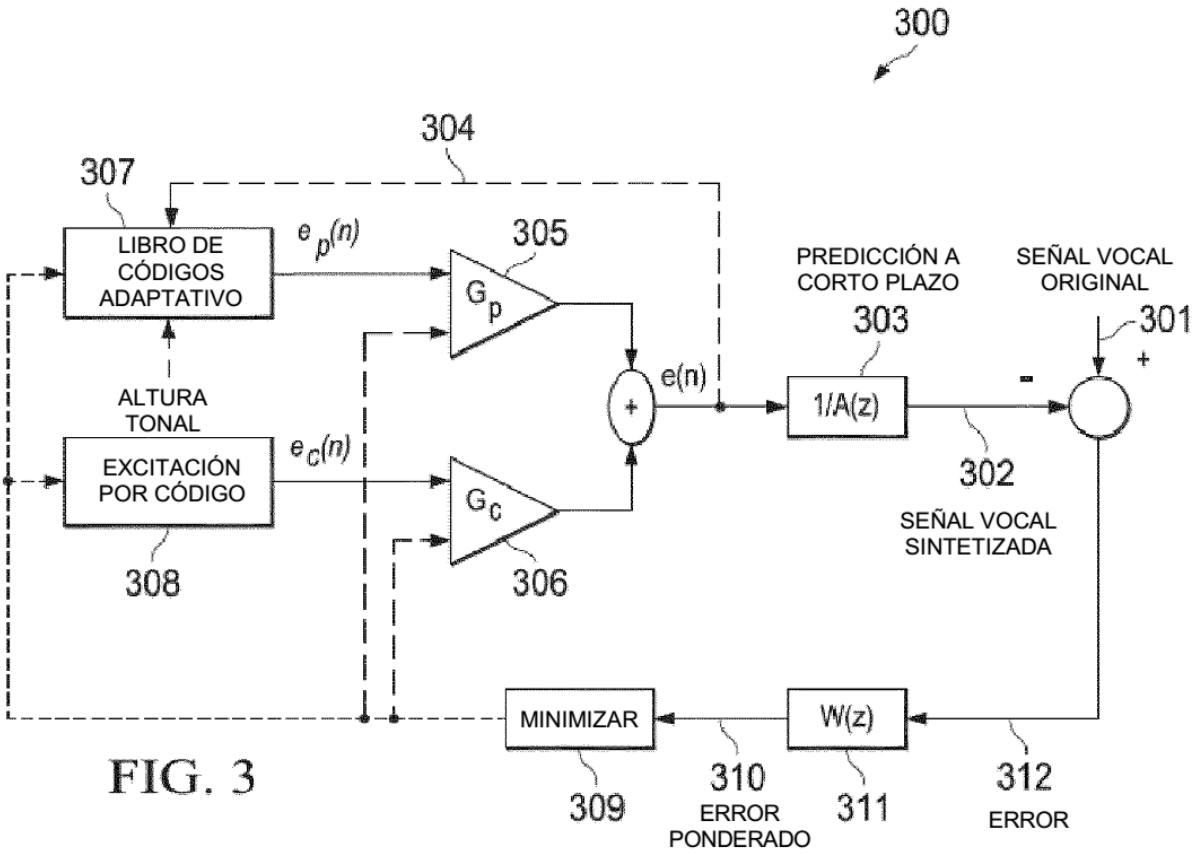

300
304
307
LIBRO DE CÓDIGOS ADAPTATIVO
$e_p(n)$
305
G_p
PREDICCIÓN A CORTO PLAZO
303
SEÑAL VOCAL ORIGINAL
301
+
-
e(n)
+
1/A(z)
ALTURA TONAL
EXCITACIÓN POR CÓDIGO
$e_c(n)$
G_c
302
SEÑAL VOCAL SINTETIZADA
308
306
MINIMIZAR
W(z)
309
310
ERROR PONDERADO
311
312
ERROR

FIG. 3

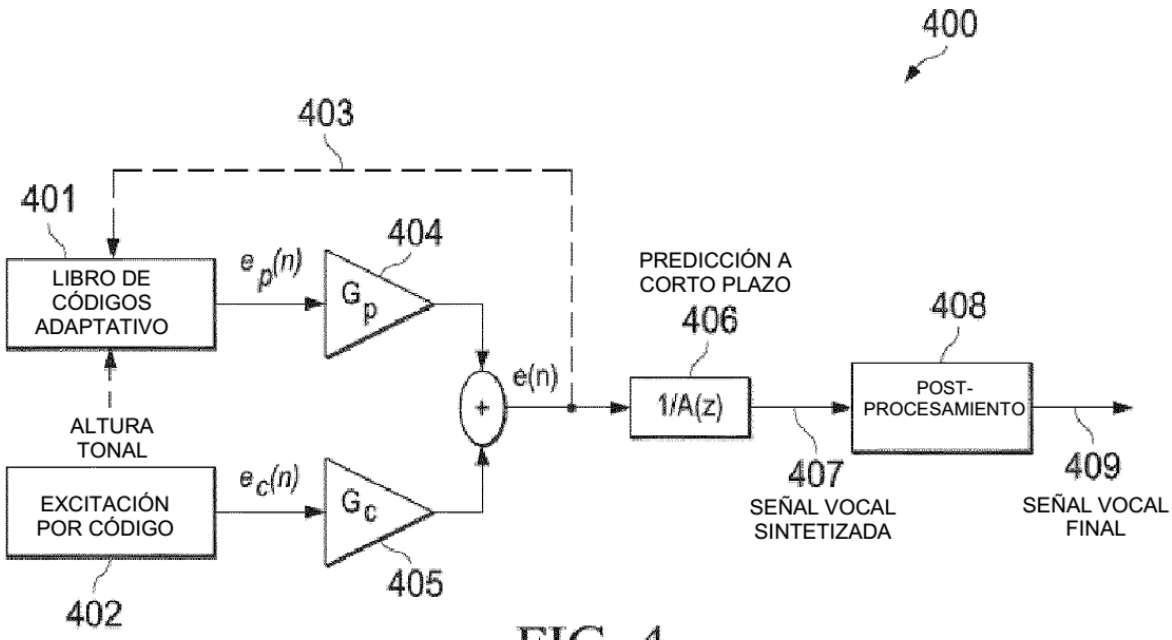

400
403
401
LIBRO DE CÓDIGOS ADAPTATIVO
$e_p(n)$
404
G_p
PREDICCIÓN A CORTO PLAZO
406
408
e(n)
+
1/A(z)
POST-PROCESAMIENTO
ALTURA TONAL
EXCITACIÓN POR CÓDIGO
$e_c(n)$
G_c
407
SEÑAL VOCAL SINTETIZADA
409
SEÑAL VOCAL FINAL
402
405

FIG. 4

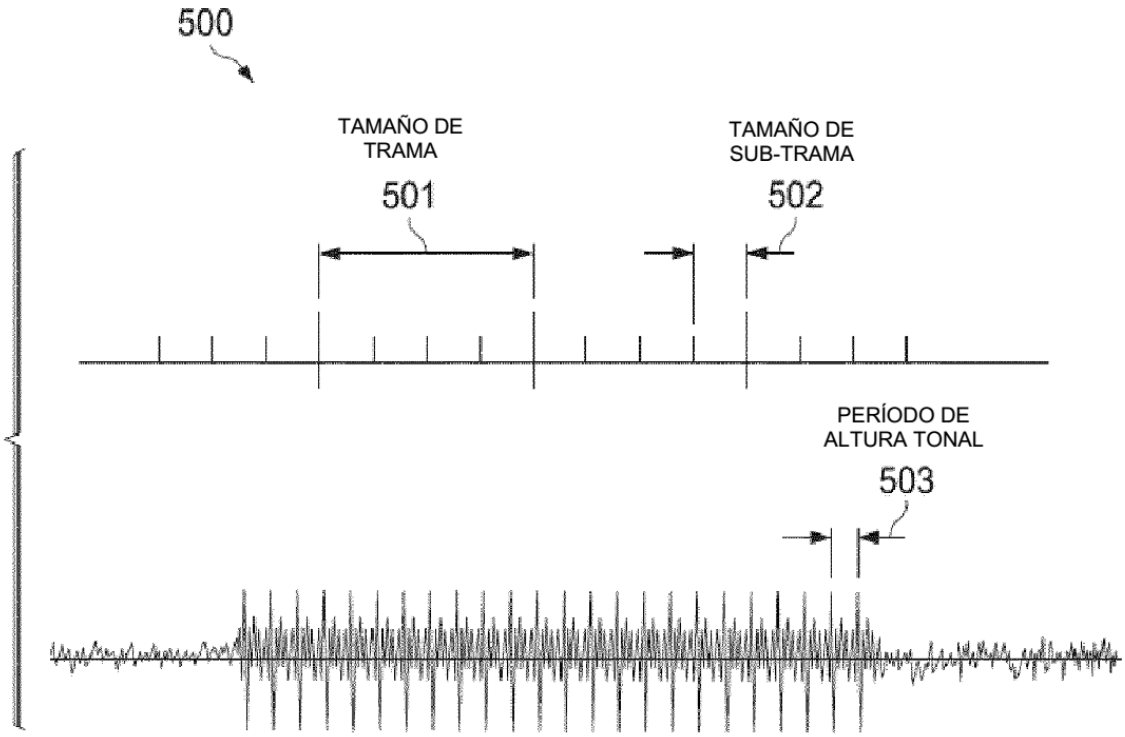

500
TAMAÑO DE TRAMA
501
TAMAÑO DE SUB-TRAMA
502
PERÍODO DE ALTURA TONAL
503

FIG. 5

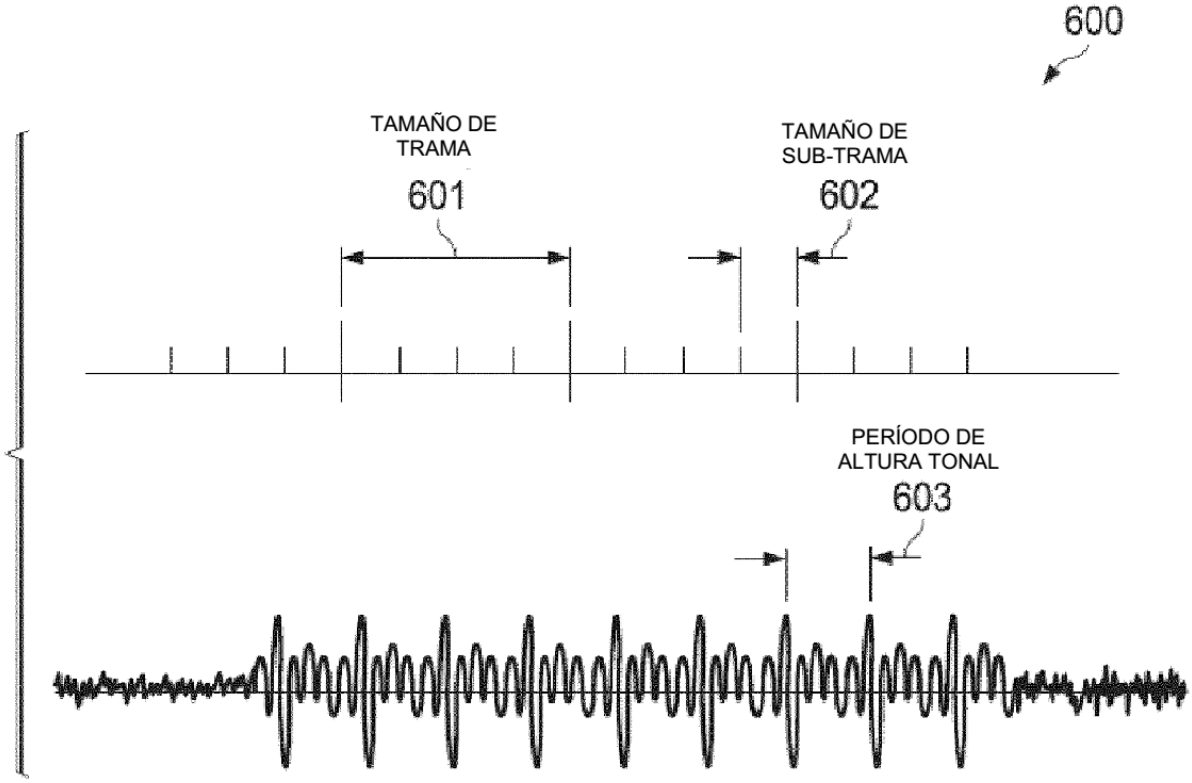
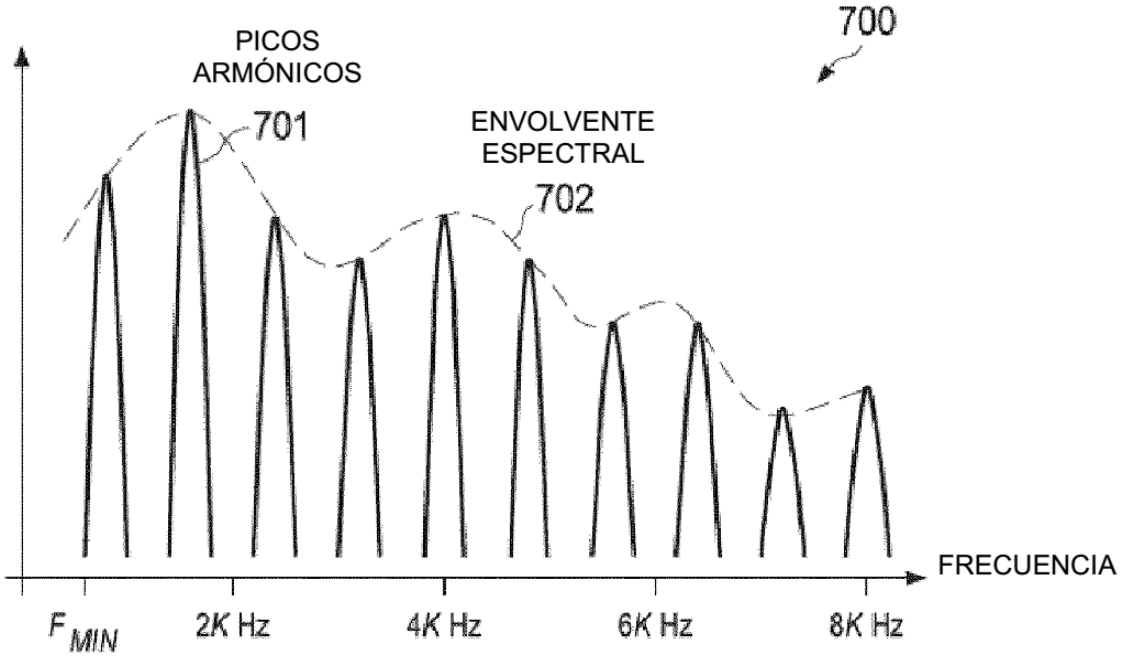
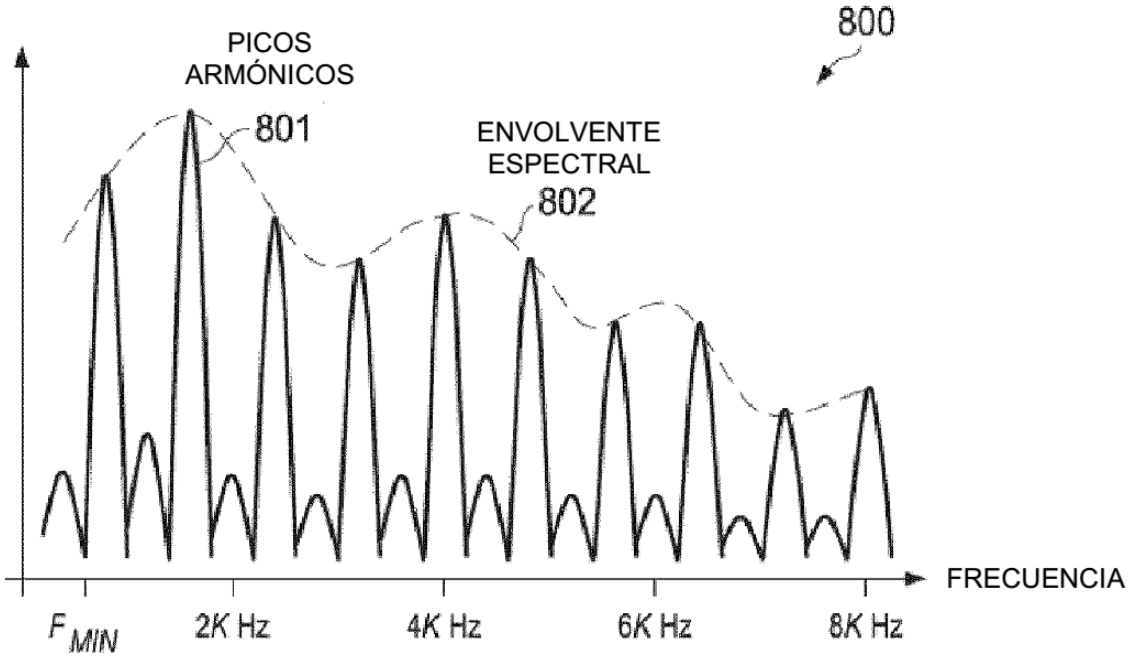


FIG. 6



EJEMPLO DE ESPECTRO DE BANDA ANCHA

FIG. 7



EJEMPLO DE UN ESPECTRO DE BANDA ANCHA REGULAR CON CODIFICACIÓN DE RETARDO DE ALTURA TONAL DOBLADA

FIG. 8

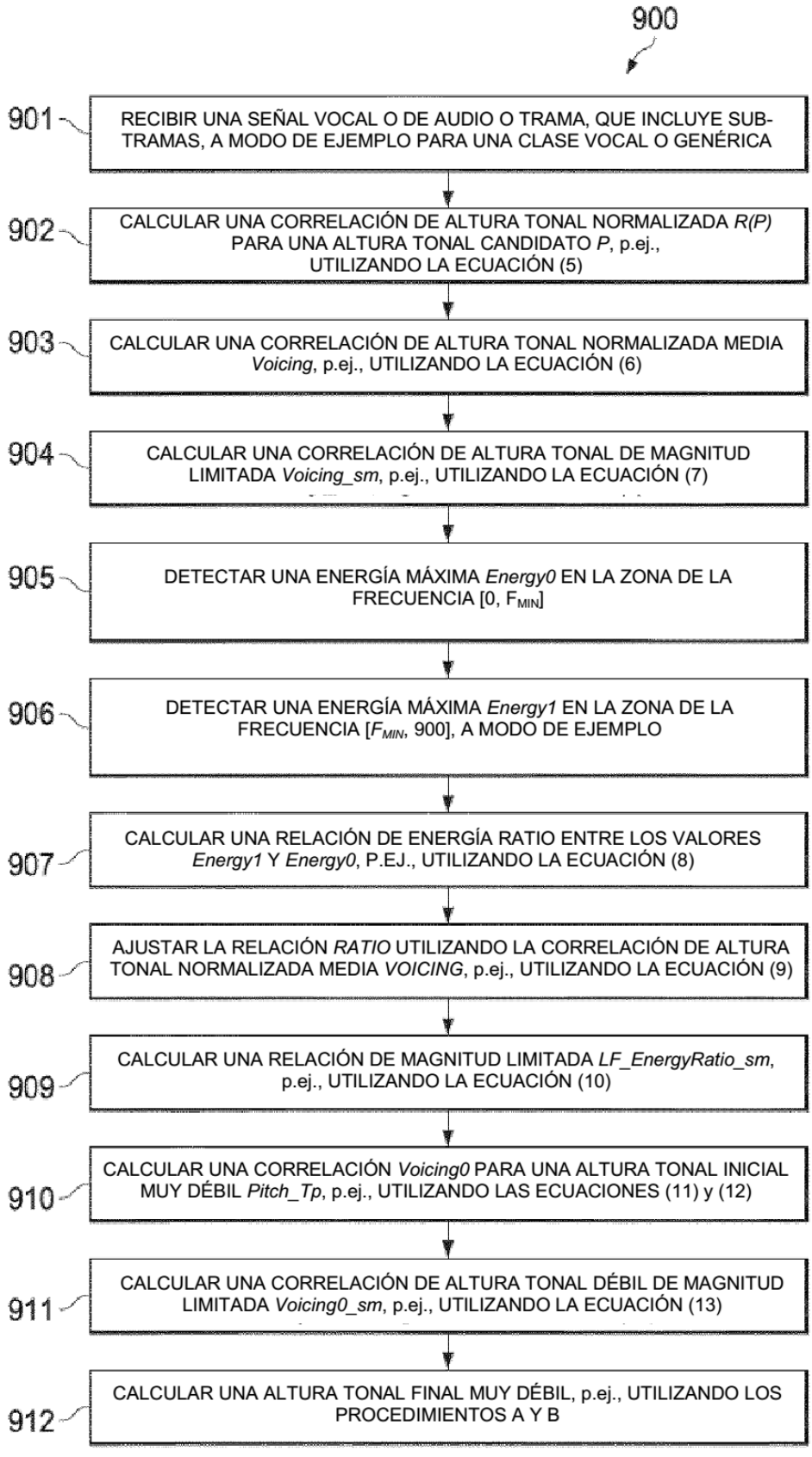

900
901
RECIBIR UNA SEÑAL VOCAL O DE AUDIO O TRAMA, QUE INCLUYE SUB-TRAMAS, A MODO DE EJEMPLO PARA UNA CLASE VOCAL O GENÉRICA
902
CALCULAR UNA CORRELACIÓN DE ALTURA TONAL NORMALIZADA *R(P)* PARA UNA ALTURA TONAL CANDIDATO *P*, p.ej., UTILIZANDO LA ECUACIÓN (5)
903
CALCULAR UNA CORRELACIÓN DE ALTURA TONAL NORMALIZADA MEDIA *Voicing*, p.ej., UTILIZANDO LA ECUACIÓN (6)
904
CALCULAR UNA CORRELACIÓN DE ALTURA TONAL DE MAGNITUD LIMITADA *Voicing_sm*, p.ej., UTILIZANDO LA ECUACIÓN (7)
905
DETECTAR UNA ENERGÍA MÁXIMA *Energy0* EN LA ZONA DE LA FRECUENCIA [0, F_{MIN}]
906
DETECTAR UNA ENERGÍA MÁXIMA *Energy1* EN LA ZONA DE LA FRECUENCIA [F_{MIN}, 900], A MODO DE EJEMPLO
907
CALCULAR UNA RELACIÓN DE ENERGÍA RATIO ENTRE LOS VALORES *Energy1* Y *Energy0*, P.EJ., UTILIZANDO LA ECUACIÓN (8)
908
AJUSTAR LA RELACIÓN *RATIO* UTILIZANDO LA CORRELACIÓN DE ALTURA TONAL NORMALIZADA MEDIA *VOICING*, p.ej., UTILIZANDO LA ECUACIÓN (9)
909
CALCULAR UNA RELACIÓN DE MAGNITUD LIMITADA *LF_EnergyRatio_sm*, p.ej., UTILIZANDO LA ECUACIÓN (10)
910
CALCULAR UNA CORRELACIÓN *Voicing0* PARA UNA ALTURA TONAL INICIAL MUY DÉBIL *Pitch_Tp*, p.ej., UTILIZANDO LAS ECUACIONES (11) y (12)
911
CALCULAR UNA CORRELACIÓN DE ALTURA TONAL DÉBIL DE MAGNITUD LIMITADA *Voicing0_sm*, p.ej., UTILIZANDO LA ECUACIÓN (13)
912
CALCULAR UNA ALTURA TONAL FINAL MUY DÉBIL, p.ej., UTILIZANDO LOS PROCEDIMIENTOS A Y B

FIG. 9

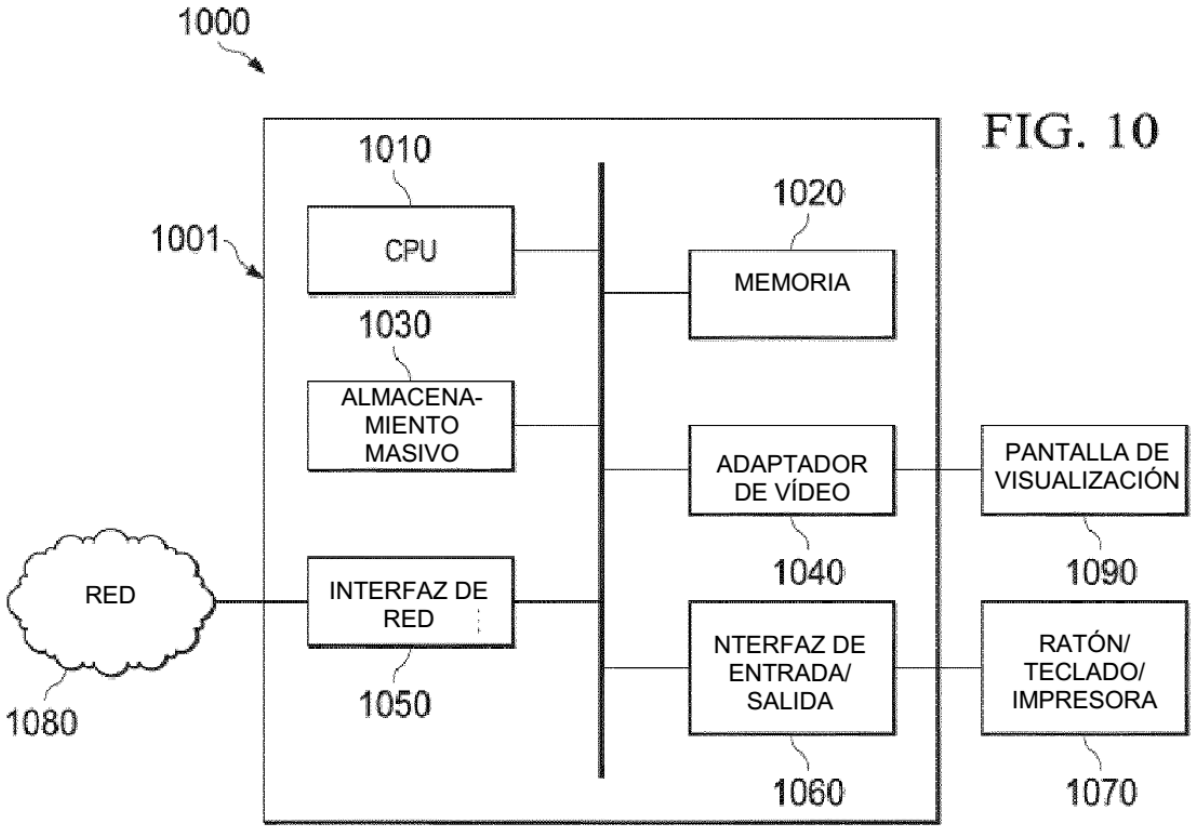

1000
1001
1010
CPU
1020
MEMORIA
1030
ALMACENA-
MIENTO
MASIVO
ADAPTADOR
DE VÍDEO
PANTALLA DE
VISUALIZACIÓN
1040
1090
RED
INTERFAZ DE
RED
NTERFAZ DE
ENTRADA/
SALIDA
RATÓN/
TECLADO/
IMPRESORA
1080
1050
1060
1070

FIG. 10