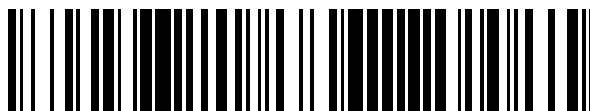


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 863 667**

51 Int. Cl.:

**G10L 19/02** (2013.01)

**G10L 19/022** (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **24.10.2006** **E 17193127 (2)**

97 Fecha y número de publicación de la concesión europea: **17.02.2021** **EP 3319086**

54 Título: **Codificación por transformada modificada distorsionada temporal de señales de audio**

30 Prioridad:

**03.11.2005 US 733512 P**

**11.08.2006 US 464176**

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

**11.10.2021**

73 Titular/es:

**DOLBY INTERNATIONAL AB (100.0%)**  
**Apollo Building, 3E, Herikerbergweg 1-35**  
**1101 CN Amsterdam Zuid-Oost, NL**

72 Inventor/es:

**VILLEMoes, LARS**

74 Agente/Representante:

**ELZABURU, S.L.P**

**ES 2 863 667 T3**

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

**DESCRIPCION**

Codificación por transformada modificada distorsionada temporal de señales de audio

**Campo de la invención**

5 La presente invención se refiere a sistemas de codificación de fuentes de audio y, en particular, a esquemas de codificación de audio que utilizan transformadas basadas en bloques.

**Antecedentes de la invención y técnica anterior**

Se conocen en la técnica diversas formas de codificar contenido de audio y vídeo. Generalmente, por supuesto, el objetivo es codificar el contenido de forma que ahorre bits sin degradar la calidad de reconstrucción de la señal.

10 Recientemente, se han desarrollado nuevos enfoques para codificar contenido de audio y vídeo, entre los cuales la codificación de audio perceptual basada en transformadas logra la mayor ganancia de codificación para señales estacionarias, es decir, cuando se pueden aplicar grandes magnitudes de transformadas. (Véase por ejemplo, T. Painter y A. Spanias: "Perceptual coding of digital audio", Actas IEEE, Vol. 88, N<sup>o</sup>. 4, abril de 2000, páginas 451-513). Las partes estacionarias de audio a menudo están bien modeladas por un número finito fijo de sinusoides estacionarias. Una vez que la magnitud de la transformada es lo suficientemente grande para resolver esos  
15 componentes, se requiere un número fijo de bits para un objetivo de deformación dado. Al aumentar aún más el tamaño de la transformada, se describirán segmentos cada vez más grandes de la señal de audio sin aumentar la demanda de bits. Sin embargo, para señales no estacionarias, es necesario reducir el tamaño de la transformada y, por tanto, la ganancia de codificación disminuirá rápidamente. Para superar este problema, para cambios abruptos y eventos transitorios, se puede aplicar el cambio de magnitud de la transformada sin incrementar significativamente el costo  
20 medio de codificación. Es decir, cuando se detecta un evento transitorio, se reduce el tamaño del bloque (tamaño de trama) de las muestras que se codificarán juntas. Para señales transitorias más persistentemente, la tasa de bits, por supuesto, aumentará drásticamente.

Un ejemplo particularmente interesante de comportamiento transitorio persistente es la variación de tono de las señales armónicas localmente, que se encuentran principalmente en las partes vocales del habla y del canto, pero  
25 también se pueden originar a partir de los vibratos y glissandos de algunos instrumentos musicales. Al tener una señal armónica, es decir, una señal que tiene picos de señal distribuidos con un espaciado igual a lo largo del eje del tiempo, el término tono describe el inverso del tiempo entre picos contiguos de la señal. Por tanto, una señal de este tipo tiene un espectro armónico perfecto, que consta de una frecuencia base igual al tono y armónicos de orden superior. En términos más generales, el tono se puede definir como el inverso del tiempo entre dos porciones vecinas de la señal  
30 correspondiente dentro de una señal localmente armónica. Sin embargo, si el tono y, por tanto, la frecuencia base varía con el tiempo, como es el caso de los sonidos vocales, el espectro se volverá cada vez más complejo y, por tanto, más ineficaz de codificar.

Un parámetro estrechamente relacionado con el tono de una señal es la deformación de la señal. Suponiendo que la  
35 señal temporal  $t$  tiene un tono igual a  $p(t)$  y que este valor de tono varía suavemente con el tiempo, la deformación de la señal temporal  $t$  viene definida por la derivada logarítmica

$$a(t) = \frac{p'(t)}{p(t)}.$$

Para una señal armónica, esta definición de deformación es insensible a la elección particular del componente armónico y a los errores sistemáticos en términos de múltiplos o fracciones del tono. La deformación mide un cambio de frecuencia en el dominio logarítmico. La unidad natural de deformación es Hertz [Hz], pero en términos musicales,  
40 una señal con deformación constante  $a(t)=a_0$ , es un barrido con una velocidad de barrido de  $a_0/\log_2$  octavas por segundo [oct/s]. Las señales vocales exhiben deformaciones de hasta 10 oct/s y una deformación media de alrededor de 2 oct/s.

Como la longitud de trama típica (longitud de bloque) de los codificadores por transformada es tan grande, tanto que el cambio de tono relativo es significativo dentro de la trama, las deformaciones o variaciones de tono de ese tamaño  
45 conducen a una codificación del análisis de frecuencia de esos codificadores. Como, para una tasa de bits constante requerida, esto solo puede superarse aumentando la tosquedad de la cuantificación, este efecto conduce a la introducción de ruido de cuantificación, que a menudo se percibe como reverberación.

Una técnica posible para superar este problema es la deformación temporal. El concepto de codificación deformada temporal se explica mejor imaginando una grabadora de cinta con velocidad variable. Al grabar la señal de audio, la  
50 velocidad se ajusta dinámicamente para lograr un tono constante en todos los segmentos sonoros. La señal de audio localmente estacionaria resultante se codifica junto con los cambios de velocidad de la cinta aplicados. En el descodificador, la reproducción se realiza con los cambios de velocidad opuestos. Sin embargo, la aplicación de la

deformación temporal simple como se describe anteriormente tiene algunos inconvenientes importantes. En primer lugar, la velocidad absoluta de la cinta termina siendo incontrolable, lo que conduce a una violación de la duración de toda la señal codificada y a limitaciones de ancho de banda. Para la reconstrucción, se debe transmitir información lateral adicional sobre la velocidad de la cinta (o equivalentemente sobre el tono de la señal), introduciendo una sobrecarga sustancial de velocidad de bits, especialmente a velocidades de bits bajas.

El enfoque común de los métodos de la técnica anterior para superar el problema de la duración incontrolable de las señales deformadas temporales es procesar segmentos consecutivos no superpuestos, es decir, tramas individuales, de la señal de forma independiente mediante una deformación temporal, de modo que se preserve la duración de cada segmento. Este enfoque se describe, por ejemplo, en Yang et. Alabama. "Pitch synchronous modulated lapped transform of the linear prediction residual of speech", actas de ICSP '98, páginas 591-594. Una gran desventaja de tal procedimiento es que aunque la señal procesada es estacionaria dentro de los segmentos, el tono exhibirá saltos en cada límite de segmento. Esos saltos conducirán evidentemente a una pérdida de eficiencia de codificación del codificador de audio subsiguiente y se introducirán discontinuidades audibles en la señal descodificada.

La deformación temporal también se realiza en varios otros esquemas de codificación. Por ejemplo, el documento US-2002/0120445 describe un esquema en el que los segmentos de señal están sujetos a ligeras modificaciones en la duración antes de la codificación de la transformada basada en bloques. Esto es para evitar grandes componentes de señal en el límite de los bloques, aceptando ligeras variaciones en la duración de los segmentos individuales.

Otra técnica que hace uso de la deformación temporal se describe en el documento US-6.169.970, donde la deformación temporal se aplica con el fin de elevar el rendimiento del pronosticador a largo plazo de un codificador vocal. A lo largo de las mismas líneas, en el documento US 2005/0131681, se describe una unidad de preprocesamiento para la codificación CELP de señales vocales que aplica una deformación lineal por partes entre intervalos no superpuestos, conteniendo cada uno de los cuales un impulso de tono blanqueado. Finalmente, se describe en (R.J. Sluijter y A.J.E.M. Janssen, "A time warper for speech signals", actas IEEE sobre Speech Coding '99, junio de 1999, páginas 150-152) cómo mejorar en la estimación del tono vocal mediante la aplicación de una función de deformación cuadrática de tiempo a una trama vocal. Otra técnica de deformación temporal se describe en Weruaga L., Kepesim : "SPEECH ANALYSISWITH THE FAST CHIRP TRANSFORM", EUSIPCO 2004, 12ª Conferencia Europea de Procesamiento de Señales, Viena, Austria, 7-10 de septiembre de 2004, septiembre de 2004 (2004-09), páginas 1011-1014, XP002619559.

Resumiendo, las técnicas de deformación según la técnica anterior comparten el problema de introducir discontinuidades en los bordes de la trama y de requerir una cantidad significativa de velocidad de bits adicional para la transmisión de los parámetros que describen la variación del tono de la señal.

**Resumen de la invención**

El objeto de esta invención es proporcionar un concepto para una codificación más eficiente de señales de audio usando la deformación temporal.

De acuerdo con un primer aspecto de la presente invención, este objeto se logra mediante un codificador de audio como se establece en la reivindicación 1.

De acuerdo con un segundo aspecto de la presente invención, este objeto se logra mediante un descodificador de la transformada deformada temporal como se establece en la reivindicación 2.

De acuerdo con un tercer aspecto de la presente invención, este objeto se logra mediante el método de codificación de audio como se establece en la reivindicación 3.

De acuerdo con un cuarto aspecto de la presente invención, este objeto se logra mediante un método de descodificación de la transformada deformada temporal como se establece en la reivindicación 4.

De acuerdo con un quinto aspecto de la presente invención, este objeto se logra mediante un programa de ordenador como se establece en la reivindicación 5.

Realizaciones particulares se definen en las reivindicaciones dependientes.

La presente invención se basa en el hallazgo de que una representación espectral de una señal de audio, que tiene tramas de audio consecutivas, se puede derivar de manera más eficiente cuando se estima una deformación temporal común para cualesquiera dos tramas vecinas, de modo que una siguiente transformada en bloque pueda utilizar adicionalmente la información deformada.

Por lo tanto, las funciones de ventana necesarias para la aplicación satisfactoria de un procedimiento de superposición y adición durante la reconstrucción se pueden derivar y aplicar, ya anticipando el re muestreo de la señal debido a la

deformación temporal. Por lo tanto, la elevada eficiencia de la codificación por transformada basada en bloques de señales deformadas temporalmente se puede utilizar sin introducir discontinuidades audibles.

Por tanto, la presente invención ofrece una solución atractiva a los problemas de la técnica anterior. Por un lado, el problema relacionado con la segmentación de la señal de audio se supera mediante una técnica particular de superposición y adición, que integra las operaciones de deformación temporal con la creación de ventanas e introduce un desplazamiento temporal de la transformada en bloque. Las transformadas continuas en el tiempo resultantes tienen una capacidad de reconstrucción perfecta y sus contrapartes discretas temporales sólo están limitadas por la calidad de la técnica de re muestreo aplicada del descodificador durante la reconstrucción. Esta propiedad da como resultado una convergencia de alta velocidad de bits del esquema de codificación de audio resultante. Es principalmente posible lograr una transmisión sin pérdidas de la señal disminuyendo la tosquedad de la cuantificación, es decir, aumentando la velocidad de transmisión de bits. Esto no se puede lograr, por ejemplo, con métodos de codificación puramente paramétricos.

Una ventaja adicional de la presente invención es una fuerte disminución de la demanda de velocidad de bits de la información adicional requerida para ser transmitida para revertir la deformación temporal. Esto se logra transmitiendo información lateral de los parámetros de la deformación en lugar de información lateral del tono. Esto tiene la ventaja adicional de que la presente invención exhibe sólo un grado leve de dependencia de parámetros en oposición a la dependencia crítica de la detección de tono correcta para muchos métodos de codificación de audio basados en parámetros de tono. Esto se debe a que la transmisión de los parámetros de tono requiere la detección de la frecuencia fundamental de una señal localmente armónica, lo que no siempre es fácilmente factible. El esquema de la presente invención es, por tanto, muy robusto, ya que evidentemente la detección de un armónico superior no falsifica los parámetros de la deformación a transmitir, dada la definición anterior de los parámetros de la deformación.

Para resumir, una representación espectral de una señal de audio que tiene tramas de audio consecutivas puede ser derivada de manera más eficiente, cuando se estima una deformación temporal común para dos tramas vecinas, de manera que una transformada en bloque siguiente pueda utilizar adicionalmente la información de la deformación. Por lo tanto, las funciones de ventana requeridas para la aplicación satisfactoria de un procedimiento de superposición y adición durante la reconstrucción se pueden derivar y aplicar, anticipando las funciones de ventana el re muestreo de la señal debido a la deformación temporal. Por lo tanto, la elevada eficiencia de la codificación de la transformada basada en bloques de las señales deformadas temporalmente se puede utilizar sin introducir discontinuidades audibles.

El término "tono" debe interpretarse en un sentido general. Este término también cubre una *variación de tono* en conexión con lugares relacionados con la información de la deformación. Puede haber una situación en la que la información de la deformación no dé acceso al tono absoluto, sino a la información del tono relativo o normalizado. Entonces, dada una información de la deformación, puede llegar a una descripción del tono de la señal, cuando acepta obtener una forma de curva de tono correcta sin valores en el eje y.

### Breve descripción de los dibujos

Las realizaciones preferidas de la presente invención que se describen a continuación hacen referencia a los dibujos adjuntos, en los que:

La figura 1 muestra un ejemplo de mapeos de deformación de acuerdo con la invención;

Las figuras 2-2b muestran la aplicación de una ventana de acuerdo con la invención dependiente de la deformación;

Las figuras 3a, 3b muestran un ejemplo de re muestreo de acuerdo con la invención;

Las figuras 4a, 4b muestran un ejemplo de síntesis de señal de acuerdo con la invención en el lado del descodificador;

Las figuras 5a, 5b muestran un ejemplo de creación de ventanas de acuerdo con la invención en el lado del descodificador;

Las figuras 6a, 6b muestran un ejemplo de deformación temporal de acuerdo con la invención en el lado del descodificador;

La figura 7 muestra un ejemplo de un procedimiento de acuerdo con la invención de superposición y adición en el lado del descodificador;

La figura 8 muestra un ejemplo de un codificador de audio de acuerdo con la invención;

La figura 9 muestra un ejemplo de un descodificador de audio de acuerdo con la invención;

La figura 10 muestra un ejemplo adicional de un descodificador de acuerdo con la invención;

La figura 11 muestra un ejemplo de una ejecución compatible con versiones anteriores de los conceptos de la invención.

La figura 12 muestra un diagrama de bloques de una ejecución de la codificación de acuerdo con la invención;

La figura 13 muestra un diagrama de bloques para un ejemplo de descodificación de acuerdo con la invención;

5 La figura 14 muestra un diagrama de bloques de una realización adicional de descodificación de acuerdo con la invención;

Las figuras 15a, 15b muestran una ilustración de la eficiencia de codificación alcanzable ejecutando el concepto de acuerdo con la invención.

**Descripción detallada de las realizaciones preferidas**

10 Las realizaciones descritas a continuación son meramente ilustrativas de los principios de la presente invención para la codificación de señales de audio por transformada deformada temporal. Se entiende que las modificaciones y variaciones de las disposiciones y los detalles descritos en este documento serán evidentes para los expertos en la técnica. Por lo tanto, la intención es que esté limitada únicamente por el alcance de las reivindicaciones de patente más adelante y no por los detalles específicos presentados a modo de descripción y explicación de las realizaciones en este documento.

A continuación, se revisan brevemente las ideas y conceptos básicos de deformación y transformadas en bloques para determinar el concepto de acuerdo con la invención, que se describirá con más detalle a continuación, haciendo referencia a las figuras adjuntas.

20 En general, los detalles de la transformada deformada temporal son más fáciles de derivar en el dominio de las señales continuas en el tiempo. Los siguientes párrafos describen la teoría general, que luego se especializan y convierten a su aplicación de acuerdo con la invención en señales temporales discretas. La etapa principal en esta conversión es reemplazar el cambio de coordenadas realizado en señales continuas en el tiempo con un re muestreo no uniforme de señales discretas en el tiempo, de tal manera que se preserve la densidad media de la muestra, es decir, que la duración de la señal de audio no se altere.

25  $s = \psi(t)$  describe un cambio temporal de coordenadas por una función creciente estrictamente diferencial continua  $\psi$ , mapeando el intervalo  $I$  en el eje  $t$  sobre el intervalo  $J$  en el eje  $s$ .

$\psi(t)$  es, por tanto, una función que se puede utilizar para transformar el eje tiempo de una cantidad dependiente del tiempo, que es equivalente a un re muestreo en el caso discreto en el tiempo. Cabe señalar que en la siguiente descripción, el intervalo  $I$  en el eje  $t$  es un intervalo en el dominio normal en el tiempo y el intervalo  $J$  en el eje  $s$  es un intervalo en el dominio de tiempo deformado.

Dada una base ortonormal  $\{v_\alpha\}$  para señales de energía finita en el intervalo  $J$ , se obtiene una base ortonormal  $\{u_\alpha\}$  para señales de energía finita en el intervalo  $I$  por la regla

$$u_\alpha(t) = \psi'(t)^{1/2} v_\alpha(\psi(t)). \quad (1)$$

35 Dado un intervalo de tiempo infinito  $I$ , la especificación local de la deformación temporal se puede lograr segmentando  $I$  y luego construyendo  $\psi$ , pegando piezas re escaladas de mapeos deformados normalizados.

*Un mapeo deformado normalizado* es una función continuamente diferenciable y estrictamente creciente que mapea el intervalo unitario  $[0,1]$  sobre sí mismo. Partiendo de una secuencia de puntos de segmentación  $t = t_k$  donde  $t_{k+1} > t_k$ , y de una secuencia correspondiente de mapeos de deformación normalizados  $\psi_k$ , se construye

$$\psi(t) = d_k \psi_k \left( \frac{t - t_k}{t_{k+1} - t_k} \right) + s_k, \quad t_k \leq t \leq t_{k+1}, \quad (2)$$

40 donde  $d_k = s_{k+1} - s_k$  y la secuencia  $d_k$  se ajusta de manera que  $\psi(t)$  se hace continuamente diferenciable. Esto define  $\psi(t)$  a partir de la secuencia de mapeos deformados normalizados  $\psi_k$ , hasta un cambio afín de escala del tipo  $A\psi(t) + B$ .

Sea  $\{v_{k,n}\}$  una base ortonormal para señales de energía finita en el intervalo  $J$ , adaptada a la segmentación  $s_k = \psi(t_k)$  en el sentido de que hay un entero  $K$ , el *factor de superposición*, tal que  $v_{k,n}(s) = 0$  si  $s < s_k$  o  $s > s_{k+K}$

La presente invención se centra en los casos  $K \geq 2$ , ya que el caso  $K = 1$  corresponde a los métodos según la técnica anterior sin superposición. Cabe señalar que actualmente no se conocen muchas construcciones para  $K \geq 3$ . Se desarrollará un ejemplo particular para el concepto de acuerdo con la invención para el caso  $K = 2$  a continuación, incluidas las bases trigonométricas locales que también se utilizan en transformadas discretas del coseno modificadas (MDCT) y otras transformadas discretas de lapso de tiempo.

5 Sea la construcción  $\{v_{k,n}\}$  a partir de la segmentación *local*, en el sentido de que existe un entero  $p$ , tal que  $v_{k,n}(s)$  no depende de  $S_l$  para  $l < k - p$  o  $l > k + K + p$ . Finalmente, sea la construcción tal que un cambio afín de segmentación a  $As_k + B$  de lugar a un cambio de base  $A^{-1/2}v_{k,n}((s - B) / A)$ . Entonces

$$u_{k,n}(t) = \psi'(t)^{1/2} v_{k,n}(\psi(t)) \quad (3)$$

10 es una base ortonormal deformada temporal para señales de energía finita en el intervalo  $I$ , que está bien definida a partir de los puntos de segmentación  $t_k$  y la secuencia de mapeos deformados normalizados  $\psi_k$ , independiente de la inicialización de las secuencias de los parámetros  $s_k$  y  $d_k$  en (2). Se adapta a la segmentación dada en el sentido de que  $u_{k,n}(t) = 0$  si  $t < t_k - p$  o  $t > t_{k+K} + p$ , y se define localmente en el sentido de que  $u_{k,n}(t)$  no depende de  $t_l$  para  $l < k - p$  o  $l > k + K + p$ , ni de los mapeos deformados normalizados  $\psi_l$  para  $l < k - p$  o  $l > k + K + p$ .

15 Las formas de onda de síntesis (3) son continuas, pero no necesariamente diferenciables, debido al factor Jacobiano  $(\psi'(t))^{1/2}$ . Por esta razón, y para la reducción de la carga computacional en el caso discreto en el tiempo, también se puede construir un sistema *biortogonal* derivado. Supóngase que existen constantes  $0 < C_1 < C_2$  tal que

$$C_1 \eta_k \leq \psi'(t) \leq C_2 \eta_k, \quad t_k \leq t \leq t_{k+K} \quad (4)$$

para una secuencia  $\eta_k > 0$ . Entonces

$$\left\{ \begin{array}{l} f_{k,n}(t) = \eta_k^{1/2} v_{k,n}(\psi(t)); \\ g_{k,n}(t) = \psi'(t) \eta_k^{-1/2} v_{k,n}(\psi(t)). \end{array} \right\} \quad (5)$$

20 define un par biortogonal si de bases de Riesz para el espacio de señales de energía finita en el intervalo  $I$ . Por consiguiente,  $f_{k,n}(t)$  así como  $g_{k,n}(t)$  pueden usarse para el análisis, mientras que es particularmente ventajoso usar  $f_{k,n}(t)$  como formas de onda de síntesis y  $g_{k,n}(t)$  como formas de onda de análisis.

25 Basándose en las consideraciones generales anteriores, un ejemplo del concepto de acuerdo con la invención se derivará en los párrafos siguientes para el caso de segmentación uniforme  $t_k = k$  y factor de superposición  $K = 2$ , utilizando una base local del coseno adaptada a la segmentación resultante en el eje  $S$ .

30 Cabe señalar que las modificaciones necesarias para tratar las segmentaciones no uniformes son obvias de modo que el concepto de acuerdo con la invención también es aplicable a tales segmentaciones no uniformes. En cuanto al ejemplo propuesto por M.W. Wickerhauser, "Adapted wavelet analysis from theory to software", A. K. Peters, 1994, Capítulo 4, un punto de partida para construir una base local del coseno es una función de corte ascendente  $p$  tal que  $p(r) = 0$  para  $r < -1$ ,  $p(r) = 1$  para  $r > 1$ , y  $p(r)^2 + p(-r)^2 = 1$  en la región activa  $-1 \leq r \leq 1$ .

Dada una segmentación  $S_k$ , se puede construir una ventana en cada intervalo  $S_k \leq S \leq S_{k+2}$  de acuerdo con

$$w_k(s) = \rho \left( \frac{s - c_k}{\epsilon_k} \right) \rho \left( \frac{c_{k+1} - s}{\epsilon_{k+1}} \right), \quad (6)$$

35 con puntos medios de corte  $C_k = (S_k + S_{k+1}) / 2$  y radios de corte  $\epsilon_k = (S_{k+1} - S_k) / 2$ . Esto corresponde a la construcción del punto medio de Wickerhauser.

Con  $l_k = C_{k+1} - C_k = \epsilon_k + \epsilon_{k+1}$ , resulta una base ortonormal a partir de

$$v_{k,n}(s) = \sqrt{\frac{2}{l_k}} w_k(s) \cos \left[ \frac{\pi(n + \frac{1}{2})}{l_k} (s - c_k) \right], \quad (7)$$

donde el índice de frecuencia  $n = 0, 1, 2, \dots$ . Es fácil verificar que esta construcción obedece a las condiciones de localidad con  $p = 0$  e invariancia afín descrita anteriormente. La base deformada resultante (3) en el eje  $t$  puede en este caso reescribirse en la forma

$$u_{k,n}(t) = \sqrt{2\phi'_k(t-k)} b_k(\phi_k(t-k)) \cos \left[ \pi(n + \frac{1}{2})(\phi_k(t-k) - m_k) \right], \quad (8)$$

para  $k \leq t \leq k+2$ , donde  $\phi_k$  se define pegando  $\psi_k$  e  $\psi_{k+1}$  para formar un mapeo continuamente diferenciable del intervalo  $[0,2]$  sobre sí mismo,

$$\phi_k(t) = \begin{cases} 2m_k \psi_k(t), & 0 \leq t \leq 1; \\ 2(1-m_k) \psi_{k+1}(t-1) + 2m_k, & 1 \leq t \leq 2. \end{cases} \quad (9)$$

Esto se obtiene poniendo

$$m_k = \frac{1}{2} \phi_k(1) = \frac{\psi'_{k+1}(0)}{\psi'_k(1) + \psi'_{k+1}(0)}. \quad (10)$$

La construcción de  $\psi_k$  se ilustra en la figura 1, mostrando el tiempo normalizado en el eje  $x$  y el tiempo deformado en el eje  $y$ . La figura 1 se describirá particularmente para el caso  $k = 0$ , es decir, para construir  $\phi_0(t)$  y por lo tanto, para derivar una función de deformación para una primera trama 10, que dura desde el tiempo normalizado 0 hasta el tiempo normalizado 1 y para una segunda trama 12 que dura desde el tiempo normalizado 1 hasta el tiempo normalizado 2. Se supone además que la primera trama 10 tiene una función de deformación 14 y que la segunda trama 12 tiene una función de deformación 16, derivada con el objetivo de lograr un tono igual dentro de las tramas individuales, cuando el eje de tiempo se transforma como lo indican las funciones de deformación 14 y 16. Cabe señalar que la función de deformación 14 corresponde a  $\psi_0$  y la función de deformación 16 corresponde a  $\psi_1$ . De acuerdo con la ecuación 9, se construye una función de deformación combinada  $\phi_0(t)$  18 pegando los mapeos de deformación 14 y 16 para formar un mapeo continuamente diferenciable del intervalo  $[0,2]$  sobre sí mismo. Como resultado, el punto (1,1) se transforma en (1, a), donde a, corresponde a  $2m_k$  en la ecuación 9.

Dado que el concepto de acuerdo con la invención está dirigido a la aplicación de la deformación temporal en un escenario de superposición y adición, el ejemplo de construcción de la siguiente función de deformación combinada para la trama 12 y para la siguiente trama 20 también se muestra en la figura 1. Cabe señalar que, siguiendo el principio de superposición y adición, para una reconstrucción de la trama 12, se requiere el conocimiento de ambas funciones de deformación 18 y 22.

Cabe señalar además que pegar dos funciones de deformación derivadas independientemente no es necesariamente la única forma de derivar una función combinada de deformación adecuada  $\phi$  (18, 22) ya que  $\phi$  puede muy bien derivarse también directamente fijando una función de deformación adecuada a dos tramas consecutivas. Se prefiere tener una consistencia afín de las dos funciones de deformación en la superposición de sus dominios de definición.

De acuerdo con la ecuación 6, la función de ventana en la ecuación 8 viene definida por

$$b_k(r) = \rho \left( \frac{r - m_k}{m_k} \right) \rho \left( \frac{1 + m_k - r}{1 - m_k} \right), \quad (11)$$

que aumenta de cero a uno en el intervalo  $[0, 2m_k]$  y disminuye de uno a cero en el intervalo  $[2m_k, 2]$ .

También se puede derivar una versión biortogonal de (8) si existen constantes  $0 < C_1, < C_2$ , tales que

$$C_1 \leq \phi'_k(t) \leq C_2, \quad 0 \leq t \leq 2,$$

para todos los  $k$ . Elegir  $nk = l_k$  en (4) conduce a la especialización de (5) a

$$\left. \begin{cases} f_{k,n}(t) = \sqrt{2} b_k(\phi_k(t-k)) \cos\left[\pi\left(n + \frac{1}{2}\right)(\phi_k(t-k) - m_k)\right]; \\ g_{k,n}(t) = \sqrt{2} \phi'_k(t-k) b_k(\phi_k(t-k)) \cos\left[\pi\left(n + \frac{1}{2}\right)(\phi_k(t-k) - m_k)\right]. \end{cases} \right\} \quad (12)$$

Así, para el caso de tiempo continuo, se derivan las funciones de síntesis y análisis (ecuación 12), que dependen de la función deformada combinada. Esta dependencia permite deformar el tiempo dentro de un escenario de superposición y adición sin pérdida de información sobre la señal original, es decir, permitiendo una reconstrucción perfecta de la señal.

Puede observarse que, para propósitos de ejecución, las operaciones realizadas dentro de la ecuación 12 pueden descomponerse en una secuencia de etapas de proceso individuales consecutivas. Una forma particularmente atractiva de hacerlo es realizar primero una ventana de la señal, seguida de un re muestreo de la señal de ventana y finalmente una transformada.

Como es habitual, las señales de audio se almacenan y transmiten digitalmente como valores de muestreo discretos muestreados con una frecuencia de muestreo dada, el ejemplo dado para la ejecución del concepto de acuerdo con la invención se desarrollará a continuación para la aplicación en el caso discreto.

La transformada discreta del coseno modificada deformada temporal (TWMDCT) se puede obtener a partir de una base del coseno local deformada temporal mediante la discretización de integrales de análisis y formas de onda de síntesis. La siguiente descripción se basa en la base biortogonal (véase ecuación 12). Los cambios necesarios para tratar el caso ortogonal (8) consisten en una ponderación adicional en el dominio en el tiempo por el factor Jacobiano  $\sqrt{\phi'_k(t-k)}$ . En el caso especial en el que no se aplique la deformación, ambas construcciones se reducen a la MDCT ordinaria. Sea  $L$  el tamaño de la transformada y supóngase que la señal  $x(t)$  a analizar tiene una banda limitada por  $q\pi L$  (rad/s) para algunas  $q < 1$ . Esto permite que la señal sea descrita por sus muestras en el período de muestreo  $1/L$ .

Los coeficientes de análisis vienen dados por

$$\begin{aligned} c_{k,n} &= \int_k^{k+2} x(t) g_{k,n}(t) dt \\ &= \sqrt{2} \int_k^{k+2} x(t) b_k(\phi_k(t-k)) \cos\left[\pi\left(n + \frac{1}{2}\right)(\phi_k(t-k) - m_k)\right] \phi'_k(t-k) dt \end{aligned} \quad (13)$$

Definir la porción de señal en ventana  $x_k(r) = x(t+k) b_k(\Phi_k(t))$  y realizar las sustituciones  $\tau = t - k$  y  $r = \Phi_k(\tau)$  en la integral (13) conduce a

$$c_{k,n} = \int_0^2 x_k(\phi_k^{-1}(r)) \cos\left[\pi\left(n + \frac{1}{2}\right)(r - m_k)\right] dr \quad (14)$$

Una forma particularmente atractiva de discretizar esta integral enseñada por la presente invención es elegirlos puntos de muestreo  $r = r_v = m_k + (v + 1/2) / L$ , donde  $v$  es un valor entero. Suponiendo una deformación leve y la limitación de banda descrita anteriormente, esto da la aproximación

$$c_{k,n} \approx \frac{\sqrt{2}}{L} \sum_v X_k(v) \cos\left[\frac{\pi}{L}\left(n + \frac{1}{2}\right)(v + \frac{1}{2})\right], \quad n = 0, 1, \dots, L-1, \quad (15)$$

donde

$$X_k(v) = x_k(\phi_k^{-1}(r_v)) \quad (16)$$

El intervalo de suma en (15) está definido por  $0 \leq r_v < 2$ . Incluye  $v = 0, 1, \dots, L-1$  y se extiende más allá de este intervalo en cada extremo, de manera que el número total de puntos sea  $2L$ . Téngase en cuenta que debido a las ventanas, el resultado es insensible al tratamiento de los casos extremos, lo que puede ocurrir si  $m_k = (v_0 + 1/2) / L$  para algún entero  $v_0$ .

5 Como es bien sabido que la suma (ecuación 15) se puede calcular mediante operaciones de plegado elementales seguida por un DCT de tipo IV, puede ser apropiado descomponer las operaciones de la ecuación 15 en una serie de operaciones y transformadas posteriores para hacer uso de realizaciones de hardware y software eficientes ya existentes, particularmente de DCT (transformada discreta del coseno). De acuerdo con la integral discretizada, una  
 10 señal de tiempo discreta dada se puede interpretar como las muestras equidistantes en períodos de muestreo  $1/L$  de  $x(t)$ . Un primer paso de la creación de las ventanas conduciría a:

$$x_k \left( \frac{p + \frac{1}{2}}{L} \right) = x \left( \frac{p + \frac{1}{2}}{L} + k \right) b_k \left( \phi_k \left( \frac{p + \frac{1}{2}}{L} \right) \right) \quad (17)$$

para  $p = 0, 1, 2, \dots, 2L-1$ . Antes de la transformada en bloque como se describe en la ecuación 15 (introduciendo un desplazamiento adicional dependiendo de  $m_k$ ), se requiere un re muestreo, mapeo

$$x_k \left( \frac{p + \frac{1}{2}}{L} \right) \mapsto x_k \left( \phi_k^{-1} \left( m_k + \frac{v + \frac{1}{2}}{L} \right) \right). \quad (18)$$

15 La operación de re muestreo se puede realizar por cualquier método de re muestreo no equidistante,

Resumiendo, el MDCT deformado temporal de acuerdo con la invención puede descomponerse en una operación de ventanas, un re muestreo y una transformada en bloque.

A continuación, se describirán brevemente las etapas individuales haciendo referencia a las figuras 2 a 3b. Las figuras 2 a 3b muestran las etapas de la codificación MDCT deformada temporal considerando solo dos bloques de señal en ventana de una señal de tono generada sintéticamente. Cada trama individual comprende 1024 muestras de modo que cada una de las dos tramas combinadas consideradas 24 y 26 (tramas originales 30 y 32 y tramas originales 32 y 34) consta de 2048 muestras, de modo que las dos tramas combinadas en ventana tienen una superposición de 1024 muestras. Las figuras 2 a 2b muestran en el eje x el tiempo normalizado de 3 tramas a procesar. El rango de las primeras 30 tramas va de 0 a 1, el rango de la segunda trama 32 va de 1 a 2 y el rango de las 3 tramas va de 2 a 3 en el eje del tiempo. Por tanto, en el dominio de tiempo normalizado, cada unidad de tiempo corresponde a una trama completa que tiene 1024 muestras de la señal. Las ventanas de análisis normalizado abarcan los intervalos de tiempo normalizados  $[0,2]$  y  $[1,3]$ . El objetivo de las siguientes consideraciones es recuperar la trama intermedia 32 de la señal. Como la reconstrucción de las tramas externas de la señal (30, 34) requiere datos de segmentos de señal de ventana contigua, esta reconstrucción no debe considerarse aquí. Puede observarse que los mapeos de deformación combinados mostrados en la figura 1 son mapeos de deformación derivados de la señal de la figura 2, que ilustran la combinación de acuerdo con la invención de tres mapeos de deformación normalizados subsiguientes (curvas de puntos) en dos mapeos de deformación superpuestos (curvas sólidas). Como se explicó anteriormente, los mapeos de deformación combinados de acuerdo con la invención 18 y 22 se derivan para el análisis de señales. Además, cabe señalar que debido a la invariancia afín de la deformación, esta curva representa un mapeo deformado con la misma deformación que en los dos segmentos originales.  
 20  
 25  
 30  
 35

La figura 2 ilustra la señal original mediante un gráfico sólido. Su tren de impulso estilizado tiene un tono que crece linealmente con el tiempo, por lo tanto, tiene una deformación positiva y decreciente considerando que la deformación se define como la derivada logarítmica del tono. En la figura 2, las ventanas de análisis de acuerdo con la invención como derivadas usando la ecuación 17 se superponen como curvas de puntos. Cabe señalar que la desviación de las ventanas simétricas normalizadas (como por ejemplo en MDCT) es mayor donde la deformación es mayor, es decir, en el primer segmento  $[0,1]$ . La definición matemática de las ventanas solo se da volviendo a muestrear las ventanas de la ecuación 11, el re muestreo realizado como se expresa por el segundo factor del lado derecho de la ecuación 17.  
 40

Las figuras 2a y 2b ilustran el resultado del sistema de ventanas de acuerdo con la invención, aplicando las ventanas de la figura 2 a los segmentos de señal individuales.  
 45

Las figuras 3a y 3b ilustran el resultado del re muestreo dependiente de los parámetros de la deformación de los bloques de señal de ventana de las figuras 2a y 2b, el re muestreo realizado como lo indican los mapeos de deformación dados por las curvas sólidas de la figura 1. El intervalo de tiempo normalizado  $[0,1]$  se asigna al intervalo de tiempo deformado  $[0, a]$ , que es equivalente a una compresión de la mitad izquierda del bloque de señal con ventana. En consecuencia, se realiza una expansión de la mitad derecha del bloque de señal con ventana, mapeando  
 50

el interno [1,2] a [a,2]. Dado que el mapeo de deformación se deriva de la señal con el objetivo de derivar la señal deformada con tono constante, el resultado de la deformación (re muestreo de acuerdo con la ecuación 18) es un bloque de señal con ventana que tiene un tono constante. Cabe señalar que un desajuste entre el mapeo deformado y la señal conduciría a un bloque de señal con un tono aún variable en este punto, lo que no perturbaría la reconstrucción final.

El desplazamiento de la siguiente transformada en bloque está marcado por círculos de modo que el intervalo [m, m + 1] corresponde a las muestras discretas  $v = 1, 0, \dots, L-1$  con  $L = 1024$  en la fórmula 15. Esto significa de manera equivalente que las formas de onda moduladoras de la transformada en bloque comparten un punto de simetría par en m y un punto de simetría impar en m + 1. Además, es importante tener en cuenta que a es igual a 2m, de modo que m es el punto medio entre 0 y a y m + 1 es el punto medio entre a y 2. Resumiendo, las figuras 3a y 3b describen la situación después del re muestreo de acuerdo con la invención descrito por la ecuación 18 que, por supuesto, depende de los parámetros de la deformación.

Las muestras en el dominio de la transformada deformada temporal de las señales de las figuras 3a y 3b se cuantifican y codifican entonces y pueden transmitirse junto con la información del lado de deformación que describe mapeos de deformación normalizados  $\psi_k$  a un descodificador. Como la cuantificación es una técnica comúnmente conocida, la cuantificación que utiliza una regla de cuantificación específica no se ilustra en las siguientes figuras, centrándose en la reconstrucción de la señal en el lado del descodificador.

En una realización de la presente invención, el descodificador recibe la secuencia del mapeo de deformación junto con las muestras en el dominio de transformada deformada temporal descodificadas  $d_{k,n}$ , donde puede suponerse que  $d_{k,n} = 0$  para  $n \geq L$  debido a la limitación de banda supuesta de la señal. Al igual que en el lado del codificador, el punto de partida para lograr la síntesis de tiempo discreto será considerar la reconstrucción de tiempo continuo utilizando la síntesis de formas de onda de la ecuación 12:

$$y(t) = \sum_{n,k} d_{n,k} f_{n,k}(t) = \sum_k y_k(t-k) \quad (19)$$

donde

$$y_k(u) = z_k(\phi_k(u)) \quad (20)$$

y con

$$z_k(r) = \sqrt{2} b_k(r) \sum_{n=0}^{L-1} d_{k,n} \cos\left[\pi\left(n + \frac{1}{2}\right)(r - m_k)\right]. \quad (21)$$

La ecuación (19) es el procedimiento usual de superposición y adición de una síntesis de transformada en ventana. Como en la etapa de análisis, es ventajoso muestrear la ecuación (21) en los puntos  $r = r_v = m_k + (v + 1/2) / L$ , dando lugar a

$$z_k(r_v) = \sqrt{2} b_k(r_v) \sum_{n=0}^{L-1} d_{k,n} \cos\left[\frac{\pi}{L}\left(n + \frac{1}{2}\right)(v + \frac{1}{2})\right] \quad (22)$$

que se calcula fácilmente mediante las siguientes etapas: Primero, una DCT de tipo IV seguido de una extensión en 2L en muestras según el parámetro de desplazamiento  $m_k$  de acuerdo con la regla  $0 \leq r_v < 2$ . A continuación, se realiza la creación de ventanas con la ventana  $b_k(r_v)$ . Una vez que se encuentra  $z_k(r_v)$ , el re muestreo

$$z_k\left(m_k + \frac{v + \frac{1}{2}}{L}\right) \mapsto z_k\left(\phi_k\left(\frac{p + \frac{1}{2}}{L}\right)\right) \quad (23)$$

da el segmento de señal  $y_k$  en puntos de muestra equidistantes  $(p + 1/2) / L$  listo para la operación de superposición y adición descrita en la fórmula (19).

El método de re muestreo se puede volver a ser elegido con bastante libertad y no tiene que ser el mismo que en el codificador. En una realización de la presente invención, se utilizan métodos basados en interpolación de ranura, donde el orden de las funciones de ranura se puede ajustar como una función de un parámetro de limitación de banda q para lograr un compromiso entre la complejidad computacional y la calidad de la reconstrucción. Un valor común del parámetro q es  $q = 1/3$ , un caso en el que las ranuras cuadráticas suelen ser suficientes.

La decodificación se ilustrará a continuación mediante las figuras 4a a 7 para la señal mostrada en las figuras 3a y 3b. Se enfatizará nuevamente que la transformada en bloque y la transmisión de los parámetros de la transformada no se describen en este documento, ya que es una técnica comúnmente conocida. Como comienzo para el proceso de decodificación, las figuras 4a y 4b muestran una configuración, en la que ya se ha realizado la transformada en bloque inversa, dando como resultado las señales mostradas en las figuras 4a y 4b. Una característica importante de la transformada en bloque inversa es la adición de componentes de señal que no están presentes en la señal original de las figuras 3a y 3b, lo que se debe a las propiedades de simetría de las funciones de síntesis ya explicadas anteriormente. En particular, la función de síntesis tiene simetría par con respecto a m y simetría impar con respecto a  $m + 1$ . Por lo tanto, en el intervalo  $[0, a]$ , se añaden componentes de señal positiva en la transformada en bloque inversa, mientras que en el intervalo  $[a, 2]$ , se añaden componentes de señal negativos. Adicionalmente, la función de ventana de acuerdo con la invención utilizada para la operación de ventana de síntesis se superpone como una curva de puntos en las figuras 4a y 4b.

La definición matemática de esta ventana de síntesis en el dominio en el tiempo deformado viene dada por la ecuación 11. Las figuras 5a y 5b muestran la señal, todavía en el dominio en el tiempo deformado, después de la aplicación del sistema de ventanas de acuerdo con la invención.

Las figuras 6a y 6b finalmente muestran el resultado del re muestreo dependiente del parámetro de la deformación de las señales de las figuras 5a y 5b.

Finalmente, la figura 7 muestra el resultado de la operación de superposición y adición, siendo la etapa final en la síntesis de la señal, (véase la ecuación 19). La operación de superposición y adición es una superposición de las formas de onda de las figuras 6a y 6b. Como ya se mencionó anteriormente, la única trama que se reconstruirá completamente es la trama intermedia 32 y, una comparación con la situación original de la figura 2 muestra que la trama intermedia 32 se reconstruye con alta fidelidad. La cancelación precisa de los componentes perturbadores de la señal de adición introducidos durante la transformada en bloque inverso solo es posible ya que es una propiedad crucial de la presente invención que los dos mapeos combinados 14 y 22 en la figura 1 difieran solo por un mapeo afín dentro del intervalo de tiempo normalizado superpuesto  $[1, 2]$ . Una consecuencia de esto es que existe una correspondencia entre las porciones de señal y las ventanas en los segmentos de tiempo deformados  $[a, 2]$  y  $[1, b]$ . Al considerar las figuras 4a y 4b, un estiramiento lineal de los segmentos  $[1, b]$  en  $[a, 2]$  hará, por tanto, que los gráficos de señal y las mitades de ventana describan el conocido principio de cancelación por alias en el dominio en el tiempo de la MDCT normalizada. La señal, que ya ha sido cancelada por alias, puede entonces simplemente mapearse en el intervalo de tiempo normalizado  $[1, 2]$  por un mapeo de deformación inverso común.

Cabe señalar que, de acuerdo con una realización adicional de la presente invención, la reducción adicional de la complejidad computacional se puede lograr mediante la aplicación de una etapa de pre filtrado en el dominio en frecuencia. Esto se puede realizar mediante una simple ponderación previa de los valores de muestra transmitidos  $d_{k,n}$ . Un filtrado previo de este tipo se describe, por ejemplo, en M. Unser, A. Aldroubi y M. Eden, "B-spline signal processing part II-efficient design and applications". Una realización requiere que se aplique un re muestreo B-ligero a la salida de la transformada en bloque inversa antes de la operación de ventana. Dentro de esta realización, el re muestreo opera sobre una señal derivada de la ecuación 22 que tiene  $d_{k,n}$  modificado. La aplicación de la función de ventana  $b_k(r_v)$  tampoco se realiza. Por lo tanto, en cada extremo del segmento de la señal, el re muestreo debe tener en cuenta las condiciones límite en términos de periodicidades y simetrías inducidas por la elección de la transformada en bloque. La ventana requerida se realiza luego del re muestreo usando la ventana  $b_k(\Phi_k((p + 1/2) / L))$ .

Resumiendo, de acuerdo con una primera realización de un decodificador de acuerdo con la invención, la MDCT deformada temporal inversa comprende, cuando se descompone en etapas individuales:

- Transformada inversa
- Ventanas
- Re muestreo
- Superposición y adición.

De acuerdo con una segunda realización de la presente invención, la MDCT deformada temporal inversa comprende:

- Ponderación espectral

- Transformada inversa
- Re muestreo
- Ventanas
- Superposición y adición.

5 Puede observarse que en un caso en el que no se aplique la deformación, que es el caso en el que todos los mapeos de deformación normalizados son triviales, ( $\psi_k(t) = t$ ), la realización de la presente invención como se detalló anteriormente coincide exactamente con la MDCT habitual.

Se describirán ahora realizaciones adicionales de la presente invención que incorporan las características mencionadas anteriormente haciendo referencia a las figuras 8 a 15.

10 La figura 8 muestra un ejemplo de un codificador de audio de acuerdo con la invención que recibe una señal digital de audio 100 como entrada y genera un flujo de bits para ser transmitido a un descodificador que incorpora el concepto de codificación de transformada deformada temporal de acuerdo con la invención. La señal de entrada digital de audio 100 puede ser una señal de audio natural o una señal de audio pre procesada, donde, por ejemplo, el preprocesamiento podría ser una operación de blanqueo para blanquear el espectro de la señal de entrada. El  
 15 codificador de acuerdo con la invención incorpora un extractor de parámetros de la deformación 101, un transformador de la deformación 102, una calculadora del modelo perceptual 103, un codificador de la deformación 104, un codificador 105 y un multiplexor 106. El extractor 101 de los parámetros de la deformación estima una secuencia de los parámetros de la deformación, que se introduce en el transformador de la deformación 102 y en el codificador de la deformación 104. El transformador de la deformación 102 deriva una representación espectral deformada temporal de la señal de entrada de audio digital 100. La representación espectral deformada temporal se introduce en el  
 20 codificador 105 para cuantificación y posiblemente otra codificación, como por ejemplo codificación diferencial. El codificador 105 es controlado adicionalmente por la calculadora del modelo perceptual 103. Tal, por ejemplo, tosquedad de la cuantificación puede incrementarse cuando se vayan a codificar los componentes de la señal que están enmascarados principalmente por otros componentes de la señal. El codificador de la deformación 104 codifica la secuencia de los parámetros de la deformación para reducir su tamaño durante la transmisión dentro del flujo de bits. Esto podría comprender, por ejemplo, la cuantificación de los parámetros o, por ejemplo, técnicas de codificación  
 25 diferencial o de codificación de entropía así como esquemas de codificación aritmética.

El multiplexor 106 recibe la secuencia de los parámetros de la deformación codificada del codificador de deformación 104 y una representación espectral de deformación temporal codificada de la señal 100 de la entrada de audio digital para multiplexar ambos datos en el flujo de bits emitido por el codificador.  
 30

La figura 9 ilustra un ejemplo de un descodificador de transformada deformada temporal que recibe un flujo de bits 200 compatible para derivar una señal de audio reconstruida como salida. El descodificador comprende un desmultiplexor 201, un descodificador de la deformación 202, un descodificador 203 y un transformador de la deformación inversa 204. El desmultiplexor desmultiplexa el flujo de bits en la secuencia codificada de los parámetros de la deformación, que se introduce en el descodificador de la deformación 202. El desmultiplexor desmultiplexa además la representación codificada de la representación espectral deformada temporal de la señal de audio, que se introduce en el descodificador 203 que es el inverso del correspondiente codificador 105 del codificador de audio de la figura 8. El descodificador de la deformación 202 deriva una reconstrucción de la secuencia de los parámetros de la deformación y el descodificador 203 deriva una representación espectral deformada temporal de la señal de audio original. La representación de la secuencia de parámetros de la deformación, así como la representación espectral deformada temporal, se ingresan en el transformador de la deformación inversa 204 que deriva una señal de salida de audio digital que ejecuta el concepto de acuerdo con la invención de codificación por transformada superpuesta deformada temporal de las señales de audio.  
 35  
 40

La figura 10 muestra una realización adicional de un descodificador por transformada deformada temporal en la que la secuencia de los parámetros de la deformación se deriva en el propio descodificador. La realización alternativa mostrada en la figura 10 comprende un descodificador 203, un estimador de la deformación 301 y un transformador de deformación inversa 204. El descodificador 203 y el transformador de deformación inversa 204 comparten las mismas funcionalidades que los dispositivos correspondientes de la realización anterior y, por lo tanto, la descripción de estos dispositivos dentro de las diferentes realizaciones es totalmente intercambiable. El estimador de la deformación 301 deriva la deformación real de la salida de la representación espectral deformada temporal por el descodificador 203 combinando las estimaciones del tono en el dominio en frecuencia anteriores con una estimación del tono en el dominio en frecuencia actual. Por tanto, la secuencia de los parámetros de la deformación se señala implícitamente, lo que tiene la gran ventaja de que se puede ahorrar más velocidad de bits, ya que no tiene que transmitirse la información de los parámetros de la deformación adicional en la entrada del flujo de bits al descodificador. Sin embargo, la señalización implícita de los datos deformados está limitada por la resolución temporal de la transformada.  
 45  
 50  
 55

La figura 11 ilustra la compatibilidad con versiones anteriores del concepto de acuerdo con la invención, cuando los descodificadores de la técnica anterior no son capaces usar el concepto de acuerdo con la invención de descodificación deformada temporal. Tal descodificador ignoraría la información adicional en los parámetros de la deformación, descodificando por consiguiente el flujo de bits en una señal de dominio en frecuencia alimentada a un transformador inverso 401 que no realiza ninguna deformación. Dado que el análisis de frecuencia realizado por la transformada deformada temporal en codificadores de acuerdo con la invención está bien alineado con la transformada que no incluye ninguna deformación temporal, un descodificador que ignore los datos de la deformación seguirá produciendo una salida de audio significativa. Esto se hace a costa de una calidad de audio degradada debido a la deformación temporal, que no se invierte en los descodificadores según la técnica anterior.

La figura 12 muestra un diagrama de bloques del método de acuerdo con la invención de la transformada deformada temporal. La transformada deformada temporal de acuerdo con la invención comprende la creación de ventanas 501, el re muestreo 502 y una transformada en bloque 503. En primer lugar, la señal de entrada se aplica en ventanas con una secuencia de ventanas superpuestas dependiendo de la secuencia de los parámetros de la deformación que sirve como entrada adicional para cada una de las etapas de codificación individuales, 501 a 503. Cada segmento de señal de entrada en ventana se re muestrea posteriormente en la etapa de re muestreo 502, en la que el re muestreo se realiza según lo indicado por la secuencia de los parámetros de la deformación.

Dentro de la etapa de transformada en bloque 503, una transformada en bloque se deriva típicamente usando una conocida transformada trigonométrica discreta. Por lo tanto, la transformada se realiza en el segmento de señal en ventana y re muestreada. Cabe señalar que la transformada en bloque también depende de un valor de desplazamiento, que se deriva de la secuencia de los parámetros de la deformación. Por lo tanto, la salida consiste en una secuencia de tramas en el dominio de la transformada.

La figura 13 muestra un diagrama de flujo de un método de transformada deformada temporal inversa. El método comprende las etapas de transformada por bloque inverso 601, ventanas 602, re muestreo 603 y superposición y adición 604. Cada trama de una señal en el dominio de la transformada se convierte en una señal en el dominio en el tiempo mediante la transformada por bloque inverso 601. Correspondiente a la etapa de codificación, la transformada por bloque depende de un valor de desplazamiento derivado de la secuencia de parámetros recibidos que sirve como entrada adicional a la transformada por bloque inverso 601, ventanas 602 y el re muestreo 603. El segmento de señal derivado por la transformada por bloque 601 se sitúa posteriormente en ventana en la etapa de ventanas 602 y se re muestrea en el re muestreo 603 usando la secuencia de los parámetros de la deformación. Finalmente, al superponer y agregar 604, el segmento en ventana y re muestreado se agrega a los segmentos previamente transformados inversamente en una operación usual de superposición y adición, dando como resultado una reconstrucción de la señal de salida en el dominio en el tiempo.

La figura 14 muestra una realización alternativa de un transformador de deformación temporal inversa de acuerdo con la invención, que se realiza para reducir adicionalmente la complejidad computacional. El descodificador comparte parcialmente las mismas funcionalidades con el descodificador de la figura 13. Por lo tanto, la descripción de los mismos bloques funcionales en ambas realizaciones son completamente intercambiables. La realización alternativa difiere de la realización de la figura 13 en que realiza una ponderación previa espectral 701 antes de la transformada en bloque inverso 601. Esta pre ponderación espectral fija es equivalente a un filtrado en el dominio en el tiempo con periodicidades y simetrías inducidas por la elección de la transformada en bloque. Dicha operación de filtrado es parte de ciertos métodos de re muestreo ligeros, lo que permite una reducción de la complejidad computacional del re muestreo modificado posterior 702. Tal re muestreo ahora se realizará en un dominio de señal con periodicidades y simetrías inducidas por la elección de la transformada en bloque. Por lo tanto, una etapa de creación de ventanas modificada 703 de reducción se realiza después del re muestreo 702. Finalmente, al superponer y añadir 604, el segmento dividido y vuelto a muestrear se le añade al segmento previamente transformado inverso en un procedimiento de superposición y adición habitual que da la señal de salida en el dominio en el tiempo reconstruida.

Las figuras 15a y 15b muestran la fuerza del concepto de acuerdo con la invención de codificación deformada temporal, mostrando representaciones espectrales de la misma señal con y sin deformación temporal aplicada. La figura 15a ilustra una trama de líneas espectrales que se originan a partir de una transformada discreta del coseno modificada de tamaño de la transformada 1024 de un segmento de señal vocal masculina muestreado a 16 kHz. La resolución de frecuencia resultante es de 7,8 Hz y solo las primeras 600 líneas se trazan para esta ilustración, correspondientes a un ancho de banda de 4,7 kHz. Como se puede ver de la frecuencia fundamental y del gráfico, el segmento es un sonido vocal con un tono medio de aproximadamente 155 Hz. Como puede verse además en la figura 15a, los pocos primeros armónicos de la frecuencia de tono son claramente distinguibles, pero hacia las frecuencias altas, el análisis se vuelve cada vez más denso y encriptado. Esto se debe a la variación del tono dentro de la longitud del segmento de la señal a analizar. Por lo tanto, la codificación de los rangos de frecuencia media a alta requiere una cantidad sustancial de bits para no introducir artefactos audibles durante la descodificación. Por el contrario, al fijar la velocidad de bits, inevitablemente se producirá una cantidad sustancial de deformación debido a la demanda de aumentar la tosquedad de la cuantificación.

La figura 15b ilustra una trama de líneas espectrales que se originan a partir de una transformada discreta del coseno modificada deformada temporal de acuerdo con la presente invención. Obviamente, se ha utilizado la misma señal de

audio masculina original que en la figura 15a. Los parámetros de la transformada son los mismos que para la figura 15a, pero el uso de una transformada deformada temporal adaptada a la señal tiene el efecto dramático visible sobre la representación espectral. El carácter disperso y organizado de la señal en el dominio de la transformada deformada temporal produce una codificación con un rendimiento de deformación de la velocidad mucho mejor, incluso cuando se tiene en cuenta el costo de codificar los datos deformados adicionales.

Como ya se mencionó, la transmisión de los parámetros de la deformación en lugar de la transmisión de la información del tono o velocidad tiene la gran ventaja de reducir drásticamente la velocidad de bits adicional requerida. Por consiguiente, en los siguientes párrafos, se detallan diversos esquemas de acuerdo con la invención de transmisión de la información de los parámetros de deformación requeridos.

Para una señal con deformación  $a(t)$  en un tiempo  $t$ , la elección óptima de la secuencia de mapeo deformada normalizada  $\psi_k$  para las bases del coseno local (véase (8), (12) se obtiene resolviendo

$$\frac{\psi_k''(t-k)}{\psi_k'(t-k)} = a(t), \quad k \leq t \leq k+1 \quad (24)$$

Sin embargo, la cantidad de información requerida para describir esta secuencia de mapeo deformada es demasiado grande y es difícil la definición y medición de los valores puntuales de  $a(t)$ . Para fines prácticos, un *intervalo de actualización deformado*  $\Delta t$  se establece sobre ello y cada mapeo deformado  $\psi_k$  se describe mediante los parámetros  $N = 1 / \Delta t$ .

Un intervalo de actualización deformado de alrededor de 10-20 ms suele ser suficiente para las señales vocales. De manera similar a la construcción en (9) de  $\Phi_k$  de  $\psi_k$  y  $\psi_{k+1}$ , un mapeo deformado normalizado continuamente diferenciable puede ser reconstruido por  $N$  mapeos deformados normalizados por medio de operaciones de ajuste de re escalado afines adecuadas. Ejemplos de prototipos de mapeos deformados normalizados incluyen

$$\left\{ \begin{array}{l} \text{Cuadrático:} \quad t \mapsto \left(1 - \frac{a}{2}\right)t + \frac{a}{2}t^2; \\ \text{Exponencial:} \quad t \mapsto \frac{\exp(at) - 1}{\exp(a) - 1}; \\ \text{Möbius:} \quad t \mapsto \frac{t}{\alpha + (1 - \alpha)t}, \quad \alpha = \frac{4 + a}{4 - a} \end{array} \right\} \quad (25)$$

donde  $a$  es un parámetro deformado. Al definir la deformación de un mapeo  $h(t)$  por  $h''/h'$ , los tres mapeos logran una deformación igual a  $a$  en  $t = 1/2$ . El mapeo exponencial tiene una deformación constante en todo el intervalo  $0 \leq t \leq 1$ , y para valores pequeños de  $a$ , los otros dos mapeos muestran una desviación muy pequeña de este valor constante. Para un mapeo deformado dado aplicado en el descodificador para el re muestreo (23), se requiere su inverso en el codificador para el re muestreo (ecuación 18). Una parte principal del esfuerzo de inversión se origina en la inversión de los mapeos deformados normalizados. La inversión de un mapeo cuadrático requiere operaciones de raíz cuadrada, la inversión de un mapeo exponencial requiere un logaritmo y la inversa del mapeo racional de Moebius es un mapeo de Moebius con parámetro deformado negado. Dado que las funciones exponenciales y las divisiones son comparativamente caras, un enfoque en la máxima facilidad computacional en el descodificador conduce a la elección preferida de una secuencia de mapeo deformado cuadrática por partes  $\psi_k$ .

El mapeo de deformación normalizado  $\psi_k$  está completamente definido por  $N$  parámetros de la deformación  $a_k(0)$ ,  $a_k(1)$ , ...,  $a_k(N-1)$  por los requisitos que

- es un mapeo deformado normalizado;
- se ensambla mediante copias re escaladas de uno de los prototipos de mapeos deformados suaves (25)
- es continuamente diferenciable

• satisfice

$$\frac{\psi_k''\left(\frac{l+\frac{1}{2}}{N}\right)}{\psi_k'\left(\frac{l+\frac{1}{2}}{N}\right)} = a_k(l), \quad l = 0, 1, \dots, N-1 \quad (26)$$

5 La presente invención enseña que los parámetros pueden ser cuantificados linealmente, típicamente, a un tamaño de etapa de aproximadamente 0,5 Hz. A continuación, se codifican los valores enteros resultantes. Alternativamente, la derivada  $\psi_k'$  se puede interpretar como una curva de tono normalizada donde los valores

$$\frac{\psi_k'(l\Delta t)}{\psi_k'(0)} - 1, \quad l = 1, 2, \dots, N, \quad (27)$$

10 se cuantifican a un tamaño de etapa fijo, normalmente 0,005. En este caso, los valores enteros resultantes se codifican adicionalmente por diferencia, secuencialmente o de manera jerárquica. En ambos casos, la velocidad de bits de la información lateral resultante es típicamente de unos pocos cientos de bits por segundo, que es solo una fracción de la velocidad requerida para describir los datos del tono en un códec vocal.

15 Un codificador con grandes recursos computacionales puede determinar la secuencia de datos deformada que de manera óptima reduce el costo de codificación o maximiza una medida de escasez de líneas espectrales. Un procedimiento menos costoso es utilizar métodos bien conocidos para el seguimiento del tono que dan como resultado una función de tono medida  $p(t)$  y que se aproxima la curva de tono con una función lineal por partes de  $p_0(t)$  en aquellos intervalos donde existe una pista de tono y no presenta grandes saltos en los valores del tono. La secuencia de la deformación estimada viene dada por

$$a_k(l) = \frac{2 p_0((l+1)\Delta t + k) - p_0(l\Delta t + k)}{\Delta t p_0((l+1)\Delta t + k) + p_0(l\Delta t + k)} \quad (28)$$

20 dentro de los intervalos de seguimiento del tono. Fuera de esos intervalos, la deformación se fija a cero. Téngase en cuenta que un error sistemático en las estimaciones del tono tal como la duplicación del período del tono, tiene muy poco efecto en las estimaciones de la deformación.

25 Como se ilustra en la figura 10, en una realización alternativa de la presente invención, la secuencia de los parámetros de la deformación puede derivarse de los datos en el dominio de la transformada descodificados mediante un estimador de la deformación. El principio es calcular una estimación del tono en el dominio de la frecuencia para cada trama de datos de la transformada o de los tonos de los bloques de señales descodificados subsiguientes. La información de la deformación se deriva luego de una fórmula similar a la fórmula 28.

30 La aplicación del concepto de acuerdo con la invención se ha descrito principalmente aplicando la deformación en un escenario de un único canal de audio. Por supuesto, el concepto de acuerdo con la invención no se limita en modo alguno al uso dentro de tal escenario monofónico. Además, puede resultar extremadamente ventajoso utilizar la alta ganancia de codificación que se puede conseguir mediante el concepto de acuerdo con la invención dentro de aplicaciones de codificación multicanal, en las que el canal único o el canal múltiple que debe transmitirse puede codificarse utilizando el concepto de acuerdo con la invención.

35 Además, la deformación podría definirse generalmente como una transformación del eje x de una función arbitraria que depende de x. Por lo tanto, el concepto de acuerdo con la invención también se puede aplicar a escenarios donde las funciones o la representación de señales están deformadas que no dependen explícitamente del tiempo. Por ejemplo, también se puede realizar la deformación de una representación de la frecuencia de una señal.

Además, el concepto de acuerdo con la invención también se puede aplicar ventajosamente a señales que están segmentadas con una longitud de segmento arbitraria y no con la misma longitud como se describe en los párrafos anteriores.

El uso de las funciones básicas y el criterio presentados en los párrafos anteriores debe entenderse además como un ejemplo ventajoso de aplicación del concepto de acuerdo con la invención. Para otras aplicaciones también se pueden utilizar diferentes funciones básicas, así como diferentes criterios.

5 Dependiendo de ciertos requisitos de realización de los métodos de acuerdo con la invención, los métodos de acuerdo con la invención se pueden realizar en hardware o en software. La realización se puede realizar utilizando un medio de almacenamiento digital, en particular un disco, DVD o CD que tenga almacenadas señales de control interpretables electrónicamente, que cooperan con un sistema de ordenador programable de manera que se lleven a cabo los métodos de acuerdo con la invención. Generalmente, la presente invención es, por tanto, un producto de programa de ordenador con un código de programa almacenado en un soporte interpretable por ordenador, siendo el código de programa operativo para realizar los métodos de acuerdo con la invención cuando el producto de programa de ordenador se ejecuta en un ordenador. En otras palabras, los métodos de acuerdo con la invención son, por lo tanto, un programa de ordenador que tiene un código de programa para realizar al menos uno de los métodos de acuerdo con la invención cuando el programa de ordenador se ejecuta en un ordenador.

10 Si bien lo anterior se ha mostrado y descrito particularmente con referencia a realizaciones particulares de lo mismo, los expertos en la técnica entenderán que se pueden realizar diversos cambios en la forma y detalles sin apartarse del alcance de la misma. Debe entenderse que se pueden realizar diversos cambios para adaptarse a las diferentes realizaciones. El alcance de la presente invención viene definido por las reivindicaciones adjuntas.

**REIVINDICACIONES**

1. Codificador de audio para recibir una señal de entrada de audio (100) y para generar un flujo de bits a transmitir a un descodificador, que comprende:

5 un extractor de los parámetros de la deformación (101) para estimar una secuencia de parámetros de deformación, en el que la secuencia de los parámetros de la deformación comprende una secuencia de parámetros de deformación, en el que cada parámetro de la deformación describe un intervalo de longitud finita de la señal de audio;

un transformador de deformación (102) para recibir la secuencia de los parámetros de la deformación y para derivar una representación espectral deformada temporal de la señal de entrada de audio (100);

10 una calculadora de modelo perceptual (103) para recibir la señal de entrada de audio (100):

un codificador de deformación (104) para codificar la secuencia de los parámetros de la deformación para reducir su tamaño durante una transmisión de una secuencia de parámetros de deformación codificada dentro de un flujo de bits;

15 un codificador (105) para recibir la representación espectral deformada temporal para su cuantificación a fin de obtener una representación espectral deformada temporal codificada de la señal de entrada de audio, en el que el codificador (105) está controlado por la calculadora de modelo perceptual (103); y

un multiplexor (106) para recibir y multiplexar la secuencia de los parámetros de la deformación codificada y la representación espectral deformada temporal codificada de la señal de entrada de audio.

2. Descodificador de la transformada deformada temporal para derivar una señal de audio reconstruida. que comprende:

20 un desmultiplexor (201) para desmultiplexar un flujo de bits en una secuencia de parámetros de deformación codificada y una representación codificada de la representación espectral de la deformación temporal;

25 un descodificador de deformación (202) para descodificar la secuencia de los parámetros de la deformación codificada para derivar una reconstrucción de la secuencia de los parámetros de la deformación, en la que la secuencia de los parámetros de la deformación comprende una secuencia de parámetros de deformación, en el que cada parámetro de la deformación describe un intervalo de longitud finito de la señal de audio;

un descodificador (203) para descodificar la representación codificada de la representación espectral deformada temporal para derivar una representación espectral deformada temporal de una señal de audio; y

30 un transformador de deformación temporal inverso (204) para recibir la reconstrucción de la secuencia de los parámetros de la deformación y la representación espectral deformada temporal de una señal de audio y para derivar la señal de salida de audio reconstruida utilizando una codificación de la transformada superpuesta deformada temporal

3. Método de codificación de audio, que comprende:

recibir una señal de entrada de audio (100);

35 estimar (101) una secuencia de parámetros de deformación, en el que la secuencia de los parámetros de la deformación comprende una secuencia de parámetros de deformación, en la que cada parámetro de la deformación describe un intervalo de longitud finita de la señal de audio;

derivar (102) una representación espectral deformada temporal de la señal de entrada de audio (100) utilizando la secuencia de los parámetros de la deformación;

40 codificar (104) la secuencia de los parámetros de la deformación para reducir su tamaño durante una transmisión de una secuencia de parámetros de deformación codificada dentro de un flujo de bits;

cuantificar (105) la representación espectral deformada temporal para obtener una representación espectral deformada temporal codificada de la señal de entrada de audio, en la que la cuantificación (105) está controlada por una calculadora de modelo perceptual (103); y

45 multiplexar (106) la secuencia de los parámetros de la deformación codificada y la representación espectral de la deformación temporal codificada de la señal de entrada de audio.

4. Método de descodificación de la transformada deformada temporal para derivar una señal de audio reconstruida, que comprende:

desmultiplexar (201) un flujo de bits en una secuencia de parámetros de deformación codificada y una representación codificada de la representación espectral deformada temporal;

5           descodificar (202) la secuencia de los parámetros de la deformación codificada para derivar una reconstrucción de la secuencia de los parámetros de la deformación, en el que la secuencia de los parámetros de la deformación comprende una secuencia de parámetros de deformación, en la que cada parámetro de la deformación describe un intervalo de longitud finita de la señal de audio;

10          descodificar (203) la representación codificada de la representación espectral deformada temporal para derivar una representación espectral deformada temporal de una señal de audio; y

derivar (204) la señal de salida de audio reconstruida usando una codificación de la transformada superpuesta deformada temporal usando la reconstrucción de la secuencia de los parámetros de la deformación y la representación espectral deformada temporal de la señal de audio.

15          5. Programa de ordenador que tiene un código de programa adaptado para realizar, cuando se ejecuta en un ordenador, cualquiera de los métodos de las reivindicaciones 3 o 4.

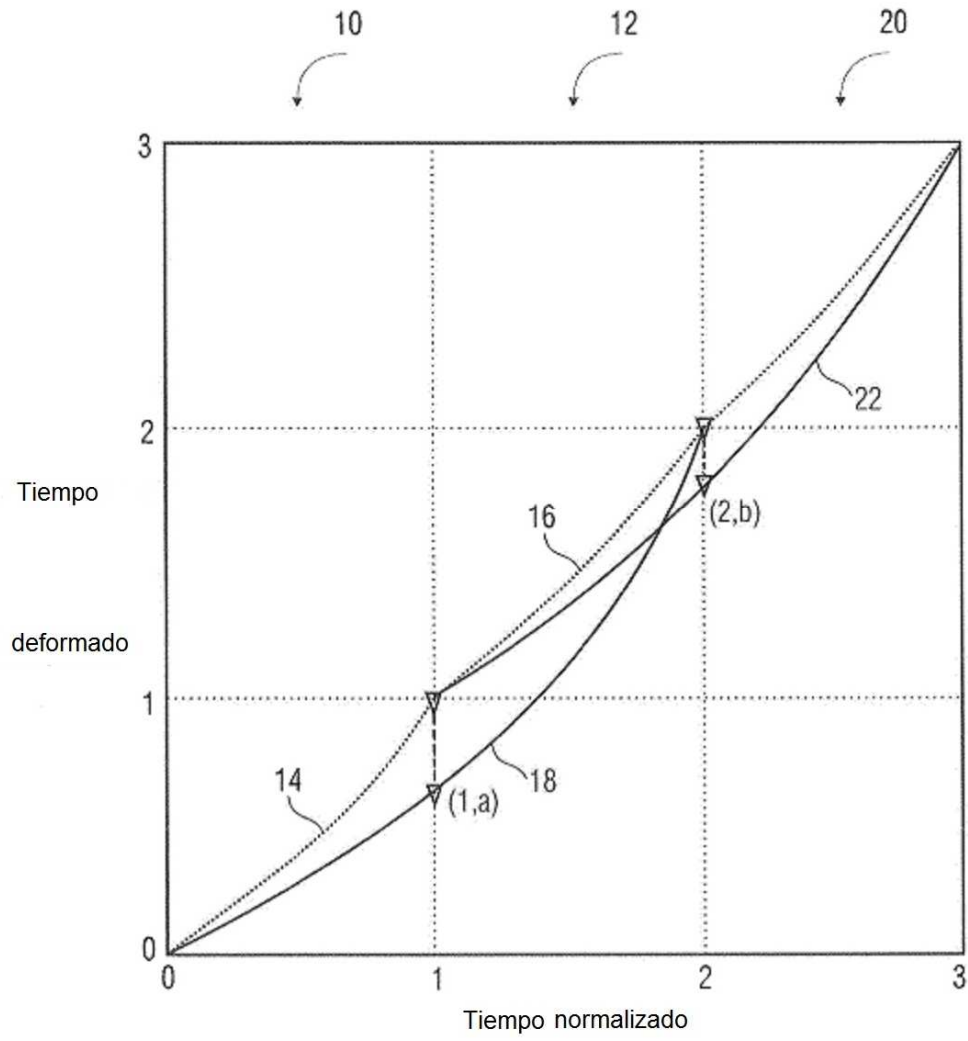


FIG 1

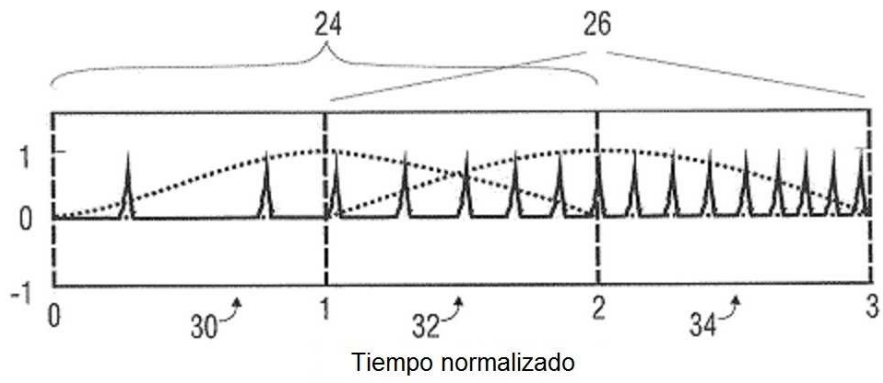


FIG 2

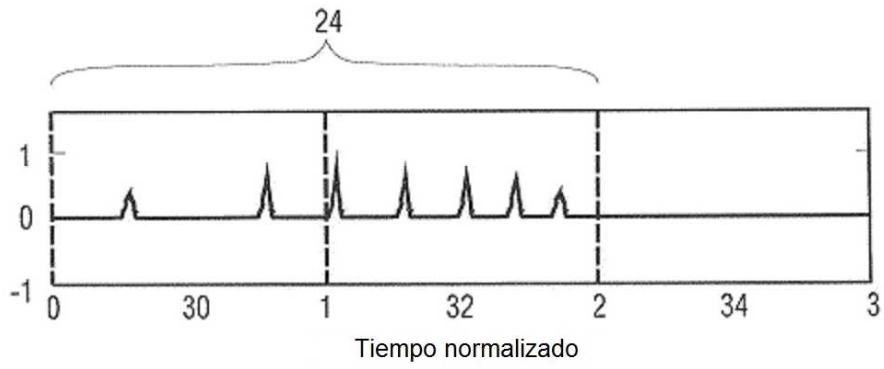


FIG 2A

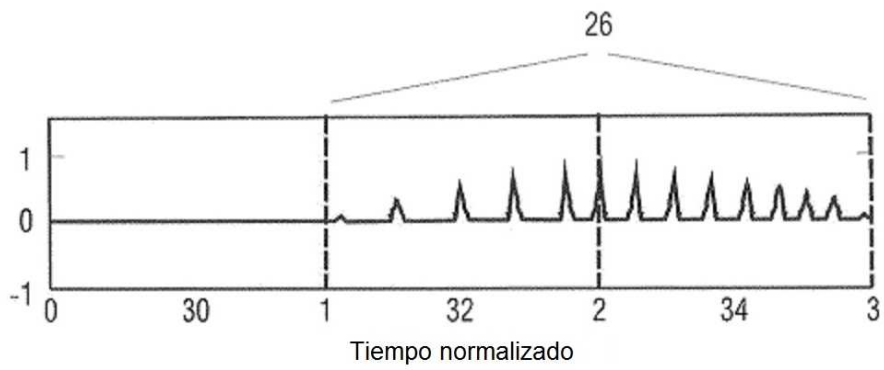


FIG 2B

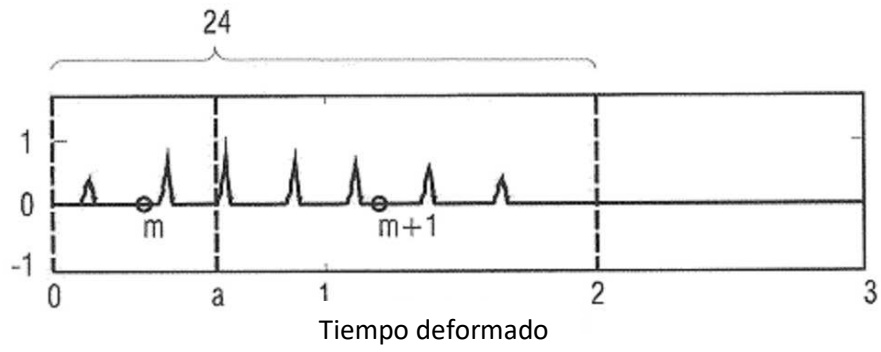


FIG 3A

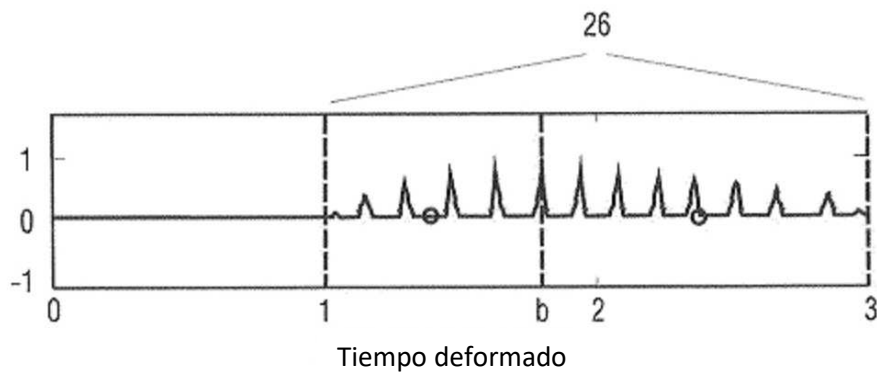


FIG 3B

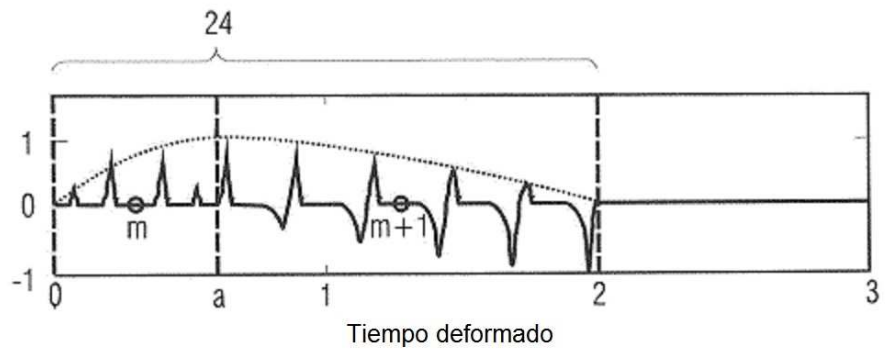


FIG 4A

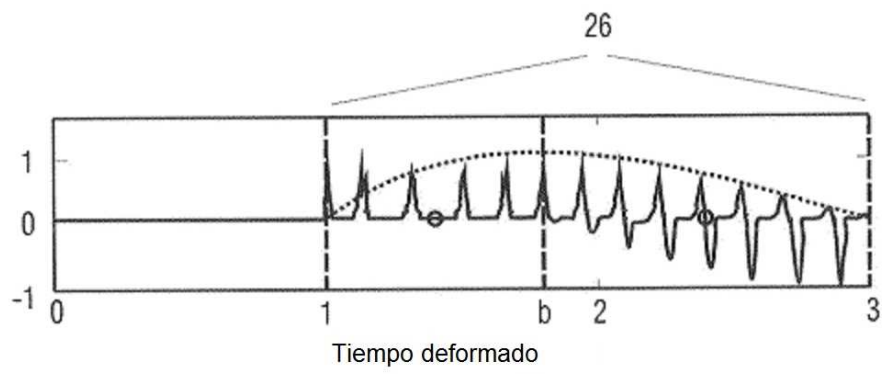


FIG 4B

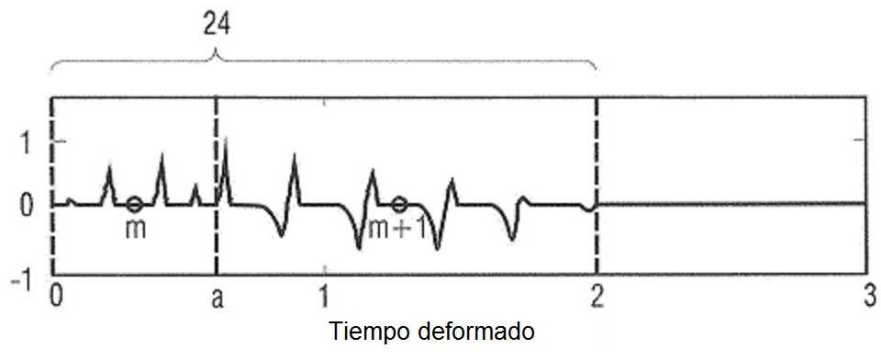


FIG 5A

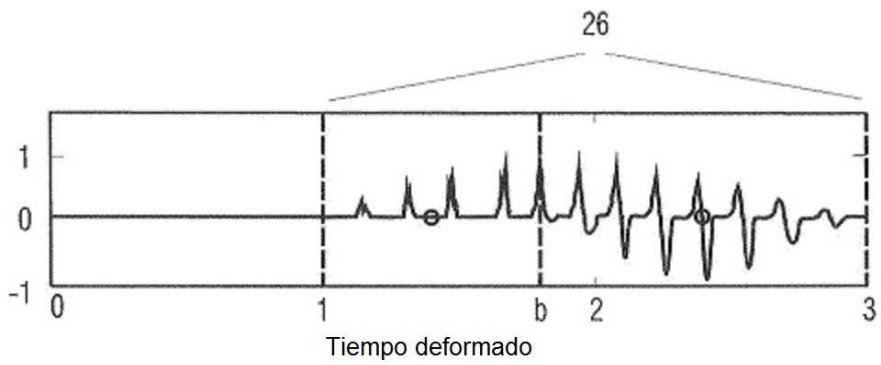


FIG 5B

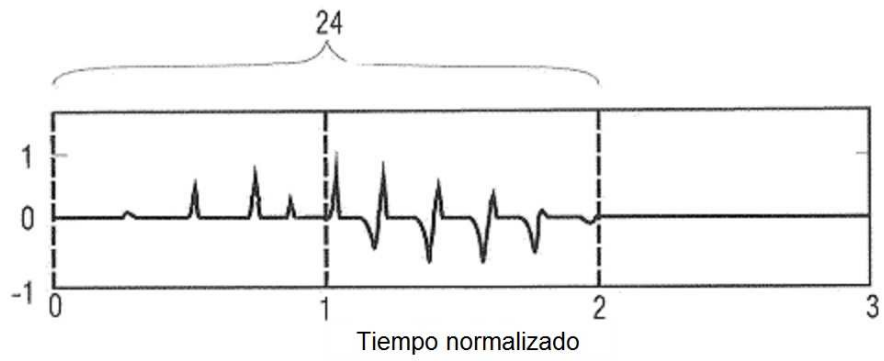


FIG 6A

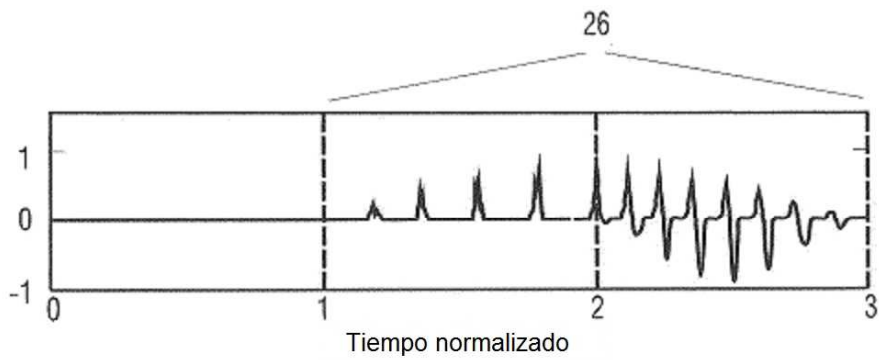


FIG 6B

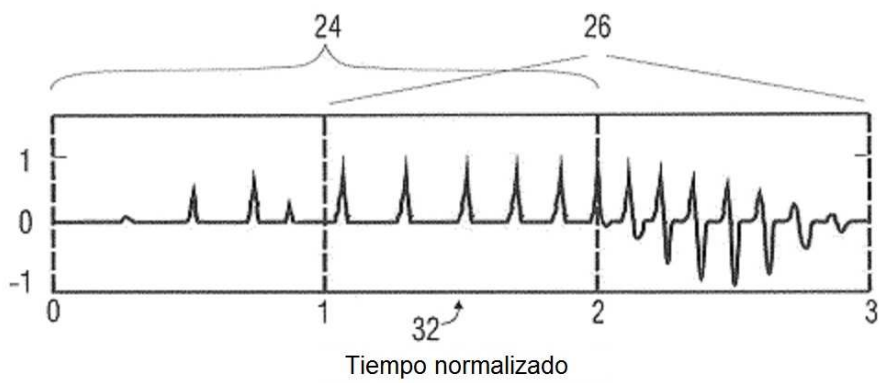


FIG 7

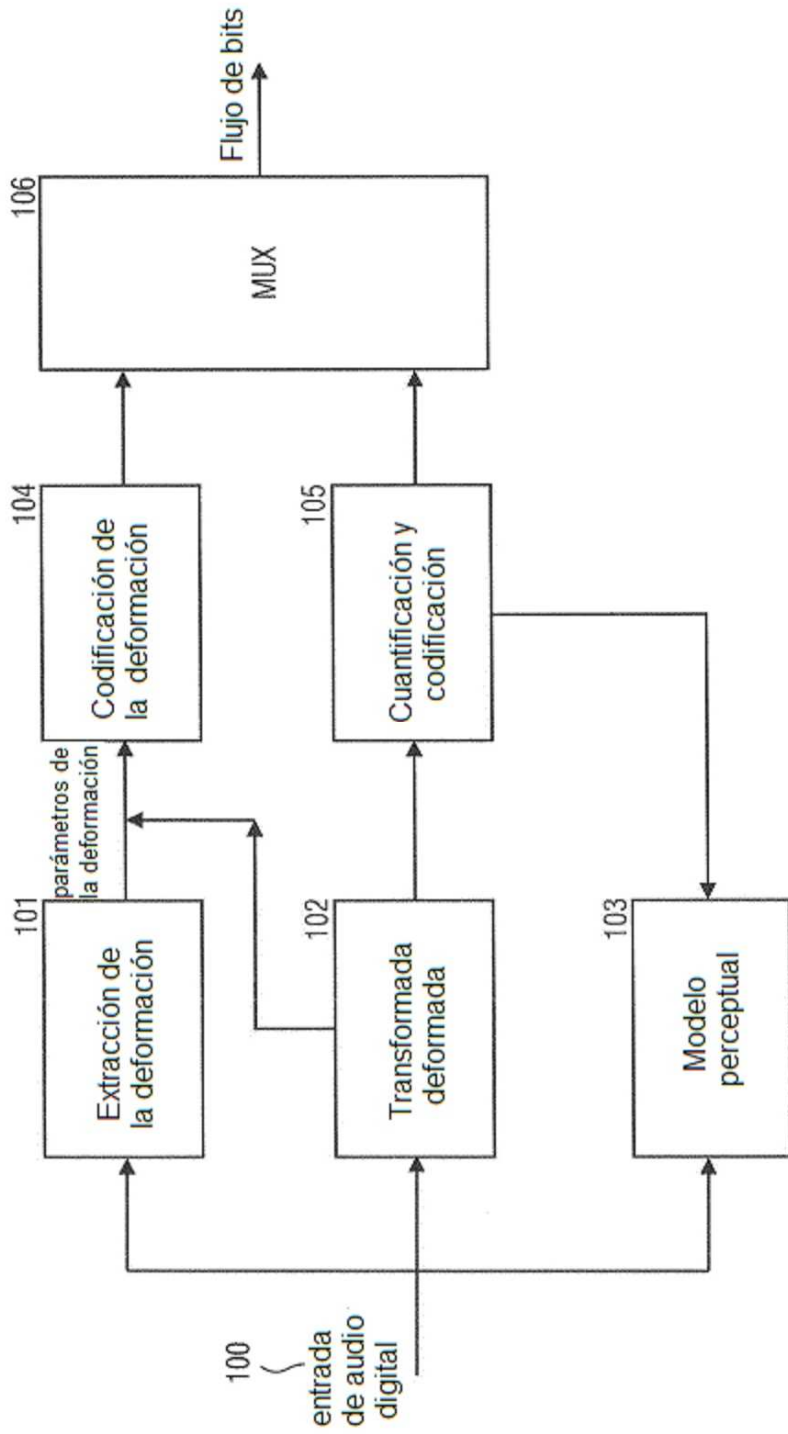


FIG 8

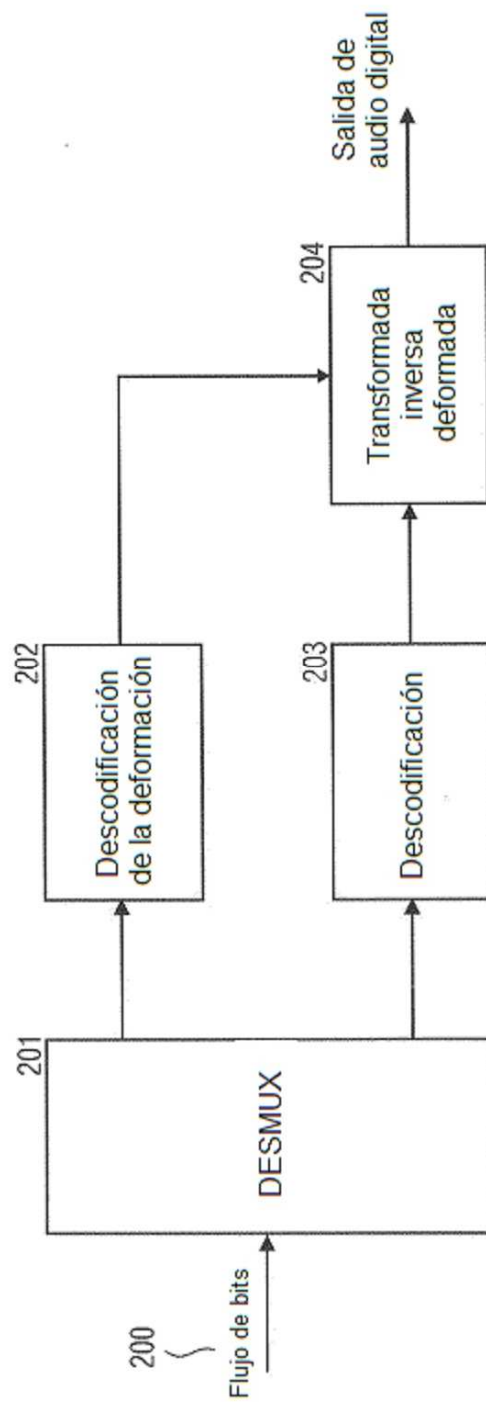


FIG 9

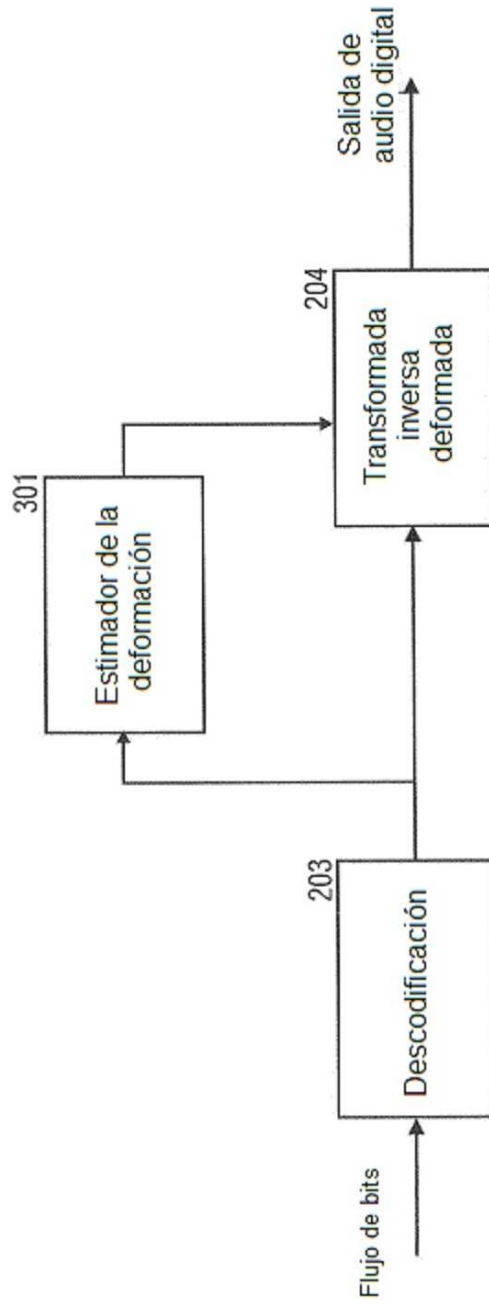


FIG 10

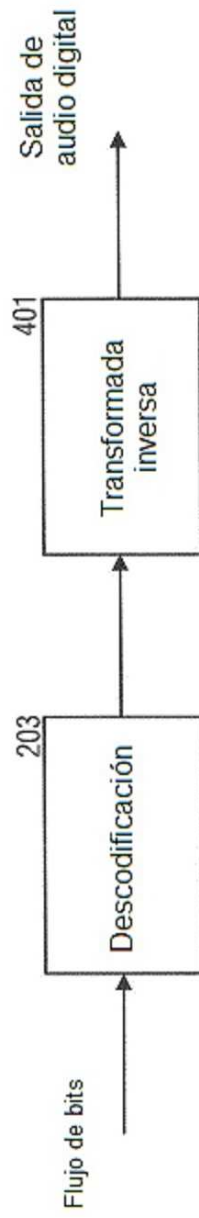


FIG 11

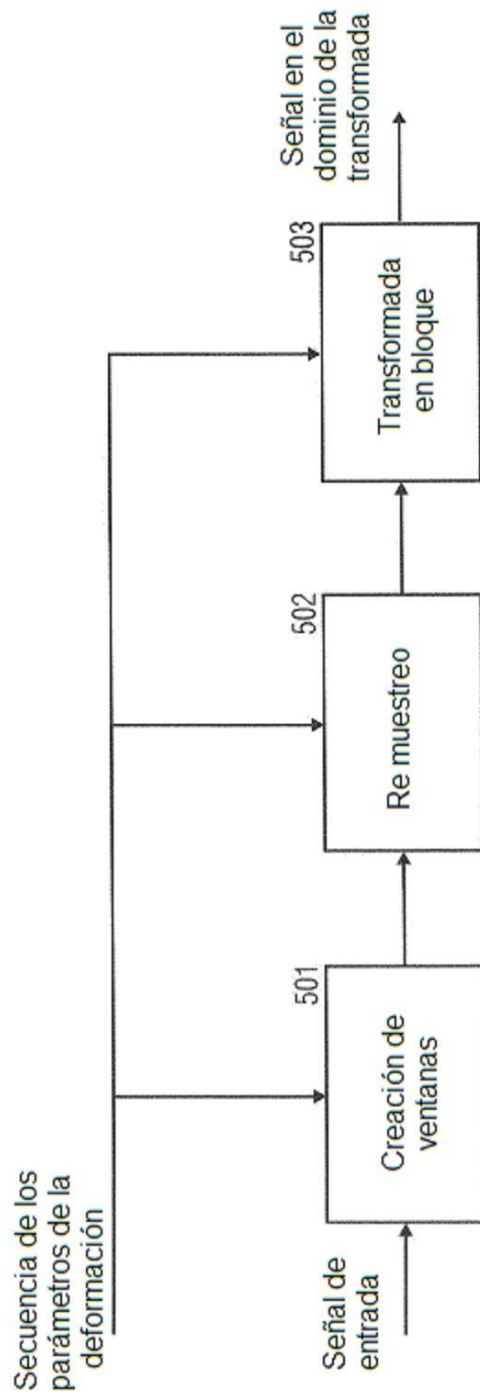


FIG 12

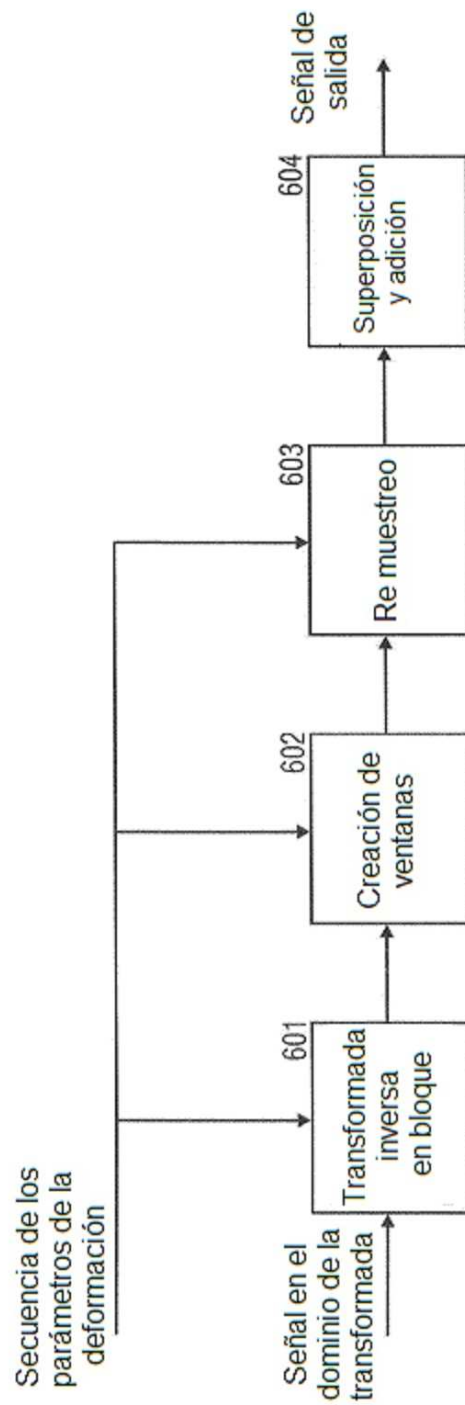


FIG 13

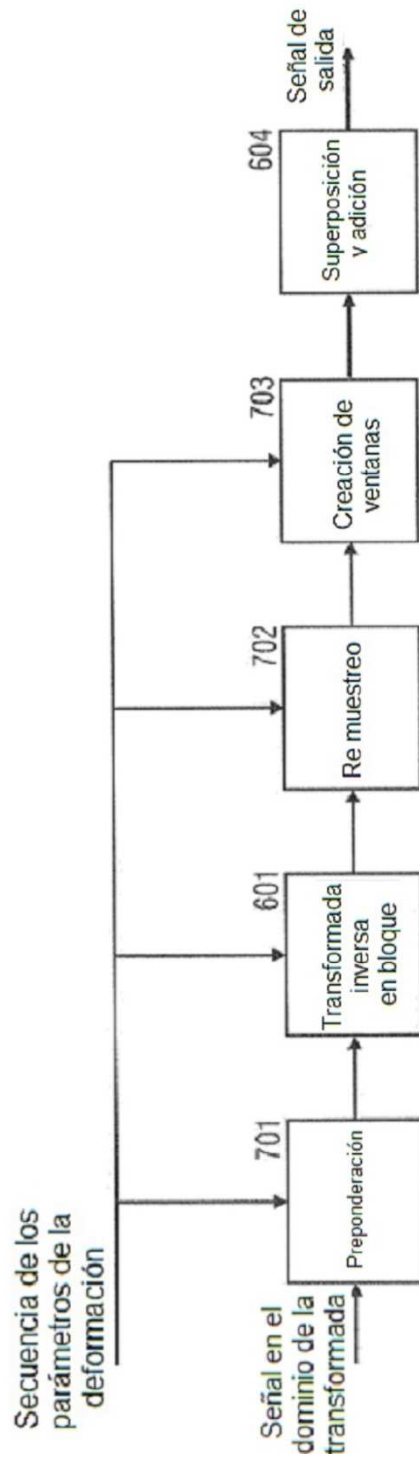


FIG 14

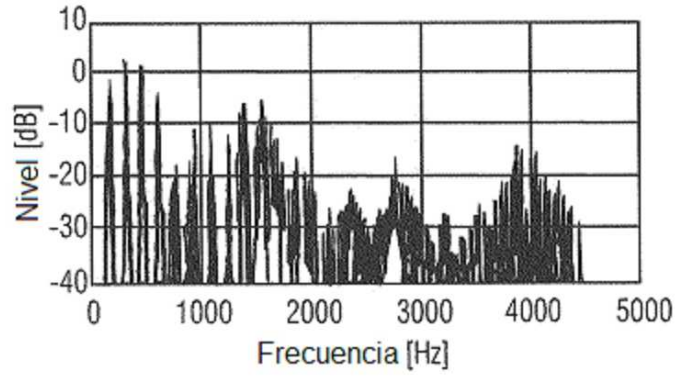


FIG 15A

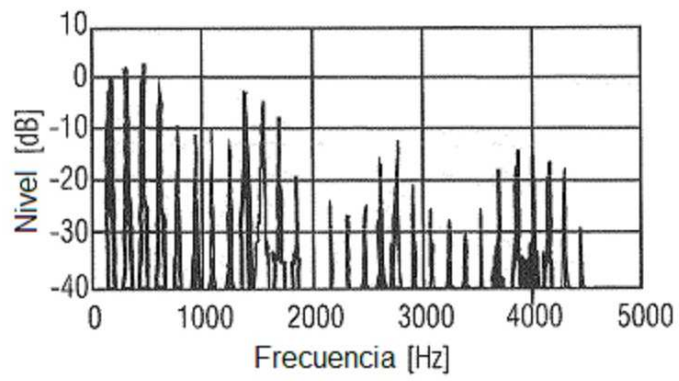


FIG 15B