

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 646 814**

51 Int. Cl.:

**G10L 19/02** (2013.01)

**G10L 19/022** (2013.01)

**G10L 19/002** (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **24.10.2006 E 10183308 (5)**

97 Fecha y número de publicación de la concesión europea: **27.09.2017 EP 2306455**

54 Título: **Codificación de señales de audio por transformada modificada deformada en el tiempo**

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:  
**18.12.2017**

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:

**SALVA FERRER, Joan**

**ES 2 646 814 T3**

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

## DESCRIPCIÓN

Codificación de señales de audio por transformada modificada deformada en el tiempo

## 5 CAMPO DE LA INVENCION

**[0001]** La presente invención se refiere a sistemas de codificación de origen de audio y en particular a esquemas de codificación de audio usando transformadas basadas en bloques.

## 10 ANTECEDENTES DE LA INVENCION

**[0002]** En la técnica se conocen varias maneras para codificar contenido de audio y vídeo. En general, por supuesto, el objetivo es codificar el contenido ahorrando bits sin degradar la calidad de reconstrucción de la señal.

15 **[0003]** Recientemente, se han desarrollado nuevos enfoques para codificar contenido de audio y vídeo, entre los que la codificación de audio perceptual basada en transformadas consigue la mayor ganancia de codificación para señales estacionarias, es decir, cuando pueden aplicarse grandes tamaños de transformada. (Véase por ejemplo T. Painter y A. Spanias: "Perceptual coding of digital audio", Proceedings of the IEEE, vol. 88, n.º 4, abril del 2000, páginas 451 a 513). Con frecuencia, las partes estacionarias de audio se modelan bien mediante un número  
20 finito fijo de sinusoides estacionarias. Una vez que el tamaño de transformada es lo suficientemente grande para resolver esas componentes, se requiere un número fijo de bits para un objetivo de distorsión dado. Aumentando adicionalmente el tamaño de transformada, se describirán segmentos más y más grandes de la señal de audio sin aumentar la demanda de bits. Sin embargo, para señales no estacionarias, resulta necesario reducir el tamaño de transformada y, por lo tanto, la ganancia de codificación disminuirá rápidamente. Para superar este problema, en  
25 caso de cambios bruscos y eventos transitorios, puede aplicarse conmutación de tamaño de transformada sin aumentar de manera significativa el coste de codificación medio. Es decir, cuando se detecta un evento transitorio, se disminuye el tamaño de bloque (tamaño de trama) de las muestras que van a codificarse conjuntamente. Por supuesto, para señales transitorias más persistentes, la tasa de bits aumentará drásticamente.

30 **[0004]** Un ejemplo interesante particular para el comportamiento transitorio persistente es la variación de tono de señales localmente armónicas, que se encuentra principalmente en las partes de voz del habla y el canto, pero que también pueden originarse a partir de los vibratos y glissandos de algunos instrumentos musicales. Teniendo una señal armónica, es decir una señal que tiene picos de señal distribuidos con igual separación a lo largo del eje de tiempo, el término tono describe la inversa del tiempo entre picos adyacentes de la señal. Por tanto, una señal de  
35 este tipo tiene un espectro armónico perfecto, que consiste en una frecuencia base igual al tono y armónicos de orden superior. En términos más generales, el tono puede definirse como la inversa del tiempo entre dos partes de señal correspondientes contiguas en una señal localmente armónica. Sin embargo, si el tono, y con él la frecuencia base, varía con el tiempo, tal como es el caso en los sonidos vocales, el espectro se hará cada vez más complejo y por tanto, más ineficiente de codificar.

40 **[0005]** Un parámetro estrechamente relacionado con el tono de una señal es la deformación de la señal. Suponiendo que la señal en el tiempo  $t$  tiene un tono igual a  $p(t)$  y que este valor de tono varía suavemente a lo largo del tiempo, la deformación de la señal en el tiempo  $t$  se define por la derivada logarítmica.

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

45

**[0006]** Para una señal armónica, esta definición de deformación es insensible a la selección particular de la componente armónica y 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 la deformación es el hercio [Hz], aunque en  
50 términos musicales, una señal con una deformación constante  $a(t)=a_0$  es un barrido con una tasa de barrido de  $a_0/\log 2$  octavas por segundo [oct/s]. Las señales del habla muestran deformaciones de hasta 10 oct/s y una deformación promedia de aproximadamente 2 oct/s.

**[0007]** Debido a que la longitud de trama (longitud de bloque) típica de los codificadores de transformada es  
55 tan grande que el cambio de tono relativo es significativo en la trama, las variaciones de tono o deformaciones de ese tamaño dan como resultado una aleatorización del análisis de frecuencia de esos codificadores. Puesto que, para una tasa de bits constante requerida, esto solo puede superarse aumentando la imprecisión de la cuantificación, este efecto resulta en la introducción de ruido de cuantificación, que con frecuencia se percibe como reverberación.

60

**[0008]** Una posible técnica para superar este problema es la deformación en el tiempo. El concepto de codificación deformada en el tiempo se explica de la mejor forma imaginando un magnetófono con velocidad

variable. Cuando se graba la señal de audio, la velocidad se ajusta dinámicamente para conseguir un tono constante en todos los segmentos vocales. La señal de audio resultante localmente estacionaria se codifica junto con los cambios de velocidad de cinta aplicados. En el descodificador, se realiza entonces una reproducción con los cambios de velocidad opuestos. Sin embargo, la aplicación de la deformación simple en el tiempo tal como se describió anteriormente tiene algunas desventajas significativas. En primer lugar, la velocidad de cinta absoluta termina siendo incontrolable, lo que conlleva a un incumplimiento de la duración de toda la señal codificada y de las limitaciones de ancho de banda. Para su reconstrucción, ha de transmitirse información secundaria adicional sobre la velocidad de cinta (o de manera equivalente sobre el tono de la señal), introduciendo una tara de tasa de bits sustancial, especialmente a tasas de bits bajas.

**[0009]** El enfoque común de los métodos de la técnica anterior para superar el problema de la duración incontrolable de señales deformadas en el tiempo es procesar de forma independiente segmentos consecutivos de la señal que no se solapan, es decir, tramas individuales, mediante una deformación en el tiempo, de modo que se conserve la duración de cada segmento. Este enfoque se describe por ejemplo en Yang et. al. "Pitch synchronous modulated lapped transform of the linear prediction residual of speech", Proceedings of ICSP '98, páginas 591 a 594. Una gran desventaja de un procedimiento de este tipo es que, aunque la señal procesada es estacionaria en los segmentos, el tono mostrará saltos en cada límite de segmento. Esos saltos causarán una pérdida de la eficacia de codificación del codificador de audio posterior y se introducirán discontinuidades audibles en la señal descodificada.

**[0010]** La deformación en el tiempo también se aplica en otros esquemas de codificación. Por ejemplo, el documento US-2002/0120445 describe un esquema, en el que segmentos de señal se someten a ligeras modificaciones de duración antes de la codificación de transformada basada en bloques. Esto es para evitar grandes componentes de señal en el límite de los bloques, aceptando ligeras variaciones de duración de los segmentos únicos.

Otra técnica que usa la deformación en el tiempo se describe en el documento US 6.169.970, en la que se aplica deformación en el tiempo para aumentar el rendimiento del predictor a largo plazo de un codificador del habla. En la misma línea, en el documento US 2005/0131681, se describe una unidad de procesamiento previo para la codificación CELP de señales del habla que aplica una deformación lineal por trozos entre intervalos que no se solapan, conteniendo cada uno un pulso de tono blanqueado. Finalmente, en (R.J. Sluijter y A.J.E.M. Janssen, "A time warper for speech signals" IEEE workshop on Speech Coding'99, junio de 1999, páginas 150 a 152) se describe cómo mejorar la estimación de tono del habla mediante la aplicación de una función de deformación en el tiempo cuadrática a una trama del habla. En resumen, las técnicas de deformación de la técnica anterior comparten los problemas de introducir discontinuidades en los bordes de trama y de necesitar una cantidad significativa de tasa de bits adicional para la transmisión de los parámetros que describen la variación del tono de la señal.

**[0011]** Otra técnica de deformación en el tiempo se encuentra descrita en Weruaga L., Kepesi M.: "SPEECH ANALYSIS WITH THE FAST CHIRP TRANSFORM", EUSIPCO 2004, 12th European Signal Processing Conference, Vienna, Austria, 7-10 Sep. 2004, Septiembre 2004 (2004-09), páginas 1011-1014, XP002619559.

## DESCRIPCIÓN DE LA INVENCIÓN

**[0012]** Es el objetivo de esta invención proporcionar un concepto para una codificación más eficaz de señales de audio usando deformación en el tiempo. El objeto de la presente invención se consigue mediante las reivindicaciones independientes. En las reivindicaciones dependientes se definen realizaciones específicas. El término "*tono*" debe interpretarse en un sentido general. Este término también cubre una *variación de tono* en conexión con lugares en relación a la información de deformación. Puede haber una situación en la que la información de deformación no de acceso a un tono absoluto, pero a información de tono relativo o normalizado. Así, dada una información de deformación puede llegarse a una descripción del tono de la señal, cuando se acepta obtener una forma de curva de tono correcta sin valores en el eje y.

## BREVE DESCRIPCIÓN DE LOS DIBUJOS

**[0013]** A continuación se describen realizaciones preferidas de la presente invención con referencia a los dibujos adjuntos, en los que:

La **Figura 1** muestra un ejemplo de mapas de deformación de la invención;

Las **Figuras 2 a, 2b** muestran la aplicación de una ventana dependiente de la deformación de la invención;

Las **Figuras 3 a, 3b** muestran un ejemplo del remuestreo de la invención;

Las **Figuras 4 a, 4 b** muestran un ejemplo de la síntesis de señales de la invención en el lado del descodificador;

Las **Figuras 5 a, 5 b** muestran un ejemplo de la aplicación de función ventana de la invención en el lado del descodificador;

Las **Figuras 6 a, 6 b** muestran un ejemplo de la deformación en el tiempo de la invención en el lado del descodificador;

La **Figura 7** muestra un ejemplo de un procedimiento de solapamiento y adición de la invención en el lado del descodificador;

La **Figura 8** muestra un ejemplo de un codificador de audio de la invención;  
 La **Figura 9** muestra un ejemplo de un decodificador de audio de la invención;  
 La **Figura 10** muestra otro ejemplo de un decodificador de la invención;  
 La **Figura 11** muestra un ejemplo de una implementación compatible hacia atrás de los conceptos de la invención;  
 5 La **Figura 12** muestra un diagrama de bloques para una aplicación de la codificación de la invención;  
 La **Figura 13** muestra un diagrama de bloques para un ejemplo de la decodificación de la invención;  
 La **Figura 14** muestra un diagrama de bloques de otra realización de la decodificación de la invención;  
 Las **Figuras 15 a, 15 b** muestran una ilustración de la eficacia de codificación que se puede conseguir mediante la  
 aplicación del concepto de la invención.

10

**DESCRIPCIÓN DETALLADA DE LAS REALIZACIONES PREFERIDAS**

**[0014]** Las realizaciones descritas a continuación son meramente ilustrativas de los principios de la presente  
 invención para la codificación por transformada deformada en el tiempo de señales de audio. Se entiende que las  
 15 modificaciones y variaciones de las disposiciones y los detalles descritos en el presente documento serán evidentes  
 para otros expertos en la técnica. Por tanto, se pretende limitarlas solamente al alcance de las reivindicaciones  
 inminentes de la patente y no a los detalles específicos presentados a modo de descripción y explicación de las  
 realizaciones en el presente documento.

**[0015]** A continuación, se revisan brevemente los conceptos e ideas básicos de transformadas de bloque y  
 deformación para motivar el concepto de la invención, que se analizará posteriormente con más detalle, haciendo  
 referencia a las figuras adjuntas.

**[0016]** Por lo general, los detalles específicos de la transformada deformada en el tiempo pueden derivarse  
 25 de la manera más sencilla en el dominio de señales de tiempo continuo. Los siguientes párrafos describen la teoría  
 general, que a continuación se especializará y convertirá posteriormente en su aplicación de la invención a señales  
 de tiempo discreto. La etapa principal en esta conversión es sustituir el cambio de coordenadas realizado en señales  
 de tiempo continuo con un remuestreo no uniforme de señales de tiempo discreto de modo que se conserve la  
 densidad de muestra media, es decir, que no se altere la duración de la señal de audio.

30

**[0017]** Supongamos que  $s=\psi(t)$  describe un cambio de coordenada de tiempo descrito por una función  $\psi$   
 estrictamente creciente continuamente diferenciable, que mapea el intervalo  $I$  del eje  $t$  con el intervalo  $J$  del eje  $s$ .

**[0018]**  $\psi(t)$  es por tanto una función que puede usarse para transformar el eje de tiempo de una cantidad  
 35 dependiente del tiempo, lo que es equivalente a un remuestreo en el caso discreto de tiempo. Debería indicarse que  
 en el siguiente análisis, el intervalo  $I$  del eje  $t$  es un intervalo en el dominio de tiempo normal y el intervalo  $J$  del eje  $x$   
 es un intervalo en el dominio de tiempo deformado.

**[0019]** Dada una base  $\{v_\alpha\}$  ortonormal para señales de energía finita en el intervalo  $J$ , se obtiene una base  
 40  $\{u_\alpha\}$  ortonormal para señales de energía finita en el intervalo  $I$  por medio de la regla

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

**[0020]** Dado un intervalo  $I$  de tiempo infinito, puede conseguirse la especificación local de deformación en el  
 45 tiempo segmentando  $I$  y, a continuación, construyendo  $\psi$  uniendo entre sí partes reajustadas a escala de mapas de  
 deformación normalizados.

**[0021]** Un *mapa de deformación normalizado* es una función estrictamente creciente y continuamente  
 diferenciable que mapea el intervalo  $[0,1]$  de la unidad consigo misma. Comenzando desde una secuencia de puntos  
 50  $t=t_k$  de segmentación donde  $t_{k+1} > t_k$ , y una secuencia correspondiente de mapas  $\psi_k$  de deformación normalizados, se  
 construye:

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

55 donde  $d_k = s_{k+1} - s_k$  y la secuencia  $d_k$  se ajusta de modo que  $\psi(t)$  se convierte en continuamente diferenciable. Esto  
 define  $\psi(t)$  a partir de la secuencia de mapas  $\psi_k$  de deformación normalizados hasta un cambio afín de escala del  
 tipo  $A\psi(t)+B$ .

**[0022]** Supongamos que  $\{v_{k,n}\}$  es 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 número entero  $K$ , el *factor de solapamiento*, de modo que  $v_{k,n}(s) = 0$  si  $s < s_k$  o  $s > s_{k+K}$

5 **[0023]** La presente invención se centra en casos en los que  $K \geq 2$ , debido a que el caso en el que  $K = 1$  corresponde a los métodos de la técnica anterior sin solapamiento. Debería indicarse que actualmente no se conocen muchas construcciones para  $K \geq 3$ . A continuación se desarrollará un ejemplo particular para el concepto de la invención para el caso en el que  $K = 2$ , que incluye bases trigonométricas locales que también se usan en transformadas de coseno discretas modificadas (MDCT, *Modified Discrete Cosine Transform*) y otras transformadas discretas solapadas en el tiempo.

10 **[0024]** Supongamos que la construcción de  $\{v_{k,n}\}$  de la segmentación es *local*, en el sentido de que hay un número entero  $p$ , de modo que  $v_{k,n}(s)$  no depende de  $s_i$  para  $|k-p| > k+K+p$ . Por último, supongamos que la construcción es tal que un cambio afín de segmentación a  $As_k+B$  da como resultado un cambio de la base a  $A^{-1/2}v_{k,n}((s-B)/A)$ . Entonces

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

15 es una base ortonormal deformada en el tiempo para señales de energía finita en el intervalo  $I$ , que está bien definida a partir de los puntos  $t_k$  de segmentación y la secuencia de mapas  $\psi_k$  de deformación normalizados independiente de la inicialización de las secuencias  $s_k$  y  $d_k$  de los parámetros en (2). Está adaptada a la  
 20 segmentación proporcionada en el sentido de que  $u_{k,n}(t) = 0$  si  $t < t_k$  o  $t > t_{k+K}$ , y está definida localmente en el sentido de que  $u_{k,n}(t)$  no depende ni de  $t_i$  para  $|k-p| > k+K+p$ , ni de los mapas  $\psi_i$  de deformación normalizados para  $|k-p| > k+K+p$ .

25 **[0025]** Las formas de onda (3) de síntesis son continuas pero no necesariamente diferenciables, debido al factor jacobiano  $(\psi'(t))^{1/2}$ . Por este motivo, y para reducir la carga computacional en el caso de tiempo discreto, también puede construirse un sistema *biortogonal* derivado. Supongamos que hay constantes  $0 < C_1 < C_2$  de modo que

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

30 para una secuencia  $\eta_k > 0$ . Entonces

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

define un par biortogonal de bases de Riesz para el espacio de señales de energía finita en el intervalo  $I$ .

35 **[0026]** Así,  $f_{k,n}(t)$  así como  $g_{k,n}(t)$  pueden usarse para 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.

40 **[0027]** Basándose en las consideraciones generales anteriores, se derivará un ejemplo del concepto de la invención en los párrafos posteriores para el caso de una segmentación uniforme  $t_k = k$  y un factor de solapamiento  $K=2$ , usando una base de coseno local adaptada a la segmentación resultante en el eje  $s$ .

45 **[0028]** Debería indicarse que las modificaciones necesarias para tratar con segmentaciones no uniformes son obvias de modo que el concepto de la invención es también aplicable a dichas segmentaciones no uniformes. Tal como propone por ejemplo M.W. Wickerhauser, "Adapted wavelet analysis from theory to software", A. K. Peters, 1994, capítulo 4, un punto de inicio para formar una base de coseno local es una función  $\rho$  de corte ascendente de modo que  $\rho(r) = 0$  para  $r < -1$ ,  $\rho(r) = 1$  para  $r > 1$ , y  $\rho(r)^2 + \rho(-r)^2 = 1$  en la región activa  $-1 \leq r \leq 1$ .

50 **[0029]** Dada una segmentación  $s_k$ , puede construirse entonces una ventana en cada intervalo  $s_k \leq s \leq s_{k+2}$  según

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

con puntos medios de corte  $c_k=(s_k+s_{k+1})/2$  y radios de corte  $\varepsilon_k=(s_{k+1}-s_k)/2$ . Esto corresponde a la construcción de punto medio de Wickerhauser.

5 **[0030]** Con  $l_k=c_{k+1}-c_k=\varepsilon_k+\varepsilon_{k+1}$ , se obtiene como resultado una base ortonormal a partir de

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

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

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

15 para  $k \leq k+2$ , donde  $\phi_k$  se define uniendo entre sí  $\psi_k$  y  $\psi_{k+1}$  para formar un mapa continuamente diferenciable del intervalo  $[0,2]$  consigo mismo,

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

**[0031]** Esto se obtiene poniendo

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

20

**[0032]** La construcción de  $\psi_k$  se ilustra en la figura 1, que muestra el tiempo normalizado en el eje x y el tiempo deformado en el eje y. La figura 1 debería considerarse particularmente para el caso en el que  $k = 0$ , es decir, para formar  $\phi_0(t)$  y por tanto 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 asume además que la primera trama 10 tiene una función 14 de deformación y la segunda trama 12 tiene una función 16 de deformación, derivadas con el fin de conseguir el mismo tono en las tramas individuales, cuando el eje de tiempo se transforma tal como se indica por las funciones 14 y 16 de deformación. Debería indicarse que la función 14 de deformación corresponde a  $\psi_0$  y la función 16 de deformación corresponde a  $\psi_1$ . Según la ecuación 9, se construye una función 18  $\phi_0(t)$  de deformación combinada uniendo entre sí los mapas 14 y 16 de deformación para formar un mapa continuamente diferenciable del intervalo  $[0,2]$  consigo mismo. Como resultado, el punto (1,1) se transforma en (1,a), donde a corresponde a  $2m_k$  en la ecuación 9.

35 **[0033]** Puesto que el concepto de la invención se dirige a la aplicación de deformación en el tiempo en un escenario de solapamiento y adición, el ejemplo de formar la siguiente función deformada combinada para la trama 12 y la trama 20 siguiente también se da en la figura 1. Debería indicarse que tras el principio de solapamiento y adición, para la reconstrucción completa de la trama 12, es necesario conocer ambas funciones 18 y 22 de deformación.

40

**[0034]** Debería indicarse además que unir entre sí dos funciones de deformación derivadas independientemente no es necesariamente la única manera de derivar una función  $\phi$  de deformación combinada adecuada, (18, 22) debido a que  $\phi$  también puede derivarse muy bien ajustando directamente una función de deformación adecuada a dos tramas consecutivas. Se prefiere tener una consistencia afín de las dos funciones de deformación en el solapamiento de sus dominios de definición.

45

**[0035]** Según la ecuación 6, la función ventana en la ecuación 8 se define mediante

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

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

[0036] También puede derivarse una versión biortogonal de (8) si hay constantes  $0 < C_1 < C_2$ , de modo que

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

5 para todo  $k$ . La selección  $\eta_k = l_k$  en (4) lleva a la especialización de (5) a

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

[0037] Por tanto, para el caso de tiempo continuo, se derivan funciones de síntesis y análisis (ecuación 12),  
10 que son dependientes de la función deformada combinada. Esta dependencia permite la deformación en el tiempo en un escenario de solapamiento y adición sin pérdida de información en la señal original, es decir, permite una reconstrucción perfecta de la señal.

[0038] Puede indicarse que para fines de implementación, las operaciones realizadas en la ecuación 12  
15 pueden descomponerse en una secuencia de etapas de procesamiento individuales consecutivas. Una forma particularmente interesante de hacerlo es realizar primero una aplicación de función ventana a la señal, seguida por un remuestreo de la señal a la que se le ha aplicado una función ventana y finalmente por una transformación.

[0039] Como es habitual, las señales de audio se almacenan y transmiten digitalmente como valores de  
20 muestra discretos muestreados con una frecuencia de muestreo determinada, el ejemplo proporcionado para la implementación del concepto de la invención se desarrollará a continuación adicionalmente para la aplicación en el caso discreto.

[0040] La transformada de coseno discreta modificada deformada en el tiempo (TWMDCT, *time-warped*  
25 *modified discrete cosine transform*) puede obtenerse a partir de una base de coseno local deformado en el tiempo discretizando formas de onda de síntesis e integrales de análisis. La siguiente descripción se basa en la base biortogonal (véase la ecuación 12). Los cambios necesarios para tratar con el caso ortogonal (8) consisten en una ponderación en el dominio de tiempo adicional por el factor jacobiano  $\sqrt{\phi'_k(t-k)}$ . En el caso especial en el que no se aplica deformación, ambas construcciones se reducen a la MDCT convencional. Supongamos que  $L$  es el tamaño  
30 de transformada y supongamos que la señal  $x(t)$  que va a analizarse está limitada con respecto a la banda por  $q\pi L$  (rad/s) para algún  $q < 1$ . Esto permite describir la señal mediante sus muestras en el periodo de muestreo  $1/L$ .

[0041] Los coeficientes de análisis vienen dados por

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

35 [0042] Definir la parte  $x_k(\tau) = x(\tau+k) b_k(\phi_k(\tau))$  de la señal a la que se le ha aplicado una función ventana y realizar sustituciones  $\tau = t-k$  y  $r = \phi_k(\tau)$  en la integral (13) lleva a

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

[0043] Una forma particularmente interesante de discretizar esta integral enseñada por la presente invención

$$r = r_v = m_k + (v + \frac{1}{2}) / L,$$

es seleccionar los puntos de muestra, donde a  $v$  se le da un valor de número entero.

40 Suponiendo deformación ligera y la limitación de banda descrita anteriormente, esto proporciona la aproximación

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

donde

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

[0044] El intervalo de suma en (15) se define por  $0 \leq \nu < 2$ . Incluye  $\nu=0, 1, \dots, L-1$  y se extiende más allá de este intervalo en cada extremo de modo que el número total de puntos es  $2L$ . Obsérvese que debido a la aplicación de función ventana, el resultado es insensible al tratamiento de los casos de borde, que pueden producirse si

$$m_k = (\nu_0 + \frac{1}{2}) / L$$

para algún número  $\nu_0$  entero.

[0045] Como es bien sabido que la suma (ecuación 15) puede calcularse mediante operaciones de plegado elementales seguidas por una DCT de tipo IV, puede ser apropiado descomponer las operaciones de la ecuación 15 en una serie de operaciones y transformaciones posteriores para utilizar implementaciones de hardware y software eficaces ya existentes, particularmente de DCT (*Discrete Cosine Transform*, transformada de coseno discreta). Según la integral discretizada, una señal de tiempo discreto determinada puede interpretarse como las muestras equidistantes en periodos  $1/L$  de muestreo de  $x(t)$ . Una primera etapa de aplicación de función ventana llevaría entonces 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 transformación por bloques tal como se describe mediante la ecuación 15 (introduciendo un desfase adicional que depende de  $m_k$ ), es necesario realizar un remuestreo, mapeando

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

[0046] La operación de remuestreo puede realizarse mediante cualquier método adecuado para remuestreo no equidistante.

[0047] En resumen, la MDCT deformada en el tiempo de la invención puede descomponerse en una operación de aplicación de función ventana, un remuestreo y una transformada de bloque.

[0048] Las etapas individuales se describirán brevemente a continuación con referencia a las figuras 2 a 3b. Las figuras 2 a 3b muestran las etapas de codificación MDCT deformada en el tiempo considerando solo dos bloques de señales con aplicación de función ventana de una señal con tono generado sintéticamente. Cada trama individual comprende 1024 muestras de modo que cada una de las dos tramas 24 y 26 combinadas consideradas (tramas 30 y 32 originales y tramas 32 y 34 originales) consiste en 2048 muestras de modo que las dos tramas combinadas a las que se les ha aplicado una función ventana tienen un solapamiento de 1024 muestras. Las figuras 2 a 2b muestran en el eje x el tiempo normalizado de las 3 tramas que van a procesarse. La primera trama 30 oscila entre 0 y 1, la segunda trama 32 oscila entre 1 y 2, y la trama 3 oscila entre 2 y 3 en el eje de tiempo. Por tanto, en el dominio de tiempo normalizado, cada unidad de tiempo corresponde a una trama completa que tiene 1024 muestras de señal. Las ventanas de análisis normalizadas abarcan los intervalos  $[0,2]$  y  $[1,3]$  de tiempo normalizados. El objetivo de las siguientes consideraciones es recuperar la trama 32 intermedia de la señal. Puesto que la reconstrucción de las tramas (30, 34) de señal externas requiere datos de segmentos de señal con aplicación de función ventana adyacentes, no se considerará esta reconstrucción en este caso. Puede observarse que los mapas de deformación combinados mostrados en la figura 1 son mapas de deformación derivados de la señal de la figura 2, que ilustra la combinación de la invención de tres mapas de deformación normalizados posteriores (curvas con puntos) en dos mapas de deformación de solapamiento (curvas continuas). Tal como se explicó anteriormente, los mapas 18 y 22 de deformación combinados de la invención se derivan para el análisis de señal. Además, puede observarse que debido a la invariante de deformación afín, esta curva representa un mapa deformado con la misma deformación que en los dos segmentos originales.

[0049] La figura 2 ilustra la señal original mediante un gráfico continuo. Su tren de pulsos estilizado tiene un tono que crece linealmente con el tiempo, así, tiene una deformación positiva y decreciente considerando que esa deformación está definida de modo que es la derivada logarítmica del tono. En la figura 2, las ventanas de análisis de la invención derivadas usando la ecuación 17 están superpuestas como curvas con puntos. Debería indicarse



que la desviación de ventanas simétricas convencionales (como por ejemplo en MDCT) es la mayor cuando la deformación es la mayor, es decir, en el primer segmento [0,1]. La definición matemática de las ventanas solas viene dada por el remuestreo de las ventanas de la ecuación 11, remuestreo aplicado tal y como se expresa en el segundo factor del lado derecho de la ecuación 17.

5

**[0050]** Las figuras 2a y 2b ilustran el resultado de la aplicación de una función ventana de la invención, aplicando las ventanas de la figura 2 a los segmentos de señal individuales.

**[0051]** Las figuras 3a y 3b ilustran el resultado del remuestreo dependiente de los parámetros de deformación de los bloques de señales con aplicación de función ventana de las figuras 2a y 2b, el remuestreo realizado tal y como lo indican los mapas de deformación dados por las curvas continuas de la figura 1. El intervalo [0,1] de tiempo normalizado se mapea con el intervalo [0,a] de tiempo deformado, que es equivalente a una compresión de la mitad izquierda del bloque de señales con aplicación de función ventana. En consecuencia, se realiza una expansión de la mitad derecha del bloque de señales con aplicación de función ventana, mapeando el intervalo [1,2] a [a,2]. Debido a que el mapa de deformación se deriva de la señal con el objetivo de derivar la señal deformada con tono constante, el resultado de la deformación (remuestreo según la ecuación 18) es un bloque de señales con aplicación de función ventana que tiene un tono constante. Debería indicarse que una falta de coincidencia entre el mapa deformado y la señal llevaría a un bloque de señales con un tono aún variable en este punto, que no afectaría a la reconstrucción final.

10

**[0052]** El desfase de la siguiente transformada de bloque está marcado por círculos, de modo que el intervalo [m, m+1] corresponde a las muestras  $v = 1, 0, \dots, L-1$  discretas con  $L = 1024$  en la fórmula 15. Esto significa de forma equivalente que las formas de onda de modulación de la transformada de bloque comparten un punto de simetría par en m y un punto de simetría impar en m+1. Además es importante indicar 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. En resumen, las figuras 3a y 3b describen la situación después del remuestreo de la invención descrito por la ecuación 18 que, evidentemente, depende de los parámetros de deformación.

15

**[0053]** Entonces se cuantifican y codifican muestras de dominio de transformada deformada en el tiempo de las señales de las figuras 3a y 3b y pueden transmitirse conjuntamente con información secundaria de deformación que describe mapas  $\psi_k$  de deformación normalizados a un descodificador. Debido a que la cuantificación es una técnica comúnmente conocida, la cuantificación utilizando una regla de cuantificación específica no se ilustra en las siguientes figuras, centrándose en la reconstrucción de la señal en el lado del descodificador.

20

**[0054]** En una realización de la presente invención, el descodificador recibe la secuencia de mapas de deformación junto con muestras  $d_{k,n}$  de dominio de transformada deformada en el tiempo descodificadas, donde puede suponerse que  $d_{k,n} = 0$  para  $n \geq L$  debido a la limitación de banda supuesta de la señal. Como en el lado del codificador, el punto de partida para conseguir síntesis de tiempo discreto debería ser considerar la reconstrucción de tiempo continuo usando las formas de onda de síntesis de la ecuación 12:

25

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

donde

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

y con

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

30

**[0055]** La ecuación (19) es el procedimiento de solapamiento y adición habitual de una síntesis de transformada a la que se aplicó una función ventana. Al igual que en la etapa de análisis, es ventajoso muestrear

la ecuación (21) en los puntos  $r = r_v = m_k + (v + \frac{1}{2}) / L$ , dando lugar a

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

que se calcula fácilmente mediante las siguientes etapas: en primer lugar, una DCT de tipo IV seguida por la extensión en  $2L$  en muestras que dependen del parámetro  $m_k$  de desfase según la regla  $0 \leq r_v < 2$ . A continuación, se

35

realiza una aplicación de función ventana con la ventana  $b_k(r_v)$ . Una vez encontrado  $z_k(r_v)$ , el remuestreo

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

proporciona el segmento  $y_k$  de señal en puntos  $(p + \frac{1}{2})/L$  de muestra equidistantes listos para la operación de solapamiento y adición descrita en la fórmula (19).

5

**[0056]** El método de remuestreo puede elegirse de nuevo con amplia libertad y no tiene que ser el mismo que en el codificador. En una realización de la presente invención se usan métodos basados en interpolación por splines, en los que el orden de las funciones de splines puede ajustarse como una función de un parámetro  $q$  de limitación de banda para conseguir 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 con frecuencia serán suficientes splines cuadráticos.

**[0057]** La decodificación se ilustrará a continuación mediante las figuras 4a a 7 para la señal mostrada en las figuras 3a y 3b. Se destacará de nuevo que la transformada de bloque y la transmisión de los parámetros de transformada no se describe en este punto, ya que es una técnica comúnmente conocida. Como inicio para el proceso de decodificación, las figuras 4a y 4b muestran una configuración en la que ya se ha realizado la transformada de bloque inversa, dando como resultado las señales mostradas en las figuras 4a y 4b. Una característica importante de la transformada de bloque inversa es la adición de componentes de señal no presentes en la señal original de las figuras 3a y 3b, que se debe a las propiedades de simetría de las funciones de síntesis ya explicadas anteriormente. En particular, la función de síntesis tiene una simetría par con respecto a  $m$  y una simetría impar con respecto a  $m+1$ . Por tanto, en el intervalo  $[0,a]$  se añaden componentes de señal positivas en la transformada de bloque inversa, mientras que en el intervalo  $[a,2]$  se añaden componentes de señal negativas. Adicionalmente, la función ventana de la invención usada para la operación de aplicación de función ventana de síntesis se superpone como una curva de puntos en las figuras 4a y 4b.

25

**[0058]** La definición matemática de esta ventana de síntesis en el dominio de tiempo deformado viene determinada por la ecuación 11. Las figuras 5a y 5b muestran la señal, aún en el dominio de tiempo deformado, tras la aplicación de función ventana de la invención.

**[0059]** Las figuras 6a y 6b muestran finalmente el resultado del remuestreo dependiente de los parámetros de deformación de las señales de las figuras 5a y 5b.

**[0060]** Finalmente, la figura 7 muestra el resultado de la operación de solapamiento-y-adición, que es la etapa final en la síntesis de la señal (véase la ecuación 19). La operación de solapamiento-y-adición es una superposición de las formas de onda de las figuras 6a y 6b. Tal como ya se mencionó anteriormente, la única trama que debe reconstruirse completamente es la trama 32 intermedia, y, una comparación con la situación original de la figura 2 muestra que la trama 32 intermedia se reconstruye con alta fidelidad. La cancelación precisa de los componentes de señal de adición perturbadora introducida durante la transformada de bloque inversa solo es posible porque es una propiedad crucial de la presente invención que los dos mapas 14 y 22 deformados combinados en la figura 1 difieren solo por un mapa afín dentro del intervalo  $[1,2]$  de tiempo normalizado de solapamiento. Una consecuencia de esto es que existe una correspondencia entre partes de señal y ventanas en los segmentos  $[a,2]$  y  $[1,b]$  de tiempo deformado. Cuando se consideran las figuras 4a y 4b, un alargamiento lineal de los segmentos  $[1,b]$  en  $[a,2]$  hará que los gráficos de señal y las mitades de ventana describan el principio bien conocido de cancelación de solapamiento en el dominio de tiempo de la MDCT convencional. La señal, en la que ya se ha cancelado el solapamiento, puede mapearse entonces simplemente sobre el intervalo  $[1,2]$  de tiempo normalizado mediante un mapa de deformación inverso común.

**[0061]** Puede observarse que, según una realización adicional de la presente invención, puede conseguirse una reducción adicional de complejidad computacional mediante la aplicación de una etapa de filtrado previo en el dominio de frecuencia. Esto puede implementarse mediante una ponderación previa simple de los valores  $d_{k,n}$  de muestra transmitidos. 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 implementación requiere aplicar un remuestreo de B-splines al resultado de la transformada de bloque inversa antes de la operación de aplicación de función ventana. En esta realización, el remuestreo opera sobre una señal tal como se deriva mediante la ecuación 22 habiendo modificado  $d_{k,n}$ . La aplicación de la función  $b_k(r_v)$  ventana tampoco se realiza. Por tanto, en cada extremo del segmento de señal, el remuestreo debe considerar las condiciones de borde en términos de periodicidades y simetrías inducidas por la selección de la transformada de bloque. La aplicación de función ventana requerida se

realiza entonces después del remuestreo usando la ventana  $b_k \left( \phi_k \left( (p + \frac{1}{2}) / L \right) \right)$ .

**[0062]** En resumen, según una primera realización de un decodificador de la invención, la MDCT deformada en el tiempo inversa comprende, cuando se descompone en etapas individuales:

- transformada inversa
- aplicación de función ventana
- 5 • remuestreo
- solapamiento y adición.

**[0063]** Según una segunda realización de la presente invención, la MDCT deformada en el tiempo inversa comprende:

- 10 • ponderación espectral
- transformada inversa
- remuestreo
- aplicación de función ventana
- solapamiento y adición.

15

**[0064]** Puede observarse que en caso de que no se aplique deformación, es decir, el caso en el que todos los mapas de deformación normalizados sean triviales, ( $\psi_k(t)=t$ ), la realización de la presente invención tal como se detalló anteriormente coincide exactamente con la MDCT habitual.

20 **[0065]** Ahora se describirán realizaciones adicionales de la presente invención que incorporan las características mencionadas anteriormente con referencia a las figuras 8 a 15.

**[0066]** La figura 8 muestra un ejemplo de un codificador de audio de la invención que recibe una señal 100 de audio digital como entrada y que genera un flujo de bits que va a transmitirse a un decodificador que incorpora el concepto de codificación de transformada deformada en el tiempo de la invención. La señal 100 de entrada de audio digital puede ser o bien una señal de audio natural o bien una señal de audio previamente procesada, en la que por ejemplo el procesamiento previo podría ser una operación de blanqueo para blanquear el espectro de la señal de entrada. El codificador de la invención incorpora un extractor 101 de parámetros de deformación, un transformador 102 de deformación, un calculador 103 de modelo perceptual, un codificador 104 de deformación, un codificador 25 105, y un multiplexor 106. El extractor 101 de parámetros de deformación estima una secuencia de parámetros de deformación, que se introduce en el transformador 102 de deformación y en el codificador 104 de deformación. El transformador 102 de deformación deriva una representación espectral deformada en el tiempo de la señal 100 de entrada de audio digital. La representación espectral deformada en el tiempo se introduce en el codificador 105 para cuantificación y otra posible codificación, como por ejemplo codificación diferencial. El codificador 105 se controla adicionalmente mediante el calculador 103 de modelo perceptual. Así, por ejemplo, la imprecisión de cuantificación puede aumentarse cuando deben codificarse componentes de señal que están enmascaradas principalmente mediante otras componentes de señal. El codificador 104 de deformación codifica la secuencia de parámetros de deformación para reducir su tamaño durante la transmisión en el flujo de bits. Esto podría comprender por ejemplo la cuantificación de los parámetros o, por ejemplo, técnicas de codificación por entropía o codificación diferencial, así como esquemas de codificación aritmética.

**[0067]** El multiplexor 106 recibe la secuencia de parámetros de deformación codificada del codificador 104 de deformación y una representación espectral deformada en el tiempo codificada de la señal 100 de entrada de audio digital para multiplexar ambos datos en el flujo de bits emitido por el codificador.

45

**[0068]** La figura 9 ilustra un ejemplo de un decodificador de transformada deformada en el tiempo que recibe un flujo 200 de bits compatible para derivar una señal de audio reconstruida como salida. El decodificador comprende un demultiplexor 201, un decodificador 202 de deformación, un decodificador 203, y un transformador 204 de deformación inverso. El demultiplexor demultiplexa el flujo de bits en la secuencia de parámetros de deformación codificada, que se introduce en el decodificador 202 de deformación. El demultiplexor demultiplexa además la representación codificada de la representación espectral deformada en el tiempo de la señal de audio, que se introduce en el decodificador 203 que es el inverso del codificador 105 correspondiente del codificador de audio de la figura 8. El decodificador 202 de deformación deriva una reconstrucción de la secuencia de parámetros de deformación y el decodificador 203 deriva una representación espectral deformada en el tiempo de la señal de audio original. La representación de la secuencia de parámetros de deformación, así como la representación espectral deformada en el tiempo, se introducen en el transformador 204 de deformación inverso que deriva una señal de salida de audio digital que implementa el concepto de la invención de codificación de transformada solapada deformada en el tiempo de señales de audio.

60 **[0069]** La figura 10 muestra una realización adicional de un decodificador de transformada deformada en el tiempo en el que la secuencia de parámetros de deformación se deriva en el propio decodificador. La realización alternativa mostrada en la figura 10 comprende un decodificador 203, un estimador 301 de deformación, y un transformador 204 de deformación inverso. El decodificador 203 y el transformador 204 de deformación inverso comparten las mismas funcionalidades que los dispositivos correspondientes de la realización anterior y por tanto la descripción de estos dispositivos en diferentes realizaciones es completamente intercambiable. El estimador 301 de deformación deriva la deformación real de la representación espectral deformada en el tiempo emitida por el

65

descodificador 203 combinando estimaciones de tono de dominio de frecuencia anteriores con una estimación de tono de dominio de frecuencia actual. Por tanto, la secuencia de parámetros de deformación se señala de manera implícita, lo que tiene la gran ventaja de que puede ahorrarse la tasa de bits adicional porque no es necesario transmitir información de parámetros de deformación adicional en el flujo de bits introducido en el descodificador. Sin embargo, la señalización implícita de datos deformados está limitada por la resolución de tiempo de la transformada.

**[0070]** La figura 11 ilustra la compatibilidad hacia atrás del concepto de la invención, cuando se usan descodificadores de la técnica anterior que no admiten el concepto de la invención de la descodificación deformada en el tiempo. Un descodificador de este tipo no consideraría la información de parámetros de deformación adicional, descodificando así el flujo de bits en una señal de dominio de frecuencia alimentada a un transformador 401 inverso que no implementa ninguna deformación. Puesto que el análisis de frecuencia realizado por la transformación deformada en el tiempo en codificadores de la invención está bien alineado con la transformada que no incluye ninguna deformación en el tiempo, un descodificador que ignore los datos de deformación aún produciría una salida de audio significativa. Esto se realiza a expensas de una calidad de audio degradada debida a la deformación en el tiempo, que no se invierte en los descodificadores de la técnica anterior.

**[0071]** La figura 12 muestra un diagrama de bloques del método de la invención de transformación deformada en el tiempo. La transformación deformada en el tiempo de la invención comprende la aplicación 501 de función ventana, remuestreo 502, y una transformación 503 de bloque. En primer lugar, se aplica una función ventana a la señal de entrada con una secuencia de ventanas de solapamiento que depende de la secuencia de parámetros de deformación que sirve como entrada adicional para cada una de las etapas 501 a 503 de codificación individuales. Cada segmento de señal de entrada al que se le ha aplicado una función ventana se remuestra posteriormente en la etapa 502 de remuestreo, en la que se realiza un remuestreo tal como se indica mediante la secuencia de parámetros de deformación.

**[0072]** En la etapa 503 de transformación de bloque, una transformada de bloque se deriva normalmente usando una transformada trigonométrica discreta bien conocida. La transformada se realiza por tanto en el segmento de señal al que se le ha aplicado una función ventana y remuestreado. Debe observarse que la transformada de bloque también depende de un valor de desfase, que se deriva de la secuencia de parámetros de deformación. Por tanto, la salida consiste en una secuencia de tramas de dominio de transformada.

**[0073]** La figura 13 muestra un diagrama de flujo de un método de transformada deformada en el tiempo inversa. El método comprende las etapas de transformación 601 de bloque inversa, aplicación 602 de una función ventana, remuestreo 603 y solapamiento y adición 604. Cada trama de una señal de dominio de transformada se convierte en una señal de dominio de tiempo por medio de la transformación 601 de bloque inversa. Correspondiendo a la etapa de codificación, la transformada de bloque depende de un valor de desfase derivado de la secuencia de parámetros recibida que sirve como entrada adicional a la transformación 601 de bloque inversa, la aplicación 602 de una función ventana y el remuestreo 603. Al segmento de señal derivado por la transformación 601 de bloque se le aplica posteriormente una función ventana en la etapa de aplicación 602 de función ventana y se remuestra en el remuestreo 603 usando la secuencia de parámetros deformados. Finalmente, en el solapamiento y adición 604 el segmento al que se le ha aplicado una función ventana y remuestreado se añade a los segmentos transformados previamente de manera inversa en una operación de solapamiento y adición habitual, dando como resultado una reconstrucción de la señal de salida de dominio de tiempo.

**[0074]** La figura 14 muestra una realización alternativa de un transformador deformado en el tiempo inverso de la invención, que se aplica para reducir adicionalmente la complejidad computacional. El descodificador comparte parcialmente las mismas funcionalidades con el descodificador de la figura 13. Por tanto, la descripción de los mismos bloques funcionales en ambas realizaciones es completamente intercambiable. La realización alternativa difiere de la realización de la figura 13 en que aplica una ponderación 701 previa espectral antes de la transformación 601 de bloque inversa. Esta ponderación previa espectral fija es equivalente a un filtrado de dominio de tiempo con periodicidades y simetrías inducidas por la selección de la transformada de bloque. Una operación de filtrado de este tipo es parte de ciertos métodos de remuestreo basados en splines, permitiendo una reducción de la complejidad computacional del remuestreo 702 modificado posterior. Tal remuestreo debe realizarse ahora en un dominio de señal con las periodicidades y simetrías inducidas por la selección de la transformada de bloque. Por tanto, una etapa 703 de aplicación de función ventana modificada se realiza después del remuestreo 702. Finalmente, en el solapamiento y adición 604, el segmento al que se le ha aplicado una función ventana y remuestreado se añade al segmento transformado previamente de manera inversa en un procedimiento de solapamiento y adición habitual proporcionando la señal de salida de dominio de tiempo reconstruida.

**[0075]** Las figuras 15a y 15b muestran la intensidad del concepto de la invención de codificación deformada en el tiempo, mostrando representaciones espectrales de la misma señal con y sin aplicación de deformación en el tiempo. La figura 15a ilustra una trama de líneas espectrales que se originan a partir de una transformada de coseno discreta modificada con un tamaño de transformada de 1024 de un segmento de señal de habla masculina muestreado a 16 kHz. La resolución de frecuencia resultante es de 7,8 Hz y solo se trazan las primeras 600 líneas para su ilustración, correspondientes a un ancho de banda de 4,7 kHz. Tal como puede observarse a partir de la frecuencia fundamental y el gráfico, el segmento es un sonido vocal con un tono medio de aproximadamente 155

Hz. Tal como puede observarse además a partir de la figura 15a, los pocos primeros armónicos de la frecuencia de tono son claramente distinguibles, pero hacia frecuencias altas, el análisis se hace cada vez más denso y aleatorio. Esto se debe a la variación del tono en la longitud del segmento de señal que se va a analizar. Por tanto, la codificación de los intervalos de frecuencia media a alta requiere una cantidad sustancial de bits con el fin de no introducir artefactos audibles en la decodificación. Por el contrario, cuando se fija la tasa de bits, inevitablemente se obtendrá una cantidad sustancial de distorsión a partir de la demanda de aumentar la imprecisión de la cuantificación.

**[0076]** La figura 15b ilustra una trama de líneas espectrales que se originan a partir de una transformada de coseno discreta modificada deformada en el tiempo según la presente invención. Evidentemente, se ha usado la misma señal de audio masculina original que en la figura 15a. Los parámetros de transformada son los mismos que para la figura 15a, aunque el uso de una transformada deformada en el tiempo adaptada a la señal tiene el efecto drástico visible sobre la representación espectral. El carácter poco denso y organizado de la señal en el dominio de transformada deformada en el tiempo proporciona una codificación con un rendimiento de distorsión de tasa mucho mejor, incluso cuando se considera el coste de codificar los datos de deformación adicionales.

**[0077]** Tal y como se ha mencionado anteriormente, la transmisión de parámetros de deformación en lugar de la transmisión de información de velocidad o tono tiene la gran ventaja de disminuir de manera drástica la tasa de bits requerida. Por tanto, en los párrafos siguientes, se detallan varios esquemas de la invención para transmitir la información de parámetros de deformación requerida.

**[0078]** Para una señal con una deformación  $\alpha(t)$  en un tiempo  $t$ , la selección óptima de la secuencia  $\psi_k$  de mapa de deformación normalizado para las bases de coseno locales (véanse (8), (12)) se obtiene resolviendo

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

**[0079]** Sin embargo, la cantidad de información requerida para describir esta secuencia de mapas de deformación es demasiado grande y la definición y medición de valores puntuales de  $\alpha(t)$  es difícil. Con fines prácticos, se decide un intervalo  $\Delta t$  de actualización de deformación y cada mapa  $\psi_k$  de deformación se describe por  $N=1/\Delta t$  parámetros. Un intervalo de actualización de deformación de aproximadamente 10 a 20 ms es normalmente suficiente para señales del habla. De manera similar a la construcción en (9) de  $\phi_k$  a partir de  $\psi_k$  y  $\psi_{k+1}$ , puede reconstruirse un mapa de deformación normalizado continuamente diferenciable mediante N mapas de deformación normalizados a través de operaciones de reajuste a escala afines adecuadas. Ejemplos prototipo de mapas de deformación normalizados incluyen

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

donde  $\alpha$  es un parámetro de deformación. Definiendo la deformación de un mapa  $h(t)$  por  $h''/h'$ , los tres mapas consiguen una deformación igual a  $\alpha$  en  $t=1/2$ . El mapa exponencial tiene una deformación constante en todo el intervalo  $0 \leq t \leq 1$ , y para valores pequeños de  $\alpha$ , los otros dos mapas muestran una desviación muy pequeña con respecto a este valor constante.

**[0080]** Para un mapa de deformación determinado aplicado en el decodificador para el remuestreo (23), se requiere su inversa en el codificador para el remuestreo (ecuación 18). Una parte principal del esfuerzo para la inversión se origina a partir de la inversión de los mapas de deformación normalizados. La inversión de un mapa cuadrático requiere operaciones de raíz cuadrada, la inversión de un mapa exponencial requiere un logaritmo, y la inversa del mapa de Moebius racional es un mapa de Moebius con parámetro de deformación negado. Debido a que las funciones exponenciales y divisiones son comparablemente caras, una atención a una facilidad máxima de cálculo en el decodificador lleva a la selección preferida de una secuencia  $\psi_k$  de mapas de deformación cuadráticos por partes.

**[0081]** El mapa  $\psi_k$  de deformación normalizado se define entonces completamente por  $N$  parámetros  $\alpha_k(0), \alpha_k(1), \dots, \alpha_k(N-1)$  de deformación con los requisitos de que

- 50 • es un mapa de deformación normalizado;

- se reconstruye mediante copias reajustadas a escala de uno de los mapas (25) de deformación prototipo uniformes;
- es continuamente diferenciable;
- cumple

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

5 [0082] La presente invención enseña que los parámetros de deformación pueden cuantificarse linealmente, normalmente hasta un tamaño de paso de aproximadamente 0,5 Hz. A continuación se codifican los valores enteros resultantes. Como alternativa, la derivada  $\psi_k'$  puede interpretarse como una curva de tono normalizada en la que se cuantifican los valores

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

10 hasta un tamaño de paso fijo, normalmente de 0,005. En este caso, a los valores enteros resultantes se les aplica una codificación diferencial adicional, secuencialmente o de manera jerárquica. En ambos casos, la tasa de bits de información secundaria resultante es normalmente de unos pocos cientos de bits por segundo, que es solo una fracción de la tasa requerida para describir datos de tono en un códec del habla.

15 [0083] Un codificador con amplios recursos computacionales puede determinar la secuencia de datos de deformación que reduce de manera óptima el coste de codificación o maximiza una medida de dispersión de líneas espectrales. Un procedimiento menos caro es usar métodos bien conocidos para el seguimiento del tono que dan como resultado una función  $p(t)$  de tono medida y luego aproximan la curva de altura tonal con una función  $p_0(t)$  lineal por partes en aquellos intervalos en los que existe el seguimiento de tono y no muestra grandes saltos en los valores de tono. La secuencia de deformación estimada viene determinada entonces por

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

en los intervalos de seguimiento de tono. Fuera de esos intervalos la deformación se establece a cero. Obsérvese que un error sistemático en las estimaciones de tono, tal como una duplicación del periodo de tono, tiene un efecto  
25 muy reducido sobre las estimaciones de deformación.

[0084] Tal como se ilustra en la figura 10, en una realización alternativa de la presente invención, la secuencia de parámetros deformados puede derivarse de los datos de dominio de transformada descodificados por un estimador de deformación. El principio es calcular una estimación de tono de dominio de frecuencia para cada  
30 trama de datos de transformada o de tonos de bloques de señales descodificadas posteriores. La información de deformación se deriva entonces de una fórmula similar a la fórmula 28.

[0085] La aplicación del concepto de la invención se ha descrito principalmente aplicando la deformación en el tiempo de la invención en un escenario de canal de audio único. El concepto de la invención no está limitado  
35 evidentemente de ningún modo al uso en un escenario monofónico de este tipo. Puede ser además extremadamente ventajoso usar la ganancia de codificación alta que se puede conseguir mediante el concepto de la invención en aplicaciones de codificación multicanal, en las que el canal único o múltiple que tiene que transmitirse puede codificarse usando el concepto de la invención.

40 [0086] Además, la deformación podría definirse en general como una transformación del eje x de una función arbitraria que depende de x. Por tanto, el concepto de la invención también puede aplicarse a escenarios en los que funciones o la representación de señales están deformadas de forma que no dependen explícitamente del tiempo. Por ejemplo, también puede aplicarse la deformación de una representación de frecuencia de una señal.

45 [0087] Además, el concepto de la invención también puede aplicarse ventajosamente a señales que están segmentadas con una longitud de segmento arbitraria y no con una longitud igual, tal como se describió en los párrafos anteriores.

[0088] El uso de las funciones base y la discretización presentadas en los párrafos anteriores debe  
50 entenderse además como un ejemplo ventajoso de aplicar el concepto de la invención. Para otras aplicaciones, también pueden usarse diferentes funciones base así como diferentes discretizaciones. Dependiendo de ciertos

requisitos de aplicación de los métodos de la invención, los métodos de la invención pueden aplicarse en hardware o en software. La aplicación puede realizarse usando un medio de almacenamiento digital, en particular un disco, DVD o un CD que tenga señales de control legibles electrónicamente almacenadas en el mismo, que funcionen conjuntamente con un sistema informático programable de modo que se realicen los métodos de la invención. En general, la presente invención es, por tanto, un producto de programa informático con un código de programa almacenado en un soporte legible por ordenador, estando operativo el código de programa para realizar los métodos de la invención cuando el producto de programa informático se ejecuta en un ordenador. En otras palabras, los métodos de la invención son, por tanto, un programa informático que tienen un código de programa para realizar al menos uno de los métodos de la invención cuando el programa informático se ejecuta en un ordenador.

10

**[0089]** Aunque lo anterior se ha mostrado y descrito particularmente con referencia a realizaciones particulares de la misma, los expertos en la técnica entenderán que pueden realizarse otros cambios diferentes en la forma y detalles sin apartarse del alcance de la misma. Debe entenderse que varios cambios pueden realizarse. El alcance de la presente invención se encuentra definido por las reivindicaciones adjuntas.

15

**REIVINDICACIONES**

1. Codificador de audio de la invención que recibe una señal de audio digital como entrada (100) y que genera un flujo de bits que va a transmitirse a un decodificador, que incorpora:
- 5 un extractor de parámetros de deformación (101) para estimar una secuencia de parámetros de deformación de la señal de entrada de audio (100);  
 un transformador de deformación (102) para recibir la secuencia del parámetro de deformación y para derivar una representación espectral deformada en el tiempo de la señal de entrada de audio(100);
- 10 un calculador de modelo perceptual (103) para recibir la señal de entrada de audio (100);  
 un codificador de deformación (104) para codificar la secuencia del parámetro de deformación para reducir su tamaño durante la transmisión dentro del flujo de bits;  
 un codificador (105) para recibir la representación espectral deformada en el tiempo para su cuantificación para obtener una representación espectral deformada en el tiempo codificada de la señal de entrada de audio, en el cual
- 15 el codificador (105) es controlado mediante el calculador de modelo perceptual (103); y  
 un multiplexor (106) para recibir y multiplexar la secuencia del parámetro de deformación codificada y la representación espectral deformada en el tiempo codificada de la señal de entrada de audio.
- 20 2. Codificador de audio según la reivindicación 1, en el cual la representación espectral deformada en el tiempo codificada de la señal de entrada de audio comprende una representación de la señal de entrada de audio que comprende una primera trama, una segunda trama que sigue a la primera trama, y una tercera trama que sigue a la segunda trama,
- 25 comprendiendo el extractor de parámetros de deformación (101) un estimador de deformación para estimar primera información de deformación para la primera y la segunda trama y para estimar una segunda información de deformación para la segunda trama y la tercera trama, describiendo la información de deformación una información de tono de la señal de audio;
- 30 comprendiendo el transformador de deformación (102) un analizador espectral para derivar primeros coeficientes espectrales para la primera y la segunda trama usando la primera información de deformación y para derivar segundos coeficientes espectrales para la segunda y la tercera trama usando la segunda información de deformación; y  
 comprendiendo el multiplexor (106) una interfaz de salida para emitir la representación de la señal de audio que
- 35 incluye los primeros y los segundos coeficientes espectrales
3. Codificador de audio según la reivindicación 2, en el que el estimador de deformación está operativo para estimar la información de deformación de modo que un tono en una representación deformada de tramas, la representación deformada derivada de tramas transformando el eje de tiempo de la señal de audio en las tramas tal
- 40 como se indica por la información de deformación, es más constante que un tono en las tramas.
4. Codificador de audio según la reivindicación 2, en el que el estimador de deformación está operativo para estimar la información de deformación de modo que se combinan una primera información de deformación intermedia de una primera trama correspondiente y una segunda información de deformación intermedia de una
- 45 segunda trama correspondiente, usando una regla de combinación.
5. Codificador de audio según la reivindicación 4, en el que la regla de combinación es de manera que se concatenan secuencias de parámetros de deformación reajustados a escala de la primera información de deformación intermedia con secuencias de parámetros de deformación reajustados a escala de la segunda
- 50 información de deformación intermedia.
6. Codificador de audio según la reivindicación 5, en el que la regla de combinación es de manera que la información de deformación resultante comprende una secuencia de parámetros de deformación continuamente diferenciable.
- 55 7. Codificador de audio según la reivindicación 2, en el que el analizador espectral está adaptado para derivar los coeficientes espectrales usando una representación ponderada de dos tramas aplicando una función ventana a las dos tramas, en el que la función ventana depende de la información de deformación.
- 60 8. Decodificador de transformada deformada en el tiempo para derivar una señal de audio reconstruida, comprendiendo:  
 un demultiplexor (201) para demultiplexar un flujo de bits en una secuencia de parámetros de deformación codificada y una representación codificada de la representación espectral deformada en el tiempo;  
 un decodificador de deformación (292) para decodificar la secuencia de parámetros de deformación codificada para
- 65 derivar una reconstrucción de la secuencia de parámetros de deformación;  
 un decodificador (203) para decodificar la representación codificada de la representación espectral deformada en el



tiempo para derivar una representación espectral deformada en el tiempo de una señal de audio; y un transformador de deformación inverso (204) para recibir la reconstrucción de la secuencia de parámetros de deformación y la representación espectral deformada en el tiempo de la señal de audio y para derivar la señal de salida de audio reconstruida utilizando una codificación de transformada solapada deformada en el tiempo.

5

9. Un decodificador de transformada deformada en el tiempo según la reivindicación 8, en el que el decodificador (203) esté configurado para reconstruir una señal de audio que tiene una primera trama, una segunda trama que sigue a la primera trama y una tercera trama que sigue a la segunda trama, usando primera información de deformación, describiendo la primera información de deformación una información de tono de la señal de audio para la primera y la segunda trama, segunda información de deformación, describiendo la segunda información de deformación una información de tono de la señal de audio para la segunda y la tercera trama, primeros coeficientes espectrales para la primera y la segunda trama y segundos coeficientes espectrales para la segunda y la tercera trama, comprendiendo el descodificador (203) un procesador de valores espectrales para derivar una primera trama combinada usando los primeros coeficientes espectrales y la primera información de deformación, teniendo la primera trama combinada información sobre la primera y sobre la segunda trama, y para derivar una segunda trama combinada usando los segundos coeficientes espectrales y la segunda información de deformación, teniendo la segunda trama combinada información sobre la segunda y la tercera trama; y un sintetizador para reconstruir la segunda trama usando la primera trama combinada y la segunda trama combinada.

10

10. Un decodificador de transformada deformada en el tiempo según la reivindicación 9, en el que el procesador de valores espectrales está operativo para usar funciones de base de coseno para derivar las tramas combinadas, dependiendo las funciones de base de coseno de la información de deformación de manera que usar las funciones de base de coseno sobre los coeficientes espectrales proporciona una representación no ponderada deformada en el tiempo de una trama combinada.

25

11. Un decodificador de transformada deformada en el tiempo según la reivindicación 9, en el que el procesador de valores espectrales está operativo para usar una función ventana para aplicar pesos a valores de muestra de las tramas combinadas, dependiendo la función ventana de la información de deformación de manera que cuando se aplican los pesos a la representación no ponderada deformada en el tiempo de una trama combinada, proporciona una representación deformada en el tiempo de una trama combinada.

30

12. Un decodificador de transformada deformada en el tiempo según la reivindicación 9, en el que el procesador de valores espectrales está operativo para usar información de deformación para derivar una trama combinada transformando el eje de tiempo de representaciones de tramas combinadas como se indica por la información de deformación.

35

13 Método de codificación de audio, comprendiendo:

recibir una señal de entrada de audio (100);

40

estimar (101) una secuencia de parámetros de deformación de la señal de entrada de audio;

derivar (102) una representación espectral deformada en el tiempo de la señal de entrada de audio (100) utilizando la secuencia de parámetros de deformación;

codificar (104) la secuencia de parámetros de deformación para reducir su tamaño durante la transmisión en el flujo de bits;

45

cuantificar (105) la representación espectral deformada en el tiempo para obtener una representación espectral deformada en el tiempo codificada de la señal de entrada de audio, donde la cuantificación (105) esté controlada por un calculador de modelo perceptual (103); y

multiplexar (106) la secuencia de parámetros de deformación codificada y la representación espectral deformada en el tiempo codificada de la señal de entrada de audio.

50

14. Método de decodificación de transformada deformada en el tiempo para la derivación de una señal de audio reconstruida, comprendiendo:

demultiplexar (201) un flujo de bits en una secuencia de parámetros de deformación codificada y una representación codificada de la representación espectral deformada en el tiempo;

55

decodificar (202) la secuencia de parámetros de deformación para derivar una reconstrucción de la secuencia de parámetros de deformación;

decodificar (203) la representación codificada de la representación espectral deformada en el tiempo para derivar una representación espectral deformada en el tiempo de una señal de audio; y

60

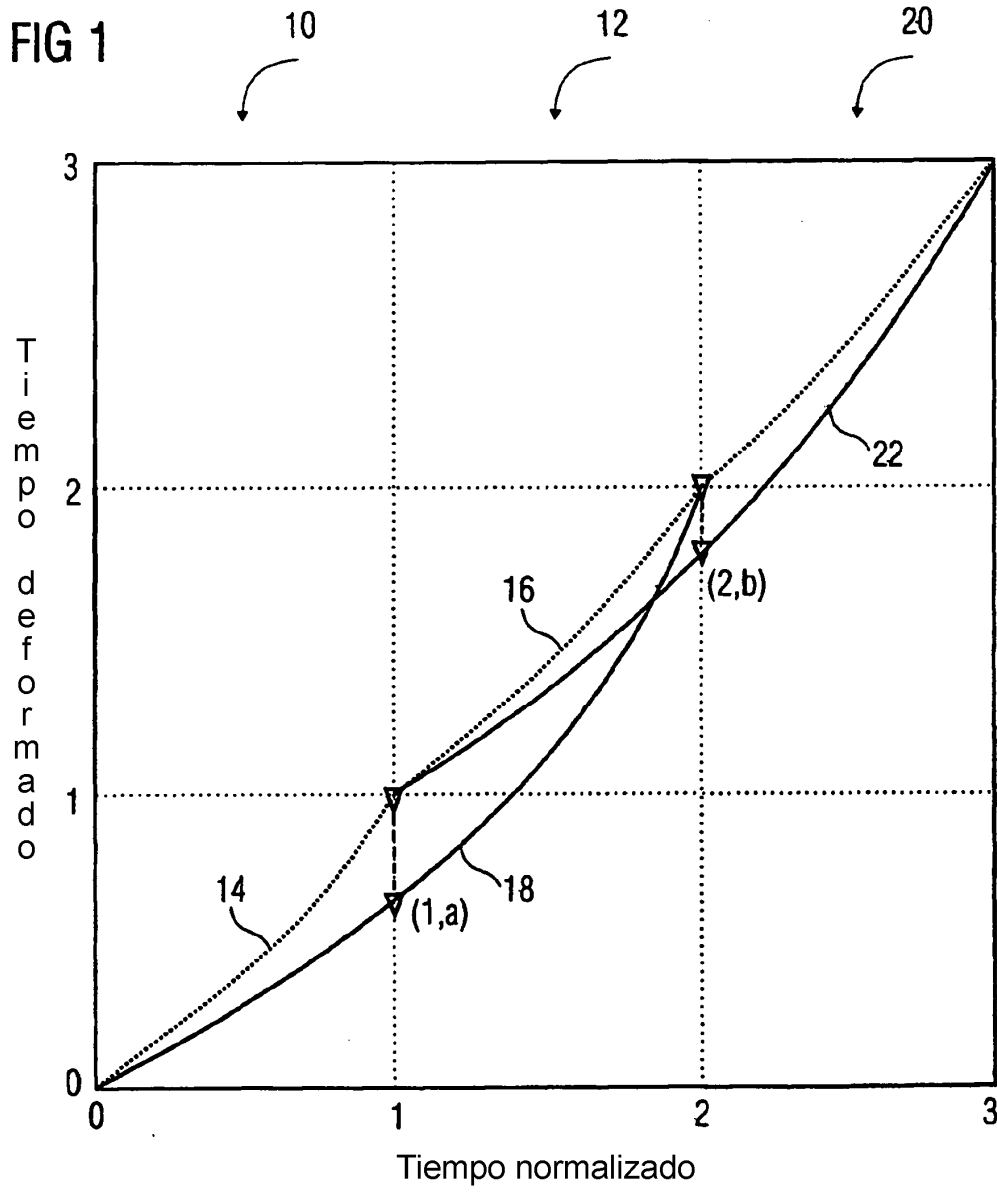
derivar (204) la señal de salida de audio reconstruida utilizando una codificación de transformada solapada deformada en el tiempo mediante la reconstrucción de la secuencia de parámetros de deformación y la representación espectral deformada en el tiempo de la señal de audio.

15. Programa informático que tiene un código de programa adaptado para realizar, cuando se ejecuta en

65

un ordenador, cualquiera de los métodos de las reivindicaciones 13 o 14.

FIG 1



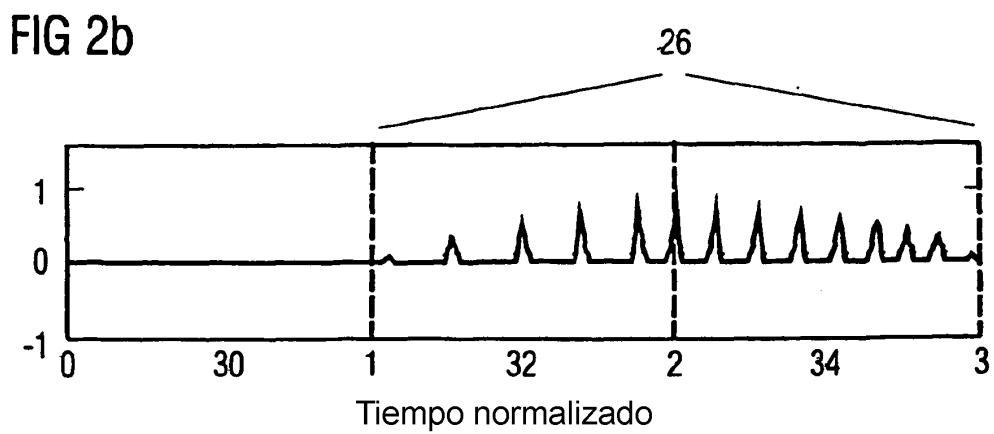
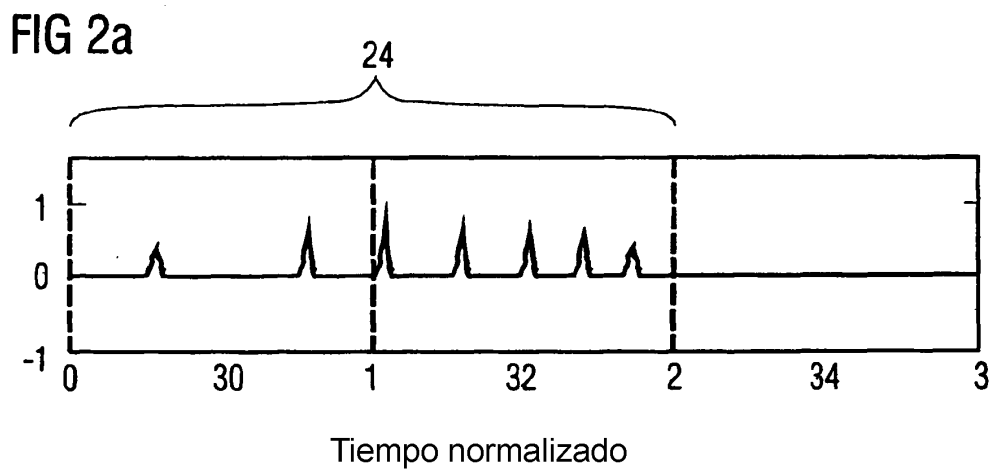
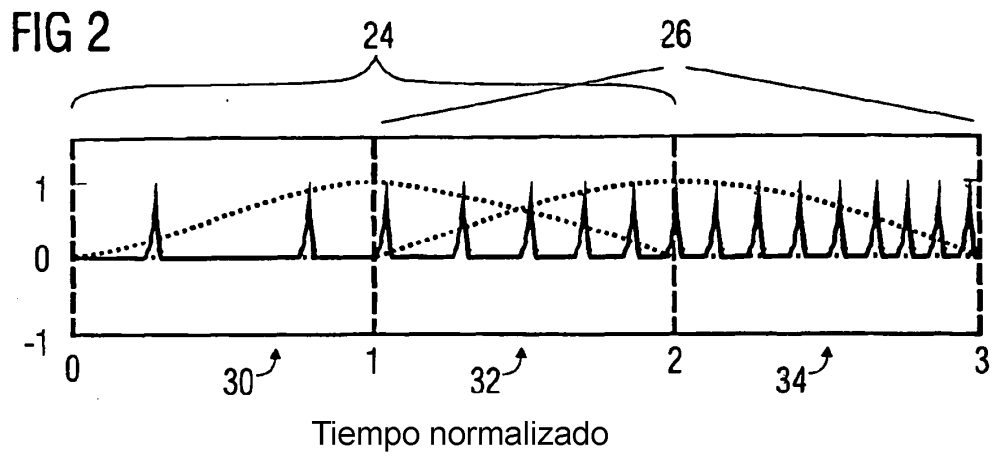


FIG 3a

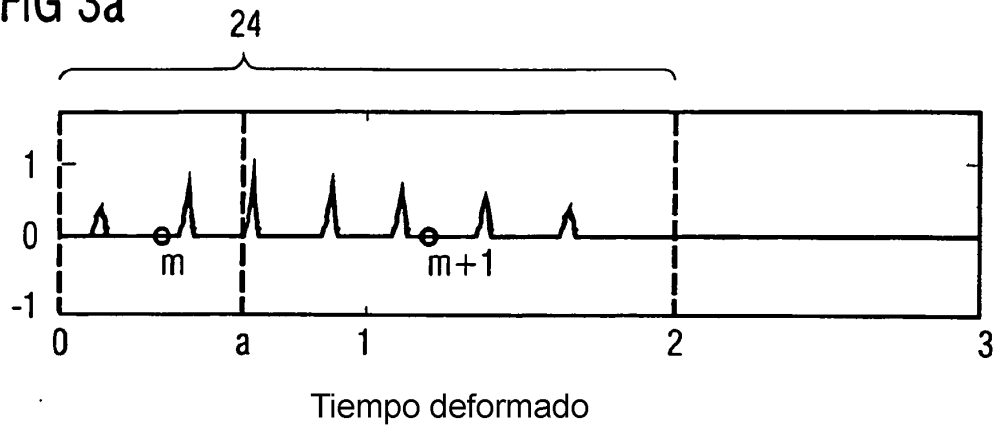


FIG 3b

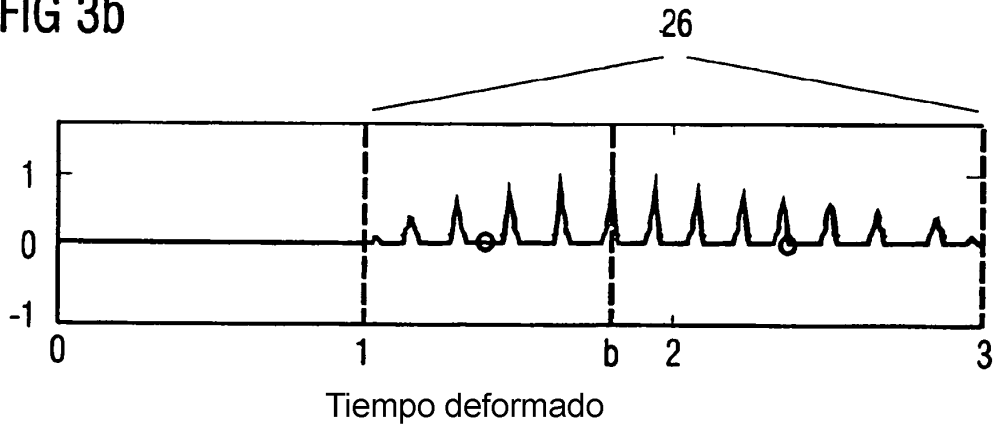


FIG 4a

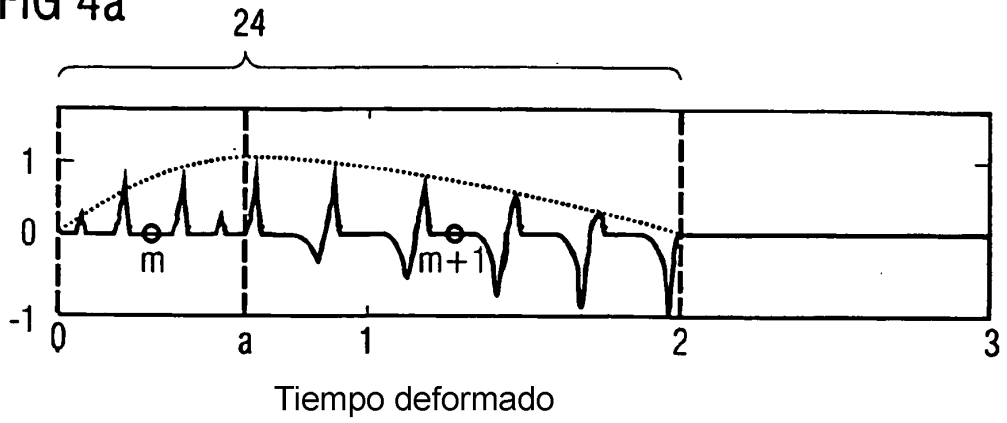


FIG 4b

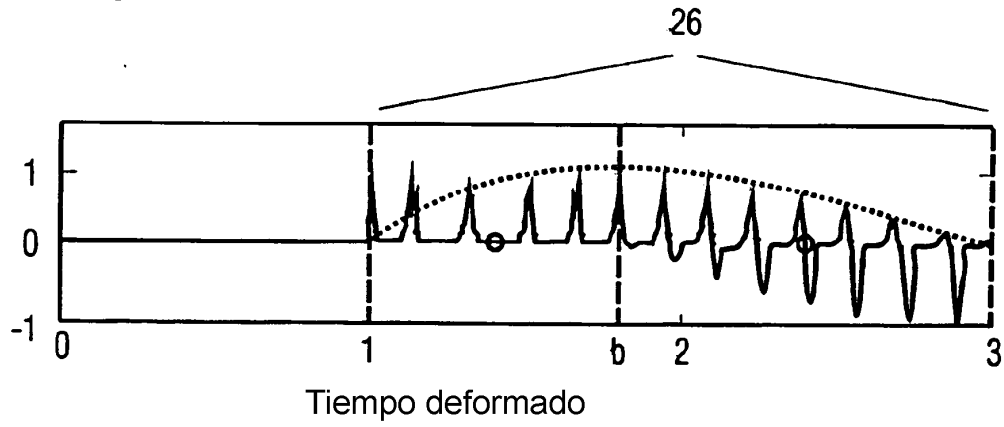


FIG 5a

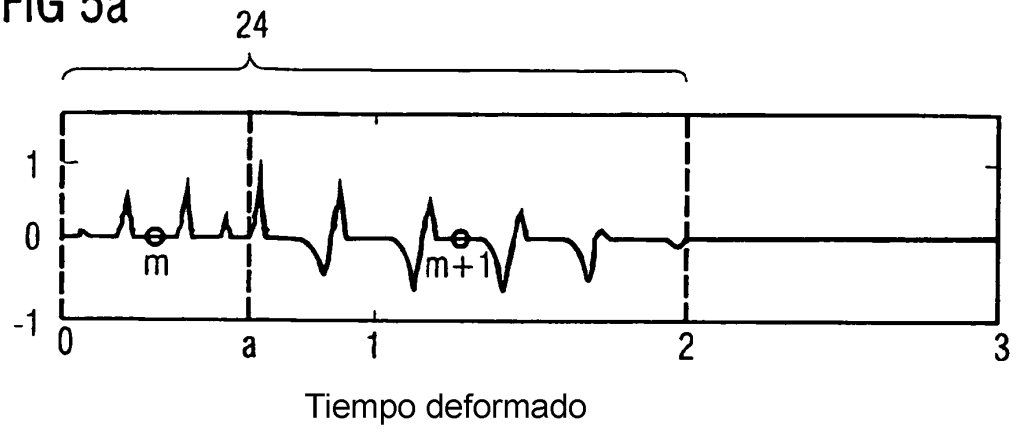
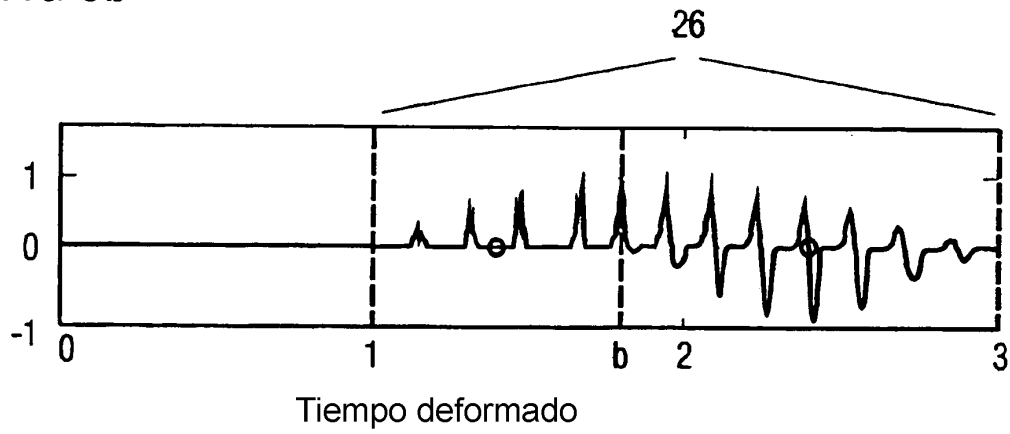


FIG 5b



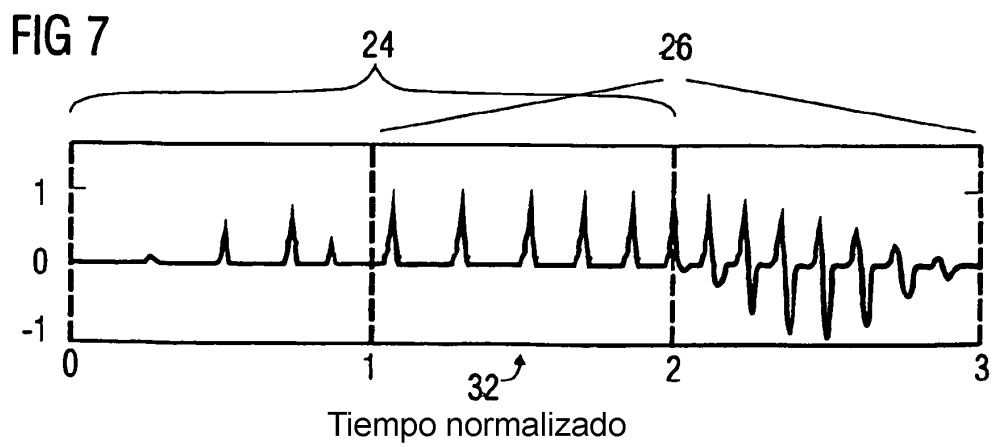
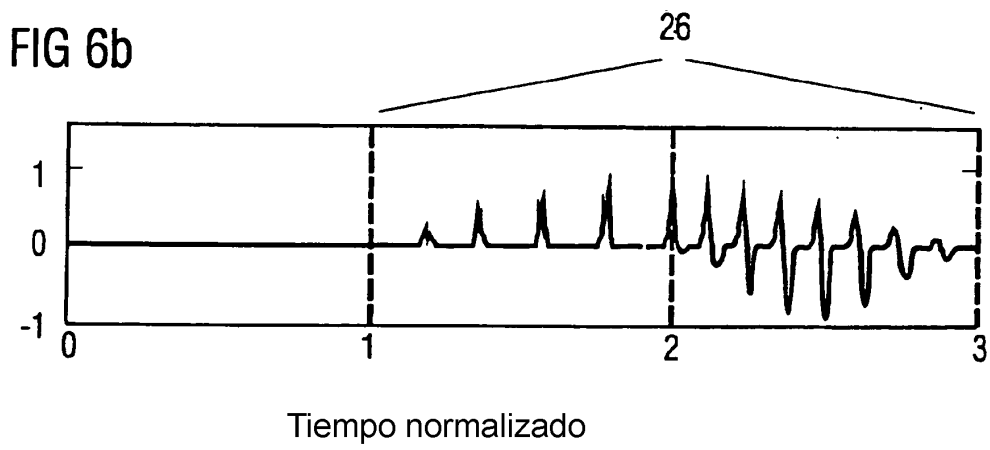
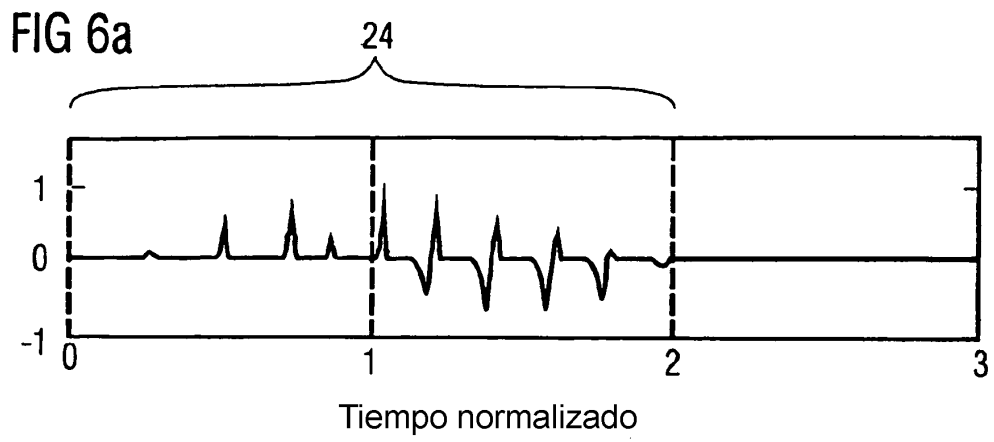


FIG 8

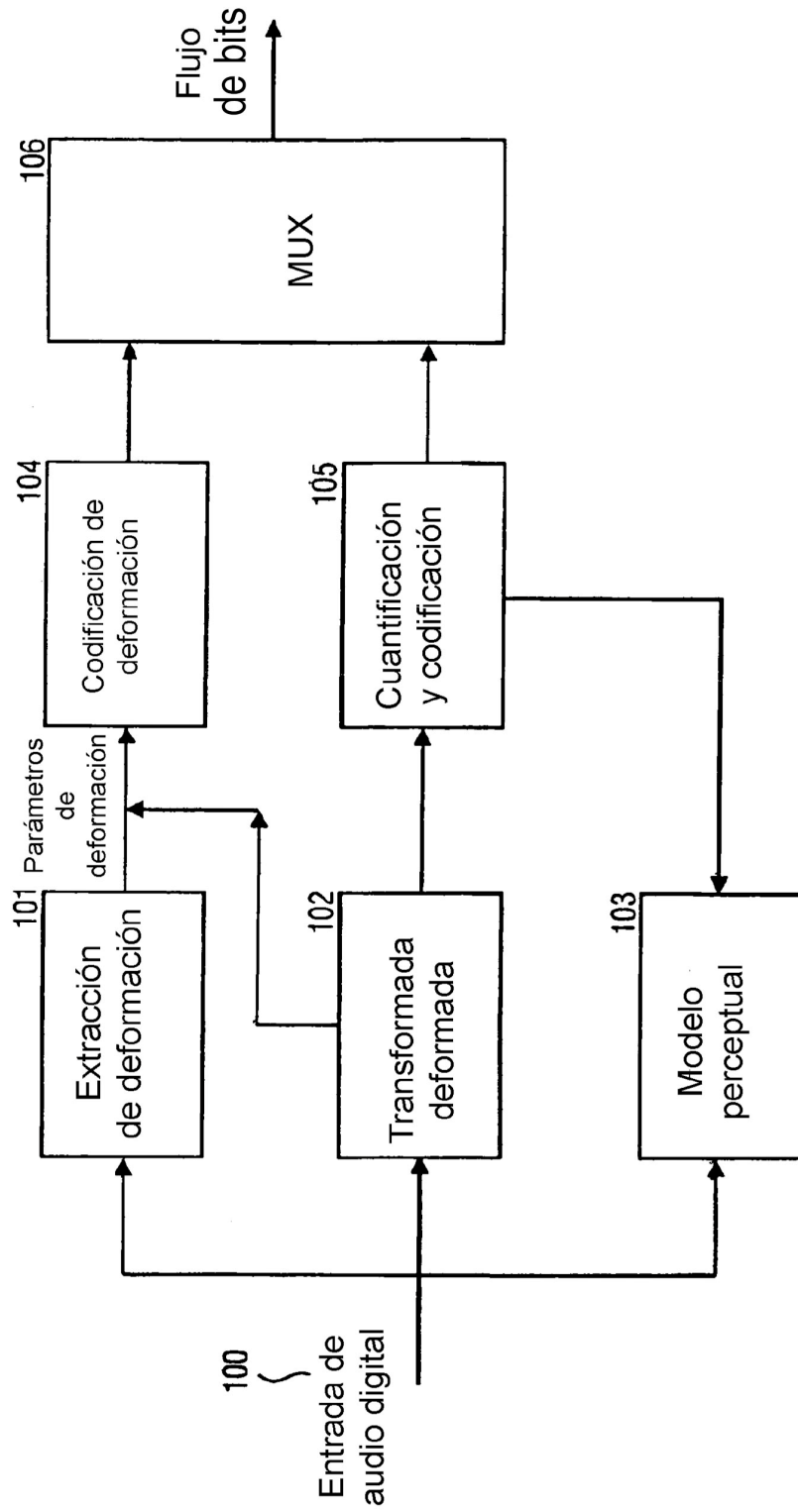




FIG 9

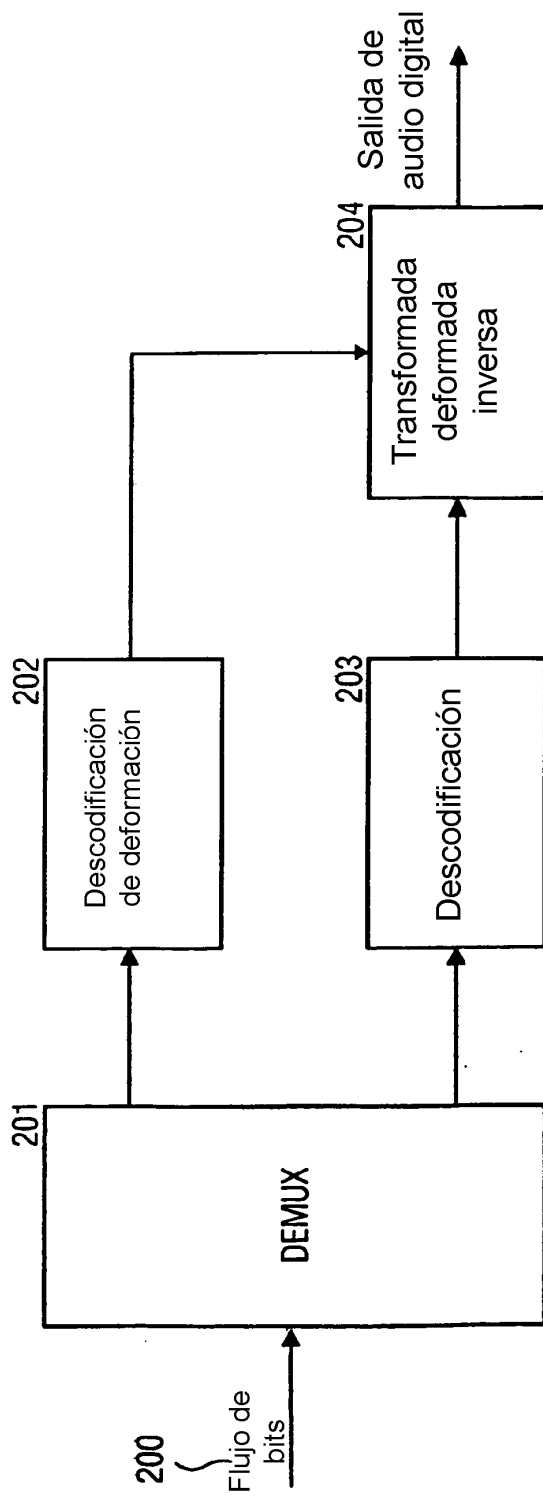


FIG 10

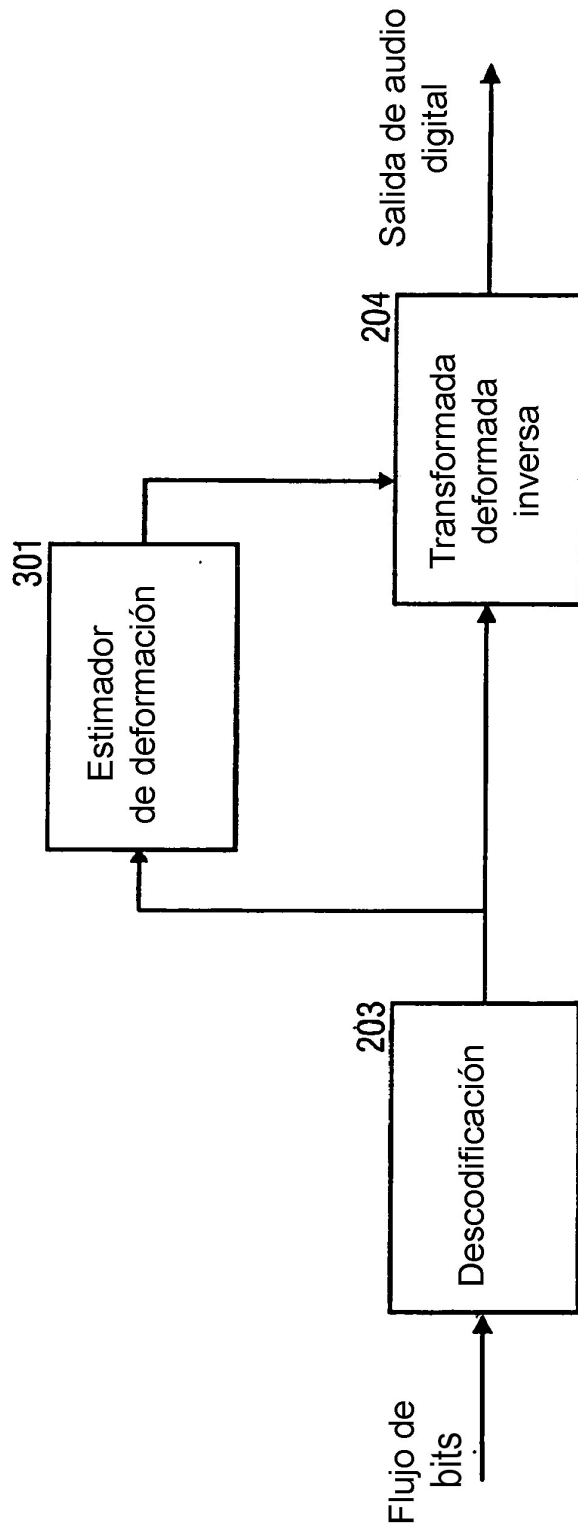


FIG 11

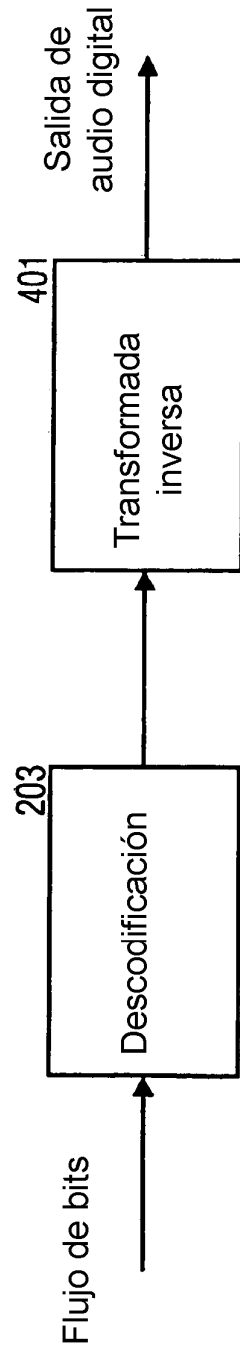


FIG 12

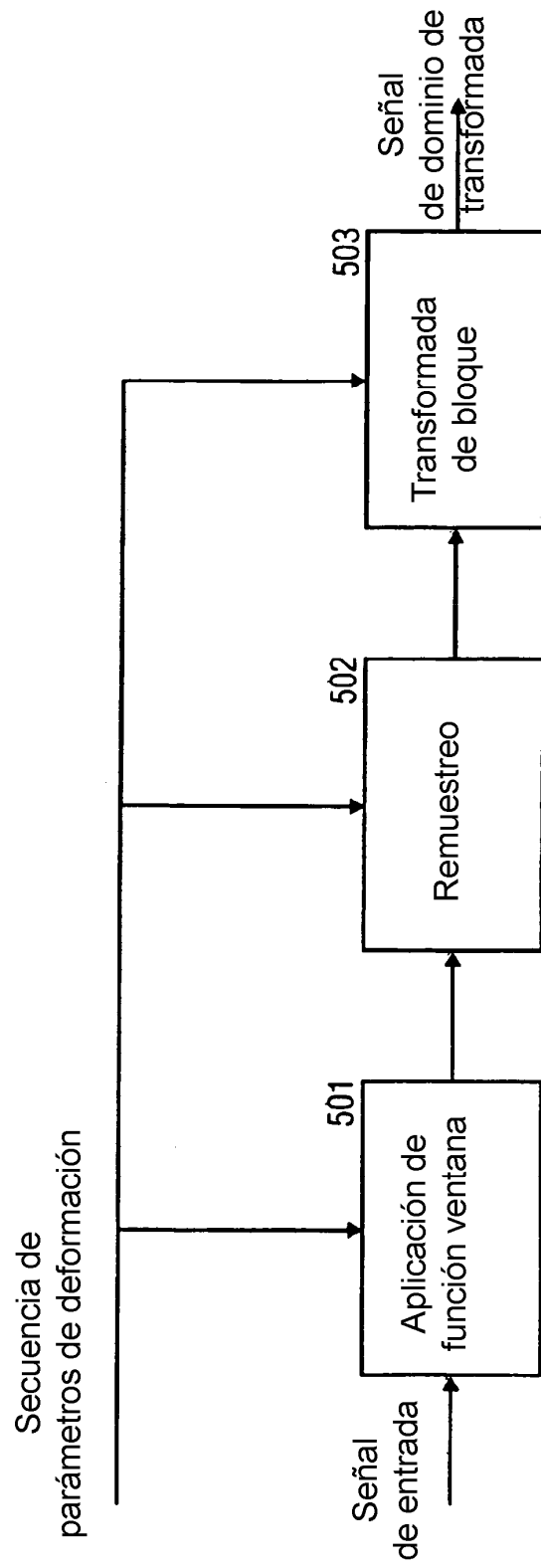


FIG 13

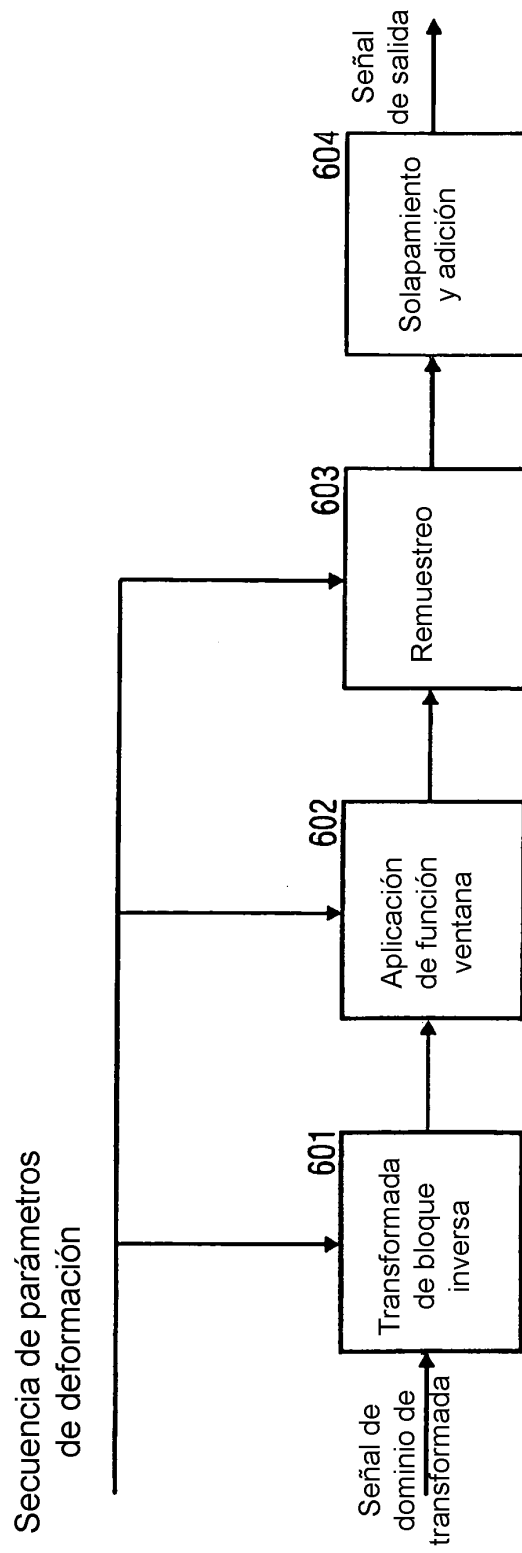


FIG 14

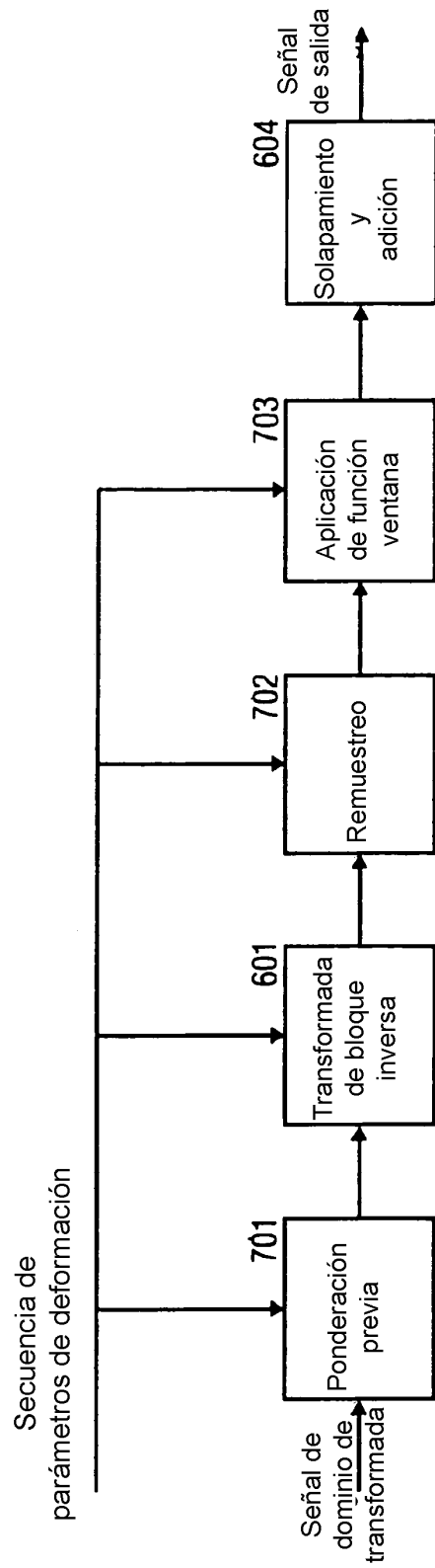


FIG 15a

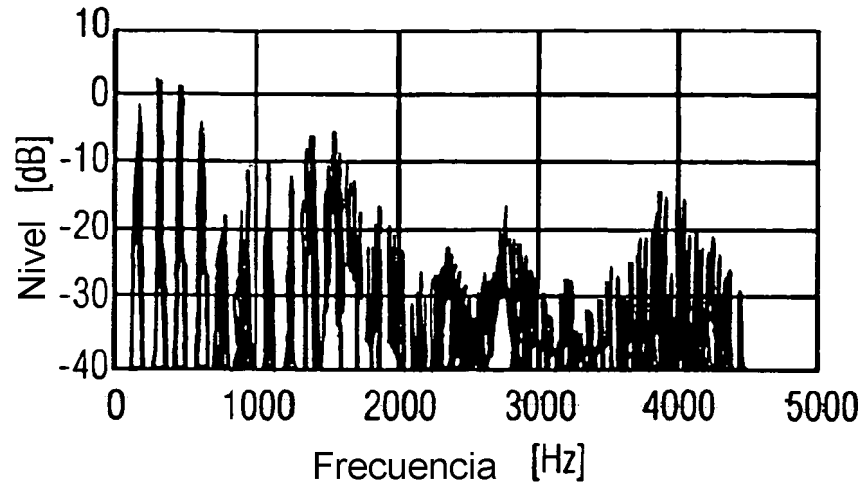


FIG 15b

