

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 375 192**

51 Int. Cl.:
G10L 19/02 (2006.01)
H04B 1/66 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 96 Número de solicitud europea: **08828229 .8**
96 Fecha de presentación: **26.08.2008**
97 Número de publicación de la solicitud: **2186087**
97 Fecha de publicación de la solicitud: **19.05.2010**

54 Título: **CODIFICACIÓN POR TRANSFORMACIÓN MEJORADA DE HABLA Y SEÑALES DE AUDIO.**

30 Prioridad:
27.08.2007 US 968159 P
11.04.2008 US 44248

45 Fecha de publicación de la mención BOPI:
27.02.2012

45 Fecha de la publicación del folleto de la patente:
27.02.2012

73 Titular/es:
TELEFONAKTIEBOLAGET L M ERICSSON
(PUBL)
STOCKHOLM S-164 83, SE

72 Inventor/es:
BRIAND, Manuel y
TALEB, Anisse

74 Agente: **de Elzaburu Márquez, Alberto**

ES 2 375 192 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 por transformación mejorada de habla y señales de audio

Campo técnico

5 La presente invención se refiere de manera general al procesamiento de señal tal como la compresión de señal y la codificación de audio, y más concretamente a la codificación de habla y audio por transformación mejorada y los dispositivos correspondientes.

Antecedentes

10 Un codificador es un dispositivo, circuitería, o programa informático que es capaz de analizar una señal tal como una señal de audio y sacar una señal en una forma codificada. La señal resultante se usa a menudo para propósitos de transmisión, almacenamiento, y/o cifrado. Por otra parte, un descodificador es un dispositivo, circuitería, o programa informático que es capaz de invertir la operación del codificador, en el que se recibe la señal codificada y saca una señal descodificada.

15 En la mayoría de los codificadores de última tecnología tales como los codificadores de audio, cada trama de la señal de entrada se analiza y se transforma desde el dominio del tiempo al dominio de la frecuencia. El resultado de este análisis es cuantificado y codificado y luego transmitido o almacenado dependiendo de la aplicación. En el lado de recepción (o cuando se usa la señal codificada almacenada) un procedimiento de descodificación correspondiente seguido por un procedimiento de síntesis hace posible restaurar la señal en el dominio del tiempo.

20 Los códec (codificador-descodificador) se emplean a menudo para la compresión/descompresión de información tal como datos de audio y vídeo para la transmisión eficiente sobre canales de comunicación de ancho de banda limitado.

25 Los denominados codificadores por transformación o de manera más general, los códec de transformación normalmente se basan en torno a una transformación del dominio del tiempo a frecuencia tal como una DCT (Transformada de Coseno Discreta), una Transformada de Coseno Discreta Modificada (MDCT) o alguna otra transformada envuelta que permita una mejor eficiencia de codificación relativa a las propiedades del sistema de audición. Una característica común de los códec de transformación es que funcionan en bloques superpuestos de muestras es decir tramas superpuestas. Los coeficientes de codificación que resultan de un análisis de transformación o un análisis sub-banda equivalente de cada trama normalmente se cuantifican y almacenan o transmiten al lado de recepción como una secuencia de bits. El descodificador, tras la recepción de la secuencia de bits, realiza la descuantificación y la transformación inversa para reconstruir las tramas de señal.

30 Los denominados codificadores perceptuales usan un modelo de codificación con pérdidas para el destino de recepción es decir el sistema auditivo humano, más que un modelo de señal fuente. La codificación de audio perceptual de esta manera supone la codificación de las señales de audio, incorporando conocimiento psicoacústico del sistema auditivo, para optimizar/reducir la cantidad de bits necesarios para reproducir fielmente la señal de audio original. Además, la codificación perceptual intenta eliminar, es decir no transmitir, o aproximar las partes de la señal que el destinatario humano no percibiría, es decir la codificación con pérdidas como opuesta a la codificación sin pérdidas de la señal fuente. El modelo típicamente se conoce como el modelo psicoacústico. En general, los codificadores perceptuales tendrán una relación señal a ruido (SNR) más baja que tendrá un codificador de forma de onda, y una calidad percibida más alta que un codificador sin pérdidas que funciona a una tasa de bit equivalente.

35 Un codificador perceptual usa un patrón de enmascaramiento de estímulos para determinar el menor número de bits necesarios para codificar es decir cuantificar cada sub-banda de frecuencia, sin introducir ruido de cuantificación audible.

40 Los codificadores perceptuales existentes que funcionan en el dominio de la frecuencia normalmente usan una combinación del denominado Umbral Absoluto de Audición (ATH) y tanto el esparcimiento de enmascaramiento tonal como de tipo ruido para calcular el denominado Umbral de Enmascaramiento (MT) [1]. En base a este umbral de enmascaramiento instantáneo, los modelos psicoacústicos existentes calculan los factores de escala que se usan para dar forma al espectro original de manera que el ruido de codificación se enmascare por los componentes de alto nivel de energía por ejemplo el ruido introducido por el codificador es inaudible [2].

45 El modelado perceptual se ha usado extensamente en la codificación de audio de alta tasa de bit. Los codificadores estandarizados, tales como de Capa III de MPEG-1 [3], la Codificación de Audio Avanzada MPEG-2 [4], logran "calidad CD" a tasas de 128 kbps y 64 kbps respectivamente para audio de banda ancha. Sin embargo, estos códec son por definición forzados a subestimar la cantidad de enmascaramiento para asegurar que la distorsión permanece inaudible. Además, los codificadores de audio de banda ancha normalmente usan un modelo de audición de alta complejidad (psicoacústico), que no es muy fiable a bajas tasas de bit (por debajo de 64 kbps).

55 El documento de la técnica anterior US2004/0131204 revela un codificador perceptual que divide una señal de audio en bloques de tiempo sucesivos, cada bloque de tiempo se divide en bandas de frecuencia, y un factor de escala se

asigna a cada banda de frecuencia. Los bits por bloque aumentan con los valores de factor de escala y las variaciones banda a banda en los valores de factor de escala. Se determina un factor de escala preliminar para cada banda de frecuencia, y se optimizan los factores de escala para cada banda de frecuencia.

Resumen

5 Debido a los problemas anteriormente mencionados, hay una necesidad de un modelo psicoacústico mejorado fiable a tasas de bit bajas manteniendo una funcionalidad de baja complejidad.

La presente invención supera estas y otras desventajas de las adaptaciones de la técnica anterior.

10 De acuerdo con la invención, se proporciona un método de codificación por transformación perceptual de señales de audio, como se fija en adelante en la reivindicación 1, y una adaptación para la codificación por transformación perceptual de las señales de audio, como se fija en adelante en la reivindicación 8.

Las ventajas adicionales ofrecidas por la invención se apreciarán cuando se lea la descripción más adelante de las realizaciones de la invención.

Breve descripción de los dibujos

15 La invención, junto con otros objetos y ventajas de la misma, se puede comprender mejor mediante la referencia a la siguiente descripción tomada junto con los dibujos anexos, en los cuales:

La Fig. 1 ilustra el codificador ejemplar adecuado para codificar audio en la banda completa;

La Fig. 2 ilustra un descodificador ejemplar adecuado para descodificar en la banda completa;

La Fig. 3 ilustra un codificador por transformación perceptual genérico;

La Fig. 4 ilustra un descodificador por transformación perceptual genérico;

20 La Fig. 5 ilustra un diagrama de flujo de un método en un modelo psicoacústico de acuerdo con la presente invención;

La Fig. 6 ilustra un diagrama de flujo adicional de una realización preferente de un método de acuerdo con la presente invención;

La Fig. 7 ilustra otro diagrama de flujo de una realización de un método de acuerdo con la presente invención.

25 **Abreviaturas**

ATH Umbral Absoluto de Audición

BS Espectro Bark

DCT Transformada Discreta de Coseno

DFT Transformada Discreta de Fourier

30 ERB Ancho de banda Rectangular Equivalente

IMDCT Transformada Discreta de Coseno Modificada Inversa

MT Umbral de Enmascaramiento

MDCT Transformada Discreta de Coseno Modificada

SF Factor de Escala

35 **Descripción detallada**

La presente invención principalmente se refiere a la codificación por transformación, y específicamente a la codificación sub-banda.

Para simplificar la comprensión de la siguiente descripción de las realizaciones de la presente invención, se describirán más adelante algunas definiciones clave.

40 El procesamiento de señal en telecomunicación algunas veces utiliza compresión-expansión como un método de mejora de la representación de señal con rango dinámico limitado. El término es una combinación de compresión y expansión, que indica de esta manera que el rango dinámico de una señal se comprime antes de la transmisión y se expande al valor original en el receptor. Esto permite a las señales con un rango dinámico grande que sean

transmitidas sobre instalaciones que tiene una capacidad de rango dinámico más pequeño.

A continuación, la invención se describirá en relación a una realización de códec ejemplar y no limitante específico adecuado para la extensión de códec de la banda completa G.722 de la ITU-T, ahora renombrada G.719 de la ITU-T. En este ejemplo particular, el códec se presenta como un códec de audio basado en transformada de baja complejidad, el cual funciona preferentemente a una velocidad de muestreo de 48 kHz y ofrece la gama de ancho de banda de audio completa desde 20 Hz hasta 20 kHz. El codificador procesa las señales PCM lineales de 16 bit de entrada en tramas de 20 ms y el códec tiene un retardo total de 40 ms. El algoritmo de codificación preferentemente se basa en codificación por transformación con tiempo de resolución adaptativo, asignación de bit adaptativa y cuantificación de vector de retícula de baja complejidad. Además, el descodificador puede sustituir los componentes de espectro no codificados por o bien ruido de relleno adaptativo de señal o bien extensión del ancho de banda.

La Fig. 1 es un diagrama de bloques de un codificador ejemplar adecuado para codificación de audio en la banda completa. La señal de entrada muestreada a 48 kHz se procesa a través de un detector transitorio. Dependiendo de la detección de un transitorio, se aplica una transformación de resolución de alta frecuencia o una de resolución de baja frecuencia (resolución de tiempo alta) en la trama de la señal de entrada. La transformación adaptativa se basa preferentemente en una Transformada Discreta de Coseno Modificada (MDCT) en caso de tramas estacionarias. Para tramas no estacionarias, se usa una transformada de resolución temporal sin una necesidad de retardo adicional y con muy poco sobredimensionamiento en complejidad. Las tramas no estacionarias preferentemente tienen una resolución temporal equivalente a tramas de 5 ms (aunque se puede seleccionar cualquier resolución arbitraria).

Puede ser beneficioso agrupar los coeficientes espectrales obtenidos en bandas de longitudes desiguales. La norma de cada banda se puede estimar y la envolvente espectral resultante que consta de las normas de todas las bandas se cuantifica y codifica. Los coeficientes entonces se normalizan por las normas cuantificadas. Las normas cuantificadas se ajustan además en base a la ponderación espectral adaptativa y se usan como entrada para la asignación de bits. Los coeficientes espectrales normalizados son el vector de retícula cuantificado y codificado en base a los bits asignados para cada banda de frecuencia. El nivel de los coeficientes espectrales no codificados se estima, codifica y transmite al descodificador. La codificación Huffman se aplica preferentemente a los índices de cuantificación tanto para los coeficientes espectrales codificados así como las normas codificadas.

La Fig. 2 es un diagrama de bloques de un descodificador ejemplar adecuado para descodificar audio en la banda completa. La bandera transitoria se descodifica primero lo cual indica la configuración de la trama, es decir estacionaria o transitoria. La envolvente espectral se descodifica y se usan los mismos, algoritmos de ajuste de norma, bit exacto y de asignación de bits en el descodificador para recalcular la asignación de bits, lo cual es esencial para la descodificación de los índices de cuantificación de los coeficientes de transformación normalizados.

Después de la descuantificación, se regeneran los coeficientes espectrales no codificados de baja frecuencia (los bit cero asignados), preferentemente usando un libro de códigos de relleno espectral construido a partir de los coeficientes espectrales recibidos (coeficientes espectrales con asignación de bits no cero).

El índice de ajuste del nivel de ruido se puede usar para ajustar el nivel de los coeficientes regenerados. Los coeficientes espectrales no codificados de alta frecuencia se regeneran preferentemente usando extensión de ancho de banda.

Los coeficientes espectrales descodificados y los coeficientes espectrales regenerados se mezclan y conducen a un espectro normalizado. La envolvente espectral descodificada se aplica conduciendo al espectro de la banda completa descodificado.

Finalmente, se aplica la transformada inversa para recuperar la señal descodificada en el dominio del tiempo. Esto se realiza preferentemente mediante la aplicación o bien de la Transformada Discreta de Coseno Modificada (IMDCT) para modos estacionarios, o bien la inversa de la transformada de resolución temporal más alta para el modo transitorio.

El algoritmo adaptado para la extensión de la banda completa se basa en la tecnología de codificación por transformación adaptativa. Ella funciona sobre tramas de audio de entrada y salida de 20ms. Debido a que la ventana de transformación (en base a la longitud de la función) es de 40ms y se usa una superposición del 50 por ciento entre las tramas de entrada y salida sucesivas, el tamaño efectivo del almacenador temporal adelantado es de 20ms. Por lo tanto, el retardo algorítmico total es de 40 ms el cual es la suma del tamaño de trama más el tamaño adelantado. Todos los otros retardos adicionales experimentados en el uso de un códec de banda completa G.722.1 (G.719 de la ITU-T) son debidos o bien a retardos de cálculo y/o bien de transmisión de red.

Un esquema de codificación general y típico en relación con un codificador por transformación perceptual se describirá con referencia a la Fig. 3. El esquema de descodificación correspondiente se presentará con referencia a la Fig. 4.

El primer paso del esquema o proceso de codificación consta de un procesamiento en el dominio del tiempo llamado *ventanaje* de la señal, que provoca una segmentación en el tiempo de una señal de audio de entrada.

La transformada del dominio del tiempo a frecuencia usada por el códec (tanto el codificador como el decodificador) podría ser, por ejemplo:

- la Transformada Discreta de Fourier (DFT), de acuerdo con la Ecuación 1,

$$X[k] = \sum_{n=0}^{N-1} w[n] \times x[n] \times e^{-j2\pi \frac{nk}{N}}, k \in \left[0, \dots, \frac{N}{2} - 1\right], \quad (1)$$

5 donde $X[k]$ es la DFT de la señal de entrada de la ventana $x[n]$. N es el tamaño de la ventana $w[n]$, n es el índice de tiempo y k el índice de contenedor de frecuencia.

- la Transformada Discreta de Coseno (DCT),

- la Transformada Discreta de Coseno Modificada (MDCT), de acuerdo con la Ecuación 2,

$$X[k] = \sum_{n=0}^{2N-1} w[n] \times x[n] \times \cos\left[\frac{\pi}{N} \left(n + \frac{N+1}{2}\right) \left(k + \frac{1}{2}\right)\right], k \in [0, \dots, N-1], \quad (2)$$

10 donde $X[k]$ es la MDCT de una señal de entrada de la ventana $x[n]$. N es el tamaño de la ventana $w[n]$, n es el índice de tiempo y k el índice de contenedor de frecuencia.

En base a cualquiera de estas representaciones en frecuencia de la señal de audio de entrada, un códec de audio perceptual aspira a la descomposición del espectro, o su aproximación, con respecto a las bandas críticas de los sistemas de audición por ejemplo la denominada escala Bark, o una aproximación de la escala Bark, o alguna otra escala de frecuencia. Para la comprensión adicional, la escala Bark es una escala estandarizada de frecuencia, en que cada "Bark" (llamado después Barkhausen) constituye un ancho de banda crítico.

15

Este paso se puede lograr mediante una agrupación de frecuencia de los coeficientes de transformación de acuerdo con una escala perceptual establecida de acuerdo con las bandas críticas, ver Ecuación 3.

$$X_b[k] = \{X[k]\}, k \in [k_b, \dots, k_{b+1} - 1], b \in [1, \dots, N_b], \quad (3)$$

20 donde N_b es el número de bandas de frecuencia o psicoacústicas, k el índice de contenedor de frecuencia, y b es un índice relativo.

Como se estableció previamente, un códec de transformación perceptual se basa en la estimación del Umbral de Enmascaramiento $MT[b]$ para derivar una función de formación de frecuencia por ejemplo los Factores de Escala $SF[b]$, aplicados a los coeficientes de transformación $X_b[k]$ en el dominio de sub-banda psicoacústico. El espectro escalado $X_{s_b}[k]$ se puede definir de acuerdo con la Ecuación 4 de más abajo

25

$$X_{s_b}[k] = X_b[k] \times MT[b], k \in [k_b, \dots, k_{b+1} - 1], b \in [1, \dots, N_b] \quad (4)$$

donde N_b es el número de bandas de frecuencia o psicoacústicas, k el índice de contenedor de frecuencia, y b es un índice relativo.

30 Finalmente, el codificador perceptual entonces puede explotar el espectro escalado perceptualmente para propósitos de codificación. Como se muestra en la Fig. 3, un proceso de cuantificación y codificación puede realizar la reducción de redundancia, el cual puede ser capaz de centrarse en los coeficientes más relevantes perceptualmente del espectro original mediante el uso del espectro escalado.

En la etapa de decodificación (ver Fig. 4) la operación inversa se logra mediante el uso de la descuantificación y decodificación del flujo binario recibido por ejemplo la secuencia de bits.

35 Este paso es seguido por la Transformada inversa (MDCT – IMDCT inversa o DFT – IDFT inversa, etc.) para devolver la señal al dominio del tiempo. Finalmente, el método de superponer-añadir se usa para generar la señal de audio reconstruida perceptualmente, es decir la codificación con pérdidas dado que solamente se decodifican los coeficientes relevantes perceptualmente.

40 Para tener en cuenta las limitaciones del sistema de audición, la invención realiza un procesamiento en frecuencia adecuado que permite el escalamiento de los coeficientes de transformación de manera que la codificación no modifique la percepción final.

Consecuentemente, la presente invención permite el modelado psicoacústico para cumplir los requerimientos de las aplicaciones de complejidad muy baja. Esto se logra mediante el uso del cálculo directo y simplificado de los factores de escala. Posteriormente, una compresión-expansión/ expansión de los factores de escala permite la codificación de audio de la banda completa de tasa de bit baja con calidad de audio perceptual alta. En resumen, la técnica de la presente invención permite optimizar perceptualmente la asignación de bits del cuantificador de manera que todos los coeficientes relevantes perceptualmente se cuantifican independientemente de la señal original o del rango dinámico del espectro.

Más adelante, se describirán las realizaciones de los métodos y adaptaciones para las mejoras del modelo psicoacústico de acuerdo con la presente invención.

A continuación, se describirán los detalles del modelado psicoacústico usados para derivar los factores de escala que se pueden usar para una codificación perceptual eficiente.

Con referencia a la Fig. 5, se describirá una realización general de un método de acuerdo con la presente invención. Básicamente, se proporciona una señal de audio por ejemplo una señal de habla para la codificación. Es procesada de acuerdo con los procedimientos estándar, como se describió previamente, provocando de esta manera una señal de audio de entrada segmentada en el tiempo y ventanada. Los coeficientes de transformación se determinan inicialmente en el paso 210 para la señal de audio de entrada segmentada en el tiempo de esta manera. Posteriormente, los coeficientes agrupados perceptualmente o las sub-bandas de frecuencia perceptuales se determinan en el paso 212, por ejemplo de acuerdo con la escala Bark o alguna otra escala. Para cada tal coeficiente o sub-banda determinada, se determina un umbral de enmascaramiento en el paso 214. Además, los factores de escala se calculan para cada sub-banda o coeficiente en el paso 216. Finalmente, los factores de escala calculados de esta manera se adaptan en el paso 218 para impedir la pérdida de energía debida a la codificación de las sub-bandas relevantes perceptualmente, es decir las sub-bandas que afectan realmente la experiencia de audición en una persona o aparato que recibe.

Esta adaptación mantendrá por lo tanto la energía de las sub-bandas relevantes y por lo tanto maximizará la calidad percibida de la señal de audio descodificada.

Con referencia a la Fig. 6, se describirá una realización específica adicional de un modelo psicoacústico de acuerdo con la presente invención. La realización permite los cálculos de los Factores de Escala, SF[b] para cada sub-banda psicoacústica, b, definida por el modelo. Aunque la realización se describe con énfasis en la denominada escala Bark, solamente con un ajuste menor es igualmente aplicable a cualquier escala perceptual. Si pérdida de generalidad, consideremos una resolución de frecuencia alta para las frecuencias bajas (grupos de pocos coeficientes de transformación) e inversamente para las frecuencias altas. El número de coeficientes por sub-banda se puede definir por una escala perceptual, por el ejemplo el Ancho de Banda Rectangular Equivalente (ERB) que se considera como una buena aproximación de la denominada escala Bark, o por la resolución de frecuencia del cuantificador usado más tarde. Una solución alternativa puede ser usar una combinación de las dos dependiendo del esquema de codificación usado.

Con los coeficientes de transformación $X[k]$ como entrada, el análisis psicoacústico en primer lugar calcula el Espectro Bark BS[b] (en dB) definido de acuerdo con la Ecuación 5:

$$BS[b] = 10 \times \log_{10} \left(\sum_{k=k_b}^{k_{b+1}-1} |X[k]|^2 \right), b \in [1, \dots, N_b] \quad (5)$$

donde N_b es el número de sub-bandas psicoacústicas, k el índice de contenedor de frecuencia, y b es un índice relativo.

En base a la determinación de los coeficientes perceptuales o sub-bandas críticas por ejemplo el Espectro Bark, el modelo psicoacústico de acuerdo con la presente invención realiza el cálculo de baja complejidad anteriormente mencionado de los Umbrales de Enmascaramiento MT.

El primer paso consiste en derivar los Umbrales de Enmascaramiento MT a partir del Espectro Bark mediante la consideración de un enmascaramiento medio. No se hace diferencia entre los componentes de tono y de ruido en la señal de audio. Esto se logra mediante una disminución de la energía de 29 dB para cada sub-banda b, ver la Ecuación 6 más adelante,

$$MT[b] = BS[b] - 29, b \in [1, \dots, N_b] \quad (6).$$

El segundo paso se basa en el efecto de esparcimiento del enmascaramiento de frecuencia descrito en [2]. El modelo psicoacústico, presentado por este medio, tiene en cuenta tanto el esparcimiento hacia delante como hacia atrás dentro de una ecuación simplificada como se define por la siguiente

$$\begin{cases} MT[b] = \max(MT[b], MT[b-1] - 12.5), b \in [2, \dots, N_b] \\ MT[b] = \max(MT[b], MT[b+1] - 25), b \in [1, \dots, N_b - 1] \end{cases} \quad (7).$$

El paso final entrega un Umbral de Enmascaramiento para cada sub-banda mediante la saturación de los valores previos con el denominado Umbral Absoluto de Audición ATH como se define por la Ecuación 8

$$MT[b] = \max(ATH[b], MT[b]), b \in [1, \dots, N_b] \quad (8).$$

5 El ATH se define comúnmente como el nivel de volumen en el cual un sujeto puede detectar un sonido particular el 50% del tiempo. A partir de los Umbrales de Enmascaramiento calculados MT , el modelo de baja complejidad propuesto de la presente invención aspira a calcular los Factores de Escala, $SF[b]$, para cada sub-banda psicoacústica. El cálculo de los SF se basa tanto en un paso de normalización, como en un paso de compresión-expansión/expansión adaptativo.

10 En base al hecho de que los coeficientes de transformación están agrupados de acuerdo con una escala no lineal (ancho de banda más grande para las frecuencias altas), la energía acumulada en todas las sub-bandas para el cálculo de los MT se puede normalizar después de la aplicación del esparcimiento de enmascaramiento. El paso de normalización se puede escribir como la Ecuación 9

$$MT_{norm}[b] = MT[b] - 10 \times \log_{10}(L[N_b]), b \in [1, \dots, N_b] \quad (9),$$

15 donde $L[1, \dots, N_b]$ son la longitud (número de coeficientes de transformación) de cada sub-banda psicoacústica b.

Los Factores de Escala SF entonces se derivan a partir de los Umbrales de Enmascaramiento normalizados con la suposición de que los MT normalizados, MT_{norm} son equivalentes al nivel de ruido de codificación, que se puede introducir por el esquema de codificación considerado. Entonces definimos los Factores de Escala $SF[b]$ como el opuesto de los valores MT_{norm} de acuerdo con la Ecuación 10.

$$SF[b] = -MT_{norm}[b], b \in [1, \dots, N_b] \quad (10).$$

20 Entonces, los valores de los Factores de Escala se reducen de manera que el efecto de enmascaramiento se limita a una cantidad predeterminada. El modelo puede prever un rango dinámico variable (adaptativamente a la tasa de bit) o fijo de los Factores de Escala para $a = 20$ dB:

$$SF[b] = \alpha \times \frac{(SF[b] - \min(SF))}{(\max(SF) - \min(SF))}, b \in [1, \dots, N_b] \quad (11)$$

25 También es posible enlazar este valor dinámico con la velocidad de transmisión de datos disponible. Entonces, para hacer que el cuantificador se centre en los componentes de frecuencia baja, los Factores de Escala se pueden ajustar de manera que no pueda aparecer pérdida de energía para las sub-bandas relevantes perceptualmente. Típicamente, los valores de los SF bajos (más bajos que 6 dB) para las sub-bandas más bajas (frecuencias por debajo de 500 Hz) se aumentan de manera que se considerarán por el esquema de codificación como relevantes perceptualmente.

30 Con referencia a la Fig. 7 se describirá una realización adicional. Están presentes los mismos pasos según se describió con referencia a la Fig. 5. Además, los coeficientes de transformación determinados a partir del paso 210 se normalizan en el paso 211, antes de que sean usados para determinar los coeficientes o sub-bandas perceptuales en el paso 212. Además, el paso 218 de adaptación de los factores de escala está además comprendiendo un paso 219 de compresión-expansión de manera adaptativa de los factores de escala, y el paso 220 de alisar de manera adaptativa los factores de escala. Estos dos pasos 219, 220 se pueden incluir naturalmente en las realizaciones de las Fig. 5 y 6 también.

40 De acuerdo con esta realización, el método de acuerdo con la invención realiza adicionalmente una asignación adecuada de la información espectral para la gama del cuantificador usada por el códec de transformación de dominio. La dinámica de las normas espectrales de entrada se asigna adaptativamente a la gama del cuantificador para optimizar la codificación de las partes dominantes de la señal. Esto se logra mediante el cálculo de una función ponderada, que es capaz de o bien de comprimir-expandir, o bien expandir las normas espectrales originales a la gama del cuantificador. Esto permite la codificación de audio de la banda completa con calidad de audio alta a varias velocidades de transmisión de datos (velocidades media y baja) sin modificar la percepción final. Una gran ventaja de la invención es también el cálculo de baja complejidad de la función ponderada para cumplir los requerimientos

de las aplicaciones de muy baja complejidad (y bajo retardo).

5 De acuerdo con la realización, la señal a asignar al cuantificador corresponde con la norma (raíz cuadrática media) de la señal de entrada en un dominio espectral transformado (por ejemplo el dominio de la frecuencia). La descomposición de frecuencia de la sub-banda (límites de sub-banda) de estas normas (sub-bandas con índice p) tiene que asignar al cuantificador la resolución de frecuencia (sub-bandas con índice b). Las normas son entonces ajustadas en nivel y una norma dominante se calcula para cada sub-banda b de acuerdo con la norma colindante (alisada hacia delante y hacia atrás) y una energía mínima absoluta. Los detalles de la operación se describen a continuación.

10 Inicialmente, las normas ($Spe(p)$) se asignan al dominio espectral. Esto se realiza de acuerdo con la siguiente operación lineal, ver *Ecuación 12*

$$BSpe(b) = \frac{1}{H_b} \sum_{p \in J_b} Spe(p) + T_b, \quad b = 0, \dots, B_{MAX} - 1 \quad (12),$$

donde B_{MAX} es el número máximo de sub-bandas (20 para esta implementación específica). Los valores de H_b , T_b y J_b se definen en la **Tabla 1** que se basa en un cuantificador que usa 44 sub-bandas espectrales. J_b es un intervalo de suma que corresponde a los números de sub-banda del dominio transformado.

15

Tabla 1 Constante de asignación de espectro

b	J_b	H_b	T_b	$A(b)$
0	0	1	3	8
1	1	1	3	6
2	2	1	3	3
3	3	1	3	3
4	4	1	3	3
5	5	1	3	3
6	6	1	3	3
7	7	1	3	3
8	8	1	3	3
9	9	1	3	3
10	10, 11	2	4	3
11	12, 13	2	4	3
12	14, 15	2	4	3
13	16, 17	2	5	3
14	18, 19	2	5	3
15	20, 21, 22, 23	4	6	3
b	J_b	H_b	T_b	$A(b)$
16	24, 25, 26	3	6	4
17	27, 28, 29	3	6	5
18	30, 31, 32, 33, 34	5	7	7
19	35, 36, 37, 38, 39, 40, 41, 42, 43	9	8	11

El espectro asignado $BSpec(b)$ se reenvía alisado de acuerdo con la *Ecuación 13*

$$BSpe(b) = \max(BSpe(b), BSpe(b-1) - 4), \quad b = 1, \dots, B_{MAX}, \quad (13)$$

y hacia atrás alisado de acuerdo con la Ecuación 14 de más abajo

$$BSpe(b) = \max(BSpe(b), BSpe(b+1) - 4), \quad b = B_{MAX} - 1, \dots, 0 \quad (14)$$

A la función resultante se le pone un umbral y se vuelve a normalizar de acuerdo con la Ecuación 15

$$BSpe(b) = T(b) - \max(BSpe(b), A(b)), \quad b = 0, \dots, B_{MAX} - 1 \quad (15)$$

donde $A(b)$ se da por la **Tabla 1**. La función resultante, la Ecuación 16 de más adelante, además se comprime-expande o expande adaptativamente dependiendo del rango dinámico del espectro ($a=4$ en esta implementación específica)

$$BSpe(b) = \frac{\alpha}{\max\{BSpe(b)\} - \min\{BSpe(b)\}} [BSpe(b) - \min\{BSpe(b)\}] \quad (16)$$

De acuerdo con la dinámica de la señal (mín. y máx.) la función de ponderación se calcula de manera que comprime-expande la señal si su dinámica excede la gama del cuantificador, y extiende la señal si su dinámica no cubre la gama completa del cuantificador.

Finalmente, mediante el uso de la asignación del dominio de sub-banda inverso (en base a los límites originales en el dominio transformado), la función de ponderación se aplica a las normas originales para generar las normas ponderadas que alimentarán el cuantificador.

Una realización de una adaptación para habilitar las realizaciones del método de la presente invención se describirán con referencia a la Fig. 8. La adaptación comprende una unidad de entrada/salida I/O para la transmisión y la recepción de señales de audio o representaciones de señales de audio para el procesamiento. Además la adaptación comprende los medios de determinación de la transformación 310 adaptados para determinar los coeficientes de transformación representativos de una transformación tiempo a frecuencia de una señal de audio de entrada segmentada en el tiempo recibida, o la representación de tal señal de audio. De acuerdo con una realización adicional la unidad de determinación de la transformación se puede adaptar a o conectar a una unidad de norma 311 adaptada para normalizar los coeficientes determinados. Esto se indica por la línea de puntos en la Fig. 8. Además, la adaptación comprende una unidad 312 para la determinación de un espectro de sub-bandas perceptuales para la señal de audio de entrada, o la representación de la misma, en base a los coeficientes de transformación determinados, o los coeficientes de transformación normalizados. Una unidad de enmascaramiento 314 se proporciona para la determinación de los umbrales de enmascaramiento MT para cada dicha sub-banda en base a dicho espectro determinado. Finalmente, la adaptación comprende una unidad 316 para el cálculo de los factores de escala para cada dicha sub-banda en base a dichos umbrales de enmascaramiento determinados. Esta unidad 316 se puede dotar con o conectar a medios de adaptación 318 para la adaptación de dichos factores de escala calculados para cada sub-banda citada para impedir pérdida de energía para las sub-bandas relevantes perceptualmente. Para una realización específica, la unidad de adaptación 318 comprende una unidad 319 para comprimir-expandir adaptativamente los factores de escala determinados, y una unidad 320 para alisar adaptativamente los factores de escala determinados.

La adaptación anteriormente descrita puede ser incluida en o ser conectable a un codificador o adaptación de codificador en un sistema de telecomunicación.

Las ventajas de la presente invención comprenden:

el cálculo de baja complejidad con audio de la banda completa de alta calidad

la resolución de frecuencia flexible adaptada al cuantificador

la compresión-expansión/ expansión adaptativa de los factores de escala

Se entenderá por aquellos expertos en la técnica que se pueden hacer varias modificaciones y cambios a la presente invención sin salirse del alcance de la misma, el cual se define por las reivindicaciones adjuntas.

Referencias

[1] J. D. Johnston, "Estimación de la Entropía Perceptual Usando Criterios de Enmascaramiento de Ruido", Actas de ICASSP, páginas 2524-2527, mayo 1988.

[2] J. D. Johnston, "Codificación de transformación de señales de audio usando criterios de ruido perceptual"

Comun. Áreas de Selecc. J. del IEEE, vol. 6, páginas 314-323, 1988

[3] ISO/IEC JTC/SC29/WG 11, CD 11172-3, "Codificación de Imágenes en Movimiento y Audio Asociado para Medios de Almacenamiento Digitales en hasta 1,5 Mbit/s, Parte 3 AUDIO", 1993.

[4] ISO/IEC 13818-7, "Codificación de Audio Avanzada, AAC, MPEG-2", 1997.

REIVINDICACIONES

1. Un método de codificación por transformación perceptual de señales de audio en un sistema de telecomunicación, dicho método que comprende los pasos de:
- 5 determinar los coeficientes de transformación (210) representativos de una transformación de tiempo a frecuencia de una señal de audio de entrada segmentada en el tiempo;
- determinar un espectro de sub-bandas perceptuales (212) para dicha señal de audio de entrada en base a dichos coeficientes de transformación determinados;
- determinar los umbrales de enmascaramiento (214) para cada dicha sub-banda en base a dicho espectro determinado;
- 10 calcular los factores de escala (216) para cada dicha sub-banda en base a dichos umbrales de enmascaramiento determinados;
- dicho método que se **caracteriza por** el paso de:
- adaptar dichos factores de escala calculados (218) para cada dicha sub-banda para impedir pérdida de energía para las sub-bandas relevantes perceptualmente.
- 15 2. El método de acuerdo con la reivindicación 1, **caracterizado por** dicho paso de adaptación (218) que comprende la realización de la compresión-expansión adaptativa (219), y, el alisamiento (220) de dichos factores de escala para cada sub-banda citada.
3. El método de acuerdo con la reivindicación 2, **caracterizado por** realizar dicho paso de adaptación en base a una gama del cuantificador predeterminada.
- 20 4. El método de acuerdo con la reivindicación 1, **caracterizado por** dicho paso de determinación del umbral de enmascaramiento (214) que además comprende la normalización de dichos umbrales de enmascaramiento determinados, y calcular posteriormente dichos factores de escala en base a dichos umbrales de enmascaramiento normalizados.
5. El método de acuerdo con la reivindicación 2, **caracterizado por** el paso inicial adicional de normalización de los coeficientes de transformación determinados (211), y la realización de todos los pasos en base a dichos coeficientes de transformación normalizados.
- 25 6. El método de acuerdo con la reivindicación 1, **caracterizado porque** dicho espectro se basa al menos parcialmente en el espectro Bark.
7. El método de acuerdo con la reivindicación 4, **caracterizado por** dicho paso de normalización que comprende el cálculo de la raíz cuadrática media de dicha señal de audio de entrada en un dominio espectral transformado.
- 30 8. Una adaptación para la codificación por transformación perceptual de las señales de audio en un sistema de telecomunicación, que comprende:
- los medios de determinación de transformación (310) para determinar los coeficientes de transformación representativos de una transformación de tiempo a frecuencia de una señal de audio de entrada segmentada en el tiempo;
- 35 los medios de espectro (312) para determinar un espectro de sub-bandas perceptual para dicha señal de audio de entrada en base a dichos coeficientes de transformación determinados;
- los medios de enmascaramiento (314) para determinar los umbrales de enmascaramiento para cada sub-banda citada en base a dicho espectro determinado;
- 40 los medios de factor de escala (316) para calcular los factores de escala para cada sub-banda citada en base a dichos umbrales de enmascaramiento determinados;
- caracterizada porque** dicha adaptación además comprende:
- los medios de adaptación (318) para adaptar dichos factores de escala calculados para cada sub-banda citada para impedir la pérdida de energía para las sub-bandas relevantes perceptualmente.
- 45 9. La adaptación de acuerdo con la reivindicación 8, **caracterizada porque** dichos medios de adaptación (318) comprenden medios adicionales para realizar la compresión-expansión adaptativa (319) y el alisamiento (320) de dichos factores de escala para cada sub-banda citada.
10. La adaptación de acuerdo con la reivindicación 8, **caracterizada por** medios adicionales para la normalización (311) de dichos coeficientes de transformación determinados.

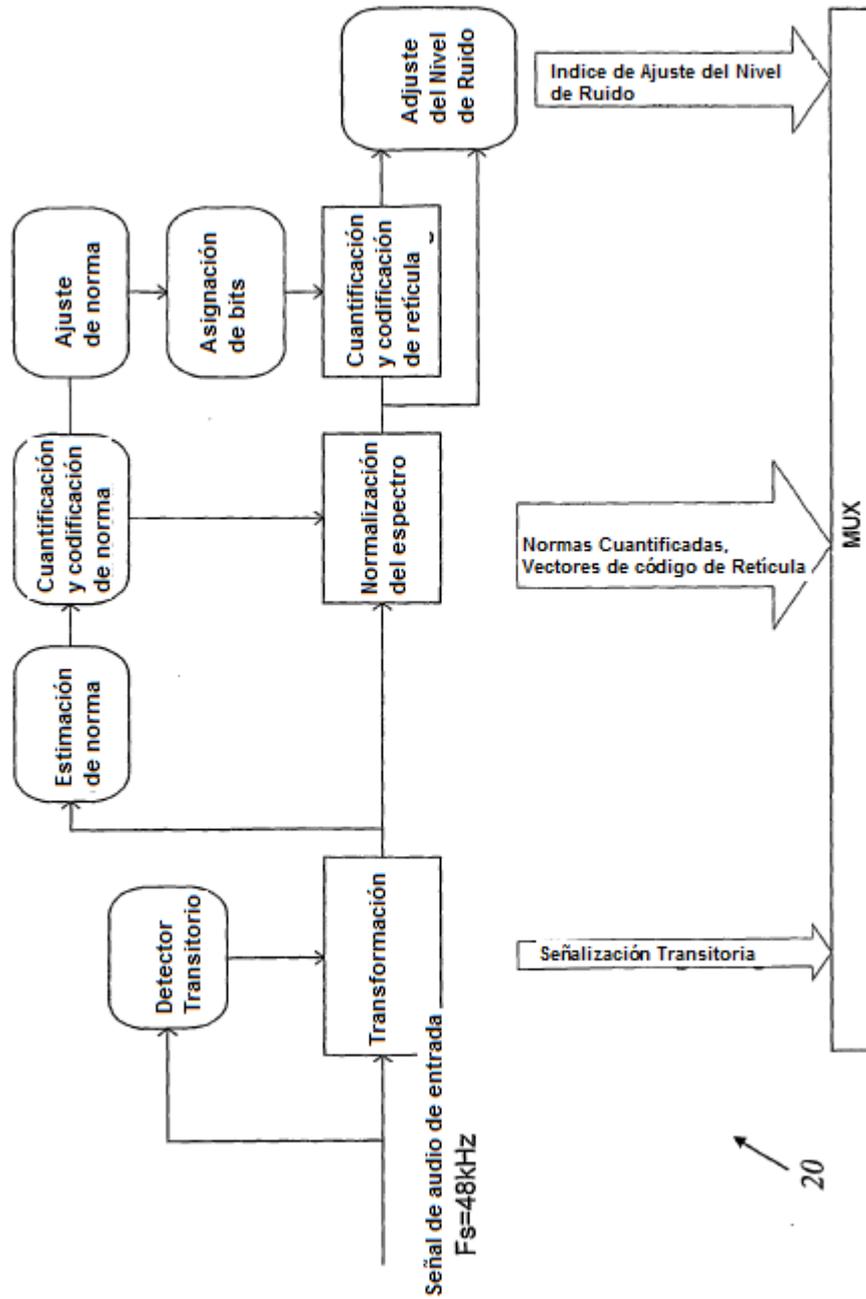


Fig. 1

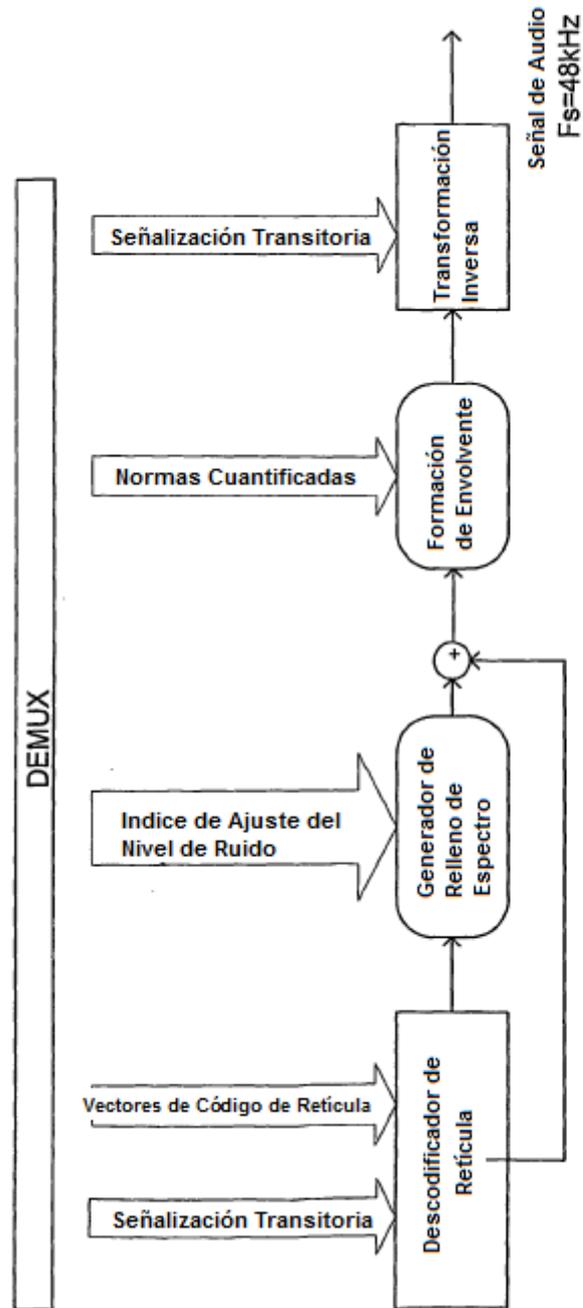


Fig. 2

40

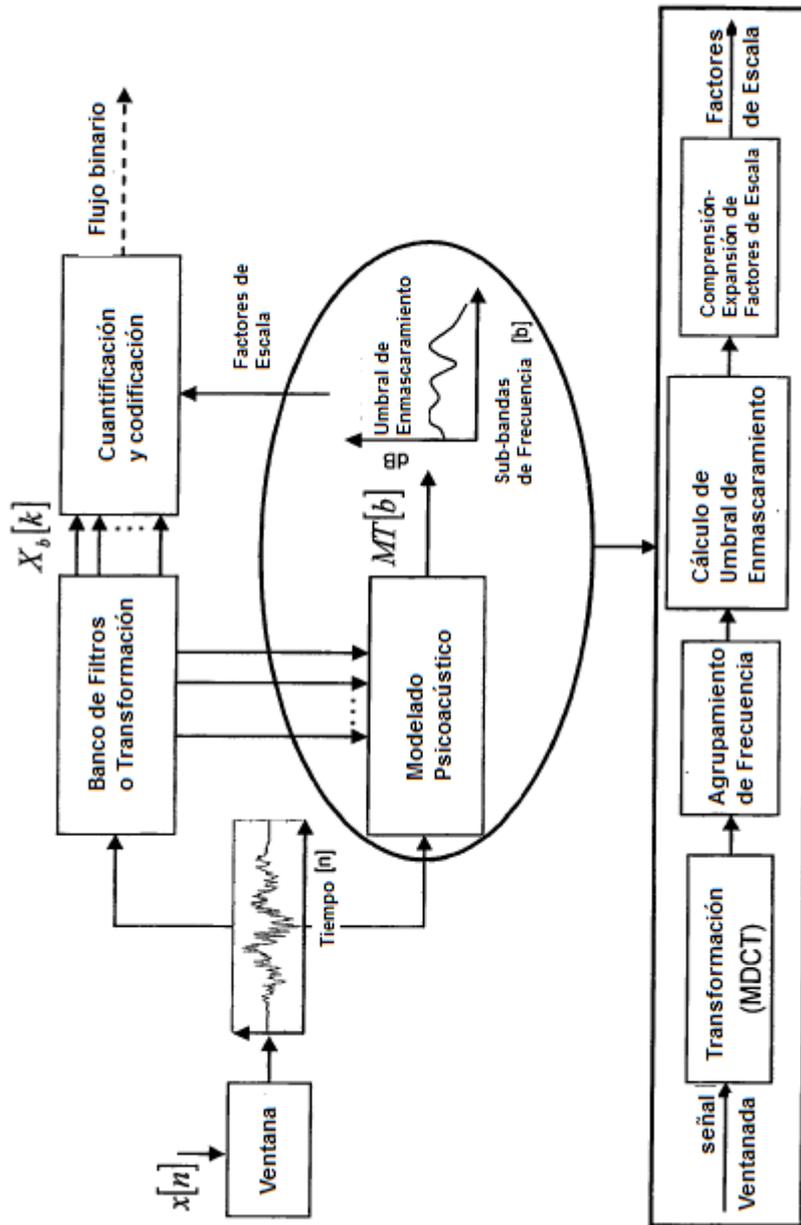


FIG. 3

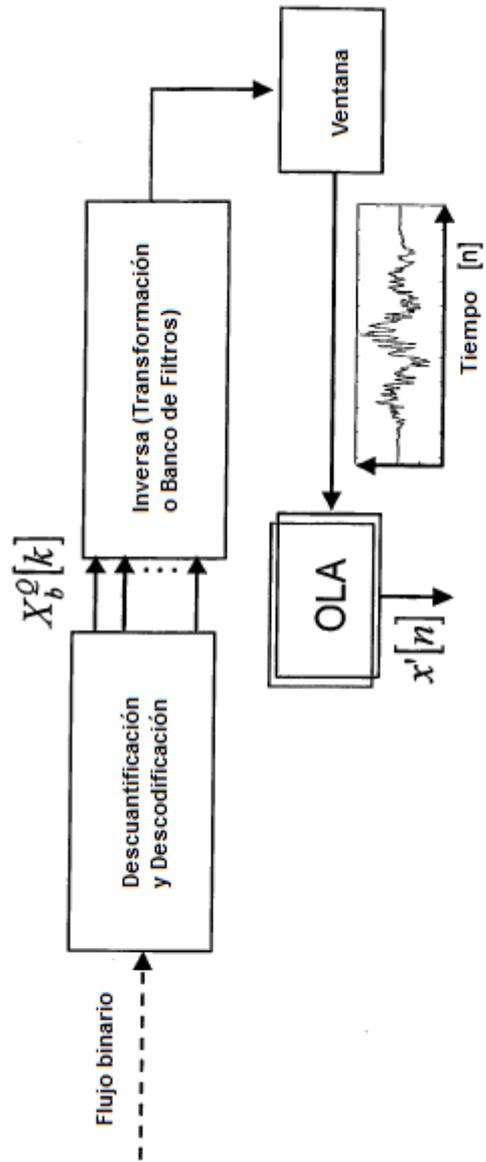


FIG. 4

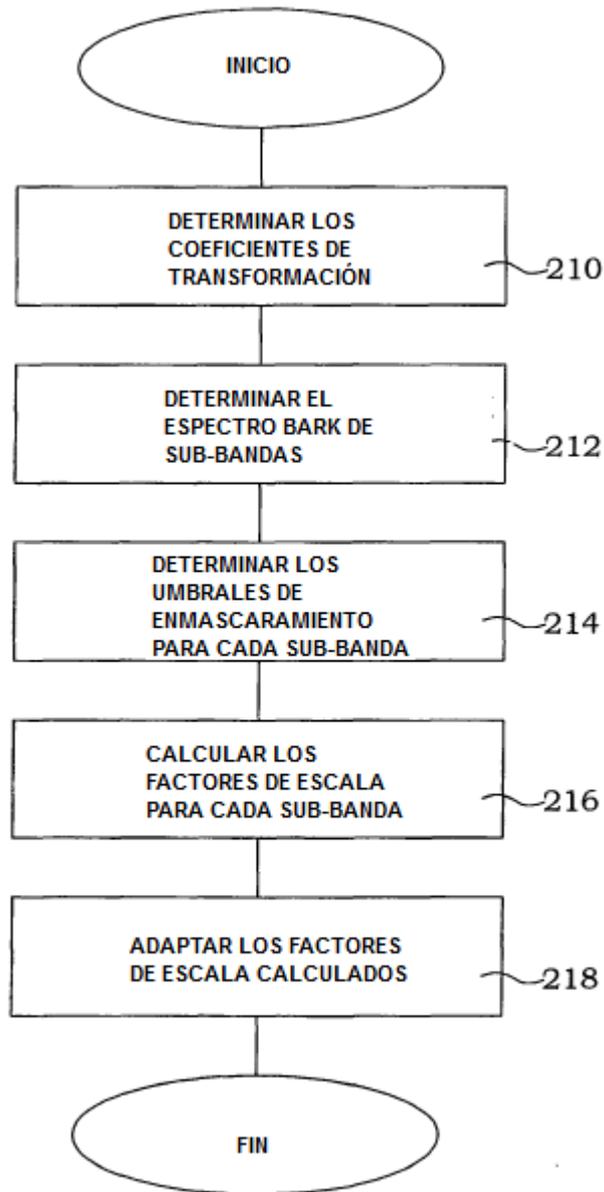


Fig. 5

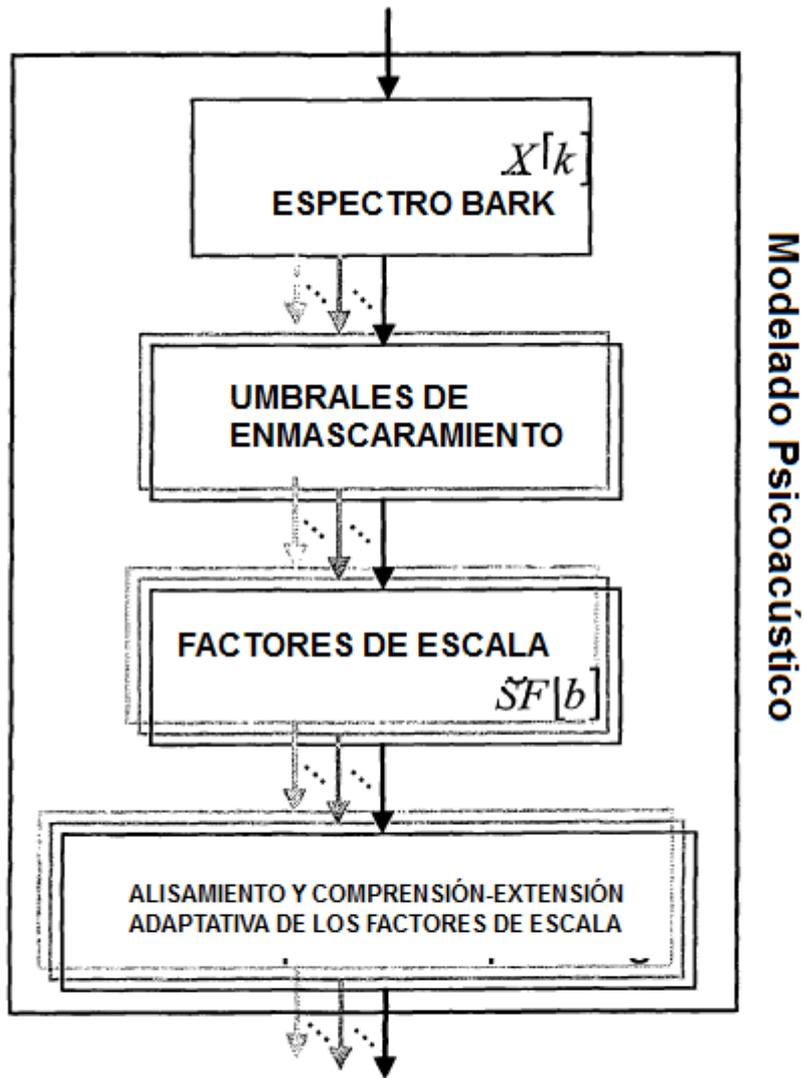


Fig. 6

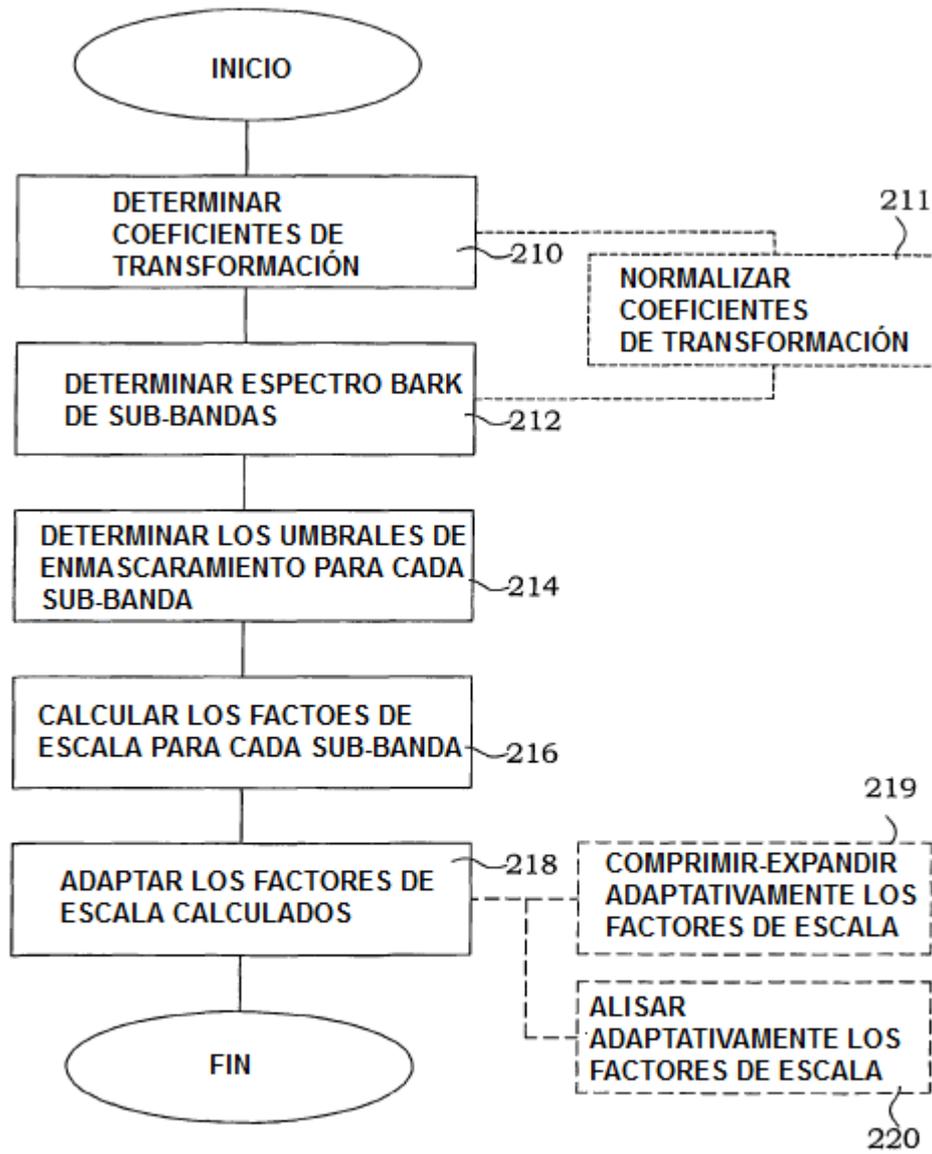


Fig. 7

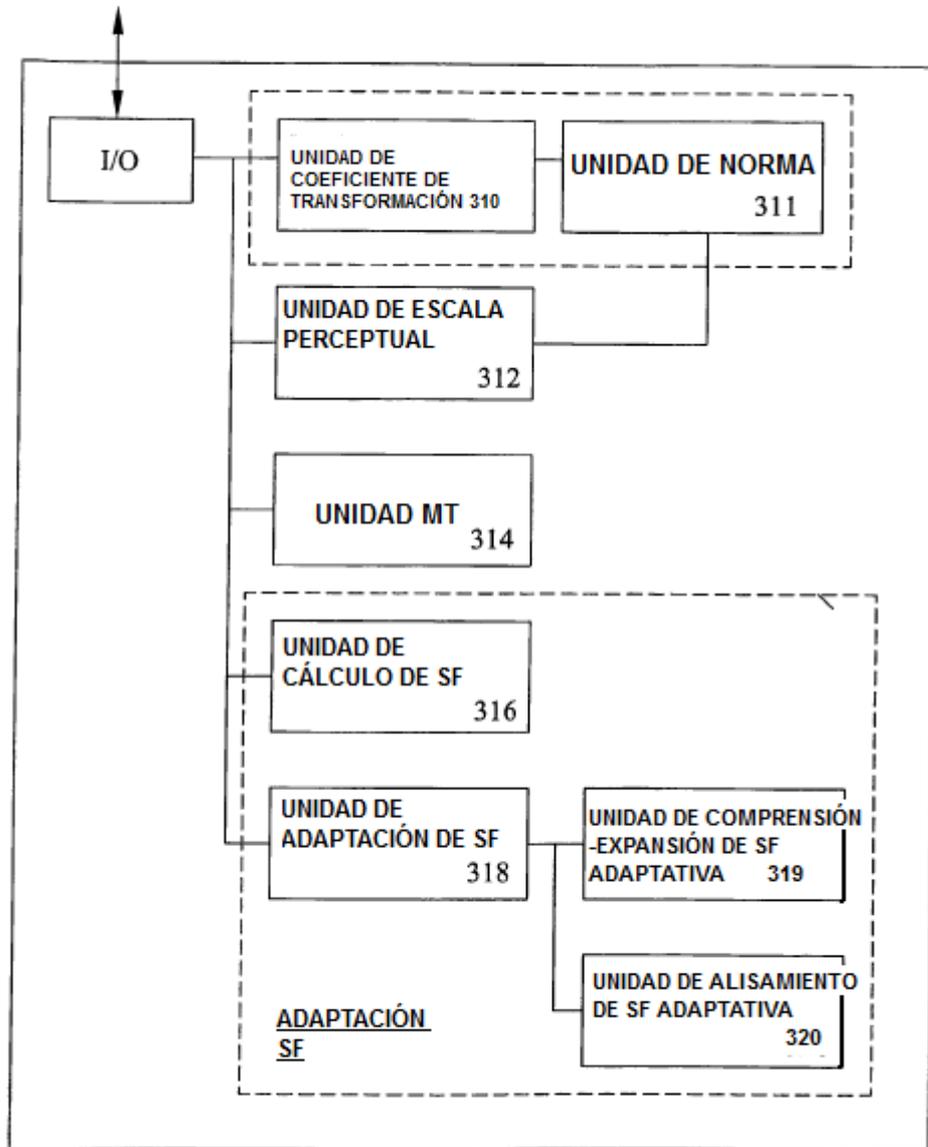


Fig. 8