

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 658 942**

51 Int. Cl.:

G10L 19/022 (2013.01)

G10L 19/02 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **25.08.2008 PCT/SE2008/050959**

87 Fecha y número de publicación internacional: **05.03.2009 WO09029032**

96 Fecha de presentación y número de la solicitud europea: **25.08.2008 E 08828335 (3)**

97 Fecha y número de publicación de la concesión europea: **15.11.2017 EP 2186088**

54 Título: **Análisis espectral/síntesis de baja complejidad utilizando resolución temporal seleccionable**

30 Prioridad:

27.08.2007 US 968125 P

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

13.03.2018

73 Titular/es:

**TELEFONAKTIEBOLAGET LM ERICSSON (PUBL)
(100.0%)
NARVIKSGATAN 5
S-164 33 KISTA, SE**

72 Inventor/es:

TALEB, ANISSE

74 Agente/Representante:

ELZABURU, S.L.P

ES 2 658 942 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

Análisis espectral/síntesis de baja complejidad utilizando resolución temporal seleccionable

Sector técnico

5 La presente invención se refiere, en general, a procesamiento de señal, tal como compresión de señal y codificación de audio, y más concretamente a la codificación de audio y descodificación de audio, y a los dispositivos correspondientes.

Antecedentes

10 Un codificador es un dispositivo, circuitos o un programa informático que puede analizar una señal, tal como una señal de audio, y entregar una señal en forma codificada. La señal resultante se utiliza a menudo con fines de transmisión, almacenamiento y/o cifrado. Por otra parte, un descodificador es un dispositivo, circuitos o un programa informático que puede invertir la operación del codificador, porque recibe la señal codificada y entrega una señal descodificada.

15 En la mayor parte de los codificadores de la técnica actual, tales como los codificadores de audio, cada trama de la señal de entrada se analiza en el dominio de frecuencia. El resultado de este análisis es cuantificado y codificado, y a continuación transmitido o almacenado, dependiendo de la aplicación. En el lado de recepción (o cuando se utiliza la señal codificada almacenada), un correspondiente procedimiento de descodificación seguido por un procedimiento de síntesis hacen posible recuperar la señal en el dominio de tiempo.

A menudo se utilizan códecs para la compresión/descompresión de información, tal como datos de audio y de video, para su transmisión eficiente sobre canales de comunicación limitados en ancho de banda.

20 En particular, existe en el mercado una gran necesidad de transmitir y almacenar señales de audio a bajas velocidades binarias manteniendo al mismo tiempo una alta calidad de audio. Por ejemplo, en casos en que los recursos de transmisión o el almacenamiento están limitados, el funcionamiento a bajas velocidades binarias es un factor de coste esencial. Este es el caso habitualmente, por ejemplo, en aplicaciones de descarga continua y de mensajería en los sistemas de comunicación móvil.

25 En la figura 1 se muestra esquemáticamente un ejemplo general de un sistema de transmisión de audio que utiliza codificación y descodificación de audio. El sistema global comprende básicamente un codificador de audio 10 y un módulo de transmisión (TX) 20 en el lado de transmisión, y un módulo de recepción (RX) 30 y un descodificador de audio 40 en el lado de recepción.

30 Es generalmente admitido que hay que tener un especial cuidado para tratar con señales no estacionarias, en particular para la aplicación de codificación de audio, y en general, para la compresión de señales. En la codificación de audio, puede surgir un artefacto conocido como distorsión de pre-eco en los denominados codificadores de transformada.

35 Los codificadores de transformada, o más generalmente los códecs (codificadores-descodificadores) de transformada están basados normalmente en una transformada en los dominios de tiempo a frecuencia, tal como una DCT (Discrete Cosine Transform, transformada de coseno discreta), una transformada de coseno discreta modificada (MDCT) u otra transformada traslapada. Una característica común de los códecs de transformada es que funcionan sobre bloques solapados de muestras: tramas solapadas. Los coeficientes de codificación resultantes de un análisis de transformada o de un análisis de sub-banda equivalente de cada trama son normalmente cuantificados y almacenados o transmitidos al lado de recepción como un flujo de bits. El descodificador, al recibir el

40 flujo de bits, lleva a cabo la descuantificación y la transformación inversa con el fin de reconstruir las tramas de la señal.

Generalmente se producen pre-ecos cuando una señal con una entrada abrupta comienza cerca del final de un bloque de transformada inmediatamente posterior a una zona de baja energía.

45 Esta situación se produce, por ejemplo, cuando se codifica el sonido de instrumentos de percusión, por ejemplo, castañuelas, glockenspiel. En un algoritmo basado en bloques, cuando se cuantifican los coeficientes de transformada, la transformada inversa en el lado del descodificador dispersará la distorsión del ruido de cuantificación homogéneamente en el tiempo. Esto tiene como resultado una distorsión no enmascarada en la zona de baja energía que sigue en el tiempo a la entrada de ataque de la señal, tal como se muestra en las figuras 2A y B, donde la figura 2A muestra el sonido de percusión original y la figura 2B muestra la señal transformada-codificada que presenta la dispersión temporal del ruido de codificación que conduce a una distorsión de pre-eco.

50

El pre-enmascaramiento temporal es una propiedad psicoacústica de la audición humana que tiene el potencial de enmascarar esta distorsión; si bien esto es posible solamente cuando el tamaño del bloque de transformada es lo suficientemente pequeño como para que se produzca dicho pre-enmascaramiento.

Atenuación de artefactos de pre-eco (técnica anterior)

Para evitar este artefacto indeseable, han sido propuestas y aplicadas satisfactoriamente varias tecnologías. Algunas de estas tecnologías se han estandarizado y están generalizadas en las aplicaciones comerciales.

Técnicas de reserva de bits

- 5 La idea que subyace a la técnica de reserva de bits es preservar algunos bits de las tramas que son "fáciles" de codificar en el dominio de frecuencia. Los bits preservados se utilizan a continuación para alojar tramas muy exigentes, tales como tramas transitorias. Esto tiene como resultado una velocidad binaria instantánea variable, pudiendo realizarse algún ajuste de manera que la velocidad binaria promedio sea constante. Sin embargo, el inconveniente principal es que de hecho se requieren reservas muy grandes para tratar con determinados
- 10 transitorios y esto conduce a un retardo muy grande que hace que esta tecnología tenga poco interés para aplicación conversacional. Además, esta metodología solo atenúa ligeramente el artefacto de pre-eco.

Modificación de la ganancia y modelado del ruido temporal

- 15 El enfoque de modificación de la ganancia aplica un suavizado de los picos transitorios en el dominio de tiempo antes del análisis espectral y de la codificación. La envolvente de modificación de ganancia se envía como información complementaria y se realiza su aplicación inversa en la señal transformada inversa, modelando de ese modo el ruido de codificación temporal. Un inconveniente principal de la técnica de modificación de ganancia está en su modificación de la ventana de análisis del banco de filtros (por ejemplo, MDCT), introduciendo por lo tanto un ensanchamiento de la respuesta de frecuencia del banco de filtros. Esto puede conducir a problemas a bajas frecuencias, especialmente si el ancho de banda excede el de la banda crítica.

- 20 El modelado de ruido temporal (TNS, Temporal Noise Shaping) está inspirado por la técnica de modificación de ganancia. La modificación de ganancia se aplica en el dominio de frecuencia y trabaja sobre los coeficientes espectrales. TNS se aplica solamente durante flancos de ataque de entrada susceptibles a pre-ecos. La idea es aplicar predicción lineal (*LP*, linear prediction) a través de la frecuencia en lugar del tiempo. Esto está motivado por el hecho de que durante los transitorios y en las señales impulsivas en general, la ganancia de codificación en el
- 25 dominio de frecuencia se maximiza mediante la utilización de técnicas de *LP*. TNS se ha estandarizado en AAC y está demostrado que proporciona una buena atenuación de los artefactos de pre-eco. Sin embargo, la utilización de TNS implica análisis *LP* y filtrado, lo que aumenta significativamente la complejidad del codificador y del descodificador. Adicionalmente, los coeficientes *LP* tienen que ser cuantificados y enviados como información complementaria, lo que implica mayor complejidad y sobrecarga en la velocidad binaria.

Conmutación entre ventanas

- 30 La figura 3 muestra la conmutación entre ventanas (MPEG-1, capa III "mp3"), donde se requieren ventanas de transición "inicio" y "final" entre ventanas largas y cortas para conservar las propiedades de PR (reconstrucción perfecta). Esta técnica fue introducida en primer lugar por Edler [1] y es popular para la supresión de pre-eco, en particular en el caso de algoritmos de codificación de transformada basados en MDCT. La conmutación entre
- 35 ventanas se basa en la idea de cambiar la resolución temporal de la transformada tras la detección de un transitorio. Habitualmente, esto implica modificar la longitud del bloque de análisis, de una duración larga durante señales estacionarias a una duración corta cuando se detectan transitorios. La idea se basa en dos consideraciones:

- Una ventana corta aplicada a la trama corta que contiene el transitorio minimizará la dispersión temporal del ruido de codificación y permitirá que el pre-enmascaramiento temporal tenga efecto y haga la distorsión inaudible.
 - Asignar mayores velocidades binarias a las zonas temporales cortas que contienen el transitorio.
- 40

- Aunque la conmutación entre ventanas ha sido muy satisfactoria, presenta inconvenientes significativos. Por ejemplo, el modelo perceptual y los módulos de codificación sin pérdidas del códec tienen que soportar diferentes resoluciones temporales, lo que se traduce normalmente en una mayor complejidad. Además, cuando se utilizan transformadas traslapadas, tales como MDCT, y para satisfacer los requisitos de reconstrucción perfecta, la conmutación entre ventanas tiene que introducir ventanas de transición entre bloques cortos y largos, tal como se muestra en la figura 3. La necesidad de ventanas de transición genera inconvenientes adicionales, en concreto un mayor retardo debido al hecho de que conmutar las ventanas no se puede hacer instantáneamente, y asimismo a las malas propiedades de localización de frecuencias de las ventanas de transición, que conducen a una dramática reducción en la ganancia de codificación.
- 45

- 50 El documento WO 2006/030289 describe la codificación de audio multicanal. Un sistema de codificación de audio digital de baja velocidad binaria incluye un codificador que asigna libros de códigos a grupos de índices de cuantificación en base a sus propiedades locales teniendo como resultado rangos de aplicación de libros de códigos que son independientes de los límites de cuantificación de los bloques. Este documento describe un banco de filtros de resolución, o un banco de filtros de resolución tri-modo, que es conmutable selectivamente entre modos alto y
- 55 bajo de resolución de frecuencias, o modos alto, bajo e intermedio, tal como cuando se detecta un transitorio en una trama. El resultado es una señal de audio multicanal que tiene una velocidad binaria significativamente inferior para

una transmisión o un almacenamiento eficientes. El descodificador es esencialmente inverso a la estructura y los procedimientos del codificador, y tiene como resultado una señal de audio reproducida que no se puede distinguir de manera audible de la señal original.

5 El documento "Modified Discrete Cosine Transform - its Implications for Audio Coding and Error Concealment", de Wang Y et al., fue publicado en Journal of the Audio Engineering Society, volumen 51, número 1/2, enero/febrero de 2003 (XP001178776) y describe un estudio de la transformada de coseno discreta modificada y sus implicaciones para codificación de audio y ocultación de errores desde la perspectiva del análisis de frecuencias de Fourier. Se describe una relación entre la transformada de coseno discreta modificada y la transformada de Fourier discreta mediante la transformada de Fourier discreta desplazada, que proporciona una posible implementación rápida de la transformada de coseno discreta modificada utilizando una rutina de transformada rápida de Fourier. Se analiza y se muestra con ejemplos el concepto de cancelación del repliegue en el dominio de tiempo, las propiedades simétricas y las no ortogonales de la transformada de coseno discreta modificada.

Compendio

La presente invención supera estos y otros inconvenientes de las disposiciones de la técnica anterior.

15 Por lo tanto, existe una necesidad general de técnicas y dispositivos mejorados de procesamiento de señal, y más particularmente una necesidad especial de una nueva estrategia de códec de audio para tratar la distorsión de pre-eco.

Un objetivo general de la presente invención es dar a conocer un procedimiento y un dispositivo mejorados para procesamiento de señales que actúa sobre tramas solapadas de una señal de entrada en el dominio de tiempo.

20 En particular, es deseable proporcionar un codificador de audio mejorado.

Otro objetivo de la invención es dar a conocer un procedimiento y un dispositivo mejorados para procesamiento de señal que actúa en base a coeficientes espectrales representativos de una señal en el dominio de tiempo.

Es particularmente deseable proporcionar un descodificador de audio mejorado.

25 Estos y otros objetivos se consiguen mediante la invención tal como se define mediante las reivindicaciones de patente adjuntas.

Un primer aspecto de la invención se refiere a un procedimiento y un dispositivo para procesamiento de señal que actúa sobre tramas solapadas de una señal de entrada.

30 La invención se basa en el concepto de utilizar una trama replegada en el dominio de tiempo, como base para segmentación en tiempo y análisis espectral, llevando a cabo segmentación en el tiempo en base a la trama replegada en el dominio de tiempo y llevando a cabo el análisis espectral en base a los segmentos de tiempo resultantes.

Por lo tanto, la resolución temporal de la transformada de tiempo a frecuencia "segmentada" global se puede modificar simplemente adaptando la segmentación en tiempo para obtener un número adecuado de segmentos de tiempo en base a los cuales se aplica el análisis espectral.

35 Más específicamente, una idea básica es llevar a cabo un repliegue en el dominio de tiempo (TDA, time-domain aliasing) en base a una trama solapada, para generar una correspondiente trama replegada en el dominio de tiempo, y llevar a cabo segmentación en tiempo en base a la trama replegada en el dominio de tiempo para generar por lo menos dos segmentos, denominados asimismo subtramas. En base a estos segmentos, se lleva a cabo a continuación el análisis espectral para obtener, para cada segmento, coeficientes representativos del contenido en frecuencia del segmento.

40 El conjunto global de coeficientes, denominados asimismo coeficientes espectrales, para todos los segmentos proporciona un mosaico de tiempo-frecuencia seleccionable de la trama de señal original.

45 La descomposición instantánea en segmentos puede ser utilizada, por ejemplo, para atenuar el efecto de pre-eco, por ejemplo en el caso de transitorios, o en general para proporcionar una representación eficiente de la señal que permita una codificación eficiente en velocidad binaria, de la trama en cuestión.

Un primer aspecto de la invención está relacionado en particular con un codificador de audio configurado para funcionar de acuerdo con los anteriores principios básicos.

50 Un segundo aspecto de la invención se refiere a un procesamiento de señal de un procedimiento y un dispositivo, que funciona en base a coeficientes espectrales representativos de una señal en el dominio de tiempo. Este aspecto de la invención se refiere básicamente a operaciones inversas naturales del procesamiento de señal del primer aspecto de la invención. En resumen, se lleva a cabo análisis espectral segmentado inverso en base a diferentes subconjuntos de coeficientes espectrales para generar, para cada subconjunto de coeficientes espectrales, una subtrama transformada inversa denominada asimismo un segmento. La segmentación en tiempo inversa se lleva a

cabo en base a subtramas transformadas inversas solapadas para combinar estas subtramas en una trama replegada en el dominio de tiempo. Se lleva a cabo repliegue inverso en el dominio de tiempo en base a la trama replegada en el dominio de tiempo para permitir la reconstrucción de la señal en el dominio de tiempo.

5 El segundo aspecto de la invención está relacionado en particular con un descodificador de audio configurado para funcionar de acuerdo con los principios básicos anteriores.

Se apreciarán otras ventajas ofrecidas por la invención al leer la siguiente descripción de realizaciones de la invención.

Breve descripción de los dibujos

10 La invención, junto con otros aspectos y ventajas de la misma, se comprenderá mejor al hacer referencia a la siguiente descripción detallada, tomada junto con los dibujos adjuntos, en los cuales:

La figura 1 es un diagrama de bloques esquemático que muestra un ejemplo general de un sistema de transmisión de audio que utiliza codificación y descodificación de audio.

La figura 2A muestra un sonido de percusión original, y la figura 2B muestra una señal codificada transformada que muestra la dispersión temporal del ruido de codificación, que conduce a una distorsión de pre-eco.

15 La figura 3 muestra una técnica convencional de conmutación entre ventanas para codificación basada en transformada.

La figura 4A muestra esquemáticamente la transformada MDCT (transformada de coseno discreta modificada) directa general.

20 La figura 4B muestra esquemáticamente la transformada MDCT (transformada de coseno discreta modificada) inversa general.

La figura 5 es un diagrama esquemático que muestra la descomposición de la transformada MDCT (transformada de coseno discreta modificada) en dos fases en cascada.

La figura 6 es un diagrama de flujo esquemático que muestra un ejemplo de un procedimiento para procesamiento de señal de acuerdo con una realización preferida a modo de ejemplo de la invención.

25 La figura 7 es un diagrama de bloques esquemático de un dispositivo de procesamiento de señal general.

La figura 8 es un diagrama de bloques esquemático de un dispositivo de acuerdo con otra realización preferida a modo de ejemplo de la invención.

La figura 9 es un diagrama de bloques esquemático de un dispositivo de acuerdo con otra realización preferida a modo de ejemplo de la invención.

30 La figura 10 es un diagrama esquemático de un ejemplo de reordenamiento del repliegue en el dominio de tiempo de acuerdo con una realización a modo de ejemplo de la invención.

La figura 11 es un diagrama esquemático que muestra un ejemplo de segmentación en dos segmentos de tiempo, incluyendo relleno de ceros, de acuerdo con una realización a modo de ejemplo de la invención.

35 La figura 12 muestra diagramas de las dos funciones base para la segmentación de la figura 11, que se refieren a una frecuencia normalizada de 0,25 junto con correspondientes diagramas de respuesta de frecuencia.

La figura 13 muestra diagramas de las funciones base MDCT originales en relación con la frecuencia normalizada de 0,25 junto con los correspondientes diagramas de respuesta de frecuencia.

La figura 14 es un diagrama esquemático que muestra un ejemplo de segmentación en cuatro segmentos de tiempo, incluyendo relleno de ceros, de acuerdo una realización a modo de ejemplo de la invención.

40 La figura 15 es un diagrama esquemático que muestra un ejemplo de segmentación en ocho segmentos de tiempo, incluyendo relleno de ceros, de acuerdo una realización a modo de ejemplo de la invención.

La figura 16 muestra una realización de una transformada global resultante para el caso de cuatro segmentos, de acuerdo con una realización a modo de ejemplo de la invención.

45 La figura 17 muestra un modo, a título de ejemplo, de obtención de una segmentación no uniforme por medio de un enfoque jerárquico.

La figura 18 muestra un ejemplo de conmutación instantánea a una resolución temporal más fina tras la detección de un transitorio.

La figura 19 es un diagrama de bloques que muestra un ejemplo básico de un dispositivo de procesamiento de señal para funcionamiento basado en coeficientes espectrales representativos de una señal en el dominio de tiempo.

La figura 20 es un diagrama de bloques de un codificador a modo de ejemplo adecuado para la extensión de banda completa.

5 La figura 21 es un diagrama de bloques de un descodificador a modo de ejemplo adecuado para la extensión de banda completa.

La figura 22 es un diagrama de bloques esquemático de un ejemplo particular de un transformador inverso y la implementación asociada para segmentación en tiempo inversa y reordenamiento opcional.

Descripción detallada

10 En todos los dibujos, se utilizarán los mismos caracteres de referencia para elementos correspondientes o similares.

Para una mejor comprensión de la invención, puede ser útil comenzar con una breve introducción a la codificación por transformación, y especialmente a la codificación por transformación basada en las denominadas transformadas traslapadas.

15 Tal como se ha mencionado anteriormente, los códecs de transformada se basan normalmente en una transformada de los dominios de tiempo a frecuencia, tal como una DCT (transformada de coseno discreta), una transformada traslapada tal como una transformada de coseno discreta modificada (MDCT) o una transformada traslapada modulada (MLT, Modulated Lapped Transform).

20 Por ejemplo, la transformada de coseno discreta modificada (MDCT) es una transformada relacionada con la de Fourier basada en la transformada de coseno discreta de tipo IV (DCT-IV), con la propiedad adicional de ser traslapada: está diseñada para ser ejecutada sobre bloques consecutivos de un conjunto mayor de datos, donde los bloques consecutivos están solapados, denominados tramas solapadas, de tal modo que la última mitad de un bloque coincide con la primera mitad del bloque siguiente, tal como se muestra esquemáticamente en la figura 4A. Este solapamiento, además de las calidades de compactación de energía de la DCT, hacen la MDCT especialmente atractiva para aplicaciones de compresión de señales, dado que contribuye a evitar el surgimiento de artefactos a partir de los límites de bloque. De este modo, se utiliza una MDCT en MP3, AC-3, Ogg Vorbis y AAC para compresión de audio, por ejemplo.

25 Como transformada traslapada, la MDCT es algo diferente cuando se compara con otras transformadas relacionadas con la de Fourier. De hecho, la MDCT tiene como mucho la mitad de salidas que de entradas. Formalmente, la MDCT es un mapeo lineal de \mathfrak{R}^{2N} en \mathfrak{R}^N (donde \mathfrak{R} indica el conjunto de los números reales).

30 Matemáticamente, los números reales x_0, x_1, \dots, x_{2N} se transforman en los números reales X_0, X_1, \dots, X_N , según la fórmula:

$$X_k = \sum_{n=0}^{2N-1} x_n \cos \left[\frac{\pi}{N} \left(n + \frac{1}{2} + \frac{N}{2} \right) \left(k + \frac{1}{2} \right) \right]$$

Esta fórmula anterior, dependiendo del convenio, puede contener un coeficiente de normalización adicional.

35 La MDCT inversa se conoce como IMDCT. Debido a que las dimensiones de la salida y de la entrada son diferentes, a primera vista puede parecer que la MDCT no sería invertible. Sin embargo, se consigue una inversibilidad perfecta sumando las IMDCT solapadas de los bloques solapados subsiguientes, es decir, las tramas solapadas, provocando que se cancelen los errores y se puedan recuperar los datos originales; esta técnica se conoce como cancelación del repliegue en el dominio de tiempo (TDAC, time-domain aliasing cancellation), y se muestra esquemáticamente en la figura 4B.

40 En resumen, para la transformada directa, se mapean 2N muestras (de una de las tramas solapadas) a N coeficientes espectrales, y para la transformada inversa, se mapean N na. coeficientes espectrales a 2N muestras en el dominio de tiempo (de una de las tramas solapadas reconstruidas) que son sumadas-solapadas para formar una señal de salida en el dominio de tiempo.

La IMDCT transforma N números reales Y_0, Y_1, \dots, Y_N en los números reales y_0, y_1, \dots, y_{2N} , de acuerdo con la fórmula:

45
$$y_n = \frac{1}{N} \sum_{k=0}^{N-1} Y_k \cos \left[\frac{\pi}{N} \left(n + \frac{1}{2} + \frac{N}{2} \right) \left(k + \frac{1}{2} \right) \right]$$

En una típica aplicación de compresión de señal, las propiedades de la transformada se mejoran adicionalmente utilizando una función ventana w_n que es multiplicada por la señal de entrada a la transformada directa x_n y la señal

de salida de la transformada inversa y_n . En principio, x_n y y_n podrían utilizar ventanas diferentes, pero por simplicidad se considera solamente el caso de ventanas idénticas.

Existen varias ventanas ortogonales y bi-ortogonales de propósito general. En el caso ortogonal, las condiciones de reconstrucción perfecta (PR) generalizadas se pueden reducir a restricciones Nyquist y de fase lineal sobre la ventana, es decir:

$$\begin{aligned} w(2N-1-n) &= w(n) \\ w^2(n) + w^2(n+N) &= 1, \\ n &= 0 \dots N-1 \end{aligned}$$

Cualquier ventana que satisfaga las condiciones de reconstrucción perfecta (PR) puede ser utilizada para generar el banco de filtros. Sin embargo, para obtener una alta ganancia de codificación, la respuesta de frecuencia resultante del banco de filtros debería ser lo más selectiva posible.

La referencia [2] designa mediante MLT (Modulated Lapped Transform, transformada traslapada modulada) el banco de filtros MDCT que hace uso de la ventana seno, que se define como:

$$w(n) = \sin \left[\left(n + \frac{1}{2} \right) \frac{\pi}{2N} \right]$$

Esta ventana particular, la denominada ventana seno, es la más popular en codificación de audio. Aparece, por ejemplo, en el banco de filtros híbridos MPEG-1 capa III (MP3), así como en MPEG-2/4 AAC.

Una de las propiedades atractivas que ha contribuido a la utilización generalizada de la MDCT para codificación de audio es la disponibilidad de algoritmos rápidos basados en FFT. Esto hace de la MDCT un banco de filtros viable para implementaciones en tiempo real.

Es bien sabido que la MDCT con una longitud de ventana de $2N$ se puede descomponer en dos fases en cascada. La primera fase consiste en la operación de repliegue en el dominio de tiempo (TDA) seguida por una segunda fase basada en la DCT de tipo IV, tal como se muestra en la figura 5.

La operación TDA está dada explícitamente por la siguiente operación matricial:

$$\tilde{x} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & -\mathbf{J}_N & -\mathbf{I}_N \\ \mathbf{I}_N & -\mathbf{J}_N & \mathbf{0} & \mathbf{0} \end{bmatrix} x_w,$$

donde x_w indica la trama de entrada en el dominio de tiempo en ventana:

$$x_w(n) = w(n) \cdot x(n),$$

las matrices \mathbf{I}_N y \mathbf{J}_N designan la identidad y las matrices de inversión temporal de orden N .

$$\mathbf{I}_N = \begin{bmatrix} 1 & & & 0 \\ & \ddots & & \\ & & \ddots & \\ 0 & & & 1 \end{bmatrix}, \quad \mathbf{J}_N = \begin{bmatrix} 0 & & & 1 \\ & \ddots & & \\ & & \ddots & \\ 1 & & & 0 \end{bmatrix}.$$

Un primer aspecto de la invención se refiere al procesamiento de señal que actúa sobre tramas solapadas de una señal de entrada. Un concepto clave es utilizar una trama replegada en el dominio de tiempo como base para la segmentación en tiempo y el análisis espectral, y llevar a cabo la segmentación en tiempo en base a la trama replegada en el dominio de tiempo y el análisis espectral en base a los segmentos de tiempo resultantes. Los segmentos de tiempo, o segmentos para abreviar, se denominan asimismo subtramas. Esto es lo lógico dado que un segmento en una trama se puede denominar una subtrama. Las expresiones "segmento" y "subtrama" se utilizarán, en general, de manera intercambiable a lo largo de la descripción.

La figura 6 es un diagrama de flujo esquemático que muestra un ejemplo de un procedimiento para procesamiento de señal de acuerdo con una realización preferida a modo de ejemplo de la invención. Tal como se indica en la etapa S1, el procedimiento puede involucrar una etapa opcional de pre-procesamiento, tal como se explicará y ejemplificará más adelante. En la etapa S2, se lleva a cabo una operación de repliegue en el dominio de tiempo (TDA) en base a una seleccionada de las tramas solapadas para generar una denominada trama TDA correspondiente que puede, opcionalmente, ser procesada en una o varias fases, tal como se indica en S3, antes de que se lleve a cabo la segmentación en tiempo. En cualquier caso, la segmentación en tiempo se lleva a cabo en base a la trama replegada en el dominio de tiempo (que puede haber sido procesada) para generar por lo menos dos segmentos de tiempo, tal como se indica en la etapa S4. En la etapa S5, se ejecuta el denominado análisis

espectral en base a los segmentos para obtener, para cada segmento, coeficientes representativos del contenido de frecuencia del segmento. Preferentemente, el análisis espectral se basa en aplicar una transformada sobre cada uno de los segmentos para producir, para cada segmento, un conjunto correspondiente de coeficientes espectrales. Es posible asimismo aplicar una etapa opcional de pos-procesamiento (no mostrada).

- 5 El análisis espectral está basado en una transformada de coseno discreta modificada (MDCT). Por lo tanto, la resolución temporal de la transformada de tiempo a frecuencia segmentada global se puede modificar simplemente adaptando la segmentación en tiempo para obtener un número adecuado de segmentos de tiempo en base a los cuales se aplica el análisis espectral. El procedimiento de segmentación se puede adaptar para producir segmentos no solapados, segmentos solapados, segmentos de longitud no uniforme y/o segmentos de longitud uniforme. De este modo, se puede obtener cualquier mosaico de tiempo-frecuencia arbitrario de la trama de señal original.

El procedimiento de procesamiento de señal global actúa normalmente trama a trama sobre tramas solapadas de una señal de entrada en el dominio de tiempo, y las etapas anteriores de repliegue en el tiempo, segmentación, análisis espectral y procesamiento opcional previo, intermedio y posterior se repiten preferentemente para cada una de una serie de tramas solapadas.

- 15 Preferentemente, el procesamiento de señal propuesto por la presente invención incluye análisis de señal, compresión de señal y/o codificación de audio. En un codificador de audio, por ejemplo, los coeficientes espectrales serán cuantificados normalmente en un flujo de bits para su almacenamiento y/o transmisión.

- 20 La figura 7 es un diagrama de bloques esquemático de un dispositivo de procesamiento de señal general. El dispositivo comprende básicamente una unidad de repliegue en el dominio de tiempo (TDA) 12, una unidad de segmentación en tiempo 14 y un analizador espectral 16. En el ejemplo básico de la figura 7, una trama considerada de una serie de tramas solapadas es replegada en el dominio de tiempo en la unidad TDA 12 para generar una trama replegada en el dominio de tiempo, y la unidad de segmentación en tiempo 14 actúa sobre la trama replegada en el dominio de tiempo para generar una serie de segmentos de tiempo, denominados asimismo subtramas. El analizador espectral 16 está configurado para un análisis espectral segmentado basado en estos segmentos, con el fin de generar, para cada segmento, un conjunto de coeficientes espectrales. Los coeficientes espectrales colectivos de todos los segmentos representan un mosaico de tiempo-frecuencia de la trama del dominio de tiempo procesada, con una resolución temporal superior a la normal.

- 30 Dado que el ejemplo utiliza una trama replegada en el dominio de tiempo como base para el análisis espectral, existe la posibilidad de conmutación instantánea entre análisis espectral no segmentado en base a la trama replegada en el dominio de tiempo, denominado procesamiento de resolución en todas las frecuencias, y análisis espectral segmentado basado en segmentos relativamente más cortos, denominado procesamiento de resolución temporal aumentada.

- 35 Preferentemente, dicha conmutación instantánea se lleva a cabo mediante una funcionalidad de conmutación 17 que depende de la detección de un transitorio de señal en la señal de entrada. El transitorio se puede detectar en el dominio de tiempo, en el dominio de tiempo replegado o incluso en el dominio de frecuencia. Habitualmente, una trama transitoria se procesa con una resolución temporal mayor que una trama estacionaria, que puede a continuación ser procesada utilizando un procesamiento normal de todas las frecuencias.

Existe asimismo la posibilidad de conmutar instantáneamente la resolución temporal utilizando un número mayor o menor de segmentos de tiempo para el análisis espectral.

- 40 Preferentemente, el repliegue en el dominio de tiempo, la segmentación en tiempo y el análisis espectral se repiten para cada una de una serie de tramas solapadas consecutivas.

El dispositivo de procesamiento de señal de la figura 7 puede ser parte de un codificador de audio, tal como el codificador de audio 10 de la figura 1 o de la figura 20, que utiliza codificación de transformada para el análisis espectral.

- 45 En base al procedimiento "directo" anterior, la cadena de operaciones inversas para mapear un conjunto de coeficientes espectrales a una trama en el dominio de tiempo es fácil y naturalmente evidente para un experto en la materia.

- 50 En resumen, en otro ejemplo, se lleva a cabo un análisis espectral inverso en base a diferentes subconjuntos de coeficientes espectrales con el fin de generar, para cada subconjunto de coeficientes espectrales, una subtrama transformada inversa, denominada asimismo segmento. A continuación se lleva a cabo una segmentación en tiempo inversa en base a las subtramas transformadas inversas solapadas para combinar estas subtramas en una trama replegada en el dominio de tiempo, y se lleva a cabo un repliegue inverso en el dominio de tiempo en base a la trama replegada en el dominio de tiempo para permitir la reconstrucción de la señal en el dominio de tiempo.

- 55 El repliegue inverso en el dominio de tiempo se lleva a cabo habitualmente para reconstruir una primera trama en el dominio de tiempo, y el procedimiento global puede a continuación sintetizar la señal en el dominio de tiempo en base a sumar-solapar la primera trama en el dominio de tiempo con una posterior segunda trama en el dominio de

tiempo reconstruida. Se puede hacer referencia, por ejemplo, a las operaciones generales de solapar-sumar de la figura 4B.

5 Preferentemente, el procesamiento de señal inverso incluye por lo menos una de síntesis de señal y descodificación de audio. El análisis espectral inverso se puede basar en cualquiera de una serie de diferentes transformadas inversas, preferentemente transformadas traslapadas. Por ejemplo, en aplicaciones de descodificación de audio, es beneficioso utilizar la transformada MDCT inversa.

A continuación se explica una visión general más detallada y una explicación de la cadena inversa de operaciones así como implementaciones preferidas.

10 La figura 8 es un diagrama de bloques esquemático de un dispositivo de acuerdo con otra realización preferida a modo de ejemplo de la invención. Además de los bloques básicos de la figura 7, el dispositivo de la figura 8 incluye además una o varias unidades de procesamiento, tales como la unidad de aplicación de función ventana 11 y la unidad de reordenamiento 13.

15 En el ejemplo de la figura 8, la unidad opcional de aplicación de la función ventana 11 lleva a cabo la aplicación de la función ventana en base a una de las tramas solapadas para generar una trama en ventana, que se envía a la unidad TDA 12 para el repliegue en el dominio de tiempo. Básicamente, la aplicación de la función ventana se puede llevar a cabo para mejorar las propiedades de selectividad en frecuencias de la transformada. La forma de la ventana se puede optimizar para satisfacer determinados criterios de selectividad en frecuencias, se pueden utilizar diversas técnicas de optimización y estas son bien conocidas por los expertos en la materia.

20 Para mantener una coherencia temporal completa de la señal de entrada, es beneficioso aplicar un reordenamiento del repliegue en el dominio de tiempo. Por esta razón, puede estar dispuesta una unidad de reordenamiento 13 para reordenar la trama replegada en el dominio de tiempo con el fin de generar una trama replegada en el dominio de tiempo reordenada, que se envía a la unidad de segmentación 14. De este modo, se puede llevar a cabo segmentación en base a la trama replegada en el dominio de tiempo reordenada. El analizador espectral 16 funciona preferentemente sobre los segmentos generados a partir de la unidad de segmentación en tiempo 14 para obtener un análisis espectral segmentado con una resolución temporal superior a la normal.

25 La figura 9 es un diagrama de bloques esquemático de un dispositivo de acuerdo con otra realización preferida a modo de ejemplo de la invención. El ejemplo de la figura 9 es similar al de la figura 8, excepto en que en la figura 9 se indica explícitamente que la segmentación en tiempo está basada en un conjunto de funciones ventana adecuadas, y que el análisis espectral está basado en la aplicación de transformada sobre segmentos de la trama replegada en el dominio de tiempo reordenada.

En un ejemplo particular, la segmentación involucra añadir relleno de ceros a la trama replegada en el dominio de tiempo reordenada y dividir la señal resultante en segmentos relativamente cortos y preferentemente solapados.

Preferentemente, el análisis espectral se basa en aplicar una transformada traslapada, tal como MDCT o MLT, a cada uno de dichos segmentos solapados.

35 A continuación se describirá la invención haciendo referencia a otras realizaciones ejemplares y no limitativas.

Tal como se ha mencionado, la invención se basa en el concepto de utilizar la señal replegada en el tiempo (salida de la operación de repliegue en el dominio de tiempo) como una nueva trama de señal en la que se aplica el análisis espectral. Al cambiar la resolución temporal de la transformada que se aplica después del repliegue temporal para obtener el coeficiente (por ejemplo, MDCT), por ejemplo el DCT_{IV} , la invención permite obtener un análisis espectral sobre segmentos de tiempo arbitrarios con muy poca sobrecarga en complejidad, así como instantáneamente, es decir, sin retardo adicional.

Para obtener un análisis de señal con una resolución temporal predeterminada es suficiente aplicar directamente las transformadas ortogonales de longitudes apropiadas preferentemente sobre segmentos solapados de la señal de entrada en ventana, replegada temporalmente.

45 La salida de cada una de esas transformadas de menor longitud conducirá a un conjunto de coeficientes representativo del contenido de frecuencia de cada segmento en cuestión. El conjunto de coeficientes para todos los segmentos proporcionará instantáneamente un mosaico de tiempo-frecuencia arbitrario de la trama de señal original.

Esta descomposición instantánea puede ser utilizada para atenuar el efecto de pre-eco, por ejemplo en el caso de transitorios, así como para proporcionar una representación eficiente de la señal que permita una codificación eficiente en velocidad binaria, de la trama en cuestión.

50 Los segmentos solapados de la señal en ventana replegada en tiempo no tienen que tener la misma longitud. Debido a la correspondencia temporal entre segmentos en el dominio replegado en tiempo y el dominio de tiempo normal, el nivel deseado del análisis de resolución temporal determinará el número de segmentos así como la longitud de cada uno de los segmentos sobre los que se lleva a cabo el análisis de frecuencia.

La mejor aplicación de la invención es junto con un detector de transitorios y/o en el contexto de codificación mediante medir la ganancia de codificación obtenida para un conjunto dado de segmentaciones en tiempo, incluyendo esto estimaciones de ganancia de codificación tanto en lazo abierto como en lazo cerrado para cada prueba de segmentación en tiempo.

- 5 La invención es útil, por ejemplo, junto con el estándar ITU-T G.722.1, y especialmente para el estándar "ITU-T G.722.1 fullband extension for 20 kHz full-band audio", ahora renombrado el estándar ITU-T G.719, tanto para codificación como para decodificación, tal como se ejemplificará más adelante.

10 La invención permite una conmutación instantánea de la resolución temporal de la transformada global (por ejemplo, en base a MDCT). Por lo tanto, a diferencia de la conmutación entre ventanas, la invención no requiere ningún retardo.

La invención tiene una complejidad muy baja y no se requiere ningún banco de filtros adicional. La invención utiliza preferentemente la misma transformada que la MDCT, es decir DCT de tipo IV. La invención gestiona de manera eficiente la supresión de artefactos de pre-eco mediante conmutar instantáneamente a una mayor resolución temporal.

- 15 La invención permitiría asimismo construir esquemas de codificación de lazo cerrado/abierto basados en segmentaciones en tiempo adaptativas de señal.

Para una mejor comprensión de la invención, se describirán a continuación ejemplos más detallados de operaciones individuales (posiblemente opcionales) de procesamiento de señal así como otros ejemplos de implementaciones globales.

- 20 Si existen requisitos estrictos sobre coherencia temporal, se recomienda el denominado reordenamiento.

Reordenamiento TDA

25 Para mantener la coherencia temporal de la señal de entrada, la salida de la operación de repliegue en el dominio de tiempo tiene que ser reordenada antes de su posterior procesamiento. La operación de ordenamiento es necesaria, sin ordenamiento las funciones base del banco de filtros resultante tendrán respuestas incoherentes en tiempo y frecuencia. Se muestra un ejemplo de operación de reordenamiento en la figura 10, e implica intercambiar las mitades superior e inferior de la señal de salida TDA $\tilde{x}(n)$. Este reordenamiento es solamente conceptual y en realidad no hay cálculos involucrados. La invención no se limita al ejemplo mostrado en la figura 10. Por supuesto, se pueden implementar otros tipos de reordenamiento.

Realización simple - mejora de la resolución temporal

30 Una primera realización simple muestra cómo duplicar la resolución temporal según la presente invención. Por consiguiente, se aplica un análisis de tiempo-frecuencia $v(n)$, para duplicar la resolución temporal, $v(n)$ se divide en dos segmentos preferentemente en solapamiento. Dado que $v(n)$ es una señal limitada en el tiempo, se añade una cantidad de relleno de ceros al comienzo y al final de $v(n)$. Preferentemente, la señal de entrada es una señal en ventana replegada en el tiempo reordenada, de longitud N. La longitud del relleno de ceros depende de la longitud de la señal $v(n)$ y de la cantidad deseada de segmentos, en este caso, dado que se desean dos segmentos solapados, la longitud del relleno de ceros es igual a la cuarta parte de la longitud de $v(n)$ y se añaden al inicio y al final de $v(n)$. Utilizar dicho relleno de ceros conduce a dos segmentos solapados al 50 % de la misma longitud que la longitud de $v(n)$.

40 Preferentemente, los segmentos solapados resultantes son sometidos a función ventana, tal como se ejemplifica en la figura 11. Se debe observar que aunque la forma de la ventana se puede optimizar hasta cierto punto para la aplicación deseada, ésta tiene que observar los requisitos de reconstrucción perfecta. Esto se puede ver en la figura 11, donde la mitad derecha de la ventana del segundo segmento tiene un valor 1 para la parte que aplica a la señal $v(n)$ y el valor 0 para el relleno de ceros añadido.

45 Cada uno de los segmentos obtenidos tiene una longitud de exactamente N. Aplicar la MDCT sobre cada segmento conduce a $N/2$ coeficientes; es decir, un total de N coeficientes, por lo que el banco de filtros resultante está muestreado críticamente, ver la figura 11. Debido a las limitaciones sobre las formas de las ventanas, la operación es invertible y aplicar las operaciones inversas sobre los dos conjuntos de coeficientes MDCT (coeficientes MDCT de los segmentos 1 y 2) conducirá de nuevo a la señal $v(n)$.

50 Para esta realización, las funciones base del banco de filtros resultantes tienen una localización temporal mejorada pero pierden en la localización en frecuencia, lo que es una consecuencia bien conocida del principio de incertidumbre tiempo-frecuencia.

55 La figura 12 muestra las dos funciones básicas que están relacionadas con la frecuencia normalizada a 0,25. Claramente, la dispersión temporal se limita bastante, si bien se observa asimismo que existe una propagación de la dispersión temporal que se debe al solapamiento de las dos secciones de la señal replegado en el tiempo. Esta propagación en el dominio de tiempo es una consecuencia de la cancelación del repliegue en el dominio de tiempo y

estaría siempre presente. Sin embargo, se puede atenuar mediante una elección adecuada (optimización numérica) de las funciones ventana. La figura 12 muestra asimismo las respuestas de frecuencia. A modo de comparación, las funciones base MDCT originales se muestran en la figura 13, si bien éstas corresponden a un muestreo mucho más estrecho del dominio de frecuencia, y su extensión temporal es mucho mayor. La figura 13 muestra las funciones base originales correspondientes al banco de filtros MLT (MDCT + ventana seno).

Resoluciones temporales superiores

Se puede obtener una mayor resolución temporal dividiendo en más segmentos la señal replegada en tiempo reordenada. Las figuras 14 y 15 muestran cómo se consigue esto para cuatro y ocho segmentos, respectivamente. La figura 14 muestra una resolución temporal superior mediante división en cuatro segmentos, y la figura 15 muestra una resolución temporal superior mediante división en ocho segmentos. Tal como se comprenderá, se puede utilizar cualquier número adecuado de segmentos de tiempo, dependiendo de la resolución temporal deseada.

En general, la unidad de segmentación en tiempo está configurada para generar un número seleccionable de N de segmentos en base a una trama replegada en el dominio de tiempo, donde N es un entero igual o mayor que 2.

Para el caso de cuatro segmentos, la figura 16 muestra una realización de la transformada global resultante. La aplicación de la función ventana sobre una trama de entrada se lleva a cabo en una unidad 11 de aplicación de la función ventana, el repliegue en tiempo se lleva a cabo en una unidad 12 de repliegue en el dominio de tiempo, y el reordenamiento opcional se lleva a cabo en una unidad de reordenamiento 13. A continuación se realiza un análisis espectral segmentado aplicando la post-función ventana sobre cuatro segmentos utilizando unidades de post-función ventana 14 y transformadas segmentadas mediante las unidades de transformada 16. Preferentemente, la transformada segmentada global se basa en MDCT segmentada, utilizando repliegue temporal y DCT_{IV} para cada segmento.

Mosaico no uniforme en el dominio de tiempo

Con esta invención es posible asimismo obtener segmentaciones en tiempo no uniformes, según el mismo concepto. Existen por lo menos dos posibles maneras de llevar a cabo dicha operación. Un primer procedimiento está basado en una segmentación en tiempo no uniforme de la señal replegada en el tiempo reordenada. De este modo, las ventanas utilizadas para segmentar la señal tienen longitudes diferentes.

Un segundo procedimiento está basado en un enfoque jerárquico. La idea es aplicar en primer lugar segmentación en tiempo gruesa y a continuación volver a aplicar la invención a los segmentos gruesos resultantes hasta que se obtiene el mosaico deseado.

La figura 17 muestra un ejemplo de cómo se puede implementar este segundo procedimiento. Para este ejemplo, en primer lugar la señal se divide en dos segmentos de tiempo según la presente invención; a continuación uno de los segmentos se vuelve a dividir en dos segmentos. Un ejemplo de una transformada adecuada es la transformada MDCT, utilizando repliegue en tiempo y DCT_{IV} para cada segmento considerado.

Funcionamiento con detección de transitorios

La invención puede ser utilizada para atenuar los artefactos de pre-eco y es en este caso donde se asocia mejor con un detector de transitorios, tal como se ejemplifica en la figura 18. Al detectar un transitorio, el detector de transitorios activaría un indicador (IsTransient (es transitorio)). El indicador del detector de transitorios utilizaría a continuación el mecanismo de conmutación 17 para conmutar instantáneamente de un proceso normal de resolución de todas las frecuencias (análisis espectral no segmentado) a una resolución temporal superior (análisis espectral segmentado), tal como se representa en la figura 18. Con esta realización es posible entonces analizar señales transitorias con una resolución temporal mucho más fina, eliminando por lo tanto los molestos artefactos de pre-eco.

Operaciones de codificación en lazo cerrado/lazo cerrado.

La invención se puede utilizar asimismo como un medio para encontrar el mosaico de tiempo-frecuencia óptimo para el análisis de una señal antes de la codificación. Se pueden utilizar dos modos como ejemplo, lazo cerrado y lazo abierto. En un funcionamiento en lazo abierto, un dispositivo externo decidiría el mejor mosaico (en términos de eficiencia de codificación) de tiempo-frecuencia para una trama de señal determinada, y utilizaría la invención para analizar la señal según el mosaico óptimo. En un funcionamiento en lazo cerrado, se utiliza un conjunto de mosaicos predefinidos, se analiza la señal para cada uno de estos mosaicos y ésta se codifica según el mosaico. Para cada mosaico se calcula una medida de la fidelidad. Se selecciona el mosaico que conduce a la mejor fidelidad. El mosaico seleccionado, junto con los coeficientes codificados correspondientes a este mosaico, se transmiten al descodificador.

Tal como se ha mencionado, los principios y conceptos descritos anteriormente para el procedimiento directo permiten a un experto en la materia realizar una cadena inversa de operaciones en un procedimiento inverso.

La figura 19 es un diagrama de bloques que muestra un ejemplo básico de un dispositivo de procesamiento de señal para funcionamiento basado en coeficientes espectrales representativos de una señal en el dominio de tiempo. El

dispositivo incluye un transformador inverso 42, una unidad 44 para segmentación en tiempo inversa, una unidad de TDA inverso 46 y un sumador-solape 48 opcional.

Básicamente, es deseable sintetizar una señal en el dominio de tiempo a partir de un flujo de bits cuantificado y codificado. Una vez que se han recuperado los coeficientes espectrales, se lleva a cabo un análisis espectral inverso en el transformador inverso 42 en base a diferentes subconjuntos de coeficientes espectrales con el fin de generar, para cada subconjunto de coeficientes espectrales, una subtrama transformada inversa, denominada asimismo un segmento. La unidad 44 para la segmentación en tiempo inversa funciona en base a subtramas transformadas inversas solapadas, para combinar estas subtramas en una trama replegada en el dominio de tiempo. La unidad de TDA inverso 46 lleva a cabo a continuación un repliegue inverso en el dominio de tiempo en base a la trama replegada en el dominio de tiempo para permitir la reconstrucción de la señal en el dominio de tiempo.

El repliegue inverso en el dominio de tiempo se lleva a cabo habitualmente para reconstruir una primera trama en el dominio de tiempo, y el procedimiento global puede a continuación sintetizar la señal en el dominio de tiempo en base a sumar-solapar la primera trama del dominio de tiempo con una posterior segunda trama del dominio de tiempo reconstruida, utilizando el sumador-solape 48.

Se pueden incluir fases de procesamiento anterior, intermedia y posterior en el dispositivo de la figura 19.

El análisis espectral inverso se puede basar en cualquiera de una serie de diferentes transformadas inversas, preferentemente transformadas traslapadas. Por ejemplo, en aplicaciones de descodificación de audio, es beneficioso utilizar la transformada MDCT inversa (IMDCT).

Preferentemente, el dispositivo de procesamiento de señal está configurado para síntesis de señal y/o descodificación de audio para reconstruir la señal de audio en el dominio de tiempo. En una realización preferida de la invención, el dispositivo de procesamiento de señal de la figura 19 es parte de un descodificador de audio, tal como el descodificador de audio 40 de la figura 1 o a figura 21.

A continuación, se describirá la invención en relación con una realización de códec específica a modo de ejemplo y no limitativa, adecuada para la extensión del códec de banda completa ITU-T G.722.1, es decir el códec ITU-T G.719. En este ejemplo particular, el códec se presenta como un códec de audio basado en transformada, de baja complejidad, que funciona preferentemente a una velocidad de muestreo de 48 kHz y ofrece un ancho de banda de audio completo que abarca desde 20 Hz hasta 20 kHz. El codificador procesa señales PCM lineales de 16 bits de entrada en tramas de 20 ms y el códec tiene un retardo global de 40 ms. El algoritmo de codificación está basado preferentemente en codificación de transformación con resolución temporal adaptativa, asignación de bits adaptativa y cuantificación de vectores de red de baja complejidad. Además, el descodificador puede sustituir los componentes del espectro no codificados por relleno de ruido adaptativo de la señal o extensión del ancho de banda.

La figura 20 es un diagrama de bloques de un codificador a modo de ejemplo adecuado para la extensión de banda completa. La señal de entrada muestreada a 48 kHz es procesada por medio de un detector de transitorios. Dependiendo de la detección de un transitorio, se aplica una transformada de resolución de alta frecuencia o de resolución de baja frecuencia (alta resolución temporal) en la trama de la señal de entrada. La transformada adaptativa se basa preferentemente en una transformada de coseno discreta modificada (MDCT) en el caso de tramas estacionarias. Para tramas no estacionarias, se utiliza una transformada de mayor resolución temporal sin necesidad de retardo adicional y con muy poca sobrecarga en complejidad. Las tramas no estacionarias tienen preferentemente 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. Se estima la norma de cada banda y la envolvente espectral resultante compuesta por las normas de todas las bandas es cuantificada y codificada. A continuación, los coeficientes son normalizados mediante las normas cuantificadas. Las normas cuantificadas se ajustan adicionalmente en base a una ponderación espectral adaptativa y se utilizan como entrada para la asignación de bits. Los coeficientes espectrales normalizados son vectores de red cuantificados y codificados en base a los bits asignados para cada banda de frecuencia. El nivel de los coeficientes espectrales no codificados es estimado, codificado y transmitido al descodificador. Preferentemente se aplica codificación Huffman a los índices de cuantificación tanto para los coeficientes espectrales codificados como para las normas codificadas.

La figura 21 es un diagrama de bloques de un descodificador a modo de ejemplo, adecuado para extensión de banda completa. En primer lugar se descodifica el indicador de transitorios, que indica la configuración de la trama, es decir, estacionaria o transitoria. Se descodifica la envolvente espectral y se utilizan los mismos ajustes de norma y asignación de bits, exactos en bits, en el descodificador para recalcular la asignación de bits que es esencial para descodificar los índices de cuantificación de los coeficientes de transformada normalizados.

Después de la descuantificación, se generan coeficientes espectrales no codificados de baja frecuencia (bits cero asignados), preferentemente utilizando un libro de códigos de relleno espectral construidos a partir de los coeficientes espectrales recibidos (coeficientes espectrales con asignación de bits distinta de cero).

Se puede utilizar un índice de ajuste del nivel de ruido para ajustar el nivel de los coeficientes regenerados. Los coeficientes espectrales no codificados de alta frecuencia son regenerados preferentemente utilizando extensión del ancho de banda.

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

Finalmente, se aplica la transformada inversa para recuperar la señal descodificada en el dominio de tiempo. Esto se realiza preferentemente aplicando la transformada de coseno discreta modificada inversa (IMDCT) para modos estacionarios, o bien la inversa de la transformada de mayor resolución temporal para el modo transitorio.

10 El algoritmo adaptado para extensión de banda completa está basado en tecnología de codificación de transformada adaptativa. Funciona en tramas de 20 ms de audio de entrada y de salida. Debido a que la ventana de transformada (longitud de función base) es de 40 ms y se utiliza un solape del 50 por ciento entre sucesivas tramas de entrada y salida, el tamaño de la memoria tampón anticipada es de 20 ms. Por lo tanto, el retardo algorítmico global es de 40 ms, que es la suma del tamaño de la trama más el tamaño anticipado. Todos los demás retardos adicionales experimentados en la utilización de un códec de banda completa G.722.1 son debidos a retardos computacionales y/o de transmisión de la red.

15 La figura 22 es un diagrama de bloques esquemático de un ejemplo particular de un transformador inverso y la implementación asociada para segmentación en tiempo inversa y reordenamiento opcional de acuerdo con una realización preferida de la invención. El transformador inverso está basado en DCT_{IV} en cascada con repliegue temporal inverso. Cuatro denominados sub-espectros $z_l^q(k)$, donde $l = 0, 1, 2, 3$, son procesados por el transformador inverso, y cada sub-espectro es, en primer lugar, sometido a transformada inversa por medio de un respectivo DCT_{IV} en el dominio replegado del dominio de tiempo, y a continuación sometidos a repliegue temporal inverso, es decir sometidos a repliegue inverso en el dominio de tiempo, para proporcionar una transformada de tipo MDCT inversa global para cada sub-espectro. La longitud de la señal resultante \tilde{x}_l^{qw} para cada índice de subtrama l es igual al doble de la longitud del espectro de entrada, es decir $L/2$.

20 Las señales sometidas a repliegue inverso en el dominio de tiempo resultantes para cada subtrama l son sometidas a función ventana utilizando la misma configuración de ventanas que las del codificador. Las señales en ventana resultantes se suman solapadas. Se debe observar que la ventana para la primera $m = 0$ y la última $m = 3$ subtramas es cero. Esto se debe a que se utiliza relleno de ceros en el codificador. Estos dos bordes de trama tienen que ser calculados y en la práctica se desechan. La señal resultante de las operaciones de sumar-solapar de todas las subtramas $v^q(n)$ se reordena utilizando la operación inversa llevada a cabo en el codificador, lo que conduce a la señal $\tilde{x}^q(n)$, $n=0, \dots, L-1$.

25 La salida de la transformada inversa, en modo estacionario o transitorio, es de longitud L . Antes de someter la señal a una función ventana (no mostrada en la figura 22) es primero sometida a repliegue inverso en el dominio de tiempo (ITDA, inverse time domain realiesed), conduciendo a una señal de longitud $2L$, según:

$$\tilde{x}^{wq} = \begin{bmatrix} 0 & \mathbf{I}_{L/2} \\ 0 & -\mathbf{J}_{L/2} \\ -\mathbf{J}_{L/2} & 0 \\ -\mathbf{I}_{L/2} & 0 \end{bmatrix} \tilde{x}^q$$

La señal resultante se somete a función ventana para cada trama r , según:

$$\tilde{x}^{(r)}(n) = h(n)\tilde{x}_{(r)}^{wq}(n), \quad n = 0, \dots, 2L-1,$$

donde $h(n)$ es una función ventana.

40 Finalmente, la señal de banda completa de salida se construye sumando con solapamiento las señales $\tilde{x}^{(r)}(n)$ para dos tramas consecutivas:

$$x^{(r)}(n) = \tilde{x}^{(r-1)}(n+L) + \tilde{x}^{(r)}(n), \quad n = 0, \dots, 2L-1.$$

Las realizaciones descritas anteriormente se proporcionan tan solo como ejemplos, y se deberá entender que la presente invención no se limita a las mismas.

45 Referencias

[1] B. Edler, "Codierung von Audiosignalen mit überlappender Transformation und adaptiven Fensterfunktionen" Frequenz, páginas 252-256, 1989.

[2] H. Malvar, "Lapped Transforms for efficient transform/subband coding". IEEE Trans. Acous., Speech, and Sig. Process., volumen 38, número 6, páginas 969-978, junio de 1990.

[3] J. Herre and J.D. Johnston, "Enhancing the performance of perceptual audio coders by using temporal noise Shaping (TNS)", in Proc. 101st Conv. Aud. Eng. Soc., preprint #4384, noviembre de 1996.

REIVINDICACIONES

1. Un procedimiento para procesamiento de señal que actúa sobre tramas solapadas de una señal de audio en el dominio de tiempo, comprendiendo dicho procedimiento las etapas de:
- 5 - llevar a cabo (S2) un repliegue en el dominio de tiempo (TDA) en base a una trama solapada, que tiene una longitud de $2N$, para generar una correspondiente trama replegada en el dominio de tiempo (\tilde{x}) con una longitud N ;
- dicho procedimiento, caracterizado por:
- reordenar (13) la trama replegada en el dominio de tiempo (\tilde{x}) para generar una trama replegada en el dominio de tiempo reordenada (v);
- 10 - llevar a cabo (S4) segmentación en tiempo, que comprende relleno de ceros de la trama replegada en el dominio de tiempo reordenada (v) de longitud N para producir una trama que tiene una longitud mayor que N , y dividir a continuación la trama producida resultante por lo menos en dos segmentos solapados que tienen cada uno una longitud igual o menor que N ; y
- 15 - llevar a cabo (S5) análisis espectral en base a dichos por lo menos dos segmentos solapados aplicando, en cada uno de dichos por lo menos dos segmentos solapados, una transformada de coseno discreta modificada (MDCT) al segmento para obtener, para cada segmento, un conjunto correspondiente de coeficientes representativos del contenido de frecuencias del segmento.
2. El procedimiento según la reivindicación 1, en el que dicha MDCT (16) está formada por una fase de operación de repliegue en el dominio de tiempo (TDA) seguida por una segunda fase basada en una transformada de coseno discreta (DCT) de tipo IV y cada segmento tiene una longitud menor que N .
- 20 3. El procedimiento según la reivindicación 1, que comprende la etapa de conmutar (17), de acuerdo con la detección de un transitorio de señal en dicha señal de audio, entre:
- análisis espectral no segmentado (116) basado en dicha trama replegada en el dominio de tiempo (x), denominado proceso de resolución en todas las frecuencias; y
- 25 - análisis espectral segmentado (14, 16) basado en dichos por lo menos dos segmentos solapados, denominado proceso de resolución temporal aumentada.
4. Un dispositivo (30, 40) para procesamiento de señal que actúa sobre tramas solapadas de una señal de audio, comprendiendo dicho dispositivo:
- medios para llevar a cabo repliegue en el dominio de tiempo (12) (TDA) en base a una trama solapada, que tiene una longitud de $2N$, para generar una trama replegada en el dominio de tiempo que tiene una longitud N ;
- 30 el dispositivo caracterizado por:
- medios para reordenar (13) la trama replegada en el dominio de tiempo con el fin de generar una trama replegada en el dominio de tiempo reordenada (v);
- medios para llevar a cabo segmentación (14) en tiempo, que comprende relleno de ceros de la trama replegada en el dominio de tiempo reordenada (v) de longitud N para producir una trama que tiene una longitud mayor que N , y dividir a continuación la trama producida resultante por lo menos en dos segmentos solapados que tienen cada uno una longitud igual o menor que N ; y
- 35 - un analizador espectral (16) configurado para llevar a cabo un análisis espectral segmentado, en base a dichos por lo menos dos segmentos solapados aplicando, sobre cada uno de dichos por lo menos dos segmentos solapados, una transformada de coseno discreta modificada (MDCT) para el segmento con el fin de obtener, para cada segmento, un correspondiente conjunto de coeficientes representativos del contenido de frecuencias del segmento.
- 40 5. El dispositivo (30, 40) según la reivindicación 4, en el que dicha MDCT está formada por una fase de operación de repliegue en el dominio de tiempo (TDA) seguida por una segunda fase basada en una transformada de coseno discreta (DCT) de tipo IV, y cada segmento tiene una longitud menor que N .
- 45 6. El dispositivo (30, 40) según la reivindicación 4, que comprende medios para conmutar (17), dependiendo de la detección de un transitorio de señal en dicha señal de audio, entre análisis espectral no segmentado (116) en base a dicha trama replegada en el dominio de tiempo (\tilde{x}), y análisis espectral segmentado (14, 16) en base a dichos por lo menos dos segmentos.

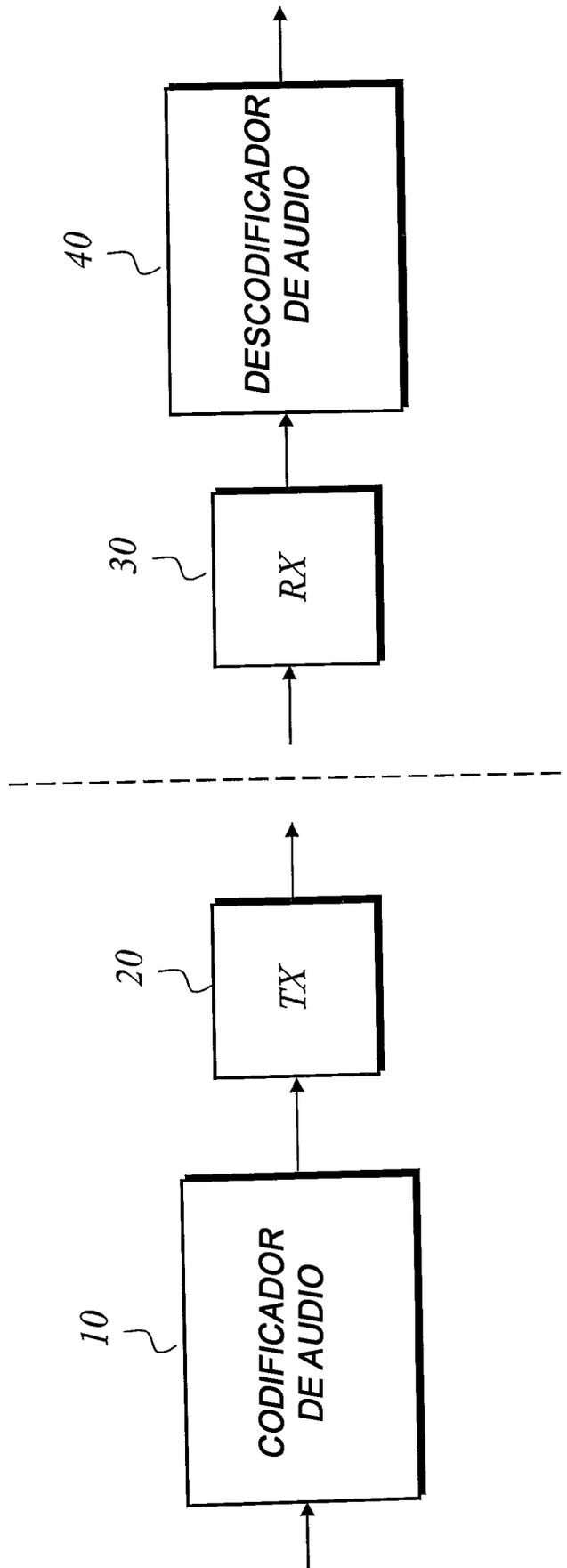


Fig. 1

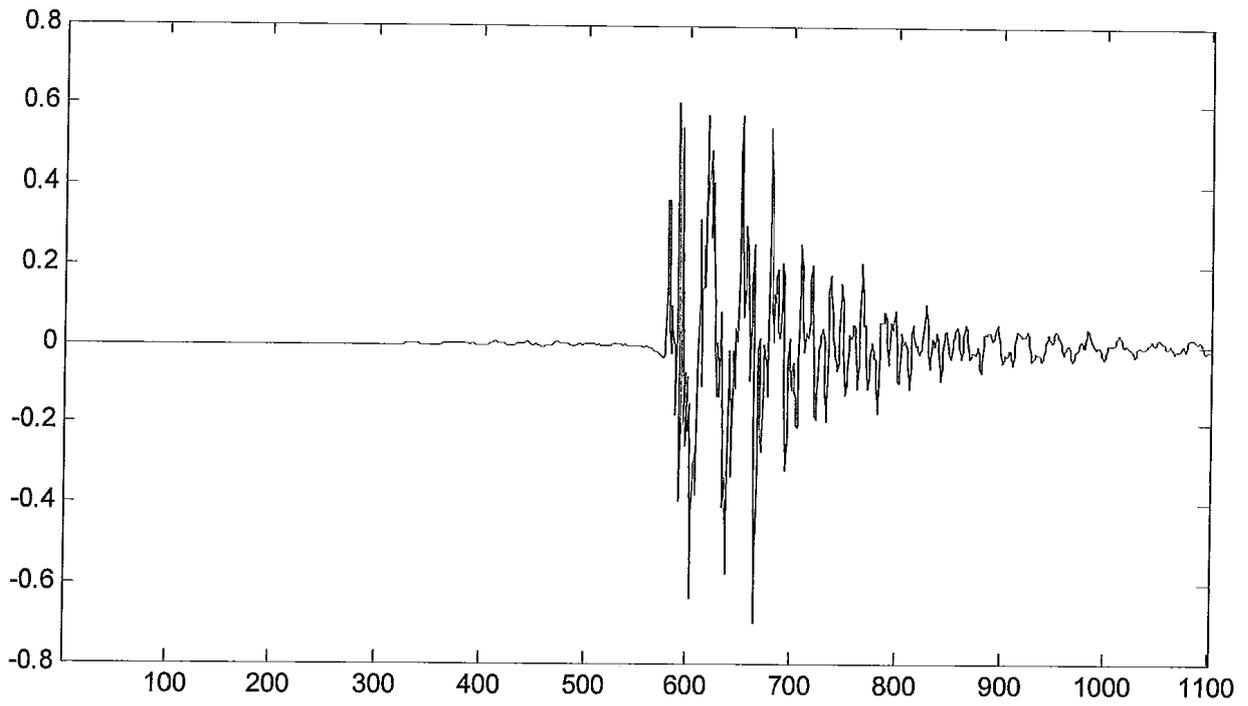


Fig. 2A

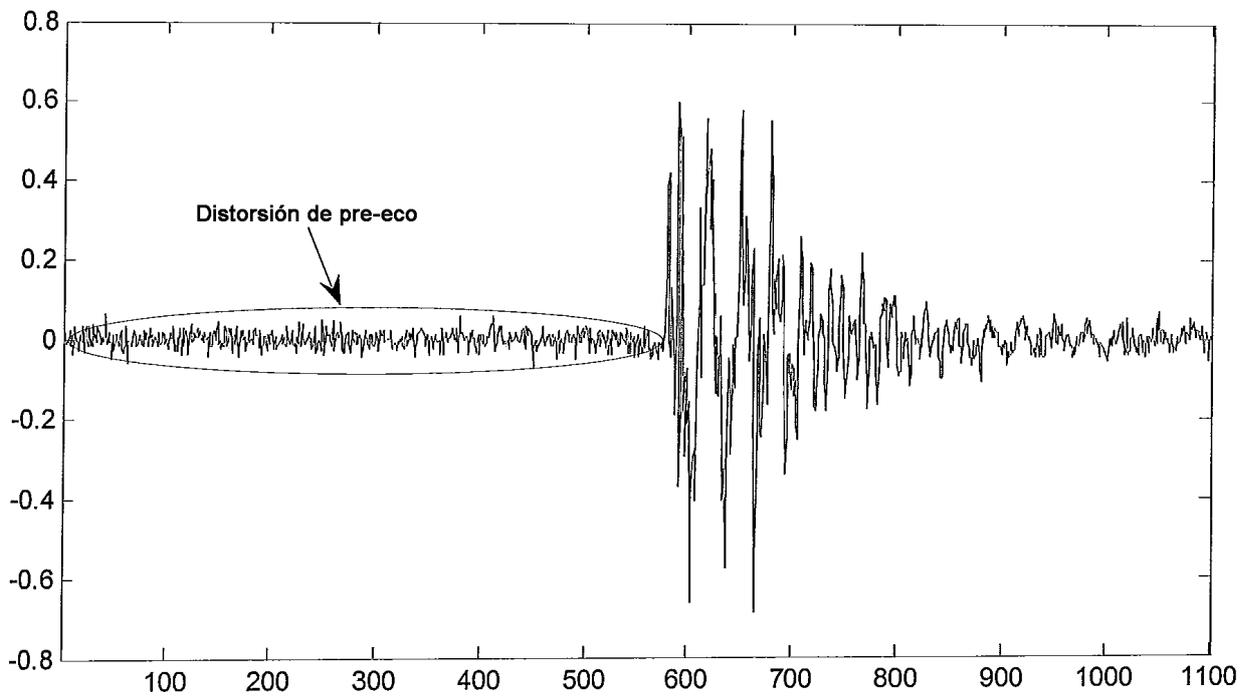


Fig. 2B

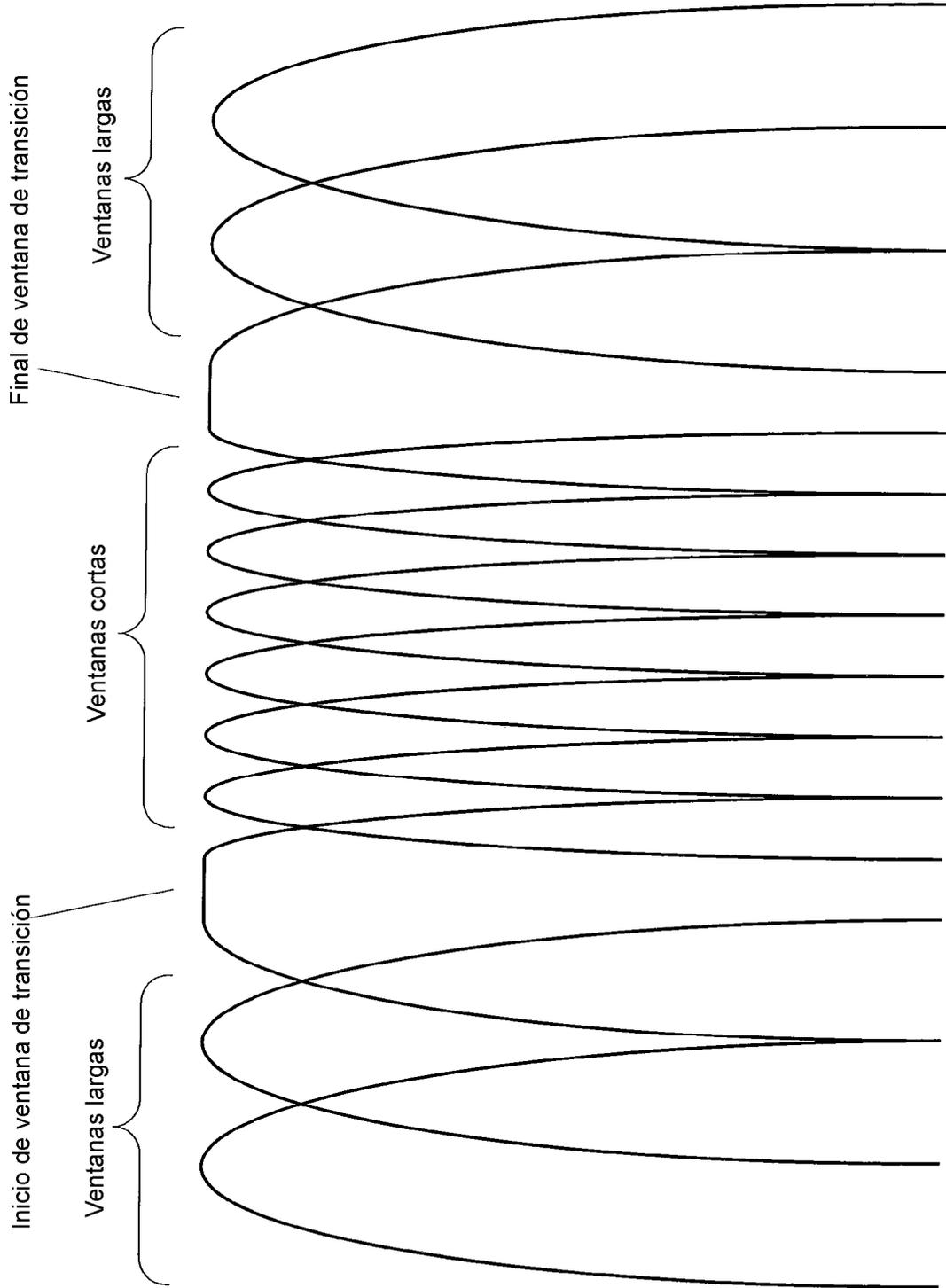


Fig. 3

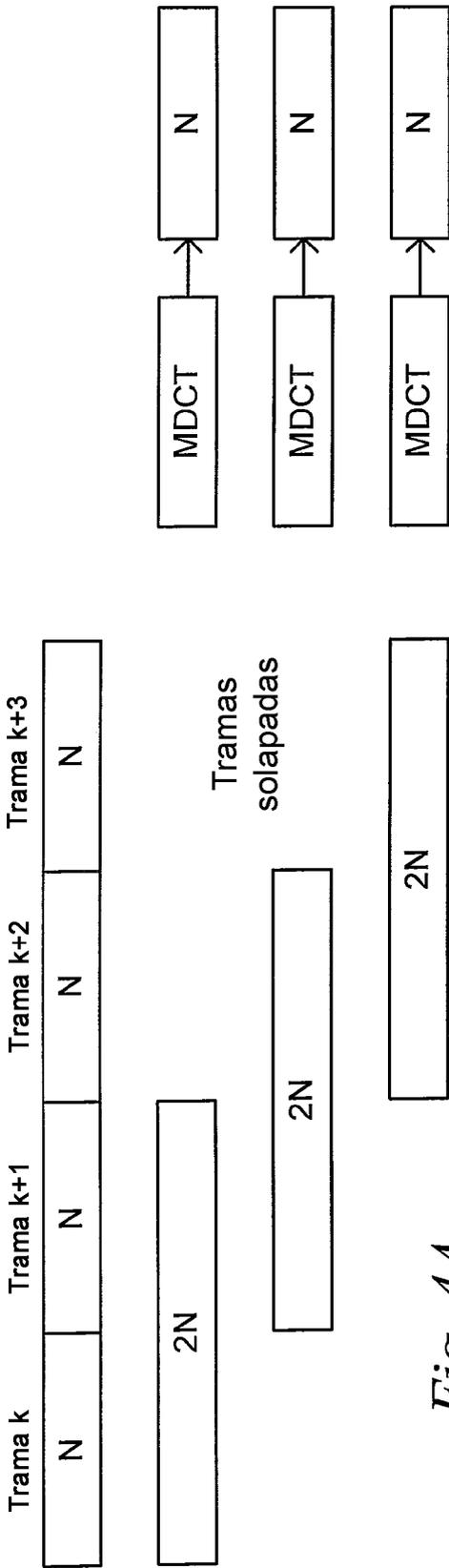


Fig. 4A

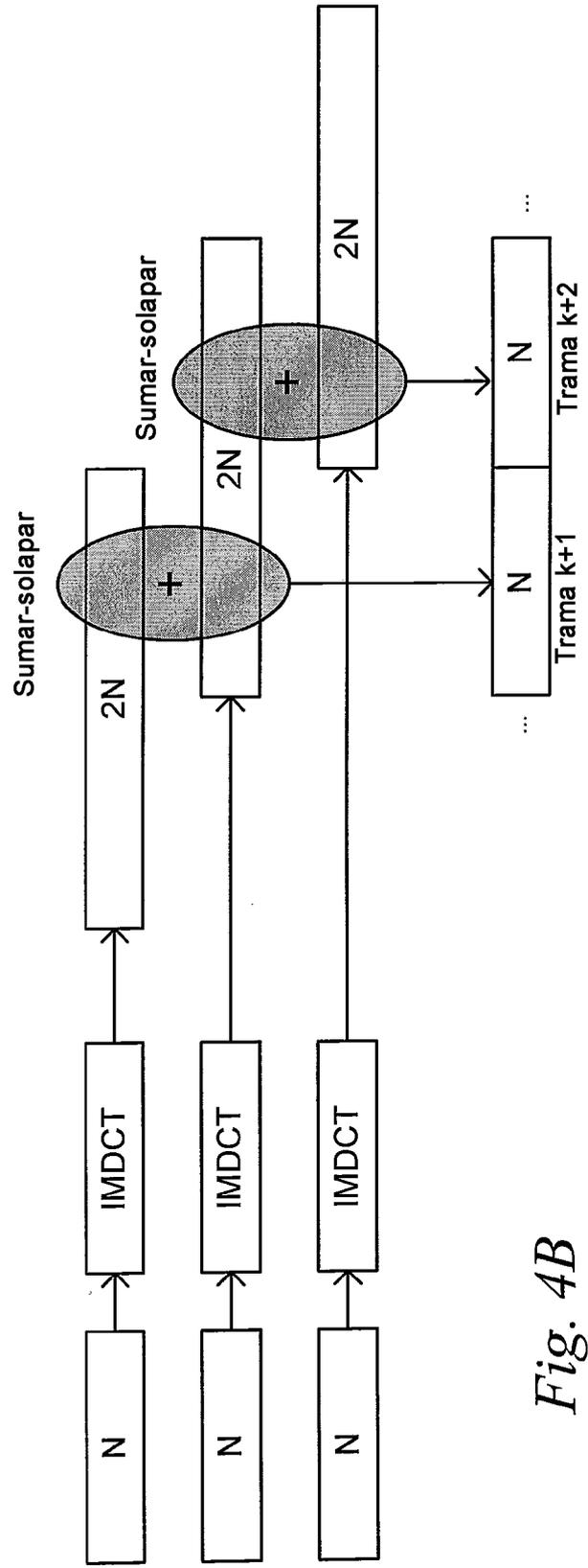
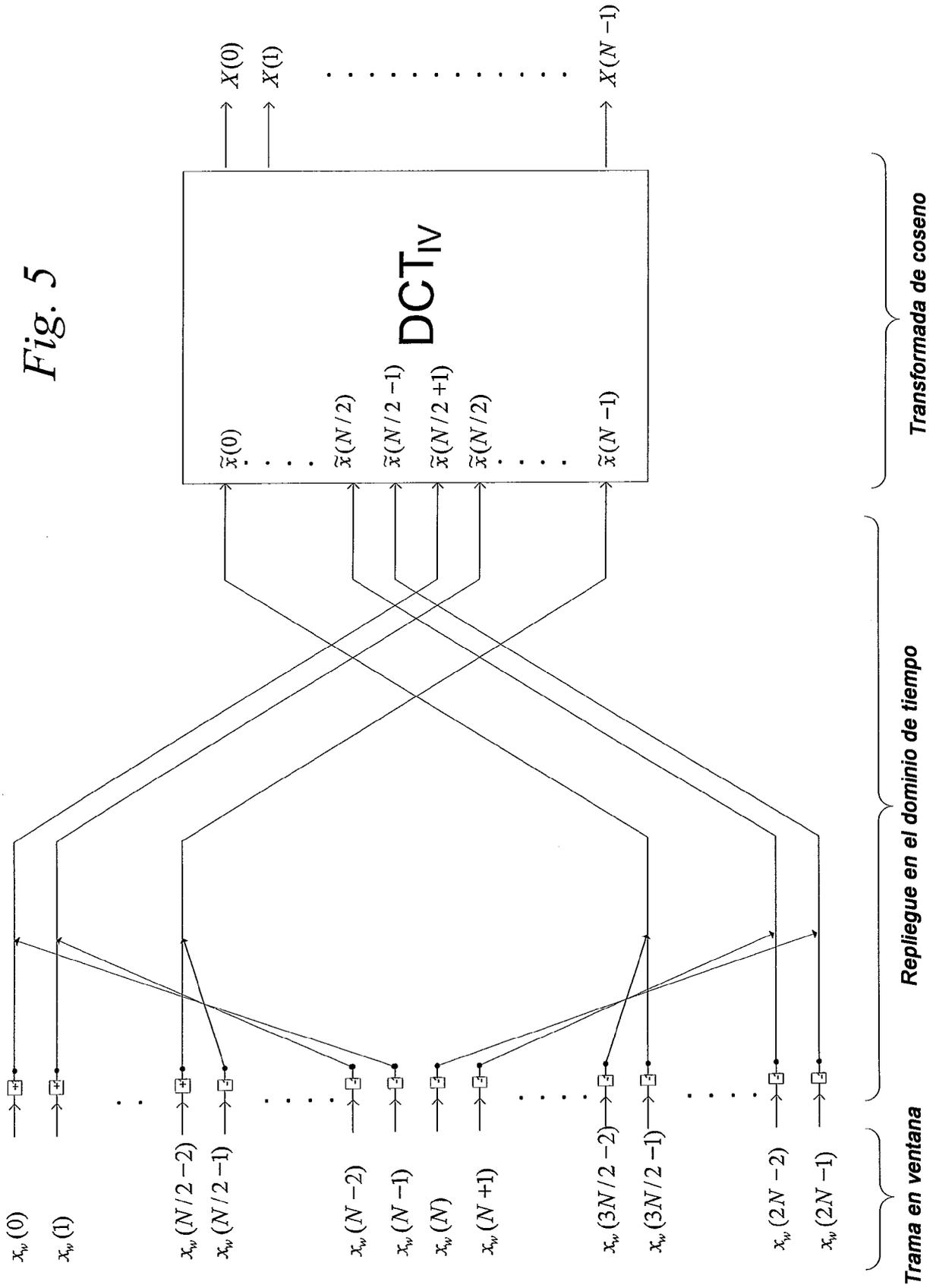


Fig. 4B

Fig. 5



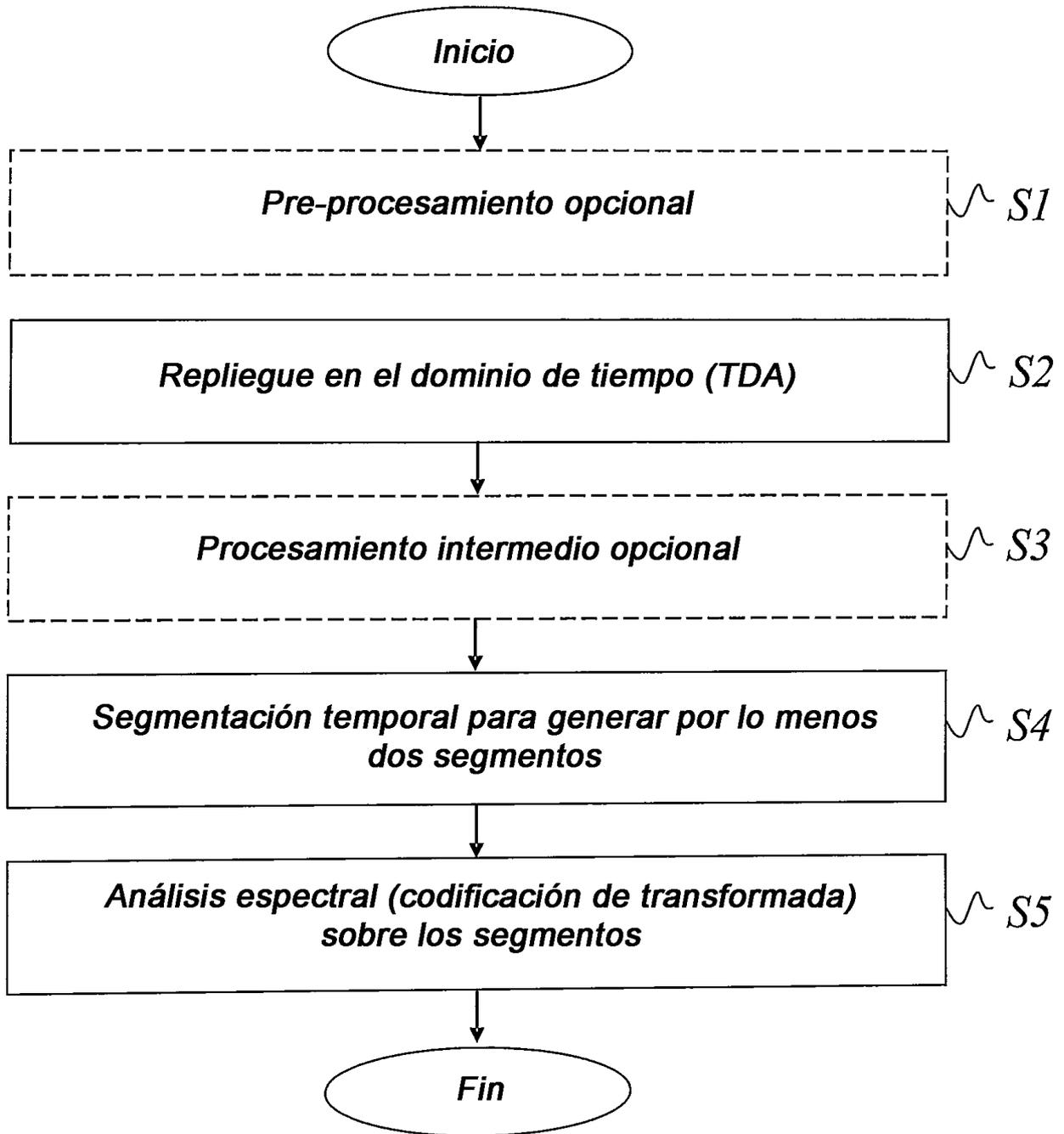


Fig. 6

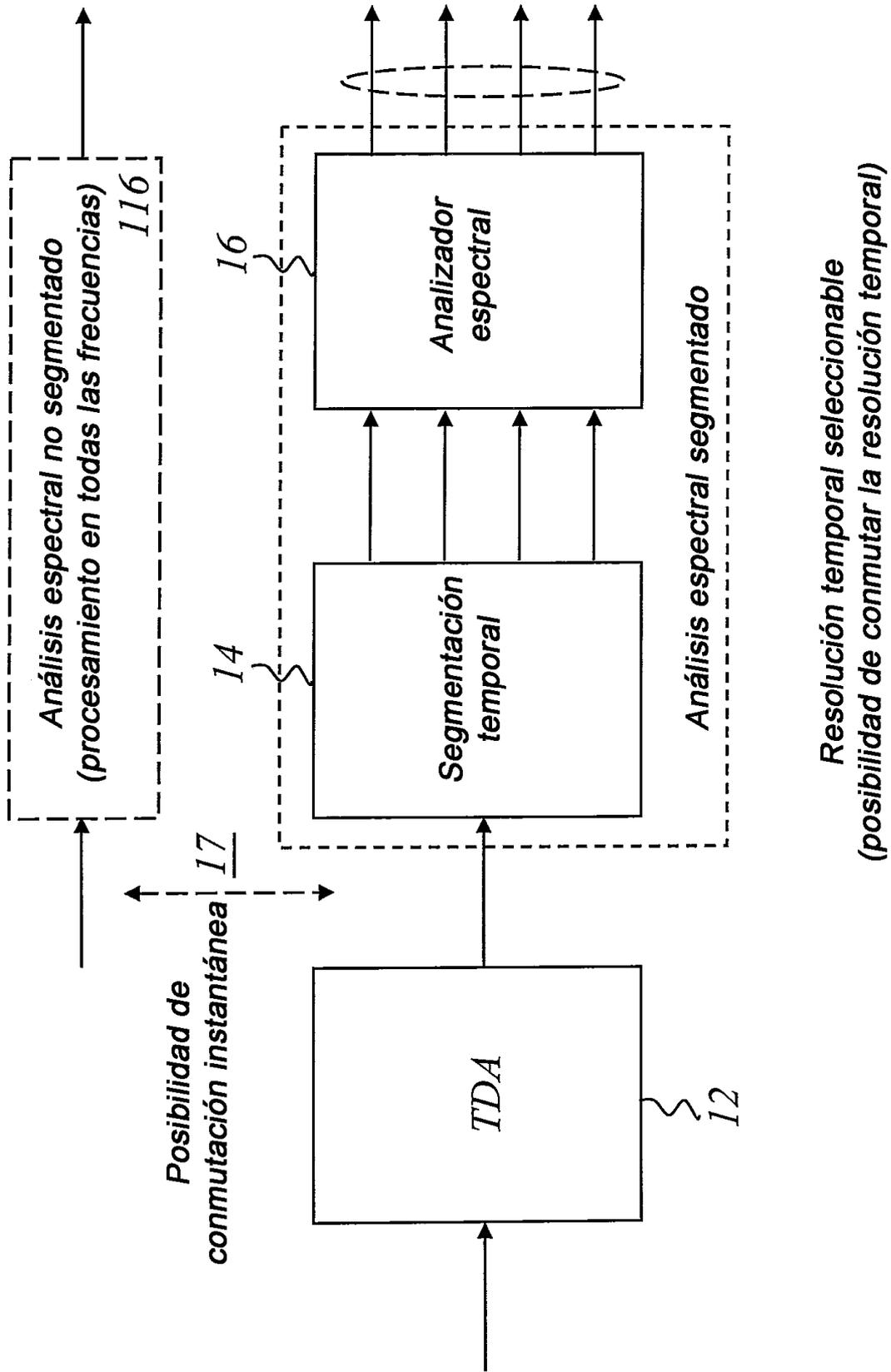


Fig. 7

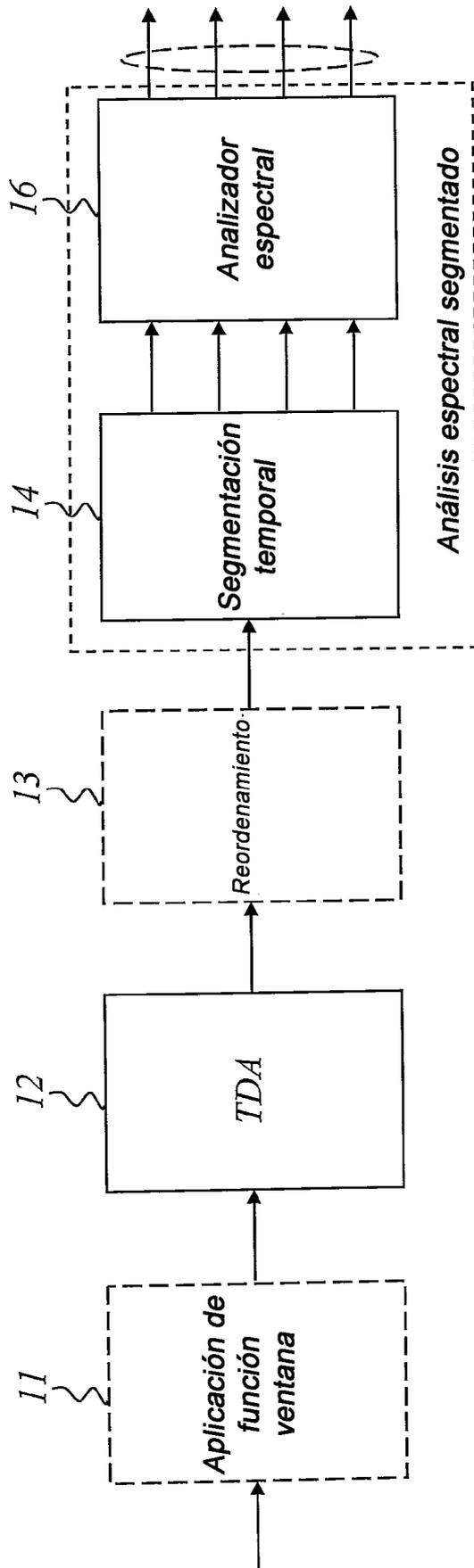


Fig. 8

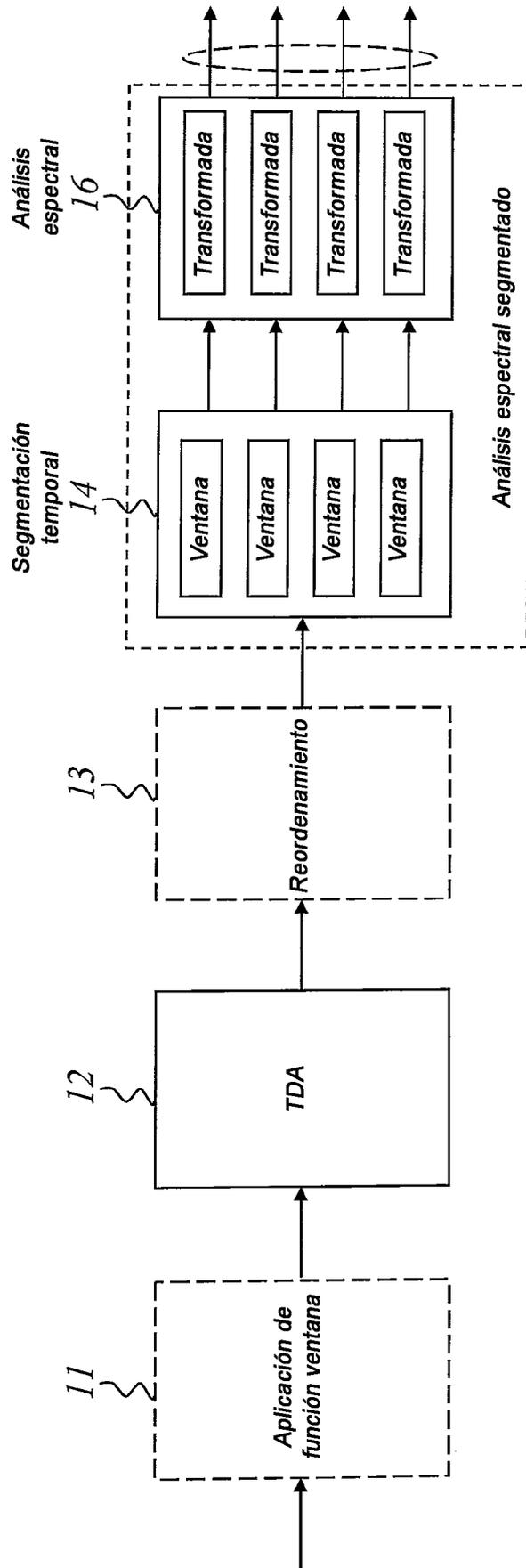


Fig. 9

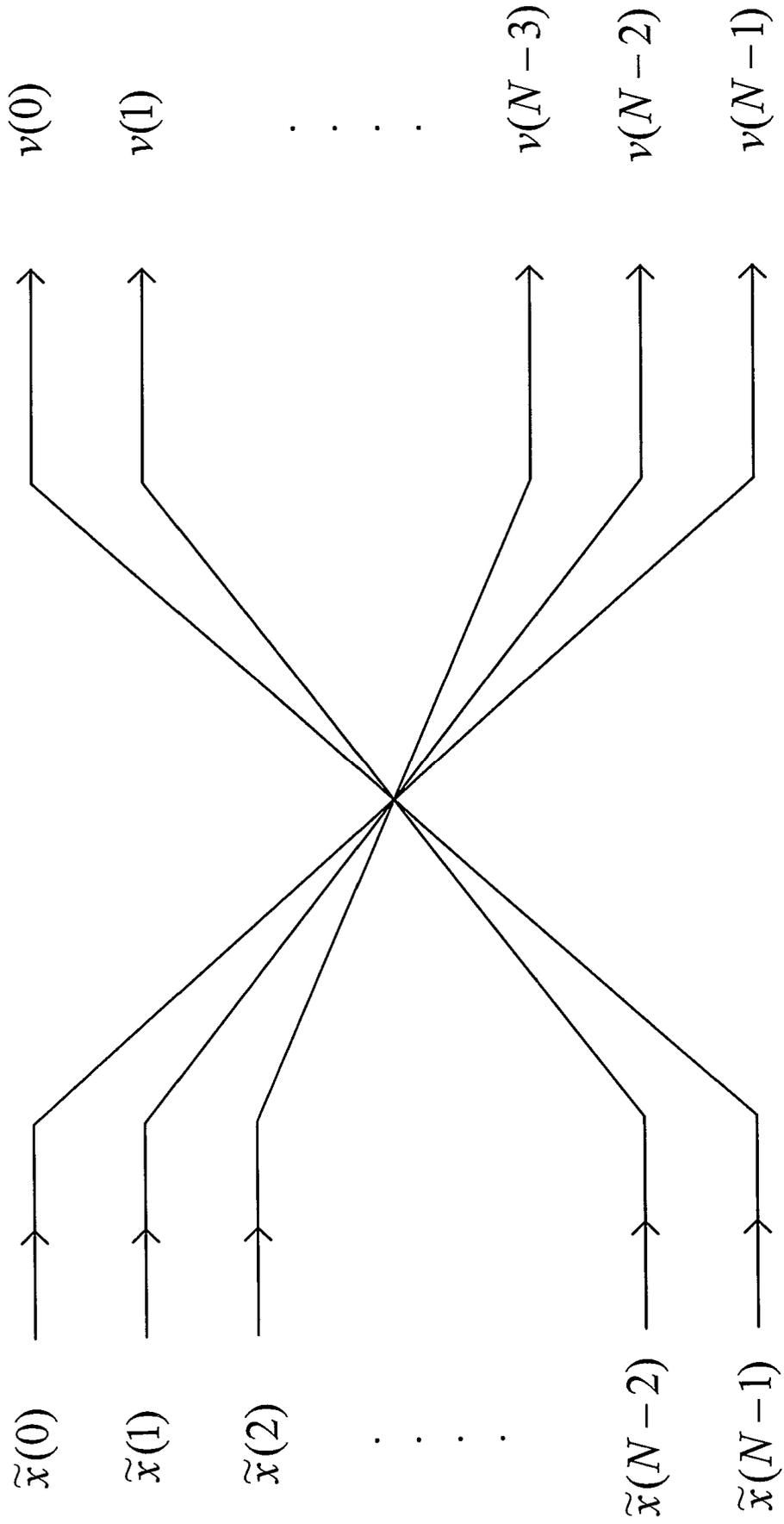


Fig. 10

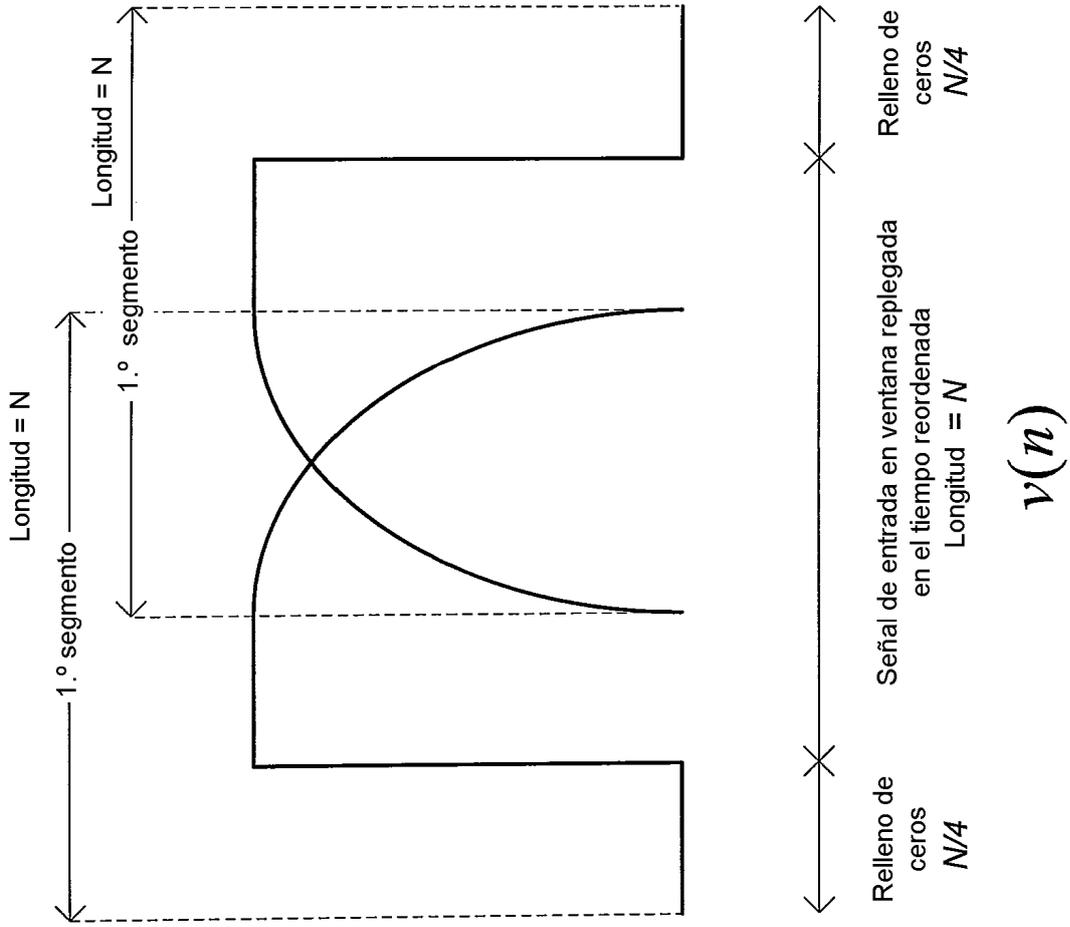


Fig. 11

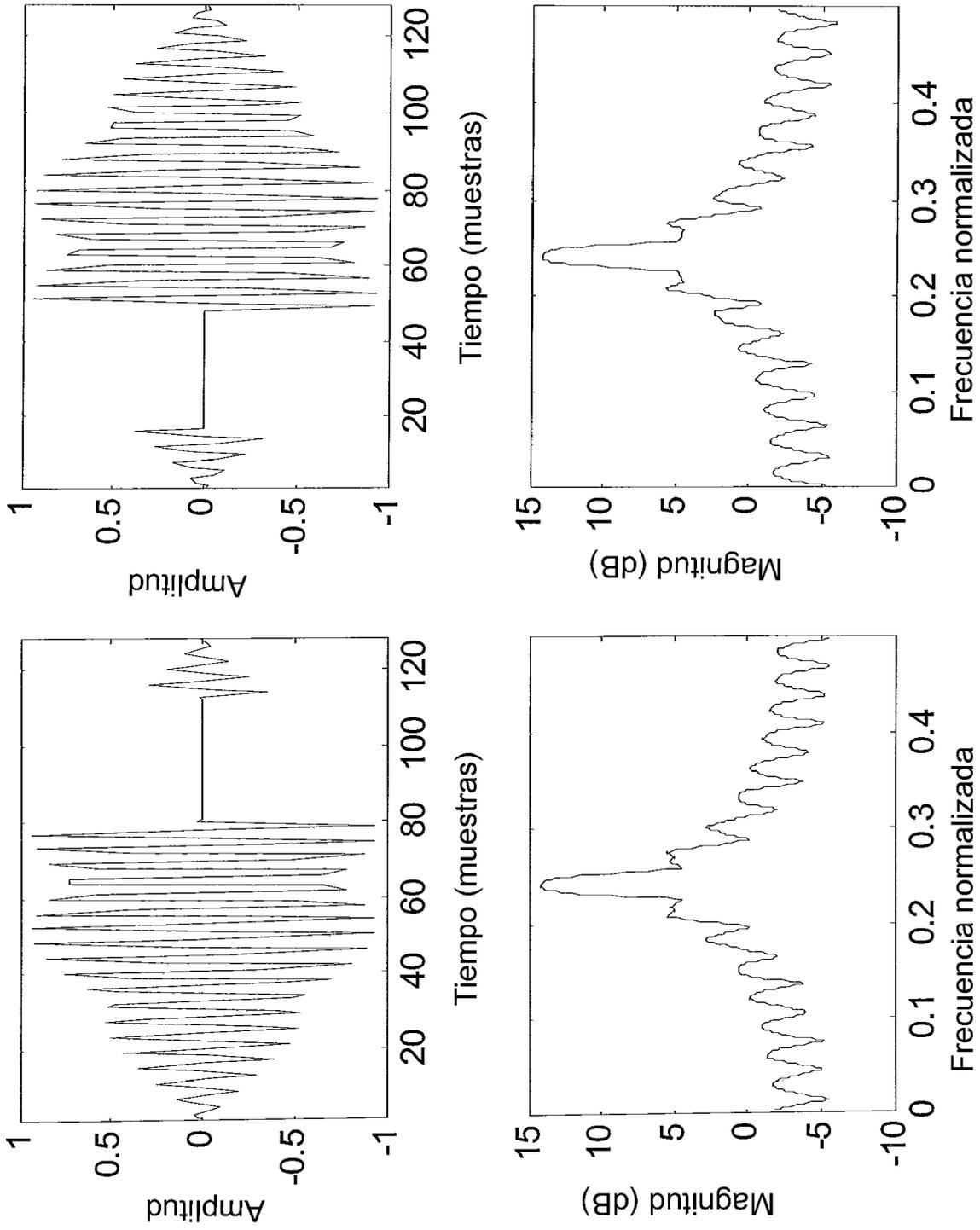


Fig. 12

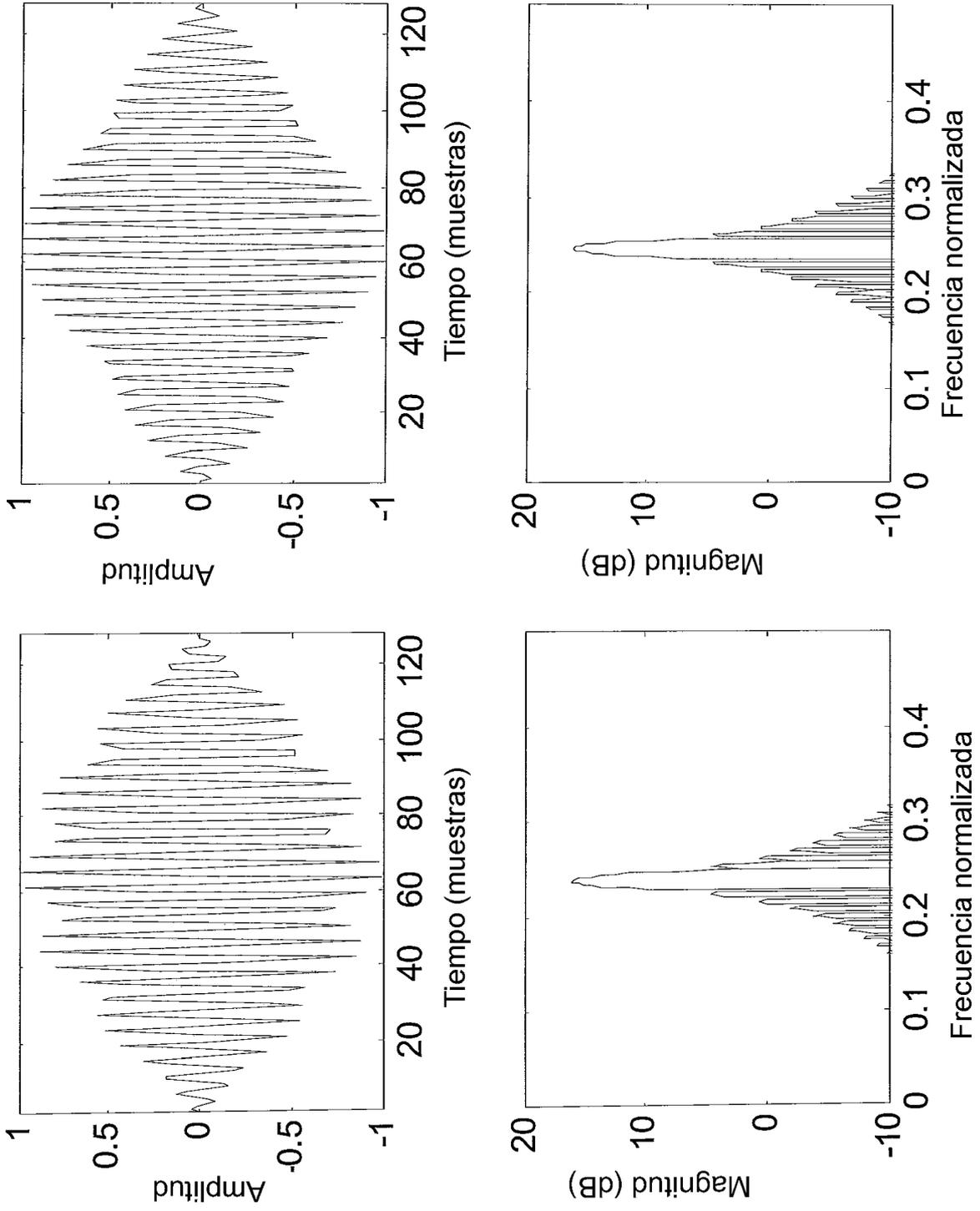
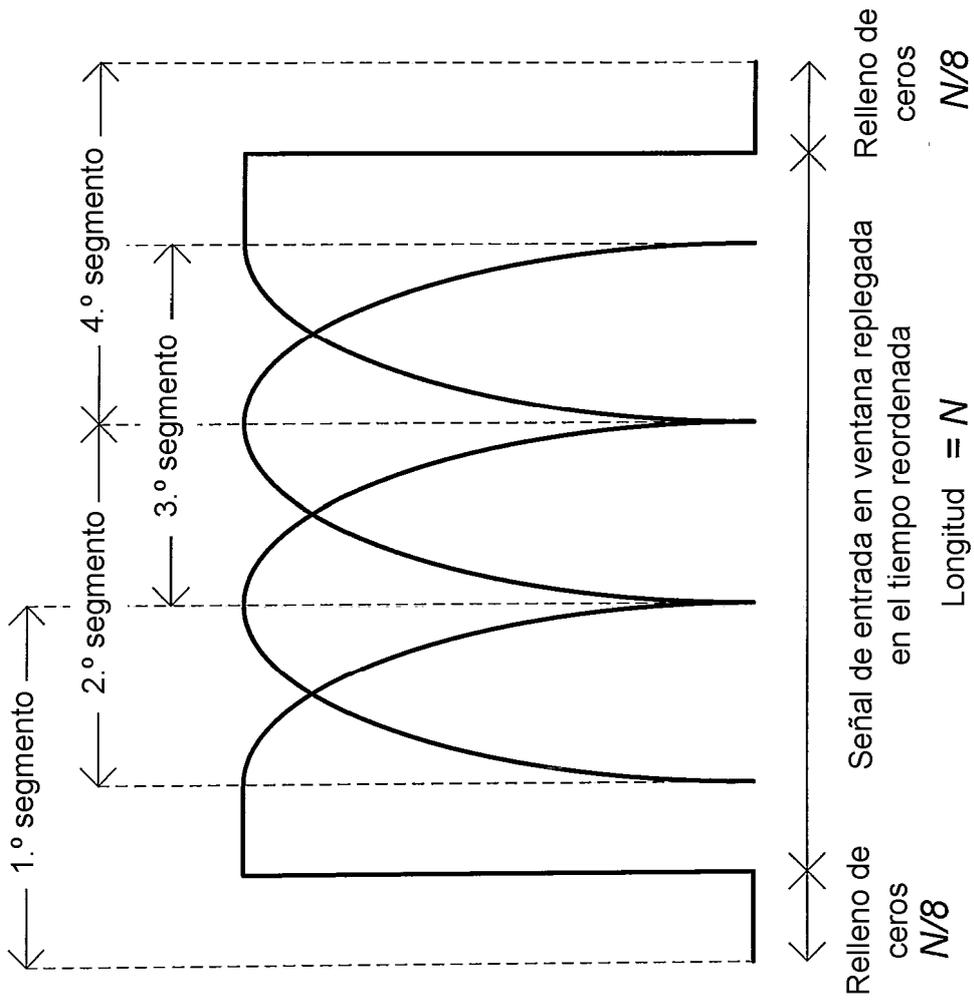


Fig. 13



$$v(n)$$

Fig. 14

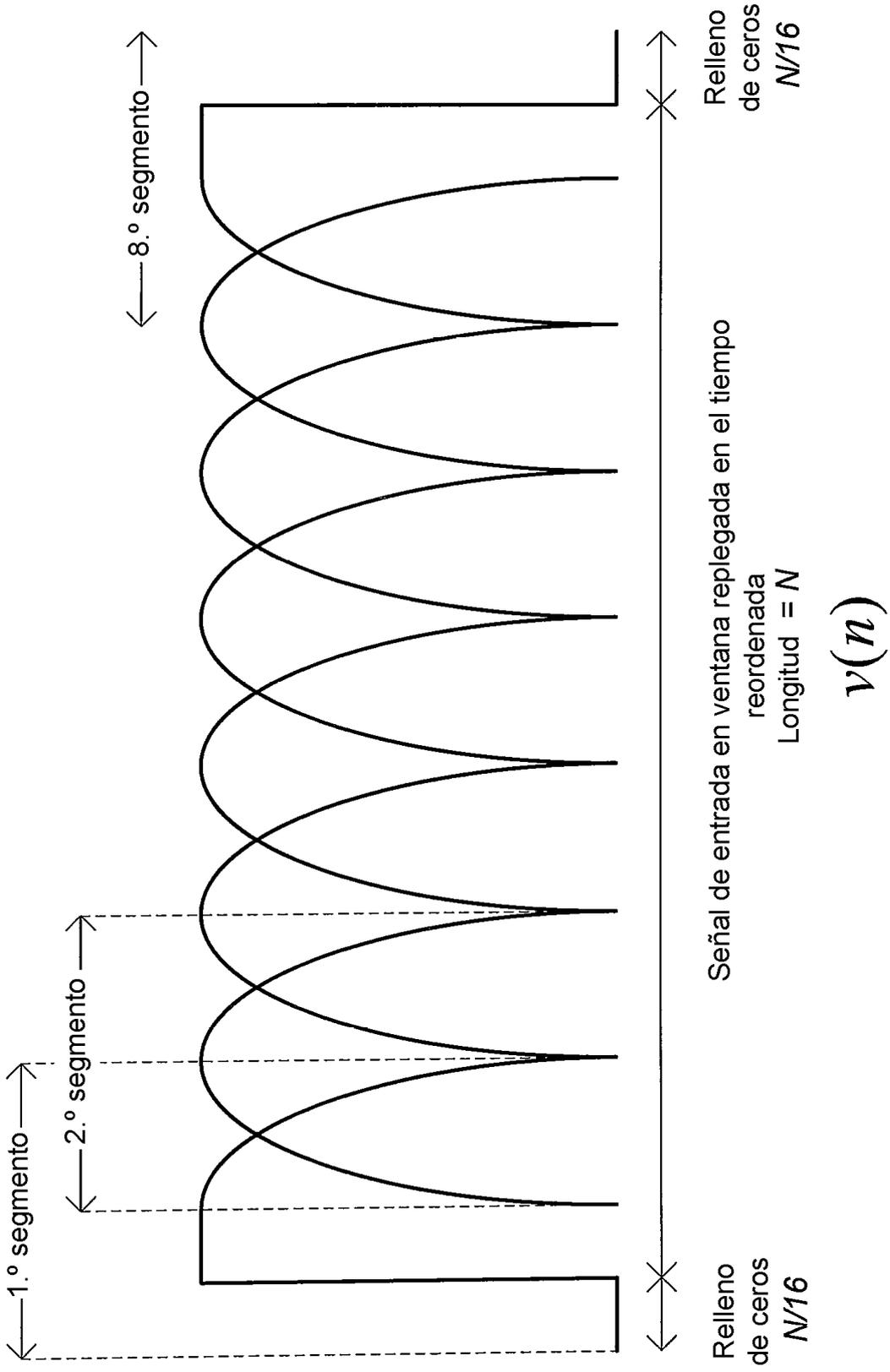


Fig. 15

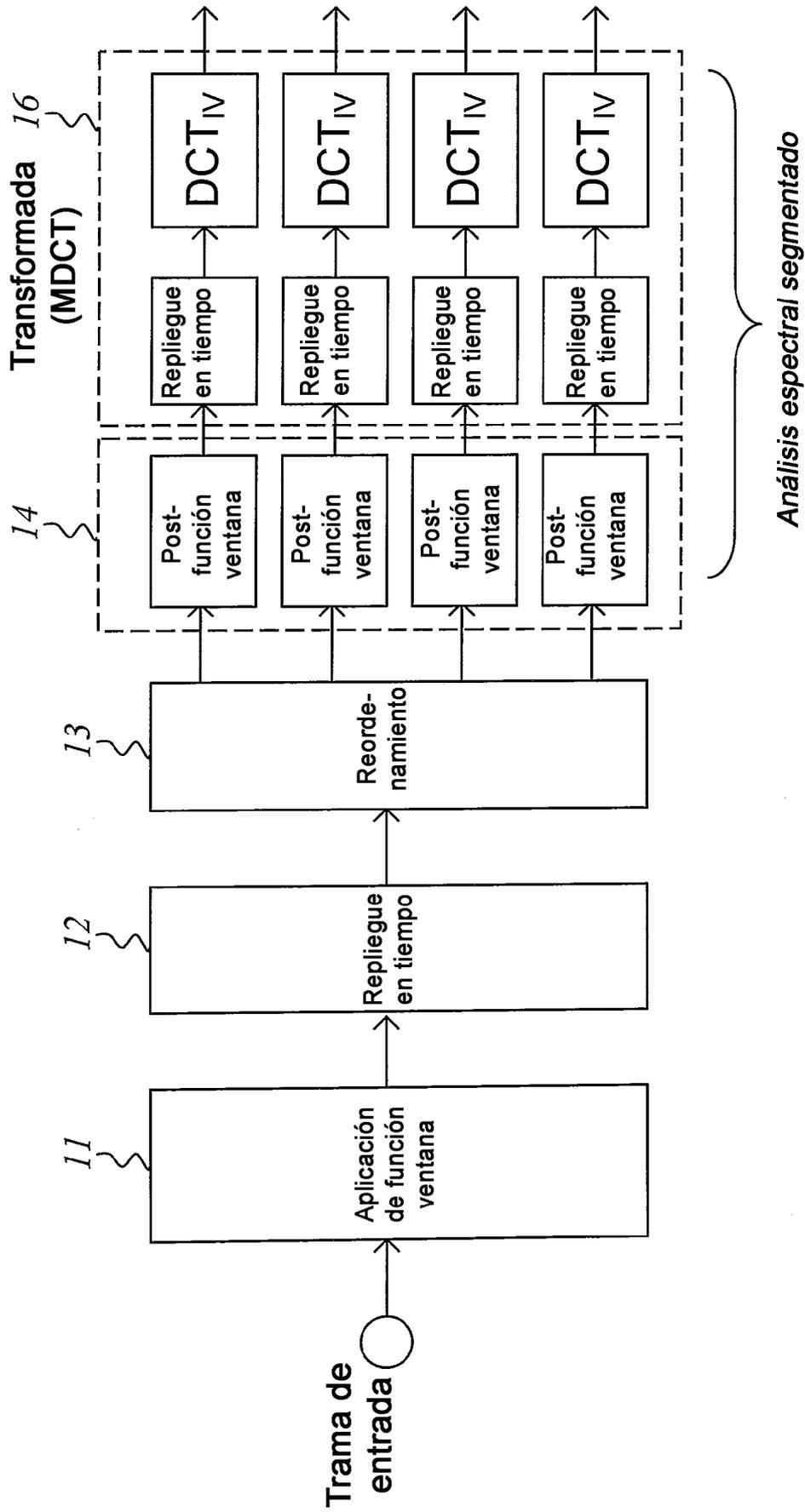


Fig. 16

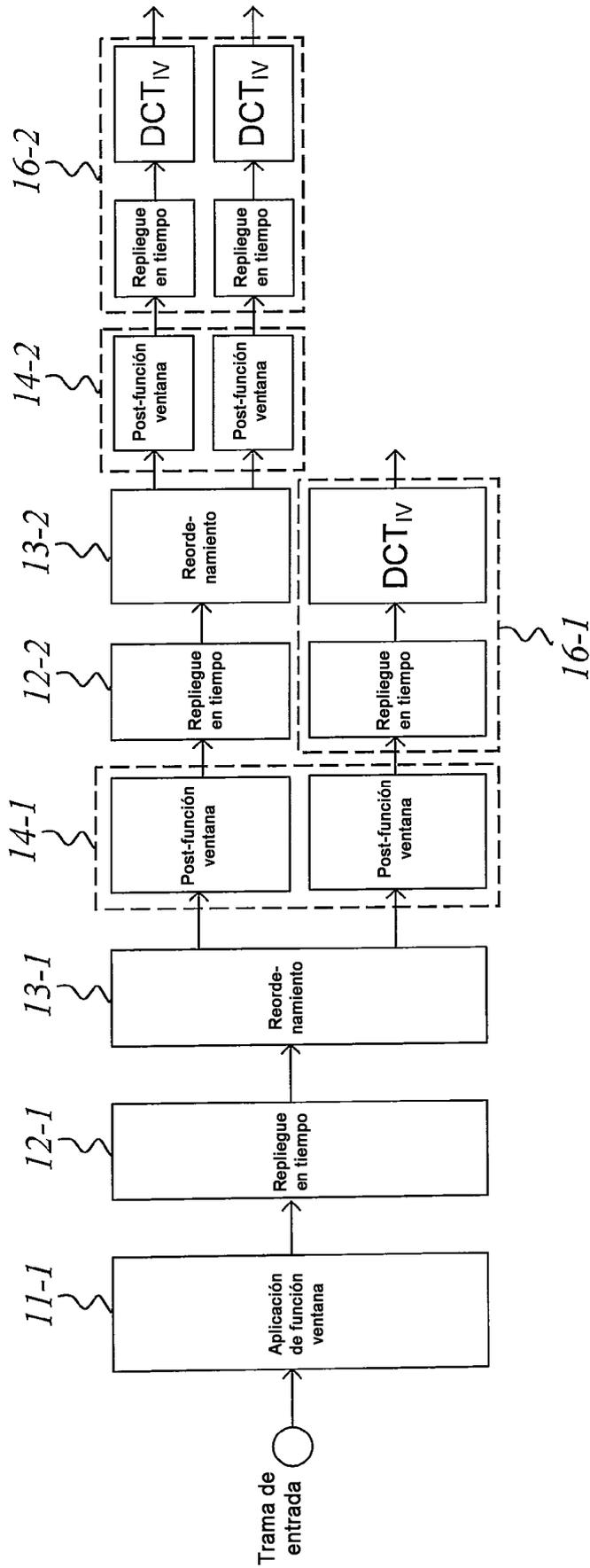


Fig. 17

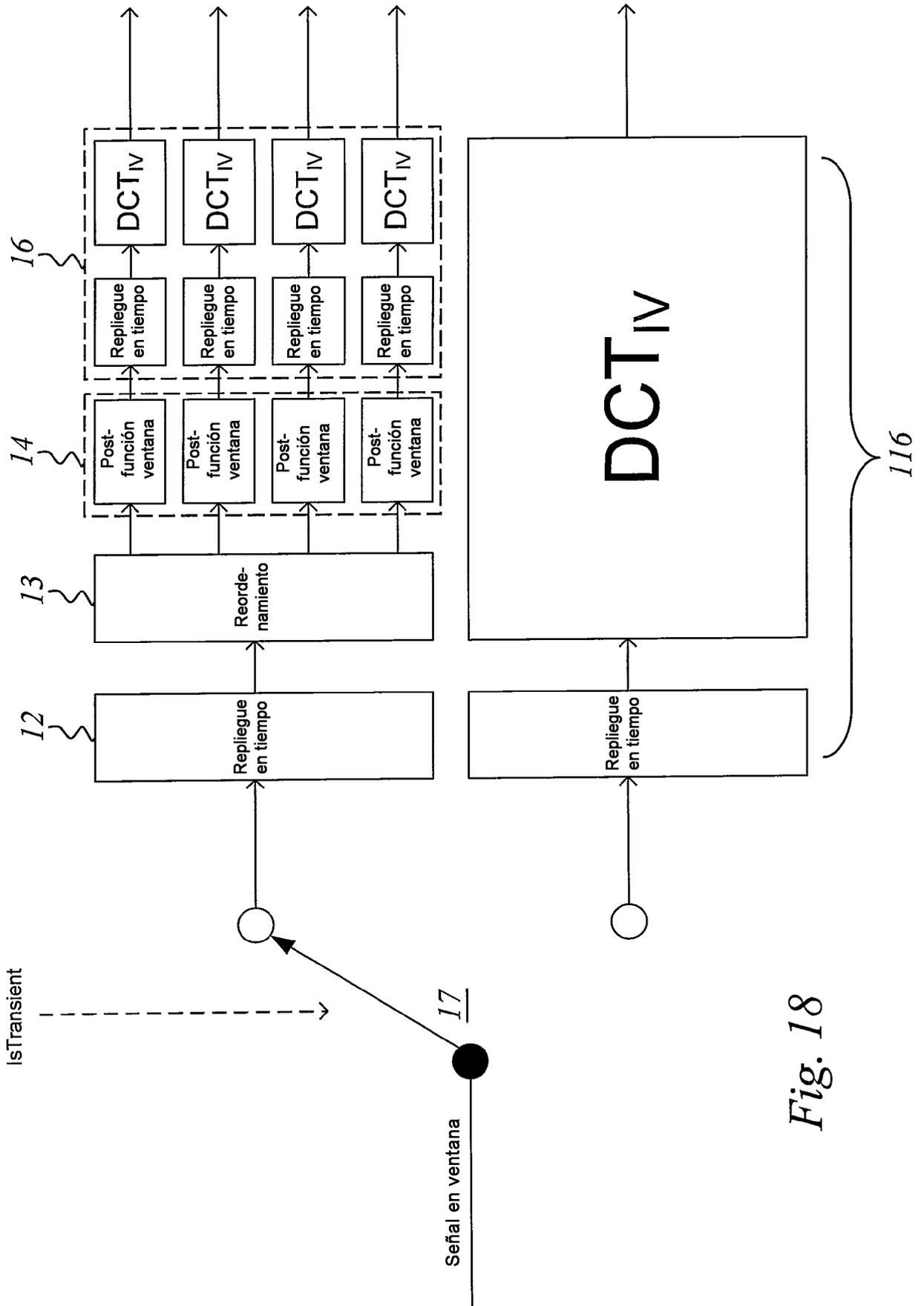


Fig. 18

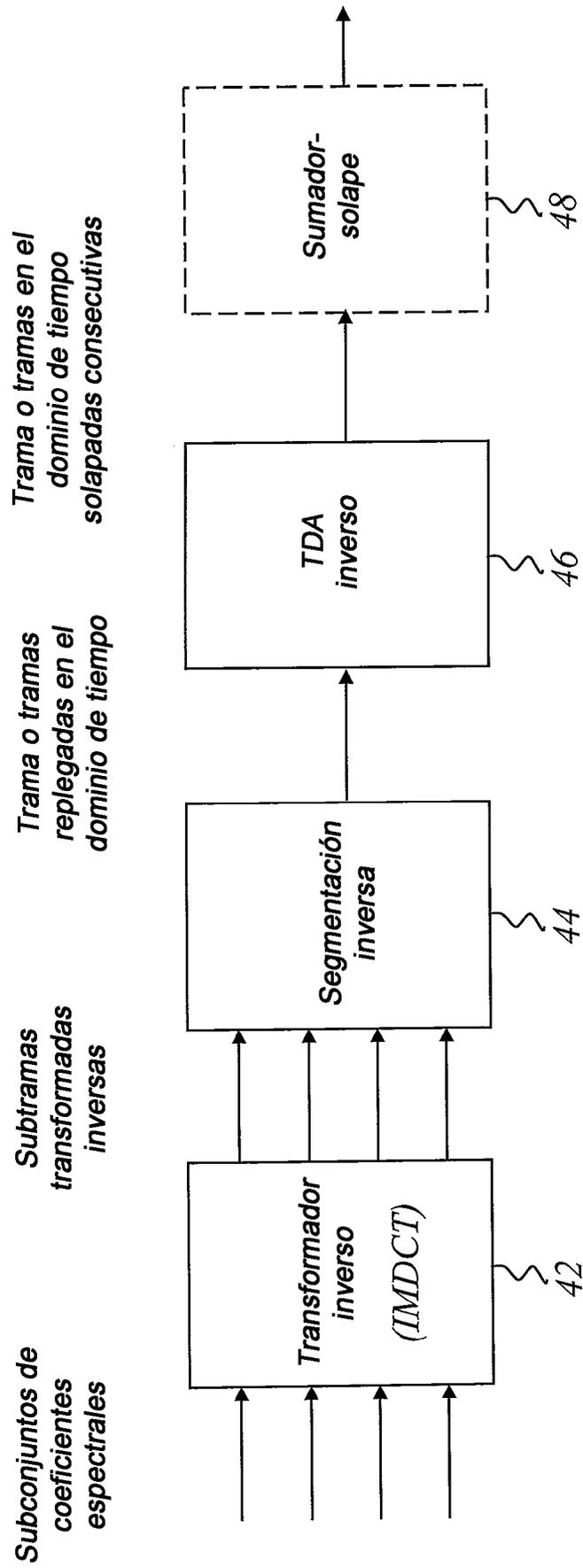


Fig. 19

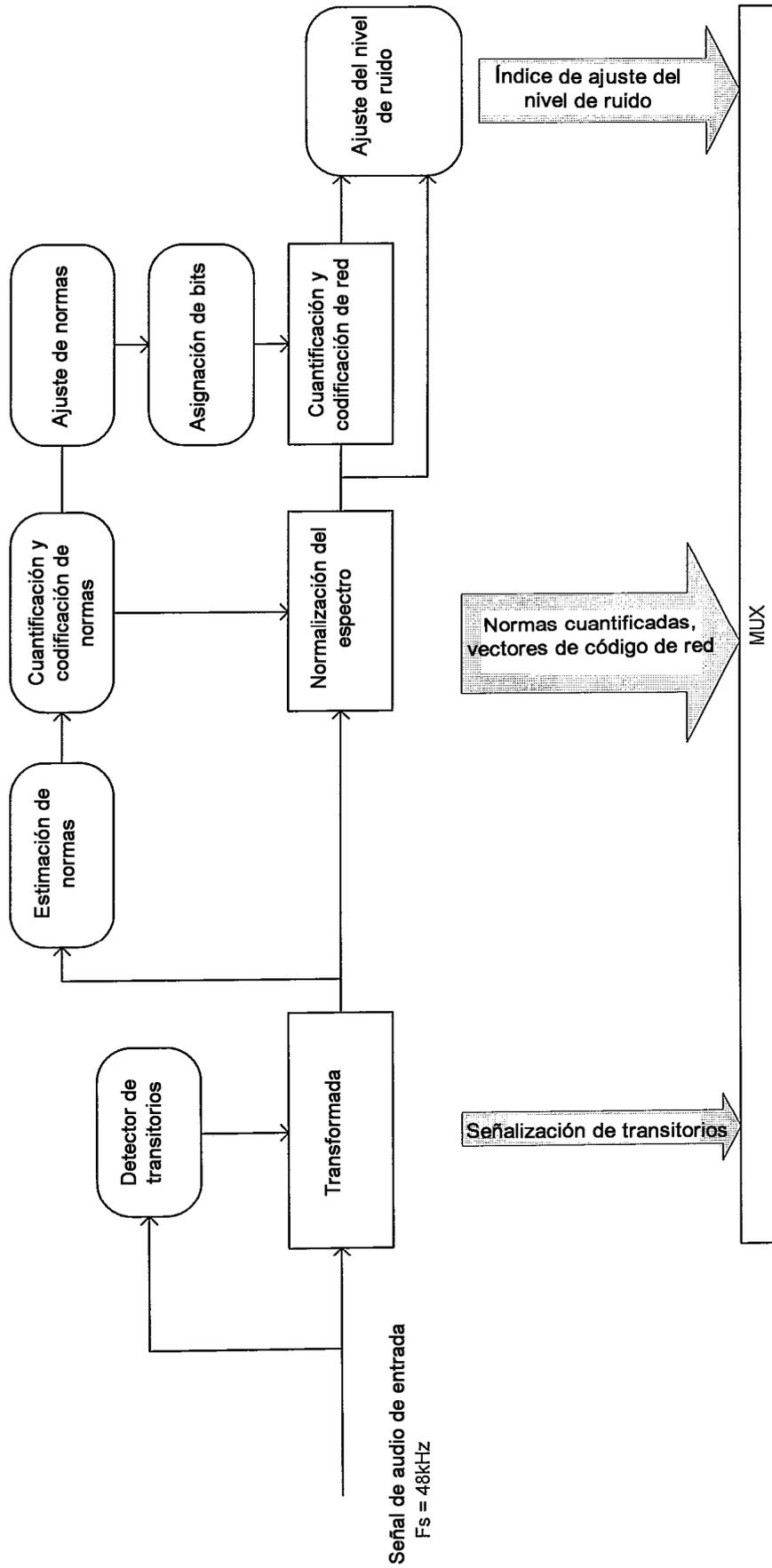


Fig. 20

10

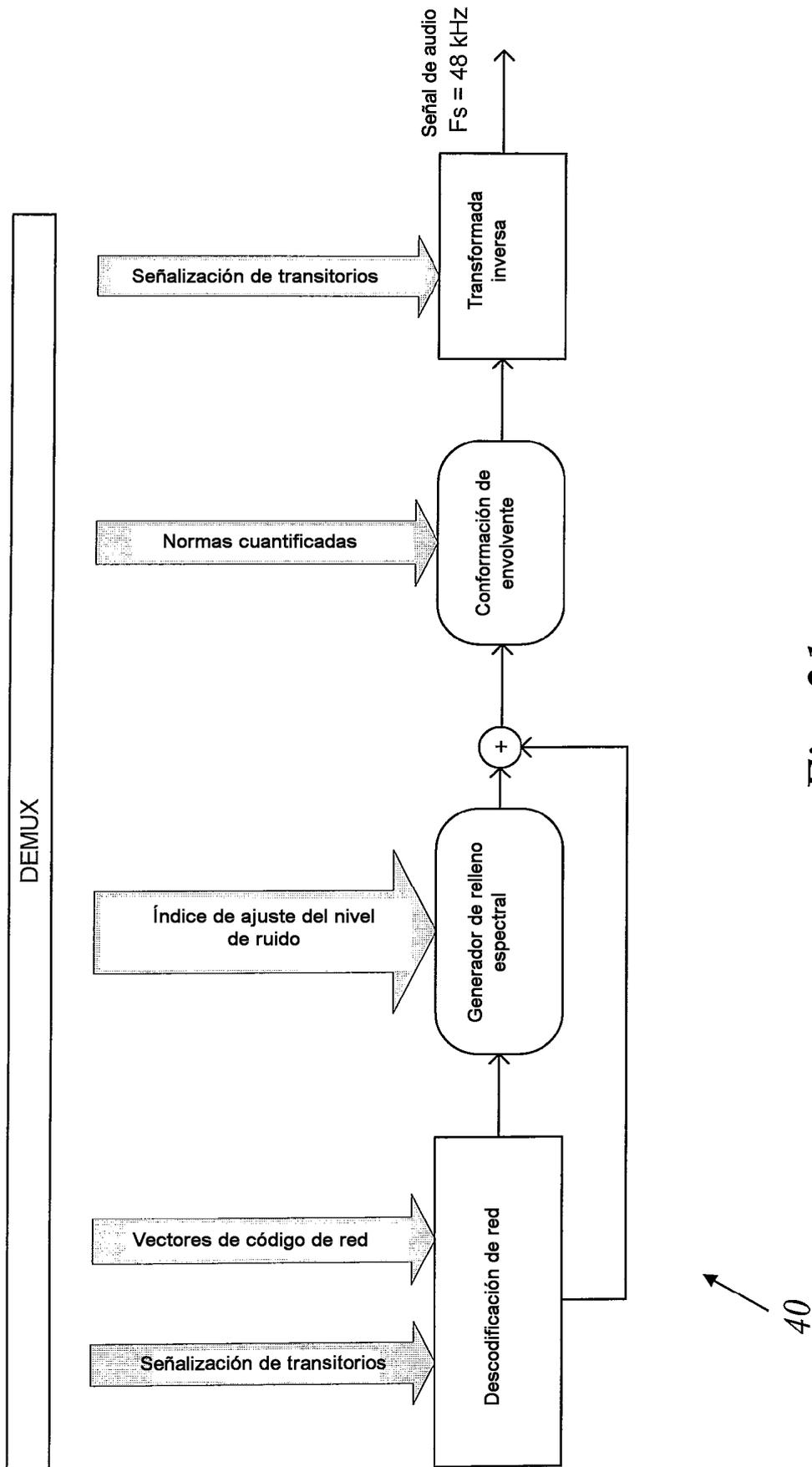


Fig. 21

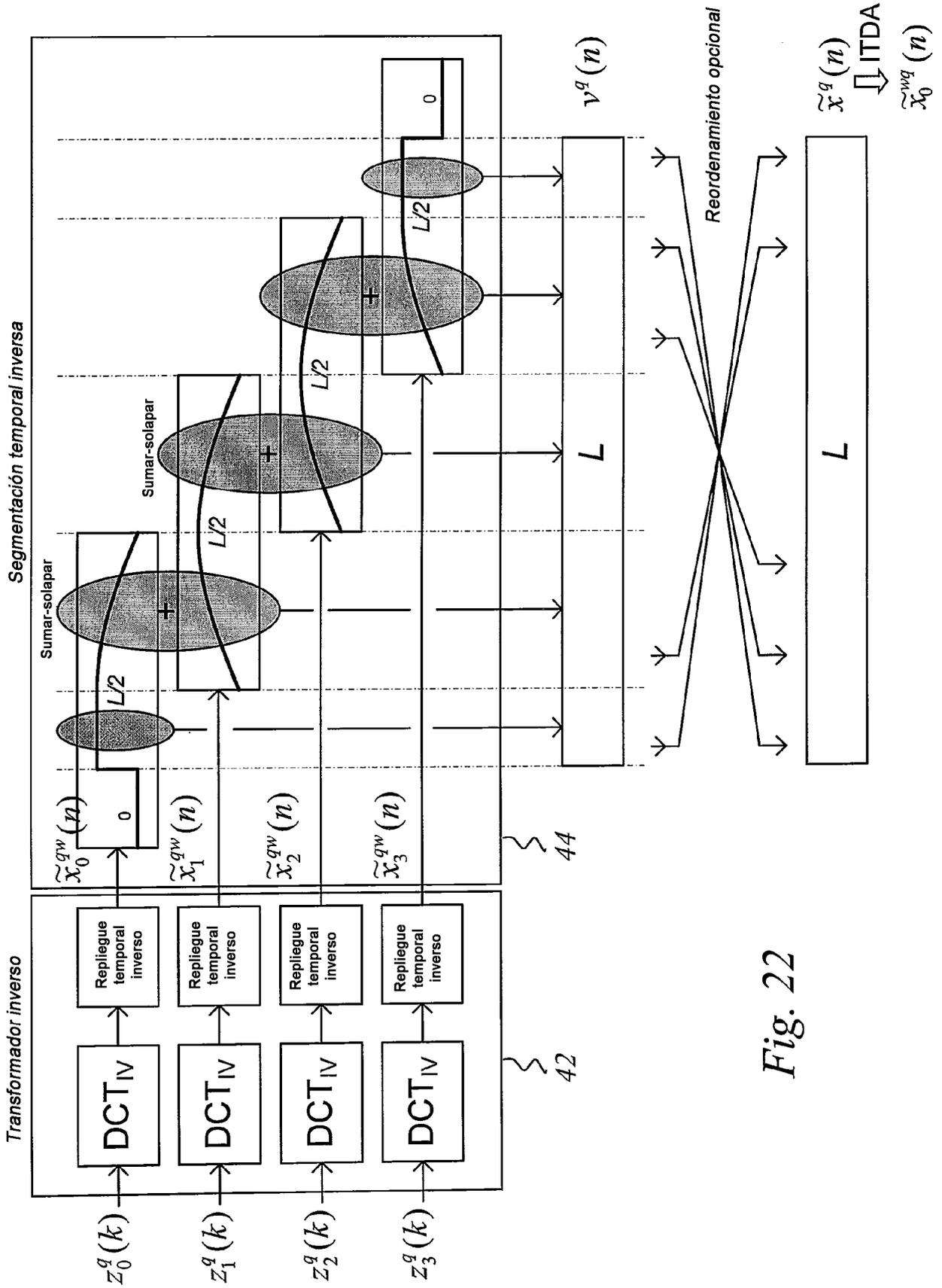


Fig. 22