

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 773 794**

51 Int. Cl.:

G10L 19/008 (2013.01)

G10L 19/02 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **20.01.2017 PCT/EP2017/051214**

87 Fecha y número de publicación internacional: **27.07.2017 WO17125563**

96 Fecha de presentación y número de la solicitud europea: **20.01.2017 E 17700707 (7)**

97 Fecha y número de publicación de la concesión europea: **08.01.2020 EP 3405949**

54 Título: **Aparato y procedimiento para estimar una diferencia de tiempos entre canales**

30 Prioridad:

22.01.2016 EP 16152453

22.01.2016 EP 16152450

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

14.07.2020

73 Titular/es:

**FRAUNHOFER-GESELLSCHAFT ZUR
FÖRDERUNG DER ANGEWANDTEN
FORSCHUNG E.V. (100.0%)
Hansastraße 27c
80686 München , DE**

72 Inventor/es:

**BAYER, STEFAN;
FOTOPOULOU, ELENI;
MULTRUS, MARKUS;
FUCHS, GUILLAUME;
RAVELLI, EMMANUEL;
SCHNELL, MARKUS;
DÖHLA, STEFAN;
JÄGERS, WOLFGANG;
DIETZ, MARTIN y
MARKOVIC, GORAN**

74 Agente/Representante:

SALVÀ FERRER, Joan

ES 2 773 794 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

Aparato y procedimiento para estimar una diferencia de tiempos entre canales

- 5 **[0001]** La presente solicitud se relaciona con el tratamiento estéreo o, en general, con el tratamiento multicanal, en el que una señal multicanal tiene dos canales tales como un canal izquierdo y un canal derecho en el caso de una señal estéreo o más de dos canales, tales como tres, cuatro, cinco o cualquier otro número de canales.
- 10 **[0002]** La voz estéreo y particularmente la voz estéreo conversacional ha recibido mucha menos atención científica que el almacenamiento y la difusión de la música estereofónica. De hecho, la transmisión monofónica todavía es usada principalmente hoy en día en las comunicaciones de voz. No obstante, con el aumento de la anchura de banda y la capacidad de las redes, se prevé que las comunicaciones basadas en tecnologías estereofónicas tendrán mayor difusión y brindarán una mejor experiencia auditiva.
- 15 **[0003]** La codificación eficaz de material de audio estereofónico ha sido estudiada durante largo tiempo en la codificación audio perceptual de la música para el almacenamiento y la difusión eficientes. Con altas velocidades binarias, cuando es fundamental conservar la forma de onda, se ha empleado durante mucho tiempo el estéreo suma-resta, conocido como estéreo M/S (*mid/side*, central/lateral). Con bajas velocidades binarias, se introdujo el estéreo en intensidad y más recientemente la codificación con estéreo paramétrico. Esta última técnica fue adoptada
20 en diferentes estándares tales como HeAACv2 y Mpeg USAC. Genera una submezcla de la señal de dos canales y asocia la información lateral espacial compacta.
- [0004]** La codificación de estéreo conjunto se construye normalmente sobre una resolución de alta frecuencia, es decir, baja resolución de tiempo, transformación tiempo-frecuencia de la señal y entonces no es
25 compatible con el tratamiento con bajo retardo y en el dominio del tiempo que realiza la mayoría de los codificadores de voz. Además, la velocidad binaria generada es usualmente alta.
- [0005]** Por otra parte, el estéreo paramétrico emplea un banco de filtros adicional situado en el extremo delantero del codificador como preprocesador y en el extremo trasero del decodificador como posprocesador. Por lo
30 tanto, el estéreo paramétrico se puede usar con codificadores de voz convencionales como ACELP ya que está hecho en MPEG USAC. Además, la parametrización del escenario auditivo se puede lograr con una cantidad mínima de información lateral, que es adecuada para bajas velocidades binarias. No obstante, el estéreo paramétrico no está específicamente diseñado, por ejemplo, en MPEG USAC, para bajo retardo y no suministra una calidad homogénea en diferentes escenarios conversacionales. En la representación paramétrica convencional de la
35 escena espacial, la anchura de la imagen estéreo se reproduce artificialmente mediante un decorrelacionador aplicado en los dos canales sintetizados y controlado por parámetros IC (*Inter-channel Coherence*, coherencia entre canales) calculados y transmitidos por el codificador. Para la mayoría de la voz estéreo, esta forma de ensanchar la imagen estéreo no es adecuada para recrear el efecto ambiental natural de la voz que es un sonido bastante directo dado que está producido por una única fuente ubicada en una posición específica en el espacio (a veces, con alguna
40 reverberación proveniente de la sala). Por el contrario, los instrumentos musicales tienen mucha más anchura natural que la voz, y se pueden imitar mejor mediante la decorrelación de los canales.
- [0006]** También se presentan inconvenientes cuando la voz se registra con micrófonos no coincidentes, como en la configuración A-B cuando los micrófonos están distantes o en el registro o la representación binaural. Tales
45 escenarios se pueden prever en la captura de voz en teleconferencias o en la creación de un escenario auditivo virtual con hablantes distantes en la MCU (*Multipoint Control Unit*, unidad de control multipunto). El momento de llegada de la señal entonces es distinto entre un canal y el otro a diferencia de los registros realizados en micrófonos coincidentes como X-Y (registro de intensidad) o M-S (registro central-lateral). Posteriormente, el cálculo de la coherencia de esos dos canales no alineados en el tiempo se puede estimar de forma incorrecta lo cual hace
50 fracasar la síntesis artificial del efecto ambiental.
- [0007]** Las referencias de la técnica anterior relacionadas con el tratamiento estéreo son la patente de EE.UU. 5.434.948 o la patente de EE.UU. 8.811.621. Otra técnica anterior se conoce a partir del documento WO2012/105886A1.
- 55 **[0008]** El documento WO 2006/089570 A1 describe un esquema de codificador/decodificador multicanal transparente o casi transparente. Un esquema de codificador/decodificador multicanal genera además una señal residual del tipo forma de onda. Esta señal residual se transmite junto con uno o más parámetros multicanal a un decodificador. Al contrario que un decodificador multicanal puramente paramétrico, el decodificador mejorado genera
60 una señal de salida multicanal que tiene una calidad de salida mejorada debido a la señal residual adicional. En el lado del codificador, tanto un canal izquierdo como un canal derecho son filtrados por un banco de filtros de análisis. A continuación, para cada señal de subbanda, se calcula un valor de alineación y un valor de ganancia para una subbanda. Dicha alineación se realiza entonces antes del tratamiento adicional. En el lado del decodificador, se

realiza una desalineación y un tratamiento de ganancia y a continuación las señales correspondientes son sintetizadas por un banco de filtros de síntesis con el fin de generar una señal izquierda decodificada y una señal derecha decodificada.

5 **[0009]** En dichas aplicaciones de tratamiento estéreo, el cálculo de una diferencia entre canales o una diferencia de tiempos entre canales entre una señal de un primer canal y una señal de un segundo canal resulta útil para realizar normalmente un procedimiento de alineación en el tiempo de banda ancha. No obstante, sí existen otras aplicaciones para el uso de una diferencia de tiempos entre canales entre un primer canal y un segundo canal, en el que estas aplicaciones están en el almacenamiento o la transmisión de datos paramétricos, tratamiento
10 estéreo/multicanal que comprende una alineación en el tiempo de dos canales, una diferencia de tiempos de estimación de llegada para determinar la posición de un hablante en una sala, filtrado espacial formador de haces, descomposición en primer plano/segundo plano o la ubicación de una fuente de sonido mediante, por ejemplo, triangulación acústica, por mencionar solo algunas.

15 **[0010]** En todas estas aplicaciones, es necesaria una determinación eficiente, exacta y robusta de una diferencia de tiempos entre canales entre una señal de un primer canal y de un segundo canal.

[0011] En realidad, existen estas determinaciones conocidas con el nombre de "GCC-PHAT" o, dicho de otra forma, transformada de fase con correlación cruzada generalizada. Normalmente, se calcula un espectro de
20 correlación cruzada entre las señales de los dos canales y, a continuación, se aplica una función de ponderación al espectro de correlación cruzada para obtener el llamado espectro de correlación cruzada generalizada antes de realizar una transformada espectral inversa tal como una DFT inversa en el espectro de correlación cruzada generalizada con el fin de encontrar una representación en el dominio del tiempo. Esta representación en el dominio del tiempo representa valores para ciertos retardos de tiempo y el pico más alto de la representación en el dominio
25 del tiempo corresponde entonces normalmente al retardo o a la diferencia de tiempos, es decir, el retardo de tiempo entre canales de la diferencia entre las señales de los dos canales.

[0012] No obstante, se ha mostrado que, sobre todo en señales que son diferentes, por ejemplo, de la voz
30 limpia sin ninguna reverberación ni ruido de fondo, la robustez de esta técnica general no resulta óptima.

[0013] Por lo tanto, un objeto de la presente invención es proporcionar un concepto mejorado para estimar una diferencia de tiempos entre canales entre las señales de dos canales.

[0014] Este objeto se logra mediante un aparato para estimar una diferencia de tiempos entre canales según
35 la reivindicación 1, o un procedimiento para estimar una diferencia de tiempos entre canales según la reivindicación 15 o un programa informático según la reivindicación 16.

[0015] La presente invención se basa en el hallazgo de que el suavizado del espectro de correlación cruzada a lo largo del tiempo que está controlado mediante una característica espectral del espectro de la señal del primer
40 canal o la señal del segundo canal mejora significativamente la robustez y la exactitud de la determinación de la diferencia de tiempos entre canales.

[0016] En realizaciones preferidas, se determina una característica de tonalidad/ruido del espectro, y en el caso de una señal del tipo tono, el suavizado es más intenso, mientras que, en el caso de una señal con ruido, el
45 suavizado se hace menos intenso.

[0017] Preferentemente, se usa una medida de uniformidad espectral, y en el caso de señales del tipo tono, la medida de uniformidad espectral será baja y el suavizado se volverá más intenso, y en el caso de señales del tipo ruido, la medida de uniformidad espectral será alta tal como aproximadamente 1 o próxima a 1 y el suavizado será
50 débil.

[0018] Por lo tanto, según la presente invención, un aparato para estimar una diferencia de tiempos entre canales entre una señal del primer canal y una señal del segundo canal comprende un calculador para calcular un espectro de correlación cruzada para un bloque de tiempo para la señal del primer canal en el bloque de tiempo y la
55 señal del segundo canal en el bloque de tiempo. El aparato comprende además un estimador de características espectrales para estimar una característica de un espectro de la señal del primer canal y la señal del segundo canal para el bloque de tiempo, y, adicionalmente, un filtro de suavizado para suavizar el espectro de correlación cruzada a lo largo del tiempo usando la característica espectral para obtener un espectro de correlación cruzada suavizado. A continuación, el espectro de correlación cruzada suavizado se trata adicionalmente mediante un procesador con el
60 fin de obtener el parámetro de diferencia de tiempos entre canales.

[0019] Para realizaciones preferidas relacionadas con el tratamiento adicional del espectro de correlación cruzada suavizado, se realiza una operación adaptativa de fijación de umbral, en la cual se analiza la representación

en el dominio del tiempo del espectro de correlación cruzada generalizado suavizado con el fin de determinar un umbral variable, que depende de la representación en el dominio del tiempo y se compara un pico de la representación en el dominio del tiempo con el umbral variable, en el que se determina una diferencia de tiempos entre canales como un retardo asociado con un pico que está en una relación predeterminada con el umbral como, por ejemplo, mayor que el umbral.

[0020] En una realización, el umbral variable se determina como un valor que es igual a un múltiplo entero de un valor entre el mayor, por ejemplo, diez por ciento de los valores de la representación en el dominio del tiempo o, alternativamente, en una realización adicional para la determinación variable, el umbral variable se calcula mediante la multiplicación del umbral variable y el valor, en el que el valor depende de una característica de la relación señal-ruido de las señales del primer canal y del segundo, en el que el valor se vuelve más alto para una relación señal-ruido más alta y se vuelve más bajo para una relación señal-ruido más baja.

[0021] Como se indicó anteriormente, el cálculo de la diferencia de tiempos entre canales se puede usar en muchas aplicaciones diferentes tales como el almacenamiento o la transmisión de datos paramétricos, el tratamiento/codificación estéreo/multicanal, la alineación en el tiempo de dos canales, la estimación de la diferencia de tiempo de llegada para determinar la posición de un altavoz en una sala con dos micrófonos y una configuración de micrófonos conocida, con el fin de formación de haces, filtrado espacial, descomposición en primer plano/segundo plano o la determinación de la ubicación de una segunda fuente, por ejemplo, mediante triangulación acústica sobre la base de diferencias de tiempos de dos o tres señales.

[0022] Sin embargo, a continuación, se describe una implementación y uso preferidos del cálculo de la diferencia de tiempos entre canales con fines de alineación en el tiempo de banda ancha de dos señales estéreo en un proceso de codificación de una señal multicanal que tiene al menos dos canales.

[0023] Un aparato para codificar una señal multicanal que tiene al menos dos canales comprende un determinador de parámetros para determinar un parámetro de alineación de banda ancha, por una parte, y una pluralidad de parámetros de alineación de banda estrecha, por otra. Estos parámetros son usados por un alineador de señales para alinear los al menos dos canales usando estos parámetros para obtener canales alineados. A continuación, un procesador de señal calcula una señal central y una señal lateral usando los canales alineados y la señal central y la señal lateral son después codificadas y enviadas en una señal de salida codificada que además tiene, como información lateral paramétrica, el parámetro de alineación de banda ancha y la pluralidad de parámetros de alineación de banda estrecha.

[0024] En el lado del decodificador, un decodificador de señal decodifica la señal central codificada y la señal lateral codificada para obtener una señal central y una señal lateral decodificada. Estas señales son tratadas a continuación por un procesador de señales para calcular un primer canal decodificado y un segundo canal decodificado. Estos canales decodificados se desalinean a continuación usando la información sobre el parámetro de alineación de banda ancha y la información sobre la pluralidad de parámetros de banda estrecha incluidos en una señal multicanal codificada para obtener la señal multicanal decodificada.

[0025] En una implementación específica, el parámetro de alineación de banda ancha es un parámetro de diferencia de tiempos entre canales y la pluralidad de parámetros de alineación de banda estrecha son diferencias de fases entre canales.

[0026] La presente invención se basa en el hallazgo de que específicamente para señales de voz en las que hay más de un hablante, pero también para otras señales de audio en las que hay varias fuentes de audio, los diferentes lugares de las fuentes de audio que establecen correspondencias en dos canales de la señal multicanal se pueden tener en cuenta para usar un parámetro de alineación de banda ancha tal como un parámetro de diferencia de tiempos entre canales que se aplica a la totalidad del espectro ya sea de uno o de los dos canales. Además de este parámetro de alineación de banda ancha, se ha descubierto que varios parámetros de alineación de banda estrecha que difieren de subbanda a subbanda producen además una mejor alineación de la señal en los dos canales.

[0027] Por lo tanto, una alineación de banda ancha correspondiente al mismo retardo en cada subbanda junto con una alineación de fase correspondiente a diferentes rotaciones de fase para diferentes subbandas produce una alineación óptima de ambos canales antes de que estos dos canales se conviertan seguidamente en una representación central/lateral que a continuación es codificada adicionalmente. Debido al hecho de que se ha obtenido una alineación óptima, la energía en la señal central es lo más alta posible, por una parte, y, por otra parte, la energía en la señal lateral es lo más pequeña posible, de modo que se puede obtener un resultado óptimo de codificación con una velocidad binaria lo más baja posible o una calidad de audio lo más alta posible para una determinada velocidad binaria.

- [0028]** Específicamente para un material de voz conversacional, normalmente parece que existen hablantes activos en dos lugares diferentes. Además, la situación es tal que, normalmente, solo un hablante está hablando desde el primer lugar y a continuación el segundo hablante está hablando desde el segundo lugar o localización. La influencia de las diferentes localizaciones en los dos canales tales como un primer canal o canal izquierdo y un segundo canal o canal derecho se refleja mediante diferentes tiempos de llegada y, por lo tanto, un cierto retardo entre ambos canales debido a las diferentes localizaciones, y este retardo cambia cada cierto tiempo. Generalmente, esta influencia se refleja en las señales de los dos canales como una desalineación de banda ancha que se puede corregir mediante el parámetro de alineación de banda ancha.
- 5
- [0029]** Por otra parte, se puede dar cuenta de otros efectos, particularmente provenientes de la reverberación o fuentes adicionales de ruido, mediante parámetros individuales de alineación de fase para bandas individuales que están superpuestas en los diferentes tiempos de llegada de banda ancha o la desalineación de banda ancha de ambos canales.
- 10
- [0030]** Habida cuenta de esto, el uso de ambos, un parámetro de alineación de banda ancha y una pluralidad de parámetros de alineación de banda estrecha además del parámetro de alineación de banda ancha producen una alineación de canales óptima en el lado del codificador para obtener una buena y muy compacta representación central/lateral mientras que, por otra parte, la desalineación correspondiente posterior a una decodificación en el lado del decodificador produce una buena calidad de audio para una determinada velocidad binaria o en una pequeña velocidad binaria para una determinada calidad de audio requerida.
- 15
- [0031]** Una ventaja de la presente invención es que proporciona un nuevo esquema de codificación estéreo mucho más adecuado para una conversión de la voz estéreo que los esquemas de codificación estéreo existentes. Según la invención, las tecnologías de estéreo paramétrico y las tecnologías de codificación de estéreo conjunto se combinan de forma particular al aprovechar la diferencia de tiempos entre canales que se produce en los canales de una señal multicanal específicamente en el caso de fuentes de voz, pero también en el caso de otras fuentes de audio.
- 20
- [0032]** Diversas realizaciones proporcionan ventajas útiles, tal como se expondrá más adelante.
- 25
- [0033]** El nuevo procedimiento es un enfoque híbrido que mezcla elementos de un estéreo M/S convencional y un estéreo paramétrico. En un M/S convencional, los canales se submezclan pasivamente para generar una señal central y una lateral. El proceso se puede extender adicionalmente al rotar el canal usando una transformada de Karhunen-Loeve (KLT), también conocida como PCA (*Principal Component Analysis*, análisis de componentes principales) antes de sumar y restar los canales. La señal central está codificada en una codificación de código primario mientras que la lateral se transmite a un codificador secundario. El estéreo M/S desarrollado puede usar además la predicción de la señal lateral por el canal central codificado en la trama presente o en la anterior. El objetivo principal de la rotación y la predicción es maximizar la energía de la señal central al mismo tiempo que se minimiza la energía de la lateral. El estéreo M/S preserva la forma de onda y en este aspecto es muy robusto ante cualquier escenario estéreo, pero puede ser muy costoso en términos de consumo de bits.
- 30
- [0034]** Para una máxima eficiencia con baja velocidad binaria, el estéreo paramétrico calcula y codifica parámetros, como ILD (*Inter-channel Level Differences*, diferencias de niveles entre canales), IPD (*Inter-channel Phase Differences*, diferencias de fases entre canales), ITD (*Inter-channel Time Differences*, diferencias de tiempos entre canales) e IC (*Inter-channel Coherence*, coherencia entre canales). Estos representan de manera compacta la imagen estéreo y son pistas del escenario auditivo (localización de la fuente, panorámica, anchura del estéreo, etc.). Entonces el objetivo es parametrizar el escenario estéreo y codificar solo una señal de submezcla que puede estar en el decodificador y, con la ayuda de las pistas estéreo transmitidas, espacializarla nuevamente.
- 35
- [0035]** El enfoque de los autores de la invención mezcló los dos conceptos. En primer lugar, las pistas ITD e IPD se calculan y aplican a los dos canales. El objetivo es representar la diferencia de tiempo en banda ancha y la diferencia de fases en bandas con diferente frecuencia. A continuación, los dos canales se alinean en tiempo y fase y después se realiza la codificación M/S. Se encontró que las ITD e IPD fueron útiles para modelizar la voz estéreo y son una buena alternativa a la rotación basada en KLT en M/S. A diferencia de una codificación puramente paramétrica, el efecto ambiental no es ya modelizado por las IC sino directamente por la señal lateral que se codifica y/o se predice. Se encontró que este enfoque es más robusto especialmente cuando se manejan señales de voz.
- 40
- [0036]** El cálculo y el tratamiento de las ITD es una parte fundamental de la invención. Las ITD ya fueron aprovechadas en la técnica anterior de BCC (*Binaural Cue Coding*, codificación de referencia binaural), pero de una forma que resultaba ineficiente ya que las ITD cambian a lo largo del tiempo. Para evitar este inconveniente, se diseñó una división en ventanas específica para suavizar las transiciones entre dos ITD diferentes y poder cambiar sin interrupciones entre un hablante y otro ubicados en lugares diferentes.
- 45
- 50
- 55
- 60

[0037] Realizaciones adicionales están relacionadas con el procedimiento en el que, en el lado del codificador, la determinación de parámetros para determinar la pluralidad de parámetros de alineación de banda estrecha se realiza usando canales que ya han sido alineados con el parámetro de alineación de banda ancha determinado anteriormente.

5

[0038] De manera correspondiente, la desalineación de banda estrecha en el lado del decodificador se realiza antes de llevar a cabo la desalineación de banda ancha usando el normalmente único parámetro de alineación de banda ancha.

10 **[0039]** En realizaciones adicionales, se prefiere que, ya sea en el lado del codificador, pero aún más importante, en el lado del decodificador, se realice alguna clase de operación de división en ventanas y superposición-suma o alguna clase de desvanecimiento cruzado desde un bloque al siguiente después de todas las alineaciones y, específicamente, después de una alineación en el tiempo usando el parámetro de alineación de banda ancha. Así se evita cualquier defecto audible tal como chasquidos cuando el parámetro de alineación en el
15 tiempo o de banda ancha cambia de un bloque a otro.

[0040] En otras realizaciones se aplican diferentes resoluciones espectrales. En particular, las señales de los canales se someten a una conversión espectral en el tiempo con una resolución de alta frecuencia tal como un espectro DFT mientras que parámetros tales como los parámetros de alineación de banda estrecha se determinan
20 para bandas de parámetros que tienen una resolución espectral más baja. Normalmente, una banda de parámetros tiene más de una línea espectral que el espectro de la señal y normalmente tiene un conjunto de líneas espectrales del espectro DFT. Asimismo, las bandas de parámetros aumentan desde bajas frecuencias hasta altas frecuencias con el fin de dar cuenta de problemas psicoacústicos.

25 **[0041]** Realizaciones adicionales se relacionan con el uso adicional de un parámetro de nivel tal como una diferencia entre niveles u otros procedimientos para tratar la señal lateral tales como parámetros de relleno estéreo, etc. La señal codificada se puede representar mediante la señal lateral real misma, o mediante una señal residual de predicción que se realiza usando la señal central de la trama actual o cualquier otra trama, o mediante una señal lateral o una señal residual de predicción lateral solo en un subconjunto de bandas y parámetros de predicción solo
30 para las bandas restantes, o incluso mediante parámetros de predicción para todas las bandas sin ninguna información de señal lateral de resolución de alta frecuencia. Por lo tanto, en la última alternativa mencionada, la señal lateral codificada solo es representada por un parámetro de predicción para cada banda de parámetros o solo un subconjunto de bandas de parámetro de modo que para las bandas de parámetros restantes no existe ninguna información sobre la señal lateral original.

35

[0042] Asimismo, es preferible tener la pluralidad de parámetros de alineación de banda estrecha de modo que no todas las bandas de parámetros reflejen la anchura de banda completa de la señal de banda ancha sino solo un conjunto de bandas bajas tal como el 50 por ciento más bajo de las bandas de parámetros. Por otra parte, los parámetros de relleno estéreo no se usan para el par de bandas bajas, dado que, para estas bandas, se transmite la
40 propia señal lateral o una señal residual de predicción con el fin de asegurar que, al menos para las bandas bajas, esté disponible una representación correcta de forma de onda. Por otra parte, la señal lateral no se transmite en una representación exacta de forma de onda para las bandas altas con el fin de disminuir todavía más la velocidad binaria, sino que la señal lateral se representa normalmente mediante parámetros de relleno estéreo.

45 **[0043]** Asimismo, es preferible realizar el análisis de parámetros y la alineación completos dentro de uno y el mismo dominio de la frecuencia sobre la base del mismo espectro DFT. Para este fin es preferible además usar la tecnología GCC-PHAT con el objeto de determinar la diferencia de tiempos entre canales. En una realización preferida de este procedimiento, el suavizado de un espectro de correlación sobre la base de una información sobre una forma espectral, siendo la información preferentemente una medida de uniformidad espectral, se realiza de
50 forma que el suavizado será débil en el caso de señales del tipo ruido y el suavizado será más marcado en el caso de señales del tipo tono.

[0044] Asimismo, es preferible realizar una rotación de fase especial, en la que se tengan en cuenta las amplitudes de los canales. En particular, la rotación de fase se distribuye entre los dos canales con fines de
55 alineación en el lado del codificador, y naturalmente, con fines de desalineación en el lado del decodificador, en el que un canal que tiene una amplitud más alta se considera un canal principal y resultará menos afectado por la rotación de fase, es decir, se rotará menos que un canal con una amplitud más baja.

[0045] Asimismo, el cálculo de suma-resta se realiza usando un cambio de escala de la energía con un factor
60 de escala que se obtiene de las energías de los dos canales y además está limitado a un cierto intervalo con el fin de asegurar que el cálculo central/lateral no esté afectando demasiado a la energía. Por otra parte, no obstante, cabe destacar que, con fines de la presente invención, esta clase de conservación de la energía no es tan crucial como en los procedimientos de la técnica anterior, dado que el tiempo y la fase fueron alineados de antemano. Por

lo tanto, las fluctuaciones de energía debido al cálculo de una señal central y una señal lateral de izquierda y derecha (en el lado del codificador) o debido al cálculo de una señal izquierda y derecha desde la señal central y lateral (en el lado del decodificador) no son tan importantes como en la técnica anterior.

[0046] A continuación, se exponen realizaciones preferidas de la presente invención con respecto a los 5 dibujos adjuntos, en los que:

la Fig. 1 es un diagrama de bloques de una implementación preferida de un aparato para codificar una señal multicanal;

10 la Fig. 2 es una realización preferida de un aparato para decodificar una señal multicanal codificada;

la Fig. 3 es una ilustración de diferentes resoluciones de frecuencia y otros aspectos relacionados con la frecuencia para ciertas realizaciones;

15 la Fig. 4a ilustra un diagrama de flujo de los procedimientos realizados en el aparato para codificar con el fin de alinear los canales;

la Fig. 4b ilustra una realización preferida de los procedimientos realizados en el dominio de la frecuencia;

20 la Fig. 4c ilustra una realización preferida de los procedimientos realizados en el aparato para codificar usando una ventana de análisis con porciones de relleno con ceros e intervalos de superposición;

la Fig. 4d ilustra un diagrama de flujo para procedimientos adicionales realizados dentro del aparato para codificar;

25 la Fig. 4e ilustra un diagrama de flujo para mostrar una implementación preferida de una estimación de diferencia de tiempos entre canales;

la Fig. 5 ilustra un diagrama de flujo que ilustra una realización adicional de procedimientos realizados en el aparato para codificar;

30

la Fig. 6a ilustra un diagrama de bloques de una realización de un codificador;

la Fig. 6b ilustra un diagrama de flujo de una realización correspondiente a un decodificador;

35 la Fig. 7 ilustra un escenario de ventana preferido con ventanas seno con baja superposición y relleno con ceros para un análisis y síntesis tiempo-frecuencia en estéreo;

la Fig. 8 ilustra una tabla que muestra el consumo de bits de diferentes valores de parámetros;

40 la Fig. 9a ilustra los procedimientos realizados por un aparato para decodificar una señal multicanal codificada en una realización preferida;

la Fig. 9b ilustra una implementación preferida del aparato para decodificar una señal multicanal codificada;

45 la Fig. 9c ilustra un procedimiento realizado en el contexto de una desalineación de banda ancha en el contexto de la decodificación de una señal multicanal codificada;

la Fig. 10a ilustra la realización de un aparato para estimar una diferencia de tiempos entre canales;

50 la Fig. 10b ilustra una representación esquemática de un tratamiento adicional de señales en el que se aplica la diferencia de tiempos entre canales;

la Fig. 11a ilustra los procedimientos realizados por el procesador de la Fig. 10a;

55 la Fig. 11b ilustra procedimientos adicionales realizados por el procesador en la Fig. 10a;

la Fig. 11c ilustra una implementación adicional del cálculo de un umbral variable y del uso del umbral variable en el análisis de la representación en el dominio del tiempo;

60 la Fig. 11d ilustra una primera realización para la determinación del umbral variable;

la Fig. 11e ilustra una implementación adicional de la determinación del umbral;

la Fig. 12 ilustra una representación en el dominio del tiempo para un espectro de correlación cruzada suavizado para una señal de voz limpia;

la Fig. 13 ilustra una representación en el dominio del tiempo de un espectro de correlación cruzada suavizado para una señal de voz que tiene ruido y efecto ambiental.

[0047] La Fig. 10a ilustra la realización de un aparato para estimar una diferencia de tiempos entre canales entre la señal de un primer canal tal como un canal izquierdo y la señal de un segundo canal tal como un canal derecho. Estos canales se introducen en un convertidor tiempo-espectral 150 que se ilustra adicionalmente, con respecto a la Fig. 4e como elemento 451.

[0048] Asimismo, las representaciones en el dominio del tiempo de las señales del canal izquierdo y del canal derecho entran en un calculador 1020 para calcular un espectro de correlación cruzada para un bloque de tiempo de la señal del primer canal en el bloque de tiempo y la segunda señal del canal en el bloque de tiempo. Asimismo, el aparato comprende un estimador de características espectrales 1010 para estimar una característica de un espectro de la señal del primer canal o de la señal del segundo canal para el bloque de tiempo. El aparato comprende además un filtro de suavizado 1030 para suavizar el espectro de correlación cruzada a lo largo del tiempo usando la característica espectral para obtener un espectro de correlación cruzada suavizado. El aparato comprende además un procesador 1040 para tratar el espectro de correlación cruzada suavizado con el fin de obtener la diferencia de tiempos entre canales.

[0049] En particular, las funcionalidades del estimador de características espectrales también se reflejan en la Fig. 4e, elementos 453, 454 en una realización preferida.

Además, las funcionalidades del calculador de espectro de correlación cruzada 1020 también están reflejadas por el elemento 452 en la Fig. 4e descrita más adelante en una realización preferida.

De manera correspondiente, las funcionalidades del filtro de suavizado 1030 también están reflejadas por el elemento 453 en el contexto de la Fig. 4e que se describirá más adelante. Además, las funcionalidades del procesador 1040 también están descritas en el contexto de la Fig. 4e en una realización preferida como los elementos 456 a 459.

Preferentemente, la estimación de características espectrales calcula un nivel de ruido o una tonalidad del espectro en la que una implementación preferida es el cálculo de una medida de uniformidad espectral próxima a 0 en el caso de señales tonales o sin ruidos y próxima a 1 en el caso de señales ruidosas o de tipo ruido.

Particularmente, a continuación, se configura el filtro de suavizado para aplicar un suavizado más intenso con un primer grado de suavizado a lo largo del tiempo en el caso de una primera característica menos ruidosa o una primera característica más tonal, o para aplicar un suavizado más débil con un segundo grado de suavizado a lo largo del tiempo en el caso de una segunda característica más ruidosa o una segunda característica menos tonal.

Particularmente, el primer suavizado es mayor que el segundo grado de suavizado, en el que la primera característica ruidosa es menos ruidosa que la segunda característica ruidosa o la primera característica tonal es más tonal que la segunda característica tonal. La implementación preferida es la medida de uniformidad espectral.

Asimismo, como se ilustra en la Fig. 11a, el procesador se implementa preferentemente para normalizar el espectro de correlación cruzada suavizado como se ilustra en 456 en las Fig. 4e y 11a antes de realizar el cálculo de la representación en el dominio del tiempo en la etapa 1031 correspondiente a las etapas 457 y 458 en la realización de la Fig. 4e. No obstante, como también se plantea en la Fig. 11a, el procesador también puede operar sin la normalización en la etapa 456 en la Fig. 4e. Entonces el procesador está configurado para analizar la representación en el dominio del tiempo como se ilustra en el bloque 1032 de la Fig. 11a con el fin de hallar la diferencia de tiempos entre canales. Este análisis se puede realizar de cualquier forma conocida y dará como resultado una robustez mejorada, dado que el análisis se realiza sobre la base del espectro de correlación cruzada que se suaviza según la característica espectral.

Como se ilustra en la Fig. 11b, una implementación preferida del análisis en el dominio del tiempo 1032 es un filtrado de paso bajo de la representación en el dominio del tiempo que se ilustra en 458 en la Fig. 11b correspondiente al elemento 458 de la Fig. 4e y un tratamiento posterior adicional 1033 usando una operación de búsqueda/selección de picos máximos dentro de la representación en el dominio del tiempo con filtro de paso bajo.

Como se ilustra en la Fig. 11c, la implementación preferida de la operación de selección o búsqueda

de picos consiste en realizar esta operación usando un umbral variable. Particularmente, el procesador está configurado para realizar la operación de selección o búsqueda de picos dentro de la representación en el dominio del tiempo obtenida del espectro de correlación cruzada suavizado mediante la determinación 1034 de un umbral variable a partir de la representación en el dominio del tiempo y mediante la comparación de un pico o varios picos de la representación en el dominio del tiempo (obtenida con o sin normalización espectral) con el umbral variable, en el que la diferencia de tiempos entre canales se determina como un retardo asociado con un pico que está en una relación predeterminada con el umbral tal que es mayor que el umbral variable.

5 [0058] Como se ilustra en la Fig. 11d, una realización preferida ilustrada en el seudocódigo relacionado con la Fig. 4e-4b que se describe más adelante consiste en la clasificación 1034a de valores según su magnitud. A continuación, como se ilustra en el elemento 1034b en la Fig. 11d, se determina, por ejemplo, el 10% o el 5% más alto de los valores.

10 [0059] A continuación, como se ilustra en la etapa 1034c, un número, tal como el número 3, se multiplica por el valor más bajo del 10% o 5% más alto con el fin de obtener el umbral variable.

15 [0060] Como se señaló, preferentemente, se determina el 10% o el 5% más alto, pero también puede ser útil determinar el número más bajo del 50% más alto de los valores y usar un número para la multiplicación más alto tal como 10. Naturalmente, se determina incluso una cantidad más pequeña tal como el 3% más alto de los valores y a continuación se multiplica el valor más bajo en este 3% más alto de los valores por un número que es, por ejemplo, igual a 2,5 o 2, es decir, menor que 3. Por lo tanto, se pueden usar diferentes combinaciones de números y porcentajes en la realización ilustrada en la Fig. 11d. Además de los porcentajes, los números también pueden variar, y generalmente se prefieren números mayores que 1,5.

20 [0061] En una realización adicional ilustrada en la Fig. 11e, la representación en el dominio del tiempo se divide en subbloques como se ilustra en el bloque 1101, y estos subbloques se indican en la Fig. 13 en 1300. Aquí se usan aproximadamente 16 subbloques para el intervalo válido de modo que cada subbloque tiene una extensión de retardo de 20. No obstante, la cantidad de subbloques puede ser mayor que este valor o menor y preferentemente mayor que 3 y menor que 50.

25 [0062] En la etapa 1102 de la Fig. 11e, se determina el pico en cada subbloque, y en la etapa 1103, se determina el pico promedio en todos los subbloques. A continuación, en la etapa 1104, se determina un valor de multiplicación a que depende de una relación señal-ruido, por una parte, y en una realización adicional, depende de la diferencia entre el umbral y el pico máximo como se indica a la izquierda del bloque 1104. Según estos valores de entrada, se determina uno de entre preferentemente tres valores de multiplicación diferentes en los que el valor de multiplicación puede ser igual a_{bajo} , a_{alto} y $a_{mínimo}$.

30 [0063] A continuación, en la etapa 1105, el valor de multiplicación a determinado en el bloque 1104 se multiplica por el umbral promedio con el fin de obtener el umbral variable que a continuación se usa en la operación de comparación en el bloque 1106. Para la operación de comparación, una vez más se puede usar la representación en el dominio del tiempo que entra en el bloque 1101 o los picos ya determinados en cada subbloque como se señala en el bloque 1102.

35 [0064] A continuación, se detallan realizaciones adicionales en relación con la evaluación y detección de un pico dentro de la función de correlación cruzada en el dominio del tiempo.

40 [0065] La evaluación y detección de un pico dentro de la función de correlación cruzada en el dominio del tiempo obtenida a partir del procedimiento de correlación cruzada generalizada (GCC-PHAT) con el fin de estimar la ITD (*Inter-channel Time Difference*, diferencia de tiempos entre canales) no es siempre sencilla debido a diferentes escenarios de entrada. La entrada de voz limpia puede producir una función de correlación cruzada con baja desviación con un pico intenso, mientras que la voz en un ambiente reverberante ruidoso puede producir un vector con alta desviación y picos con magnitud baja pero todavía destacada que indican la existencia de ITD. Se describe un algoritmo para la detección de picos que es adaptable y flexible para albergar diferentes escenarios de entrada.

45 [0066] Debido a restricciones de retardo, el sistema en general puede manejar la alineación en el tiempo de canales hasta cierto límite, esto es, ITD_MAX. El algoritmo propuesto está diseñado para detectar si existe una ITD válida en los siguientes casos:

50 - **ITD válida debido a pico destacado.** Está presente un pico destacado dentro de los límites [-ITD_MAX, ITD_MAX] de la función de correlación cruzada.

55 - **Sin correlación.** Cuando no existe correlación entre los dos canales, no existe pico destacado. Se deberá definir un umbral, por encima del cual el pico es suficientemente intenso para ser considerado un valor ITD válido. En caso

contrario, no se señalará ningún manejo de ITD, lo cual significa que ITD se ajusta a cero y no se realiza ninguna alineación en el tiempo.

5 - **ITD fuera de límites.** Se deberán evaluar los picos intensos de la función de correlación cruzada fuera de la región [-ITD_MAX, ITD_MAX] con el fin de determinar si existen ITD que se encuentran fuera de la capacidad de manejo del sistema. En este caso no se señalará ningún manejo de ITD y por lo tanto no se realiza ninguna alineación en el tiempo.

10 **[0067]** Para determinar si la magnitud de un pico es suficientemente alta para ser considerada un valor de diferencia de tiempos, es necesario definir un umbral adecuado. Para los diferentes escenarios de entrada, la salida de la función de correlación cruzada varía según los diferentes parámetros, por ejemplo, el ambiente (ruido, reverberación, etc.), la configuración de micrófonos (AB, M/S, etc.). Por lo tanto, resulta esencial definir el umbral de manera adaptable.

15 **[0068]** En el algoritmo propuesto, el umbral se define calculando en primer lugar la media de un cálculo aproximado de la envolvente de la magnitud de la función de correlación cruzada dentro de la región [-ITD_MAX, ITD_MAX] (Fig. 13), y a continuación se pondera el promedio de manera correspondiente según la estimación de SNR.

20 **[0069]** A continuación, se describe el algoritmo paso a paso.

[0070] La salida de la DFT inversa del GCC-PHAT, que representa la correlación cruzada en el dominio del tiempo, se reorganiza desde retardos negativos a positivos (Fig. 12).

25 **[0071]** El vector de correlación cruzada se divide en tres áreas principales: el área de interés, que es [-ITD_MAX, ITD_MAX] y el área fuera de los límites ITD_MAX, que son retardos menores que -ITD_MAX (max_low) y mayores que ITD_MAX (max_high). Los picos máximos de las áreas “fuera del límite” se detectan y se guardan para compararlos con el pico máximo detectado en el área de interés.

30 **[0072]** Para determinar si está presente una ITD válida, se considera el área del subvector [-ITD_MAX, ITD_MAX] de la función de correlación cruzada. El subvector se divide en N subbloques (Fig. 13).

[0073] Para cada subbloque, se determinan y se guardan la magnitud pico máxima peak_sub y la posición de retardo equivalente index_sub.

35 **[0074]** Se determina el máximo de los máximos locales peak_max y se comparará con el umbral para determinar la existencia de un valor ITD válido.

40 **[0075]** El valor máximo peak_max se compara con max_low y max_high. Si peak_max es menor que cualquiera de los dos, entonces no se señala ningún manejo de ITD y no se realiza ninguna alineación en el tiempo. Debido al límite de manejo de ITD del sistema, no es necesario evaluar las magnitudes de los picos fuera del límite.

[0076] Se calcula la media de las magnitudes de los picos:

45
$$peak_{media} = \frac{\sum_N peak_{sub}}{N}$$

[0077] A continuación, se calcula el umbral thres ponderando peak_mean con un factor de ponderación dependiente de SNR a_w.

50
$$thres = a_w peak_{mean}, \quad \text{donde } a_w = \begin{cases} a_{bajo}, & SNR \leq SNR_{umbral} \\ a_{alto}, & SNR > SNR_{umbral} \end{cases}$$

55 **[0078]** En los casos en que $SNR \ll SNR_{umbral}$ y $|thres - peak_{max}| \ll \epsilon$, la magnitud del pico también se compara con un umbral ligeramente más relajado ($a_w = a_{minimo}$) con el fin de evitar rechazar un pico destacado con altos picos vecinos. Los factores de ponderación podrían ser, por ejemplo, $a_{alto} = 3$, $a_{bajo} = 2,5$ y $a_{minimo} = 2$, mientras que SNR_{umbral} podría ser, por ejemplo, 20 dB y el límite $\epsilon = 0,05$.

[0079] Los intervalos preferidos son 2,5 a 5 para a_{alto} ; 1,5 a 4 para a_{bajo} ; 1,0 a 3 para $a_{\text{mínimo}}$; 10 a 30 dB para $\text{SNR}_{\text{umbral}}$; y 0,01 a 0,5 para ϵ , en los que a_{alto} es mayor que a_{bajo} que es mayor que $a_{\text{mínimo}}$.

[0080] Si $\text{peak_max} > \text{thres}$ el retardo equivalente es devuelto como la ITD estimada, y en caso contrario no se señala ningún manejo de ITD ($\text{ITD}=0$).

[0081] Más adelante se describen realizaciones adicionales con respecto a la Fig. 4e.

[0082] A continuación, se expone una implementación preferida de la presente invención dentro del bloque 1050 de la Fig. 10b con el fin de un tratamiento adicional de la señal con respecto a las Fig. 1 a 9e, es decir, en el contexto de un tratamiento/codificación estéreo/multicanal y alineación en el tiempo de dos canales.

[0083] No obstante, como se mencionó y como se ilustró en la Fig. 10b, existen muchos otros campos, en los que también se puede realizar el tratamiento adicional de una señal usando la diferencia de tiempos entre canales determinada.

[0084] La Fig. 1 ilustra un aparato para codificar una señal multicanal que tiene al menos dos canales. La señal multicanal 10 se introduce en un determinador de parámetro 100, por una parte, y en un alineador de señales 200, por otra. El determinador de parámetro 100 determina, por una parte, un parámetro de alineación de banda ancha y, por otra parte, una pluralidad de parámetros de alineación de banda estrecha de la señal multicanal. Estos parámetros salen por una línea de parámetros 12. Asimismo, estos parámetros también salen por una línea de parámetros adicional 14 a una interfaz de salida 500 tal como se ilustra. En la línea de parámetros 14, se envían parámetros adicionales tales como los parámetros de nivel desde el determinador de parámetros 100 hasta la interfaz de salida 500. El alineador de señales 200 está configurado para alinear los al menos dos canales de la señal multicanal 10 usando el parámetro de alineación de banda ancha y la pluralidad de parámetros de alineación de banda estrecha recibidos por vía de la línea de parámetros 10 para obtener canales alineados 20 a la salida del alineador de señales 200. Estos canales alineados 20 se envían a un procesador de señales 300 que está configurado para calcular una señal central 31 y una señal lateral 32 de los canales alineados recibidos por la línea 20. El aparato para codificar comprende además un codificador de señales 400 para codificar la señal central desde la línea 31 y la señal lateral desde la línea 32 para obtener una señal central codificada en la línea 41 y una señal lateral codificada en la línea 42. Estas dos señales son enviadas a la interfaz de salida 500 para generar una señal multicanal codificada en la línea de salida 50. La señal codificada en la línea de salida 50 comprende la señal central codificada de la línea 41, la señal lateral codificada de la línea 42, los parámetros de alineación de banda estrecha y los parámetros de alineación de banda ancha de la línea 14 y, opcionalmente, un parámetro de nivel de la línea 14 y, además opcionalmente, un parámetro de relleno estéreo generado por el codificador de señales 400 y enviado a la interfaz de salida 500 por la línea de parámetros 43.

[0085] Preferentemente, el alineador de señales está configurado para alinear los canales de la señal multicanal usando el parámetro de alineación de banda ancha, antes de que el determinador de parámetros 100 calcule realmente los parámetros de banda estrecha. Por lo tanto, en esta realización, el alineador de señales 200 envía los canales alineados de banda ancha de nuevo al determinador de parámetros 100 por la línea de conexión 15. A continuación, el determinador de parámetros 100 determina la pluralidad de parámetros de alineación de banda estrecha, a partir de uno anterior, con respecto a las características de la señal multicanal alineada con características de banda ancha. En otras realizaciones, no obstante, los parámetros se determinan sin esta secuencia específica de procedimientos.

[0086] La Fig. 4a ilustra una implementación preferida, en la que se realiza la secuencia específica de etapas seguidas por la línea de conexión 15. En la etapa 16, el parámetro de alineación de banda ancha se determina usando los dos canales y se obtiene el parámetro de alineación de banda ancha tal como un parámetro de diferencia de tiempos entre canales o ITD. A continuación, en la etapa 21, los dos canales se alinean mediante el alineador de señales 200 de la Fig. 1 usando el parámetro de alineación de banda ancha. A continuación, en la etapa 17, se determinan los parámetros de banda estrecha usando los canales alineados dentro del determinador de parámetros 100 para determinar una pluralidad de parámetros de alineación de banda estrecha tales como una pluralidad de parámetros de diferencia de fases entre canales para bandas diferentes de la señal multicanal. A continuación, en la etapa 22, se alinean los valores espectrales en cada banda de parámetros usando el parámetro de alineación de banda estrecha correspondiente para esta banda específica. Cuando este procedimiento de la etapa 22 se realiza para cada banda, para la cual está disponible un parámetro de alineación de banda estrecha, entonces los canales alineados primeros y segundos o izquierdos/derechos están disponibles para el tratamiento adicional de señales por el procesador de señales 300 de la Fig. 1.

[0087] La Fig. 4b ilustra una implementación adicional del codificador multicanal de la Fig. 1 en la que se realizan diversos procedimientos en el dominio de la frecuencia.

[0088] Específicamente, el codificador multicanal comprende además un convertidor tiempo-espectro 150

para convertir una señal multicanal en el dominio del tiempo en una representación espectral de los al menos dos canales dentro del dominio de la frecuencia.

5 **[0089]** Asimismo, como se ilustra en 152, el determinador de parámetros, el alineador de señales y el procesador de señales ilustrados en 100, 200 y 300 en la Fig. 1 operan todos en el dominio de la frecuencia.

[0090] Asimismo, el codificador multicanal y, específicamente, el procesador de señales comprende además un convertidor espectro-tiempo 154 para generar una representación en el dominio del tiempo de la señal central, por lo menos.

10

[0091] Preferentemente, el convertidor espectro-tiempo convierte además una representación espectral de la señal lateral también determinada por los procedimientos representados en el bloque 152 en una representación en el dominio del tiempo, y el codificador de señales 400 de la Fig. 1 a continuación está configurado para codificar adicionalmente la señal central y/o la señal lateral como señales en el dominio del tiempo según la implementación
15 específica del codificador de señales 400 de la Fig. 1.

[0092] Preferentemente, el convertidor tiempo-espectro 150 de la Fig. 4b está configurado para implementar las etapas 155, 156 y 157 de la Fig. 4c. Específicamente, la etapa 155 comprende proporcionar una ventana de análisis con al menos una porción con relleno con ceros en uno de sus extremos y, específicamente, una porción
20 con relleno con ceros en la porción de ventana inicial y una porción con relleno con ceros en la porción de ventana terminal como se ilustra, por ejemplo, en la Fig. 7 más adelante. Además, la ventana de análisis tiene adicionalmente intervalos de superposición o porciones superpuestas en una primera mitad de la ventana y en una segunda mitad de la ventana y, además, preferentemente, una parte central que es un intervalo sin superposición, según sea el caso.

25

[0093] En la etapa 156, cada canal se divide en ventanas usando la ventana de análisis con intervalos de superposición. Específicamente, cada canal se divide en ventanas usando la ventana de análisis de forma que se obtiene un primer bloque del canal. A continuación, se obtiene un segundo bloque del mismo canal que tiene un cierto intervalo de superposición con el primer bloque y así sucesivamente, de forma que después de, por ejemplo,
30 cinco operaciones de división en ventanas están disponibles cinco bloques de muestras divididas en ventanas de cada canal que a continuación se transforman individualmente en una representación espectral tal como se ilustra en 157 en la Fig. 4c. El mismo procedimiento se realiza también para los otros canales de modo que, al final de la etapa 157, está disponible una secuencia de bloques de valores espectrales y, específicamente, valores espectrales complejos tales como valores espectrales DFT o muestras de subbandas complejas.

35

[0094] En la etapa 158, que es realizada por el determinador de parámetros 100 de la Fig. 1, se determina un parámetro de alineación de banda ancha y en la etapa 159, que se realiza mediante el alineador de señales 200 de la Fig. 1, se realiza un desplazamiento circular usando el parámetro de alineación de banda ancha. En la etapa 160, nuevamente realizada por el determinador de parámetros 100 de la Fig. 1, se determinan los parámetros de
40 alineación de banda estrecha para bandas/subbandas individuales y en la etapa 161, los valores espectrales alineados son rotados para cada banda usando los parámetros de alineación de banda estrecha correspondientes para las bandas específicas.

[0095] La Fig. 4d ilustra adicionalmente los procedimientos realizados mediante el procesador de señales
45 300. Específicamente, el procesador de señales 300 está configurado para calcular una señal central y una señal lateral como se ilustra en la etapa 301. En la etapa 302, se puede realizar algún tratamiento adicional de la señal lateral y a continuación, en la etapa 303, cada bloque de la señal central y la señal lateral se transforma nuevamente al dominio del tiempo y, en la etapa 304, se aplica una ventana de síntesis a cada bloque obtenido en la etapa 303 y, en la etapa 305, se realiza una operación de superposición-suma para la señal central por un lado y una operación
50 de superposición-suma para la señal lateral por otro lado para finalmente obtener las señales centrales/laterales en el dominio del tiempo.

[0096] Específicamente, las operaciones de las etapas 304 y 305 producen una especie de desvanecimiento cruzado desde un bloque de la señal central o la señal lateral en el bloque siguiente de la señal central y la señal
55 lateral se realiza de modo que, incluso cuando ocurre cualquier cambio de parámetros tal como del parámetro de diferencia de tiempos entre canales o del parámetro de diferencia de fases entre canales, sin embargo no será audible en las señales central/lateral en el dominio del tiempo obtenidas mediante la etapa 305 en la Fig. 4d.

[0097] La nueva codificación estéreo con bajo retardo es una codificación conjunta estéreo central/lateral
60 (M/S) que aprovecha algunas pistas espaciales, en la que el canal central está codificado por un codificador principal mononúcleo, y el canal lateral está codificado en un codificador de núcleo secundario. Los principios del codificador y del decodificador se ilustran en las Fig. 6a, 6b.

[0098] El tratamiento estéreo se realiza principalmente en FD (*Frequency Domain*, dominio de la frecuencia). Opcionalmente, parte del tratamiento estéreo se puede realizar en TD (*Time Domain*, dominio del tiempo) antes del análisis de frecuencia. Así sucede para el cálculo ITD, que se puede calcular y aplicar antes del análisis de frecuencia para alinear los canales en el tiempo antes de proseguir con el análisis y el tratamiento estéreo.

5 Alternativamente, el tratamiento ITD se puede realizar directamente en el dominio de la frecuencia. Dado que los codificadores de voz usuales como ACELP no contienen ninguna descomposición interna tiempo-frecuencia, la codificación estéreo añade un banco de filtros complejo modulado adicional por medio de un banco de filtros de síntesis y análisis antes del codificador de núcleo y otra etapa de banco de filtros de síntesis y análisis después del decodificador de núcleo. En la realización preferida, se emplea una DFT sobremuestreada con una región con baja
10 superposición. No obstante, en otras realizaciones, se puede usar cualquier descomposición valorada como compleja de tiempo-frecuencia con resolución temporal similar.

[0099] El tratamiento estéreo consiste en calcular las pistas espaciales: la diferencia de tiempos entre canales (ITD), las diferencias de fases entre canales (IPD) y las diferencias de niveles entre canales (ILD). La ITD y las IPD se usan en la señal de entrada estéreo para alinear los dos canales L (izquierdo) y R (derecho) en tiempo y en fase. La ITD se calcula en banda ancha o en el dominio del tiempo mientras que las IPD y las ILD se calculan para todas o para una parte de las bandas de parámetros, correspondientes a una descomposición no uniforme del espacio de la frecuencia. Una vez que los dos canales están alineados se aplica un estéreo conjunto M/S, en el que a continuación la señal lateral se predice adicionalmente desde la señal central. La ganancia de predicción se
15 obtiene de las ILD.
20

[0100] La señal central se codifica a continuación mediante un codificador de núcleo primario. En la realización preferida, el codificador de núcleo primario sigue la norma 3GPP EVS, o una codificación obtenida de la misma que puede conmutar entre un modo de codificación de voz, ACELP, y un modo de música sobre la base de una transformación MDCT. Preferentemente, ACELP y el codificador basado en MDCT están sustentados respectivamente por módulos TD-BWE (*Time Domain BandWidth Extension*, extensión de anchura de banda en el dominio del tiempo) y/o IGF (*Intelligent Gap Filling*, relleno inteligente de intervalos).
25

[0101] La predicción de la señal lateral primero es realizada por el canal central usando ganancias de predicción obtenidas de las ILD. A continuación, se puede predecir el residuo mediante una versión con retardo de la señal central o codificarse directamente por medio de un codificador de núcleo secundario, llevado a cabo en la realización preferida en el dominio MDCT. El tratamiento estéreo en el codificador se puede resumir mediante la Fig. 5 tal como se explicará más adelante.
30

[0102] La Fig. 2 ilustra un diagrama de bloques de la realización de un aparato para decodificar una señal multicanal codificada recibida en la línea de entrada 50.
35

[0103] En particular, la señal es recibida por una interfaz de entrada 600. Están conectados con la interfaz de entrada 600 un decodificador de señales 700 y un desalineador de señales 900. Asimismo, un procesador de señales 800 está conectado con un decodificador de señales 700 por una parte y está conectado con un desalineador de señales por otra parte.
40

[0104] En particular, la señal multicanal codificada comprende una señal central codificada, una señal lateral codificada, información sobre el parámetro de alineación de banda ancha e información sobre la pluralidad de parámetros de banda estrecha. Por lo tanto, la señal multicanal codificada en la línea 50 puede ser exactamente la misma señal que la salida de la interfaz de salida de 500 en la Fig. 1.
45

[0105] No obstante, es importante destacar aquí que, al contrario de lo que se ilustra en la Fig. 1, el parámetro de alineación de banda ancha y la pluralidad de parámetros de alineación de banda estrecha incluidos en la señal codificada en cierta forma pueden ser exactamente los parámetros de alineación usados por el alineador de señales 200 en la Fig. 1 pero, alternativamente, también pueden ser sus valores inversos, es decir, parámetros que pueden ser usados exactamente por las mismas operaciones realizadas por el alineador de señales 200 pero con valores inversos de modo que se obtiene la desalineación.
50

[0106] Por lo tanto, la información sobre los parámetros de alineación puede ser los parámetros de alineación usados por el alineador de señales 200 en la Fig. 1 o pueden ser valores inversos, es decir, verdaderos "parámetros de desalineación". Además, estos parámetros normalmente serán cuantificados en una forma determinada tal como se expondrá más adelante con respecto a la Fig. 8.
55

[0107] La interfaz de entrada 600 de la Fig. 2 separa la información sobre el parámetro de alineación de banda ancha y la pluralidad de parámetros de alineación de banda estrecha de las señales centrales/laterales codificadas y envía esta información por la línea de parámetros 610 al desalineador de señales 900. Por otra parte, la señal central codificada se envía al decodificador de señales 700 por la línea 601 y la señal lateral codificada se
60

envía al decodificador de señales 700 por la línea de señal 602.

[0108] El decodificador de señales está configurado para decodificar la señal central codificada y para decodificar la señal lateral codificada para obtener una señal central decodificada en la línea 701 y una señal lateral decodificada en la línea 702. Estas señales son usadas por el procesador de señales 800 para calcular una señal decodificada del primer canal o una señal izquierda decodificada y para calcular una señal decodificada del segundo canal o una señal derecha decodificada del canal derecho a partir de la señal central decodificada y la señal lateral decodificada, y el primer canal decodificado y el segundo canal decodificado son salidas en las líneas 801, 802, respectivamente. El desalineador de señales 900 está configurado para desalinear el primer canal decodificado en la línea 801 y el canal derecho decodificado 802 usando la información sobre el parámetro de alineación de banda ancha y adicionalmente usando la información sobre la pluralidad de parámetros de alineación de banda estrecha para obtener una señal multicanal decodificada, es decir, una señal decodificada que tiene al menos dos canales decodificados y desalineados en las líneas 901 y 902.

[0109] La Fig. 9a ilustra una secuencia preferida de etapas realizadas por el desalineador de señales 900 de la Fig. 2. Específicamente, la etapa 910 recibe los canales izquierdo y derecho alineados tal como están disponibles en las líneas 801, 802 de la Fig. 2. En la etapa 910, el desalineador de señales 900 desalinea subbandas individuales usando la información sobre los parámetros de alineación de banda estrecha con el fin de obtener canales primero y segundo o izquierdo y derecho con desalineación de fase decodificados en 911a y 911b. En la etapa 912, los canales son desalineados usando el parámetro de alineación de banda ancha de modo que, en 913a y 913b, se obtienen canales desalineados en fase y tiempo.

[0110] En la etapa 914 se realiza cualquier tratamiento adicional que comprende el uso de una operación de división en ventanas o cualquier superposición-suma o, generalmente, cualquier operación de desvanecimiento cruzado con el fin de obtener, en 915a o 915b, una señal decodificada con reducción de artefactos o libre de artefactos, es decir, obtener canales decodificados que no tienen ningún artefacto aunque haya habido, normalmente, parámetros de desalineación variables en el tiempo para la banda ancha, por una parte, y para la pluralidad de bandas estrechas, por otra parte.

[0111] La Fig. 9b ilustra una implementación preferida del decodificador multicanal ilustrado en la Fig. 2.

[0112] En particular, el procesador de señales 800 de la Fig. 2 comprende un convertidor tiempo-espectro 810.

[0113] El procesador de señales comprende además un convertidor central/lateral a izquierda/derecha 820 con el fin de calcular a partir de una señal central M y una señal lateral S una señal izquierda L y una señal derecha R.

[0114] No obstante, es importante señalar que, para calcular L y R mediante la conversión central/lateral-izquierda/derecha en el bloque 820, no hay que usar necesariamente la señal lateral S. En su lugar, como se expondrá más adelante, las señales izquierda/derecha se calculan inicialmente usando solo un parámetro de ganancia obtenido de un parámetro de diferencia de niveles entre canales ILD. Generalmente, también se puede considerar que la ganancia de predicción es una forma de ILD. La ganancia se puede obtener de ILD pero también se puede calcular directamente. Es preferible no calcular ILD, sino calcular la ganancia de predicción directamente y transmitir y usar la ganancia de predicción en el decodificador en vez del parámetro ILD.

[0115] Por lo tanto, en esta implementación, la señal lateral S solo se usa en el actualizador de canales 830 que opera con el fin de proporcionar una mejor señal izquierda/derecha usando la señal lateral transmitida S como se ilustra mediante la línea de paso 821.

[0116] Por lo tanto, el convertidor 820 opera usando un parámetro de nivel obtenido mediante una entrada de parámetro de nivel 822 y sin usar realmente la señal lateral S, pero el actualizador de canales 830 opera a continuación usando la lateral 821 y, según la implementación específica, usando un parámetro de relleno estéreo recibido por la línea 831. El alineador de señales 900 comprende entonces un desalineador de fase y un elemento de cambio de escala de energía 910. El elemento de cambio de escala de energía es controlado mediante un factor de cambio de escala obtenido mediante un calculador de factor de cambio de escala 940. El calculador de factor de cambio de escala 940 es alimentado por la salida del actualizador de canales 830. Sobre la base de los parámetros de alineación de banda estrecha recibidos mediante la entrada 911, se realiza la desalineación de fase y, en el bloque 920, sobre la base del parámetro de alineación de banda ancha recibido mediante la línea 921, se realiza la desalineación en el tiempo. Finalmente, se realiza la conversión espectro-tiempo 930 con el fin de obtener finalmente la señal decodificada.

[0117] La Fig. 9c ilustra una secuencia adicional de etapas realizadas normalmente dentro de los bloques

920 y 930 de la Fig. 9b en una realización preferida.

[0118] Específicamente, los canales desalineados de banda estrecha son introducidos en la funcionalidad de desalineación de banda ancha correspondiente al bloque 920 de la Fig. 9b. En el bloque 931 se realiza una DFT o cualquier otra transformada. Después del cálculo real de las muestras en el dominio del tiempo, se realiza una división en ventanas de síntesis opcional usando una ventana de síntesis. Preferentemente la ventana de síntesis es exactamente igual a la ventana de análisis, o procede de la ventana de análisis, por ejemplo, mediante interpolación o diezmado, pero depende en cierto modo de la ventana de análisis. Preferentemente, esta dependencia es tal que los factores de multiplicación definidos por dos ventanas de superposición suman hasta uno por cada punto en el intervalo de superposición. Por lo tanto, después de la ventana de síntesis en el bloque 932, se realiza una operación de superposición y una operación de suma posterior. Alternativamente, en lugar de la división en ventanas de síntesis y la operación de superposición/suma, se realiza un desvanecimiento cruzado entre los bloques posteriores para cada canal, con el fin de obtener, como ya se expuso en el contexto de la Fig. 9a, una señal decodificada con artefactos reducidos.

[0119] Si se considera la Fig. 6b, resulta claro que las operaciones reales de decodificación para la señal central, es decir, el "decodificador EVS" por una parte, y para la señal lateral, la cuantificación vectorial inversa VQ^{-1} y la operación MDCT inversa (IMDCT) corresponden al decodificador de señales 700 de la Fig. 2.

[0120] Asimismo, las operaciones DFT en el bloque 810 corresponden al elemento 810 en la Fig. 9b y las funcionalidades del tratamiento estéreo inverso y el desplazamiento inverso en el tiempo corresponden a los bloques 800, 900 de la Fig. 2 y las operaciones de la DFT inversa 930 en la Fig. 6b corresponden a la operación correspondiente en el bloque 930 de la Fig. 9b.

[0121] A continuación, se expone con mayor detalle la Fig. 3. En particular, la Fig. 3 ilustra un espectro DFT que tiene líneas espectrales individuales. Preferentemente, el espectro DFT o cualquier otro espectro ilustrado en la Fig. 3 es un espectro complejo y cada línea es una línea espectral compleja que tiene magnitud y fase o que tiene una parte real y una parte imaginaria.

[0122] Además, el espectro también está dividido en bandas de parámetros diferentes. Cada banda de parámetros tiene al menos una y preferentemente más de una línea espectral. Además, las bandas de parámetros aumentan desde frecuencias bajas a altas. Normalmente, el parámetro de alineación de banda ancha es un solo parámetro de alineación de banda ancha para todo el espectro, es decir, para un espectro que comprende todas las bandas 1 a 6 en la realización ejemplar en la Fig. 3.

[0123] Asimismo, la pluralidad de parámetros de alineación de banda estrecha se proporciona de forma tal que existe un solo parámetro de alineación para cada banda de parámetros. Esto significa que el parámetro de alineación para una banda siempre se aplica a todos los valores espectrales dentro de la banda correspondiente.

[0124] Asimismo, además de los parámetros de alineación de banda estrecha, también se proporcionan parámetros de nivel para cada banda de parámetros.

[0125] Contrariamente a los parámetros de nivel que se proporcionan para cada una de las bandas de parámetros desde la banda 1 hasta la banda 6, es preferible proporcionar la pluralidad de parámetros de alineación de banda estrecha solo para un número limitado de bandas bajas tales como las bandas 1, 2, 3 y 4.

[0126] Además, se proporcionan parámetros de relleno estéreo para un cierto número de bandas excluyendo las bandas bajas tales como, en la realización ejemplar, las bandas 4, 5 y 6, mientras que existen valores espectrales de señal lateral para las bandas bajas de parámetros 1, 2 y 3, y en consecuencia, no existen parámetros de relleno estéreo para estas bandas bajas en el que la adaptación de forma de onda se obtiene usando tanto la señal lateral en sí misma como una señal residual de predicción que representa la señal lateral.

[0127] Como ya se señaló, existen más líneas espectrales en las bandas altas, tales como, en la realización de la Fig. 3, siete líneas espectrales en la banda de parámetros 6 frente a solo tres líneas espectrales en la banda de parámetros 2. Naturalmente, sin embargo, el número de bandas de parámetros, el número de líneas espectrales y el número de líneas espectrales dentro de una banda de parámetros y también los diferentes límites para ciertos parámetros serán diferentes.

[0128] No obstante, la Fig. 8 ilustra una distribución de los parámetros y el número de bandas para las cuales se proporcionan parámetros en una determinada realización en el que existen, en contraste con la Fig. 3, realmente 12 bandas.

[0129] Como se ilustra, el parámetro de nivel ILD se proporciona para cada una de las 12 bandas y es

cuantificado hasta una exactitud de cuantificación representada por cinco bits por banda.

[0130] Asimismo, los parámetros de alineación de banda estrecha IPD solo se proporcionan para las bandas bajas hasta una frecuencia de borde de 2,5 kHz. Además, la diferencia de tiempos entre canales o el parámetro de alineación de banda ancha solo se proporciona como un único parámetro para todo el espectro, pero con una exactitud de cuantificación muy elevada representada por ocho bits para toda la banda.

[0131] Asimismo, se proporcionan parámetros de relleno estéreo cuantificados de manera bastante aproximada representados por tres bits por banda y no para las bandas bajas debajo de 1 kHz dado que, para las bandas bajas, se incluyen valores codificados realmente de señal lateral o valores espectrales residuales de señal lateral.

[0132] A continuación, se resume con respecto a la Fig. 5 un tratamiento preferido en el lado del codificador. En una primera etapa, se realiza un análisis DFT del canal izquierdo y del canal derecho. Este procedimiento corresponde a las etapas 155 a 157 de la Fig. 4c. En la etapa 158, se calcula el parámetro de alineación de banda ancha y, particularmente, la diferencia de tiempos entre canales (ITD) del parámetro de alineación de banda ancha. Como se ilustra en 170, se realiza un desplazamiento en el tiempo de L y R en el dominio de la frecuencia. Alternativamente, este desplazamiento en el tiempo se puede realizar en el dominio del tiempo. A continuación, se realiza una DFT inversa, el desplazamiento en el tiempo se realiza en el dominio del tiempo y se realiza una DFT directa adicional con el fin una vez más de tener representaciones espectrales después de la alineación usando el parámetro de alineación de banda ancha.

[0133] Los parámetros ILD, es decir, parámetros de nivel y parámetros de fase (parámetros IPD), se calculan para cada banda de parámetros en las representaciones desplazadas L y R como se ilustra en la etapa 171. Esta etapa corresponde a la etapa 160 de la Fig. 4c, por ejemplo. Las representaciones L y R desplazadas en el tiempo se rotan como una función de los parámetros de diferencia de fases entre canales como se ilustra en la etapa 161 de la Fig. 4c o la Fig. 5. A continuación, la señal central y la señal lateral se calculan como se ilustra en la etapa 301 y, preferentemente, además con una operación de conservación de energía como se expone más adelante. En una etapa siguiente 174, se realiza una predicción de S con M como una función de ILD y opcionalmente con una señal M pasada, es decir, una señal central de una trama anterior. Posteriormente, se realiza la DFT inversa de la señal central y la señal lateral que corresponde a las etapas 303, 304, 305 de la Fig. 4d en la realización preferida.

[0134] En la etapa final 175, la señal central en el dominio del tiempo M y, opcionalmente, la señal residual se codifica como se ilustra en la etapa 175. Este procedimiento corresponde a lo realizado por el codificador de señales 400 en la Fig. 1.

[0135] En el decodificador en el tratamiento de estéreo inverso, la señal *lateral* (*Side*) se genera en el dominio DFT y se predice en primer lugar a partir de la señal *Central* (*Mid*) como:

$$\tilde{Side} = g \cdot Mid$$

en la que g es una ganancia calculada para cada banda de parámetros y es función de la diferencia de nivel entre canales (ILD) transmitida.

[0136] La señal residual de la predicción $Side - g \cdot Mid$ se puede refinar a continuación de dos formas diferentes:

- Mediante una codificación secundaria de la señal residual:

$$\tilde{Side} = g \cdot Mid + g_{cod} \cdot (Side - g \cdot Mid)$$

en el que g_{cod} es una ganancia global transmitida para todo el espectro

- Mediante una predicción residual, conocida como relleno estéreo, que predice el espectro lateral residual con el espectro de la señal *central* anterior decodificada de la trama DFT anterior:

$$\tilde{Side} = g \cdot Mid + g_{pred} \cdot Mid \cdot z^{-1}$$

en la que g_{pred} es una ganancia de predicción transmitida por banda de parámetros.

[0137] Las dos clases de refinamiento de codificación se pueden combinar dentro del mismo espectro DFT. En la realización preferida, la codificación residual se aplica sobre las bandas bajas de parámetros, mientras que la

predicción residual se aplica sobre las bandas restantes. La codificación residual en la realización preferida como se ilustra en la Fig. 1 se realiza en el dominio MDCT después de sintetizar la señal lateral residual en el dominio del tiempo y transformándola mediante MDCT. A diferencia de la DFT, en MDCT se hace muestreo crítico y es más adecuada para la codificación de audio. Los coeficientes de MDCT se someten directamente a cuantificación vectorial mediante una Cuantificación Vectorial Reticular pero, alternativamente, se pueden codificar mediante Cuantificación Escalar seguida de codificación entrópica. Alternativamente, la señal lateral residual también se puede codificar en el Dominio del Tiempo mediante una técnica de codificación de la voz o directamente en el dominio DFT.

10 1. Análisis tiempo-frecuencia: DFT

[0138] Es importante que la descomposición adicional tiempo-frecuencia del tratamiento estéreo realizado mediante DFT permita un buen análisis del escenario auditivo mientras que no aumente significativamente el retardo general del sistema de codificación. Por omisión, se usa una resolución temporal de 10 ms (dos veces el entramado de 20 ms del codificador de núcleo). Las ventanas de análisis y síntesis son iguales y son simétricas. En la Fig. 7 la ventana se representa con una velocidad de muestreo de 16 kHz. Se puede observar que la región de superposición está limitada para reducir el retardo generado y que también se añade relleno con ceros para contrarrestar el desplazamiento circular cuando se aplica ITD en el dominio de la frecuencia tal como se explicará más adelante.

20 2. Parámetros estéreo

[0139] Los parámetros estéreo se pueden transmitir como máximo a la resolución temporal de la DFT estéreo. Como mínimo se pueden reducir a la resolución de entramado del codificador de núcleo, es decir, 20 ms. Por omisión, cuando no se detectan transitorios, los parámetros se calculan cada 20 ms sobre 2 ventanas DFT. Las bandas de parámetros constituyen una descomposición no uniforme y no superpuesta del espectro después de aproximadamente 2 veces o 4 veces la ERB (*Equivalent Rectangular Bandwidths*, anchuras de banda rectangulares equivalentes). Por omisión, se usa una escala ERB de 4 veces para un total de 12 bandas para una anchura de banda de frecuencia de 16 kHz (velocidad de muestreo 32 kbps, estéreo de banda superancha). La Fig. 8 resume un ejemplo de configuración, en el cual la información lateral estéreo se transmite con aproximadamente 5 kbps.

30

3. Cálculo de ITD y alineación de tiempo de canales

[0140] Las ITD se calculan por la estimación del **TDOA** (*Time Delay of Arrival*, tiempo de retardo de llegada) usando la **GCC-PHAT** (*Generalized Cross Correlation with Phase Transform*, correlación cruzada generalizada con transformada de fase):

35

$$ITD = \operatorname{argmax}(\operatorname{IDFT}\left(\frac{L_i(f)R_i^*(f)}{|L_i(f)R_i^*(f)|}\right))$$

en la que L y R son los espectros de frecuencia de los canales izquierdo y derecho respectivamente. El análisis de frecuencia se puede realizar independientemente de la DFT usada para el tratamiento estéreo posterior o se puede compartir. El pseudocódigo para calcular la ITD es el siguiente:

40

```

45 L =fft(ventana(l));
   R =fft(ventana(r));
   tmp = L .* conj( R );
   sfm_L = prod(abs(L).^(1/longitud(L)))/(media(abs(L))+eps);
   sfm_R = prod(abs(R).^(1/longitud(R)))/(media(abs(R))+eps);
   sfm = max(sfm_L,sfm_R);
50 h.cross_corr_smooth = (1-sfm)*h.cross_corr_smooth+sfm*tmp;
   tmp = h.cross_corr_smooth ./ abs( h.cross_corr_smooth+eps );
   tmp = ifft( tmp );
   tmp = tmp([longitud(tmp)/2+1:longitud(tmp) 1:longitud(tmp)/2+1]);
   tmp_sort = sort( abs(tmp) );
55 umbral = 3 * tmp_sort( vuelta(0.95*longitud(tmp_sort)) );
   xcrr_time=abs(tmp(- ( h.stereo_itd_q_max - (length(tmp)-1)/2 - 1 )):-

```

```

5      ( h.stereo_itd_q_min - (length(tmp)-1)/2 - 1 ));
      % suavizado de salida para mejor detección
      xcorr_time=[xcorr_time 0];
      xcorr_time2=filtro([0,25 0,5 0,25],1,xcorr_time);
      [m,i]=max(xcorr_time2(2:fin));
      si m > umbral
          itd = h.stereo_itd_q_max - i + 1;
      si no
          itd = 0;
10     fin

```

[0141] La Fig. 4e ilustra un diagrama de flujo para implementar el pseudocódigo ilustrado anteriormente con el fin de obtener un cálculo robusto y eficiente de una diferencia entre canales como un ejemplo de parámetro de alineación de banda ancha.

[0142] En el bloque 451, se realiza un análisis DFT de las señales en el dominio del tiempo para un primer canal (l) y un segundo canal (r). Este análisis DFT será normalmente el mismo análisis DFT que se ha expuesto en el contexto de las etapas 155 a 157 en la Fig. 5 o la Fig. 4c, por ejemplo.

[0143] A continuación, se realiza una correlación cruzada para cada segmento de frecuencias como se ilustra en el bloque 452.

[0144] Por lo tanto, se obtiene un espectro de correlación cruzada para el intervalo espectral completo de los canales izquierdo y derecho.

[0145] En la etapa 453, a continuación, se calcula una medida de uniformidad espectral a partir de la magnitud de los espectros de L y R y, en la etapa 454, se selecciona la mayor medida de uniformidad espectral. No obstante, la selección en la etapa 454 no necesariamente tiene que ser la selección de la mayor, sino que esta determinación de una sola SFM de los dos canales también puede ser la selección y el cálculo del canal izquierdo solamente o del canal derecho solamente o puede ser el cálculo del promedio ponderado de ambos valores de SFM.

[0146] En la etapa 455, el espectro de correlación cruzada es suavizado a continuación a lo largo del tiempo según las medidas de uniformidad espectral.

[0147] Preferentemente, la medida de uniformidad espectral se calcula dividiendo la media geométrica del espectro de magnitud por la media aritmética del espectro de magnitud. Por lo tanto, los valores de SFM están limitados entre cero y uno.

[0148] En la etapa 456, el espectro de correlación cruzada suavizado a continuación se normaliza por su magnitud y en la etapa 457 se calcula una DFT inversa del espectro de correlación cruzada normalizado y suavizado. En la etapa 458, preferentemente se realiza un determinado filtro en el dominio del tiempo, si bien este filtrado en el dominio del tiempo también se puede omitir según la implementación, aunque se prefiera, tal como se describirá más adelante.

[0149] En la etapa 459, se realiza una estimación de la ITD mediante selección de valores máximos (*peak-picking*) de la función de filtro generalizada de correlación cruzada y mediante la realización de ciertas operaciones de fijación de umbral.

[0150] Si no se obtiene ningún pico por encima del umbral, entonces la ITD se ajusta a cero y no se realiza alineación en el tiempo para este bloque correspondiente.

[0151] El cálculo de la ITD también se puede resumir tal como se expresa a continuación. La correlación cruzada se calcula en el dominio de la frecuencia antes de ser suavizada según la Medida de Uniformidad Espectral. La SFM está limitada entre 0 y 1. En el caso de señales del tipo ruido, la SFM será alta (es decir, alrededor de 1) y el suavizado será débil. En el caso de una señal del tipo tono, la SFM será baja y el suavizado se volverá más intenso. A continuación, se normaliza la correlación cruzada suavizada mediante su amplitud antes de ser transformada nuevamente al dominio del tiempo. La normalización corresponde a la transformada de fase de la correlación cruzada, y se sabe que funciona mejor que la correlación cruzada normal en ambientes con ruido bajo y reverberación relativamente alta. La función en el dominio de la frecuencia así obtenida se filtra en primer lugar para

obtener una selección más robusta de picos máximos. El índice correspondiente a la amplitud máxima corresponde a una estimación de la diferencia de tiempo entre el canal izquierdo y el derecho (ITD). Si la amplitud del máximo es menor que un umbral dado, entonces la estimación de la ITD no se considera fiable y se ajusta a cero.

- 5 **[0152]** Si la alineación en el tiempo se aplica en el dominio del tiempo, la ITD se calcula en un análisis DFT separado. El desplazamiento se realiza del modo siguiente:

$$\begin{cases} r(n) = r(n + ITD) \text{ si } ITD > 0 \\ l(n) = l(n - ITD) \text{ si } ITD < 0 \end{cases}$$

- 10 **[0153]** Este requiere un retardo adicional en el codificador, que en el máximo es igual a la ITD máxima absoluta que se puede manejar. La variación de la ITD a lo largo del tiempo se suaviza mediante la división en ventanas de análisis de DFT.

- 15 **[0154]** Alternativamente, la alineación en el tiempo se puede realizar en el dominio de la frecuencia. En este caso, el cálculo de ITD y el desplazamiento circular están en el mismo dominio de DFT, dominio compartido con este otro tratamiento estéreo. El desplazamiento circular viene dado por:

$$\begin{cases} L(f) = L(f) e^{-j2\pi f \frac{ITD}{2}} \\ R(f) = R(f) e^{+j2\pi f \frac{ITD}{2}} \end{cases}$$

- 20 **[0155]** El relleno con ceros de las ventanas DFT es necesario para simular un desplazamiento en el tiempo con un desplazamiento circular. El tamaño del relleno con ceros corresponde a la ITD máxima absoluta que se puede manejar. En la realización preferida, el relleno con ceros está dividido de manera uniforme en ambos lados de las ventanas de análisis, mediante la adición de 3,125 ms de ceros en ambos extremos. La ITD máxima absoluta posible es entonces 6,25 ms. En la configuración de micrófonos A-B, corresponde en el peor de los casos a una distancia máxima de aproximadamente 2,15 metros entre los dos micrófonos. La variación de ITD a lo largo del tiempo se suaviza mediante la división en ventanas de síntesis y la superposición-suma de la DFT.

- 25 **[0156]** Es importante que el desplazamiento en el tiempo esté seguido por una división en ventanas de la señal desplazada. Esta es la principal diferencia con la técnica anterior de BCC (*Binaural Cue Coding*, codificación de referencia binaural), en la que el desplazamiento en el tiempo se aplica en una señal dividida en ventanas pero que no se divide en ventanas adicionalmente en la etapa de síntesis. En consecuencia, cualquier cambio en ITD a lo largo del tiempo produce un transitorio/chasquido artificial en la señal decodificada.

4. Cálculo de IPD y rotación de canales

- 35 **[0157]** Las IPD se calculan después de la alineación de los dos canales y esto para cada banda de parámetros o al menos hasta una *ipd_max_band* dada, dependiente de la configuración estéreo.

$$IPD[b] = \text{ángulo} \left(\sum_{k=\text{band}_l[b]}^{\text{band}_r[b+1]} L[k] R^*[k] \right)$$

- 40 **[0158]** A continuación, se aplican las IPD a los dos canales para alinear sus fases:

$$\begin{cases} L'(k) = L(k) e^{-j\beta} \\ R'(k) = R(k) e^{j(IPD[b] - \beta)} \end{cases}$$

- 45 **[0159]** En las que $\beta = \text{atan2}(\sin(IPD_l[b]), \cos(IPD_l[b]) + c)$, $c = 10^{ILD_l[b]/20}$ y b es el índice de banda de parámetros al cual pertenece el índice de frecuencia k . El parámetro β es responsable de distribuir la cantidad de rotaciones de fase entre los dos canales mientras realiza la alineación de sus fases. β depende de IPD pero también del nivel de amplitud relativa de los canales, ILD. Si un canal tiene amplitud más elevada, será considerado canal principal y resultará menos afectado por las rotaciones de fase que el canal con amplitud más baja.

5. Suma-resta y codificación de señal lateral

- 50 **[0160]** La transformación de suma-resta se realiza en los espectros alineados en tiempo y fase de los dos canales de forma que la energía se conserva en la señal central.

55

$$\begin{cases} M(f) = (L'(f) + R'(f)) \cdot \alpha \cdot \sqrt{\frac{1}{2}} \\ S(f) = (L'(f) - R'(f)) \cdot \alpha \cdot \sqrt{\frac{1}{2}} \end{cases}$$

en la que $\alpha = \sqrt{\frac{L'^2 + R'^2}{(L' + R')^2}}$ está limitado entre 1/1,2 y 1,2, es decir, -1,58 y +1,58 dB. Esta limitación evita artefactos cuando se ajusta la energía de M y S. Cabe destacar que esta conservación de energía es menos importante cuando el tiempo y la fase se alinean de antemano. Alternativamente, los límites se pueden disminuir o aumentar.

[0161] La señal lateral se predice adicionalmente con M:

$$S'(f) = S(f) - g(ILD)M(f)$$

10

en la que $g(ILD) = \frac{c-1}{c+1}$ en la que $c = 10^{ILD_0(p)/20}$. Alternativamente, la ganancia de predicción óptima g se puede determinar minimizando el error cuadrático medio (ECM) de la señal residual y las ILD deducidas mediante la ecuación anterior.

15 **[0162]** La señal residual $S'(f)$ se puede modelizar de dos maneras: prediciéndola con el espectro de retardo de M o codificándola directamente en el dominio MDCT en el dominio MDCT.

6. Decodificación estéreo

20 **[0163]** La señal central X y la señal lateral S se convierten primero a los canales izquierdo y derecho L y R del modo siguiente:

$$L_i[k] = M_i[k] + gM_i[k], \text{ para } \text{band_limits}[b] \leq k < \text{band_limits}[b+1],$$

25

$$R_i[k] = M_i[k] - gM_i[k], \text{ para } \text{band_limits}[b] \leq k < \text{band_limits}[b+1],$$

en el que la ganancia g por banda de parámetros se obtiene del parámetro ILD:

$$g = \frac{c-1}{c+1} \text{ en el que } c = 10^{ILD_0(p)/20},$$

30

[0164] Para bandas de parámetros por debajo de `cod_max_band`, los dos canales se actualizan con la señal lateral decodificada:

$$L_i[k] = L_i[k] + \text{cod_gain}_i \cdot S_i[k], \text{ para } 0 \leq k < \text{band_limits}[\text{cod_max_band}],$$

35

$$R_i[k] = R_i[k] - \text{cod_gain}_i \cdot S_i[k], \text{ para } 0 \leq k < \text{band_limits}[\text{cod_max_band}],$$

[0165] Para bandas de parámetros más elevadas, se predice la señal lateral y los canales se actualizan como:

40

$$L_i[k] = L_i[k] + \text{cod_pred}_i[b] \cdot M_{i-1}[k], \text{ para } \text{band_limits}[b] \leq k < \text{band_limits}[b+1],$$

$$R_i[k] = R_i[k] - \text{cod_pred}_i[b] \cdot M_{i-1}[k], \text{ para } \text{band_limits}[b] \leq k < \text{band_limits}[b+1],$$

45 **[0166]** Finalmente, los canales se multiplican por un valor complejo con el objetivo de restaurar la energía original y la fase entre canales de la señal estéreo:

$$L_i[k] = \alpha \cdot e^{j2\pi\beta} \cdot L_i[k]$$

50

$$R_i[k] = \alpha \cdot e^{j2\pi\beta - \text{IPD}_i[b]} \cdot R_i[k]$$

en las que

$$\alpha = \sqrt{2 \cdot \frac{\sum_{k=\text{band_limits}[b]}^{\text{band_limits}[b+1]} M_i^2[k]}{\sum_{k=\text{band_limits}[b]}^{\text{band_limits}[b+1]} L_i^2[k] + \sum_{k=\text{band_limits}[b]}^{\text{band_limits}[b+1]} R_i^2[k]}}$$

en la que α está definida y limitada como se definió anteriormente, y en la que $\beta = \text{atan2}(\text{sen}(\text{IPD}_i[b]), \text{cos}(\text{IPD}_i[b]) + c)$, y en la que $\text{atan2}(x,y)$ es la tangente inversa en el cuarto cuadrante de x sobre y .

5 **[0167]** Finalmente, los canales son desplazados en el tiempo ya sea en el dominio del tiempo o de la frecuencia, según las ITD transmitidas. Los canales en el dominio del tiempo se sintetizan mediante DFT inversas y superposición-suma.

10 **[0168]** Las características específicas de la invención se relacionan con la combinación de pistas espaciales y codificación estéreo conjunto suma-resta. Específicamente, se calculan y aplican en los canales estéreo (izquierdo y derecho) las pistas espaciales IDT e IPD. Asimismo, se calcula la suma-resta (señales M/S) y preferentemente se aplica una predicción de S con M.

15 **[0169]** En el lado del decodificador, las pistas espaciales de banda ancha y de banda estrecha se combinan entre sí con codificación estéreo conjunto suma-resta. En particular, la señal lateral se predice con la señal central usando al menos una pista espacial tal como ILD y se calcula una suma-resta inversa para obtener los canales izquierdo y derecho, y, además, se aplican en los canales izquierdo y derecho las pistas espaciales de banda ancha
20 y banda estrecha.

[0170] Preferentemente, el codificador tiene una ventana y superposición-suma con respecto a los canales alineados en el tiempo después del tratamiento con el uso de la ITD. Además, el decodificador tiene adicionalmente una operación de división en ventanas y superposición-suma de las versiones desplazadas o desalineadas de los
25 canales después de aplicar la diferencia de tiempos entre canales.

[0171] El cálculo de la diferencia de tiempos entre canales con el procedimiento GCC-Phat es un procedimiento específicamente robusto.

30 **[0172]** El nuevo procedimiento es ventajoso con respecto a la técnica anterior ya que logra la codificación de velocidad binaria de audio estéreo o audio multicanal con bajo retardo. Está diseñado específicamente para ser robusto ante señales de entrada de distinta naturaleza y diferentes configuraciones del registro estéreo o multicanal. En particular, la presente invención proporciona una buena calidad para codificación de velocidad binaria de voz
estéreo.

35 **[0173]** Los procedimientos preferidos encuentran uso en la distribución de transmisiones de todo tipo de contenido de audio estéreo o multicanal tal como voz y música siempre con una baja velocidad binaria dada. Dichas áreas de aplicación son radio digital, transmisión continua en internet o aplicaciones de comunicación de audio.

40 **[0174]** Una señal de audio codificada de la invención se puede almacenar en un medio de almacenamiento digital o un medio de almacenamiento no transitorio o se puede transmitir en un medio de transmisión tal como un medio de transmisión inalámbrica o un medio de transmisión por cable tal como Internet.

45 **[0175]** Aunque algunos aspectos se han descrito en el contexto de un aparato, resulta claro que estos aspectos también representan una descripción del procedimiento correspondiente, en el que un bloque o dispositivo corresponde a una etapa de un procedimiento o una característica de una etapa de un procedimiento. De forma análoga, los aspectos descritos en el contexto de una etapa de un procedimiento también representan una descripción de un bloque o elemento o característica de un aparato correspondiente.

50 **[0176]** Según determinados requerimientos de implementación, las realizaciones de la invención se pueden implementar en hardware o en software. La implementación se puede realizar usando un medio de almacenamiento digital, por ejemplo, un disco flexible, un DVD, un CD, una memoria ROM, PROM, EPROM, EEPROM o FLASH, que tenga almacenadas en él señales de control legibles electrónicamente, que cooperan (o son capaces de cooperar) con un sistema informático programable de modo que se realice el procedimiento respectivo.

55 **[0177]** Algunas realizaciones según la invención comprenden un soporte de datos que tiene señales de control legibles electrónicamente, que son capaces de cooperar con un sistema informático programable, de forma que se lleva a cabo uno de los procedimientos descritos en la presente memoria.

60 **[0178]** Generalmente, las realizaciones de la presente invención se pueden implementar como un producto

que es un programa informático con un código de programa, estando el código de programa operativo para realizar uno de los procedimientos cuando el programa informático se ejecuta en un ordenador. El código de programa, por ejemplo, se puede almacenar en un soporte legible por una máquina.

5 **[0179]** Otras realizaciones comprenden el programa informático para realizar uno de los procedimientos descritos en la presente memoria, almacenados en un soporte legible por una máquina o un medio de almacenamiento no transitorio.

10 **[0180]** En otras palabras, una realización del procedimiento de la invención es, por lo tanto, un programa informático que tiene un código de programa para realizar uno de los procedimientos descritos en la presente memoria, cuando el programa se ejecuta en un ordenador.

15 **[0181]** Una realización adicional de la invención es, por lo tanto, un soporte de datos (o un medio de almacenamiento digital, o un medio legible por ordenador) que comprende, grabado en él, el programa informático para realizar uno de los procedimientos descritos en la presente memoria.

20 **[0182]** Una realización adicional del procedimiento de la invención es, por lo tanto, un tren de datos o una secuencia de señales que representan el programa informático para realizar uno de los procedimientos descritos en la presente memoria. El tren de datos o la secuencia de señales pueden, por ejemplo, estar configurados para ser transferidos por medio de una conexión de comunicación de datos, por ejemplo, por Internet.

25 **[0183]** Una realización adicional comprende un medio de tratamiento, por ejemplo, un ordenador, o un dispositivo lógico programable, configurado o adaptado para realizar uno de los procedimientos descritos en la presente memoria.

[0184] Una realización adicional comprende un ordenador que tiene instalado el programa informático para realizar uno de los procedimientos descritos en la presente memoria.

30 **[0185]** En algunas realizaciones, se puede usar un dispositivo lógico programable (por ejemplo, una matriz de puertas programables por campo) para realizar algunas o todas las funcionalidades de los procedimientos descritos en la presente memoria. En algunas realizaciones, una matriz de puertas programables por campo puede cooperar con un microprocesador con el fin de realizar uno de los procedimientos descritos en la presente memoria. Generalmente, los procedimientos se realizan preferentemente mediante cualquier aparato de hardware.

35 **[0186]** Las realizaciones descritas anteriormente son meramente ilustrativas de los principios de la presente invención. Se entiende que las modificaciones y variaciones de las disposiciones y los detalles descritos en la presente memoria serán evidentes para otros expertos en la materia. Por lo tanto, se pretende estar solo limitados por el alcance de las reivindicaciones de patente que se ofrecen a continuación y no por los detalles específicos presentados a modo de descripción y explicación de las realizaciones en la presente memoria.

40

REIVINDICACIONES

1. Aparato para estimar una diferencia de tiempos entre canales entre una señal de un primer canal y una señal de un segundo canal, que comprende:
 - un calculador (1020) para calcular un espectro de correlación cruzada para un bloque de tiempo de la señal del primer canal en el bloque de tiempo y de la señal del segundo canal en el bloque de tiempo;
 - un estimador de características espectrales (1010) para estimar una característica de un espectro de la señal del primer canal o de la señal del segundo canal para el bloque de tiempo;
 - un filtro de suavizado (1030) para suavizar el espectro de correlación cruzada a lo largo del tiempo usando la característica espectral para obtener un espectro de correlación cruzada suavizado; y
 - un procesador (1040) para tratar el espectro de correlación cruzada suavizado para obtener la diferencia de tiempos entre canales.
2. Aparato según la reivindicación 1, en el que el procesador (1040) está configurado para normalizar (456) el espectro de correlación cruzada suavizado usando una magnitud del espectro de correlación cruzada suavizado.
3. Aparato según la reivindicación 1 o 2, en el que el procesador (1040) está configurado para calcular (1031) una representación en el dominio del tiempo del espectro de correlación cruzada suavizado o un espectro de correlación cruzada suavizado normalizado; y para analizar (1032) la representación en el dominio del tiempo para determinar la diferencia de tiempos entre canales.
4. Aparato según una de las reivindicaciones anteriores, en el que el procesador (1040) está configurado para filtrar con paso bajo (458) la representación en el dominio del tiempo y para tratar (1033) adicionalmente un resultado del filtrado de paso bajo.
5. Aparato según una de las reivindicaciones anteriores, en el que el procesador está configurado para realizar la determinación de la diferencia de tiempos entre canales al realizar una operación de búsqueda de picos o de selección de picos dentro de una representación en el dominio del tiempo determinada a partir del espectro de correlación cruzada suavizado.
6. Aparato según una de las reivindicaciones anteriores, en el que el estimador de características espectrales (1010) está configurado para determinar, como característica espectral, una característica de ruido o una tonalidad del espectro; y en el que el filtro de suavizado (1030) está configurado para aplicar un suavizado más intenso a lo largo del tiempo con un primer grado de suavizado en el caso de una primera característica menos ruidosa o una primera característica más tonal, o para aplicar un suavizado más débil a lo largo del tiempo con un segundo grado de suavizado en el caso de una segunda característica más ruidosa o una segunda característica menos tonal, en el que el primer grado de suavizado es mayor que el segundo grado de suavizado, y en el que la primera característica ruidosa es menos ruidosa que la segunda característica ruidosa, o la primera característica tonal es más tonal que la segunda característica tonal.
7. Aparato según una de las reivindicaciones anteriores, en el que el estimador de características espectrales (1010) está configurado para calcular, como característica, una primera medida de uniformidad espectral de un espectro de la señal del primer canal y una segunda medida de uniformidad espectral de un segundo espectro de la señal del segundo canal, y para determinar la característica del espectro a partir de la primera y la segunda medida de uniformidad espectral mediante la selección de un valor máximo, mediante la determinación de un promedio ponderado o un promedio no ponderado entre las medidas de uniformidad espectral, o mediante la selección de un valor mínimo.
8. Aparato según una de las reivindicaciones anteriores, en el que el filtro de suavizado (1030) está configurado para calcular un valor del espectro de correlación cruzada suavizado para una frecuencia mediante una combinación ponderada del valor del espectro de correlación cruzada para la frecuencia del bloque de tiempo y un valor espectral de correlación cruzada para la frecuencia de al menos un bloque de tiempo pasado, en el que los factores de ponderación para la combinación ponderada se determinan mediante las características del espectro.
9. Aparato según una de las reivindicaciones anteriores, en el que el procesador (1040) está configurado

para determinar un intervalo válido y un intervalo inválido dentro de una representación en el dominio del tiempo obtenida del espectro de correlación cruzada suavizado, en el que al menos un pico máximo dentro del intervalo inválido es detectado y comparado con un pico máximo dentro del intervalo válido, en el que solo se determina la diferencia de tiempos entre canales, cuando el pico 5 máximo dentro del intervalo válido es mayor que al menos un pico máximo dentro del intervalo inválido.

10. Aparato según una de las reivindicaciones anteriores, en el que el procesador (1040) está configurado para realizar una operación de búsqueda de picos dentro de una representación en el dominio del tiempo obtenida 10 del espectro de correlación cruzada suavizado, para determinar (1034) un umbral variable a partir de la representación en el dominio del tiempo; y para comparar (1035) un pico con el umbral variable, en el que la diferencia de tiempos entre canales se determina como un retardo de tiempo asociado con un pico que está en una relación predeterminada con el umbral variable.

15 11. Aparato según la reivindicación 10, en el que el procesador está configurado para determinar el umbral variable (1334c) como un valor que es igual a un múltiplo entero de un valor entre el 10% más grande de los valores de la representación en el dominio del tiempo.

12. Aparato según una de las reivindicaciones 1 a 9, 20 en el que el procesador (1040) está configurado para determinar una amplitud de pico máxima (1102) en cada subbloque de una pluralidad de subbloques de una representación en el dominio del tiempo obtenida del espectro de correlación cruzada suavizado, en el que el procesador (1040) está configurado para calcular (1104, 1105) un umbral variable basado en una magnitud de pico media obtenida de las magnitudes máximas de pico de la pluralidad de subbloques, y 25 en el que el procesador está configurado para determinar la diferencia de tiempos entre canales como un valor de retardo de tiempo correspondiente a un pico máximo de la pluralidad de subbloques que es mayor que el umbral variable.

13. Aparato según la reivindicación 12, 30 en el que el procesador (1040) está configurado para calcular el umbral variable mediante la multiplicación (1105) del umbral promedio determinado como un pico promedio entre los picos en los subbloques y un valor, en el que el valor se determina (1104) mediante una característica SNR (relación señal-ruido) de la señal del primer canal y del segundo canal, en el que un primer valor está asociado con un primer valor SNR y un segundo valor está asociado con un segundo valor SNR, en el que el primer valor es mayor que el segundo valor, y en el que el primer 35 valor SNR es mayor que el segundo valor SNR.

14. Aparato según la reivindicación 13, en el que el procesador (1040) está configurado para usar (1104) un tercer valor ($\epsilon_{\text{mínimo}}$) que es menor que el segundo valor (ϵ_{bajo}) en el caso en que un tercer valor SNR sea menor que el segundo valor SNR y cuando una 40 diferencia entre el umbral y un pico máximo es menor que un valor predeterminado (ϵ).

15. Procedimiento para estimar una diferencia de tiempo entre canales entre una señal del primer canal y una señal del segundo canal, que comprende:

45 cálculo (1020) de un espectro de correlación cruzada para un bloque de tiempo de la señal del primer canal en el bloque de tiempo y la señal del segundo canal en el bloque de tiempo; estimación (1010) de una característica de un espectro de la señal del primer canal o de la señal del segundo canal para el bloque de tiempo; suavizado (1030) del espectro de correlación cruzada a lo largo del tiempo usando la característica espectral 50 para obtener un espectro de correlación cruzada suavizado; y tratamiento (1040) del espectro de correlación cruzada suavizado para obtener la diferencia de tiempos entre canales.

16. Programa informático para realizar, cuando se ejecuta en un ordenador o un procesador, el 55 procedimiento de la reivindicación 15.

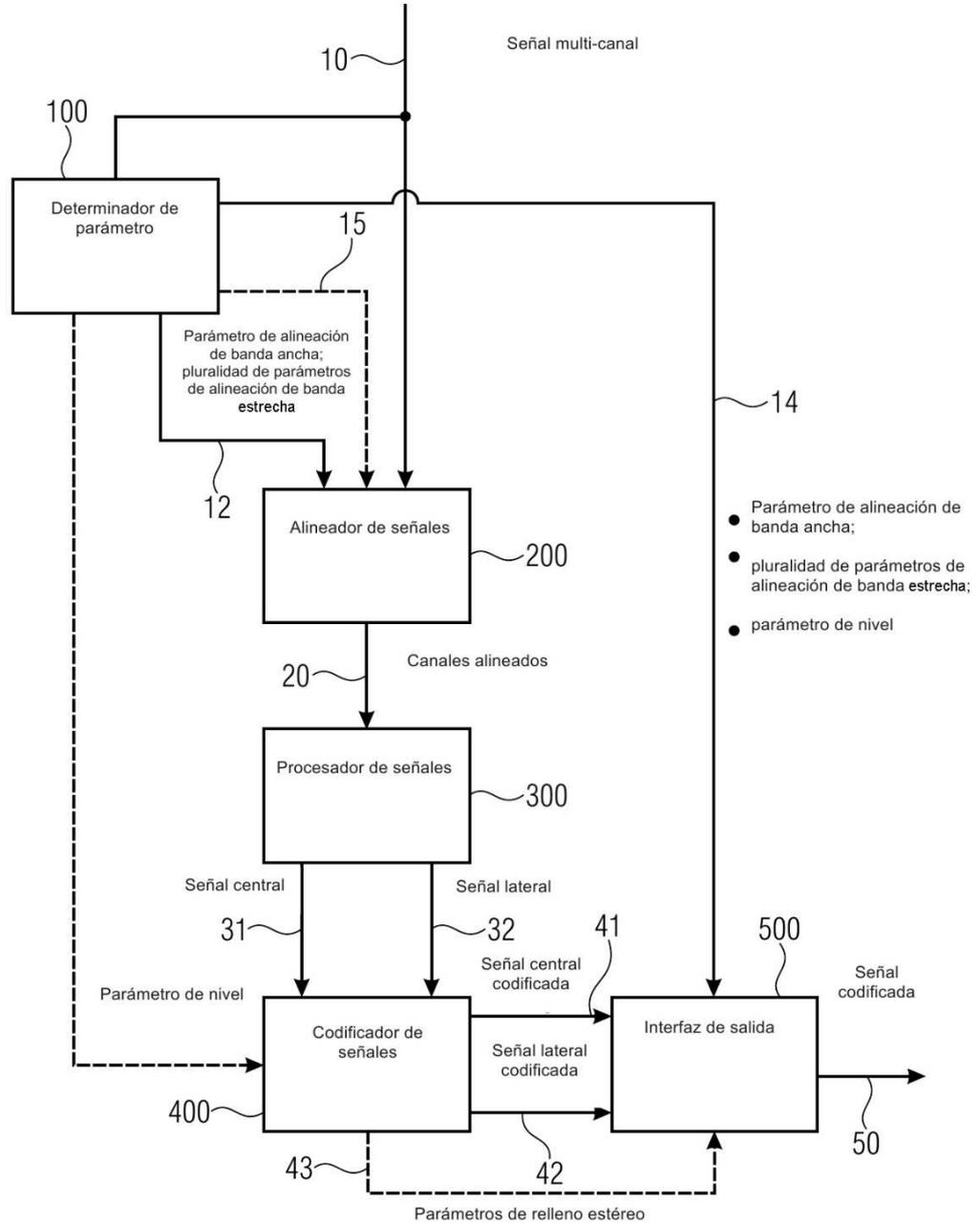


Fig. 1

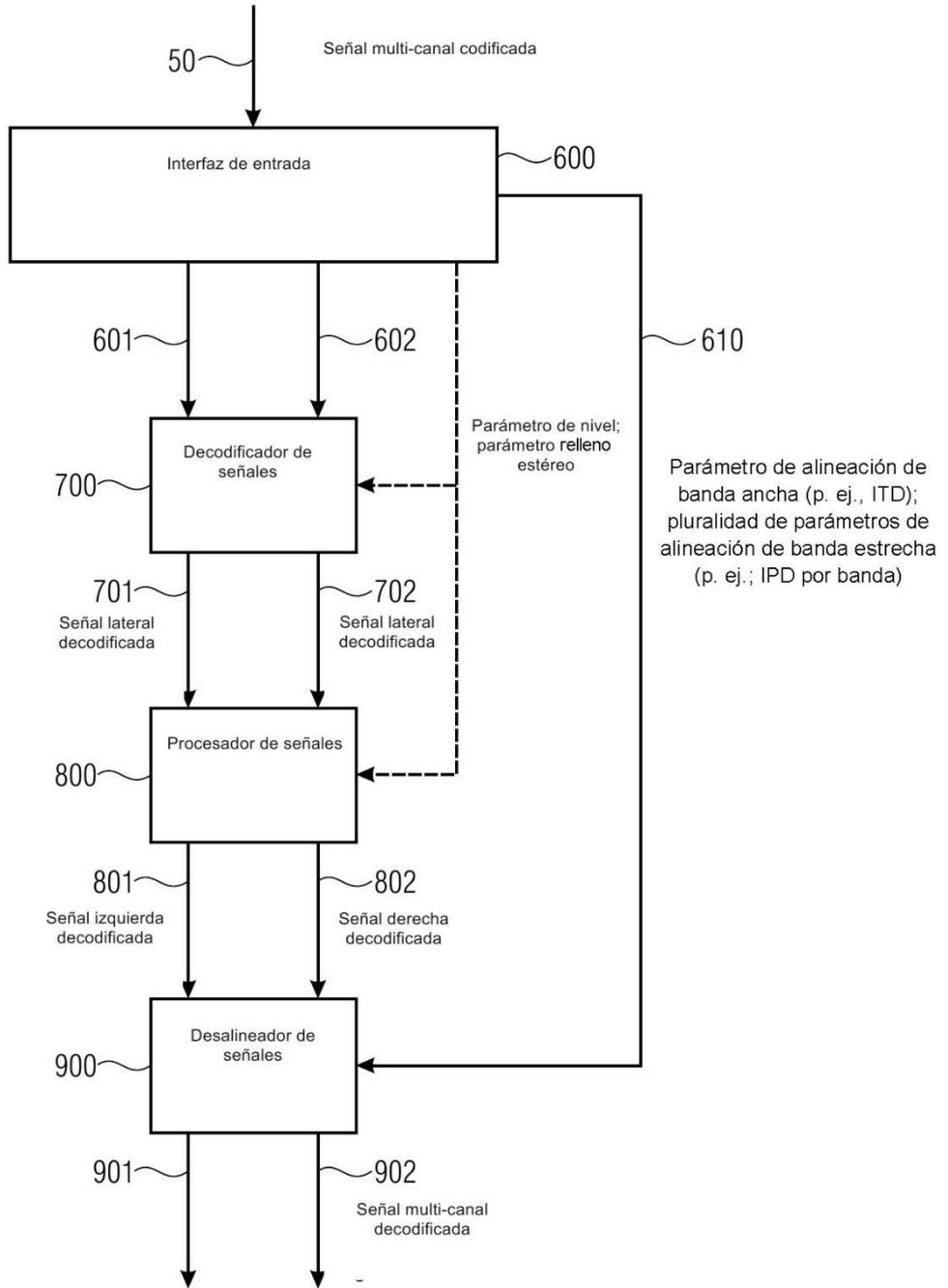
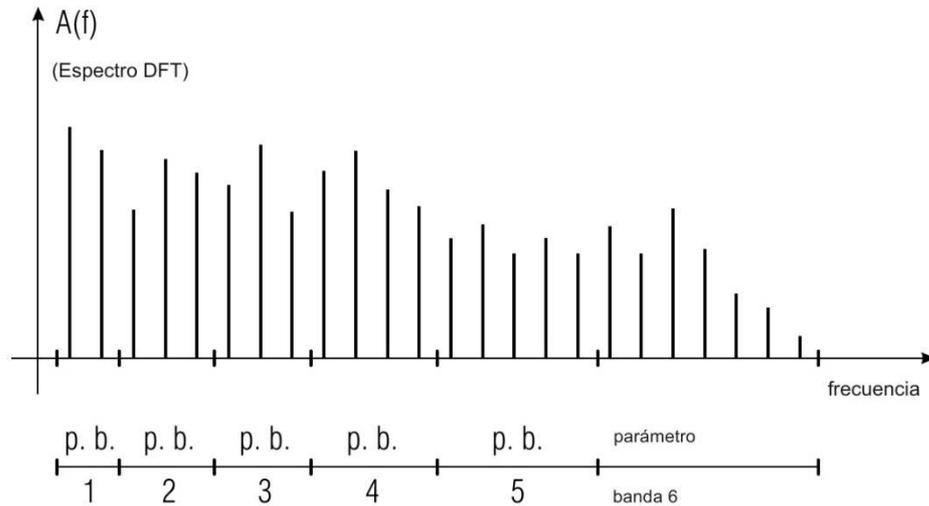


Fig. 2



- Un solo parámetro de alineación de banda ancha para todo el espectro (por ejemplo, banda 1 a banda 6);
- pluralidad de parámetros de alineación de banda estrecha para las bandas de parámetros 1, 2, 3, 4, es decir, cuatro parámetros de banda estrecha;
- parámetros de nivel para cada banda de parámetros, por ejemplo, 6 parámetros de nivel;
- parámetros de relleno estéreo para las bandas de parámetros 4, 5, 6, por ejemplo, tres parámetros de relleno estéreo;
- señal lateral (residual) para las bandas de parámetros 1, 2, 3;
- más líneas espectrales en las bandas altas, por ejemplo, siete líneas espectrales en la banda de parámetros 6 frente a tres líneas espectrales en la banda de parámetros 2.

Fig. 3

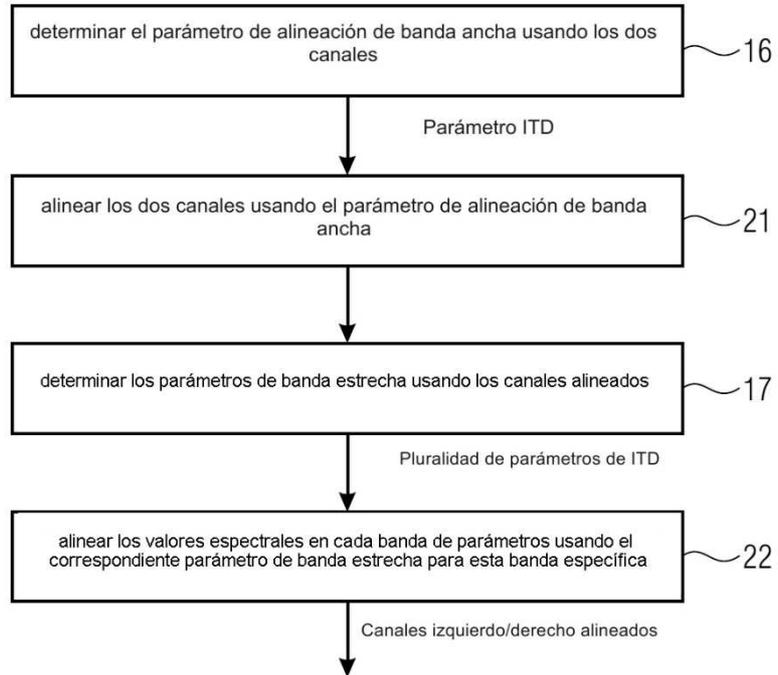


Fig. 4a

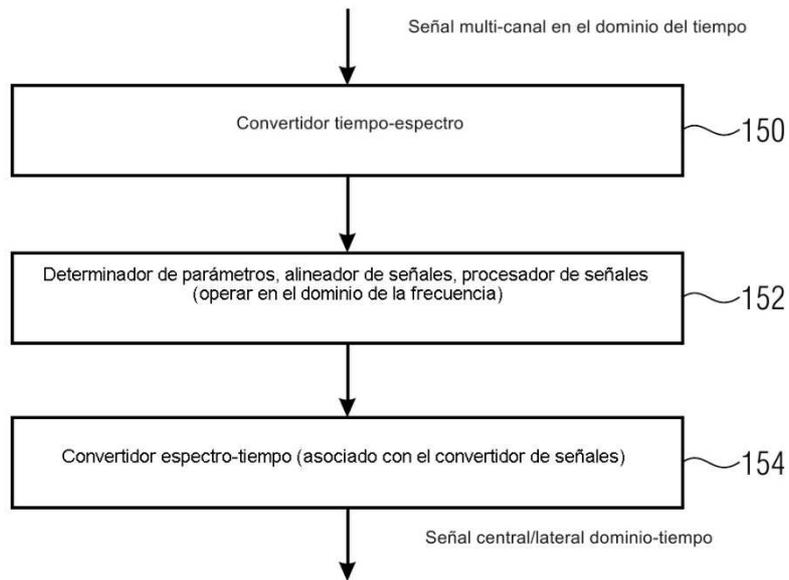


Fig. 4b

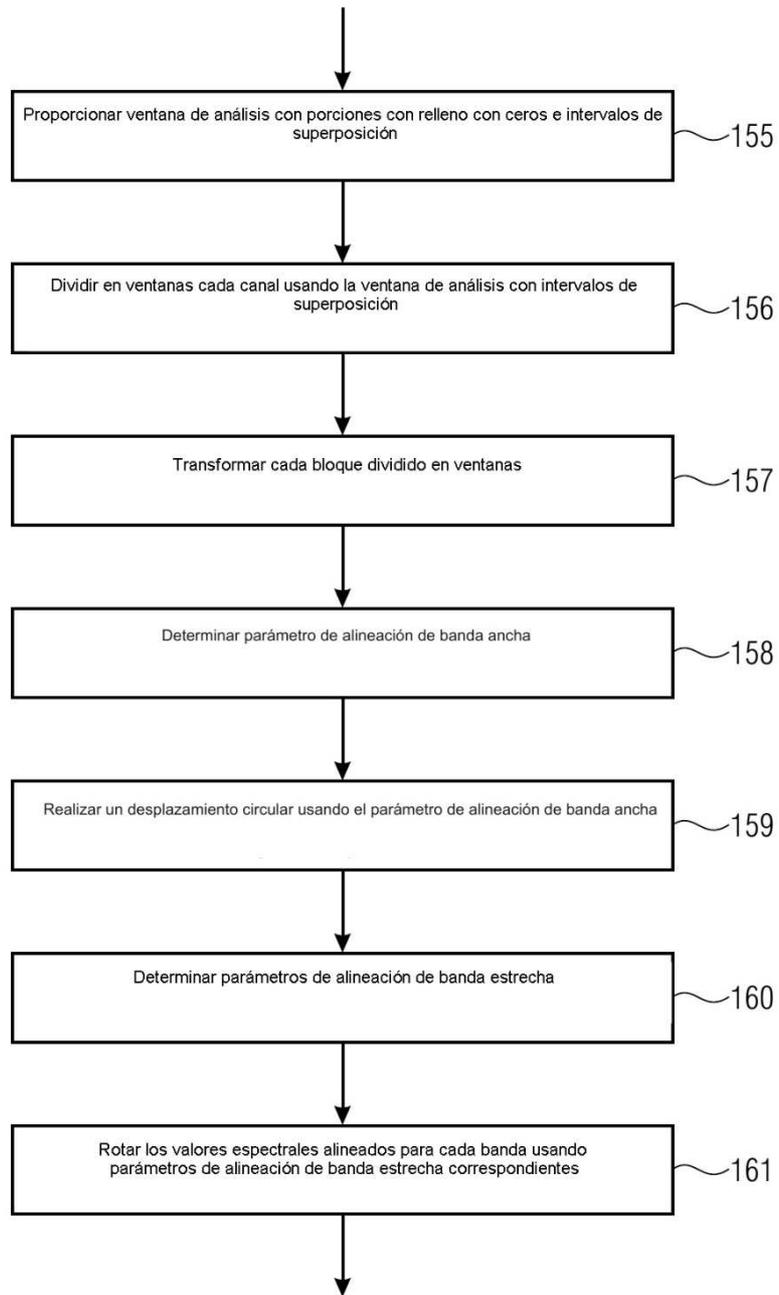


Fig. 4c

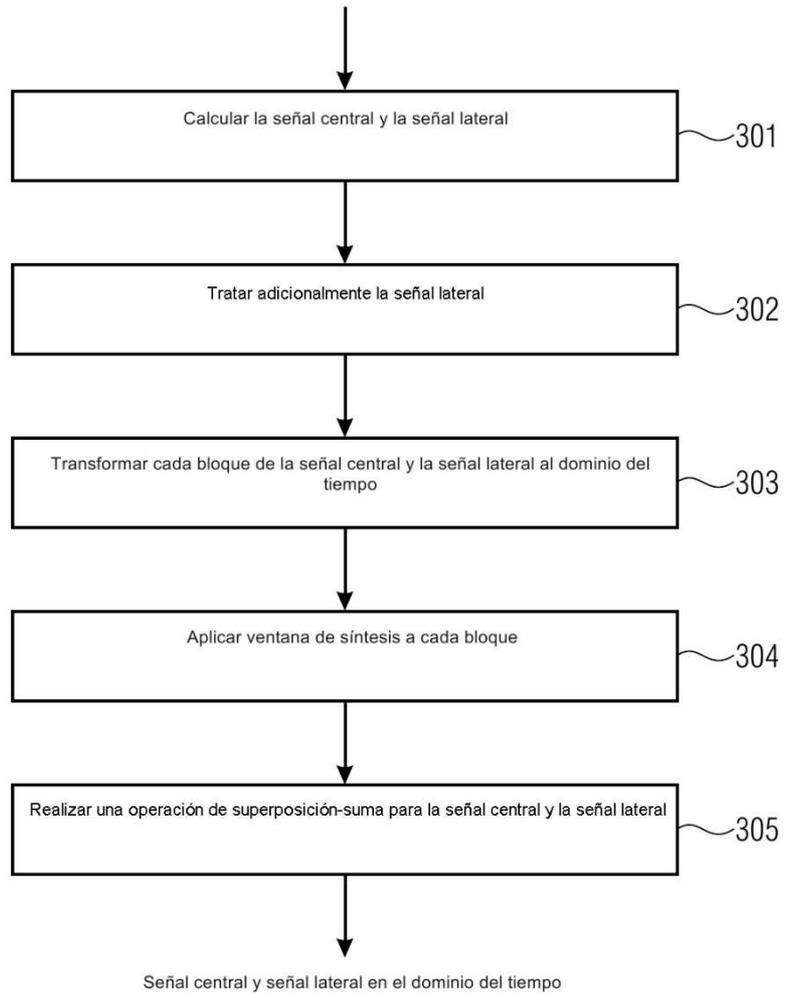
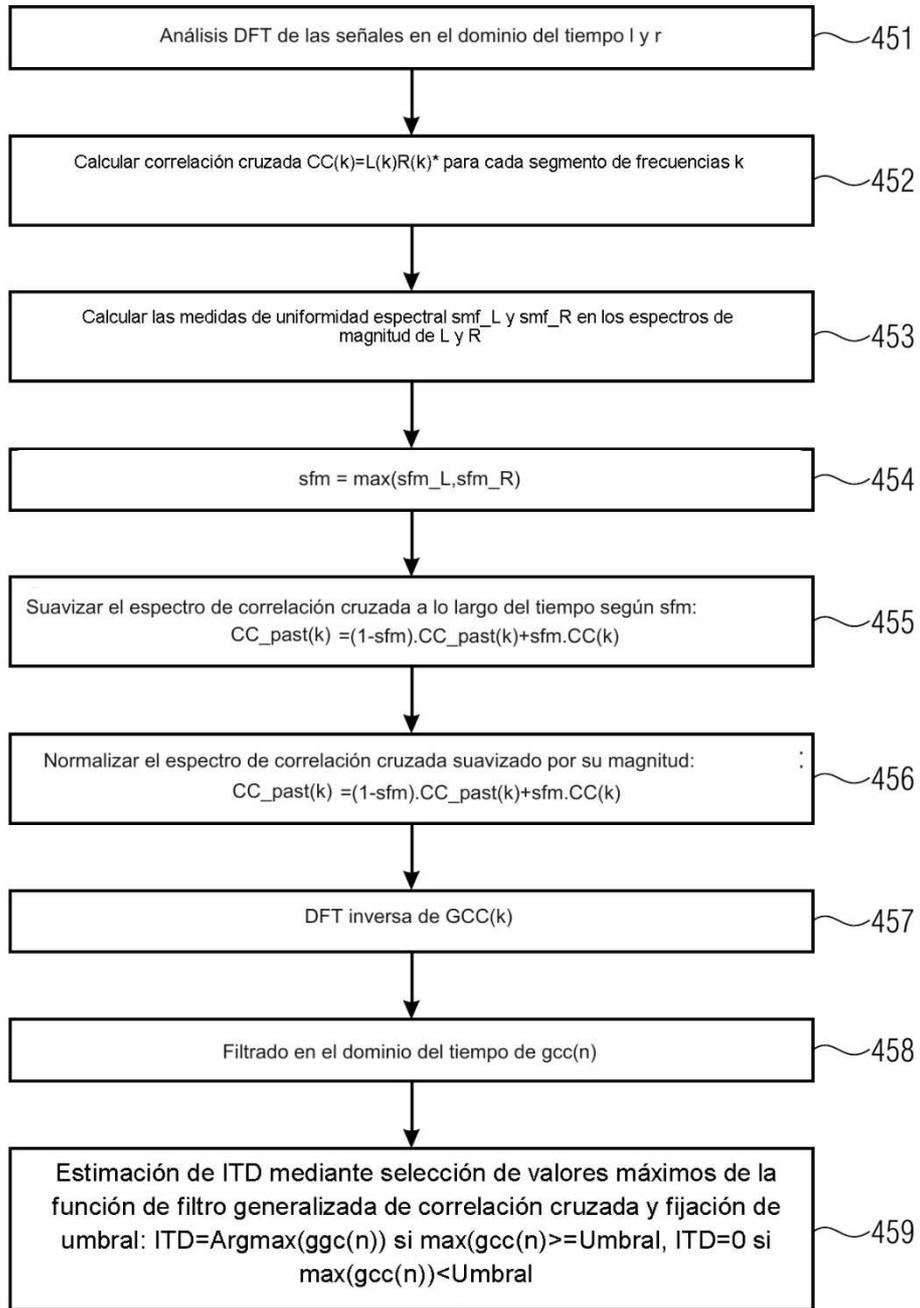


Fig. 4d



Estimación de ITD

Fig. 4e

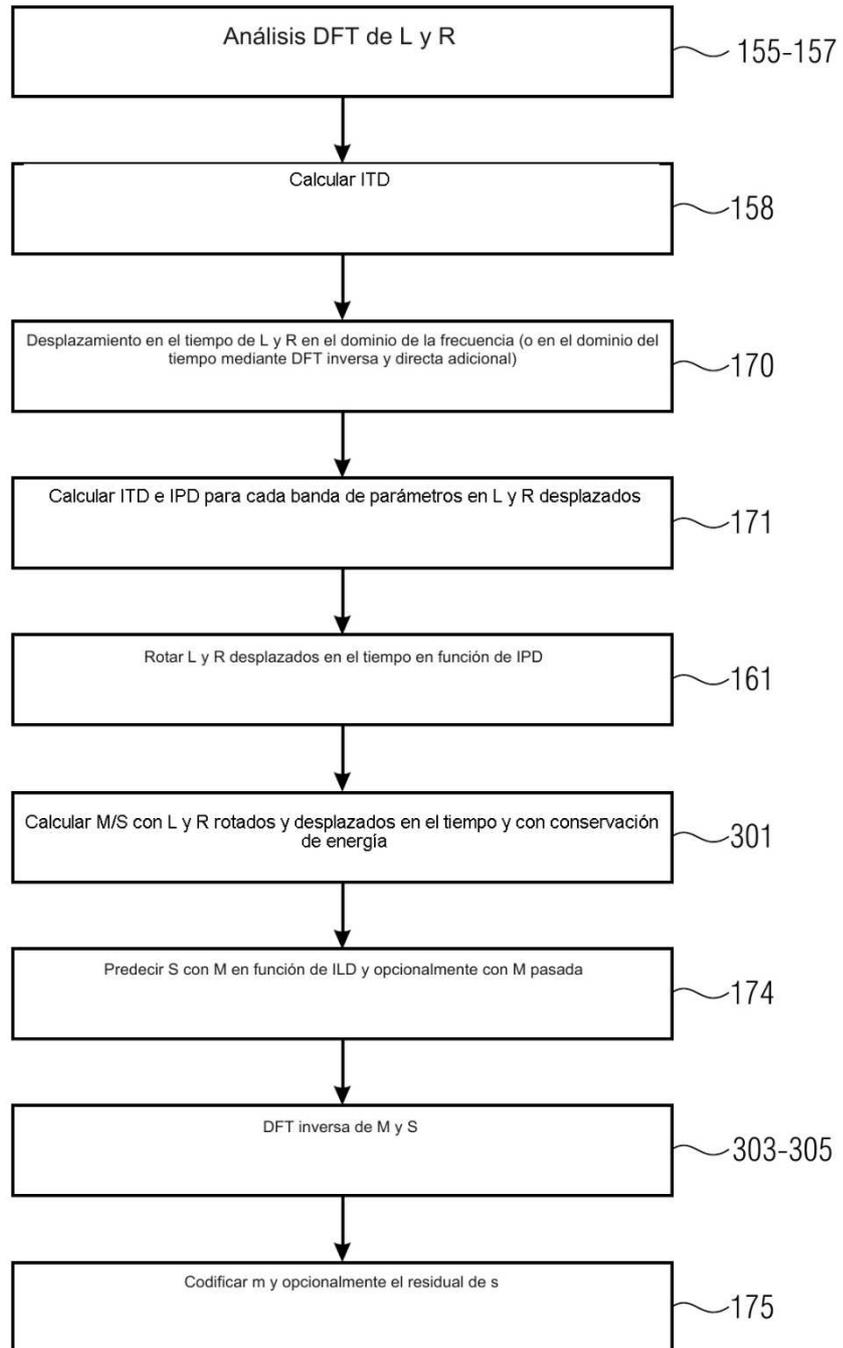


Fig. 5

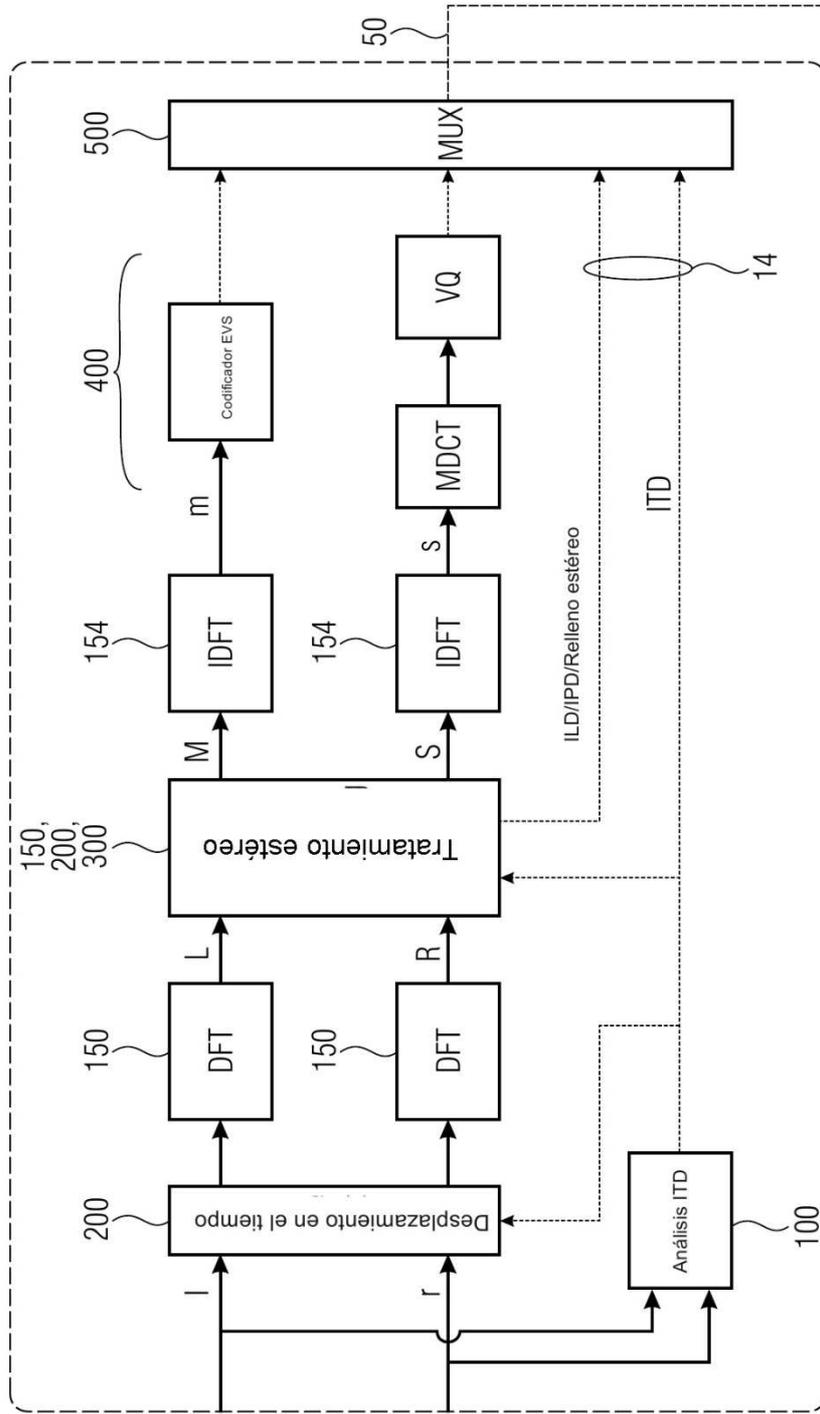


Fig. 6a

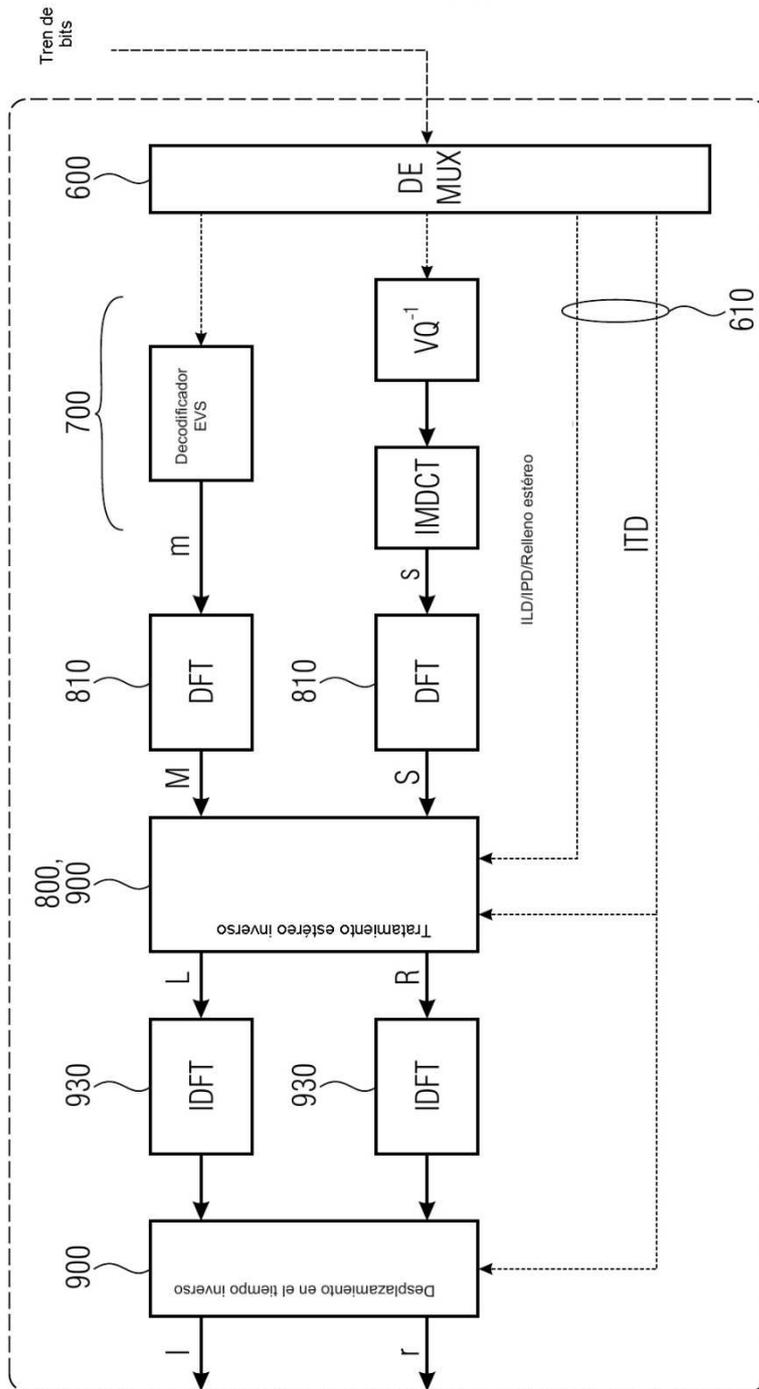


Fig. 6b

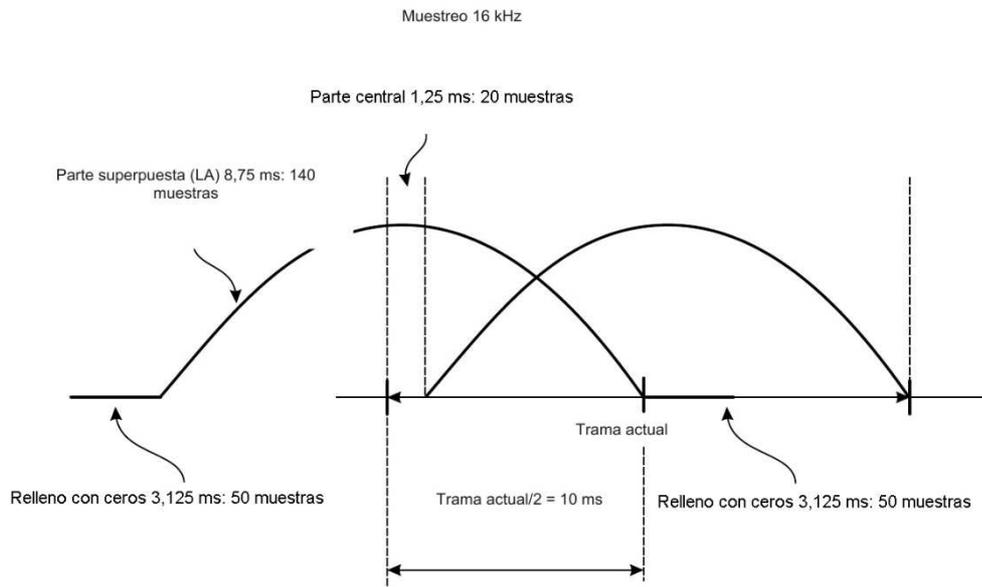


Fig. 7

Parámetros	Bits/banda	Número de bandas	Velocidad binaria
ILD	5	El total de 12 bandas	3,00 kbps
IPD	3	Hasta 2,5kHz	1,05 kbps
ITD	8	Espectro completo	0,6 kbps
Relleno estéreo	3	Desde 1kHz	0,9 kbps
total			-5 kbps

Fig. 8

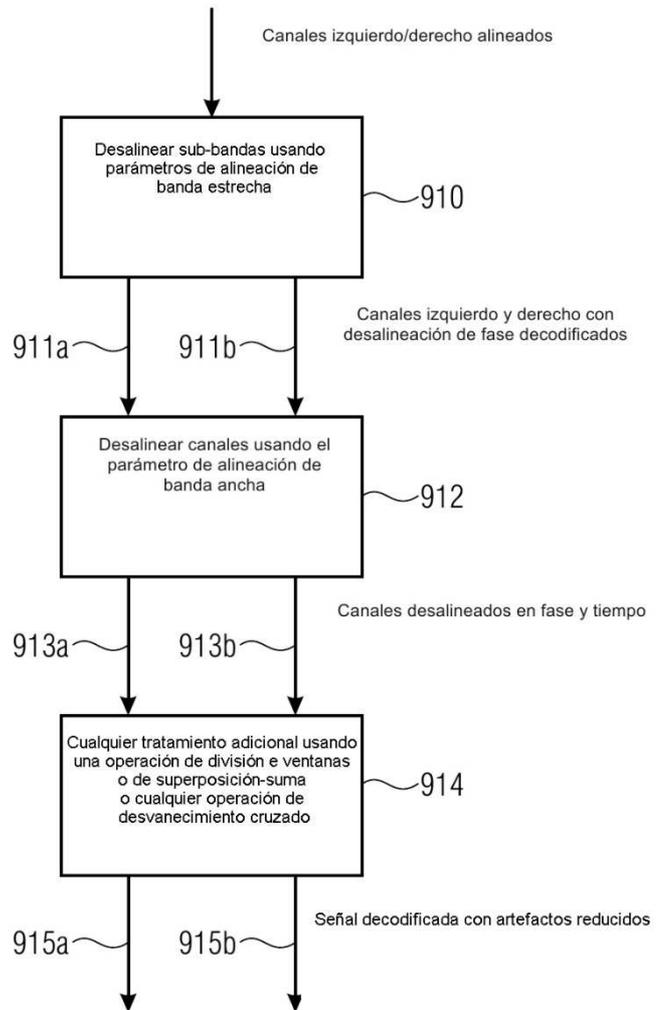


Fig. 9a

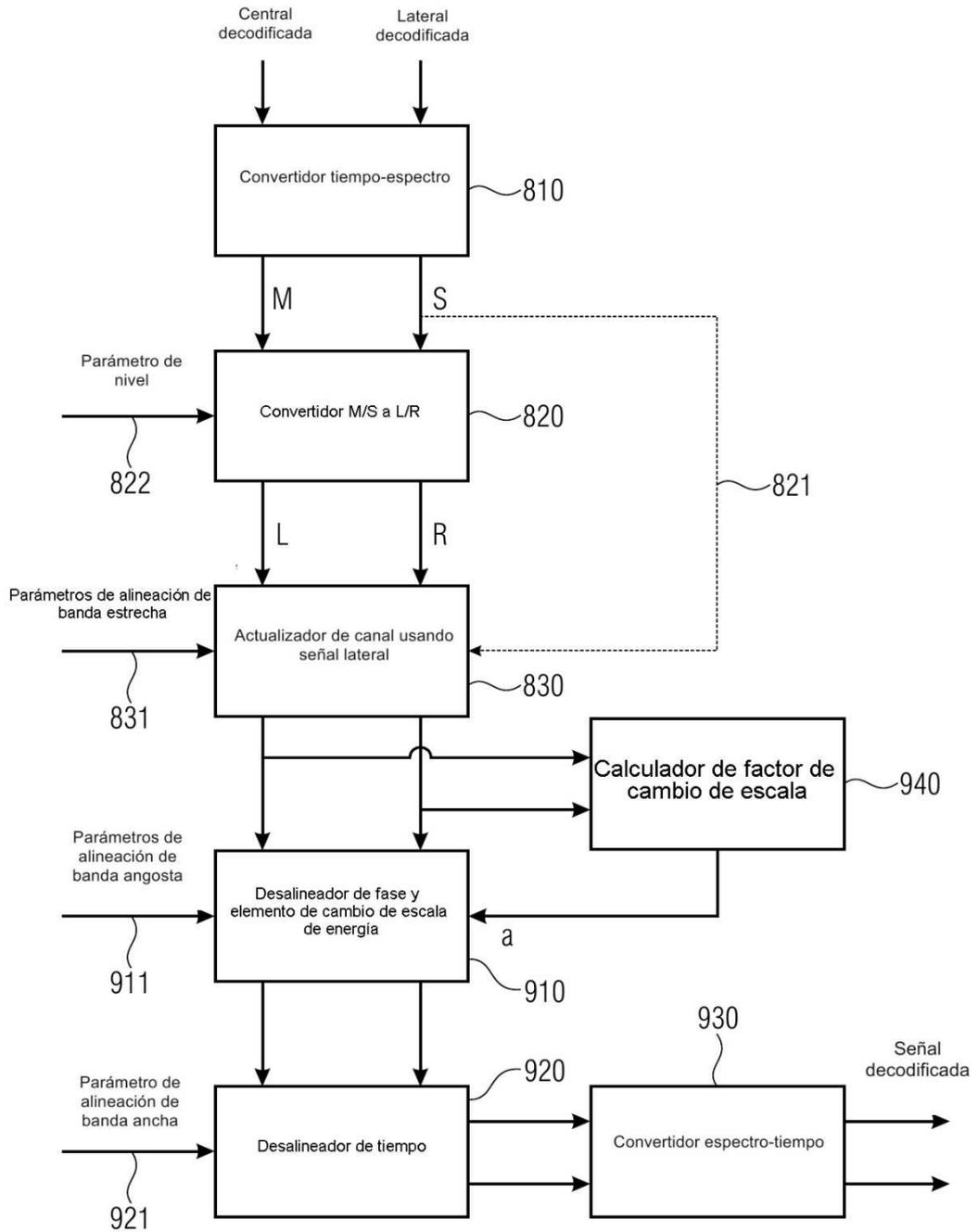


Fig. 9b

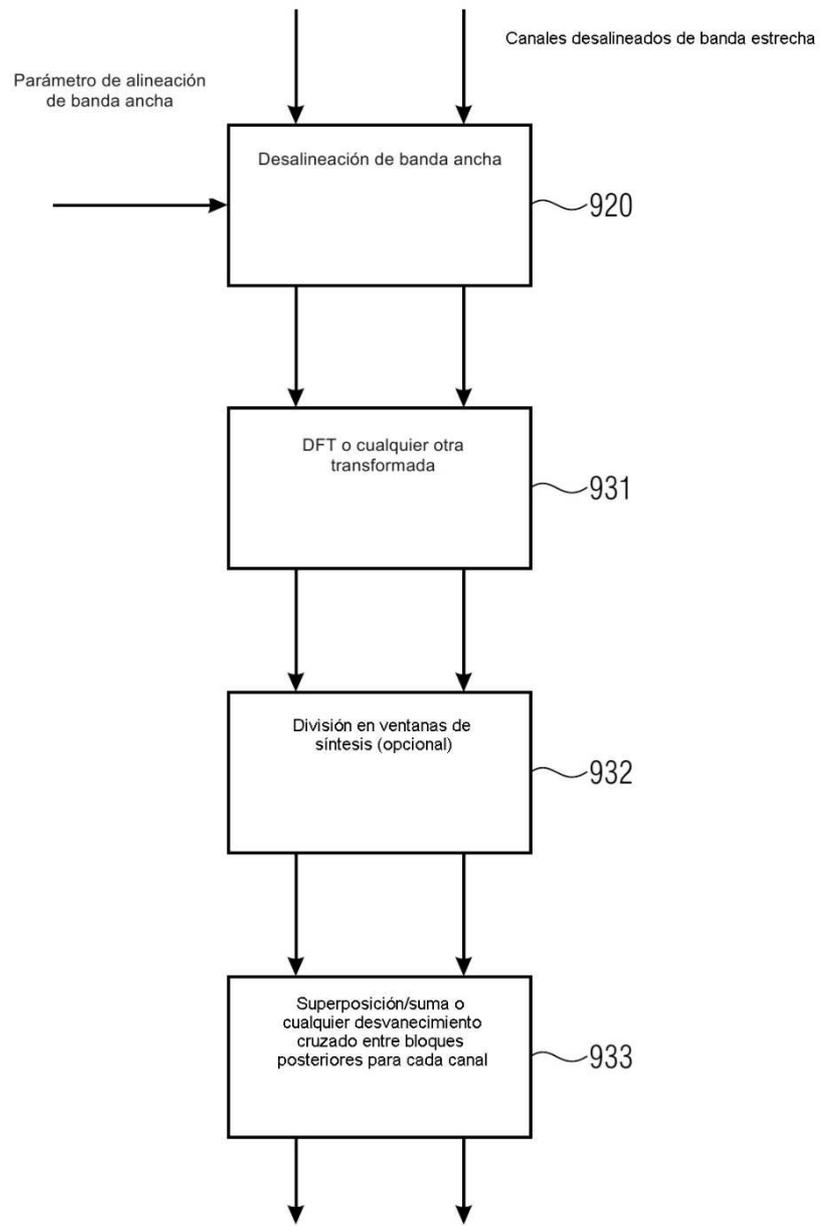


Fig. 9c

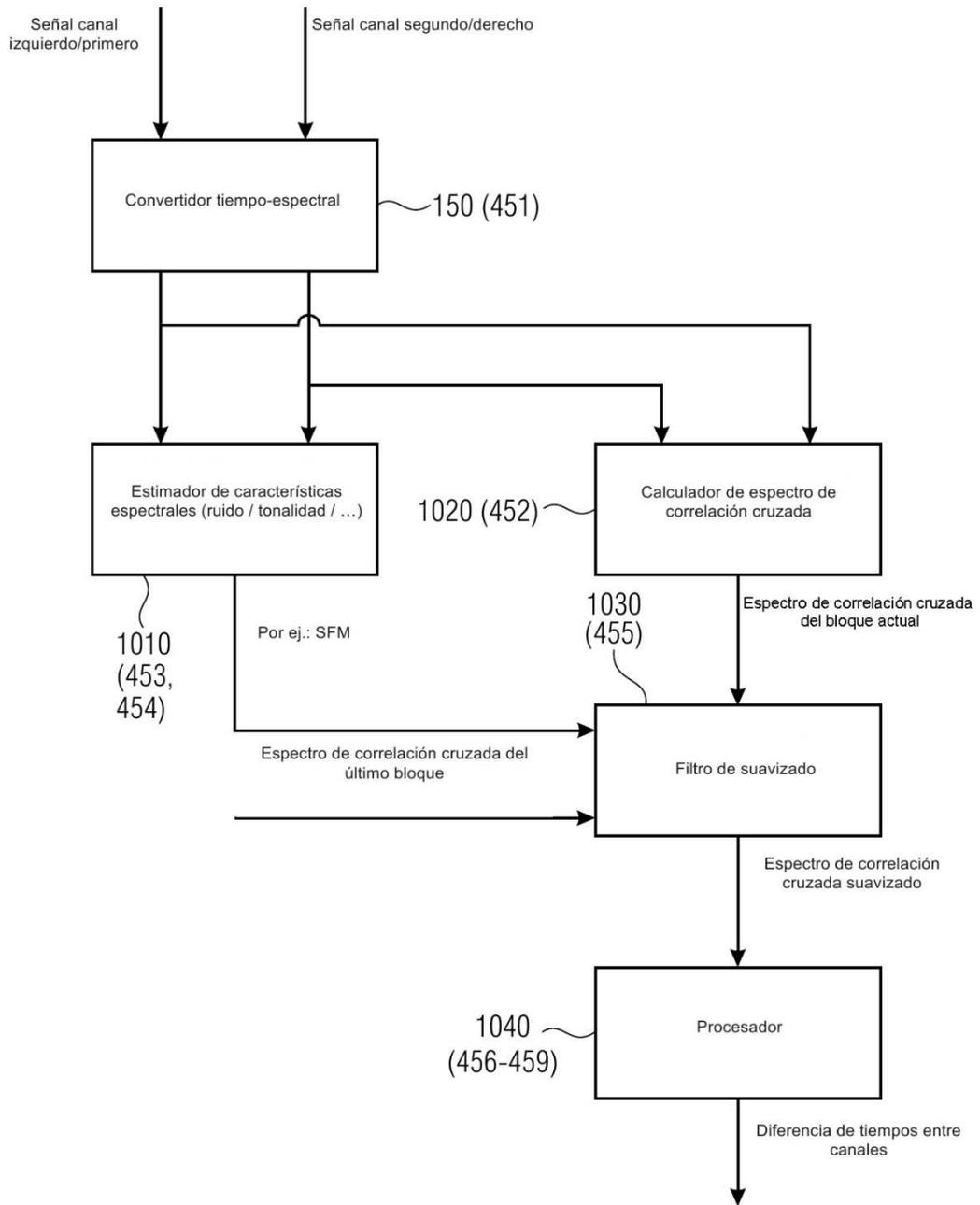
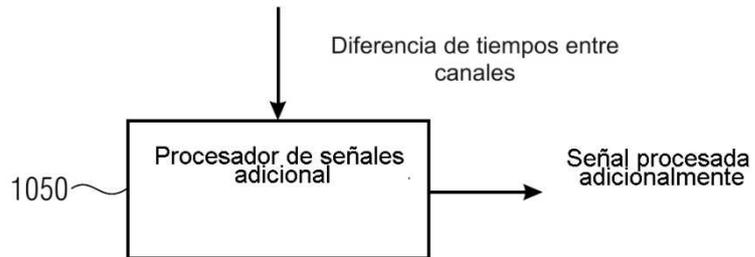


Fig. 10a



- Almacenamiento/transmisión de datos paramétricos
- Tratamiento/codificación estéreo multicanal
- Alineación en el tiempo de los dos canales
- Diferencia de tiempo de estimación de llegada para determinar la posición del altavoz en la sala con dos micrófonos y una configuración de micrófonos conocida
- Formación de haces
- Filtrado espacial
- Descomposición primer plano/segundo plano
- Ubicación de la fuente de sonido, por ejemplo, mediante triangulación acústica sobre la base de diferencias de tiempos de dos o tres señales.

Fig. 10b

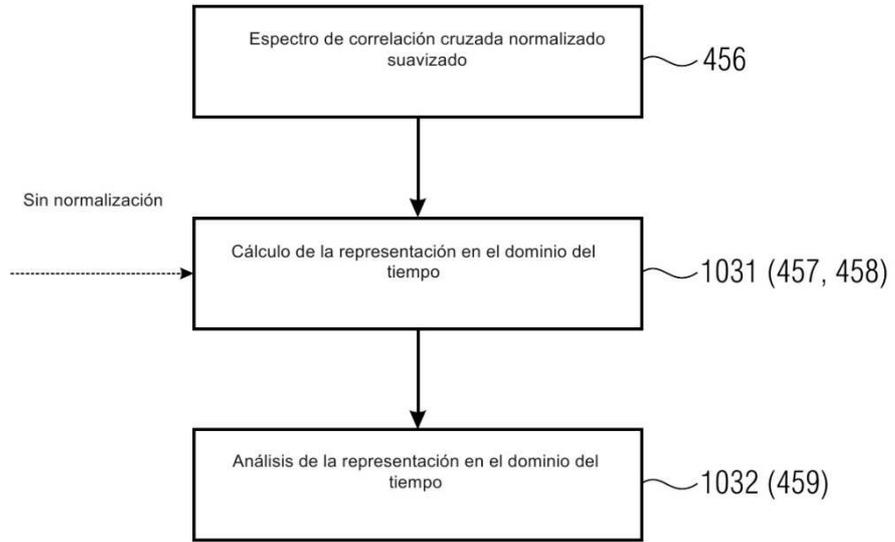


Fig. 11a

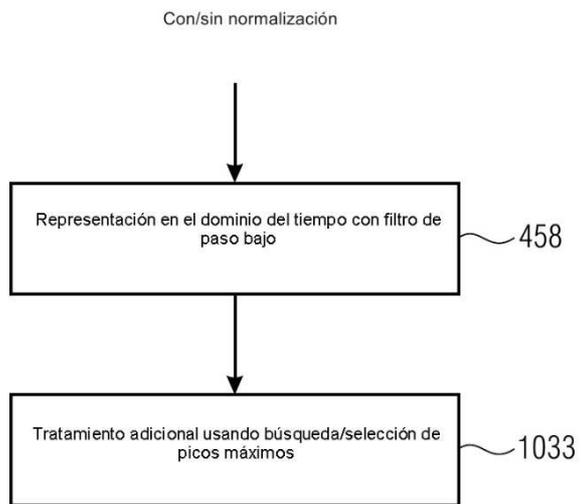


Fig. 11b

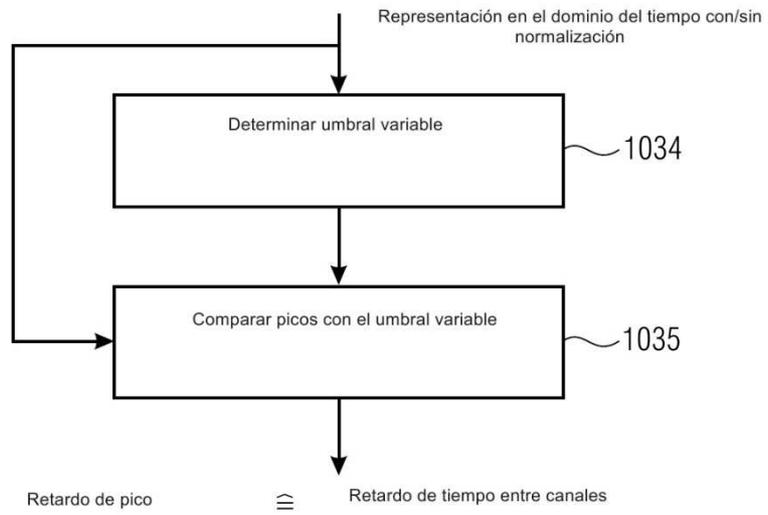


Fig. 11c

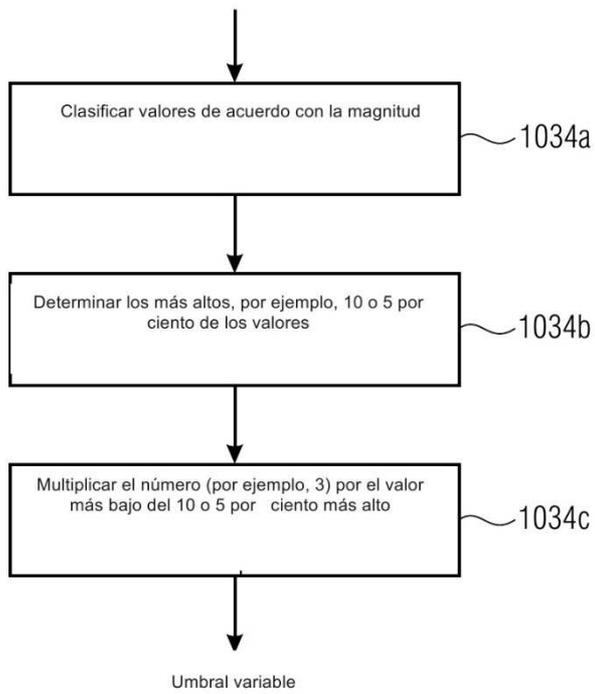


Fig. 11d

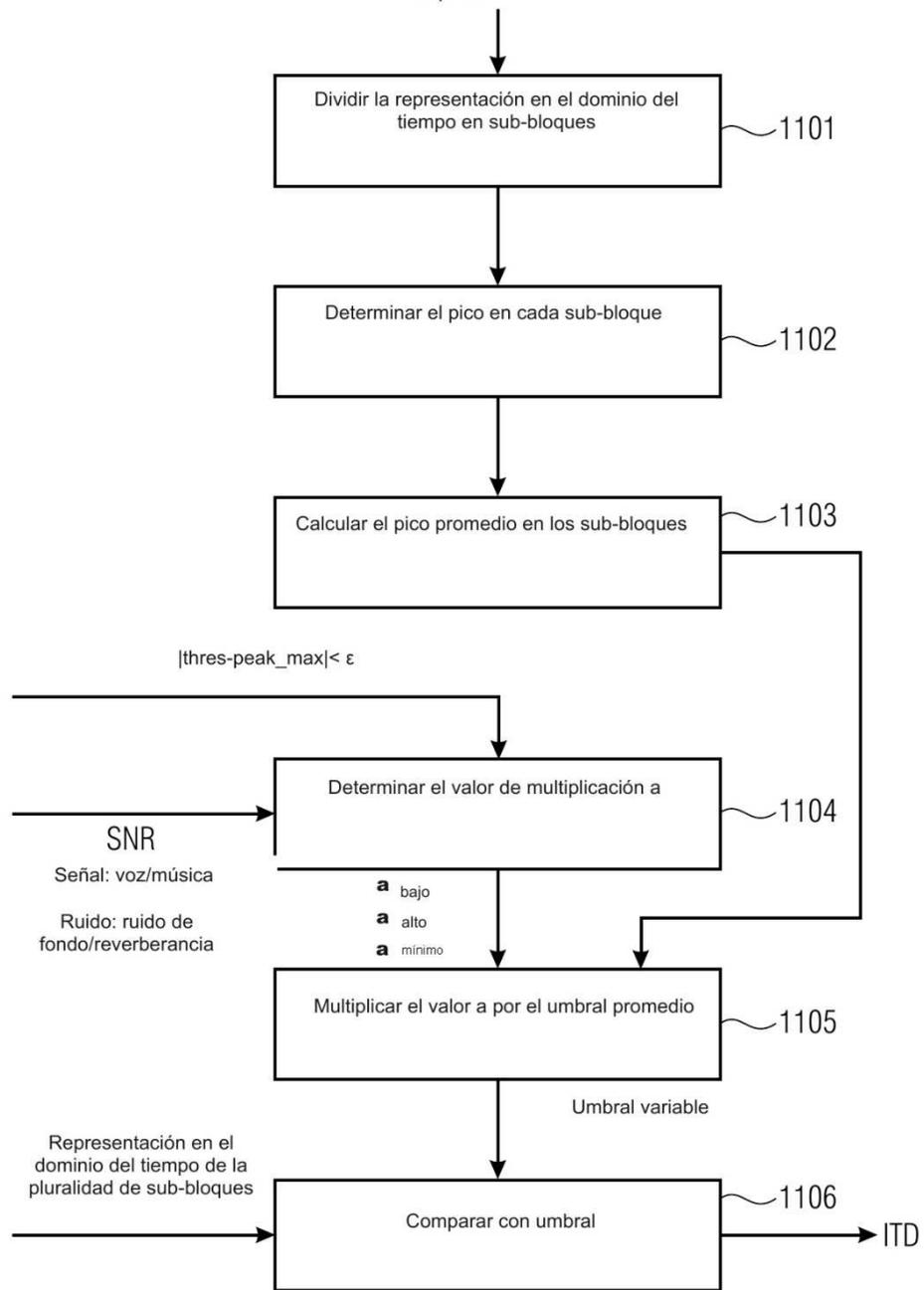


Fig. 11e

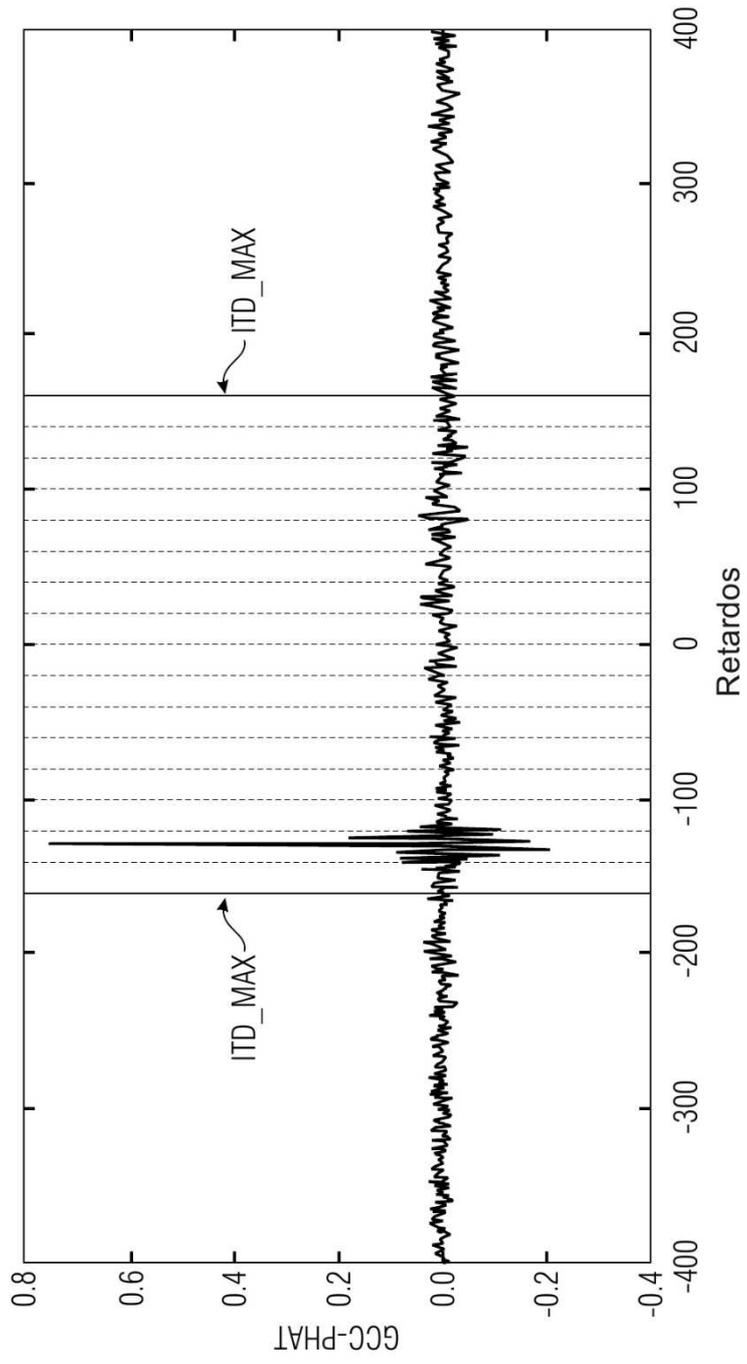


Fig. 12

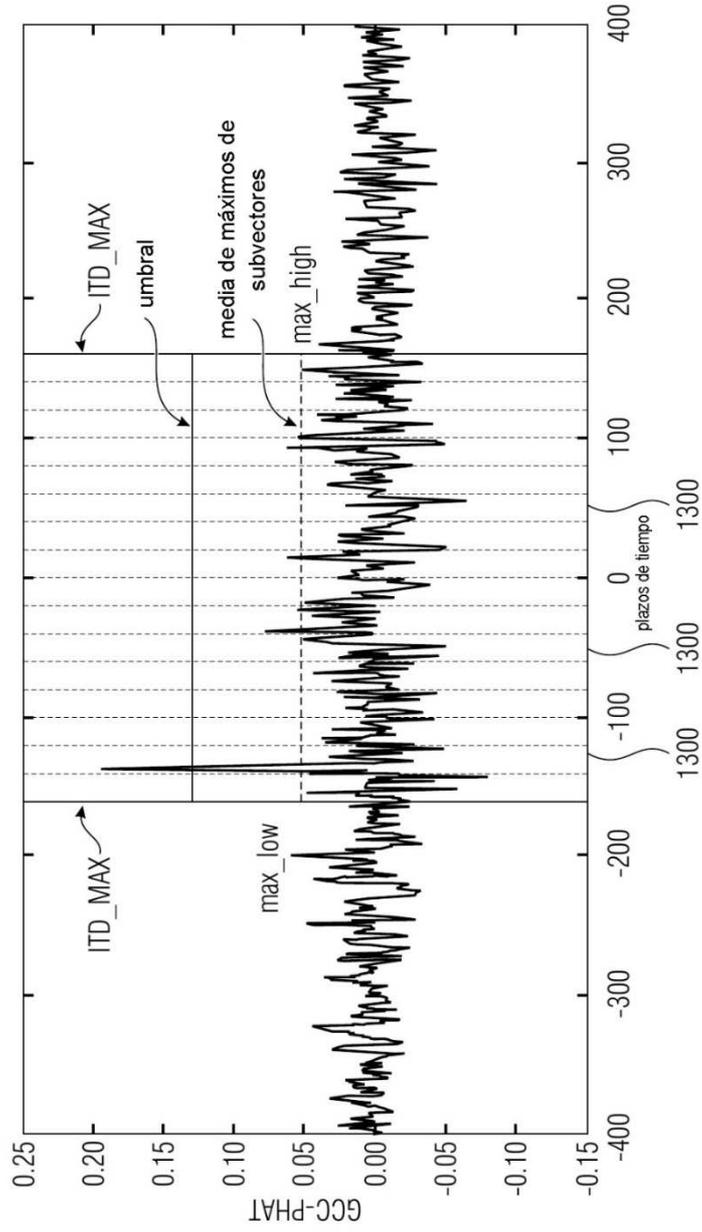


Fig. 13