(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 988 345**

(51) Int. Cl.:

G10L 19/008 (2013.01)
G10L 25/06 (2013.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Fecha de presentación y número de la solicitud europea: **03.04.2019 E 21212592 (6)**

(97) Fecha y número de publicación de la concesión europea: **21.08.2024 EP 3985665**

(54) Título: **Aparato, método o programa informático para estimar una diferencia de tiempo entre canales**

(30) Prioridad:

05.04.2018 EP 18165882

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
20.11.2024

(73) Titular/es:

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

(72) Inventor/es:

FOTOPOULOU, ELENI;
BÜTHE, JAN;
RAVELLI, EMMANUEL;
MABEN, PALLAVI;
DIETZ, MARTIN;
REUTELHUBER, FRANZ;
DÖHLA, STEFAN y
KORSE, SRIKANTH

(74) Agente/Representante:

ARIZTI ACHA, Monica

DESCRIPCIÓN

Aparato, método o programa informático para estimar una diferencia de tiempo entre canales

La presente solicitud se refiere al procesamiento estéreo o, en términos generales, el procesamiento multicanal, donde 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, como tres, cuatro, cinco o cualquier otro número de canales.

La voz estéreo y, en particular, la voz de conversación estéreo ha recibido mucha menos atención científica que el almacenamiento y transmisión de música estereofónica. En efecto, en las comunicaciones de voz, hoy en día aún se utiliza mayormente la transmisión monofónica. Sin embargo, con el aumento del ancho de banda y la capacidad de la red, se prevé que las comunicaciones basadas en tecnologías estereofónicas cobrarán más popularidad y brindarán una mejor experiencia de escucha.

Desde hace mucho tiempo se ha estudiado la codificación eficiente del material de audio estereofónico en la codificación de audio perceptual de música para el almacenamiento o transmisión eficiente. A altas tasas de bits, en las que es crucial preservar la forma de onda, se ha empleado desde hace tiempo el estéreo de suma–diferencia, que se conoce como estéreo medio lateral (M/S). En el caso de las bajas tasas de bits, se ha introducido el estéreo de intensidad y, más recientemente, la codificación paramétrica estéreo. La última técnica fue adoptada en diferentes normas tales como HeAACv2 y Mpeg USAC. Genera una mezcla descendente de la señal bicanal y asocia la información espacial lateral compacta.

La codificación conjunta estéreo se construye habitualmente en una resolución de alta frecuencia, es decir, la transformación de baja resolución temporal, de tiempo–frecuencia de la señal y por ello no es compatible con el bajo retardo y el procesamiento en el dominio del tiempo realizado en la mayoría de los codificadores de voz. Más aun, habitualmente la tasa de bits originada es alta.

Por otro lado, el estéreo paramétrico emplea un banco de filtros extra situados en el extremo anterior del codificador como preprocesador y en el extremo posterior del decodificador como postprocesador. Por lo tanto, se puede usar el estéreo paramétrico con codificadores de voz convencionales tales como ACELP, como se hace en MPEG USAC. Más aun, se puede lograr la parametrización de la escena de audio con una cantidad mínima de información lateral, lo que es adecuado para las bajas tasas de bits. Sin embargo, el estéreo paramétrico no está específicamente diseñado, por ejemplo en MPEG USAC, para el bajo retardo y no ofrece una calidad consistente para diferentes situaciones de conversación. En la representación paramétrica convencional de la escena espacial, la amplitud de la imagen estéreo es reproducida de manera artificial por un descorrelacionador aplicado a los dos canales sintetizados y controlado por parámetros de Coherencia Entre Canales (IC) computado y transmitido por el codificador. En la mayor parte de la voz estéreo, esta manera de ampliar la imagen estéreo no es apropiada para recrear el ambiente natural de la voz, que es un sonido bastante directo, ya que lo produce una única fuente situada en una posición específica en el espacio (en ocasiones con cierta reverberación del recinto). Por el contrario, los instrumentos musicales tienen una amplitud mucho más natural que la voz, que puede ser imitada mejor descorrelacionando los canales.

También se suscitan problemas cuando se registra la voz con micrófonos no coincidentes, como en la configuración A-B cuando los micrófonos están alejados uno de otro o para la grabación o renderización binaural. Se pueden contemplar esas situaciones para capturar la voz en teleconferencias o para crear una escena virtualmente de auditorio con altavoces distantes en la unidad de control multipunto. Entonces el momento de llegada de la señal es diferente de un canal a otro, a diferencia de las grabaciones realizadas con micrófonos coincidentes como X-Y (grabación de intensidad) o M-S (grabación Medio-Lateral). El cómputo de la coherencia de esos dos canales no alineados en el tiempo puede ser incorrectamente estimado, lo que hace que la síntesis de ambiente artificial falle.

Las referencias de la técnica anterior relacionadas con el procesamiento estéreo son la patente estadounidense 5.434.948 o la patente estadounidense 8.811.621.

El documento WO 2006/089570 A1 da a conocer un esquema de codificador/decodificador multicanal casi transparente o transparente. Un esquema de codificador/decodificador multicanal genera de manera adicional una señal residual de tipo forma de onda. Esta señal residual se transmite junto con uno o más parámetros multicanal a un decodificador. En contraste con el decodificador multicanal puramente paramétrico, el decodificador potenciado genera 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 filtro de análisis. Por consiguiente, por cada señal de subbanda, se calculan un valor de alineación y un valor de ganancia correspondiente a una subbanda. A continuación, se realiza esta alineación antes de continuar el procesamiento. En el lado del decodificador, se realiza un procesamiento de desalineación y de ganancia y luego las señales correspondientes son sintetizadas por un banco de filtro de síntesis a fin de generar una señal izquierda decodificada y una señal derecha decodificada.

En esas aplicaciones de procesamiento estéreo, el cálculo de una diferencia de tiempo entre canales entre una señal de un primer canal y una señal de un segundo canal es útil para ejecutar típicamente un procedimiento de alineación en el tiempo de la banda ancha. Sin embargo, sí existen otras aplicaciones para el uso de una diferencia de tiempo entre canales entre un primer canal y un segundo canal, donde estas aplicaciones están en almacenamientos o transmisión de datos paramétricos, comprendiendo el procesamiento estéreo/multicanal una alineación en el tiempo de dos canales, una diferencia de tiempo de estimación de llegada para una determinación de una posición de un altavoz en un recinto, filtrado espacial para formación de haces, descomposición de primer plano y fondo o la ubicación de una fuente de sonido, por ejemplo, mediante triangulación acústica, por nombrar solo algunas.

Para todas esas aplicaciones, es necesaria una determinación eficiente, precisa y concluyente de una diferencia de tiempo entre canales entre las señales de un primer y de un segundo canal.

Ya existen determinaciones conocidas con la denominación “GCC-PHAT” o, dicho de otro modo, correlación cruzada generalizada con la transformada de fases. Por lo general, se calcula un espectro de correlación cruzada entre las señales de los dos canales y luego se aplica una función de ponderación al espectro de correlación cruzada para obtener un denominado espectro de correlación cruzada generalizada antes de ejecutar una transformada espectral inversa tal como una DFT inversa al espectro de correlación cruzada generalizada para encontrar una representación en el dominio del tiempo. Esta representación en el dominio del tiempo representa los valores correspondientes a ciertos retardos de tiempo y el pico más alto de la representación en el dominio del tiempo corresponde entonces, por lo general, al retardo de tiempo o diferencia de tiempo, es decir, el retardo de tiempo entre canales o diferencia entre las señales de los dos canales.

En el documento US 2012/0016632 A1 se da a conocer una estrategia a modo de ejemplo para determinar un retardo de tiempo entre canales a partir de una función de correlación cruzada ponderada.

Sin embargo, se ha demostrado que, especialmente en las señales que difieren, por ejemplo, de la voz limpia sin reverberación o sonido de fondo alguno, la contundencia de esta técnica general no es óptima.

Por lo tanto, un objetivo de la presente invención es proporcionar un concepto mejorado para estimar una diferencia de tiempo entre canales entre señales de dos canales.

Este objetivo se alcanza mediante un aparato para estimar una diferencia de tiempo entre canales de acuerdo con la reivindicación 1 o un método para estimar una diferencia de tiempo entre canales de acuerdo con la reivindicación 14 o un producto de programa informático de acuerdo con la reivindicación 15.

La presente invención se basa en el hallazgo de que se ha de llevar a cabo una ponderación de un espectro de correlación cruzada alisado o no alisado para obtener un espectro de correlación cruzada ponderada utilizando un primer procedimiento de ponderación o utilizando un segundo procedimiento de ponderación dependiendo de una característica de la señal estimada por un analizador de señales, en el que el primer procedimiento de ponderación es diferente del segundo procedimiento de ponderación.

En una realización adicional, el alisado del espectro de correlación cruzada en el tiempo que está controlado por una característica espectral del espectro de la señal del primer canal o la señal del segundo canal mejora significativamente la robustez y precisión de la determinación de la diferencia de tiempo entre canales.

En las realizaciones preferidas, se determina una característica de tonalidad/ruido del espectro y, en el caso de una señal de tipo tono, el alisado es más fuerte mientras que, en el caso de una señal de ruido, el alisado realizado es menos fuerte.

Preferentemente, se utiliza una medida de la planitud del espectro y, en caso de las señales de tipo tono, la medida de planitud del espectro será baja y el alisado se volverá más fuerte y, en el caso de las señales de tipo ruido, la medida de planitud del espectro será alta, como de 1 o cerca de 1, y el alisado será débil.

Por consiguiente, de acuerdo con la presente invención, se define un aparato para estimar una diferencia de tiempo entre canales entre una señal de un primer canal y una señal de un segundo canal mediante la reivindicación 1.

En el caso de las realizaciones preferidas relacionadas con el procesamiento adicional del espectro de correlación cruzada alisado, se ejecuta una operación de determinación de umbral adaptativa, en la cual se analiza la representación en el dominio del tiempo del espectro de correlación cruzada generalizada alisado a 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 tiempo entre canales como retardo de tiempo asociado al pico que está en una relación predeterminada con el umbral, como por

ejemplo la de ser mayor que el umbral.

En una realización, se determina el umbral variable en términos de valor igual a un múltiplo entero de un valor entre los mayores, por ejemplo el diez por ciento de los valores de la representación en el dominio del tiempo o, por otro lado, en una realización adicional de la determinación variable, se calcula el umbral variable mediante una multiplicación del umbral variable por el valor, donde el valor depende de una característica de relación señal a ruido de las señales del primero y el segundo canal, donde el valor se incrementa con respecto a una relación más elevada de señal a ruido y se reduce con una relación señal a ruido más baja.

Como se señaló anteriormente, se puede usar el cálculo de la diferencia de tiempo entre canales en muchas aplicaciones diferentes tales como el almacenamiento o transmisión de datos paramétricos, un procesamiento/codificación estéreo/multicanal, una alineación de tiempo de dos canales, una diferencia de tiempo de estimación de llegada para la determinación de la posición de un altavoz en un recinto con dos micrófonos y una configuración conocida de los micrófonos, con fines de formación de haces, filtrado espacial, descomposición primer plano/fondo o una determinación de ubicación de una fuente de sonido, por ejemplo por triangulación acústica basada en diferencias de tiempo de dos o tres señales.

En lo sucesivo, sin embargo, se describe una implementación y uso preferidos del cálculo de diferencia de tiempo entre canales con fines de alineación en el tiempo de la banda ancha de las dos señales estéreo en un proceso de codificación de una señal multicanal que tiene por lo menos dos canales.

Un aparato para codificar una señal multicanal con por lo menos dos canales comprende un determinador de parámetros para determinar un parámetro de alineación de ancho de banda por un lado y una pluralidad de parámetros de alineación de banda estrecha por el otro. Estos parámetros son utilizados por un alineador de señales a fin de alinear los por lo menos dos canales usando estos parámetros para obtener canales alineados. Luego, un procesador de señales calcula una señal media y una señal lateral utilizando los canales alineados y a continuación la señal media y la señal lateral son codificadas y enviadas a una señal de salida codificada que tiene, además, información lateral paramétrica, el parámetro de alineación de banda ancha y la pluralidad de parámetros de alineación de banda estrecha.

En el lado del decodificador, un decodificador de señales decodifica la señal media codificada y la señal lateral codificada para obtener señales media y lateral decodificadas. A continuación, estas señales son procesadas por un procesador de señales para calcular un primer canal decodificado y un segundo canal decodificado. Luego, estos canales decodificados son desalineados utilizando la información sobre el parámetro de alineación de banda ancha y la información sobre la pluralidad de parámetros de banda estrecha incluida en una señal multicanal codificada para obtener la señal multicanal decodificada.

En una implementación específica, el parámetro de alineación de banda ancha es un parámetro de diferencia de tiempo entre canales y la pluralidad de parámetros de alineación de banda estrecha are diferencias de fase entre canales.

La presente invención se basa en el hallazgo de que, específicamente en el caso de las señales de voz donde hay más de un altavoz, aunque también en el caso de otras señales de audio donde hay varias fuentes de audio, se pueden tener en cuenta los diferentes lugares de las fuentes de audio que corresponden a un mapeo de los dos canales de la señal multicanal para usar un parámetro de alineación de la banda ancha tal como un parámetro de diferencia de tiempo entre canales que se aplica al espectro completo de uno o ambos canales. Además de este parámetro de alineación de la banda ancha, se ha encontrado que varios parámetros de alineación de banda estrecha que difieren de subbanda a subbanda dan lugar, además, a una mejor alineación de la señal en ambos canales.

Por consiguiente, un alineación de la banda ancha que corresponde al mismo retardo de tiempo en cada subbanda junto con una alineación de fases correspondiente a diferentes rotaciones de fases de subbandas diferentes da lugar a una alineación óptima de ambos canales antes de convertir estos dos canales en una representación media/lateral que luego es codificada. Dado que se ha obtenido una alineación óptima, la energía en la señal media es lo más alta posible por un lado y la energía en la señal lateral es lo más baja posible por el otro, para poder obtener un resultado de codificación óptima con la tasa de bits más baja posible o una calidad de audio lo más elevada posible para cierta tasa de bits.

Específicamente en el caso del material de voz de conversación, parece que por lo general hay altavoces activos en dos lugares diferentes. Además, la situación es tal que, normalmente, solo un altavoz habla desde el primer lugar y luego el segundo altavoz habla desde el segundo lugar o ubicación. La influencia de las diferentes ubicaciones sobre los dos canales, como un primer canal o canal izquierdo y un segundo canal o canal derecho, se refleja por los diferentes tiempos de llegada y, por lo tanto, un determinado retardo de tiempo entre ambos canales debido a las ubicaciones diferentes, y este retardo de tiempo cambia de una vez a otra. En general, este efecto se refleja en las

dos señales de canales en forma de desalineación de la banda ancha, que puede ser abordado por el parámetro de alineación de banda ancha.

Por otra parte, se puede dar cuenta de otros efectos, especialmente los producidos por reverberación u otras fuentes de ruido, por parámetros individuales de alineación de fases correspondientes a bandas individuales que se superponen en los diferentes tiempos de llegada de la banda ancha o la desalineación de banda ancha de ambos canales.

En vista de eso, el uso tanto de un parámetro de alineación de la banda ancha como de una pluralidad de parámetros de alineación de banda estrecha además del parámetro de alineación de banda ancha da lugar a una alineación óptima de los canales en el lado del codificador para obtener una representación media/lateral buena y muy compacta mientras que, por otro lado, una desalineación correspondiente posterior a la decodificación en el lado del decodificador da como resultado una buena calidad de audio para una determinada tasa de bits o una pequeña tasa de bits para una determinada calidad de audio necesaria.

Una ventaja de la presente invención es que proporciona un nuevo esquema de codificación estéreo mucho más adecuado para una conversión de voz estéreo que los esquemas de codificación estéreo existentes. De acuerdo con la invención, se combinan las tecnologías de estéreo paramétrico y tecnologías de codificación estéreo conjuntas, especialmente aprovechando la diferencia de tiempo entre canales que aparece en canales de una señal multicanal, específicamente en el caso de las fuentes de voz, aunque también en el caso de otras fuentes de audio.

Varias realizaciones proporcionan ventajas útiles, como se describe más adelante.

El nuevo método es una estrategia híbrida que mezcla elementos de un estéreo M/S convencional y estéreo paramétrico. En un M/S convencional, los canales son sometidos pasivamente a mezcla descendente para generar una señal media y una lateral. El proceso se puede extender adicionalmente mediante la rotación del canal utilizando una transformada de Karhunen-Loeve (KLT), lo que también se conoce como análisis de componentes principales (PCA) antes de sumar y diferenciar los canales. La señal media es codificada con una codificación de código primario, mientras que la lateral es transmitida a un codificador secundario. El estéreo M/S evolucionado puede usar además la predicción de la señal lateral por el canal medio codificado en la trama presente o la anterior. El principal objetivo de la rotación y predicción es maximizar la energía de la señal media minimizando a la vez la energía de la lateral. El estéreo M/S es preservador de la forma de onda y, en este aspecto, es muy robusto para cualquier configuración estéreo, aunque puede ser muy costoso en términos de consumo de bits.

Para la mayor eficiencia a bajas tasas de bits, el estéreo paramétrico computa y codifica parámetros, como diferencias de nivel entre canales (ILD), diferencias de fase entre canales (IPD), diferencias de tiempo entre canales (ITD) y coherencia entre canales (IC). Estos representan de manera compacta la imagen estéreo y son indicaciones de la escena auditiva (localización de la fuente, paneo, ancho del estéreo …). La finalidad es, entonces, parametrizar la escena estéreo y codificar solo una señal de mezcla descendente que puede estar en el decodificador y que, con la ayuda de las indicaciones de estéreo transmitidas, se puede espacializar una vez más.

Nuestra estrategia mezcló los dos conceptos. En primer lugar, se computan los indicios estéreo ITD e IPD y se aplican a los dos canales. La meta es representar la diferencia de tiempo en banda ancha y la fase en diferentes bandas de frecuencia. A continuación, se alinean los dos canales en tiempo y fase y luego se ejecuta la codificación M/S. Se encontró que la ITD e IPD eran útiles para modelar la voz estéreo y son un buen reemplazo de la rotación basada en KLT en M/S. A diferencia de la codificación paramétrica pura, el ambiente no es más modelado por las IC sino directamente por la señal lateral que es codificada y/o predicha. Se ha encontrado que esta estrategia es más segura, especialmente cuando se tratan señales de voz.

El cómputo y procesamiento de las ITD es una parte crucial de la invención. Las ITD ya eran aprovechadas en la codificación binaural de indicaciones (BCC) de la técnica anterior, aunque de manera ineficiente una vez que las ITD cambiaban en el tiempo. Para evitar esta deficiencia, se diseñó un enventanado específico para alisar las transiciones entre dos ITD diferentes y para poder cambiar sin dificultades de un altavoz a otro situados en un lugar diferente.

Realizaciones adicionales se relacionan con el procedimiento en que, en el lado del codificador, se ejecuta la determinación de parámetros para determinar la pluralidad de parámetros de alineación de la banda estrecha utilizando canales que ya han sido alineados con el parámetro antes determinado de alineación de la banda ancha.

De modo correspondiente, la desalineación de la banda estrecha en el lado del decodificador se realiza antes de ejecutar la desalineación de la banda ancha utilizando por lo general un único parámetro de alineación de la banda ancha.

En realizaciones adicionales, es preferible que, en el lado del codificador, aunque aún más especialmente en el lado

del decodificador, se realice alguna operación de enventanado y superposición y suma o cualquier tipo de fundido cruzado de un bloque al siguiente después de todas las alineaciones y, específicamente, después de una alineación en el tiempo utilizando el parámetro de alineación de banda ancha. Esto evita las distorsiones audibles tales como clics cuando el parámetro de alineación de tiempo o de banda ancha cambia de un bloque a otro.

En otras realizaciones, se aplican diferentes resoluciones espectrales. Específicamente, se somete a las señales de canales a una conversión tiempo–espectral con una resolución de frecuencia elevada tal como un espectro de DFT, mientras que los parámetros tales como los parámetros de alineación de banda estrecha se determinan para bandas de parámetros con una resolución espectral más baja. Por lo general, una banda de parámetros tiene más de una línea espectral que el espectro de la señal y por lo general tiene una serie de líneas espectrales del espectro de DFT. Adicionalmente, las bandas de parámetros aumentan de bajas frecuencias a altas frecuencias para compensar los problemas psicoacústicos.

Realizaciones adicionales se refieren al uso adicional de un parámetro de nivel tal como una diferencia entre niveles u otros procedimientos para procesar la señal lateral tales como parámetros de llenado estéreo, etc. La señal lateral codificada puede estar representada por la señal lateral real en sí, o por una señal residual de predicción que se ejecuta utilizando la señal media de la trama actual o cualquier otra trama, o por una señal lateral o señal lateral residual de predicción en solo una subserie de bandas y parámetros de predicción solo para el resto de las bandas, o incluso mediante parámetros de predicción correspondientes a todas las bandas sin información de señal lateral con resolución de alta frecuencia alguna. Por ende, en la última alternativa anteriormente señalada, la señal lateral codificada solo está representada por un parámetro de predicción por cada banda de parámetros o solo una subserie de bandas de parámetros, por lo que para el resto de las bandas de parámetros no existe información alguna sobre la señal lateral original.

Adicionalmente, es preferible que la pluralidad de parámetros de alineación de banda estrecha no se relacionen con todas las bandas de parámetros que reflejan el ancho de banda total de la señal de banda ancha sino solo con una serie de bandas más bajas, como por ejemplo el 50 por ciento inferior de las bandas de parámetros. Por otro lado, no se utilizan los parámetros de llenado estéreo correspondientes al par de bandas inferiores, ya que con respecto a estas bandas, se transmite la señal lateral misma o una señal residual de predicción para asegurarse de que, por lo menos para las bandas más bajas, se disponga de una representación con forma de onda correcta. Por otra parte, la señal lateral no se transmite en una representación de forma de onda exacta para las bandas superiores a fin de reducir adicionalmente la tasa de bits, aunque la señal lateral está típicamente representada por parámetros de llenado estéreo.

Además, es preferible ejecutar el análisis de parámetros y alineación completo dentro de un único dominio de frecuencia basándose en el mismo espectro de DFT. Para este fin, es preferible asimismo utilizar la correlación cruzada generalizada con tecnología de transformación de fases (GCC-PHAT) con fines de determinación de las diferencias de tiempo entre canales. En una realización preferida de este procedimiento, se ejecuta un alisado de un espectro de correlación basado en una información sobre la forma espectral, información que consiste preferentemente en una medida de la planitud espectral, de tal manera que el alisado sea tenue en el caso de las señales de ruido y que el alisado se torne más fuerte en el caso de las señales de tono.

Además, es preferible ejecutar una rotación de fase especial, donde se tomen en cuenta las amplitudes de los canales. Específicamente, la rotación de fase se distribuye entre los dos canales con el fin de alinear en el lado del codificador y, naturalmente, con la finalidad de desalinear en el lado del decodificador, donde se considera que un canal con mayor amplitud es el canal principal y ha de resultar menos afectado por la rotación de fase, es decir, que tendrá una menor rotación que un canal con amplitud más baja.

Además, el cálculo de la suma-diferencia se realiza utilizando un escalado de energía con un factor de escala que se deriva de las energías de ambos canales y está unido, además, a un determinado rango para asegurarse de que el cálculo medio/lateral no afecte demasiado la energía. Por otro lado, sin embargo, se debe tener en cuenta que, en el marco de la presente invención, este tipo de conservación de la energía no es tan crucial como en los procedimientos de la técnica anterior, ya que el tiempo y la fase se alinean de antemano. Por lo tanto, las fluctuaciones de energía debido al cálculo de una señal media y una señal lateral de izquierda y derecha (en el lado del codificador) o debido al cálculo de una señal izquierda y una derecha a partir de la media y lateral (en el lado del decodificador) no son tan significativas como en la técnica anterior.

A continuación, se describen las realizaciones preferidas de la presente invención con respecto a los dibujos adjuntos, en los cuales:

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

la figura 2 es una realización preferida de un aparato para decodificar una señal multicanal codificada;

la figura 3 es una ilustración de diferentes resoluciones de frecuencia y otros aspectos relacionados con la frecuencia correspondientes a ciertas realizaciones;

la figura 4a ilustra un diagrama de flujo de los procedimientos ejecutados en el aparato para codificar con el fin de alinear los canales;

la figura 4b ilustra una realización de los procedimientos ejecutados en el dominio de la frecuencia;

la figura 4c ilustra una realización de los procedimientos ejecutados en el aparato para codificar utilizando una ventana de análisis con porciones de relleno con ceros y rangos de superposición;

la figura 4d ilustra un diagrama de flujo de otros procedimientos ejecutados dentro del aparato para codificar;

la figura 4e ilustra un diagrama de flujo que muestra una implementación de una estimación de diferencia de tiempo entre canales;

la figura 5 ilustra un diagrama de flujo que ilustra una realización adicional de los procedimientos ejecutados en el aparato para codificar;

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

la figura 6b ilustra un diagrama de flujo de una correspondiente realización de un decodificador;

la figura 7 ilustra una configuración preferida de ventanas con ventanas de seno con baja superposición con relleno con ceros para un análisis y síntesis estéreo en tiempo–frecuencia;

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

la figura 9a ilustra los procedimientos ejecutados por un aparato para decodificar una señal multicanal codificada es una realización preferida;

la figura 9b ilustra una implementación del aparato para decodificar una señal multicanal codificada;

la figura 9c ilustra un procedimiento ejecutado en el contexto de una desalineación de banda ancha en el contexto de la decodificación de una señal multicanal codificada;

la figura 10a ilustra una realización de un aparato para estimar una diferencia de tiempo entre canales;

la figura 10b ilustra una representación esquemática del procesamiento adicional de una señal donde se aplica la diferencia de tiempo entre canales;

la figura 10c ilustra una representación esquemática del analizador de señales implementado en forma de estimador de ruido en una realización y el ponderador de acuerdo con las realizaciones de la invención;

la figura 10d ilustra una representación esquemática del ponderador de acuerdo con realizaciones de la invención;

la figura 10e ilustra una representación esquemática del procesador de acuerdo con realizaciones de la invención;

la figura 10f ilustra una representación esquemática del estimador de ruido de acuerdo con realizaciones de la invención;

la figura 11a ilustra procedimientos ejecutados por el procesador de la figura 10a;

la figura 11b ilustra procedimientos adicionales ejecutados por el procesador en la figura 10a;

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

la figura 11d ilustra una primera realización de la determinación del umbral variable;

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

la figura 11f ilustra una representación esquemática del procesador de acuerdo con las realizaciones de la invención;

la figura 12 ilustra una representación en el dominio del tiempo correspondiente a un espectro de correlación cruzada alisado para una señal de voz limpia;

la figura 13 ilustra una representación en el dominio del tiempo de un espectro de correlación cruzada alisado para una señal de voz con ruido y ambiente.

La figura 10a ilustra una realización de un aparato para estimar una diferencia de tiempo 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 son enviados a un convertidor tiempo–espectro 150 que se ilustra adicionalmente, con respecto a la figura 4e como elemento 451.

Además, las representaciones en el dominio del tiempo de las señales de los canales izquierdo y derecho son ingresadas en una calculadora 1020 para calcular un espectro de correlación cruzada correspondiente a un bloque de tiempo a partir de la señal del primer canal en el bloque de tiempo y la señal del segundo canal en el bloque de tiempo. Además, el aparato comprende un estimador de características espectrales 1010 para estimar una característica de un espectro de la señal del primer canal o la señal del segundo canal correspondiente al bloque de tiempo. El aparato comprende asimismo un filtro de alisado 1030 para alisar el espectro de correlación cruzada en el tiempo utilizando la característica espectral para obtener un espectro de correlación cruzada alisado. El aparato comprende asimismo un procesador 1040 para procesar el espectro de correlación alisado para obtener la diferencia de tiempo entre canales.

Alternativamente, en otra realización, el elemento 1030 no está presente y, por lo tanto, el elemento 1010 tampoco es necesario, como lo indica la línea discontinua 1035. El aparato comprende asimismo un analizador de señales 1037 que calcula una estimación de una característica de la señal tal como una estimación de ruido 1038. Esta estimación es enviada a un ponderador 1036 configurado para ejecutar diferentes operaciones de ponderación, dependiendo de la estimación de la característica de la señal. También se utiliza la estimación de la característica de la señal, preferentemente, para controlar el procesador 1040, por ejemplo cuando el procesador 1040 ejecuta la operación de identificación de picos. La figura 10c ilustra además el analizador de señales 1037 y el ponderador controlable 1036.

Específicamente, un aparato de acuerdo con las realizaciones de la presente invención está destinado a la estimación de una diferencia de tiempo entre canales entre la señal de un primer canal y la señal de un segundo canal. Este dispositivo comprende el analizador de señales 1037 de la figura 10a, una calculadora de espectro de correlación cruzada 1020 de la figura 10a, un ponderador 1036 para ponderar un espectro de correlación cruzada alisado o no alisado de la figura 10a y a un procesador conectado seguidamente 1040 para procesar el espectro de correlación cruzada ponderado.

El convertidor de tiempo–espectro de los elementos 150, el estimador de características espectrales 1010, el filtro de alisado 1030 no son necesarios para una implementación básica de la presente invención, aunque son preferibles para las realizaciones preferidas de la presente invención. El analizador de señales 1037 está configurado para estimar una característica de la señal tal como un nivel de ruido 1038 de la señal del primer canal o la señal del segundo canal o ambas señales o una señal derivada de la señal del primer canal o la señal del segundo canal. De esa manera, se puede derivar una característica de la señal o una estimación de la característica de la señal tal como una estimación de ruido para ser utilizada más adelante por el ponderador 1036 y, preferentemente, para ser utilizada también por el procesador 1040, solo de la señal del primer canal o canal izquierdo, solo del segundo canal o canal derecho, o se puede derivar de ambas señales. La derivación de la característica de la señal a partir de ambas señales podría ser, por ejemplo, una derivación de una característica individual de señal de la señal del primer canal, una característica de señal individual adicional de la señal del segundo canal o canal derecho y, luego, la característica de la señal final 1038 sería, por ejemplo, un promedio o un promedio ponderado entre ambos canales. En este caso, por ejemplo la ponderación se puede realizar de acuerdo con la amplitud por lo que diferentes amplitudes, por ejemplo, de las tramas de los canales, dan lugar a diferentes efectos de la correspondiente estimación individual de ruido sobre el nivel de ruido final 1038. Además, la señal derivada de la señal del primer canal y la señal del segundo canal podría ser, por ejemplo, una señal combinatoria obtenida mediante la suma de la señal del primer canal o canal izquierdo y la señal del segundo canal o canal derecho para obtener una señal combinada y, a continuación, se calcula la característica de la señal 1038 a partir de la señal combinada.

En una realización preferida, el analizador de señales 1036 se implementa en forma de estimador o analizador de ruido. Sin embargo, también se pueden ejecutar otras modalidades de análisis de señales, como análisis de tonalidad, detección de la actividad de voz, análisis de transitorios, análisis estéreo, análisis de voz/música, análisis de interferencias por oradores, análisis de la música de fondo, análisis de voz limpia o cualquier otro análisis de señal para determinar si una señal tiene una primera característica o una segunda característica de manera que se seleccione el procedimiento de ponderación que corresponde.
La combinación puede ser una combinación con factores de ponderación iguales, es decir, una combinación del canal

izquierdo sin ninguna ponderación y el canal derecho sin ninguna ponderación, lo que correspondería a factores de ponderación de 1.0 o, alternativamente, se pueden aplicar factores de ponderación diferentes. Además, se puede obtener la señal derivada del primer canal o la señal derivada del segundo canal realizando un filtrado de paso alto o un filtrado de paso bajo o se puede derivar ejecutando un procesamiento que utiliza una función de compresión de amplitud o una función de compresión inversa de la amplitud. Una función de compresión de amplitud sería una función logarítmica con un valor de potencia inferior a 1. Una función de compresión inversa sería una función exponencial o una función de potencia con un exponente superior a 1. Por consiguiente, dependiendo de ciertas implementaciones, se pueden aplicar diferentes operaciones de procesamiento a diferentes señales de canales izquierdo y derecho y se pueden combinar o no ambos canales. En la realización preferida, los canales izquierdo y el derecho se suman uno a otro preferentemente, incluso sin ninguna ponderación específica, y luego se calcula la estimación de la característica de la señal a partir del resultado del cálculo de combinación.

La calculadora 1020 para calcular un espectro de correlación cruzada correspondiente a un bloque de tiempo a partir de la señal del primer canal en el bloque de tiempo y la señal del segundo canal en el bloque de tiempo puede ser implementada de varias maneras. Una manera es calcular una correlación cruzada a partir de las señales en el dominio del tiempo en las tramas en el dominio del tiempo y luego convertir el resultado del dominio del tiempo al dominio espectral. Otra implementación consiste en usar, por ejemplo, una DFT o cualquier otra conversión del tiempo a espectral, convertir las tramas subsiguientes de la señal del primer canal y las tramas subsiguientes de la señal del segundo canal a una representación espectral en la cual las tramas subsiguientes se superponen o no se superponen. Por consiguiente, por cada bloque de tiempo de la señal del primer canal, se obtiene una representación espectral y, de modo correspondiente, por cada bloque de tiempo de la señal del segundo canal, se obtiene una representación espectral. El cálculo de correlación cruzada se realiza multiplicando un valor espectral de un determinado bin de frecuencia k y un determinado bloque de tiempo o el índice de muestreo de tiempo s por el valor conjugado complejo del valor espectral con el mismo índice k y el mismo índice s de la representación espectral del mismo bloque de tiempo del segundo canal. Se pueden emplear otros procedimientos de cálculo de correlación cruzada diferentes de los descritos anteriormente para calcular el espectro de correlación cruzada correspondiente a un bloque de tiempo.

El ponderador 1036 está configurado para ponderar el espectro de correlación cruzada obtenido por la calculadora. En una implementación, el espectro de correlación cruzada es un espectro de correlación cruzada no alisado, pero en otras realizaciones, el espectro de correlación cruzada es alisado, donde este alisado es un alisado con respecto al tiempo. Por consiguiente, para calcular el espectro de correlación cruzada alisado, se puede utilizar el espectro de correlación cruzada del último bloque junto con un espectro de correlación cruzada (bruto) del bloque actual y, dependiendo de la implementación, se puede utilizar, por ejemplo, una información de control de alisado como se proporciona, por ejemplo, por el estimador de características espectrales 1010 de la figura 10a. Sin embargo, el alisado se puede realizar asimismo utilizando una configuración de alisado predeterminada, es decir, constante o invariante en el tiempo. De acuerdo con realizaciones de la invención, se calcula el espectro de correlación cruzada ponderado empleando un primer procedimiento de ponderación 1036a o empleando un segundo procedimiento de ponderación 1036b que se ilustran, por ejemplo, en la figura 10d. Específicamente, la selección, tanto si el espectro de correlación cruzada ponderado se deriva utilizando el primer o el segundo procedimiento, se realiza dependiendo de la estimación de la característica de la señal efectuada por el analizador de señales 1037. Por consiguiente, de acuerdo con la presente invención, se utiliza una ponderación con una primera característica de ponderación para una determinada característica de la señal del primer canal o del segundo canal o la señal combinada, mientras que se aplica un segundo procedimiento de ponderación dependiendo de otra característica de la señal, como se determina por el analizador de señales 1037. El resultado del ponderador 1036 es un espectro de correlación cruzada ponderado y alisado o no alisado que luego es procesado adicionalmente por el procesador 1040 para obtener la diferencia de tiempo entre canales entre la señal del primer canal y la señal del segundo canal.

La figura 10d ilustra una implementación del analizador de señales en forma de estimador de ruido y el ponderador en conexión con el procesador 1040 de acuerdo con una realización de la invención. Específicamente, el estimador de ruido 1037 comprende una calculadora de estimación de ruido 1037a y un clasificador de estimaciones de ruido 1037b. El clasificador de estimaciones de ruido 1037b emite una señal de control 1050 que corresponde a la salida de estimaciones de ruido 1038 generada por el bloque 1037 de la figura 10a. Esta señal de control se puede aplicar a un primer conmutador 1036c o a un segundo conmutador 1036d. En esta implementación, se incluyen núcleos de procesamiento 1036a que implementan el primer procedimiento de ponderación y se provee otro núcleo de cálculo para implementar el segundo procedimiento de ponderación 1036b. Dependiendo de esta implementación, solo se incluye el conmutador 1036c y, dependiendo de la señal de control 1050, solo se selecciona el procedimiento de ponderación determinado por el conmutador 1036c, es decir que el espectro de correlación cruzada determinado por la calculadora 1020 es ingresado en el conmutador 1036c y dependiendo de la configuración del conmutador, es enviado al núcleo 1036a o al núcleo 1036b. En otra implementación, el conmutador 1036c no está presente, sino que el espectro de correlación cruzada como se determina por el bloque 1020 es alimentado a ambos núcleos de procesamiento 1036a y 1036b y, dependiendo del control del conmutador de salida 1036d, se selecciona la salida del bloque 1036a o la salida del bloque 1036b y se envía al procesador 1040. De esa manera, dependiendo de la implementación, se calcula solo un único espectro de correlación cruzada ponderado, donde la selección de cuál se

calcula se realiza según la señal de control 1050 y el conmutador de entrada. Alternativamente, se calculan ambos espectros de correlación cruzada ponderados y solo el espectro de correlación cruzada que es seleccionado por el conmutador de salida 1036d es enviado al procesador 1040. Además, puede haber un único núcleo de procesamiento sin conmutadores de entrada/salida y, dependiendo de la señal de control, se establece el procedimiento de ponderación correcto para el bloque de tiempo correspondiente. Por consiguiente, para cada bloque de tiempo, se puede calcular una estimación de ruido o señal de control 1050 y, para cada bloque de tiempo, se puede conmutar la ponderación de un procedimiento de ponderación al otro procedimiento de ponderación. En este contexto, se debe tener en cuenta que también se pueden implementar tres o más procedimientos de ponderación diferentes dependiendo de tres o más estimaciones de ruido diferentes, según el caso. Por consiguiente, la presente invención no solo incluye la selección entre dos procedimientos de ponderación diferentes, sino también incluye la selección entre tres o más procedimientos de ponderación dependiendo de una señal de control derivada de la característica de ruido de las señales de los canales primero y segundo.

En una implementación preferida, el primer procedimiento de ponderación comprende una ponderación tal que se normalice una amplitud y se mantenga una fase y el segundo procedimiento de ponderación comprende un factor de ponderación derivado del espectro de correlación cruzada alisado o no alisado utilizando una operación de potenciación con una potencia inferior a 1 o superior a 0. Además, el primer procedimiento de ponderación puede ser casi idéntico al segundo procedimiento de ponderación, excepto porque el segundo procedimiento de ponderación utiliza una potencia de entre 0 y 1, es decir, a una potencia mayor que 0 y menor que 1, mientras que el primer procedimiento de ponderación no aplica ninguna potenciación o, dicho de otro modo, aplica una potencia de 1. Por consiguiente la normalización ejecutada por el segundo procedimiento de ponderación se comprime, es decir que el factor de normalización aplicado por el primer procedimiento de ponderación tiene cierto valor y el factor de normalización aplicado por medio del segundo procedimiento de ponderación al mismo valor de correlación cruzada espectral tiene una menor magnitud. Esto se aplica a los valores espectrales más elevados del espectro de correlación cruzada. Sin embargo, en el caso de los valores pequeños del espectro de correlación cruzada, el valor de normalización correspondiente al segundo procedimiento de ponderación es mayor que el valor de normalización correspondiente al primer procedimiento de ponderación con respecto al mismo valor espectral del espectro de correlación cruzada. Esto se debe a que una operación de potenciación con una potencia inferior a 1, tal como una operación de raíz cuadrada con una potencia de 1/2 aumenta los valores pequeños pero reduce los valores elevados. Por consiguiente, los cálculos adicionales de factores de ponderación para el segundo procedimiento de ponderación también pueden comprender cualquier función de compresión tal como una función logarítmica. En una realización preferida, el primer procedimiento de ponderación opera sobre la base de la ponderación aplicada para la transformada de fase (PHAT, y el segundo procedimiento de ponderación opera sobre la base de los cálculos aplicados al procedimiento de fase de espectro de potencia cruzado modificado (MCSP).

Además, el segundo procedimiento de ponderación se implementa preferentemente de manera que comprenda una normalización, a fin de que el rango de salida del segundo procedimiento de normalización esté en un rango en el cual se ubica un rango de salida del primer procedimiento de normalización, o de manera que el rango de salida del segundo procedimiento de normalización sea igual al rango de salida del primer procedimiento de normalización. Esto se puede implementar, por ejemplo, calculando los valores absolutos de todos los valores espectrales del MCSP-espectro de correlación cruzada ponderado, mediante la suma total de todas las magnitudes de una representación espectral correspondiente a un bloque de tiempo y luego la división del resultado por el número de valores espectrales presentes en un bloque de tiempo.

En términos generales, el procesador 1040 de la figura 10a está configurado para ejecutar algunas etapas de procesamiento con respecto al espectro de correlación cruzada ponderado donde, específicamente, se ejecuta una determinada operación de identificación de picos a fin de obtener, en última instancia, la diferencia de tiempo entre canales. Preferentemente, esta operación de identificación de picos tiene lugar en el dominio del tiempo, es decir que el espectro de correlación cruzada ponderado y alisado o no alisado se convierte de la representación espectral en una representación en el dominio del tiempo y, luego, se analiza esta representación en el dominio del tiempo y, específicamente, se elige un pico o varios picos basándose en un umbral. Dependiendo de la configuración de la estimación de ruido, se ejecuta una primera operación de identificación de picos o una segunda operación de identificación de picos donde, preferentemente, ambas operaciones de identificación de picos son diferentes entre sí con respecto al umbral utilizado por la operación de identificación de picos.

La figura 10e ilustra una situación que es similar, con respecto al conmutador de entrada 1040 y el conmutador de salida 1043, al procedimiento de la figura 10d. En una implementación ilustrada en la figura 10e, se pueden aplicar ambas operaciones de identificación de picos y el resultado de la “correcta” operación de identificación de picos puede ser seleccionado por el conmutador de salida 1043. Por otro lado, el conmutador de entrada está presente y, dependiendo de la señal de control 1050, solo se selecciona el procedimiento correcto de identificación de picos, es decir, 1041 o 1042. Por consiguiente, en una implementación, no se presentan los dos conmutadores, sino que, en una implementación, se presenta el conmutador de entrada 1040 o el conmutador de salida 1043 en analogía con lo que se ha derivado anteriormente respecto a la figura 10d. En una implementación adicional, solo existe un único

núcleo de procesamiento que aplica la operación de identificación de picos con un umbral variable y se utiliza la señal de control 1050 para establecer el umbral correcto dentro del único núcleo de procesamiento. En una realización preferida, el establecimiento del umbral se realiza de tal manera que el segundo umbral sea más elevado que el primer umbral, donde el segundo umbral se utiliza, por lo tanto, cuando se ha aplicado el segundo procedimiento de ponderación ejecutado en el bloque 1036b, y donde se utiliza el primer umbral cuando se ha aplicado el primer procedimiento de ponderación en el bloque 1036a. De esa manera, cuando se detecta un nivel elevado de ruido de fondo, se aplica el segundo procedimiento de ponderación con una potencia de entre 0 y 1 o bien una operación logarítmica, es decir, un procedimiento de compresión y, luego, el umbral correspondiente a la identificación de picos debe ser más bajo en comparación con un umbral de identificación de picos que se ha de utilizar cuando se detecta un bajo nivel de ruido de fondo, es decir, cuando se aplica el primer procedimiento de ponderación que ejecuta una normalización con un factor de normalización que no se basa en una función de compresión tal como una función logarítmica o una función de potenciación con una potencia menor que 1.

Posteriormente, se ilustra una implementación preferida del analizador de señales como el estimador de ruido 1037 en la figura 10f. Básicamente, el estimador de ruido 1037 consiste en una calculadora de estimación de ruido 1037a y un clasificador de estimaciones de ruido 1037b ilustrado en la figura 10d y que también está indicado en la figura 10f. La calculadora de estimación de ruido 1037a comprende un estimador de ruido de fondo 1060 y el alisador (de tiempo) conectado a continuación 1061 que puede ser implementado, por ejemplo, en forma de filtro IIR.

La entrada a la calculadora de estimación de ruido 1037a o, específicamente, el estimador de ruido de fondo 1060 es una trama de la señal del primer canal o canal izquierdo, una trama de la señal del segundo canal o canal derecho o una señal derivada de la señal de ese canal o una señal combinada que se obtiene sumando, por ejemplo, una representación en el dominio del tiempo de la señal del primer canal y una representación en el dominio del tiempo de la señal del segundo canal en el mismo bloque de tiempo.

Con respecto al clasificador de estimaciones de ruido 1037b, la señal de entrada es transmitida a un detector de actividad de señal 1070 que controla un selector 1071. Basándose en el resultado del detector de actividad de señal 1070, el selector 1071 selecciona únicamente las tramas activas. Además, hay una calculadora de nivel de las señales 1072 conectada después del selector 1071. El nivel calculado de la señal es enviado luego a un alisador (en tiempo) 1073 que está implementado, por ejemplo, en forma de filtro IIR. Seguidamente, en el bloque 1074, tiene lugar un cálculo de relación señal a ruido y se compara el resultado, dentro de un comparador 1075, con un umbral preferentemente predeterminado que es, por ejemplo, de entre 45 dB y 25 dB y preferentemente está incluso en un rango de entre 30 y 40 dB y, más preferentemente, es de 35 dB.

La salida del comparador 1075 es el resultado de la detección que indica si hay un alto nivel de ruido o un bajo nivel de ruido o que indica que se ha de ejecutar un establecimiento de umbral de una determinada manera mediante un único procesador de procedimientos de ponderación o, cuando hay dos procesadores de procedimientos de ponderación como se ilustra en la figura 10d, entonces el resultado de la decisión obtenido del comparador 1075, es decir, la señal 1050 controla o bien el conmutador de entrada 1036c o bien el conmutador de salida 1036d a fin de enviar el espectro de correlación cruzada correctamente ponderado al procesador 1040.

Preferentemente se calcula el resultado de la detección 1050 para cada bloque de tiempo o trama. De esa manera, cuando el detector de actividad de señal 1070 indica, por ejemplo, para una trama determinada, que se trata de una trama no activa, en ese caso no se ejecuta ni el cálculo de nivel de la señal ni el alisado en el tiempo para esta trama, puesto que el selector 1071 únicamente selecciona una trama activa. Por consiguiente, en una realización no se ejecuta, en el caso de una trama inactiva, un cálculo de relación SNR y, por lo tanto, en esta realización, con respecto a esta trama inactiva, no se provee un resultado de detección en absoluto. Por consiguiente, en una implementación, se utiliza el mismo procedimiento de ponderación que se ha determinado anteriormente con respecto a la última trama activa o, de lo contrario, en el caso de una trama inactiva, se aplica el primer procedimiento de ponderación o el segundo procedimiento de ponderación, o incluso un tercer procedimiento de ponderación como solución de segunda opción. Alternativamente, se puede implementar el uso por una calculadora de relación SNR 1074, en el caso de una trama inactiva, del nivel de la señal alisada en el tiempo de la última trama activa o de aparición más reciente. De esa manera, se puede obtener el resultado de la detección incluso respecto a tramas inactivas o bien, en el caso de las tramas inactivas, se utiliza un procedimiento de ponderación alternativo (como segunda opción), o bien, en el caso de las tramas inactivas, se continúa utilizando el mismo procedimiento de ponderación que se ha determinado para la última trama activa que precede a la trama inactiva, según el caso.

En una solicitud anterior de patente [1], se introdujo un estimador de diferencia de tiempo entre canales (ITD). Este estimador se basa en la correlación cruzada generalizada con transformación de fase (GCC-PHAT), una técnica de uso generalizado en la bibliografía de TDOA (el documento inicial es [2], otra buena referencia es [3]). La diferencia de tiempo entre los dos canales se encuentra mediante la identificación de picos de la salida de la GCC. Se puede obtener una mayor robustez ya sea utilizando una gran longitud de ventana de análisis o alisando el espectro de correlación cruzada en el tiempo. La principal contribución de [1] fue hacer que este alisado se adapte a un factor de

alisado que depende de una medida de la planitud espectral.

Las etapas del estimador de ITD de [1] se pueden describir de la siguiente manera:

1. Transformada discreta de Fourier: la señal del canal izquierdo $x_L(n)$ y la señal del canal derecho $x_R(n)$ están ubicadas en tramas, enventanadas y transformadas al dominio de la frecuencia utilizando una DFT

$$X_L(k,s) = \sum_{n=0}^{N_{DFT}-1} x_L(n+sN)w(n)e^{-i2\pi\frac{kn}{N_{DFT}}}$$

$$X_R(k,s) = \sum_{n=0}^{N_{DFT}-1} x_R(n+sN)w(n)e^{-i2\pi\frac{kn}{N_{DFT}}}$$

donde n es el índice de muestreo de tiempo, s es el índice de trama, k es el índice de frecuencia, N es la longitud de la trama, N_{DFT} es la longitud de la DFT y $w(n)$ es la ventana de análisis.

2. Espectro de correlación cruzada: se computa la correlación entre los dos canales en el dominio de la frecuencia

$$C(k,s) = X_L(k,s)X_R^*(k,s)$$

3. Alisado: el espectro de correlación cruzada es alisado en el tiempo con un factor de alisado dependiendo de una medida de la planitud espectral. Se utiliza un alisado más fuerte cuando la planitud espectral es baja para que el estimador de ITD sea más fuerte en las señales tonales fijas. Se utiliza un alisado más tenue cuando la planitud espectral es elevada para que el estimador de ITD se adapte más rápidamente en las señales transitorias, es decir cuando la señal va cambiando rápidamente.

El alisado se ejecuta con

$$\tilde{C}(k,s) = \big(1 - sfm(s)\big)\tilde{C}(k,s-1) + sfm(s)C(k,s)$$

donde

$$sfm(s) = \max\big(sfm_{chan(X_L)}, sfm_{chan(X_R)}\big)$$

y

$$sfm_{chan(X)} = \frac{\prod_{k=0}^{N_{sfm}-1} X(k,s)^{\frac{1}{N_{sfm}}}}{\sum_{k=0}^{N_{sfm}-1} \frac{X(k,s)}{N_{sfm}}}$$

4. Ponderación: el espectro de correlación cruzada alisado es ponderado por la inversa de su magnitud. Esta ponderación normaliza la amplitud y mantiene solo la fase; por esta razón se denomina transformada de fase (PHAT).

$$\tilde{C}_{PHAT}(k,s) = \frac{\tilde{C}(k,s)}{\left|\tilde{C}(k,s)\right|}$$

5. Transformada inversa: se obtiene la GCC final transformando el espectro de correlación cruzada $\tilde{C}_{PHAT}(k,s)$ nuevamente al dominio del tiempo

$$GCC(n) = \frac{1}{N_{DFT}} \sum_{k=0}^{N_{DFT}-1} \tilde{C}_{PHAT}(k,s)e^{i2\pi\frac{kn}{N_{DFT}}}$$

6. Identificación de picos: la estrategia más sencilla consiste en buscar el máximo global del valor absoluto de la GCC hallada en la etapa 5. Si este máximo tiene un valor superior a determinado umbral, se estima la ITD en términos de retardo n correspondiente a este máximo. Las estrategias más avanzadas utilizan mecanismos basados en histéresis y/o persistencia ("hangover") adicionales para obtener una estimación de ITD más suave en el tiempo.

La GGC-PHAT da muy buen resultado en entornos resonantes de bajo ruido (ver, por ejemplo, [3]). Sin embargo, cuando el nivel del ruido de fondo es alto o ante la presencia de otros componentes de la señal (tales como música, transitorios, escenas estéreo complejas, tramas clasificadas como inactivas, oradores interferentes), la eficiencia de la GCC-PHAT decae significativamente. En ese caso la salida de GCC es ruidosa y no contiene un único pico acentuado. En consecuencia, la identificación de picos con frecuencia no llega a encontrar la ITD correcta. Esto ocurre porque la transformada de fase trata todas las frecuencias por igual, independientemente de la relación señal a ruido. Luego la GCC se contamina por la fase de los bins cuya estimación de ruido es baja.

Para evitar este problema, se han propuesto muchas otras ponderaciones de GCC en la bibliografía. Una de ellas resultó ser muy eficaz en nuestras señales problemáticas bajo análisis. Se propuso por primera vez en [4] y se denominó en ese momento "fase de espectro de potencia cruzada modificada" (MCSP). Su buen rendimiento en entornos de alto ruido fue confirmado más tarde en otros documentos (véase, por ejemplo, [5]). La ponderación (Etapa 4. de la técnica anterior) se modifica de la siguiente manera:

$$\tilde{C}_{MCSP}(k,s) = \frac{\tilde{C}(k,s)}{\left|\tilde{C}(k,s)\right|^{\rho}} = \tilde{C}_{PHAT}(k,s)\left|\tilde{C}(k,s)\right|^{1-\rho}$$

donde ρ es un parámetro entre 0 y 1. $\rho = 0$ corresponde al caso de la correlación cruzada normal y $\rho = 1$ corresponde al caso de la GCC-PHAT. Habitualmente se utiliza un valor por debajo de 1 aunque cercano, lo que permite modificar la GCC-PHAT poniendo más énfasis en los bins con correlación elevada, los que habitualmente corresponden a la señal, mientras que los bins con baja correlación corresponden al ruido. Más precisamente, hemos encontrado que un valor de $\rho = 0{,}8$ produjo la mayor eficiencia (fue de 0,75 en [4] y 0,78 en [5]).

Lamentablemente, esta nueva ponderación da mejor resultado que GCC-PHAT solo cuando hay presencia de un alto nivel de ruido de fondo. Situaciones alternativas en las que la nueva ponderación da posiblemente mejor resultado que GCC-PHAT son las tramas inactivas (es decir, la detección de actividad de voz detecta inactividad, lo que podría indicar un bajo nivel de voz), la presencia de transitorios, situaciones estéreo complejas, música, oradores interferentes, la presencia de música de fondo, voz que no es limpia. En entornos limpios, como la voz desprovista o con bajo nivel de ruido de fondo o música u otros componentes de la señal que puedan desviarse de la voz limpia, la GCC-PHAT da igualmente mejor resultado. Para obtener siempre los mejores resultados, se tornó necesario conmutar entre las dos estrategias dependiendo del contenido de la señal.

Para detectar la presencia de alto nivel de ruido de fondo en la señal, se utiliza un estimador de ruido junto con un detector de actividad de señal (SAD). Se puede estimar el nivel de la señal l_S en las tramas en que el SAD detecta una señal, mientras que el nivel del ruido l_N es estimado por el estimador de ruido. La presencia de alto nivel de ruido de fondo se detecta entonces comparando la relación señal a ruido $snr = l_S - l_N$ (en dB) con un umbral, por ej. si luego se detecta un alto nivel de ruido $snr < 35$.

Una vez que se sabe si la señal contiene un alto nivel de ruido de fondo o no, se toma la decisión de seleccionar la ponderación por PHAT o la ponderación por MCSP para computar la GCC (Etapa 4. en la técnica anterior). La identificación de picos (Etapa 6. En la técnica anterior) también se puede modificar dependiendo de si se ha detectado o no un alto nivel de ruido de fondo, por ejemplo, reduciendo el umbral.

A continuación se describe una realización preferida etapa por etapa.

0. Detección de alto nivel de ruido de fondo:

a. Se utiliza un estimador de ruido (por ejemplo de [6]) para estimar el nivel de ruido de fondo l_N. Se utiliza un filtro de alisado de IIR para alisar el nivel de ruido en el tiempo.

b. Se utiliza un detector de actividad de señal (por ejemplo de [6]) para clasificar una trama como activa o inactiva. Las tramas activas se utilizan luego para computar el nivel de la señal l_S, simplemente computando la energía de la señal y alisándola en el tiempo usando un filtro de alisado IIR.

c. Si la relación señal a ruido $snr = l_S - l_N$ (in dB) es inferior a un umbral (por ejemplo, 35 dB), es que se ha detectado un alto nivel de ruido de fondo.

1. Transformada discreta de Fourier: igual a la de la técnica anterior

2. Espectro de correlación cruzada: igual al de la técnica anterior

3. Alisado: igual al de la técnica anterior o como se describe en el presente documento, que se basa en la característica espectral

4. Ponderación:

Si se detecta un bajo nivel de ruido de fondo, se utiliza entonces la misma ponderación que en la técnica anterior (GCC-PHAT).

En caso de detectarse un alto nivel de ruido de fondo, se utiliza la ponderación MCSP

$$\tilde{C}_{MCSP}(k,s) = \frac{\tilde{C}(k,s)}{\left|\tilde{C}(k,s)\right|^{\rho}}$$

donde $0 < \rho < 1$ (por ejemplo, $\rho = 0{,}8$). Para mantener la salida de GCC-MCSP en el mismo rango que la salida de GCC-PHAT, se ejecuta una etapa de normalización adicional

$$\tilde{C}_{MCSP}(k,s) = \frac{\tilde{C}_{MCSP}(k,s)}{\frac{1}{N_{DFT}}\sum_{k=0}^{N_{DFT}-1}\left|\tilde{C}_{MCSP}(k,s)\right|}$$

5. Transformada inversa: igual a la de la técnica anterior

6. Identificación de picos: la identificación de picos se puede adaptar en caso de detectarse un alto nivel de ruido de fondo y se utiliza la ponderación de MCSP. Específicamente, se ha encontrado que un umbral más bajo es favorable.

Además, la figura 10a ilustra una implementación que difiere de la implementación de la figura 10c. En el ponderador 1036 de la figura 10c, el ponderador ejecuta el primer o el segundo procedimiento de ponderación. Sin embargo, en el ponderador 1036 ilustrado en la figura 10a, el ponderador únicamente ejecuta el segundo procedimiento de ponderación con respecto a la notación de la figura 10d o 10c. Esta implementación es ventajosa cuando se utiliza un filtro de alisado como el ilustrado en el bloque 1030, que ya ejecuta el primer procedimiento de ponderación con posterioridad al alisado o junto con el alisado en, por ejemplo, una única operación matemática o de hardware. Por consiguiente, en caso de ejecutar el primer procedimiento de ponderación que es la operación de normalización sin compresión alguna en el filtro de alisado, luego tanto el filtro de alisado 1030 por un lado como el ponderador real 1036 por el otro, corresponden al ponderador real para ponderar el espectro de correlación cruzada alisado o no alisado. Así, en la implementación de la figura 10a, la estimación de ruido 1038 solo es suministrada a un ponderador separado 1036 y la elección entre la salida del filtro de alisado 1030 que ya ha sido ponderada de acuerdo con el procedimiento de ponderación y la elección entre la salida del ponderador real 136 de la figura 10a se realiza mediante cierta configuración del procesador 1040 que utiliza automáticamente la salida del filtro de alisado 1030, cuando el ponderador 1036 no emite ninguna señal de salida sino que automáticamente prioriza la salida del ponderador 1036 sobre la salida del filtro de alisado 1030, cuando el ponderador 1036 produce una salida. Luego, se utiliza la estimación de ruido 1038 o, como se describe en otras figuras, la señal de control 1050 para activar o desactivar el ponderador 1036. De esa manera, se puede implementar el ponderador real para ponderar el espectro de correlación cruzada alisado o no alisado que utiliza un procedimiento de ponderación de primer orden de muchas maneras diferentes, tales como en el modo específico de activación/desactivación de la figura 10a o el modo de dos núcleos de la figura 10d con un conmutador de entrada o salida o de acuerdo con un núcleo único de procedimiento de ponderación que, dependiendo de la señal de control, selecciona uno u otro procedimiento de ponderación o adapta un procesador de ponderación general para ejecutar el primer o el segundo procedimiento de ponderación.

A continuación, se describe una realización preferida, en la que se ejecuta un alisado antes de la ponderación. En este contexto, las funcionalidades del estimador de características espectrales también están reflejadas en la figura 4e, elementos 453, 454, de una realización preferida.

Además, las funcionalidades de la calculadora de espectro de correlación cruzada 1020 también están reflejadas en el elemento 452 de la figura 4e que se describe más adelante en una realización preferida.

En consecuencia, las funcionalidades del filtro de alisado 1030 también están reflejadas en el elemento 453 en el contexto de la figura 4e que se describe más adelante. Además, también se describen las funcionalidades del procesador 1040 en el contexto de la figura 4e en una realización preferida, como elementos 456 a 459.

También se describen las realizaciones preferidas del procesador 1040 en la figura 10c.

Preferentemente, la estimación de características espectrales calcula un nivel de ruido o una tonalidad del espectro, donde una implementación preferida consiste en el cálculo de una medida de la planitud espectral que se aproxima a 0 en el caso de señales tonales o no ruidosas y que se aproxima a 1 en el caso de señales ruidosas o de tipo ruido.

Específicamente, el filtro de alisado está configurado entonces para aplicar un alisado más fuerte con un primer grado de alisado en el tiempo en el caso de una primera característica menos ruidosa o una primera característica más tonal, o para aplicar un alisado más tenue con un segundo grado de alisado en el tiempo en el caso de una segunda característica más ruidosa o una segunda característica menos tonal.

Específicamente, el primer alisado es mayor que el segundo grado de alisado, donde 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 la planitud espectral.

Además, como se ilustra en la figura 11a, el procesador es implementado preferentemente para normalizar el espectro de correlación cruzada alisado de la manera ilustrada en 456 en la figura 4e y 11a antes de la ejecución del cálculo de la representación en el dominio del tiempo en la etapa 1031 correspondiente a las etapas 457 y 458 de la realización de la figura 4e. Sin embargo, como también se esquematiza en la figura 11a, el procesador puede operar asimismo sin la normalización de la etapa 456 de la figura 4e. En ese caso, el procesador está configurado para analizar la representación en el dominio del tiempo según lo ilustrado en el bloque 1032 de la figura 11a para hallar la diferencia de tiempo entre canales. Este análisis puede ser ejecutado de cualquier manera conocida y dará lugar a una robustez mejorada, puesto que el análisis se lleva a cabo basándose en el espectro de correlación cruzada que ha sido alisado de acuerdo con la característica espectral.

Como se ilustra en la figura 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 figura 11b que corresponde al elemento 458 de la figura 4e y un procesamiento adicional subsiguiente 1033 que utiliza una operación de búsqueda/identificación de picos dentro de la representación en el dominio del tiempo filtrada por paso bajo.

Como se ilustra en la figura 11c, la implementación preferida de la operación de identificación de picos o búsqueda de picos consiste en ejecutar esta operación empleando un umbral variable. Específicamente, el procesador está configurado para ejecutar la operación de búsqueda de picos/identificación de picos dentro de la representación en el dominio del tiempo derivada del espectro de correlación cruzada alisado mediante la determinación 1034 de un umbral variable a partir de la representación en el dominio del tiempo y 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 la que la diferencia de tiempo entre canales se determina en términos de retardo de tiempo asociado a un pico que está en una relación predeterminada con el umbral, que es mayor que el umbral variable.

Como se ilustra en la figura 11d, una realización preferida ilustrada en el pseudo código relacionado con la figura 4e-b descrita más adelante consiste en la clasificación 1034a de valores de acuerdo con su magnitud. Luego, como se indica en el elemento 1034b de la figura 11d, se determinan los valores más altos, por ejemplo 10 o 5 % de los valores.

Seguidamente, como se indica en la etapa 1034c, se multiplica un número tal como el número 3 por el valor más bajo del 10 o 5 % más alto a fin de obtener el umbral variable.

Como se afirmó, se determina preferentemente el 10 o 5 % más elevado, aunque puede ser provechoso determinar el número más bajo del 50 % más elevado de los valores y utilizar un número multiplicador más elevado, como por ejemplo 10. Naturalmente, se determina incluso una cantidad más pequeña tal como el 3 % más elevado de los valores y luego se multiplica el valor más bajo entre este 3% más alto de los valores por un número que es, por ejemplo, igual a 2,5 o 2, es decir, inferior a 3. Por consiguiente, se pueden utilizar diferentes combinaciones de números y porcentajes en la realización ilustrada en la figura 11d. Además de los porcentajes, los números también pueden variar y son preferibles los números superiores a 1,5.

En una realización adicional ilustrada en la figura 11e, la representación en el dominio del tiempo se divide en subbloques, como se ilustra en el bloque 1101, y estos subbloques están indicados en 1300 en la figura 13. En este caso, se utilizan alrededor de 16 subbloques para el rango válido, por lo que cada subbloque tiene una distancia de retardo de 20. Sin embargo, el número de subbloques puede ser mayor que este valor o menor, y preferentemente superior a 3 e inferior a 50.

En la etapa 1102 de la figura 11e, se determina el pico en cada subbloque, y en la etapa 1103, se determina el pico promedio de todos los subbloques. Luego, en la etapa 1104, se determina un valor de multiplicación a que depende de una relación señal a ruido por un lado y, en una realización adicional, depende de la diferencia entre el umbral y el

pico máximo indicado a la izquierda del bloque 1104. Dependiendo de estos valores de entrada, se determina uno de, preferentemente, tres valores de multiplicación diferentes, donde el valor de multiplicación puede ser igual a a_{bajo}, a_{alto} y $a_{más\ bajo}$.

Luego, en la etapa 1105, el valor de multiplicación a determinado en el bloque 1104 se multiplica por el umbral promedio a fin de obtener el umbral variable que luego se utiliza en la operación de comparación en el bloque 1106. Para la operación de comparación, una vez más se puede utilizar la representación en el dominio del tiempo ingresada en el bloque 1101 o se pueden utilizar los picos ya determinados en cada subbloque como se indica en el bloque 1102.

A continuación, se esbozan otras realizaciones con referencia a la evaluación y detección de un pico dentro de la función de correlación cruzada en el dominio del tiempo.

La evaluación y detección de un pico dentro de la función de correlación cruzada en el dominio del tiempo que se produce como resultado del método de correlación cruzada generalizada (GCC-PHAT) para estimar la diferencia de tiempo entre canales (ITD) no siempre es sencilla debido a las diferentes situaciones de entrada. La entrada de voz limpia puede dar lugar a una función de correlación cruzada con poco desvío con un pico fuerte, mientras que la voz en un ambiente resonante y ruidoso puede producir un vector con gran desviación y picos con magnitud más baja, aunque de todos modos sobresalientes, que indican la existencia de ITD. Se describe un algoritmo de detección de picos que es adaptativo y flexible para dar lugar a las diferentes situaciones de entrada.

Debido a las restricciones por retardo, el sistema en general puede hacerse cargo del alineamiento en el tiempo de los canales hasta cierto límite, es decir ITD_MAX. El algoritmo propuesto está destinado a detectar si existe una ITD válida en los siguientes casos:

- ITD válida debido a un pico sobresaliente. Hay presencia de un pico sobresaliente dentro de los límites [-ITD_MAX, ITD_MAX] de la función de correlación cruzada.

- Falta de correlación. Cuando no hay correlación entre los dos canales, no hay pico sobresaliente. Se debe definir un umbral, por encima del cual el pico es lo suficientemente fuerte para ser considerado valor de ITD válido. De lo contrario, no se debe señalizar ningún tratamiento de la ITD, lo que significa que la ITD es ajustada a cero y no se realiza la alineación en el tiempo.

- ITD fuera de límite. Los picos fuertes de la función de correlación cruzada fuera de la región [-ITD_MAX, ITD_MAX] deben ser evaluados para determinar si existen ITD que yacen fuera de la capacidad de manejo del sistema. En este caso no se debe señalizar ningún tratamiento de la ITD y, por consiguiente, no se realiza la alineación en el tiempo.

Para determinar si la magnitud de un pico es lo suficientemente elevada para ser considerada un valor de diferencia de tiempo, se debe definir un umbral adecuado. Con las diferentes situaciones de entrada, la salida de la función de correlación cruzada varía dependiendo de diferentes parámetros, por ejemplo, el ambiente (ruido, reverberación, etc.), la configuración del micrófono (AB, M/S, etc.). Por lo tanto, es esencial definir el umbral de manera adaptativa.

En el algoritmo propuesto, el umbral se define calculando, en primer lugar, la media de un cómputo aproximado de la envolvente de la magnitud de la función de correlación cruzada dentro de la región [-ITD_MAX, ITD_MAX] (figura 13), luego se pondera el promedio de manera correspondiente dependiendo de la estimación de la SNR.

A continuación se presenta una descripción paso a paso del algoritmo.

La salida de la DFT inversa de la GCC-PHAT, que representa la correlación cruzada en el dominio del tiempo, es reordenada de retardos de tiempo negativos a positivos (figura 12).

El vector de correlación cruzada se divide en tres áreas principales: el área de interés, es decir [-ITD_MAX, ITD_MAX] y el área fuera de los límites de ITD_MAX, es decir los retardos de tiempo inferiores a –ITD_MAX (max_bajo) y superiores a ITD_MAX (max_alto). Se detectan los picos máximos de las áreas “fuera de límite” y se guardan para compararlos con el pico máximo detectado en el área de interés.

Para determinar si hay una ITD válida, se tiene en cuenta el área subvector [-ITD_MAX, ITD_MAX] de la función de correlación cruzada. El subvector se divide en N subbloques (figura 13).

Para cada subbloque, se encuentra y guarda la magnitud del pico máximo peak_sub y la posición de retardo de tiempo equivalente index_sub.

Se determina el máximo de los máximos locales peak_max y se compara con el umbral para determinar la existencia

de un valor válido de ITD.

Se compara el valor máximo peak_max con max_bajo y max_alto. Si peak_max es menor que uno de los dos, luego no se señaliza un tratamiento de itd en absoluto y no se realiza la alineación en el tiempo. Debido al límite de manejo de ITD del sistema, no es necesario evaluar las magnitudes de los picos fuera de límite.

Se calcula la media de las magnitudes de los picos:

$$peak_{media} = \frac{\Sigma_N\, peak\_sub}{N}$$

Luego se computa el umbral $thres$ ponderando $peak_{media}$ con un factor de ponderación a_w dependiente de la SNR

$$umbral = a_w peak_{media}, \text{ donde } a_w = \begin{cases} a_{bajo}, SNR \leq SNR_{umbral} \\ a_{alto,}\ SNR > SNR_{umbral} \end{cases}$$

En los casos en que $SNR \ll SNR_{umbral}$ y $|thres - peak\_max| < \varepsilon$, también se compara la magnitud del pico con un umbral ligeramente más relajado ($a_w = a_{más\ bajo}$), a fin de evitar el rechazo de un pico sobresaliente con altos picos adyacentes. Los factores de ponderación podrían ser, por ejemplo $a_{alto} = 3$, $a_{bajo} = 2{,}5$ y $a_{más\ bajo} = 2$, mientras que la SNR_{umbral} podría ser, por ejemplo, de 20dB y el límite $\varepsilon = 0{,}05$.

Los rangos preferidos son de 2,5 a 5 en el caso de a_{alto}; de 1,5 a 4 en el caso de a_{bajo}; de 1,0 a 3 en el caso de $a_{más\ bajo}$; de 10 a 30 dB para SNR_{umbral}; y de 0,01 a 0,5 para ε, donde a_{alto} es mayor que a_{bajo} que es mayor que $a_{más\ bajo}$.

Si *peak_max > umbral* se devuelve el retardo de tiempo equivalente en forma de ITD estimada, de lo contrario no se señaliza ningún manejo de ITD (ITD=0). Más adelante se describen otras realizaciones con respecto a la figura 4e.

La figura 11f ilustra la implementación preferida de la determinación de una salida válida de ITD (diferencia de tiempo entre canales).

Los subbloques de la representación en el dominio del tiempo del espectro de correlación cruzada ponderado y alisado o no alisado son ingresados a una etapa de determinación dentro del procesador 1040. Esta etapa de determinación 1120 determina un rango válido y un rango inválido dentro de una representación en el dominio del tiempo derivada del espectro de correlación cruzada ponderado y alisado o no alisado. En la etapa 1121, se determina un pico máximo dentro del rango inválido, y en la etapa 1122, se determina un pico máximo dentro del rango válido. Específicamente, se determina por lo menos un pico máximo dentro del rango inválido y se determina por lo menos un pico máximo dentro del rango válido. En el bloque 1123, se comparan los picos máximos del rango válido y el rango inválido. En el caso del pico válido, es decir, cuando el pico máximo del rango válido es mayor que el "pico inválido", el pico máximo en el rango inválido, en ese caso sí se ejecuta la determinación de ITD 1124 y se emite una salida de ITD válida. Cuando, por el contrario, se detecta que un "pico inválido" es mayor que el "pico válido" o que el pico inválido tiene el mismo tamaño que el pico válido, en ese caso no se emite una salida válida y, preferentemente, se emite un mensaje de error o se realiza cualquier acción comparable a fin de llamar la atención del procesador a este hecho.

A continuación se describe una implementación preferida de la presente invención dentro del bloque 1050 de la figura 10b en relación con un procesador adicional de señales con respecto a las figuras 1 a 9c, es decir, en el contexto de un procesamiento/codificación estéreo/multicanal y alineación en el tiempo de dos canales.

Sin embargo, como se indicó y como se ilustra en la figura 10b, existen muchos otros campos en los que también se puede ejecutar un procesamiento adicional de las señales utilizando la diferencia de tiempo entre canales determinada.

La figura 1 ilustra un aparato para codificar una señal multicanal con por lo menos dos canales. La señal multicanal 10 es ingresada en un determinador de parámetros 100 por un lado y un alineador de señales 200 por el otro. El determinador de parámetros 100 determina, por un lado, un parámetro de alineación de la banda ancha y, por el otro, una pluralidad de parámetros de alineación de banda estrecha a partir de la señal multicanal. Estos parámetros son emitidos a través de una línea de parámetros 12. Además, estos parámetros también son enviados a través de otra línea de parámetros 14 a una interfaz de salida 500 según lo ilustrado. En la línea de parámetros 14, se envían otros parámetros adicionales tales como parámetros de nivel del determinador de parámetros 100 a la interfaz de salida 500. El alineador de señales 200 está configurado para alinear dichos por lo 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 a través de la línea de parámetros 10 para obtener canales alineados 20 a la salida del alineador de señales 200. Estos canales alineados 20 son enviados a un procesador de señales 300 que está configurado para calcular

una señal media 31 y una señal lateral 32 a partir de los canales alineados recibidos a través de la línea 20. El aparato para codificar comprende asimismo un codificador de señales 400 para codificar la señal media a partir de la línea 31 y la señal lateral a partir de la línea 32 para obtener una señal media 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 media codificada obtenida 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 obtenidos de la línea 14 y, opcionalmente, un parámetro de nivel de la línea 14 y, también opcionalmente, un parámetro de relleno estéreo generado por el codificador de señales 400 y enviado a la interfaz de salida 500 a través de la línea de parámetros 43.

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 en realidad los parámetros de la banda estrecha. Por lo tanto, en esta realización, el alineador de señales 200 envía los canales alineados en la banda ancha de regreso al determinador de parámetros 100 a través de una línea de conexión 15. Luego, el determinador de parámetros 100 determina la pluralidad de parámetros de alineación de banda estrecha de una característica de banda ancha de la señal multicanal ya alineada. En otras realizaciones, sin embargo, se determinan los parámetros sin esta secuencia de procedimientos específica.

La figura 4a ilustra una implementación preferida donde se realiza la secuencia específica de etapas que incluye la línea de conexión 15. En la etapa 16, se determina el parámetro de alineación de banda ancha empleando los dos canales y se obtiene el parámetro de alineación de banda ancha tal como una diferencia de tiempo entre canales o parámetro de ITD. Luego, en la etapa 21, los dos canales son alineados por el alineador de señales 200 de la figura 1 usando el parámetro de alineación de banda ancha. Seguidamente, en la etapa 17, se determinan los parámetros de la banda estrecha utilizando 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 fase entre canales correspondientes a diferentes bandas de la señal multicanal. Luego, en la etapa 22, se alinean los valores espectrales de cada banda de parámetros utilizando el parámetro de alineación de banda estrecha correspondiente para esta banda específica. Cuando se ejecuta este procedimiento en la etapa 22 para cada banda para cual se dispone un parámetro de alineación de la banda estrecha, quedan disponibles el primer y segundo canales izquierdo y derecho alineados para continuar el procesamiento de señales por el procesador de señales 300 de la figura 1.

La figura 4b ilustra una implementación adicional del codificador multicanal de la figura 1 en la cual se ejecutan varios procedimientos en el dominio de la frecuencia.

Específicamente, el codificador multicanal comprende asimismo un convertidor tiempo–espectro 150 para convertir una señal multicanal en el dominio del tiempo a una representación espectral de dichos por lo menos dos canales en el dominio de la frecuencia.

Además, como se indica en 152, tanto el determinador de parámetros, como el alineador de señales y el procesador de señales ilustrados en 100, 200 y 300 de la figura 1 operan en el dominio de la frecuencia.

Además, el codificador multicanal y, específicamente, el procesador de señales comprende asimismo un convertidor espectro–tiempo 154 para generar una representación en el dominio del tiempo de la señal media por lo menos.

Preferentemente, el convertidor espectro a tiempo convierte asimismo una representación espectral de la señal lateral también determinada por los procedimientos representados en el bloque 152 a una representación en el dominio del tiempo y entonces el codificador de señales 400 de la figura 1 está configurado para codificar además la señal media y/o la señal lateral como señales en el dominio del tiempo dependiendo de la implementación específica del codificador de señales 400 de la figura 1.

Preferentemente, el convertidor tiempo–espectro 150 de la figura 4b está configurado para implementar las etapas 155, 156 y 157 de la figura 4c. Específicamente, la etapa 155 comprende dotar a una ventana de análisis de por lo menos una porción de relleno con ceros en un extremo de la misma y, específicamente, una porción de relleno con ceros en la porción inicial de la ventana y una porción de relleno con ceros en la porción final de la ventana según lo ilustrado, por ejemplo, más adelante en la figura 7. Además, la ventana de análisis tiene asimismo rangos de superposición o porciones de superposición en la primera mitad de la ventana y en la segunda mitad de la ventana y, además, preferentemente una parte media que es un rango sin superposición, según el caso.

En la etapa 156, se enventana cada canal utilizando la ventana de análisis con rangos de superposición. Específicamente, se enventana cada canal utilizando la ventana de análisis de tal manera que se obtenga un primer bloque del canal. Seguidamente, se obtiene un segundo bloque del mismo canal que tiene un determinado rango de superposición con el primer bloque y así sucesivamente, de tal manera que, por ejemplo, después de cinco

operaciones de enventanado, haya cinco bloques de muestras enventanadas de cada canal disponibles que luego se transforman individualmente a una representación espectral como se indica en 157 en la figura 4c. Se ejecuta el mismo procedimiento también para el otro canal de manera que, al final de la etapa 157, se disponga de una secuencia de bloques de valores espectrales y, específicamente, valores espectrales complejos tales como valores espectrales de DFT o muestras de subbandas complejas.

En la etapa 158, que se ejecuta por el determinador de parámetros 100 de la figura 1, se determina un parámetro de alineación de la banda ancha y en la etapa 159, que se ejecuta por la alineación de señales 200 de la figura 1, se realiza un desplazamiento circular usando el parámetro de alineación de banda ancha. En la etapa 160, una vez más ejecutada por el determinador de parámetros 100 de la figura 1, se determinan los parámetros de alineación de banda estrecha correspondientes a las bandas/subbandas individuales y, en la etapa 161, se hacen rotar los valores espectrales para cada banda utilizando los correspondientes parámetros de alineación de banda estrecha determinados respecto de las subbandas específicas.

La figura 4d ilustra otros procedimientos ejecutados por el procesador de señales 300. Específicamente, el procesador de señales 300 está configurado para calcular una señal media y una señal lateral como se indica en la etapa 301. En la etapa 302, se puede ejecutar algún tipo de procesamiento adicional de la señal lateral y luego, en la etapa 303, se vuelve a transformar cada bloque de la señal media y la señal lateral 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 ejecuta una operación de superposición y suma de la señal media por un lado y una operación de superposición y suma correspondiente a la señal lateral por el otro para obtener, en última instancia, las señales media/lateral en el dominio del tiempo.

Específicamente, las operaciones ejecutadas en las etapas 304 y 305 dan lugar a un tipo de fundido cruzado desde un bloque de la señal media o de la señal lateral o en el siguiente bloque de la señal media y la señal lateral de tal manera que, incluso cuando tiene lugar algún cambio de parámetro tal como el parámetro de diferencia de tiempo entre canales o el parámetro de diferencia de fase entre canales, esto de todos modos resulta inaudible en las señales media/lateral en el dominio del tiempo obtenidas en la etapa 305 de la figura 4d.

La nueva codificación estéreo con bajo retardo es una codificación estéreo conjunta medio/lateral (M/S) que explota ciertas indicaciones espaciales, donde el canal medio es codificado por un codificador de un solo núcleo primario y el canal lateral es codificado en un codificador de núcleo secundario. Los principios del codificador y decodificador están ilustrados en las figuras 6a, 6b.

El procesamiento estéreo se realiza principalmente en el dominio de la frecuencia (FD). Opcionalmente se puede ejecutar cierto procesamiento estéreo en el dominio del tiempo (TD) antes del análisis frecuencia. Éste es el caso del cómputo de la ITD, que se puede computar y aplicar antes del análisis frecuencial para alinear los canales en el tiempo antes de proseguir el análisis y procesamiento estéreo. Por otro lado, el procesamiento de la ITD se puede realizar directamente en el dominio de la frecuencia. Dado que los codificadores de voz habituales como ACELP no contienen ninguna descomposición interna de tiempo–frecuencia, la codificación estéreo agrega un banco de filtros modulado complejo adicional por medio de un banco de filtro de análisis y síntesis antes del codificador del núcleo y otra etapa de banco de filtro de análisis y síntesis después del decodificador del núcleo. En la realización preferida, se emplea una DFT sobremuestreada con escasa región de superposición. Sin embargo, en otras realizaciones, se puede emplear cualquier descomposición de tiempo–frecuencia con valor complejo y con similar resolución temporal.

El procesamiento estéreo consiste en computar las indicaciones espaciales: la diferencia de tiempo entre canales (ITD), las diferencias de fase entre canales (IPD) y las diferencias de nivel entre canales (ILD). La ITD y las IPD se utilizan en la señal estéreo de entrada para alinear los dos canales L y R (izquierdo y derecho) en el tiempo y en fase. La ITD se computa en la banda ancha o en el dominio del tiempo mientras que las IPD e ILD se computan para cada una o para partes de las bandas de parámetros, lo que corresponde a una descomposición no uniforme del espacio frecuencia. Una vez alineados los dos canales, se aplica un estéreo conjunto M/S, donde luego se predice la señal lateral a partir de la señal media. La ganancia de predicción es derivada de las ILD.

La señal media es codificada a su vez por un codificador de núcleo primario. En la realización preferida, el codificador de núcleo primario es la norma 3GPP EVS, o una codificación derivada de esta, que pueda conmutar entre un modo de codificación de voz, ACELP, y un modo de música basado en la transformación por MDCT. Preferentemente, ACELP y la codificación basada en MDCT son admitidas por los módulos de extensión de ancho de banda en el dominio del tiempo (TD-BWE) y o de llenado de brechas inteligente (IGF), respectivamente.

En primer lugar se predice la señal lateral en el canal medio utilizando las ganancias de predicción derivadas de las ILD. También se pueden predecir las residuales mediante una versión retardada de la señal media o codificarlas directamente mediante un codificador de núcleo secundario, en la realización preferida en el domino de la MDCT. El procesamiento estéreo realizado en el codificador está resumido en la figura 5, que se explica más adelante.

La figura 2 ilustra un diagrama de bloques de una realización de un aparato para decodificar una señal multicanal codificada recibida en la línea de entrada 50.

En particular, la señal se recibe por una interfaz de entrada 600. Conectados a la interfaz de entrada 600 se encuentran un decodificador de señales 700 y un desalineador de señales 900. Además, un procesador de señales 800 está conectado a un decodificador de señales 700 por un lado y está conectado al desalineador de señales por el otro.

En particular, la señal multicanal codificada comprende una señal media codificada, una señal lateral codificada, información acerca del parámetro de alineación de banda ancha e información acerca de la pluralidad de parámetros de banda estrecha. Por consiguiente, la señal multicanal codificada en la línea 50 puede ser exactamente la misma señal emitida por la interfaz de salida de 500 de la figura 1.

Sin embargo, es importante señalar aquí que, a diferencia de lo ilustrado en la figura 1, el parámetro de alineación de banda ancha y la pluralidad de parámetros de alineación de banda estrecha incluidos en la señal codificada pueden ser, en cierta forma, exactamente los parámetros de alineación utilizados por el alineador de señales 200 de la figura 1 aunque también pueden ser, por otra parte, los valores inversos de los mismos, es decir, parámetros que se pueden utilizar mediante operaciones exactamente iguales a las ejecutadas por el alineador de señales 200 aunque con valores inversos por lo que se obtiene la desalineación.

Por consiguiente, la información sobre los parámetros de alineación puede consistir en los parámetros de alineación utilizados por el alineador de señales 200 de la figura 1 o puede consistir en valores inversos, es decir, "parámetros de desalineación" reales. Además, estos parámetros son cuantificados por lo general de una determinada manera, como se describe más adelante con respecto a la figura 8.

La interfaz de entrada 600 de la figura 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 media/lateral codificadas y envía esta información a través de la línea de parámetros 610 al desalineador de señales 900. Por otro lado, la señal media codificada se envía al decodificador de señales 700 a través de la línea 601 y la señal lateral codificada se envía al decodificador de señales 700 a través de la línea de señales 602.

El decodificador de señales está configurado para decodificar la señal media codificada y para decodificar la señal lateral codificada a fin de obtener una señal media decodificada en la línea 701 y una señal lateral decodificada en la línea 702. Estas señales son utilizadas por el procesador de señales 800 para calcular una señal de primer canal decodificada o señal izquierda decodificada y para calcular una señal de segundo canal decodificada o señal de canal derecho decodificada a partir de la señal media decodificada y la señal lateral decodificada, y el primer canal decodificado y el segundo canal decodificado son emitidos por 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 utilizando la información sobre el parámetro de alineación de banda ancha y utilizando además la información sobre la pluralidad de parámetros de alineación de banda estrecha para obtener una señal multicanal decodificada, es decir, una señal decodificada que consta de por lo menos dos canales decodificados y desalineados en las líneas 901 y 902.

La figura 9a ilustra una secuencia de etapas preferida ejecutada por el desalineador de señales 900 de la figura 2. Específicamente, la etapa 910 recibe canales izquierdo y derecho alineados disponibles en las líneas 801, 802 de la figura 2. En la etapa 910, el desalineador de señales 900 desalinea las subbandas individuales utilizando la información sobre los parámetros de alineación de banda estrecha a fin de obtener los canales primero y segundo o izquierdo y derecho desalineados de fase y decodificados 911a y 911b. En la etapa 912, los canales son desalineados utilizando el parámetro de alineación de banda ancha por lo que, en 913a y 913b, se obtienen los canales desalineados en la fase y en el tiempo.

En la etapa 914, se ejecuta cualquier procesamiento adicional que comprende el uso de enventanado o cualquier operación de superposición y suma o, en general, alguna operación de fundido cruzado a fin de obtener, en 915a o 915b, una señal decodificada con reducción de distorsiones o sin distorsiones, es decir, canales decodificados que no tienen distorsiones, aunque sí ha habido, por lo general, parámetros de desalineación que varían en el tiempo correspondientes a la banda ancha por un lado y a la pluralidad de bandas estrechas por el otro.

La figura 9b ilustra una implementación preferida del decodificador multicanal ilustrado en la figura 2.

En particular, el procesador de señales 800 de la figura 2 comprende un convertidor tiempo–espectro 810.

El procesador de señales comprende además un convertidor medio/lateral a izquierdo/derecho 820 para calcular, a partir de una señal media M y una señal lateral S, una señal izquierda L y una señal derecha R.

Sin embargo, es importante que, para calcular L y R mediante la conversión de medio/lateral a izquierdo/derecho en el bloque 820, no se utilice necesariamente la señal lateral S. Por el contrario, como se explica más adelante, en un principio solo se calculan las señales izquierda/derecha utilizando un parámetro de ganancia derivado de un parámetro de diferencia de nivel entre canales ILD. En general, también se tienen en cuenta la ganancia de predicción como forma de ILD. La ganancia se puede derivar de la ILD, aunque también se puede computar directamente. Es preferible no computar más la ILD, sino computar directamente la ganancia de predicción y transmitir y utilizar la ganancia de predicción en el decodificador, en lugar del parámetro de ILD.

Por lo tanto, en esta implementación, solo se utiliza la señal lateral S en el actualizador de canales 830 que funciona para producir una mejor señal izquierda/derecha utilizando la señal lateral transmitida S como se indica mediante la línea de desvío 821.

Por lo tanto, el convertidor 820 funciona usando un parámetro de nivel obtenido mediante una entrada de parámetro de nivel 822 y sin utilizar, en realidad, la señal lateral S sino que el actualizador de canales 830 funciona entonces utilizando el lateral 821 y, dependiendo de la implementación específica, utilizando un parámetro de relleno estéreo recibido a través de la línea 831. El alineador de señales 900 comprende luego un desalineador de fase y escalador de energía 910. El escalado de energía se controla por un factor de escala derivado por una calculadora de factores de escala 940. La calculadora de factores de escala 940 está alimentada por la salida del actualizador de canales 830. Basándose en los parámetros de alineación de banda estrecha recibidos a través de la entrada 911, se ejecuta la desalineación de la fase y, en el bloque 920, sobre la base del parámetro de alineación de banda ancha recibido a través de la línea 921, se ejecuta la desalineación en el tiempo. Por último, se ejecuta una conversión de espectro a tiempo 930 para obtener, en última instancia, la señal decodificada.

La figura 9c ilustra otra secuencia de etapas que típicamente se llevan a cabo dentro de los bloques 920 y 930 de la figura 9b en una realización preferida.

Específicamente, los canales desalineados en banda estrecha son ingresados en la funcionalidad de desalineación de la banda ancha que corresponde al bloque 920 de la figura 9b. Se ejecuta una DFT o cualquier otra transformada en el bloque 931. Después del cálculo real de las muestras en el dominio del tiempo, se ejecuta un enventanado de síntesis opcional utilizando una ventana de síntesis. La ventana de síntesis es preferentemente exactamente igual a la ventana de análisis o se la deriva de la ventana de análisis, por ejemplo por interpolación o decimación, aunque depende en cierto modo de la ventana de análisis. Esta dependencia es preferentemente tal que los factores de multiplicación definidos por dos ventanas superpuestas suman uno por cada punto en el rango de superposición. Por consiguiente, después de la ventana de síntesis en el bloque 932, se ejecuta una operación de superposición y una posterior operación de suma. Por otro lado, en lugar del enventanado de síntesis y la operación de superposición y suma, se ejecuta cualquier fundido cruzado entre bloques subsiguientes para cada canal a fin de obtener, como ya se señaló en el contexto de la figura 9a, una señal decodificado con número reducido de distorsiones.

Al considerar la figura 6b, se torna evidente que las operaciones de decodificación en sí correspondientes a la señal media, es decir, el “decodificador EVS”, por un lado y, en el caso de la señal lateral, la cuantificación vectorial inversa VQ^{-1} y la operación de MDCT inversa (IMDCT) corresponden al decodificador de señales 700 de la figura 2.

Además, las operaciones de DFT efectuadas en los bloques 810 corresponden al elemento 810 de la figura 9b y las funcionalidades del procesamiento estéreo inverso y el desplazamiento de tiempo inverso corresponden a los bloques 800, 900 de la figura 2 y las operaciones de DFT inversa 930 de la figura 6b corresponden a la operación correspondiente efectuada en el bloque 930 de la figura 9b.

A continuación, se describe la figura 3 con más detalles. En particular, la figura 3 ilustra un espectro de DFT con líneas espectrales individuales. Preferentemente, el espectro de DFT o cualquier otro espectro ilustrado en la figura 3 es un espectro complejo y cada línea es una línea espectral compleja con magnitud y fase o con una parte real y una parte imaginaria.

Además, el espectro también se divide en diferentes bandas de parámetros. Cada banda de parámetros tiene por lo menos una y preferentemente más de una línea espectral. Además, las bandas de parámetros aumentan de menores a mayores frecuencias. Por lo general, el parámetro de alineación de banda ancha es un único parámetro de alineación de la banda ancha para todo el espectro, es decir, para un espectro que comprende todas las bandas 1 a 6 en la realización a modo de ejemplo de la figura 3.

Además, se incluye la pluralidad de parámetros de alineación de banda estrecha para que haya un único parámetro de alineación por cada banda de parámetros. Esto significa que el parámetro de alineación correspondiente a una banda siempre se aplica a todos los valores espectrales dentro de la banda correspondiente.

Más aun, aparte de los parámetros de alineación de banda estrecha, también se presentan parámetros de nivel para

cada banda de parámetros.

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

Además, se incluyen parámetros de llenado estéreo para un cierto número de bandas, excluyendo las bandas inferiores tales como, en la realización ilustrativa, las bandas 4, 5 y 6, mientras que hay valores espectrales de señal lateral para las bandas de parámetros inferiores 1, 2 y 3 y, en consecuencia, no existen parámetros de llenado estéreo para estas bandas inferiores donde se obtiene la coincidencia de la forma de onda utilizando la señal lateral en sí o una señal residual de predicción que representa la señal lateral.

Como ya se mencionó, existen más líneas espectrales en las bandas superiores tales como, en la realización de la figura 3, siete líneas espectrales en la banda de parámetros 6 contra 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 han de ser diferentes.

De todos modos, la figura 8 ilustra una distribución de los parámetros y el número de bandas para las cuales se proporcionan parámetros en una determinada realización en la que hay, a diferencia de la figura 3, 12 bandas en realidad.

Como se ilustra, el parámetro de nivel ILD se presenta para cada una de las 12 bandas y se cuantifica con una precisión de cuantificación representada por cinco bits por banda.

Además, los parámetros de alineación de banda estrecha IPD solo se proporcionan para las bandas inferiores hasta un límite de frecuencia de 2,5 kHz. Además, la diferencia de tiempo entre canales o parámetro de alineación de la banda ancha solo se provee en forma de parámetro único para todo el espectro, aunque con una precisión de cuantificación muy elevada representada por ocho bits para la totalidad de la banda.

Además, se proporcionan parámetros de llenado estéreo cuantificados de manera solo aproximada representados por tres bits por banda y no para las bandas inferiores por debajo de 1 kHz puesto que, para las bandas inferiores, se incluyen valores espectrales de la señal lateral o la señal lateral residual codificada en realidad.

A continuación, se resume un procesamiento preferido en el lado del codificador con respecto a la figura 5. En una primera etapa, se ejecuta un análisis de DFT del canal izquierdo y el derecho. Este procedimiento corresponde a las etapas 155 a 157 de la figura 4c. En la etapa 158, se calcula el parámetro de alineación de banda ancha y, específicamente, la diferencia de tiempo entre canales (ITD) y el parámetro de alineación de banda ancha preferido. Como se ilustra en 170, se ejecuta un desplazamiento en el tiempo de L y R (izquierda y derecha) en el dominio de la frecuencia. Alternativamente, este desplazamiento en el tiempo también se puede ejecutar en el dominio del tiempo. Luego se ejecuta una DFT inversa, se ejecuta el desplazamiento en el tiempo en el dominio del tiempo y se ejecuta una DFT directa adicional para contar, una vez más, con representaciones espectrales posteriores a la alineación utilizando el parámetro de alineación de banda ancha.

Se calculan los parámetros de ILD, es decir, los parámetros de nivel y los parámetros de fase (parámetros de IPD), para cada banda de parámetros en las representaciones desplazadas L y R, como se indica en la etapa 171. Esta etapa corresponde a la etapa 160 de la figura 4c, por ejemplo. Se rotan las representaciones L y R desplazadas en el tiempo en función de los parámetros de diferencia de fase entre canales como se indica en la etapa 161 de la figura 4c o de la figura 5. Seguidamente, se computan las señales media y lateral como se indica en la etapa 301 y, preferentemente, se efectúa una operación de conversión de energía que se describe más adelante. En una etapa subsiguiente 174, se ejecuta una predicción de S con M en función de la ILD y opcionalmente con una señal M pasada, es decir, una señal media de una trama anterior. Subsiguientemente, se ejecuta la DFT inversa de la señal media y la señal lateral que corresponde a las etapas 303, 304, 305 de la figura 4d de la realización preferida.

En la etapa final 175, se codifican la señal media en el dominio del tiempo m y, opcionalmente, la señal residual, como se ilustra en la etapa 175. Este procedimiento corresponde a lo ejecutado por el codificador de señales 400 de la figura 1.

En el decodificador en el procesamiento estéreo inverso, se genera la señal lateral en el dominio de la DFT y se predice en primer lugar a partir de la señal media de la siguiente manera:

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

donde g es una ganancia computada por cada banda de parámetros y en función de las diferencias de nivel entre canales (ILD) transmitidas.

Luego se puede refinar la residual de la predicción $Side - g \cdot Mid$ de dos maneras diferentes:

- Mediante una codificación secundaria de la señal residual:

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

donde 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 media decodificada anterior de la trama de DFT anterior:

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

donde g_{pred} es una ganancia predictiva transmitida por cada banda de parámetros.

Los dos tipos de refinación de la codificación se pueden mezclar dentro del mismo espectro de DFT. En la realización preferida, se aplica la codificación residual a las bandas de parámetros inferiores, mientras que se aplica la predicción residual al resto de las bandas. La codificación residual se ejecuta, en la realización preferida ilustrada en la figura 1, en el dominio de la MDCT después de sintetizar la señal lateral residual en el dominio del tiempo y transformarla mediante una MDCT. A diferencia de la DFT, la MDCT está sometida a muestreo crítico y es más adecuada para la codificación de audio. Los coeficientes de MDCT son directamente sometidos a cuantificación vectorial por un cuantificador vectorial de malla (*Lattice*), aunque por otro lado pueden estar cuantificados por un cuantificador escalar seguido por un codificador de entropía. Alternativamente, también se puede codificar la señal lateral residual en el dominio del tiempo mediante una técnica de codificación de voz o directamente en el dominio de la DFT.

1. Análisis de tiempo - frecuencia: DFT

Es importante que la descomposición extra de tiempo – frecuencia obtenida del procesamiento estéreo efectuado por las DFT permita un análisis de la escena auditiva satisfactorio sin aumentar significativamente el retardo total del sistema de codificación. Por defecto, se utiliza una resolución de tiempo de 10 ms (dos veces las tramas de 20 ms del codificador de núcleo). Las ventanas de análisis y síntesis son iguales y simétricas. La ventana está representada a 16 kHz de la velocidad de muestreo en la figura 7. Se puede observar que la región superpuesta es limitada para reducir el retardo generado y que también se agrega el relleno con ceros para contrarrestar el desplazamiento circular al aplicar la ITD en el dominio de la frecuencia, como se explica a continuación.

2. Parámetros estéreo

Los parámetros estéreo se pueden transmitir al máximo en la resolución temporal de la DFT estéreo. En el mínimo se pueden reducir a la resolución de trama del codificador de núcleo, es decir, 20 ms. Por defecto, cuando no se detectan transitorios, los parámetros se computan cada 20 ms a lo ancho de 2 ventanas de DFT. Las bandas de parámetros constituyen una descomposición no uniforme y no superpuesta del espectro después de aproximadamente 2 veces o 4 veces los anchos de banda rectangulares equivalentes. Por defecto, se utiliza una escala de 4 veces el ERB para un total de 12 bandas para un ancho de banda de frecuencia de 16 kHz (velocidad de muestreo de 32 kbps, estéreo de banda súper ancha). La figura 8 resume un ejemplo de configuración, en la cual la información lateral estéreo se transmite a aproximadamente 5 kbps.

3. Cómputo de ITD y alineación en tiempo de los canales

Las ITD se computan estimando el retardo de tiempo de llegada (TDOA) utilizando la correlación cruzada Generalizada con transformada de fase (GCC-PHAT):

$$ITD = argmax(IDFT(\frac{L_i(f)R^*_i(k)}{|L_i(f)R^*_i(k)|}))$$

donde L y R son los espectros de frecuencia de los canales izquierdo y derecho, respectivamente.

El análisis de frecuencia se puede ejecutar de forma independiente de la DFT empleada para el posterior procesamiento estéreo o puede ser compartido. El pseudo código para el cómputo de la ITD es el siguiente:

```
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);
h.cross_corr_alisado = (1-sfm)*h.cross_corr_alisado+sfm*tmp;
tmp = h.cross_corr_alisado ./ abs( h.cross_corr_alisado+eps );
tmp = ifft( tmp );
tmp = tmp([longitud(tmp)/2+1:longitud(tmp) 1:longitud(tmp)/2+1]);
tmp_sort = sort( abs(tmp) );
umbral = 3 * tmp_sort( vuelta(0.95*longitud(tmp_sort)) );
xcorr_time=abs(tmp(- ( h.stereo_itd_q_max - (longitud(tmp)-1)/2 - 1 ):- (
h.stereo_itd_q_min - (longitud(tmp)-1)/2 - 1 )));
 %alisado de salida para mejor detección
xcorr_time=[xcorr_time 0];
xcorr_time2=filter([0.25 0.5 0.25],1,xcorr_time);
 [m,i] = max(xcorr_time2(2:end));
if m > umbral
 itd = h.stereo_itd_q_max - i + 1;
si no
 itd = 0;
fin
```

La figura 4e ilustra un diagrama de flujo para implementar el pseudo código antes ilustrado a fin de obtener un cálculo firme y eficiente de una diferencia de tiempo entre canales como ejemplo del parámetro de alineación de banda ancha.

En el bloque 451, se ejecuta un análisis de DFT de las señales en el dominio del tiempo correspondientes a un primer canal (l) y un segundo canal (r). Este análisis de DFT es por lo general igual al análisis de DFT que se ha explicado en el contexto de las etapas 155 a 157 de la figura 5 o la figura 4c, por ejemplo.

A continuación se ejecuta una correlación cruzada para cada bin de frecuencia indicado en el bloque 452.

De esa manera, se obtiene un espectro de correlación cruzada para todo el rango espectral de los canales izquierdo y el derecho.

En la etapa 453, se calcula luego una medida de la planitud espectral a partir de los espectros de magnitud de L y R y, en la etapa 454, se selecciona la medida más grande de planitud espectral. Sin embargo, la selección de la etapa 454 no tiene que ser necesariamente la selección del más grande, sino que esta determinación de una única SFM de ambos canales también puede ser también la selección y el cálculo de sólo el canal izquierdo o sólo el canal derecho o puede ser el cálculo de un promedio ponderado de ambos valores de SFM.

En la etapa 455, se alisa luego el espectro de correlación cruzada en el tiempo dependiendo de la medida de la planitud espectral.

Preferentemente, la medida de la planitud espectral se calcula dividiendo la media geométrica del espectro de magnitud por la media aritmética del espectro de magnitud. De esa manera, los valores de la SFM son limitados entre cero y uno.

En la etapa 456, después se normaliza el espectro de correlación cruzada alisado por su magnitud y, en la etapa 457, se calcula una DFT inversa del espectro de correlación cruzada normalizado y alisado. En la etapa 458, se ejecuta preferentemente un filtro determinado en el dominio del tiempo, aunque este filtrado en el dominio del tiempo también puede ser omitido dependiendo de la implementación, si bien es preferible, como se indica más adelante.

En la etapa 459, se ejecuta una estimación de la ITD mediante la identificación de picos de la función de correlación cruzada generalizada de filtrado y ejecutando una determinada operación de determinación de umbrales.

En caso de no obtenerse un pico superior al umbral, luego la ITD es ajustada a cero y no se ejecuta una alineación en el tiempo correspondiente a este bloque.

El cómputo de la ITD también se puede resumir de la siguiente manera. La correlación cruzada se computa en el dominio de la frecuencia antes de alisarla dependiendo de la medida de la planitud espectral. La SFM está limitada entre 0 y 1. En el caso de las señales de ruido, la SFM es elevada (es decir, alrededor de 1) y el alisado es tenue. En el caso de la señal de tono, la SFM es baja y el alisado cobra fuerza. Luego se normaliza la correlación cruzada alisada por su amplitud antes de transformarla nuevamente al dominio del tiempo. La normalización corresponde a la transformada de fase de la correlación cruzada y se sabe que exhibe una mayor eficiencia que la correlación cruzada normal en ambientes de bajo ruido y resonancia relativamente elevada. La función en el dominio del tiempo así obtenida es filtrada en primer lugar para obtener una determinación de picos más eficiente. 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 ITD estimada no se considera fiable y se ajusta a cero.

Si se aplica la alineación en el tiempo en el dominio del tiempo, se computa la ITD en un análisis de DFT separado. El desplazamiento se realiza de la siguiente manera:

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

Requiere un retardo extra en el codificador, que es igual en el máximo a la máxima ITD absoluta que se puede manejar. La variación de ITD en el tiempo se alisa por el enventanado de análisis de la DFT.

Por otro lado, se puede ejecutar la alineación en el tiempo en el dominio de la frecuencia. En ese caso, el cómputo de la ITD y el desplazamiento circular están en el mismo dominio de la DFT, dominio compartido por este otro procesamiento estéreo. El desplazamiento circular está dado por:

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

Se necesita el relleno con ceros de las ventanas de DFT para simular un desplazamiento en el tiempo con un desplazamiento circular. El tamaño del relleno con ceros corresponde a la máxima ITD absoluta que se puede tratar. En la realización preferida, el relleno con ceros se divide de manera uniforme a ambos lados de la ventana de análisis, mediante la adición de 3,125 ms de ceros en ambos extremos. La máxima ITD absoluta es entonces de 6,25 ms. En una configuración de micrófonos A-B, esto corresponde, en el peor de los casos, a una distancia máxima de aproximadamente 2,15 metros entre los dos micrófonos. La variación de ITD en el tiempo se alisa por el enventanado de síntesis y superposición y suma de la DFT.

Es importante que al desplazamiento en el tiempo le siga un enventanado de la señal desplazada. Esta es una diferenciación principal con respecto a la codificación binaural de indicaciones (BCC) de la técnica anterior, donde se aplica el desplazamiento en el tiempo a la señal en ventana pero no se sigue enventanando en la etapa de síntesis. En consecuencia, todo cambio de ITD en el tiempo produce un transitorio artificial/clic en la señal decodificada.

4. Cómputo de IPD y rotación de canales

Las IPD se computan después de la alineación en el tiempo de los dos canales y esto por cada banda de parámetros o por lo menos hasta una $ipd\_\max\_band$ dada, dependiendo de la configuración estéreo.

$$IPD[b] = \text{ángulo}\left(\sum_{k=band_{limits[b]}}^{band_{limits[b+1]}} L[k]\,R^*[k]\right)$$

Luego se aplican las IPD a los dos canales para alinear sus fases:

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

donde $\beta = \text{atan2}(\sin(\text{IPD}_\text{i}[\text{b}]), \cos(\text{IPD}_\text{i}[\text{b}]) + \text{c}), \text{c} = 10^{\frac{ILD_i[b]}{20}}$ y b es el índice de la banda de parámetros al

cual pertenece el índice de frecuencia *k*. El parámetro β es responsable de distribuir la cantidad de rotación de fase entre los dos canales a la vez que alinea su fase. β depende de la IPD pero también del nivel de amplitud relativa de los canales, la ILD. Si un canal tiene una amplitud más elevada, se considera canal principal y resulta menos afectado por la rotación de fase que el canal con amplitud más baja.

5. Suma – diferencia y codificación de la señal lateral

La transformación por suma- diferencia se realiza en los espectros alineados en el tiempo y la fase de los dos canales de manera tal que se conserve la energía en la señal media.

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

donde $a = \sqrt{\frac{L'^2 + R'^2}{(L'+R')^2}}$ está limitado entre 1/1,2 y 1,2, es decir, -1,58 y +1,58 dB. La limitación evita las distorsiones al ajustar la energía de la M y S. Cabe señalar que esta conservación de la energía es menos importante cuando el tiempo y la fase se alinean de antemano. Por otra parte, se pueden aumentar o reducir los límites.

Asimismo, se predice la señal lateral S con M:

$$S'^{(f)} = S(f) - g(ILD)M(f)$$

donde $\mathrm{g(ILD)} = \frac{\mathrm{c}-1}{c+1}$, donde $\mathrm{c} = 10^{\frac{ILD_i[b]}{20}}$. Por otro lado, se puede hallar la ganancia de predicción óptima minimizando el error cuadrático medio (MSE) de la residual y las ILD deducidas mediante la ecuación anterior.

Se puede modelar la señal residual $S'^{(f)}$ de dos maneras: ya sea prediciéndola con el espectro retardado de M o codificándola directamente en el dominio de la MDCT en el dominio de la MDCT.

6. Decodificación estéreo

En primer lugar se convierte la señal media X y la señal lateral S a los canales izquierdo y derecho L y R de la siguiente manera:

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

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

Donde la ganancia g por banda de parámetros se deriva del parámetro de ILD: $\mathrm{g} = \frac{\mathrm{c}-1}{c+1}$, donde $\mathrm{c} = 10^{\frac{ILD_i[b]}{20}}$.

En el caso de las bandas de parámetros inferiores a cod_max_band, los dos canales se actualizan con la señal lateral decodificada:

$$L_i[\mathrm{k}] = L_i[k] + cod_{gain_i} \cdot S_i[k], \text{para } 0 \leq\ k < band\_limits[cod\_\max\ \_band],$$

$$R_i[\mathrm{k}] = R_i[k] - cod_{gain_i} \cdot S_i[k], \text{para } 0 \leq\ k < band\_limits[cod\_\max\ \_band],$$

En el caso de las bandas de parámetros superiores, se predice la señal lateral y se actualizan los canales de la siguiente manera:

$$L_i[\mathrm{k}] = L_i[k] + cod_{pred_i}[b] \cdot M_{i-1}[k], \text{para } band_{limits[b]} \leq\ k < band_{limits[b+1]},$$

$$R_i[\mathrm{k}] = R_i[k] - cod_{pred_i}[b] \cdot M_{i-1}[k], \text{para } band_{limits[b]} \leq k < band_{limits[b+1]},$$

Por último, se multiplican los canales por un valor complejo con el fin de restablecer la energía original y la fase entre canales de la señal estéreo:

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

$$R_i[\mathrm{k}] = a \cdot e^{j2\pi\beta - \mathrm{IPD_i[b]}} \cdot R_i[k]$$

donde

$$a = \sqrt{2 \cdot \frac{\sum_{k=band_{limits[b]}}^{band_{limits[b+1]}} M_i^2[k]}{\sum_{k=band_{limits[b]}}^{band_{limits[b+1]}-1} {L_i}^2[k] + \sum_{k=band_{limits[b]}}^{band_{limits[b+1]}-1} {R_i}^2[k]}}$$

donde a se define y limita de la manera antes definida, y donde $\beta = \mathrm{atan2}(\mathrm{sen}(\mathrm{IPD_i[b]}), \cos(\mathrm{IPD_i[b]}) + \mathrm{c})$, y donde atan2(x,y) es la tangente inversa de cuatro cuadrantes de x sobre y.

Por último, los canales son desplazados en el tiempo ya sea en el dominio del tiempo o de la frecuencia, dependiendo de las ITD transmitidas. Los canales en el dominio del tiempo son sintetizados por DFT inversas y superposición y suma.

Las características específicas de la invención se refieren a la combinación de indicaciones espaciales y codificación estéreo conjunta con suma-diferencia. Específicamente, se computan las indicaciones espaciales IDT e IPD y se aplican a los canales estéreo (izquierdo y derecho). Además, se calcula la suma-diferencia (señales M/S) y preferentemente se aplica una predicción de S con M.

En el lado del decodificador, se combinan las indicaciones espaciales de banda ancha y banda estrecha con la codificación estéreo conjunta con suma-diferencia. En particular, se predice la señal lateral con la señal media usando por lo menos una indicación espacial tal como ILD y se calcula una suma-diferencia inversa para obtener los canales izquierdo y derecho y, además, se aplican las indicaciones de banda ancha y banda estrecha a los canales izquierdo y derecho.

Preferentemente, el codificador tiene una ventana y superposición y suma con respecto a los canales alineados en el tiempo después de procesarlos utilizando la ITD. Además, el decodificador cuenta adicionalmente con una operación de enventanado y superposición y suma de las versiones desplazadas o desalineadas de los canales después de aplicar la diferencia de tiempo entre canales.

El cómputo de la diferencia de tiempo entre canales con el método de GCC-Phat es un método específicamente robusto.

El nuevo procedimiento es ventajoso con respecto a la técnica anterior puesto que obtiene la codificación por tasa de bits de audio estéreo o audio multicanal con bajo retardo. Está específicamente diseñado para ser robusto para las naturalezas diferentes de las señales de entrada y las diferentes configuraciones de la grabación multicanal o estéreo. En particular, la presente invención ofrece una buena calidad para la codificación de voz estéreo con baja tasa de bits.

Los procedimientos preferidos son de utilidad en la distribución de la transmisión de todos los tipos de contenido de audio estéreo o multicanal tal como voz y música por igual con calidad perceptual constante a una baja tasa de bits dada. Esas áreas de aplicación son aplicaciones de radio digital, transmisión por internet en tiempo real o comunicación de audio.

Si bien la presente invención ha sido descrita en términos de varias realizaciones, hay alteraciones y permutaciones que entran dentro del alcance de esta invención tal como se define en las reivindicaciones adjuntas. Se debe tener en cuenta además que hay numerosas maneras alternativas de implementación de los métodos y composiciones de la presente invención.

Si bien se han descrito algunos aspectos en el contexto de un aparato, es obvio que estos aspectos también representan una descripción del método correspondiente, en el cual un bloque o dispositivo corresponde a una etapa del método o a una característica de una etapa del método. De manera análoga, los aspectos descritos en el contexto de una etapa del método también representan una descripción de un bloque o elemento correspondiente o de una

característica de un aparato correspondiente. Algunas o todas las etapas del método pueden ser ejecutadas por (o utilizando) un aparato de hardware, como por ejemplo, un microprocesador, un ordenador programable o un circuito electrónico. En algunas realizaciones, una o más de las etapas más importantes del método pueden ser ejecutadas por ese tipo de aparato.

Dependiendo de ciertos requisitos de implementación, las realizaciones de la invención pueden ser implementadas en hardware o en software. La implementación se puede realizar empleando un medio de almacenamiento digital, por ejemplo, un disco blando, un DVD, un Blu–Ray, un CD, una ROM, una PROM, una EPROM, una EEPROM o una memoria FLASH, que tiene almacenadas en la misma señales control legibles electrónicamente, que cooperan (o tienen capacidad para cooperar) con un sistema informático programable de tal manera que se ejecute el método respectivo. Por lo tanto, el medio de almacenamiento digital puede ser legible por ordenador.

Algunas realizaciones según la invención comprenden un transportador de datos que tiene señales de control legibles electrónicamente, con capacidad para cooperar con un sistema informático programable de tal manera que se ejecute uno de los métodos descritos en el presente documento.

En general, las realizaciones de la presente invención pueden ser implementadas en forma de producto programa informático con un código de programa, donde el código de programa cumple la función de ejecutar uno de los métodos cuando se ejecuta el programa informático en un ordenador. El código de programa puede ser almacenado, por ejemplo, en un portador legible por una máquina.

Otras realizaciones comprenden el programa informático para ejecutar uno de los métodos descritos en el presente documento, almacenado en un portador legible por una máquina.

En otras palabras, una realización del método de la invención es, por lo tanto, un programa informático que tiene un código de programa para ejecutar uno de los métodos descritos en el presente documento cuando se ejecuta el programa informático en un ordenador.

Una realización adicional de los métodos de la invención es, por lo tanto, un portador de datos (o medio de almacenamiento digital, o medio legible por ordenador) que comprende, grabado en el mismo, el programa informático para ejecutar uno de los métodos descritos en el presente documento. El portador de datos, el medio de almacenamiento digital o el medio grabado son por lo general tangibles y/o no transitorios.

Una realización adicional del método de la invención es, por lo tanto, un flujo de datos o una secuencia de señales que representa el programa informático para ejecutar uno de los métodos descritos en el presente documento. El flujo de datos o la secuencia de señales puede estar configurada, por ejemplo, para ser transferida a través de una conexión de comunicación de datos, por ejemplo, a través de internet.

Una realización adicional comprende un medio de procesamiento, por ejemplo, un ordenador, un dispositivo lógico programable, configurado o adaptado para ejecutar uno de los métodos descritos en el presente documento.

Una realización adicional comprende un ordenador en la que se ha instalado el programa informático para ejecutar uno de los métodos descritos en el presente documento.

Una realización adicional según la invención comprende un aparato o un sistema configurado para transferir (por ejemplo, por vía electrónica u óptica) un programa informático para ejecutar uno de los métodos aquí descritos en el presente documento a un receptor. El receptor puede ser, por ejemplo, un ordenador, un dispositivo móvil, un dispositivo de memoria o similar. El aparato o sistema puede comprender, por ejemplo, un servidor de archivos para transferir un programa informático al receptor.

En algunas realizaciones, se puede utilizar un dispositivo lógico programable (por ejemplo, una matriz de puertas programables en el campo) para ejecutar algunas o todas las funcionalidades de los métodos descritos en el presente documento. En algunas realizaciones, una matriz de puertas programables en el campo puede cooperar con un microprocesador para ejecutar uno de los métodos descritos en el presente documento. Por lo general, los métodos son ejecutados preferentemente por cualquier aparato de hardware.

El aparato descrito en el presente documento puede ser implementado empleando un aparato de hardware o utilizando un ordenador, o utilizando una combinación de aparato de hardware y un ordenador.

Los métodos descritos en el presente documento se pueden poner en práctica empleando un aparato de hardware o utilizando un ordenador, o utilizando una combinación de aparato de hardware y computadora.

Las realizaciones anteriormente descritas 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 el presente documento han de resultar obvios para los expertos en la técnica. Por lo tanto, solo se pretende que queden limitados por el alcance de las siguientes reivindicaciones de patente y no por los detalles específicos presentados a manera de descripción y explicación de las realizaciones en el presente documento.

Referencias

[1] Solicitud de patente. "Apparatus and Method for Estimating an Inter-Channel Time Difference." Solicitud internacional Número PCT/EP2017/051214.

[2] Knapp, Charles y Glifford Carter. "The generalized correlation method for estimation of time delay." IEEE Transactions on Acoustics, Speech, and Signal Processing 24.4 (1976): 320-327.

[3] Zhang, Cha, Dinei Florêncio y Zhengyou Zhang. "Why does PHAT work well in lownoise, reverberative environments?" Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on. IEEE, 2008.

[4] Rabinkin, Daniel V., *et al.* "DSP implementation of source location using microphone arrays." Advanced signal processing algorithms, architectures, and implementations VI. Vol. 2846. International Society for Optics and Photonics, 1996.

[5] Shen, Miao y Hong Liu. "A modified cross power-spectrum phase method based on microphone array for acoustic source localization." Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference on. IEEE, 2009.

[6] 3GPP TS 26.445; Codec for Enhanced Voice Services (EVS); Detailed algorithmic description.

REIVINDICACIONES

1. Aparato para estimar una diferencia de tiempo entre canales entre la señal de un primer canal y la señal de un segundo canal de una señal estéreo o una señal multicanal que tiene más canales que el primer canal y el segundo canal, que comprende:

un analizador de señales (1037) para estimar una característica de señal (1038) de la señal del primer canal o la señal del segundo canal o ambas señales o una señal derivada de la señal del primer canal o la señal del segundo canal;

una calculadora (1020) para calcular un espectro de correlación cruzada correspondiente a 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;

un ponderador (1036) para ponderar un espectro de correlación cruzada alisado o no alisado para obtener un espectro de correlación cruzada ponderado utilizando un primer procedimiento de ponderación (1036a) o utilizando un segundo procedimiento de ponderación (1036b) dependiendo de una característica de la señal estimada por el analizador de señales (1037), en el que el primer procedimiento de ponderación es diferente del segundo procedimiento de ponderación, en el que el primer procedimiento de ponderación (1036a) comprende una ponderación de modo que se normalice una amplitud y se mantenga una fase, o en el que el segundo procedimiento de ponderación (1036b) comprende un factor de ponderación derivado del espectro de correlación cruzada alisado o no alisado utilizando una operación de potenciación con una potencia que es inferior a 1 o superior a 0 o una función logarítmica; y

un procesador (1040) para procesar el espectro de correlación cruzada ponderado para obtener la diferencia de tiempo entre canales.

2. Aparato según la reivindicación 1, en el que el analizador de señales (1037) está configurado como un estimador de ruido (1037) para estimar un nivel de ruido (1038) de la señal del primer canal o la señal del segundo canal o ambas señales o una señal derivada de la señal del primer canal o la señal del segundo canal, y en el que una primera característica de la señal es un primer nivel de ruido y una segunda característica de la señal es un segundo nivel de ruido, o en el que el analizador de señales (1037) está configurado para ejecutar un análisis de voz/ música, un análisis de orador interferente, un análisis de música de fondo, un análisis de voz limpia o cualquier otro análisis de la señal a fin de determinar si una señal tiene una primera característica o una segunda característica.

3. Aparato según una de las reivindicaciones precedentes, en el que el segundo procedimiento de ponderación (1036b) comprende una ponderación de tal manera que se normalice una amplitud y se mantenga una fase y comprende además un factor de ponderación derivado del espectro de correlación cruzada alisado o no alisado utilizando una operación de potenciación con una potencia que es inferior a 1 o superior a 0 o entre 0,79 y 0,82.

4. Aparato según una de las reivindicaciones 2 0 3,

en el que el estimador de ruido (1037) está configurado para estimar (1060) un nivel de un ruido de fondo o está configurado para alisar (1061) un nivel de ruido estimado en el tiempo o está configurado para usar un filtro de alisado de IIR, o

en el que el estimador de ruido (1037) comprende asimismo un detector de actividad de señal (1070) para clasificar el bloque de tiempo como activo o inactivo, en el que el estimador de ruido (1037) está configurado para computar (1072) el nivel de una señal utilizando uno o más bloques de tiempo activos, o en el que el estimador de ruido (1037) está configurado para señalizar (1050) un alto nivel de ruido de fondo, cuando una relación señal a ruido es inferior a un umbral, umbral que está en el rango de 45 a 25 dB.

5. Aparato según una de las reivindicaciones precedentes,

en el que el procesador (1040) está configurado para ejecutar la determinación de la diferencia de tiempo entre canales mediante la ejecución de una búsqueda de picos u operación de identificación de picos (1041, 1042) dentro de una representación en el dominio del tiempo determinada a partir del espectro de correlación cruzada alisado.

6. Aparato según una de las reivindicaciones precedentes, que comprende además:

un estimador de características espectrales (1010) para estimar una característica de un espectro de la señal del primer canal o la señal del segundo canal para el boque de tiempo; y

un filtro de alisado (1030) para alisar el espectro de correlación cruzada en el tiempo utilizando la característica espectral para obtener un espectro de correlación cruzada alisado, y en el que el ponderador (1036) está configurado para ponderar el espectro de correlación cruzada alisado,

en el que el estimador de características espectrales (1010) está configurado para determinar, como la característica espectral, un nivel de ruido o una tonalidad del espectro; y

en el que el filtro de alisado (1030) está configurado para aplicar un alisado más fuerte en el tiempo con un primer grado de alisado en el caso de una primera característica menos ruidosa o una primera característica más tonal, o para aplicar un alisado más tenue en el tiempo con un segundo grado de alisado en el caso de una segunda característica más ruidosa o una segunda característica menos tonal,

en el que el primer grado de alisado es mayor que el segundo grado de alisado, 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, o

en el que el estimador de características espectrales (1010) está configurado para calcular, como la característica, una primera medida de la planitud espectral de un espectro de la señal del primer canal y una segunda medida de la planitud espectral de un segundo espectro de la señal del segundo canal, y para determinar la característica del espectro de la primera y la segunda medida de planitud espectral mediante la selección de un valor máximo, mediante la determinación de un promedio ponderado o un promedio no ponderado entre las medidas de la planitud espectral, o mediante la selección de un valor mínimo, o

en el que el filtro de alisado (1030) está configurado para calcular un valor del espectro de correlación cruzada alisado correspondiente a una frecuencia mediante una combinación ponderada del valor del espectro de correlación cruzada correspondiente a la frecuencia del bloque de tiempo y un valor de correlación cruzada espectral correspondiente a la frecuencia de por lo menos un bloque de tiempo pasado, en el que se determinan los factores de ponderación correspondientes a la combinación ponderada por la característica del espectro.

7. Aparato según una de las reivindicaciones precedentes,

en el que el procesador (1040) está configurado para determinar (1120) un rango válido y un rango inválido dentro de una representación en el dominio del tiempo derivado del espectro de correlación cruzada ponderado alisado o no alisado,

en el que se detecta por lo menos un pico máximo dentro del rango inválido (1121) y se compara (1123) con un pico máximo dentro del rango válido, donde solo se determina la diferencia de tiempo entre canales (1124), cuando el pico máximo dentro del rango válido es mayor que por lo menos un pico máximo dentro del rango inválido.

8. Aparato según una de las reivindicaciones precedentes,

en el que el procesador (1040) está configurado

para ejecutar (1102) una operación de búsqueda de picos dentro de una representación en el dominio del tiempo derivada del espectro de correlación cruzada alisado,

para determinar (1105) una variable de un umbral fijo de la representación en el dominio del tiempo: y

para comparar (1106, 1035) un pico con el umbral variable, en el que la diferencia de tiempo entre canales se determina en forma de retardo de tiempo asociado a un pico que está en una relación predeterminada con el umbral variable.

9. Aparato según la reivindicación 8,

en el que el procesador (1040) está configurado para determinar (1105) el umbral variable como valor que es igual a un múltiplo entero de un valor entre una porción más grande tal como del 10 % de los valores de la representación en el dominio del tiempo.

10. Aparato según una de las reivindicaciones precedentes,

en el que el procesador (1040) está configurado para determinar (1102) una amplitud de un pico máximo en cada subbloque de una pluralidad de subbloques de una representación en el dominio del tiempo derivado del espectro de correlación cruzada alisado,

en el que el procesador (1040) está configurado para calcular (1105, 1034) un umbral variable sobre la base de una magnitud de pico media derivada (1103) de las magnitudes del pico máximo de la pluralidad de subbloques; y

en el que el procesador (1140) está configurado para determinar la diferencia de tiempo entre canales en términos de un valor de retardo de tiempo correspondiente a un pico máximo de la pluralidad de subbloques que es mayor que el umbral variable.

11. Aparato según la reivindicación 10,

en el que el procesador (1140) está configurado para calcular (1105) el umbral variable mediante una multiplicación del umbral medio determinado en forma de pico promedio entre los picos de los subbloques y un valor,

en el que el valor está determinado por una característica de SNR (relación señal a ruido) de las señales de los canales primero y segundo, en el que un primer valor está asociado a un primer valor de SNR y un segundo valor está asociado a un segundo valor de SNR, en el que el primer valor es mayor que el segundo valor, y en el que el primer valor de SNR es mayor que el segundo valor de SNR.

12. Aparato según una de las reivindicaciones precedentes, en el que el aparato está configurado

para ejecutar un almacenamiento o una transmisión de la diferencia de tiempo entre canales estimada, o

para ejecutar un procesamiento estéreo o multicanal o la codificación de las señales del primer y segundo canal utilizando la diferencia de tiempo entre canales estimada, o

para ejecutar un alineamiento en tiempo de las señales de los dos canales utilizando la diferencia de tiempo entre canales, o

para ejecutar una estimación de diferencia de tiempo de llegada utilizando la diferencia de tiempo entre canales estimada, o

para ejecutar una estimación de diferencia de tiempo de llegada utilizando la diferencia de tiempo entre canales para la determinación de la posición de un altavoz en un recinto con dos micrófonos y una configuración conocida de micrófono, o

para ejecutar una formación de haces utilizando la diferencia de tiempo entre canales estimada, o

para ejecutar un filtrado especial utilizando la diferencia de tiempo entre canales, o

para ejecutar una descomposición del primer plano o de fondo utilizando una diferencia de tiempo entre canales estimada, o

para ejecutar una operación de localización de una fuente de sonido utilizando la diferencia de tiempo entre canales estimada, o

para ejecutar una localización de una fuente de sonido utilizando la diferencia de tiempo entre canales estimada mediante la ejecución de una triangulación acústica sobre la base de las diferencias de tiempo entre la señal del primer canal y la señal del segundo canal o la señal del primer canal, la señal del segundo canal y por lo menos una señal adicional.

13. Aparato según la reivindicación 1, en el que el analizador de señales (1037) está configurado para determinar un nivel de ruido como característica de señal (1038), y en el que el ponderador (1036) está configurado para seleccionar o bien el primer procedimiento de ponderación o bien el segundo dependiendo del nivel de ruido.

14. Método para estimar una diferencia de tiempo entre canales entre la señal de un primer canal y la señal de un segundo canal de una señal estéreo o una señal multicanal que tiene más señales de canal que la señal

del primer canal y la señal del segundo canal, comprendiendo el método:

estimar una característica de señal de la señal del primer canal o la señal del segundo canal o ambas señales o una señal derivada de la señal del primer canal o la señal del segundo canal;

calcular un espectro de correlación cruzada correspondiente a un bloque de tiempo a partir de la señal del primer canal en el bloque de tiempo y la señal del segundo canal en el bloque de tiempo;

ponderar un espectro de correlación cruzada alisado o no alisado para obtener un espectro de correlación cruzada ponderado utilizando un primer procedimiento de ponderación o utilizando un segundo procedimiento de ponderación dependiendo de una característica de señal estimada, en el que el primer procedimiento de ponderación es diferente del segundo procedimiento de ponderación, en el que el primer procedimiento de ponderación (1036a) comprende una ponderación de modo que se normalice una amplitud y se mantenga una fase, o en el que el segundo procedimiento de ponderación (1036b) comprende un factor de ponderación derivado del espectro de correlación cruzada alisado o no alisado utilizando una operación de potenciación con una potencia que es inferior a 1 o superior a 0 o una función logarítmica; y

procesar el espectro de correlación cruzada ponderado para obtener la diferencia de tiempo entre canales.

15. Producto de programa informático que comprende instrucciones que, cuando las ejecuta un ordenador, hacen que el ordenador lleve a cabo el método según la reivindicación 14.

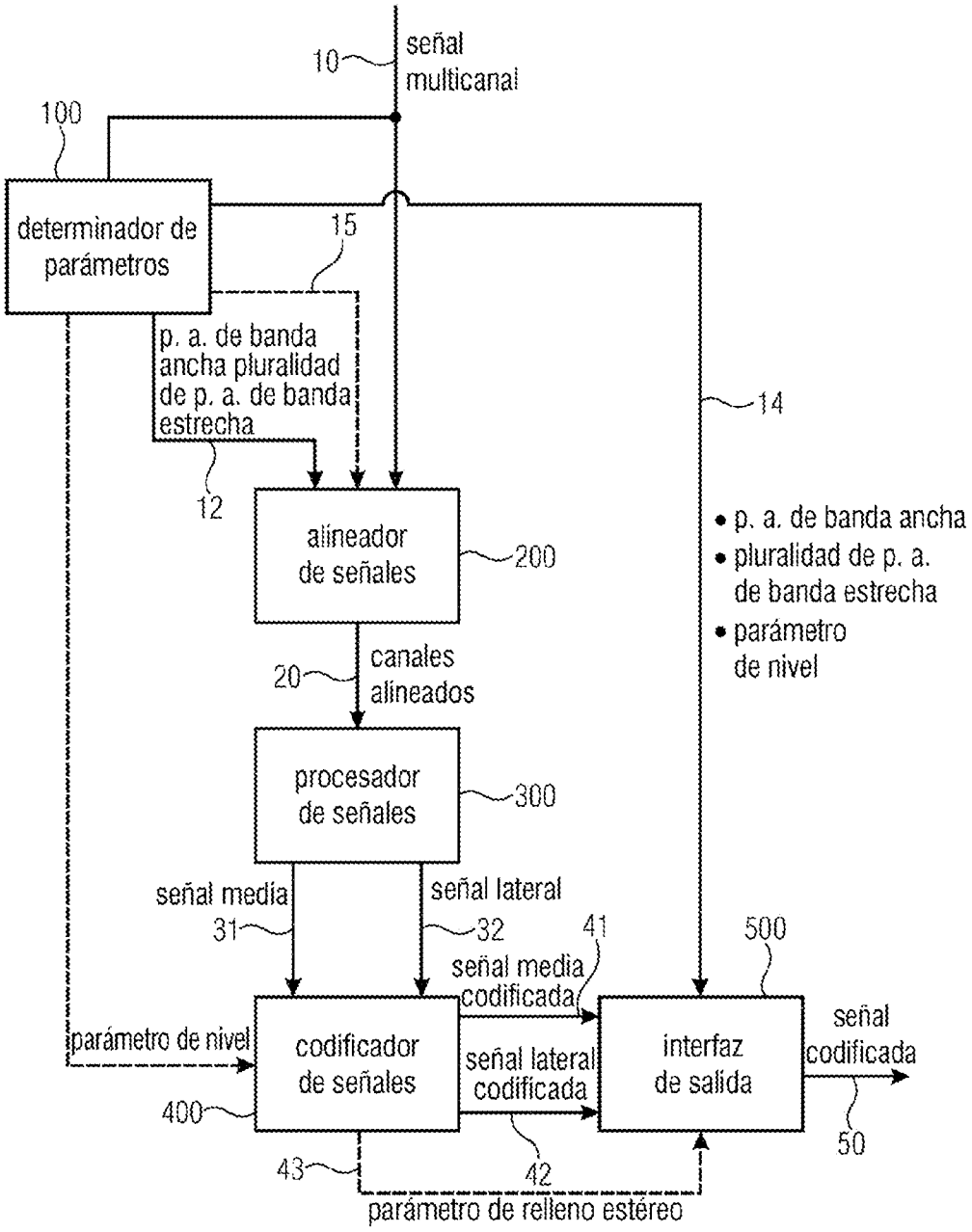

10
señal multicanal
100
determinador de parámetros
15
p. a. de banda ancha pluralidad de p. a. de banda estrecha
12
14
alineador de señales
200
• p. a. de banda ancha
• pluralidad de p. a. de banda estrecha
• parámetro de nivel
20
canales alineados
procesador de señales
300
señal media
31
señal lateral
32
41
señal media codificada
500
parámetro de nivel
codificador de señales
señal lateral codificada
interfaz de salida
señal codificada
400
50
43
42
parámetro de relleno estéreo

Fig. 1

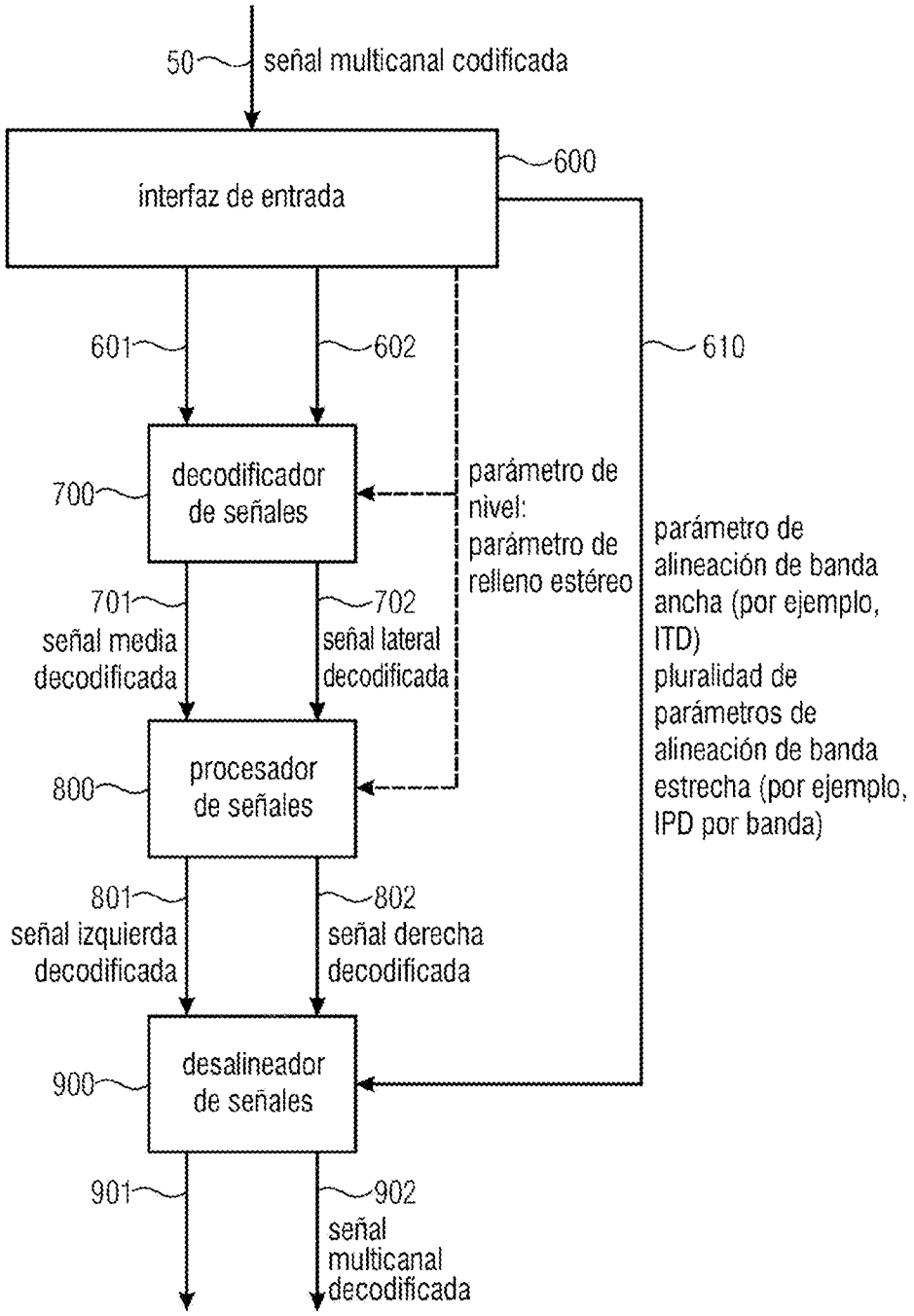
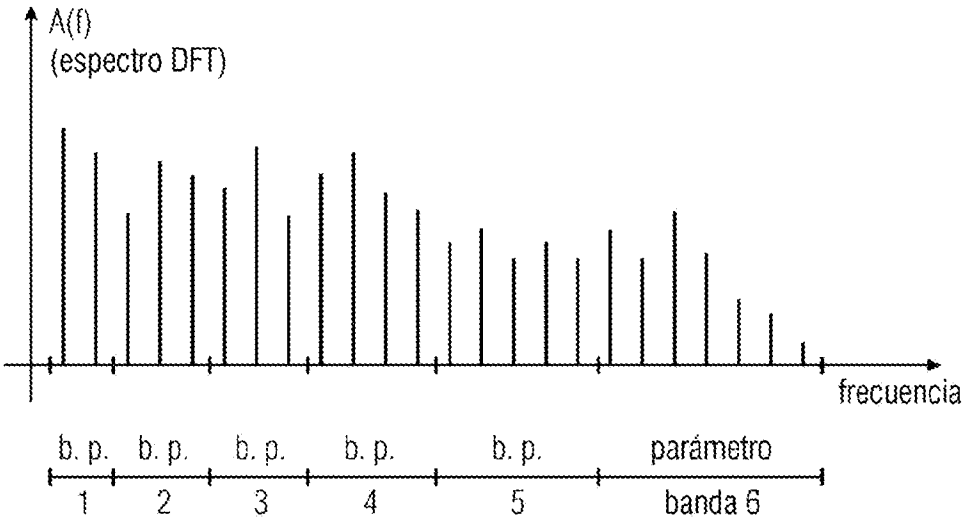

50
señal multicanal codificada
interfaz de entrada
600
601
602
610
decodificador de señales
700
parámetro de nivel: parámetro de relleno estéreo
parámetro de alineación de banda ancha (por ejemplo, ITD)
pluralidad de parámetros de alineación de banda estrecha (por ejemplo, IPD por banda)
701
señal media decodificada
702
señal lateral decodificada
procesador de señales
800
801
señal izquierda decodificada
802
señal derecha decodificada
desalineador de señales
900
901
902
señal multicanal decodificada

Fig. 2



- único parámetro de alineación de banda ancha para todo el espectro (por ejemplo, banda de p. 1 a banda de p. 6);
- pluralidad de parámetros de alineación de banda estrecha para 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 bandas de parámetros 4, 5, 6, por ejemplo, tres parámetros de relleno estéreo;
- señal lateral (residual) para bandas de parámetros 1, 2, 3;
- más líneas espectrales en la banda superior, por ejemplo, siete líneas espectrales en la banda de parámetros 5 frente a tres líneas espectrales en la banda de parámetros 2.

Fig. 3

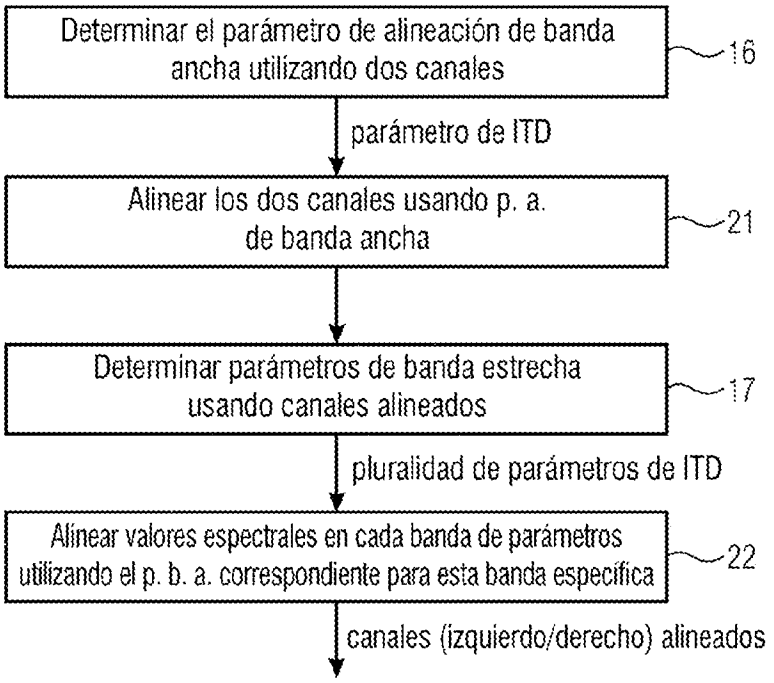

Determinar el parámetro de alineación de banda ancha utilizando dos canales
16
parámetro de ITD
Alinear los dos canales usando p. a. de banda ancha
21
Determinar parámetros de banda estrecha usando canales alineados
17
pluralidad de parámetros de ITD
Alinear valores espectrales en cada banda de parámetros utilizando el p. b. a. correspondiente para esta banda específica
22
canales (izquierdo/derecho) alineados

Fig. 4a

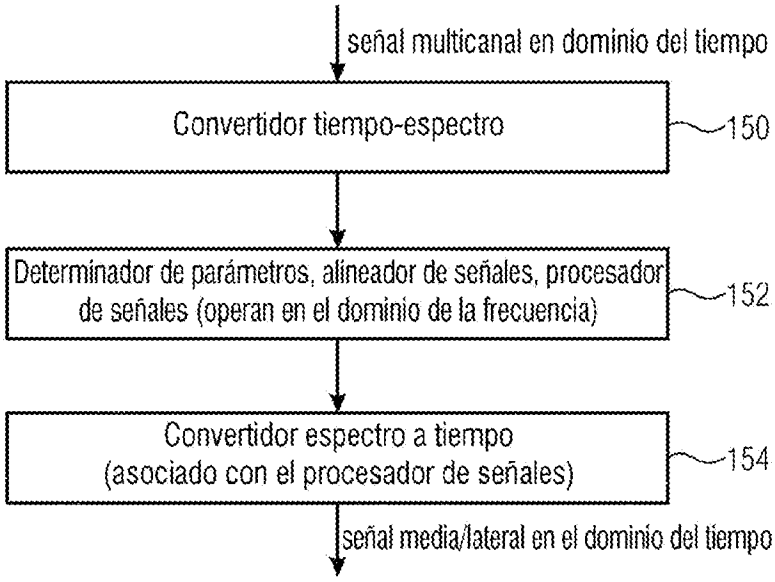

señal multicanal en dominio del tiempo
Convertidor tiempo-espectro
150
Determinador de parámetros, alineador de señales, procesador de señales (operan en el dominio de la frecuencia)
152
Convertidor espectro a tiempo (asociado con el procesador de señales)
154
señal media/lateral en el dominio del tiempo

Fig. 4b

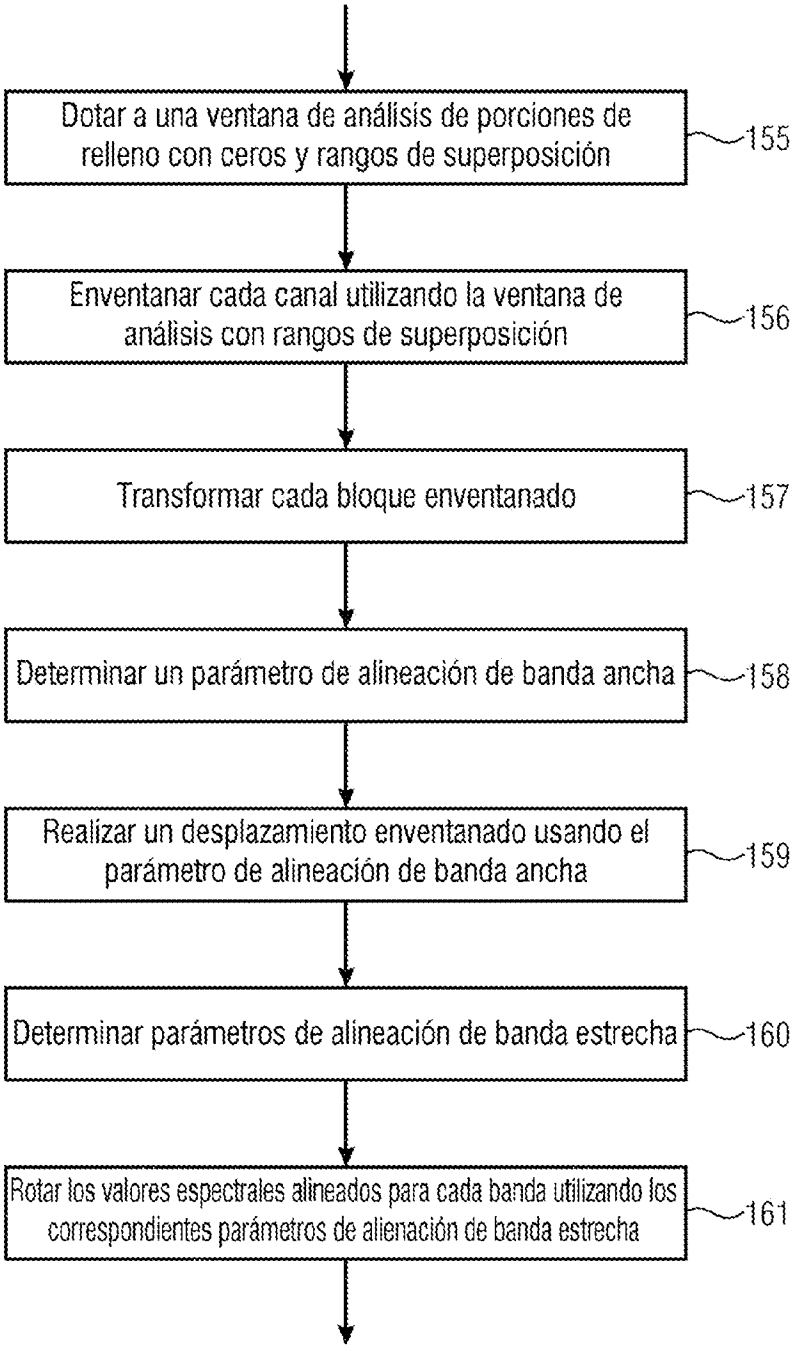

Dotar a una ventana de análisis de porciones de relleno con ceros y rangos de superposición
155
Enventanar cada canal utilizando la ventana de análisis con rangos de superposición
156
Transformar cada bloque enventanado
157
Determinar un parámetro de alineación de banda ancha
158
Realizar un desplazamiento enventanado usando el parámetro de alineación de banda ancha
159
Determinar parámetros de alineación de banda estrecha
160
Rotar los valores espectrales alineados para cada banda utilizando los correspondientes parámetros de alienación de banda estrecha
161

Fig. 4c

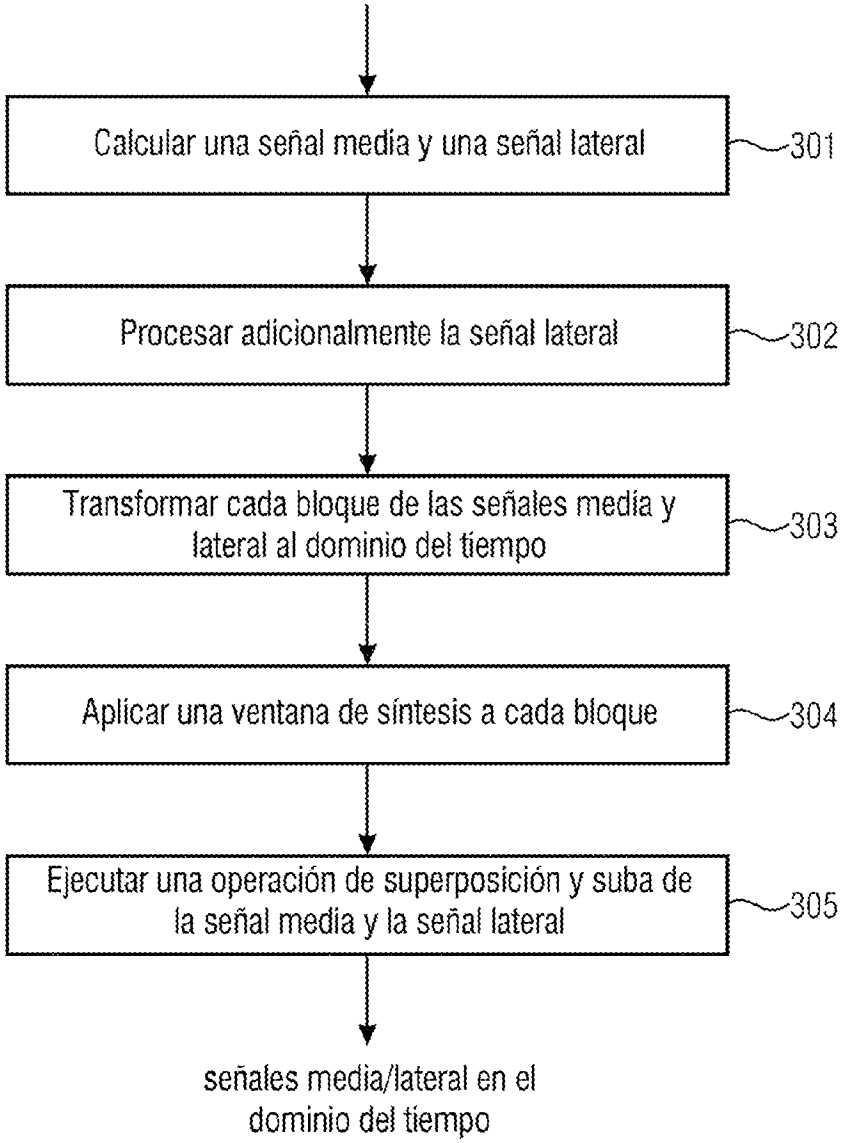
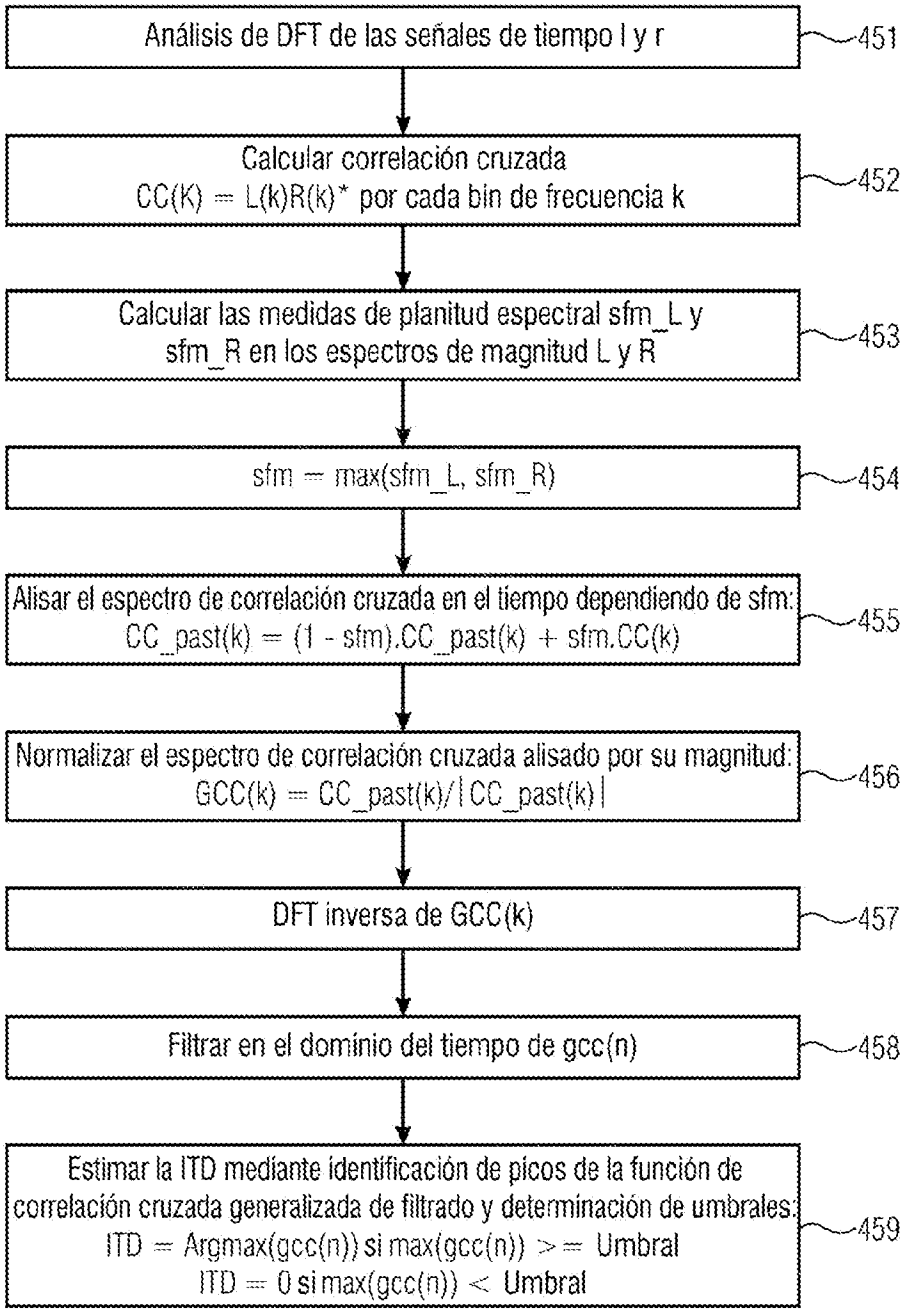

Calcular una señal media y una señal lateral
301
Procesar adicionalmente la señal lateral
302
Transformar cada bloque de las señales media y lateral al dominio del tiempo
303
Aplicar una ventana de síntesis a cada bloque
304
Ejecutar una operación de superposición y suba de la señal media y la señal lateral
305
señales media/lateral en el dominio del tiempo

Fig. 4d



estimación de ITD

Fig. 4e

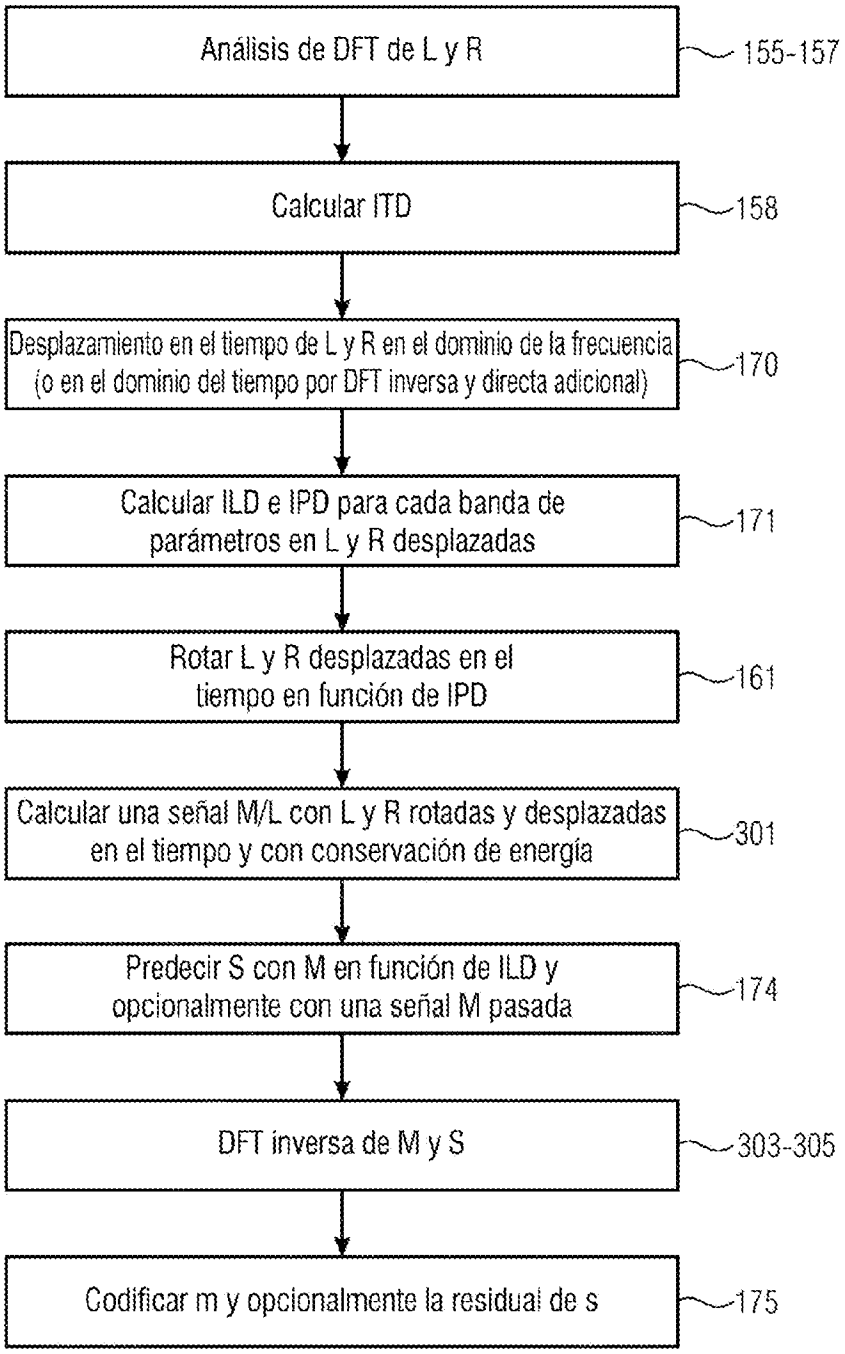

Análisis de DFT de L y R
155-157
Calcular ITD
158
Desplazamiento en el tiempo de L y R en el dominio de la frecuencia (o en el dominio del tiempo por DFT inversa y directa adicional)
170
Calcular ILD e IPD para cada banda de parámetros en L y R desplazadas
171
Rotar L y R desplazadas en el tiempo en función de IPD
161
Calcular una señal M/L con L y R rotadas y desplazadas en el tiempo y con conservación de energía
301
Predecir S con M en función de ILD y opcionalmente con una señal M pasada
174
DFT inversa de M y S
303-305
Codificar m y opcionalmente la residual de s
175

Fig. 5

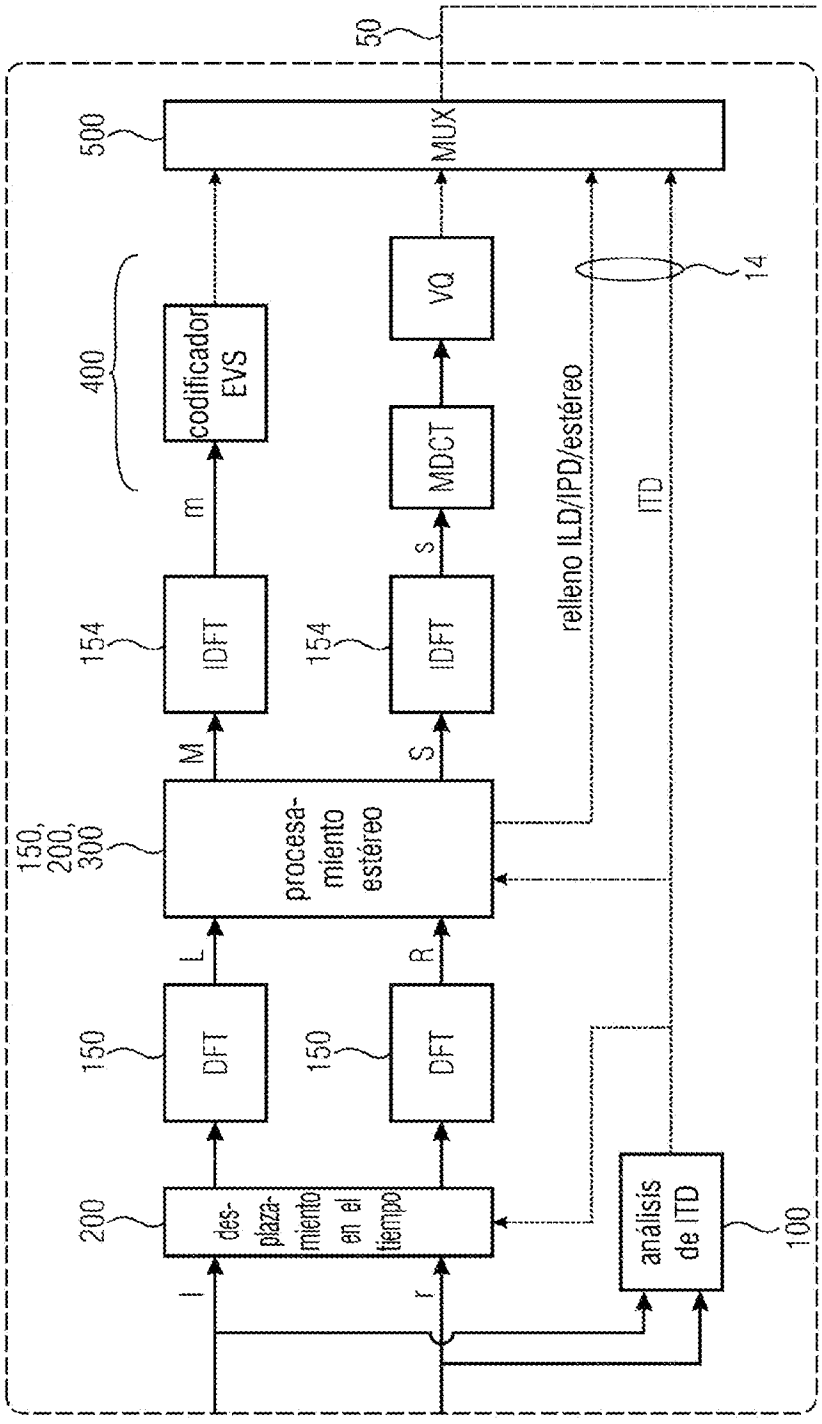

50
500
MUX
400
codificador EVS
m
154
IDFT
M
150, 200, 300
procesa-miento estéreo
L
150
DFT
200
des-plaza-miento en el tiempo
l
r
R
150
DFT
S
154
IDFT
s
MDCT
VQ
14
relleno ILD/IPD/estéreo
ITD
análisis de ITD
100

Fig. 6a

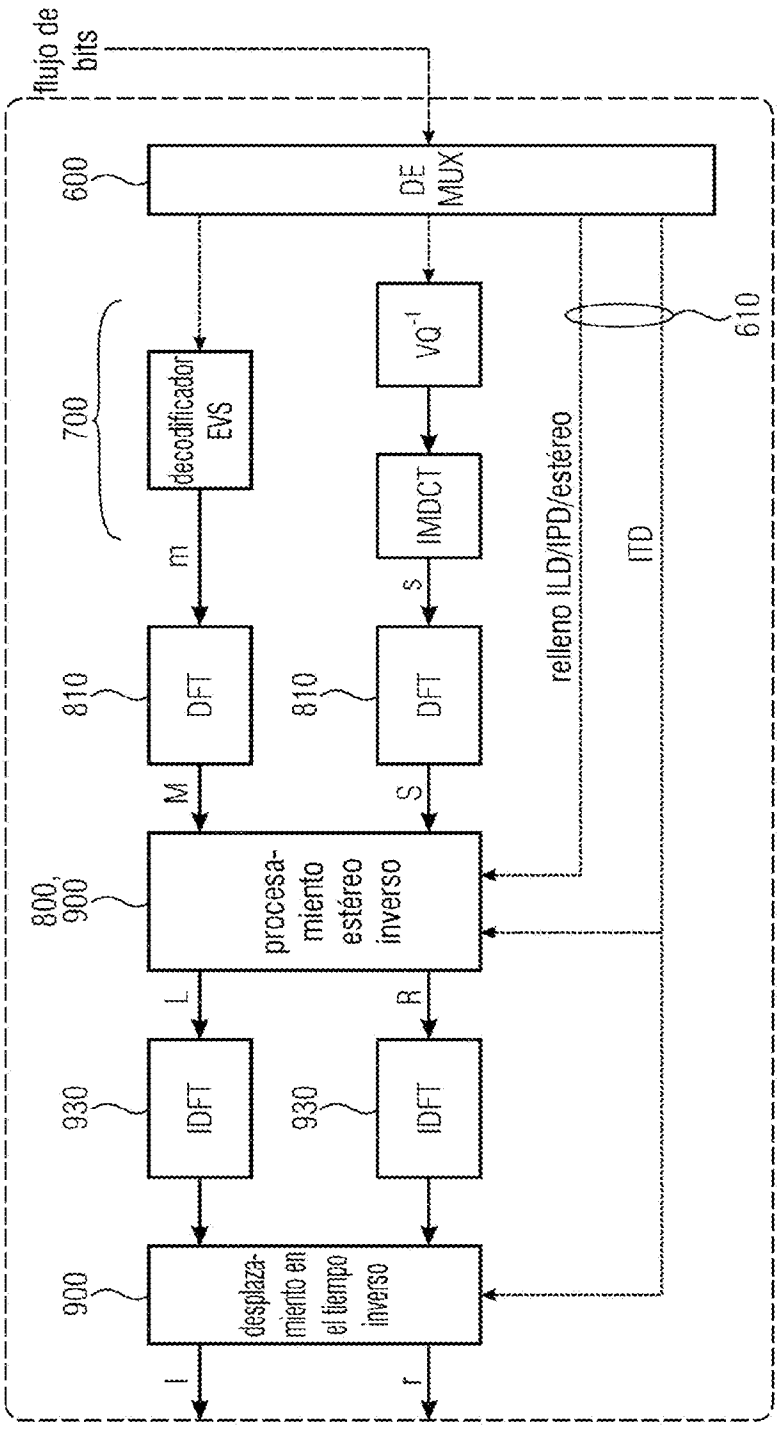

flujo de bits
600
DE MUX
610
700
decodificador EVS
VQ-1
IMDCT
m
s
relleno ILD/IPD/estéreo
ITD
810
DFT
810
DFT
M
S
800, 900
procesa-miento estéreo inverso
L
R
930
IDFT
930
IDFT
900
desplaza-miento en el tiempo inverso
l
r

Fig. 6b

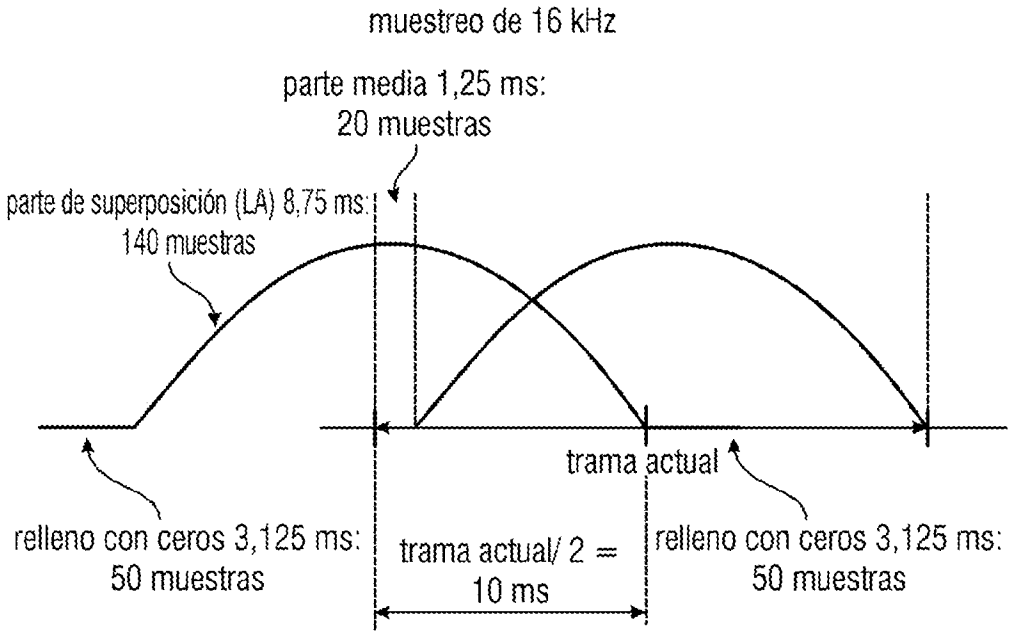


Fig. 7

parámetros	bits/banda	número de bandas	tasa de bits
ILD	5	las 12 bandas	3,00 kbps
IPD	3	hasta 2,5 kHz	1,05 kbps
ITD	8	todo el espectro	0,6 kbps
relleno estéreo	3	desde 1 kHz	0,9 kbps
total			~5 kbps

Fig. 8

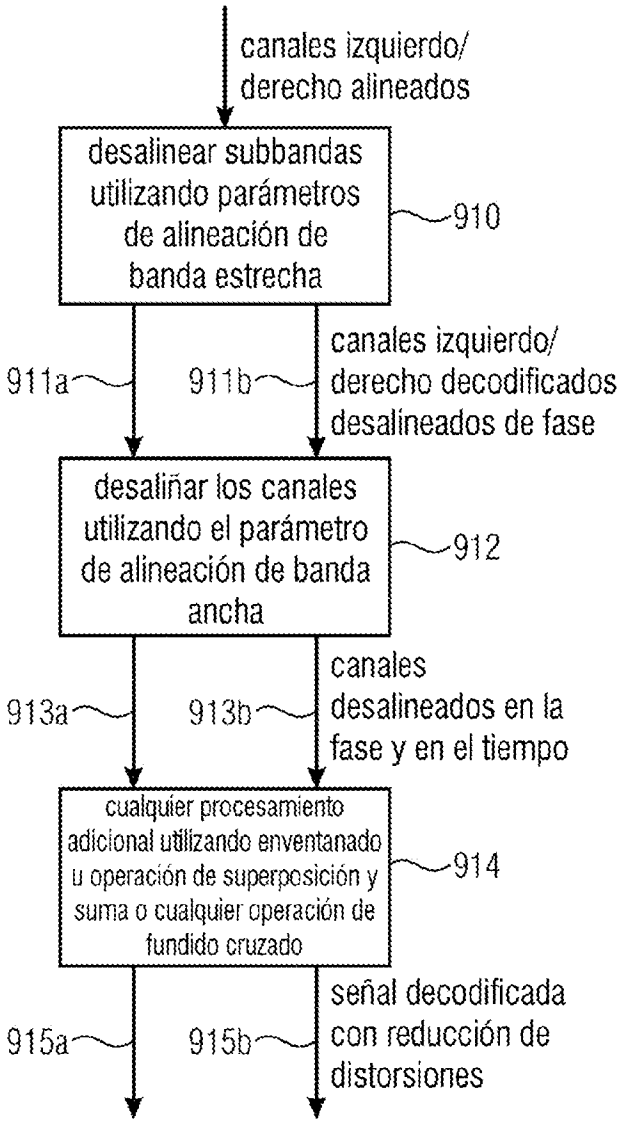

canales izquierdo/
derecho alineados
desalinear subbandas utilizando parámetros de alineación de banda estrecha
910
911a
911b
canales izquierdo/
derecho decodificados
desalineados de fase
desaliñar los canales utilizando el parámetro de alineación de banda ancha
912
913a
913b
canales
desalineados en la
fase y en el tiempo
cualquier procesamiento adicional utilizando enventanado u operación de superposición y suma o cualquier operación de fundido cruzado
914
915a
915b
señal decodificada
con reducción de
distorsiones

Fig. 9a

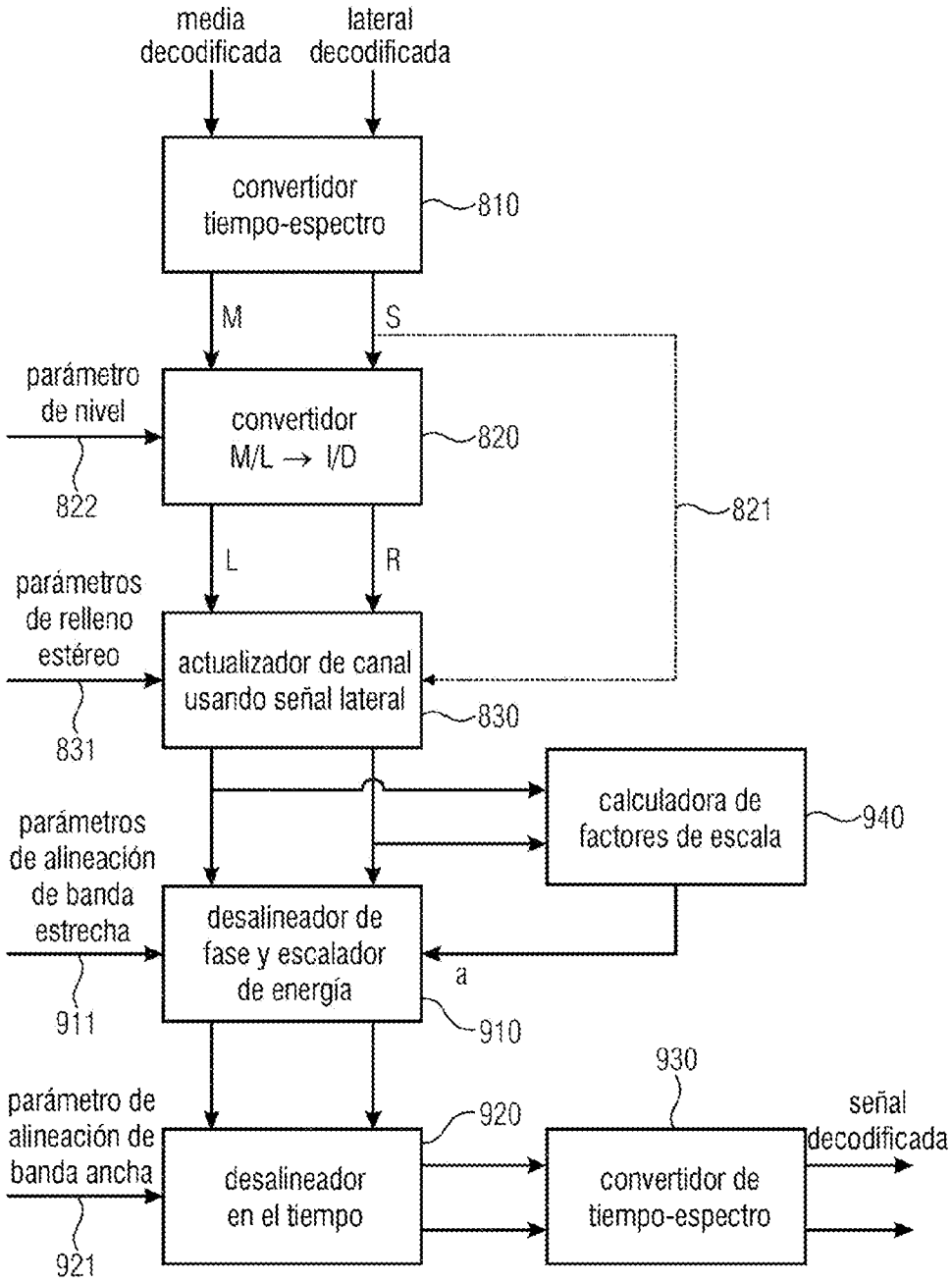

media decodificada
lateral decodificada
convertidor tiempo-espectro
810
M
S
parámetro de nivel
convertidor M/L → I/D
820
822
821
L
R
parámetros de relleno estéreo
actualizador de canal usando señal lateral
830
831
calculadora de factores de escala
940
parámetros de alineación de banda estrecha
desalineador de fase y escalador de energía
a
911
910
930
parámetro de alineación de banda ancha
920
señal decodificada
desalineador en el tiempo
convertidor de tiempo-espectro
921

Fig. 9b

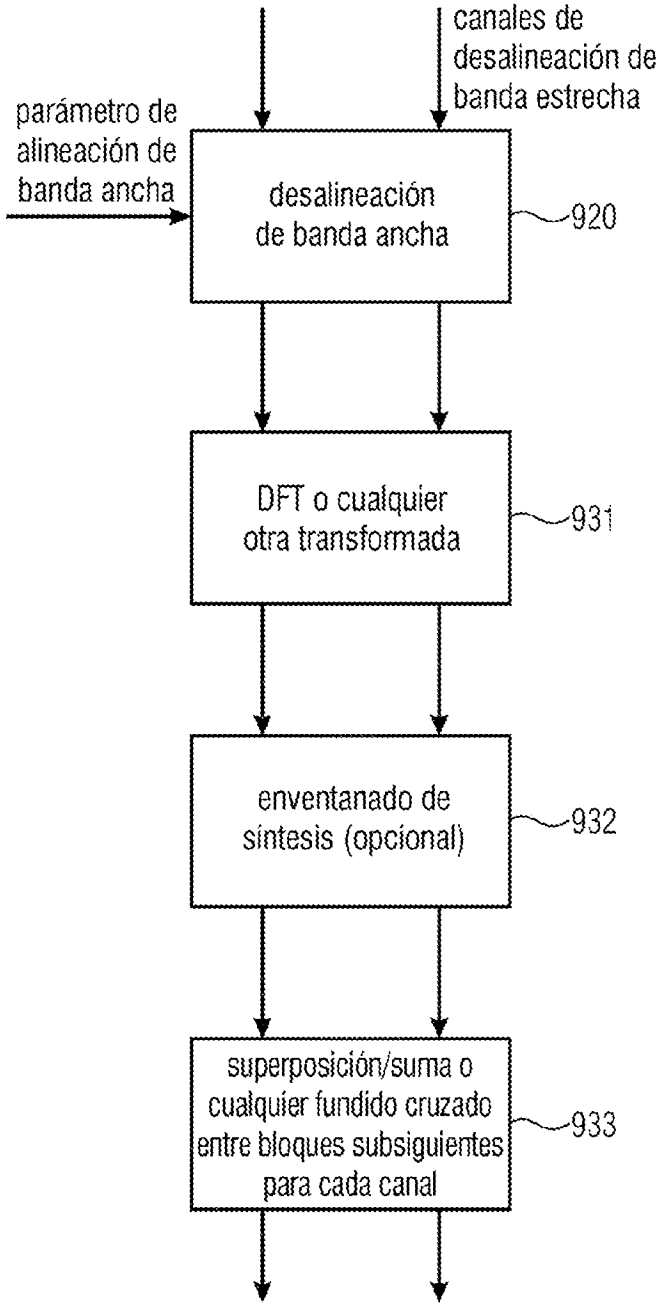

canales de desalineación de banda estrecha
parámetro de alineación de banda ancha
desalineación de banda ancha
920
DFT o cualquier otra transformada
931
enventanado de síntesis (opcional)
932
superposición/suma o cualquier fundido cruzado entre bloques subsiguientes para cada canal
933

Fig. 9c

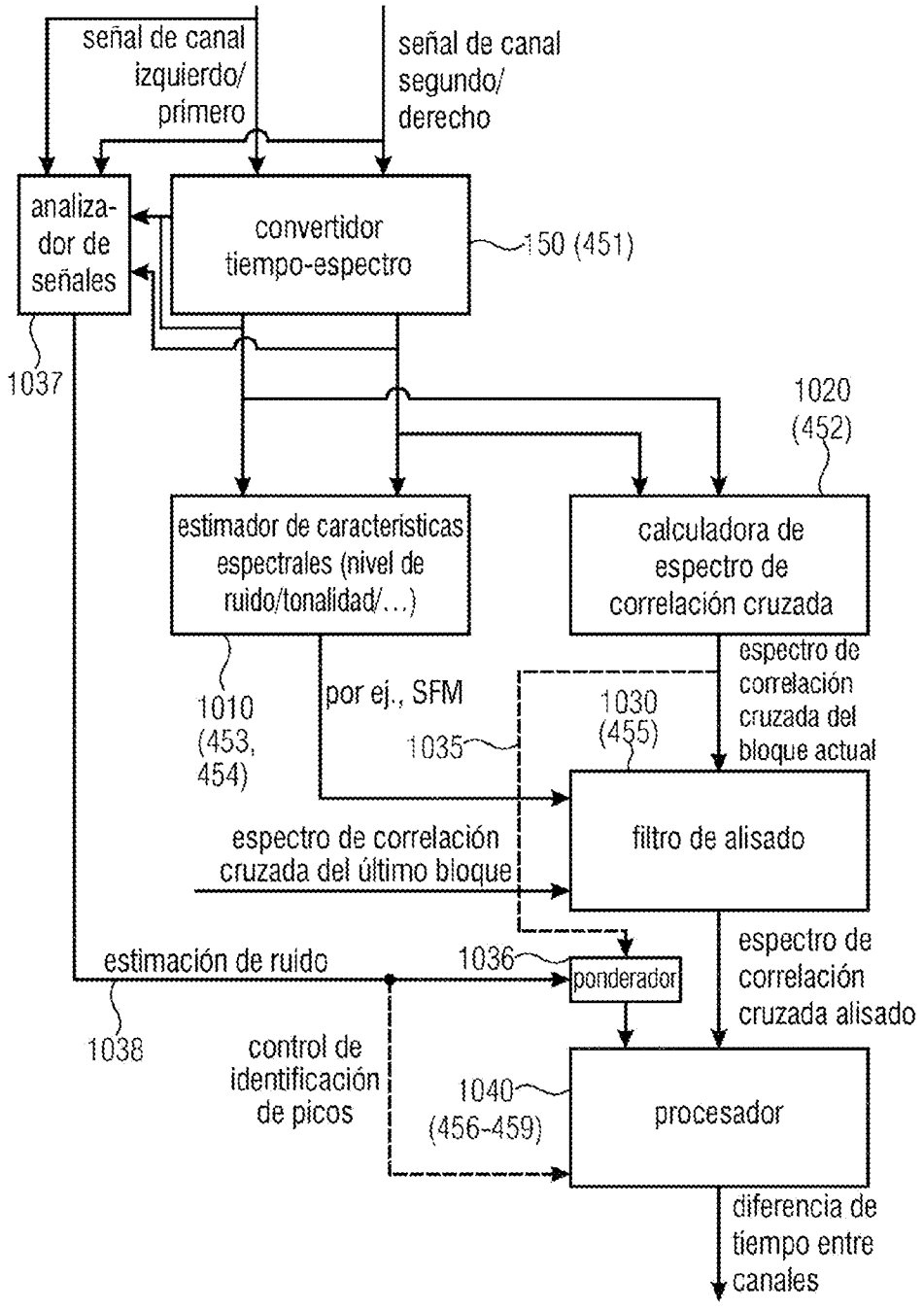
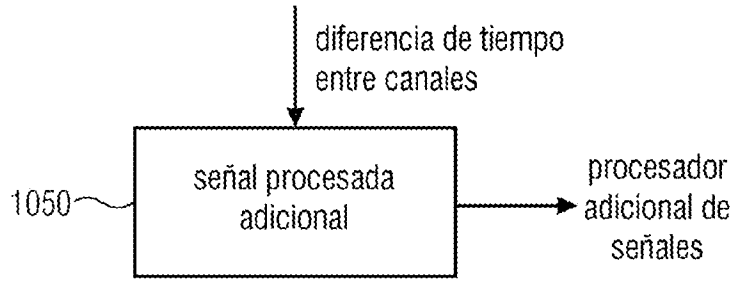

señal de canal izquierdo/ primero
señal de canal segundo/ derecho
analiza-dor de señales
convertidor tiempo-espectro
150 (451)
1037
1020 (452)
estimador de características espectrales (nivel de ruido/tonalidad/...)
calculadora de espectro de correlación cruzada
espectro de correlación cruzada del bloque actual
1010 (453, 454)
por ej., SFM
1035
1030 (455)
filtro de alisado
espectro de correlación cruzada del último bloque
estimación de ruido
1036
ponderador
espectro de correlación cruzada alisado
1038
control de identificación de picos
1040 (456-459)
procesador
diferencia de tiempo entre canales

Fig. 10a



- almacenamiento/transmisión de datos paramétricos
- procesamiento/codificación estéreo/multicanal
- alineación de tiempo de dos canales
- diferencia de tiempo de la estimación de llegada para la determinación de la posición del altavoz en un recinto con dos micrófonos y una configuración conocida de los micrófonos
- formación de haces
- filtrado espacial
- descomposición de primer plano/fondo
- ubicación de una fuente de sonido, por ejemplo mediante triangulación acústica basada en diferencias de tiempo de dos/tres señales

Fig. 10b

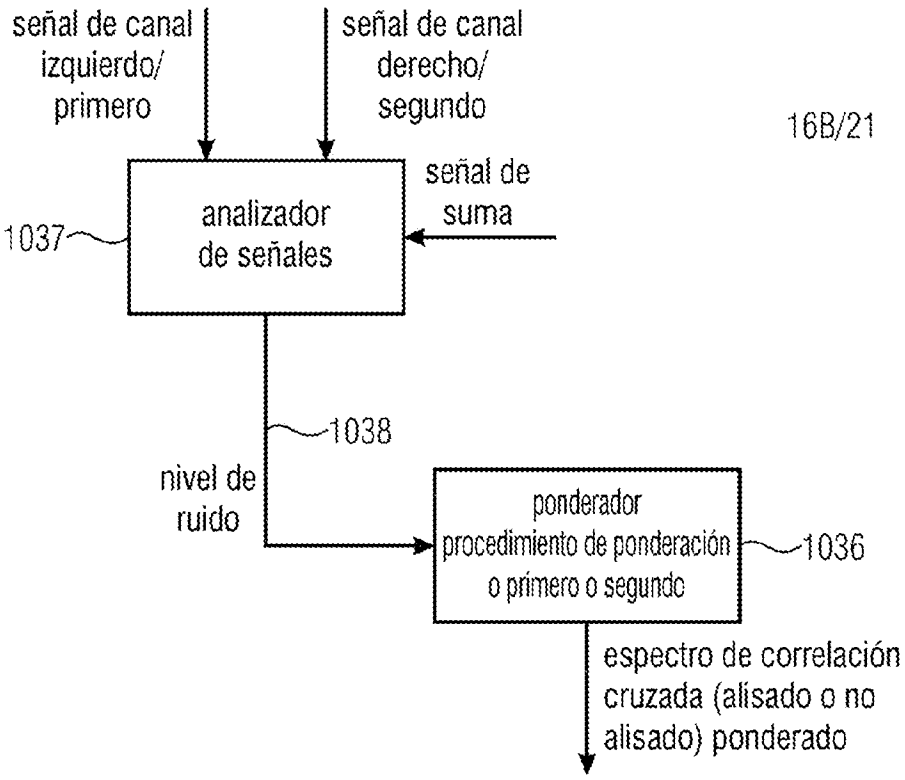

señal de canal izquierdo/ primero
señal de canal derecho/ segundo
16B/21
señal de suma
analizador de señales
1037
1038
nivel de ruido
ponderador procedimiento de ponderación o primero o segundo
1036
espectro de correlación cruzada (alisado o no alisado) ponderado

Fig. 10c

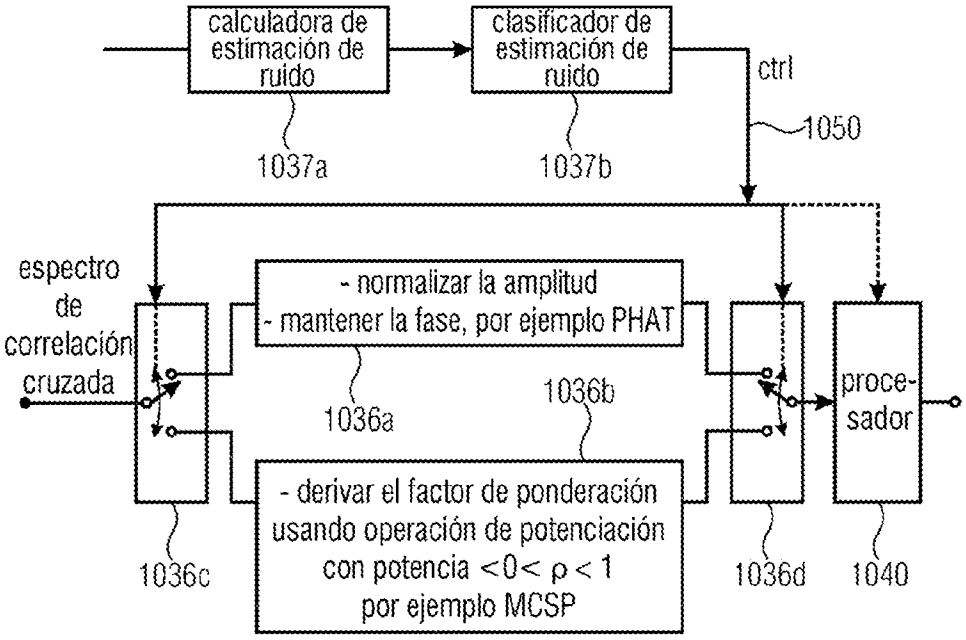
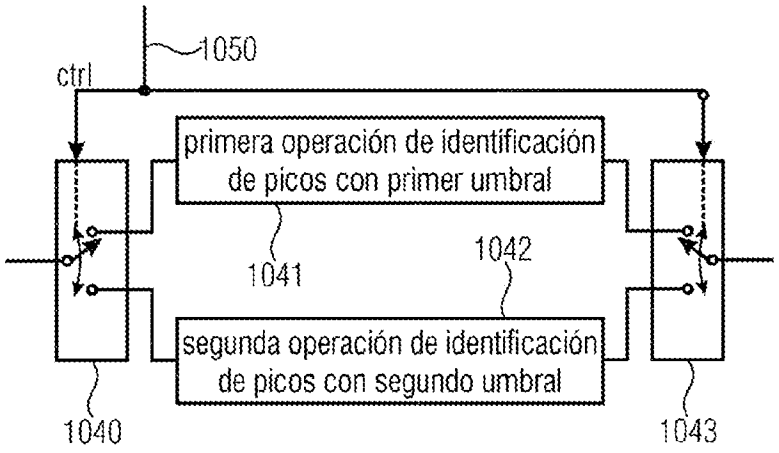


Fig. 10d



- segundo umbral inferior a primer umbral

Fig. 10e

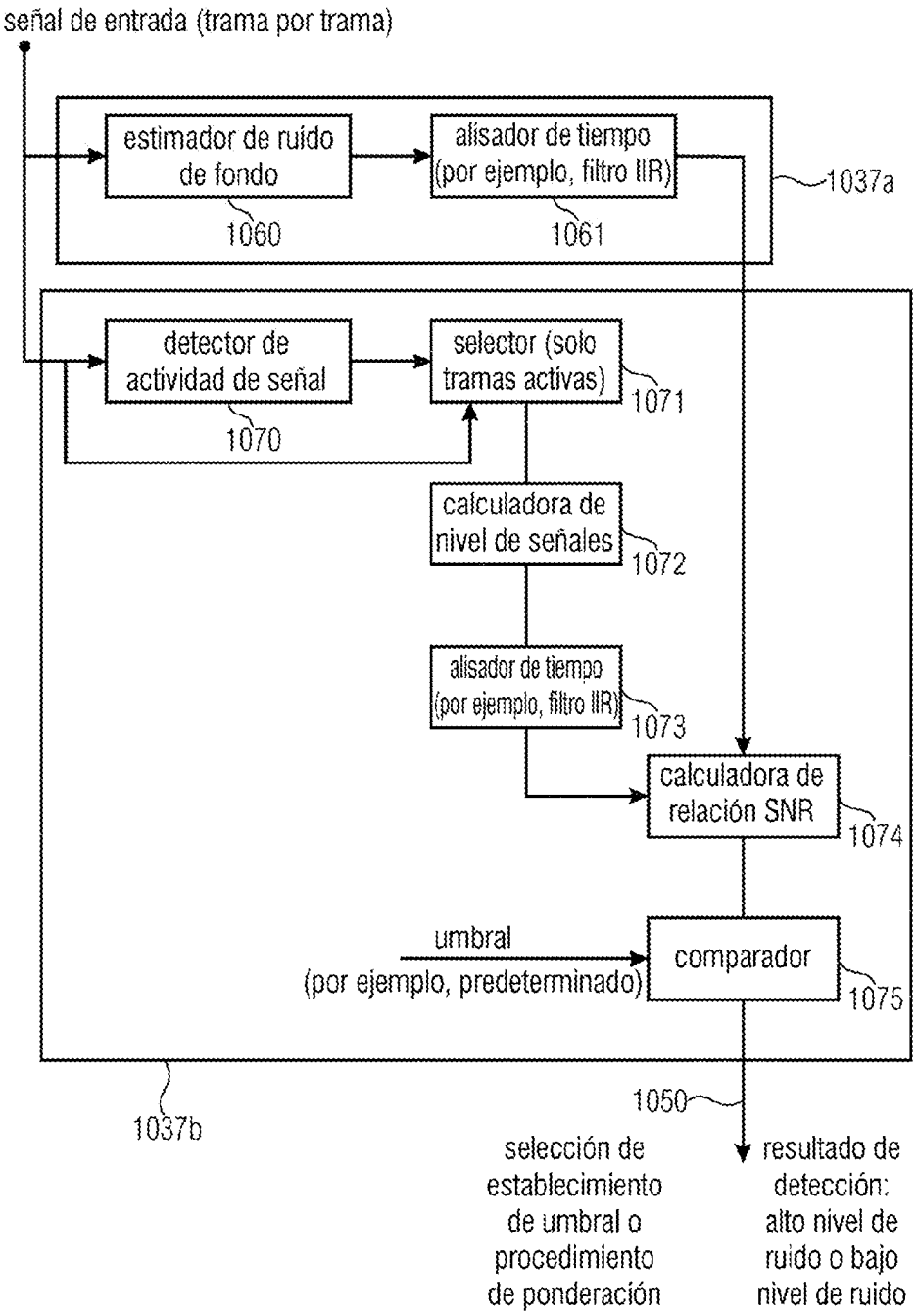

señal de entrada (trama por trama)
estimador de ruido de fondo
1060
alisador de tiempo (por ejemplo, filtro IIR)
1061
1037a
detector de actividad de señal
1070
selector (solo tramas activas)
1071
calculadora de nivel de señales
1072
alisador de tiempo (por ejemplo, filtro IIR)
1073
calculadora de relación SNR
1074
umbral (por ejemplo, predeterminado)
comparador
1075
1037b
1050
selección de establecimiento de umbral o procedimiento de ponderación
resultado de detección: alto nivel de ruido o bajo nivel de ruido

Fig. 10f

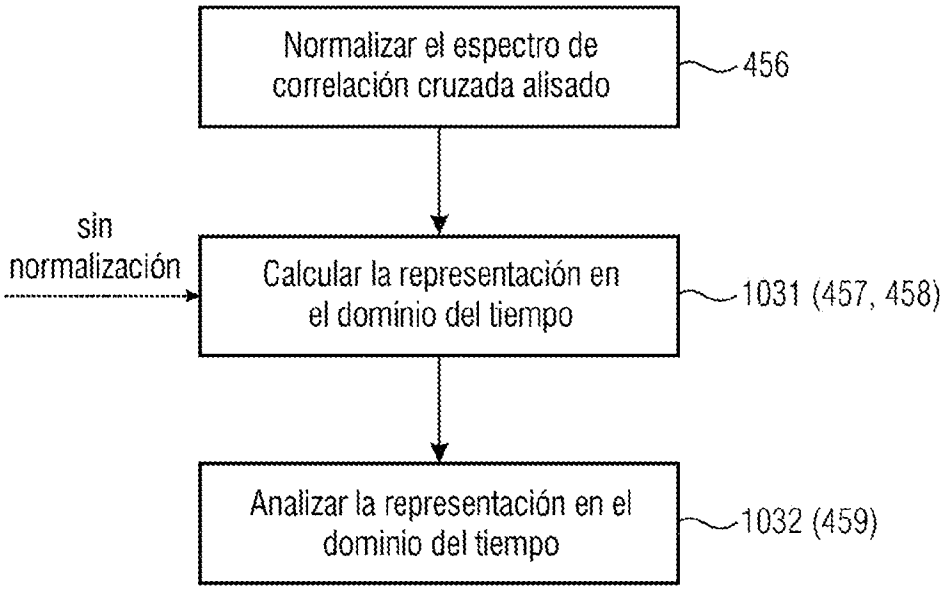


Fig. 11a

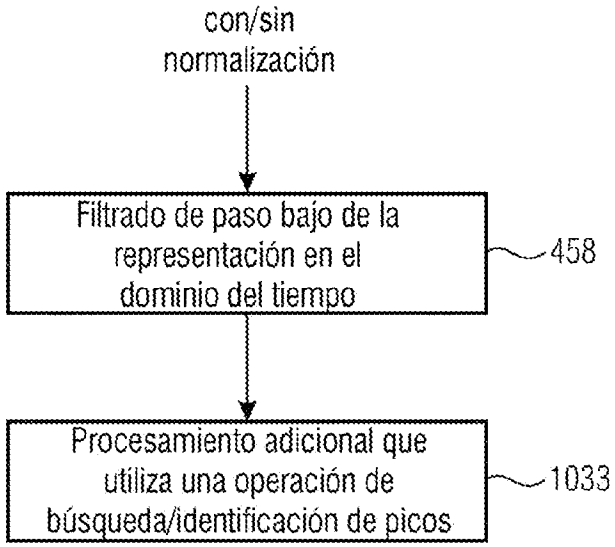


Fig. 11b

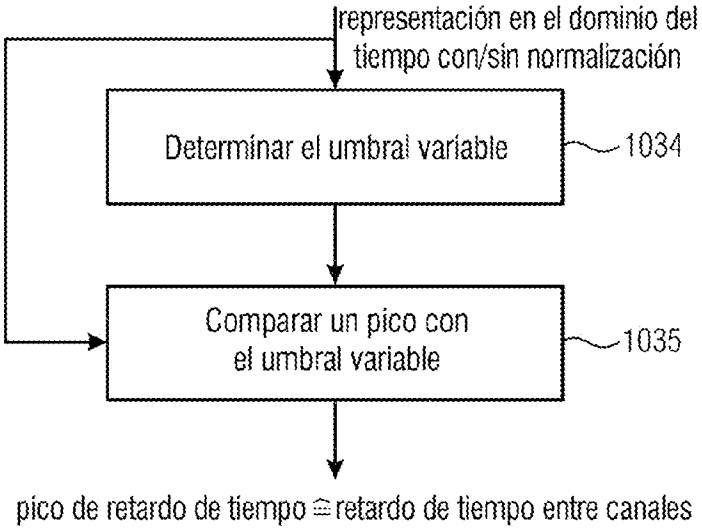


Fig. 11c

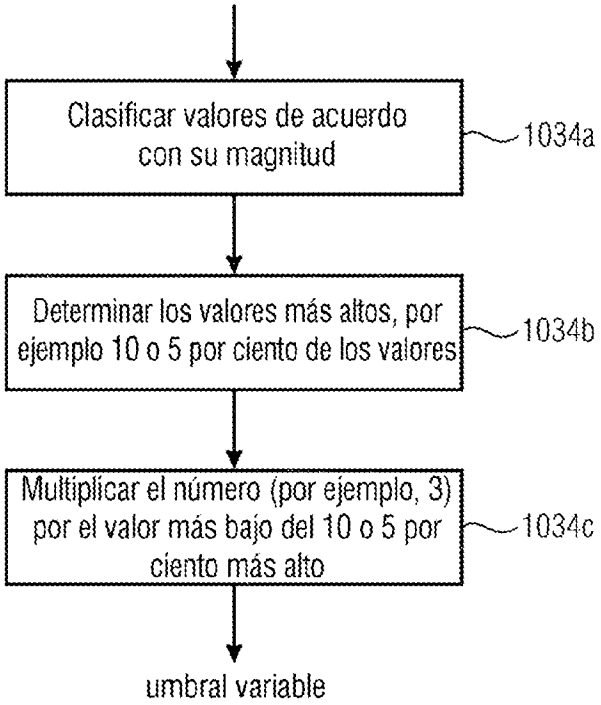


Fig. 11d

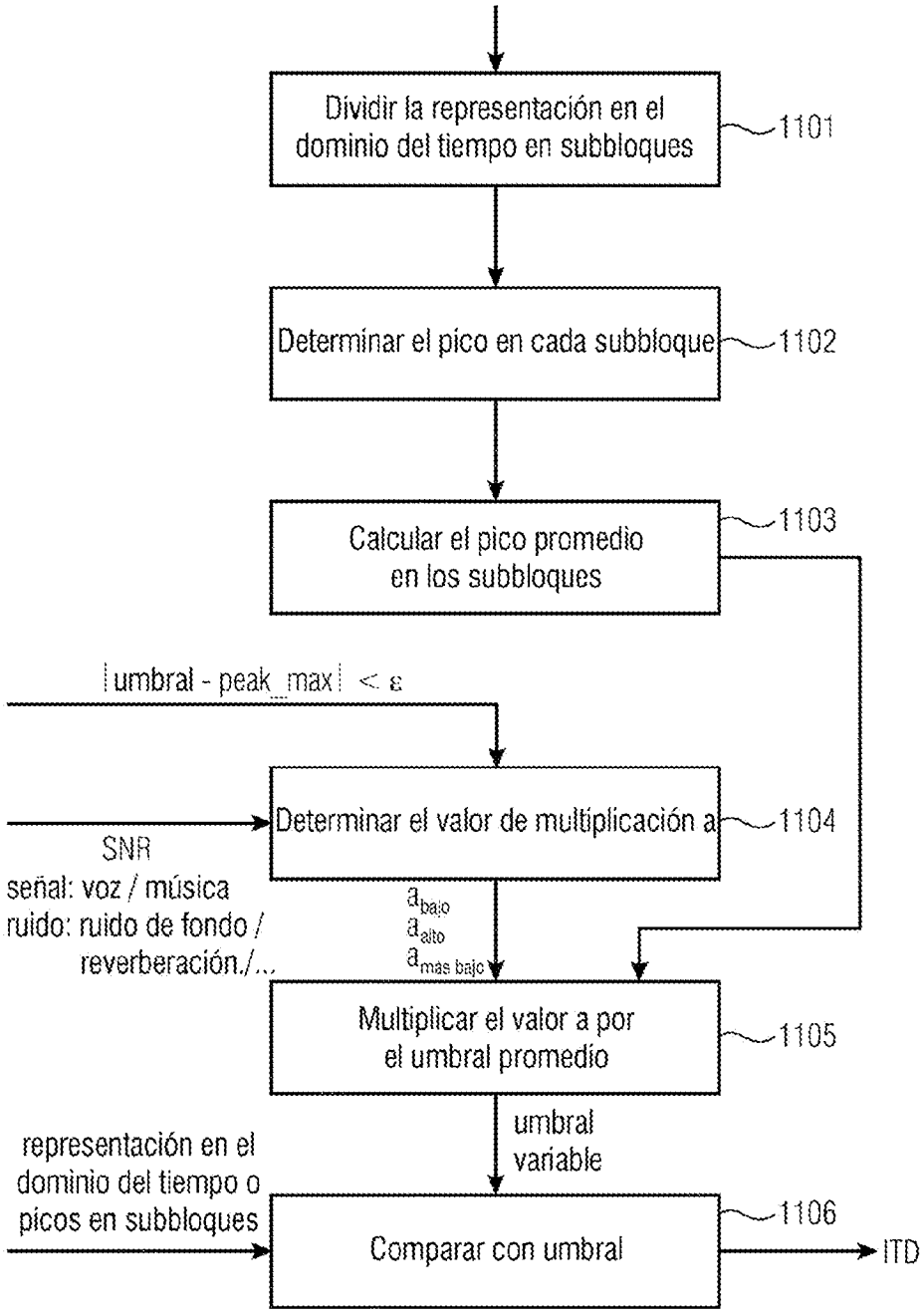

Dividir la representación en el dominio del tiempo en subbloques
1101
Determinar el pico en cada subbloque
1102
Calcular el pico promedio en los subbloques
1103
|umbral - peak_max| < ε
Determinar el valor de multiplicación a
1104
SNR
señal: voz / música
ruido: ruido de fondo / reverberación./...
a_bajo
a_alto
a_mas bajo
Multiplicar el valor a por el umbral promedio
1105
umbral variable
representación en el dominio del tiempo o picos en subbloques
Comparar con umbral
1106
ITD

Fig. 11e

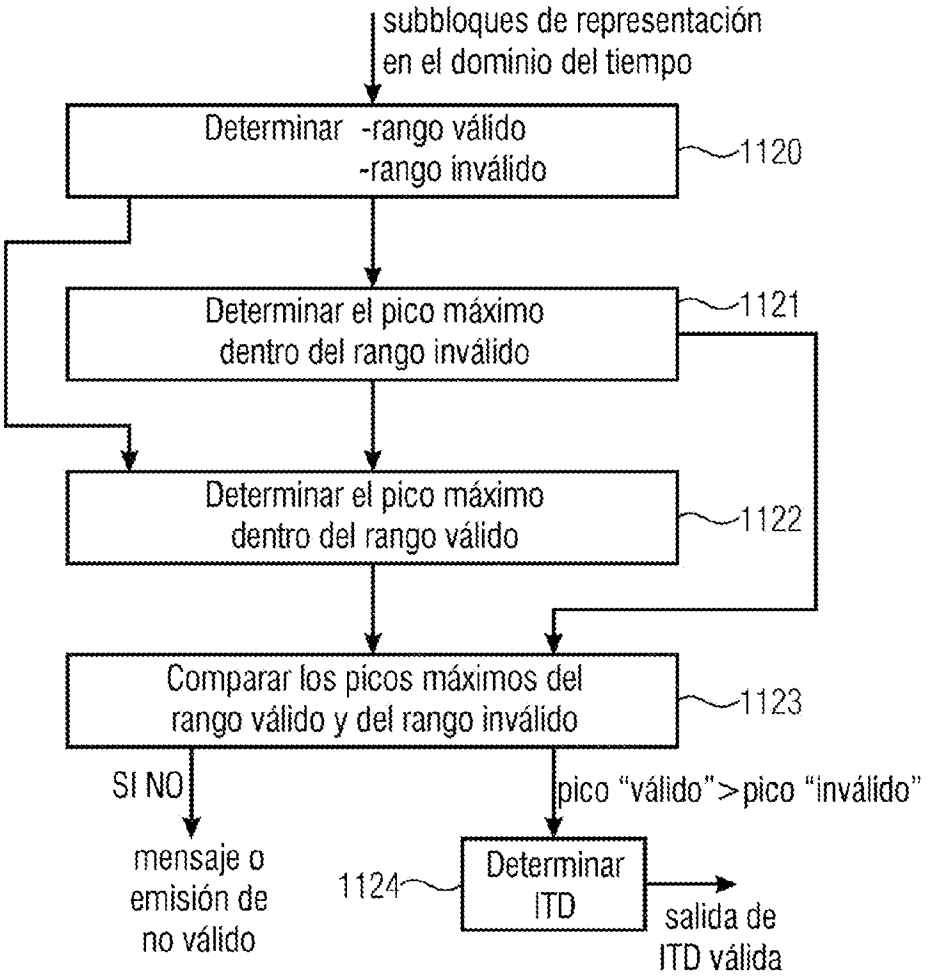

subbloques de representación
en el dominio del tiempo
Determinar -rango válido
-rango inválido
1120
Determinar el pico máximo
dentro del rango inválido
1121
Determinar el pico máximo
dentro del rango válido
1122
Comparar los picos máximos del
rango válido y del rango inválido
1123
SI NO
pico "válido">pico "inválido"
mensaje o
emisión de
no válido
1124
Determinar
ITD
salida de
ITD válida

Fig. 11f

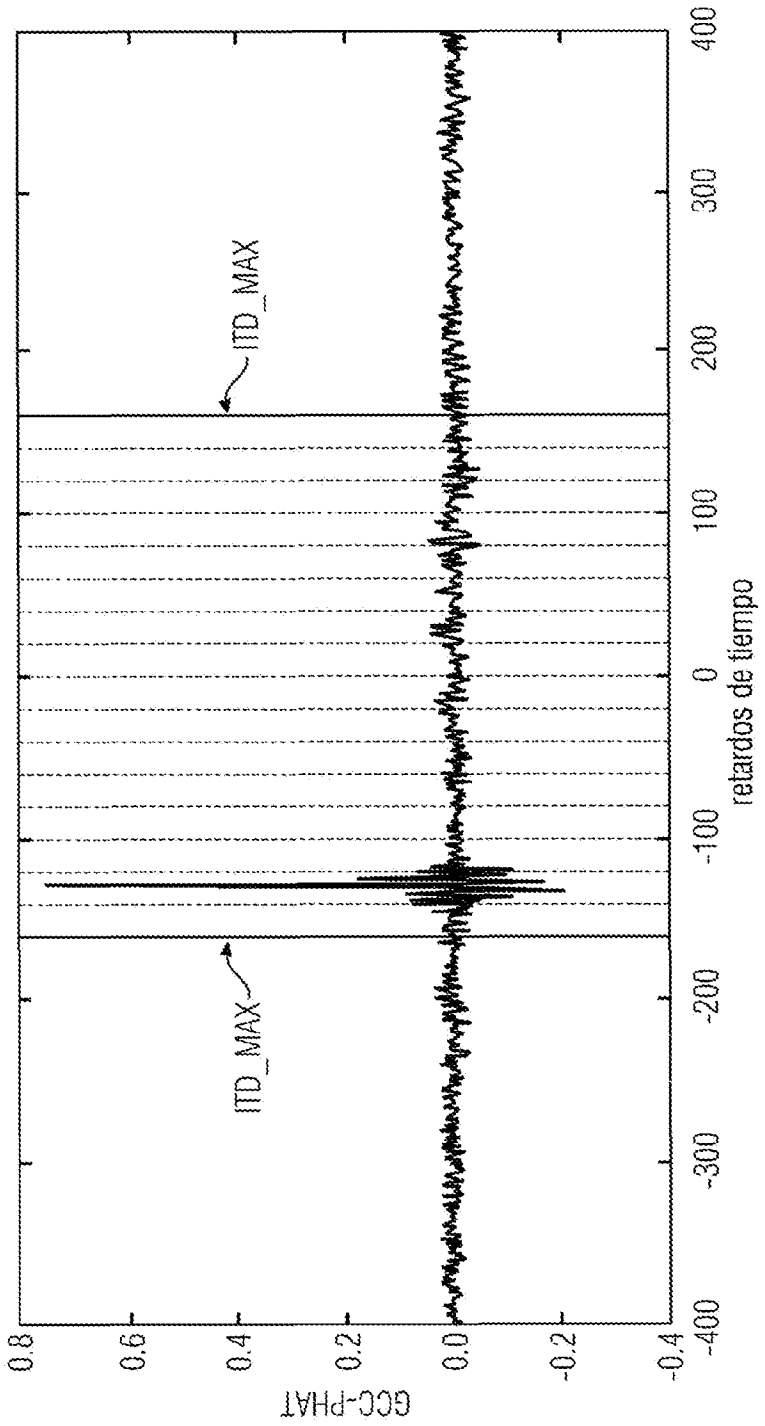

ITD_MAX
ITD_MAX
GCC-PHAT
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-400
-300
-200
-100
0
100
200
300
400
retardos de tiempo

Fig. 12

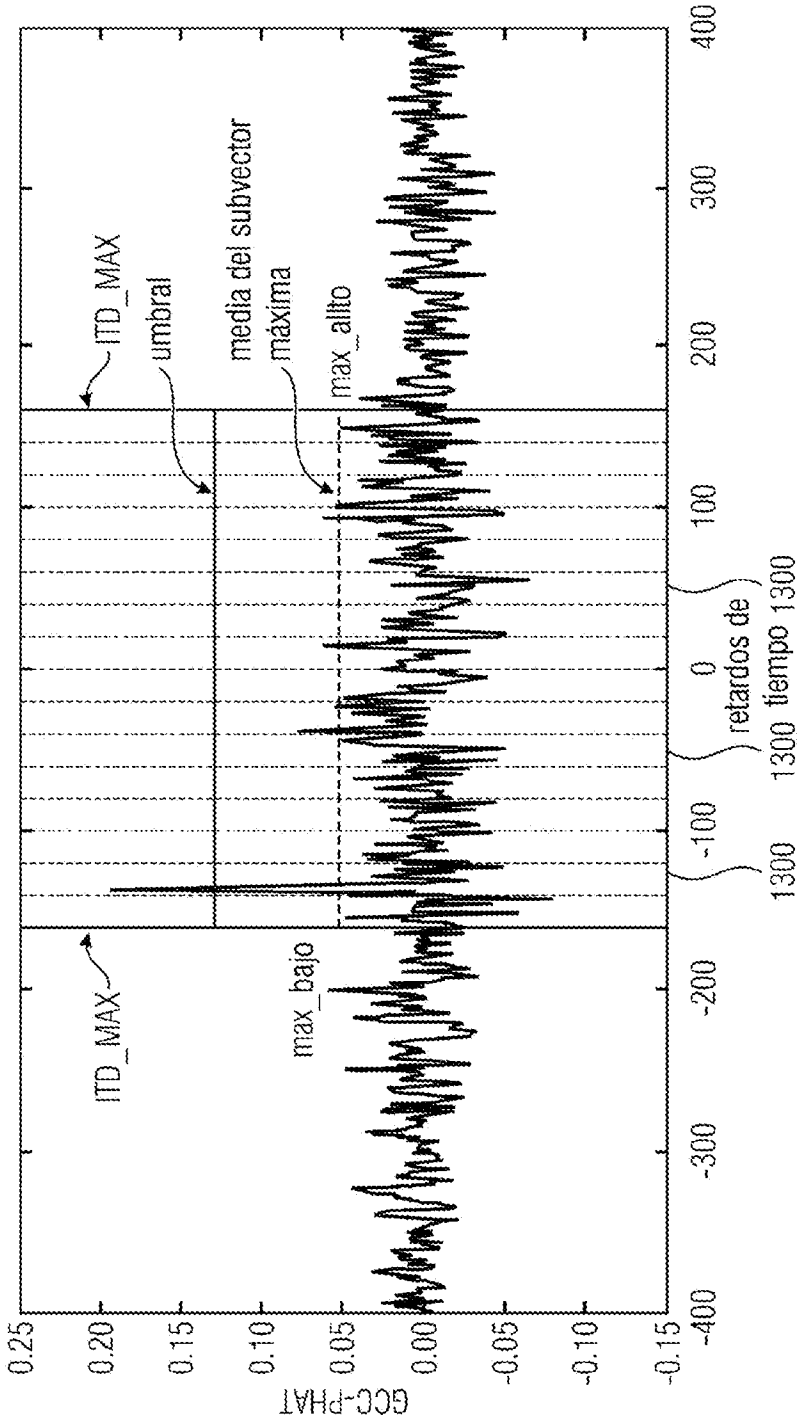

GCC-PHAT
0.25
0.20
0.15
0.10
0.05
0.00
-0.05
-0.10
-0.15
-400
-300
-200
-100
0
100
200
300
400
1300
1300
retardos de tiempo
ITD_MAX
umbral
media del subvector
máxima
max_alto
ITD_MAX
max_bajo

Fig. 13