

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 555 579**

51 Int. Cl.:

G10L 19/008 (2013.01)

G10L 19/02 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **05.04.2012 E 12713148 (0)**

97 Fecha y número de publicación de la concesión europea: **30.09.2015 EP 2834813**

54 Título: **Codificador de audio multicanal y método para codificar una señal de audio multicanal**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
05.01.2016

73 Titular/es:

**HUAWEI TECHNOLOGIES CO., LTD (100.0%)
Huawei Administration Building, Bantian,
Longgang District
Shenzhen, Guangdong 518129, CN**

72 Inventor/es:

**VIRETTE, DAVID;
LANG, YUE y
XU, JIANFENG**

74 Agente/Representante:

LEHMANN NOVO, María Isabel

ES 2 555 579 T3

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

DESCRIPCIÓN

Codificador de audio multicanal y método para codificar una señal de audio multicanal

5 CAMPO DE LA INVENCION

La presente invención se refiere a la codificación de audio y en particular a una codificación de audio espacial paramétrica también conocida como una codificación de audio multicanal paramétrica.

10 ANTECEDENTES DE LA INVENCION

La codificación de audio multicanal o estéreo paramétrica según se describe, a modo de ejemplo, en el documento de C. Faller y F. Baumgarte, titulado "Representación eficiente de señal de audio espacial utilizando una parametrización perceptual" en Proc. IEEE Workshop sobre la aplicación de procesamiento de señales para audio y acústica, octubre de 2001, páginas 199-202, utiliza pistas espaciales para sintetizar señales de audio multicanal a partir de señales de audio – normalmente mono o estéreo – de mezcla descendente. En condiciones normales, las señales de audio de mezcla descendente resultan de una superposición de una pluralidad de señales de canal de audio de una señal de audio multicanal, p.e., de una señal de audio estéreo. Estos menos canales están codificados en forma de onda e información lateral, esto es, las pistas espaciales, relacionadas con las relaciones de canales de señales originales se añade como parámetros de codificación a los canales de audio codificados. El decodificador utiliza esta información lateral para regenerar el número original de canales de audio sobre la base de los canales de audio codificados en forma de onda decodificados.

Otro método para la codificación de audio multicanal se da a conocer en el documento WO 2011/072729 A1.

Un codificador de estéreo paramétrico básico puede utilizar las diferencias de niveles entre canales (ILD) como una pista necesaria para generar la señal de estéreo a partir de la señal de audio de mezcla descendente monoaural. Codificadores más sofisticados pueden utilizar también la coherencia entre canales (ICC), que puede representar un grado de similitud entre las señales de canal de audio, esto es, canales de audio. Además, cuando se codifican señales de estéreo binaural, p.e., para señales de audio en 3D o presentación envolvente basada en auriculares, una diferencia de fase entre canales (IPD) puede desempeñar también una función importante para reproducir diferencias de fase/retardos entre los canales.

La diferencia de tiempo inter-aural (ITD) es la diferencia en el tiempo de llegada de un sonido 701 entre dos oídos 703, 705 como puede observarse a partir de la Figura 7. Es importante para la localización de sonidos, puesto que proporciona una pista para identificar la dirección 707 o el ángulo θ (theta) de incidencia de la fuente de sonido 701 (relativa a la cabeza 709). Si una señal llega a los oídos 703, 705 desde un lado, la señal tiene una ruta más larga 711 para alcanzar el oído lejano 703 (contralateral) y una ruta más corta 713 para alcanzar el oído cercano 705 (ipsilateral). Esta diferencia de longitud de recorrido da lugar a una diferencia de tiempo 715 entre las llegadas del sonido a los oídos 703, 705, lo que se detecta y ayuda al proceso de identificar la dirección 707 de la fuente de sonido 701.

La Figura 7 proporciona un ejemplo de ITD (que se indica como Δt o la diferencia de tiempo 715). Las diferencias en el tiempo de llegada en los dos oídos 703, 705 se indican por un retardo de la forma de onda del sonido. Si llega primero una forma de onda al oído izquierdo 703, el valor de ITD 715 es positivo, de no ser así, es negativo. Si la fuente de sonido 701 está directamente en frente del oyente, la forma de onda llega al mismo tiempo a ambos oídos 703, 705 y de este modo el valor de ITD 715 es cero.

Las pistas de ITD son importantes para la mayor parte de la grabación estéreo. A modo de ejemplo, la señal de audio binaural, que puede obtenerse a partir de la grabación real utilizando, a modo de ejemplo, un cabezal ficticio o síntesis binaural sobre la base del procesamiento de la denominada Función de Transferencia Relativa a la Cabeza (HRTF), se utiliza para la grabación de música o las conferencias de audio. Por lo tanto, es un parámetro muy importante para el codificador-decodificador (códec) de estéreo paramétrico de baja tasa binaria y en particular, para un codificador-decodificador que tiene como objetivo la aplicación conversacional. Un algoritmo de estimación de ITD de baja complejidad y estable se necesita para el codificador-decodificador de estéreo paramétrico de baja tasa binaria. Además, el uso de parámetros de ITD, p.ej., en adición a otros parámetros, tales como diferencias de niveles entre canales (CLDs o ILDs) y la coherencia entre canales (ICC), puede aumentar la carga de tasa binaria. Para este escenario operativo de muy baja tasa binaria específico, solamente se puede transmitir un parámetro de banda completa ITD. Cuando solamente se estima un parámetro ITD de banda completa, la limitación sobre la estabilidad se hace todavía más difícil de conseguir.

En la técnica anterior, los métodos de estimación de ITD pueden clasificarse en tres categorías principales. La estimación de ITD puede basarse en métodos en el dominio temporal. ITD se estima sobre la base de la correlación cruzada de dominio temporal entre canales. ITD corresponde al retardo en donde es máxima la correlación cruzada en el dominio temporal

$$(f * g)[n] \stackrel{\text{def}}{=} \sum_{m=-\infty}^{\infty} f^*[m]g[n + m]$$

Este método da a conocer una estimación no estable del retardo a través de varias tramas. Esto es particularmente verdadero cuando las señales de entrada f y g son señales de banda ancha con escenario operativo de sonido complejo puesto que diferentes señales de sub-banda pueden tener valores de ITD diferentes. Un ITD no estable puede dar lugar a la introducción de un chasquido (ruido) cuando el retardo se conmuta para tramas consecutivas en el decodificador. Cuando este análisis en el dominio temporal se realiza sobre la señal de banda completa, la tasa binaria de la estimación de ITD en el dominio temporal es baja, puesto que solamente se estima, codifica y transmite un parámetro ITD. Sin embargo, la complejidad es muy alta, debido al cálculo de correlación cruzada sobre las señales con alta frecuencia de muestreo.

La segunda categoría del método de estimación de ITD está basada en una combinación de métodos en el dominio temporal y de la frecuencia. En el documento de Marple, S.L., Jr. Titulado "Estimación de retardo de grupo y de retardo de fase mediante una correlación cruzada "analítica" en tiempo discreto", Signal Processing, IEEE Transactions on, vol. 47, nº 9, páginas 2604-2607, septiembre 1999, conteniendo la estimación de ITD en el dominio temporal y de la frecuencia las etapas siguientes:

1. Un análisis de Transformada de Fourier Rápida (FFT) se aplica a las señales de entrada con el fin de obtener coeficientes de la frecuencia.
2. Una correlación cruzada se calcula en el dominio de la frecuencia.
3. La correlación cruzada en el dominio de la frecuencia se convierte al dominio temporal utilizando una transformada de Fourier inversa FFT.
4. El valor de ITD se estima en el dominio temporal complejo.

Este método puede conseguir también la limitación de baja tasa binaria, puesto que solamente un ITD de banda completa es objeto de estimación, codificación y transmisión. Sin embargo, la complejidad es muy alta, debido al cálculo de la correlación cruzada y la transformada de Fourier FFT inversa que hace que este método no sea aplicable cuando la complejidad de cálculo informático está limitada.

Por último, la última categoría realiza la estimación de ITD directamente en el dominio de la frecuencia. En el documento de Baumgate, F.; Faller, C.; titulado "Codificación de pistas binaurales – parte I: fundamentos psicoacústicos y principios de diseño", Speech and Audio Processing, IEEE Transactions en vol. 11, nº 6, páginas 509-519, noviembre 2003 y en el documento Faller, C.; Baumgate, F.; titulado "Código de pistas binaurales – parte II: sistemas y aplicaciones", Procesamiento de audio y de la voz, IEEE Transactions on, vol. 11, nº 6, páginas 520-531, noviembre 2003, ITD se estima en el dominio de la frecuencia y para cada banda de frecuencias, se codifica y transmite un ITD. La complejidad de esta solución está limitada, pero la tasa binaria requerida para este método es alta, puesto que un ITD por sub-banda ha de transmitirse.

Además, la fiabilidad y estabilidad del ITD estimado depende del ancho de banda de la frecuencia de la señal de sub-bandas puesto que para un valor de ITD de sub-banda grande podría no ser compatible (fuentes de audio diferentes con diferentes posiciones podrían estar presentes en la señal de audio de banda limitada). Los sistemas de codificación de audio multicanal paramétrica, de muy baja tasa binaria, no tienen solamente la limitación sobre la tasa binaria, sino también la limitación sobre la complejidad disponible en particular para la puesta en práctica objetivo del códec en un terminal móvil en donde debe economizarse la vida útil de la batería. Los algoritmos de estimación de ITD más avanzados no pueden satisfacer ambos requisitos de baja tasa binaria y baja complejidad al mismo tiempo, mientras se mantiene una buena calidad en términos de la estabilidad de la estimación ITD.

SUMARIO DE LA INVENCION

Es el objetivo de la invención dar a conocer un concepto para un codificador de audio multicanal que proporciona, a la vez, una baja tasa binaria y una baja complejidad mientras se mantiene una buena calidad, en términos de estabilidad de la estimación de ITD.

Este objetivo se consigue mediante las características operativas establecidas en las reivindicaciones independientes. Otras formas de puesta en práctica son evidentes a partir de las reivindicaciones subordinadas, la descripción y las Figuras adjuntas.

La invención está basada en el hallazgo operativo de que la aplicación de una promediación inteligente a las diferencias entre canales, tales como ITD e IPD entre las partes diseñadas de banda limitada de dos señales de canal de audio de una señal de audio multicanal reduce, a la vez, la tasa binaria y la complejidad informática debido al procesamiento de banda limitada al mismo tiempo que se mantiene una buena calidad en términos de estabilidad

de la estimación de ITD. Una promediación inteligente discrimina las diferencias entre canales por su signo y realiza diferentes medias dependiendo de ese signo con lo que se aumenta la estabilidad del procesamiento de diferencias entre canales.

- 5 Con el fin de describir la invención en detalle, se utilizarán los términos, abreviaturas y notaciones siguientes:
- BCC: Codificación de pistas binaurales, codificación de señales estéreo o multicanal utilizando una mezcla descendente y pistas binaurales (o parámetros espaciales) para describir las relaciones entre canales.
- 10 Pistas binaurales: las pistas entre canales entre las señales de entrada en los oídos izquierdo y derecho (véase también ITD, ILD e IC).
- 15 CLD: Diferencia de niveles entre canales, lo mismo que ILD.
- FFT: Puesta en práctica rápida de la DFT, indicada como Transformada de Fourier Rápida.
- HRTF: Función de transferencia relacionada con la cabeza, transducción de creación de modelos de sonido desde una fuente a las entradas de los oídos izquierdo y derecho en un campo libre.
- 20 IC: Coherencia inter-aural, esto es, el grado de similitud entre las señales de entrada en los oídos izquierdo y derecho. Este término se refiere también, a veces, como IAC o correlación cruzada intraural (IACC)
- 25 ICC: Coherencia entre canales, correlación intercanal. Lo mismo que IC, pero con una definición más general entre cualquier par de señales (p.e., par de señales de altavoces, par de señales de entrada de auriculares, etc.).
- 30 ICPD: Diferencia de fase intercanal. Diferencia de fase media entre un par de señales.
- ICLD: Diferencia de nivel intercanal. Lo mismo que ILD, pero con una definición más general entre cualquier par de señales (p.ej., par de señales de altavoces, par de señal de entrada a auriculares, etc.).
- 35 ICTD: Diferencia de tiempo intercanal. Lo mismo que ITD, pero con una definición más general entre cualquier par de señales (p.ej., par de señales de altavoces, par de señal de entrada a auriculares, etc.).
- 40 ILD: Diferencia de nivel inter-aural, esto es, diferencia de nivel entre las señales de entrada en los oídos izquierdo y derecho. Este término se refiere también, a veces, como una diferencia de intensidad inter-aural (IID).
- IPD: Diferencia de fase inter-aural, esto es, diferencia de fase entre las señales de entrada en los oídos izquierdo y derecho.
- 45 ITD: Diferencia de tiempo inter-aural, esto es, diferencia de tiempo entre las señales de entrada en los oídos izquierdo y derecho. Este término se refiere también a veces como retardo inter-aural.
- 50 ICD: Diferencia intercanal. El término general para una diferencia entre dos canales, p.ej., una diferencia temporal, una diferencia de fase, una diferencia de nivel o una coherencia entre los dos canales.
- Mezcla: Dadas un número de señales de origen (p.ej., instrumentos registrados por separado, grabación multipista), el proceso de generar señales de estéreo o de audio multicanal previstas para una reproducción de audio espacial se indica como mezcla.
- 55 OCPD: Diferencia de fase de canal global. Una modificación de fase común de dos o más canales de audio.
- 60 Audio espacial: Señales de audio que, cuando se reproducen por intermedio de un sistema de reproducción adecuado, evocan una imagen espacial de auditorio.
- Pistas espaciales: Pistas pertinentes para la percepción espacial. Este término se utiliza para pistas entre pares de canales de una señal de audio multicanal o de estéreo (véase también ICTD, ICLD e ICC). También indicadas como parámetros espaciales o pistas binaurales.
- 65

En conformidad con un primer aspecto de la idea inventiva, la invención se refiere a un método para determinar un parámetro de codificación para una señal de canal de audio de una pluralidad de señales de canal de audio de una señal de audio multicanal, teniendo cada señal de canal de audio valores de señales de canal de audio, comprendiendo el método: determinar una transformada de frecuencia de los valores de señal de canal de audio de la señal de canal de audio; determinar una transformada de frecuencia de los valores de la señal de audio de referencia de una señal de audio de referencia, en donde la señal de audio de referencia es otra señal de canal de audio de la pluralidad de señales de canal de audio; determinar diferencias entre canales para al menos cada sub-banda de frecuencias de un subconjunto de sub-bandas de frecuencias, indicando cada diferencia entre canales una diferencia de fase o una diferencia temporal entre una parte de señal de banda limitada de la señal de canal de audio y una parte de señal de banda limitada de la señal de audio de referencia en la respectiva sub-banda de frecuencias con la que está asociada la diferencia entre canales; determinar una primera media sobre la base en los valores positivos de las diferencias entre canales y determinar una segunda media basada en valores negativos de las diferencias entre canales y determinar el parámetro de codificación basado en la primera media y en la segunda media.

En conformidad con un segundo aspecto de la idea inventiva, la invención se refiere a un método para determinar un parámetro de codificación para una señal de canal de audio de una pluralidad de señales de canal de audio de una señal de audio multicanal, teniendo cada señal de canal de audio unos valores de señales de canal de audio, comprendiendo dicho método: determinar una transformada de frecuencia de los valores de señal de canal de audio de la señal de canal de audio; determinar una transformada de frecuencia de los valores de la señal de audio de referencia de la señal de audio de referencia, en donde la señal de audio de referencia es una señal de audio de convección descendente derivada de al menos dos señales de canal de audio de la pluralidad de señales de canal de audio; determinar diferencias entre canales para al menos cada sub-banda de frecuencias de un subconjunto de sub-bandas de frecuencias, indicando cada diferencia entre canales una diferencia de fase o una diferencia temporal entre una parte de señal de banda limitada de la señal de canal de audio y una parte de señal de banda limitada de la señal de audio de referencia en la respectiva sub-banda de frecuencias, con la que está asociada la diferencia entre canales; determinar una primera media basada en valores positivos de las diferencias entre canales y determinar una segunda media basada en valores negativos de las diferencias entre canales y determinar el parámetro de codificación sobre la base de la primera media y de la segunda media.

La parte de señal de banda limitada puede ser una parte de señal en el dominio de la frecuencia. Sin embargo, la parte de señal de banda limitada puede ser una parte de señal en el dominio temporal. En este caso, un transformador de dominio de frecuencia-dominio temporal, tal como un transformador de Fourier inverso puede utilizarse a este respecto. En el dominio temporal, puede realizarse una media de retardo de parte de señal de banda limitada que corresponde a una media de fase en el dominio de la frecuencia. Para el procesamiento de señales, puede utilizarse un sistema de proceso de ventanas denominado windowing, p.ej., Hamming windowing, para establecer una ventana de la parte de señal en el dominio temporal. La parte de señal de banda limitada puede expandirse a través solamente de un contenedor de frecuencias o sobre más de un contenedor de frecuencias.

En una primera posible forma de puesta en práctica del método según el primer aspecto o según el segundo aspecto de la idea inventiva, las diferentes entre canales son diferencias de fase entre canales o diferencias de fase intercanal.

En una segunda posible forma de puesta en práctica del método según el primer aspecto de la idea inventiva como tal o en conformidad con el segundo aspecto como tal o de acuerdo con la primera forma de puesta en práctica del primer aspecto o en conformidad con la primera forma de puesta en práctica del segundo aspecto de la idea inventiva, el método comprende, además: determinar una primera desviación estándar sobre la base de valores positivos de las diferencias entre canales y determinar una segunda desviación estándar basada en valores negativos de las diferencias entre canales, en donde la determinación del parámetro de codificación está basada en la primera desviación estándar y en la segunda desviación estándar.

En una tercera posible forma de puesta en práctica del método según el primer aspecto o según el segundo aspecto como tales o en conformidad con cualquiera de las formas de puesta en práctica precedentes del primer aspecto o en conformidad con cualquiera de las formas de puesta en práctica precedentes del segundo aspecto, una sub-banda de frecuencias comprende uno o una pluralidad de contenedores de frecuencias.

En una cuarta posible forma de puesta en práctica del método en conformidad con el primer aspecto como tal o en conformidad con el segundo aspecto como tal o en conformidad con cualquiera de las formas de puesta en práctica precedentes del primer aspecto o en conformidad con cualquiera de las formas de puesta en práctica precedentes del segundo aspecto, la determinación de diferencias entre canales para al menos cada sub-banda de frecuencias de un subconjunto de sub-bandas de frecuencias comprende: determinar un espectro cruzado como una correlación cruzada a partir de la transformada de frecuencia de los valores de la señal de canal de audio y la transformada de frecuencia de los valores de señales de audio de referencia; determinar diferencias de fase entre canales para cada sub-banda de frecuencias basada en el espectro cruzado.

En una quinta posible forma de puesta en práctica del método en conformidad con la cuarta forma de puesta en

práctica del primer aspecto de la idea inventiva o en conformidad con la cuarta forma de puesta en práctica del segundo aspecto de la idea inventiva, la diferencia de fase entre canales de un contenedor de frecuencias o de una sub-banda de frecuencias se determina como un ángulo del espectro cruzado.

5 En una sexta posible puesta en práctica del método en conformidad con la cuarta o la quinta forma de puesta en práctica del primer aspecto de la idea inventiva o en conformidad con la cuarta o quinta forma de puesta en práctica del segundo aspecto de la idea inventiva, el método comprende, además: determinar diferencias temporales inter-aurales basadas en las diferencias de fase entre canales; en donde la determinación de la primera media está basada en valores positivos de las diferencias temporales inter-aurales y la determinación de la segunda media está basada en valores negativos de las diferencias temporales inter-aurales.

10 En una séptima posible forma de puesta en práctica del método en conformidad con la cuarta o la quinta forma de puesta en práctica del primer aspecto de la idea inventiva o en conformidad con la cuarta o la quinta forma de puesta en práctica del segundo aspecto de la idea inventiva, la diferencia temporal inter-aural de una sub-banda de frecuencias se determina como una función de la diferencia de fase entre canales, dependiendo dicha función de varios contenedores de frecuencias y del contenedor de frecuencias o del índice de sub-bandas de frecuencias.

15 En una octava posible forma de puesta en práctica del método en conformidad con la sexta o la séptima forma de puesta en práctica del primer aspecto de la idea inventiva o en conformidad con la sexta o la séptima forma de puesta en práctica del segundo aspecto, la determinación del parámetro de codificación comprende: contar un primer número de diferencias temporales inter-aurales positivas y un segundo número de diferencias temporales inter-aurales negativas a través del número de sub-bandas de frecuencias incluidas en el subconjunto de sub-bandas de frecuencias.

20 En una novena posible forma de realización de puesta en práctica del método en conformidad con la octava forma de realización del primer aspecto de la idea inventiva o en conformidad con la octava forma de puesta en práctica del segundo aspecto de la idea inventiva, el parámetro de codificación se determina sobre la base de una comparación entre el primer número de diferencias temporales inter-aurales positivas y el segundo número de diferencias temporales inter-aurales negativas.

25 En una décima posible forma de puesta en práctica del método en conformidad con la novena forma de realización de puesta en práctica del primer aspecto de la idea inventiva o en conformidad con la novena forma de puesta en práctica del segundo aspecto de la idea inventiva, el parámetro de codificación se determina sobre la base de una comparación entre la primera desviación estándar y la segunda desviación estándar.

30 En una undécima posible forma de puesta en práctica del método en conformidad con la novena o la décima forma de puesta en práctica del primer aspecto de la idea inventiva o en conformidad con la novena o la décima forma de puesta en práctica del segundo aspecto de la idea inventiva, el parámetro de codificación se determina sobre la base de una comparación entre el primer número de diferencias temporales inter-aurales positivas y el segundo número de diferencias temporales inter-aurales negativas multiplicado por un primer factor.

35 En la duodécima posible forma de realización del método en conformidad con la undécima forma de puesta en práctica del primer aspecto o en conformidad con la undécima forma de puesta en práctica del segundo aspecto de la idea inventiva, el parámetro de codificación se determina sobre la base de una comparación entre el primer número de diferencias temporales inter-aurales positivas y el segundo número de diferencias temporales inter-aurales negativas multiplicado por un segundo factor.

40 En una decimotercera posible forma de puesta en práctica del método en conformidad con la sexta o la séptima forma de puesta en práctica del primer aspecto de la idea inventiva o en conformidad con la sexta o la séptima forma de puesta en práctica del segundo aspecto de la idea inventiva, la determinación de parámetro de codificación comprende: contar un primer número de diferencias entre canales positivas y un segundo número de diferencias entre canales negativas a través del número de sub-bandas de frecuencias que se incluye en el subconjunto de sub-bandas de frecuencias.

45 En una decimocuarta posible forma de puesta en práctica del método en conformidad con el primer aspecto como tal o en conformidad con el segundo aspecto como tal o en conformidad con cualquiera de las formas de puesta en práctica precedentes del primer aspecto o en conformidad con cualquiera de las formas de puesta en práctica precedentes del segundo aspecto, el método se aplica en uno o en combinaciones de los codificadores siguientes: un codificador ITU-T G.722, un codificador de ITU-T G.722 Anexo B, un codificador de ITU-T G.711.1, un codificador de ITU-T G.711.1 Anexo D y un Codificador de Servicios de Voz Mejorados en 3GPP.

50 En comparación con una estimación del ITD que proporciona una estimación media del ITD de sub-bandas, los métodos en conformidad con el primero o segundo aspecto seleccionan el ITD más pertinente dentro de la sub-banda. De este modo, se consigue una baja tasa binaria y una estimación de ITD de baja complejidad al mismo tiempo que se mantiene una buena calidad en términos de estabilidad de la estimación de ITD.

En conformidad con un tercer aspecto de la idea inventiva, la invención se refiere a un codificador de audio multicanal para determinar un parámetro de codificación para una señal de canal de audio de una pluralidad de señales de canal de audio de una señal de audio multicanal, teniendo cada señal de canal de audio valores de señal de canal de audio, comprendiendo el codificador de audio espacial paramétrico: un transformador de frecuencias, tal como un dispositivo basado en la transformada de Fourier, para determinar una transformada de frecuencias de los valores de señales de canales de audio de la señal de canal de audio y para determinar una transformada de frecuencia de los valores de señales de audio de referencia de una señal de audio de referencia, en donde la señal de audio de referencia es otra señal de canal de audio de la pluralidad de señales de canal de audio, un dispositivo de determinación de diferencias entre canales para determinar las diferencias entre canales para al menos cada sub-banda de frecuencias de un subconjunto de sub-bandas de frecuencias, indicando cada diferencia intercanal una diferencia de fase o diferencia temporal entre una parte de señal de banda limitada de la señal de canal de audio y una parte de señal de banda limitada de la señal de audio de referencia en la respectiva sub-banda de frecuencias con la que está asociada la diferencia intercanal; un dispositivo de determinación de medias para determinar una primera media sobre la base de valores positivos de las diferencias entre canales y para determinar una segunda media basada en valores negativos de las diferencias entre canales y un dispositivo de determinación de parámetro de codificación para determinar el parámetro de codificación sobre la base de la primera media y de la segunda media.

En conformidad con un cuarto aspecto de la idea inventiva, la invención se refiere a un codificador de audio multicanal para determinar un parámetro de codificación para una señal de canal de audio de una pluralidad de señales de canal de audio de una señal de audio multicanal, teniendo cada señal de canal de audio valores de señal de canal de audio, comprendiendo dicho codificador de audio espacial paramétrico: un dispositivo de transformación de frecuencias tal como un dispositivo basado en la transformada de Fourier, para determinar una transformada de la frecuencia de los valores de señal de canal de audio de la señal de canal de audio y para determinar una transformada de frecuencia de valores de señal de audio de referencia de una señal de audio de referencia, en donde la señal de audio de referencia es una señal de audio de mezcla descendente derivada de al menos dos señales de canal de audio de entre la pluralidad de señales de canal de audio; un dispositivo de determinación de diferencias entre canales para determinar las diferencias entre canales para al menos cada sub-banda de frecuencias de un subconjunto de sub-bandas de frecuencias, indicando cada diferencia intercanal una diferencia de fase o una diferencia temporal entre una parte de señal de banda limitada de la señal de canal de audio y una parte de señal de banda limitada de la señal de audio de referencia en la respectiva sub-banda de frecuencias a la que está asociada la diferencia intercanal. Un dispositivo de determinación de medias para determinar una primera media basada en valores positivos de las diferencias entre canales y para la determinación de una segunda media basada en valores negativos de las diferencias entre canales y un dispositivo de determinación de parámetro de codificación para determinar el parámetro de codificación sobre la base de la primera media y de la segunda media.

En conformidad con un quinto aspecto de la idea inventiva, la invención se refiere a un programa informático con un código de programa para realizar el método en conformidad con el primer aspecto de la idea inventiva como tal o en conformidad con el segundo aspecto como tal o en conformidad con cualquiera de las reivindicaciones precedentes del primer aspecto de la idea inventiva o en conformidad con cualquiera de las reivindicaciones precedentes del segundo aspecto cuando se ejecutan en un ordenador.

El programa informático tiene una complejidad reducida y por ello, se puede poner en práctica de forma eficiente en un terminal móvil en donde debe economizarse la vida útil de la batería.

En conformidad con un sexto aspecto de la idea inventiva, la invención se refiere a un identificador de audio espacial paramétrico que está configurado para poner en práctica el método en conformidad con el primer aspecto como tal o en conformidad con el segundo aspecto como tal o en conformidad con cualquiera de las formas de puesta en práctica precedentes del primer aspecto o en conformidad con cualquiera de las formas de puesta en práctica precedentes del segundo aspecto.

En una primera posible forma de puesta en práctica del codificador de audio espacial paramétrico en conformidad con el sexto aspecto de la idea inventiva, el codificador de audio espacial paramétrico comprende un procesador que pone en práctica el método en conformidad con el primer aspecto como tal o en conformidad con el segundo aspecto como tal o en conformidad con cualquiera de las formas de puesta en práctica del primer aspecto o en conformidad con cualquiera de las formas de puesta en práctica del segundo aspecto.

En una segunda posible forma de puesta en práctica del codificador de audio espacial paramétrico en conformidad con el sexto aspecto como tal o en conformidad con la primera forma de puesta en práctica del sexto aspecto, el codificador de audio espacial paramétrico comprende un dispositivo de transformación de frecuencias tal como un dispositivo basado en la transformada de Fourier para determinar una transformada de frecuencias de los valores de señales de canal de audio de la señal de canal de audio y para determinar una transformada de frecuencias de valores de la señal de audio de referencia de una señal de audio de referencia, en donde la señal de audio de referencia es otra señal de canal de audio de entre la pluralidad de señales de canal de audio o una señal de audio de mezcla descendente derivada de al menos dos señales de canal de audio de entre la pluralidad de señales de canal de audio; un dispositivo de determinación de diferencias entre canales para determinar las diferencias entre

canales para al menos cada sub-banda de frecuencias de un subconjunto de sub-bandas de frecuencias, indicando cada diferencia entre canales una diferencia de fase o una diferencia temporal entre la parte de señal de banda limitada de la señal de canal de audio y la parte de señal de banda limitada de la señal de audio de referencia en la respectiva sub-banda, a la que está asociada la diferencia entre canales; un dispositivo de determinación de media para determinar una primera media basada en valores positivos de las diferencias entre canales y la determinación de una segunda media basada en valores negativos de las diferencias entre canales; y un dispositivo de determinación de parámetro de codificación para determinar el parámetro de codificación basado en la primera y en la segunda media.

En conformidad con un séptimo aspecto de la idea inventiva, la invención se refiere a un soporte legible por máquina tal como una memoria, en particular un disco compacto, con un programa informático que comprende un código de programa para realizar el método en conformidad con el primer aspecto como tal o en conformidad con el segundo aspecto como tal en conformidad con cualquiera de las reivindicaciones precedentes del primer aspecto o en conformidad con cualquiera de las reivindicaciones precedentes del segundo aspecto cuando se ejecutan en un ordenador.

Los métodos aquí descritos pueden ponerse en práctica como software en un Procesador de Señal Digital (DSP), un microcontrolador o en cualquier otro procesador secundario o un circuito de hardware dentro de un circuito integrado específico de la aplicación (ASIC).

La invención puede ponerse en práctica en circuitos de electrónica digital, o en equipos informáticos, firmware, software o en sus combinaciones.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

Otras formas de realización de la invención se describirán con respecto a las Figuras siguientes, en donde:

La Figura 1 ilustra un diagrama esquemático de un método para generar un parámetro de codificación para una señal de canal de audio en conformidad con una forma de puesta en práctica;

La Figura 2 ilustra un diagrama esquemático de un algoritmo de estimación de ITD en conformidad con una forma de puesta en práctica;

La Figura 3 ilustra un diagrama esquemático de un algoritmo de selección de ITD en conformidad con una forma de puesta en práctica;

La Figura 4 ilustra un diagrama de bloques de un codificador de audio paramétrico en conformidad con una forma de puesta en práctica;

La Figura 5 ilustra un diagrama de bloques de un decodificador de audio paramétrico en conformidad con una forma de puesta en práctica;

La Figura 6 ilustra un diagrama de bloques de un codificador y de un decodificador de audio estéreo paramétrico en conformidad con una forma de puesta en práctica; y

La Figura 7 ilustra un diagrama esquemático que representa los principios de las diferencias temporales inter-aurales.

DESCRIPCIÓN DETALLADA DE LAS FORMAS DE REALIZACIÓN DE LA INVENCION

La Figura 1 ilustra un diagrama esquemático de un método para generar un parámetro de codificación para una señal de canal de audio en conformidad con una forma de puesta en práctica la invención.

El método 100 x_1 es para determinar el parámetro de codificación ITD para una señal de canal de audio x_1 de entre una pluralidad de x_1 , x_2 de una señal de audio multicanal. Cada señal de canal de audio x_1 , x_2 tiene valores de señales de canal de audio $x_1[n]$, $x_2[n]$. La Figura 1 ilustra el caso estéreo en donde la pluralidad de señales de canal de audio comprende un canal de audio izquierdo x_1 y un canal de audio derecho x_2 . El método 100 comprende:

determinar 101 una transformada de frecuencias $x_1[k]$ de los valores de señal de canal de audio $x_1[n]$ de la señal de canal de audio x_1 ;

determinar 103 una transformada de frecuencias $x_2[k]$ de valores de señales de audio de referencia $x_2[n]$ de una señal de audio de referencia x_2 en donde la señal de audio de referencia es otra señal de canal de audio x_2 de entre la pluralidad de señales de canal de audio o una señal de audio de mezcla descendente derivada de al menos dos señales de canal de audio x_1 , x_2 de entre la pluralidad de señales de canal de audio;

determinar 105 diferencias entre canales ICD[b] para al menos cada sub-banda de frecuencias b de un subconjunto de sub-bandas de frecuencias, indicando cada diferencia entre canales una diferencia de fase IPD[b] o una diferencia temporal ITD[b] entre una parte de señal de banda limitada de la señal de canal de audio y una parte de señal de banda limitada de la señal de audio de referencia en la respectiva sub-banda de frecuencias b de la diferencia entre canales a la que está asociada;

determinar 107 una primera media ITD_{mean_pos} sobre la base en valores positivos de las diferencias entre canales ICD[b] y determinar una segunda media ITD_{mean_neg} sobre la base de los valores negativos de las diferencias entre canales ICD[b]; y

determinar 109 el ITD del parámetro de codificación sobre la base de la primera media y de la segunda media.

En una forma de puesta en práctica, la parte de señal de banda limitada de la señal de canal de audio y la parte de señal de banda limitada de la señal de audio de referencia se refieren a la respectiva sub-banda y sus contenedores de frecuencias en el dominio de la frecuencia.

En una forma de puesta en práctica, la parte de señal de banda limitada de la señal de canal de audio y la parte de señal de banda limitada de la señal de audio de referencia se refieren a la respectiva señal transformada en el tiempo de la sub-banda en el dominio temporal.

La parte de señal de banda limitada puede ser una parte de señal en el dominio de la frecuencia. Sin embargo, la parte de señal de banda limitada puede ser una parte de señal en el dominio temporal. En este caso, un dispositivo de transformación de dominio de la frecuencia-dominio temporal tal como un dispositivo basado en la transformada Fourier inversa puede utilizarse a este respecto. En el dominio temporal, una media de retardo de partes de la señal de banda limitada pueden realizarse a este respecto estando en correspondencia con una media de fase en el dominio de las frecuencias. Para el procesamiento de señales, se puede emplear una tecnología denominada windowing, p.ej., Hamming windowing, para establecer ventanas para la parte de señal en el dominio temporal.

La parte de señal de banda limitada puede extenderse sobre solamente un contenedor de frecuencias o sobre más de uno contenedores de frecuencias.

En una forma de realización, el método 100 se procesa como sigue:

En una primera etapa correspondiente a 101 y 103 en la Figura 1, se aplica una transformada de tiempo-frecuencia en el canal de entrada del dominio temporal, p.ej., el primer canal de entrada x_1 y el canal de referencia del dominio temporal, p.ej., el segundo canal de entrada x_2 . En caso de estéreo, estos son los canales izquierdo y derecho. En una forma de realización preferida, la transformada de tiempo-frecuencia es una Transformada de Fourier Rápida (FFT) o una Transformada de Fourier a Corto Plazo (STFT). En una forma de realización alternativa, la transformada de tiempo-frecuencia es un banco de filtros de modulación cosenoidal en un banco de filtros completo.

En una segunda etapa que corresponde a 105 en la Figura 1, se calcula un espectro cruzado para cada contenedor de frecuencias [b] de la FFT como:

$$c[b] = X_1[b]X_2^*[b],$$

en donde c[b] es el espectro cruzado de contenedor de frecuencias [b] y $X_1[b]$ y $X_2[b]$ son los coeficientes de la transformada FFT de los dos canales. Un asterisco * indica una conjugación completa. Para este caso, una sub-banda b corresponde directamente a un contenedor de frecuencias [k], un contenedor de frecuencias [b] y [k] representan exactamente el mismo contenedor de frecuencias.

Como alternativa, el espectro cruzado se calcula por sub-banda [k] como:

$$c[b] = \sum_{k=k_b}^{k_{b+1}-1} X_1[k]X_2^*[k],$$

en donde c[b] es el espectro de la sub-banda [b] y $X_1[k]$ y $X_2[k]$ son los coeficientes de la transformada FFT de los dos canales, a modo de ejemplo, los canales izquierdo y derecho en caso de estéreo. El asterisco * indica una conjugación completa. k_b es el contenedor inicial de la sub-banda [b].

El espectro cruzado puede ser una versión suavizada, que se calcula aplicando la ecuación siguiente:

$$c_{sm}[b,i]=SMW_i*c_{sm}[b,i-1]+(1-SMW_i)*c[b]$$

en donde SMW1 es el factor de alisado. i es el índice de tramas.

Las diferencias de fase entre canales (IPDs) se calculan por sub-banda sobre la base del espectro cruzado como:

$$IPD[b] = \angle c[b]$$

en donde la operación \angle es el operador de argumento para calcular el ángulo de $c[b]$. Conviene señalar que en caso de suavizado del espectro cruzado, $c_{sm}[b, i]$ se utiliza para el cálculo de IPD como:

$$IPD[b] = \angle c_{sm}[b, i]$$

En una tercera etapa que corresponde a 105 en la Figura 1, los valores de ITDs de cada contenedor de frecuencias (o sub-banda) se calculan sobre la base de IPDs.

$$ITD[b] = \frac{IPD[b]N}{\pi b}$$

en donde N es el número de contenedores de FFT.

En una cuarta etapa, que corresponde a 107 en la Figura 1, se realiza el conteo de los valores positivos y negativos de ITD. La media y la desviación estándar de ITD positivos y negativos se basa en el signo de IPD como sigue:

$$ITD_{mean_pos} = \frac{\sum_{i=0}^{i=M} ITD(i)}{Nb_{pos}} \text{ en donde } ITD(i) \geq 0$$

$$ITD_{mean_neg} = \frac{\sum_{i=0}^{i=M} ITD(i)}{Nb_{neg}} \text{ en donde } ITD(i) < 0$$

$$ITD_{std_pos} = \sqrt{\frac{\sum_{i=0}^{i=M} (ITD(i) - ITD_{mean_pos})^2}{Nb_{pos}}} \text{ en donde } ITD(i) \geq 0$$

$$ITD_{std_neg} = \sqrt{\frac{\sum_{i=0}^{i=M} (ITD(i) - ITD_{mean_neg})^2}{Nb_{neg}}} \text{ en donde } ITD(i) < 0$$

en donde Nb_{pos} y Nb_{neg} son el número de ITD positivos y negativos, respectivamente. M es el número total de ITDs que se extraen. Conviene señalar que como alternativa, si ITD es igual a 0, puede contarse en un valor de IPD negativo o no contarse en ninguna de las medias.

En una quinta etapa que corresponde a 109 en la Figura 1, ITD se selecciona a partir de valores de ITD positivos y negativos sobre la base de la media y de la desviación estándar. El algoritmo de selección se ilustra en la Figura 3.

La Figura 2 ilustra un diagrama esquemático de un algoritmo de estimación de ITD 200 en conformidad con una forma de puesta en práctica de la invención.

En una primera etapa 201, que corresponde a 101 en la Figura 1, se aplica una transformada de tiempo-frecuencia en el canal de entrada del dominio temporal, p.ej., el primer canal de entrada x_1 . En una forma de realización preferida, la transformada de tiempo-frecuencia es una Transformada de Fourier Rápida (FFT) o una Transformada de Fourier a Corto Plazo (STFT). En una forma de realización alternativa, la transformada de tiempo-frecuencia es un banco de filtros con modulación cosenoidal o un banco de filtros complejos.

En una segunda etapa 203 que corresponde a 103 en la Figura 1, se aplica una transformada de tiempo-frecuencia en el canal de referencia del dominio temporal, p.ej., el segundo canal de entrada x_2 . En una forma de realización preferida, la transformada de tiempo-frecuencia es una Transformada de Fourier Rápida (FFT) o una Transformada de Fourier a Corto Plazo (STFT). En una forma de realización alternativa, la transformada de tiempo-frecuencia es un banco de filtros con modulación cosenoidal o un banco de filtros complejos.

En una tercera etapa posterior 205 que corresponde a 105 en la Figura 1, se calcula una correlación cruzada de cada contenedor de frecuencias lo que se realiza en un número limitado de contenedores de frecuencias o sub-bandas de frecuencias. Un espectro cruzado se calcula a partir de la correlación cruzada para cada contenedor de frecuencia $[b]$ de la FFT como:

$$c[b] = X_1[b]X_2^*[b],$$

5 en donde $c[b]$ es el espectro cruzado de un contenedor de frecuencias $[b]$ y $X_1[b]$ y $X_2[b]$ son los coeficientes de FFT de los dos canales. El asterisco $*$ indica una conjugación completa. Para este caso, una sub-banda b corresponde directamente a un contenedor de frecuencias $[k]$, representando el contenedor de frecuencias $[b]$ y $[k]$ exactamente el mismo contenedor de frecuencias.

10 Como alternativa, se calcula el espectro cruzado por sub-banda $[k]$ como:

$$c[b] = \sum_{k=k_b}^{k_{b+1}-1} X_1[k]X_2^*[k],$$

15 en donde $c[b]$ es el espectro cruzado de la sub-banda $[b]$ y $X_1[k]$ y $X_2[k]$ son los coeficientes de FFT de los dos canales, a modo de ejemplo, el canal izquierdo y derecho en caso de estéreo. El asterisco $*$ indica una conjugación completa. k_b es el contenedor inicial de la sub-banda $[b]$.

El espectro cruzado puede ser una versión suavizada, que se calcula mediante la ecuación siguiente.

$$20 \quad c_{sm}[b,i] = SMW_1 * c_{sm}[b,i-1] + (1 - SMW_1) * c[b]$$

en donde SMW_1 es el factor de alisado, i es el índice de tramas.

Las diferencias de fase entre canales (IPDs) se calculan por sub-bandas sobre la base de espectro cruzado como:

$$25 \quad IPD[b] = \angle c[b]$$

en donde la operación \angle es el operador de argumento para calcular el ángulo de $c[b]$. Conviene señalar que en el caso de suavizado del espectro cruzado, $c_{sm}[b, i]$ se utiliza para el cálculo de IPD como:

$$30 \quad IPD[b] = \angle c_{sm}[b,i]$$

En una cuarta etapa posterior 207 que corresponde a 105 en la Figura 1, se calculan los valores de ITDs de cada contenedor de frecuencias (o sub-banda) sobre la base de IPDs.

$$35 \quad ITD[b] = \frac{IPD[b]N}{\pi b}$$

en donde N es el número del contenedor de FFT.

40 En una quinta etapa posterior 209 que corresponde a 107 en la Figura 1, el valor de ITD calculado de la etapa 207 se comprueba como siendo mayor que cero. Si la respuesta es afirmativa, se procesa la etapa 211 y si la respuesta es negativa, se procesa la etapa 213.

45 En la etapa 211 después de que la etapa 209 se calcula una suma sobre un número de M valores del contenedor (o sub-banda) de frecuencias de ITD, p.ej., en conformidad con la expresión " Nb_itd_pos++ , $ltd_sum_pos+=ITD$ ".

En la etapa 213 después de la etapa 209 se calcula una suma sobre un número de M valores del contenedor (o sub-banda) de frecuencias de ITD, p.ej., en conformidad con " Nb_itd_neg++ , $ltd_sum_neg+=ITD$ ".

50 En la etapa 215 después de la etapa 211, se calcula una media de ITDs positivos en conformidad con la ecuación

$$ITD_{mean_pos} = \frac{\sum_{i=0}^{i=M} ITD(i)}{Nb_{pos}} \text{ en donde } ITD(i) \geq 0$$

en donde Nb_{pos} es el número de valores de ITD positivos y M es el número total de ITDs que se extraen.

55 En la etapa opcional 219 después de la etapa 215, se calcula una desviación estándar de valores ITDs positivos en conformidad con la ecuación

$$ITD_{std_pos} = \sqrt{\frac{\sum_{i=0}^{i=M} (ITD(i) - ITD_{mean_pos})^2}{Nb_{pos}}} \text{ en donde } ITD(i) \geq 0$$

En la etapa 217 después de la etapa 213, se calcula una media de valores ITDs negativos en conformidad con la ecuación

5

$$ITD_{mean_neg} = \frac{\sum_{i=0}^{i=M} ITD(i)}{Nb_{neg}} \text{ en donde } ITD(i) < 0$$

en donde Nb_{neg} es el número de valores de ITD negativos y M es el número total de valores ITDs que son extraídos.

10 En la etapa opcional 221 después de la etapa 217, se calcula una desviación estándar de los valores ITDs negativos en conformidad con la ecuación

$$ITD_{std_neg} = \sqrt{\frac{\sum_{i=0}^{i=M} (ITD(i) - ITD_{mean_neg})^2}{Nb_{neg}}} \text{ en donde } ITD(i) < 0$$

15 En una última etapa 223 que corresponde a 109 en la Figura 1 el valor de ITD se selecciona a partir de los valores de ITD positivos y negativos sobre la base de la media y de forma opcional, sobre la base de la desviación estándar. El algoritmo seleccionado es según se ilustra en la Figura 3.

20 Este método 200 puede aplicarse a una estimación de ITD de banda completa, en cuyo caso, las sub-bandas b cubren la gama completa de frecuencias (hasta B). Las sub-bandas b se pueden elegir para seguir una descomposición perceptual del espectro como, a modo de ejemplo, las bandas críticas o la denominada Ancho de banda Rectangular Equivalente (ERB). En una forma de realización alternativa, un valor de ITD de banda completa se puede estimar sobre la base de la sub-banda b más pertinente. Y lo que es más importante, debe entenderse que las sub-bandas que son perceptualmente pertinentes para la percepción de ITD (a modo de ejemplo, entre 200 Hz y 1500 Hz).

25 La ventaja de la estimación de los valores de ITD en conformidad con el primero o segundo aspecto de la idea inventiva es que, si existen dos altavoces en los lados izquierdo y derecho del oyente, respectivamente y si están funcionando al mismo tiempo, la media simple de todos los valores de ITD proporcionará un valor próximo a cero, lo que no es correcto. Puesto que el valor de ITD cero significa que el altavoz está simplemente en frente del oyente. Aun cuando la media de todos los valores de ITD no sea cero, se estrechará la imagen en estéreo. También en esta realización, a modo de ejemplo, el método 200 seleccionará un valor ITD de entre las medias de valores de ITD positivos y negativos, sobre la base de la estabilidad del ITD extraído, lo que proporciona una mejor estimación, en términos de dirección de la fuente del sonido.

30 La desviación estándar es una forma de medir la estabilidad de los parámetros. Si la desviación estándar es pequeña, los parámetros estimados son más estables y fiables. La finalidad de utilizar la desviación estándar de valores de ITD positivos y negativos es constatar cuál es más fiable. Y seleccionar el valor fiable como el ITD de salida final. Otro parámetro similar, tal como diferencia de extremismo puede utilizarse también para comprobar la estabilidad del ITD. Por lo tanto, la desviación estándar es opcional en este caso.

35 En otra forma de puesta en práctica, el conteo positivo y negativo se realiza directamente en los IPDs, puesto que existe una relación directa entre IPD e ITD. El proceso de decisión se realiza entonces directamente sobre las medias de IPD negativas y positivas.

40 El método 100, 200 según se describe en las Figuras 1 y 2 se puede aplicar en el codificador de la extensión en estero de ITU-T G.722, ITU-T G.722 Anexo B, G.711.1 y/o G.711.1 Anexo D. Además, el método descrito puede aplicarse también para el codificador de voz y de audio para aplicación móvil según se define en el códec de 3GPP EVS (Servicios de Voz Mejorados).

45 La Figura 3 ilustra un diagrama esquemático de un algoritmo de selección de ITD en conformidad con una forma de puesta en práctica de la invención.

50 En una primera etapa 301, el número Nb_{pos} de valores positivos de ITD se comprueba con respecto al número Nb_{neg} de valores negativos de ITD. Si Nb_{pos} es mayor que Nb_{neg} , se realiza la etapa 303; si Nb_{pos} no es mayor que Nb_{neg} se realiza la etapa 305.

55

5 En la etapa 303, la desviación estándar ITD_{std_pos} de valores de ITDs positivos se comprueba con respecto a la desviación estándar ITD_{std_neg} de ITDs negativos y el número Nb_{pos} de valores de ITD positivos se comprueba con respecto al número Nb_{neg} de valores negativos de ITD multiplicado por un primer factor A, p.ej., en conformidad con: $(ITD_{std_pos} < ITD_{std_neg}) \parallel (Nb_{pos} > = A * Nb_{neg})$. Si $ITD_{std_pos} < ITD_{std_neg}$ o $Nb_{pos} > A * Nb_{neg}$, ITD se selecciona como la media de valores positivos de ITD en la etapa 307. De no ser así, la relación entre los valores de ITD positivos y negativos se comprobará, además, en la etapa 309.

10 En la etapa 309, la desviación estándar ITD_{std_neg} de valores negativos de ITD se comprueba con respecto a la desviación estándar ITD_{std_pos} de valores positivos de ITD multiplicados por un segundo factor B, p.ej., en conformidad con: $(ITD_{std_neg} < B * ITD_{std_pos})$. Si $ITD_{std_neg} < B * ITD_{std_pos}$, el valor opuesto de la media de ITD negativa se seleccionará como ITD de salida en la etapa 315. De no ser así, el valor ITD de la trama anterior (Pre_itd) se comprueba en la etapa 317.

15 En la etapa 317, el valor ITD de la trama anterior se comprueba que es mayor que cero, p.ej., en conformidad con "Pre_itd > 0". Si Pre_itd > 0, el valor ITD de salida se selecciona como la media de valores ITD positivos en la etapa 323, y de no ser así, el valor ITD de salida es el valor opuesto de la media de ITD negativa en la etapa 325.

20 En la etapa 305, la desviación estándar ITD_{std_neg} de valores ITD negativos se comprueba con respecto a la desviación estándar ITD_{std_pos} de valores ITDs negativos y el número Nb_{neg} de valores de ITDs negativos se comprueba con respecto al número Nb_{pos} de valores de ITDs positivos multiplicado por un primer factor A, p.ej., en conformidad con $(ITD_{std_neg} < ITD_{std_pos}) \parallel (Nb_{neg} > = A * Nb_{pos})$. Si $ITD_{std_neg} < ITD_{std_pos}$ o $Nb_{neg} > A * Nb_{pos}$ se selecciona ITD como la media de los valores de ITDs negativos en la etapa 311. De no ser así, la relación entre los valores de ITD negativos y positivos se comprueba, además, en la etapa 313.

25 En la etapa 313, la desviación estándar ITD_{std_pos} de valores de ITDs positivos se comprueba con respecto a la desviación estándar ITD_{std_neg} de valores de ITDs negativos multiplicado por un segundo factor B, p.ej., en conformidad con: $(ITD_{std_pos} < B * ITD_{std_neg})$. Si $ITD_{std_pos} < B * ITD_{std_neg}$. El valor opuesto de la media de ITD positiva se selecciona como ITD de salida en la etapa 319. De no ser así, el valor ITD de la trama anterior (Pre_itd) se comprueba en la etapa 321.

30 En la etapa 321, el valor ITD de la trama anterior se comprueba que es mayor que cero, p.ej., en conformidad con "Pre_itd > 0". Si Pre_itd > 0, el valor ITD de salida se selecciona como la media de los valores de ITDs negativos en la etapa 327; de no ser así, el ITD de salida es el valor opuesto de la media de ITD positiva en la etapa 329.

35 La Figura 4 ilustra un diagrama de bloques de un codificador de audio paramétrico 400 en conformidad con una forma de puesta en práctica. El codificador de audio paramétrica 400 recibe una señal de audio multicanal 401 como señal de entrada y proporciona un flujo de bits como señal de salida 403. El codificador de audio paramétrica 400 comprende un generador de parámetros 405 acoplado a la señal de audio multicanal 401 para generar un parámetro de codificación 415, un generador de señales de mezcla descendente 407 acoplado a la señal de audio multicanal 401 para generar una señal de mezcla descendente 411 o una señal suma, un codificador de audio 409 acoplado al generador de señales de mezcla descendente 407 para codificar la señal de mezcla descendente 411 para proporcionar una señal de audio codificada 413 y un combinador 417, p.ej., un dispositivo formador de pluralidades de bits acoplado al generador de parámetros 405 y el codificador de audio 409 para formar un flujo de bits 403 a partir del parámetro de codificación 415 y de la señal codificada 413.

45 El codificador de audio paramétrico 400 pone en práctica un sistema de codificación de audio para señales de audio multicanal y estéreo, que solamente transmite un canal de audio único, p.ej., la representación de mezcla descendente del canal de audio de entrada junto con los parámetros adicionales que describen las "diferentes perceptualmente importante" entre los canales de audio x_1, x_2, \dots, x_M . El sistema de codificación está en conformidad con la codificación de pistas binaurales (BCC) porque las pistas binaurales desempeñan una función importante en dicho sistema. Según se indica en la Figura, los canales de audio de entrada x_1, x_2, \dots, x_M son objeto de mezcla descendente a un canal de audio único 411, también indicado como la señal suma. Como "diferencias perceptualmente importantes" entre los canales de audio x_1, x_2, \dots, x_M , el parámetro de codificación 415, p.ej., una diferencia temporal entre canales (ICTD), una diferencia de niveles entre canales (ICLD) y/o una coherencia entre canales (ICC) se estima como una función de la frecuencia y del tiempo y se transmite como información lateral al decodificador 500 según se describe en la Figura 5.

60 El generador de parámetros 405 que pone en práctica BCC procesa la señal de audio multicanal 401 con una determina resolución en tiempo y frecuencia. La resolución en frecuencia utilizada está motivada en gran medida por la resolución de frecuencia del sistema de auditorio. La psico-acústica sugiere que la percepción espacial esté basada muy probablemente en una representación de banda crítica de la señal de acústica. Esta resolución de la frecuencia se considera utilizando un banco de filtros invertibles, con sub-bandas con ancho de banda iguales o proporcionales al ancho de banda crítico del sistema de auditorio. Es importante que la señal suma 411 transmitida contenga todos los componentes de señal de la señal de audio multicanal 401. El objetivo es que cada componente de señal sea objeto de mantenimiento completo. La adición simple de los canales de entrada de audio x_1, x_2, \dots, x_M de la señal de audio multicanal 401 suele dar lugar a la amplificación o atenuación de componentes de la señal.

Dicho de otro modo, la potencia de los componentes de señales en la suma “simple” suele ser mayor o menor que la suma de la potencia del componente de señal correspondiente de cada canal x_1, x_2, \dots, x_M . Por lo tanto, una técnica de mezcla descendente se utiliza aplicando el dispositivo de mezcla descendente 407 que ecualiza la señal suma 411 de modo que la potencia de componentes de la señal en la señal suma 411 sea aproximadamente la misma que la potencia correspondiente en todos los canales de audio de entrada x_1, x_2, \dots, x_M de la señal de audio multicanal 401. Los canales de audio de entrada x_1, x_2, \dots, x_M se descomponen en varias sub-bandas. Una de dichas sub-bandas se indica por $x_1[b]$ (conviene señalar que para simplicidad notacional no se utiliza ningún índice de sub-banda). Un procesamiento similar se aplica independientemente a todas las sub-bandas, normalmente las señales de sub-bandas se muestran en forma descendente. Las señales de cada sub-banda de cada canal de entrada se añaden y luego se multiplican con un factor de normalización de potencia.

Dada la señal suma 411, el generador de parámetros 405 sintetiza una señal de audio estéreo o multicanal 415 tal ICTD ICLD y/o ICC que se aproximan a las pistas correspondientes de la señal de audio multicanal original 401.

Cuando se consideran las respuestas de impulsos de sala binaural (BRIRs) de una sola fuente, existe una relación entre la anchura del evento de auditorio y la envolvente del oyente y el IC estimado en las partes iniciales y finales de las respuestas de impulsos de sala binaural. Sin embargo, la relación entre IC o ICC y estas propiedades para señales generales y no solamente las BRIRs no es simple. Las señales de audio multicanal y de estéreo suelen contener una mezcla compleja de señales origen simultáneamente activas superpuestas por componentes de señal reflejadas que dan lugar al registro en espacios cerrados o añadidos por el técnico de grabación para crear artificialmente una impresión espacial. Diferentes señales origen del sonido y sus reflexiones ocupan diferentes zonas en el plano del tiempo-frecuencia. Lo que antecede se refleja por ICTD, ICLD e ICC que varían como una función del tiempo y de la frecuencia. En este caso, la relación entre los valores instantáneos de ICTD, ICLD e ICC y las directrices de los eventos de auditorio y la impresión espacial no tiene carácter obvio. La estrategia del generador de parámetros 405 es sintetizar a ciegas estas pistas de modo que se aproximan a las pistas correspondientes de la señal de audio original.

En una forma de puesta en práctica, el codificador de audio paramétrica 400 utiliza bancos de filtros con sub-bandas de ancho de banda iguales a dos veces el ancho de banda rectangular equivalente. Un oyente informal declaró que la calidad de audio de BCC no mejoró notablemente cuando se eligió una más alta resolución de frecuencia. Una más baja resolución de frecuencia es favorable puesto que da lugar a menos valores de ICTD, ICLD e ICC que necesitan transmitirse al decodificador y de este modo, en una tasa binaria más baja. Con respecto a la resolución temporal, ICTD, ICLD e ICC se consideran en intervalos temporales periódicos. En una forma de puesta en práctica, ICTD, ICLD e ICC se consideran sobre cada 4 a 16 ms. Conviene señalar que a no ser que las pistas se consideren a intervalos temporales muy cortos, el efecto de prioridad no está directamente considerado.

La diferencia frecuentemente conseguida perceptualmente de pequeña magnitud entre la señal de referencia y la señal sintetizada implica que las pistas relacionadas con una amplia gama de atributos de imágenes espaciales de auditorio se consideran implícitamente sintetizando ICTD, ICLD e ICC en intervalos temporales periódicos. La tasa binaria requerida para la transmisión de estas pistas espaciales es solamente de unos pocos kb/s y por ello, el codificador de audio paramétrica 400 es capaz de transmitir señales de audio estéreo y de multicanal en tasas binarias próximas a lo que se requiere para un canal de audio único. Las Figuras 1 y 2 ilustran un método en el que se estima ICTD como el parámetro de codificación 415.

El codificador de audio paramétrico 400 comprende el generador de señales de mezcla descendente 407 para la superposición de al menos dos de las señales de canal de audio de la señal de audio multicanal 401 para obtener la señal de mezcla descendente 411, el codificador de audio 409, en particular, un codificador monoaural, para codificar la señal de mezcla descendente 411 para obtener la señal de audio codificada 413 y el combinador 417 para combinar la señal de audio codificada 413 con un parámetro de codificación correspondiente 415.

El codificador de audio paramétrico 400 genera el parámetro de codificación 415 para una señal de canal de audio de la pluralidad de señales de canal de audio indicadas como x_1, x_2, \dots, x_M de la señal de audio multicanal 401. Cada una de las señales de canal de audio x_1, x_2, \dots, x_M puede ser una señal digital que comprende valores de señal de canal de audio digital indicado como $x_1[n], x_2[n], \dots, x_M[n]$.

Una señal de canal de audio, a modo de ejemplo, para la que el codificador de audio paramétrico 400 genera el parámetro de codificación 415 es la primera señal de canal de audio x_1 con valores de señal $x_1[n]$. El generador de parámetros 405 determina el ITD del parámetro de codificación a partir de los valores de señales de canal de audio $x_1[n]$ de la primera señal de audio x_1 y a partir de los valores de señales de audio de referencia $x_2[n]$ de una señal de audio de referencia x_2 .

Una señal de canal de audio que se utiliza como una señal de audio de referencia es la segunda señal de canal de audio x_2 , a modo de ejemplo. De forma similar, cualquiera otra de las señales de canal de audio x_1, x_2, \dots, x_M puede servir como una señal de audio de referencia. En conformidad con un primer aspecto, la señal de audio de referencia es otra señal de canal de audio de las señales de canal de audio que no son iguales a la señal de canal de audio x_1 para la que se genera el parámetro de codificación 415. En conformidad con un segundo aspecto, la

señal de audio de referencia es una señal de audio de mezcla descendente derivada a partir de al menos dos señales de canal de audio de entre la pluralidad de señales de audio multicanal 401, p.ej., derivadas de la primera señal de canal de audio x_1 y de la segunda señal de canal de audio x_2 . En una forma de puesta en práctica, la señal de audio de referencia es la señal de mezcla descendente 411, también denominada señal suma generada por el dispositivo de mezcla descendente 407. En una forma de puesta en práctica, la señal de audio de referencia es la señal codificada 413 proporcionada por el codificador 409.

Una señal de audio de referencia, a modo de ejemplo, utilizada por el generador de parámetros 405, es la segunda señal de canal de audio x_2 con valores de señal $x_2[n]$.

El generador de parámetros 405 determina una transformada de frecuencia de los valores de señal de canal de audio $x_1[n]$ de la señal de canal de audio x_1 y una transformada de frecuencias de los valores de la señal de audio de referencia $x_2[n]$ de la señal de audio de referencia x_1 . La señal de audio de referencia es otra señal de canal de audio x_2 de la pluralidad de señales de canal de audio o una señal de audio de mezcla descendente derivada de al menos dos señales de canal de audio x_1, x_2 de la pluralidad de señales de canal de audio. El generador de parámetros 405 determina una diferencia entre canales para al menos cada sub-bandas de frecuencias de un subconjunto de sub-bandas de frecuencias. Cada diferencia entre canales indica una diferencia de fase IPD[b] o una diferencia de tiempo ITD[b] entre una parte de señal de banda limitada de la señal de canal de audio y una parte de señal de banda limitada de la señal de audio referencia en la respectiva sub-banda de frecuencias a la que está asociada la diferencia entre canales.

El generador de parámetros 405 determina una primera media ITD_{mean_pos} sobre la base de los valores positivos de las diferencias entre canales IPD[b] ITD[b] y una segunda media ITD_{mean_neg} sobre la base de los valores negativos de las diferencias entre canales IPD[b] ITD[b]. El generador de parámetros 405 determina el parámetro de codificación ITD sobre la base de la primera media y de la segunda media.

Una diferencia de fase entre canales (ICPD) es una diferencia de fase media entre un par de señales. Una diferencia de nivel entre canales (ICLD) es la misma que una diferencia de niveles inter-aurales (ILD), esto es, una diferencia de niveles entre las señales de entrada de los odios izquierdo y derecho, pero definida más en general entre cualquier par de señalizaciones, p.ej., un par de señales de altavoces, un par de señales de entrada en auriculares, etc. Una coherencia entre canales o una correlación entre canales es lo mismo que una coherencia inter-aural (IC), esto es, el grado de similitud entre las señales de entrada en los oídos izquierdo y derecho, pero se define más en general entre cualquier parte de señales, p.ej., par de señales de altavoces, par de señales de entrada en auriculares, etc. Una diferencia de tiempo entre canales (ICTD) es la misma que una diferencia de tiempo inter-aural (ITD), a veces también referida como retardo inter-aural, esto es, una diferencia temporal entre las señales de entrada en los oídos izquierdo y derecho, pero definida más en general entre cualquier par de señales, p.ej., par de señales en altavoces, par de señales en la entrada de auriculares, etc. Las diferencias de niveles entre canales de sub-bandas, las diferencias de fase entre canales de sub-bandas, las coherencias entre canales de sub-bandas y las diferencias de intensidad intercanal de sub-bandas se relacionan con los parámetros anteriormente especificados con respecto al ancho de banda de sub-banda.

En una primera etapa, el generador de parámetros 405 aplicada una transformada de tiempo-frecuencia en el canal de entrada en el dominio temporal, p.ej., el primer canal de entrada x_1 y el canal de referencia en el dominio temporal p.e., el segundo canal de entrada x_2 . En el caso de reproducción en estéreo, estos son los canales izquierdo y derecho. En una forma de realización preferida, la transformada de tiempo-frecuencia es una Transformada de Fourier Rápida (FFT) o una Transformada de Fourier a Corto Plazo (STFT). En una forma de realización alternativa, la transformada de tiempo-frecuencia es un banco de filtros de modulación cosenoidal o un banco de filtros complejos.

En una segunda etapa, el generador de parámetros 405 calcula un espectro cruzado para cada contenedor de frecuencias [b] de la transformada FFT como:

$$c[b] = X_1[b]X_2^*[b],$$

en donde $c[b]$ es el espectro cruzado de un contenedor de frecuencias [b] y $X_1[b]$ y $X_2[b]$ son los coeficientes de FFT de los dos canales. El asterisco * indica una conjugación completa. Para este caso, una sub-banda b corresponde directamente a un contenedor de frecuencias [k], representando el contenedor de frecuencias [b] y [k] exactamente el mismo contenedor de frecuencias.

Como alternativa, el generador de parámetros 405 calcula el espectro cruzado por sub-banda [k] como:

$$c[b] = \sum_{k=k_b}^{k_{b+1}-1} X_1[k]X_2^*[k],$$

en donde $c[b]$ es el espectro cruzado de la sub-banda $[b]$ y $X_1[k]$ y $X_2[k]$ son los coeficientes de la transformada FFT de los dos canales, a modo de ejemplo, el canal izquierdo y derecho en caso de reproducción en estéreo. El asterisco $*$ indica una conjugación completa. k_b es el contenedor inicial de la sub-banda $[b]$.

5 El espectro cruzado puede ser una versión suavizada, que se calcula mediante la ecuación siguiente.

$$c_{sm}[b,i] = SMW_1 * c_{sm}[b,i-1] + (1 - SMW_1) * c[b]$$

en donde SMW_1 es el factor de alisado, i es el índice de tramas.

10

Las diferencias de fase entre canales (IPDs) se calculan por sub-bandas sobre la base de espectro cruzado como:

$$IPD[b] = \angle c[b]$$

15 en donde la operación \angle es el operador de argumento para calcular el ángulo de $c[b]$. Conviene señalar que en el caso de suavizado del espectro cruzado, $c_{sm}[b, i]$ se utiliza para el cálculo de IPD como:

$$IPD[b] = \angle c_{sm}[b,i]$$

20 En la tercera etapa, el generador de parámetros 405 calcula los valores de ITDs de cada contenedor de frecuencias (o sub-banda) sobre la base de IPDs.

$$ITD[b] = \frac{IPD[b]N}{\pi b}$$

25 en donde N es el número del contenedor de la transformada FFT.

En la cuarta etapa, el generador de parámetros 405 realiza el conteo de los valores positivos y negativos de ITD. La media y la desviación estándar de valores de ITD positivos y negativos se basan en el signo de ITD como sigue:

$$ITD_{mean_pos} = \frac{\sum_{i=0}^{i=M} ITD(i)}{Nb_{pos}} \text{ en donde } ITD(i) \geq 0$$

$$ITD_{mean_neg} = \frac{\sum_{i=0}^{i=M} ITD(i)}{Nb_{neg}} \text{ en donde } ITD(i) < 0$$

$$ITD_{std_pos} = \sqrt{\frac{\sum_{i=0}^{i=M} (ITD(i) - ITD_{mean_pos})^2}{Nb_{pos}}} \text{ en donde } ITD(i) \geq 0$$

$$ITD_{std_neg} = \sqrt{\frac{\sum_{i=0}^{i=M} (ITD(i) - ITD_{mean_neg})^2}{Nb_{neg}}} \text{ en donde } ITD(i) < 0$$

30

en donde Nb_{pos} y Nb_{neg} son los números de ITD positivos y negativos respectivamente. M es el número total de ITDs que son objeto de extracción.

35 En la quinta etapa, el generador de parámetros 405 selecciona ITD de entre los valores de ITDs positivos y negativos sobre la base de la media y de la desviación estándar. El algoritmo de selección se representa en la Figura 3.

40 En una forma de puesta en práctica, el generador de parámetros 405 comprende:

un dispositivo de transformación de frecuencia tal como un dispositivo de transformada de Fourier, para determinar una transformada de frecuencia ($X_1[k]$) de los valores de señal de canal de audio ($x_1[n]$) de la señal de canal de audio (x_1) y para determinar una transformada de frecuencias ($X_2[k]$) de los valores de la señal de audio de referencia ($x_2[n]$) de una señal de audio de referencia (x_2), en donde la señal de audio de referencia es otra señal

de canal de audio (x_2) de la pluralidad de señales de canal de audio (x_1, x_2) o una señal de audio de mezcla descendente derivada a partir de al menos dos señales de canal de audio (x_1, x_2) de la pluralidad de señales de canal de audio;

5 un dispositivo de determinación de diferencias entre canales para determinar las diferencias entre canales (IPD[b], ITD[b]) para al menos cada sub-banda de frecuencias (b) de un subconjunto de sub-bandas de frecuencias, indicando cada diferencia entre canales una diferencia de fase (IPD[b]) o diferencia temporal (ITD[b]) entre una parte de señal de banda limitada de la señal de canal de audio y una parte de señal de banda limitada de la señal de audio de referencia en la respectiva sub-banda de frecuencias (b), con la que está asociada la diferencia entre canales;

10 un dispositivo de determinación de medias para determinar una primera media (ITD_{mean_pos}) sobre la base en valores positivos de las diferencias entre canales (IPD[b], ITD[b]) y para determinar una segunda media (ITD_{mean_neg}) basada en los valores negativos de las diferencias entre canales (IPD[b], ITD[b]) y

15 un dispositivo de determinación de parámetros de codificación para determinar el parámetro de codificación (ITD) sobre la base de la primera media y de la segunda media.

20 La Figura 5 ilustra un diagrama de bloques de un decodificador de audio paramétrico 500 en conformidad con una forma de puesta en práctica. El decodificador de audio paramétrico 500 recibe un flujo de bits 503 retransmitido a través de un canal de comunicaciones como señal de entrada y proporciona una señal de audio multicanal decodificada 501 como señal de salida. El decodificador de audio paramétrico 500 comprende un decodificador de flujo de bits 517 acoplado al flujo de bits 503 para decodificar el flujo de bits 503 en un parámetro de codificación 515 y una señal codificada 513, un decodificador 509 acoplado al decodificador de flujo de bits 517 para generar una señal suma 511 a partir de la señal codificada 513, un dispositivo de resolución de parámetros 505 acoplado al decodificador de flujos de bits 517 para la resolución de un parámetro 521 a partir del parámetro de codificación 515 y un sintetizador 505 acoplado al dispositivo de resolución de parámetros 505 y el decodificador 509 para sintetizar la señal de audio multicanal decodificada 501 a partir del parámetro 521 y de la señal suma 511.

30 El decodificador de audio paramétrico 500 genera los canales de salida la señal de audio multicanal 501 tal como ICTD, ICLD y/o ICC entre los canales que se aproximan a los de la señal de audio multicanal original. El sistema descrito es capaz de representar señales de audio multicanales en una tasa binaria solamente algo superior a la que se requiere para representar una señal de audio monoaural. Esto es así porque los valores de ICTD, ICLD e ICC estimados entre un par de canales contienen aproximadamente dos órdenes de magnitud de menos información que una forma de onda de audio. No solamente una tasa binaria baja sino también el aspecto de la compatibilidad retrospectiva es de interés. La señal suma transmitida corresponde a una conversación descendente monoaural de la señal estéreo o multicanal.

40 La Figura 6 ilustra un diagrama de bloques de un codificador de audio estéreo paramétrico 601 y un decodificador 603 en conformidad con una forma de puesta en práctica. El codificador de audio estéreo paramétrico 601 corresponde al codificador de audio paramétrico 400 según se describe con respecto a la Figura 4, pero la señal de audio multicanal 401 es una señal de audio estéreo con un canal de audio izquierdo 605 y un canal de audio izquierdo 607.

45 El codificador de audio estéreo paramétrico 601 recibe la señal de audio estéreo 605, 607 como señal de entrada y proporciona un flujo de bits como señal de salida 609. El codificador de audio estéreo paramétrico 601 comprende un generador de parámetros 611 acoplado a la señal de audio estéreo 605, 607 para generar parámetros espaciales 613, un generador de señales de mezcla descendente 615 acoplado a la señal de audio estéreo 605, 607 para generar una señal de mezcla descendente 617 o señal suma, un codificador monoaural 619 acoplado al generador de señales de mezcla descendente 615 para codificar la señal de mezcla descendente 617 para proporcionar una señal de audio codificada 621 y un combinador de flujos de bits 623 acoplado al generador de parámetros 611 y el codificador monoaural 619 para combinar el parámetro de codificación 613 y la señal de audio codificada 621 a un flujo de bits para proporcionar la señal de salida 609. En el generador de parámetros 611 los parámetros espaciales 613 se extraen y cuantifican antes de que sean multiplexados en el flujo de bits.

55 El decodificador de audio estéreo paramétrico 603 recibe el flujo de bits, esto es, la señal de salida 609 del codificador de audio estéreo paramétrico 601 transmitida a través de un canal de comunicaciones, como una señal de entrada y proporciona una señal de audio estéreo decodificada con el canal izquierdo 625 y el canal derecho 627 como la señal de salida. El decodificador de audio estéreo paramétrico 603 comprende un decodificador de flujos de bits 629 acoplado al flujo de recibido 609 para decodificar el flujo de bits 609 en parámetros de codificación 631 y una señal codificada 633, un decodificador monoaural 635 acoplado al decodificador de flujos de bits 629 para generar una señal suma 637 a partir de la señal codificada 633, un dispositivo de resolución de parámetros espaciales 639 acoplado al decodificador de flujos de bits 629 para la resolución de parámetros espaciales 641 a partir de los parámetros de codificación 631 y un sintetizador 643 acoplado al dispositivo de resolución de parámetros espaciales 639 y el decodificador monoaural 635 para sintetizar la señal de audio estéreo decodificada 625, 627 a partir de los parámetros espaciales 641 y la señal suma 637.

El procesamiento en el decodificador de audio estéreo paramétrico 603 es capaz de introducir retardos y modificar el nivel de las señales de audio de forma adaptativa en tiempo y frecuencia para generar los parámetros espaciales 631, p.ej., diferencias temporales entre canales (ICTDs) y diferencias de niveles entre canales (ICLDs). Además, el decodificador de audio estéreo paramétrico 603 realiza un filtrado adaptativo de tiempos de forma eficiente para síntesis de coherencia entre canales (ICC). En una forma de puesta en práctica, el codificador estéreo paramétrico utiliza un banco de filtros basado en la transformada de Fourier (STFT) para una puesta en práctica eficiente de los sistemas de codificación de pistas binaurales (BCC) con baja complejidad de cálculo informático. El procesamiento en el codificador de audio estéreo paramétrico 601 tiene una baja complejidad de cálculo informático y un bajo retardo, lo que hace la codificación de audio estéreo paramétrica adecuada para una puesta en práctica asequible en microprocesadores o procesadores de señales digitales para aplicaciones en tiempo real.

El generador de parámetros 611 ilustrado en la Figura 6 es funcionalmente el mismo que el generador de parámetros correspondiente 405 descrito con respecto a la Figura 4, con la excepción de que la cuantización y codificación de las pistas espaciales ha sido añadida. La señal suma 617 está codificada con un codificador de audio monoaural convencional 619. En una forma de puesta en práctica, el codificador de audio estéreo paramétrico 601 utiliza una transformada de tiempo-frecuencia basada en STFT para transformar la señal de canal de audio estéreo 605, 607 en el dominio de la frecuencia. La transformada STFT aplica una transformada de Fourier discreta (DFT) a partes dispuestas en ventanas operativas de una señal de entrada $x(n)$. Una trama de señal de N muestras se multiplica por una ventana de longitud W antes de que se aplique una transformada DFT de N puntos. Las ventanas adyacentes se solapan y se desplazan en $W/2$ muestras. La ventana se selecciona de modo que las ventanas solapantes se añadan hasta un valor constante de 1. Por lo tanto, para la transformada inversa no hay necesidad de una disposición en ventana adicional. Una transformada DFT inversa simple de tamaño N con avance temporal de tramas sucesivas de $W/2$ muestras se utiliza en el decodificador 603. Si no se modifica el espectro, se consigue una reconstrucción perfecta mediante solapamiento/adición.

Puesto que la resolución espectral uniforme de la transformada STFT no está bien adaptada a la percepción humana, los coeficientes espectrales uniformemente espaciados a la salida de la transformada STFT se agrupan en B particiones no solapantes con ancho de banda mejor adaptados a la percepción. Una partición conceptual corresponde a una "sub-banda" en conformidad con la descripción con respecto a la Figura 4. En una forma de puesta en práctica alternativa, el codificador de audio estéreo paramétrico 601 utiliza un banco de filtros no uniforme para transformar la señal de canal de audio estéreo 605, 607 en el dominio de la frecuencia.

En una forma de presentación práctica, el dispositivo de mezcla descendente 315 determina los coeficientes espectrales de una partición b o de una sub-banda b de la señal suma ecualiza $S_m(k)$ 617 mediante

$$S_m(k) = e_b(k) \sum_{c=1}^C X_{c,m}(k),$$

en donde $X_{c,m}(k)$ son los espectros de los canales de audio de entrada 605, 607 y $e_b(k)$ es un factor de ganancia calculado como

$$e_b(k) = \sqrt{\frac{\sum_{c=1}^C p_{\tilde{x}_{c,b}}(k)}{p_{\tilde{x}_b}(k)}},$$

con estimaciones de potencias de particiones,

$$p_{\tilde{x}_{c,b}}(k) = \sum_{m=A_{b-1}}^{A_b-1} |X_{c,m}(k)|^2$$

$$p_{\tilde{x}_b}(k) = \sum_{m=A_{b-1}}^{A_b-1} \left| \sum_{c=1}^C X_{c,m}(k) \right|^2.$$

Para impedir la presencia de artefactos que resultan de grandes factores de ganancia cuando la atenuación de la suma de las señales de sub-bandas es importante, los factores de ganancia $e_b(k)$ están limitados a 6 DB, esto es, $e_b(k) \leq 2$.

A partir de lo que antecede será evidente para los expertos en esta técnica que se proporciona una diversidad de

métodos, sistemas, programas informáticos en soportes de registro y dispositivos similares.

5 La presente invención soporta también un producto de programa informático que incluye un código ejecutable por ordenador o instrucciones ejecutables por ordenador que, cuando se ejecutan, hacen que al menos un ordenador ejecute las etapas de realización y cálculo aquí descritas. La presente invención soporta también un sistema configurado para ejecutar las etapas de realización y cálculo aquí descritas.

10 Numerosas alternativas, modificaciones y variaciones serán evidentes para los expertos en esta técnica considerando las enseñanzas anteriores. Por supuesto, los expertos en esta técnica reconocen fácilmente que existen numerosas aplicaciones de la invención más allá de las aquí descritas. Aunque la presente invención ha sido descrita con referencia a una o más formas de realización particulares, los expertos en esta técnica reconocen que se pueden realizar numerosos cambios sin desviarse por ello del alcance de la presente invención según se define por las reivindicaciones adjuntas. Por lo tanto, ha de entenderse que dentro del alcance de las reivindicaciones adjuntas, la invención puede ponerse en práctica de otro modo al que fue aquí concretamente descrito.

15

20

REIVINDICACIONES

- 5 **1.** Un método (100) para determinar un parámetro de codificación (ITD) para una señal de canal de audio (x_1) de una pluralidad de señales de canal de audio (x_1, x_2) de una señal de audio multicanal, teniendo cada señal de canal de audio (x_1, x_2) valores de señal de canal de audio ($x_1[n], x_2[n]$), cuyo método comprende:
- determinar (101) una transformada de frecuencia ($x_1[k]$) de los valores de señal de canal de audio ($x_1[n]$) de la señal de canal de audio (x_1);
- 10 determinar (103) una transformada de frecuencia ($x_2[k]$) de los valores de señal de audio de referencia ($x_2[n]$) de una señal de audio de referencia (x_2) en donde la señal de audio de referencia es otra señal de canal de audio (x_2) de la pluralidad de señales de canal de audio o una señal de audio de mezcla descendente derivada de al menos dos señales de canal de audio (x_1, x_2) de la pluralidad de señales de canal de audio;
- 15 determinar (105) diferencias entre canales (ICD[b]) para al menos cada sub-banda de frecuencia (b) de un subconjunto de sub-bandas de frecuencia, indicando cada diferencia entre canales una diferencia de fase (IPD[b]) o una diferencia de tiempo (ITD[b]) entre una parte de señal de banda limitada de la señal de canal de audio y una parte de señal de banda limitada de la señal de audio de referencia en la sub-banda de frecuencia respectiva (b) en donde la diferencia entre canales está asociada;
- 20 determinar (107) una primera media (ITD_{mean_pos}) basada en valores positivos de las diferencias entre canales (ICD[b]) y determinar una segunda media (ITD_{mean_neg}) basada en valores negativos de las diferencias entre canales (ICD[b]); y
- 25 determinar (109) el parámetro de codificación (ITD) sobre la base de la primera media y de la segunda media.
- 2.** El método (100) según la reivindicación 1, en donde las diferencias entre canales (ICD[b]) son diferencias de fase entre canales (IPD[b]) o diferencias de tiempo entre canales (ITD[b]).
- 30 **3.** El método (100) según la reivindicación 1 o 2, que comprende, además:
- determinar una primera desviación estándar (ITD_{std_pos}) sobre la base de los valores positivos de las diferencias entre canales (ICD[b]) y determinar una segunda desviación estándar (ITD_{std_neg}) sobre la base de valores negativos de las diferencias entre canales (ICD[b]),
- 35 en donde la determinación del parámetro de codificación (ITD) está basada en la primera desviación estándar y en la segunda desviación estándar.
- 4.** El método (100) según cualquiera de las reivindicaciones 1 a 3, en donde una sub-banda de frecuencias comprende una o una pluralidad de contenedores de frecuencias (k).
- 5.** El método (100) según cualquiera de las reivindicaciones 1 a 4, en donde la determinación de diferencias entre canales (ICD[b]) para al menos cada sub-banda de frecuencias (b) de un subconjunto de sub-bandas de frecuencia comprende:
- 45 determinar un espectro cruzado ($c[k], c[b]$) como una correlación cruzada de la transformada de frecuencia ($x_1[k]$) de los valores de señal de canal de audio ($x_1[n]$) y la transformada de frecuencia ($x_2[k]$) de los valores de la señal de audio de referencia ($x_2[n]$); y
- 50 determinar diferencias de fase entre canales (IPD[b]) para cada sub-banda de frecuencias (b) sobre la base del espectro cruzado ($c[b]$).
- 6.** El método (100) según la reivindicación 5, en donde la diferencia de fase entre canales (IPD[b]) de un contenedor de frecuencias (b) o de una sub-banda de frecuencias (b) se determina como un ángulo del espectro cruzado ($c[b]$).
- 55 **7.** El método (100) según la reivindicación 5 o 6 que comprende, además:
- determinar diferencias de tiempo entre canales (ITD[b]) basadas en las diferencias de fase entre canales (IPD[b]); en donde
- 60 la determinación de la primera media (ITD_{mean_pos}) está basada en valores positivos de las diferencias de tiempo entre canales (ITD[b]) y la determinación de la segunda media (ITD_{mean_neg}) está basada en valores negativos de las diferencias de tiempo entre canales (ITD[b]).
- 65 **8.** El método (100) según la reivindicación 6 o 7, en donde las diferencias de tiempo entre canales (ITD[b]) de una sub-banda de frecuencias (b) se determina como una función de la diferencia de fase entre canales IPD[b],

dependiendo dicha función de un número (N) de contenedores de frecuencias y del contenedor de frecuencia (k) del índice de sub-bandas de frecuencias (b).

5 **9.** El método (100) según la reivindicación 7 o la reivindicación 8, en donde la determinación (109) del parámetro de codificación (ITD) comprende:

contar un primer número ($N_{b_{pos}}$) de diferencias de tiempo entre canales positivas (ITD[b]) y un segundo número ($N_{b_{neg}}$) de diferencias de tiempo entre canales negativas (ITD[b]) sobre el número (M) de sub-bandas de frecuencias (b) incluidas dentro del subconjunto de sub-bandas de frecuencias (b).

10 **10.** El método (100) según la reivindicación 9, en donde el parámetro de codificación (ITD) se determina sobre la base de una comparación entre el primer número ($N_{b_{pos}}$) de diferencias de tiempo entre canales positivas (ITD[b]) y el segundo número ($N_{b_{neg}}$) de diferencias de tiempo entre canales negativas (ITD[b]).

15 **11.** El método (100) según la reivindicación 10, en donde el parámetro de codificación (ITD) se determina sobre la base de una comparación entre la primera desviación estándar (ITD_{std_pos}) y la segunda desviación estándar (ITD_{std_neg}).

20 **12.** El método (100) según la reivindicación 10 o la reivindicación 11, en donde el parámetro de codificación (ITD) se determina sobre la base de una comparación entre el primer número ($N_{b_{pos}}$) de diferencias entre canales positivas (ITD[b]) y el segundo número ($N_{b_{neg}}$) de diferencias de tiempo entre canales negativas (ITD[b]) multiplicado por un primer factor (A).

25 **13.** El método (100) según la reivindicación 12, en donde el parámetro de codificación (ITD) se determina sobre la base de una comparación entre la primera desviación estándar (ITD_{std_pos}) y la segunda desviación estándar (ITD_{std_neg}) multiplicado por un segundo factor (B).

30 **14.** Un codificador de audio multicanal (400, 601) para determinar un parámetro de codificación (ITD) para una señal de canal de audio (x_1) de una pluralidad de señales de canal de audio (x_1, x_2) de una señal de audio multicanal, teniendo cada señal de canal de audio (x_1, x_2) valores de señal de canal de audio ($x_1[n], x_2[n]$), cuyo codificador de audio multicanal comprende:

35 un dispositivo de transformación de frecuencia tal como un dispositivo de transformación de Fourier, para determinar una transformada de frecuencias ($x_1[k]$) de los valores de la señal de canal de audio ($x_1[n]$) de la señal de canal de audio (x_1) y para determinar una transformada de frecuencia ($x_2[k]$) de valores de señal de audio de referencia ($x_2[n]$) de una señal de audio de referencia (x_2), en donde la señal de audio de referencia es otra señal de canal de audio (x_2) de la pluralidad de señales de canal de audio o una señal de audio de mezcla descendente derivada de al menos dos señales de canal de audio (x_1, x_2) de la pluralidad de señales de canal de audio;

40 un dispositivo de determinación de diferencia entre canales para determinar las diferencias entre canales (IPD[b], ITD[b]) para al menos cada sub-banda de frecuencias (b) de un subconjunto de sub-bandas de frecuencia, indicando cada diferencia entre canales una diferencia de fase (IPD[b]) o una diferencia de tiempo (ITD[b]) entre una parte de señal de banda limitada de la señal de canal de audio y una parte de señal de banda limitada de la señal de audio de referencia en la sub-banda de frecuencias respectiva (b), estando la diferencia entre canales asociada a:

45 un dispositivo de determinación para determinar una primera media (ITD_{mean_pos}) sobre la base de los valores positivos de las diferencias entre canales (IPD[b], ITD[b]) y para determinar una segunda media (ITD_{mean_neg}) basada en valores negativos de las diferencias entre canales (IPD[b], ITD[b]); y

50 un dispositivo de determinación de parámetros de codificación para determinar el parámetro de codificación (ITD) sobre la base de la primera media y de la segunda media.

55 **15.** Un programa informático que tiene un código de programa adaptado para ejecutar el método (100) según una de las reivindicaciones 1 a 13 cuando se ejecuta en un ordenador.

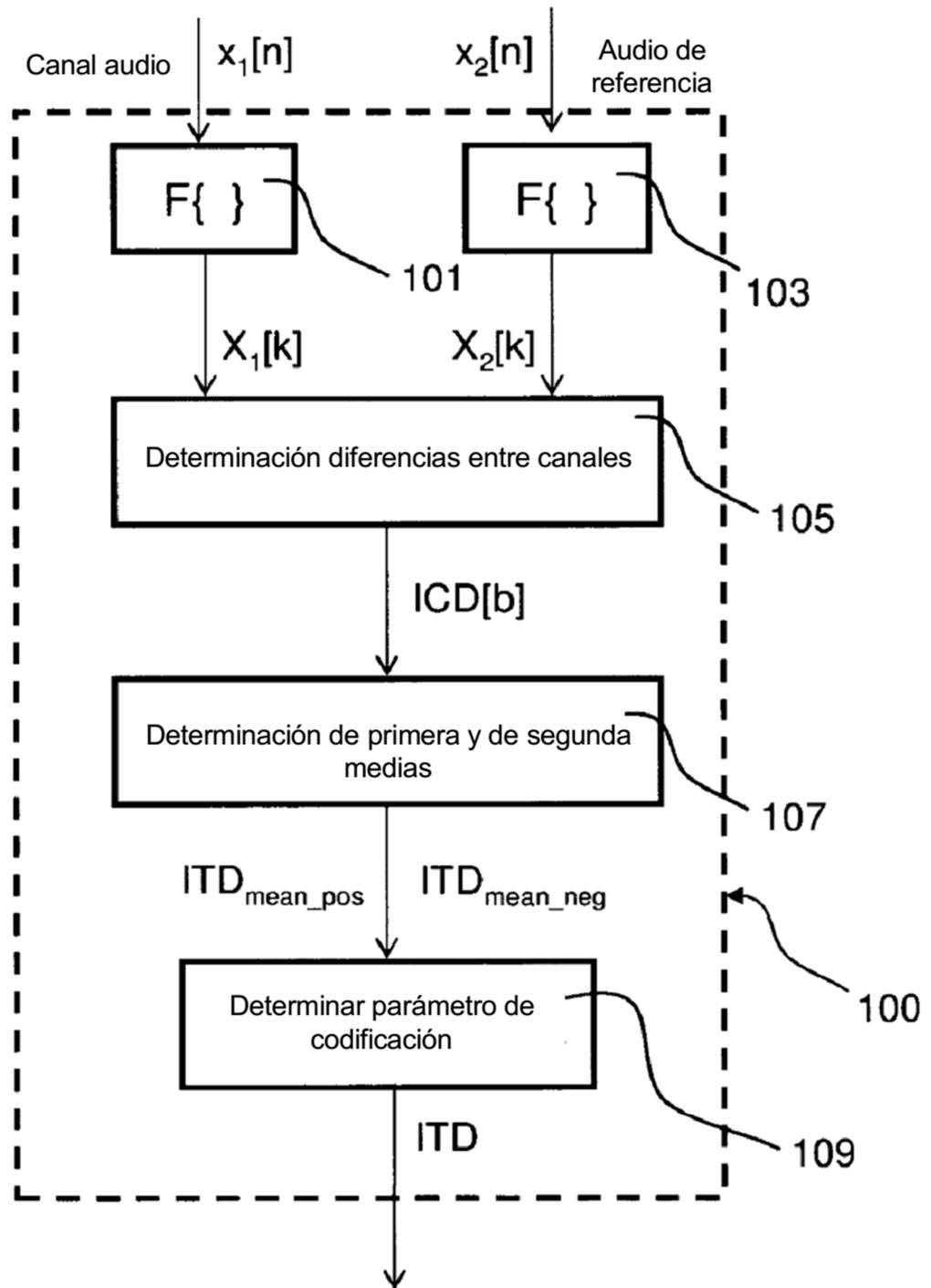


Fig. 1

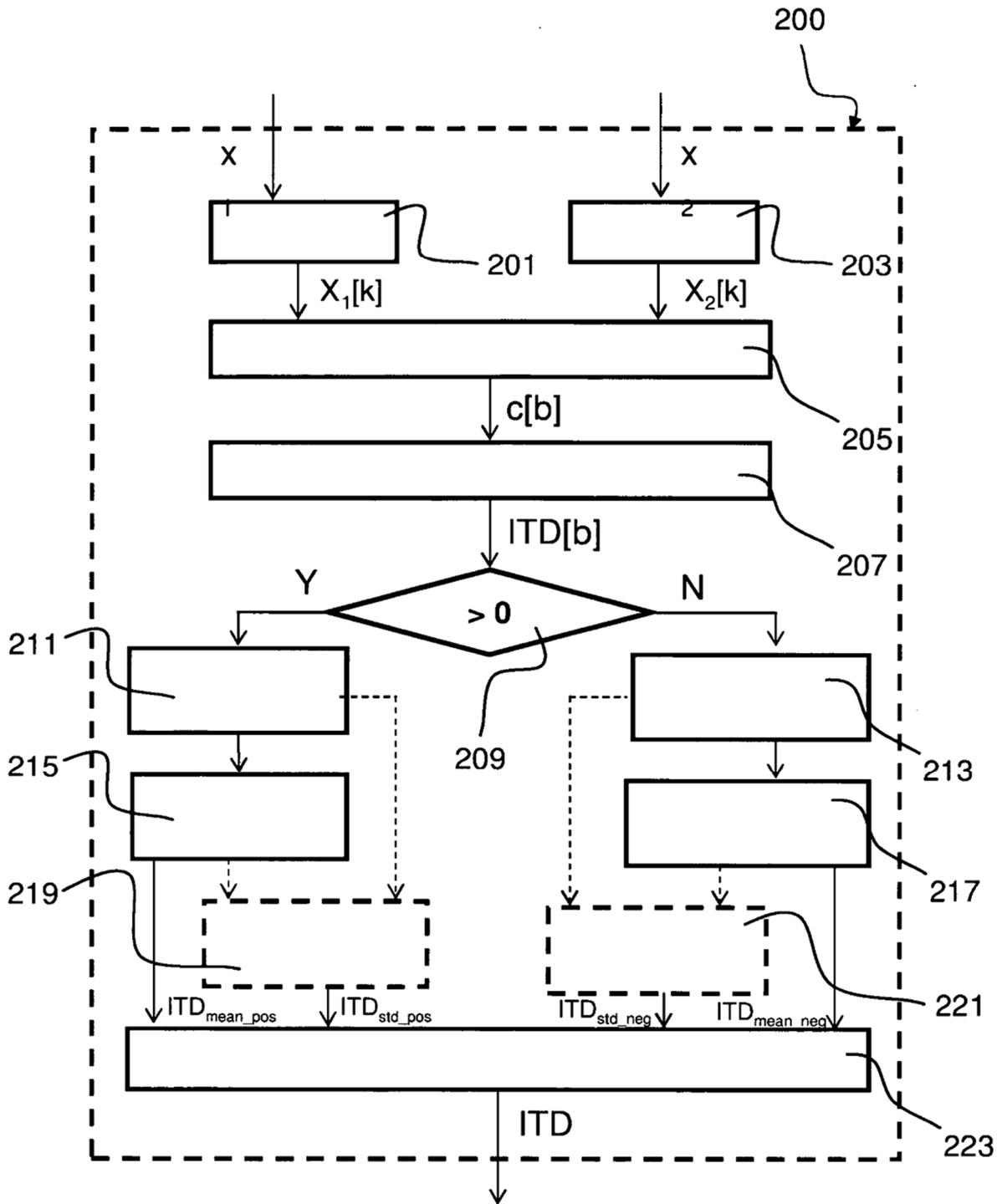


Fig. 2

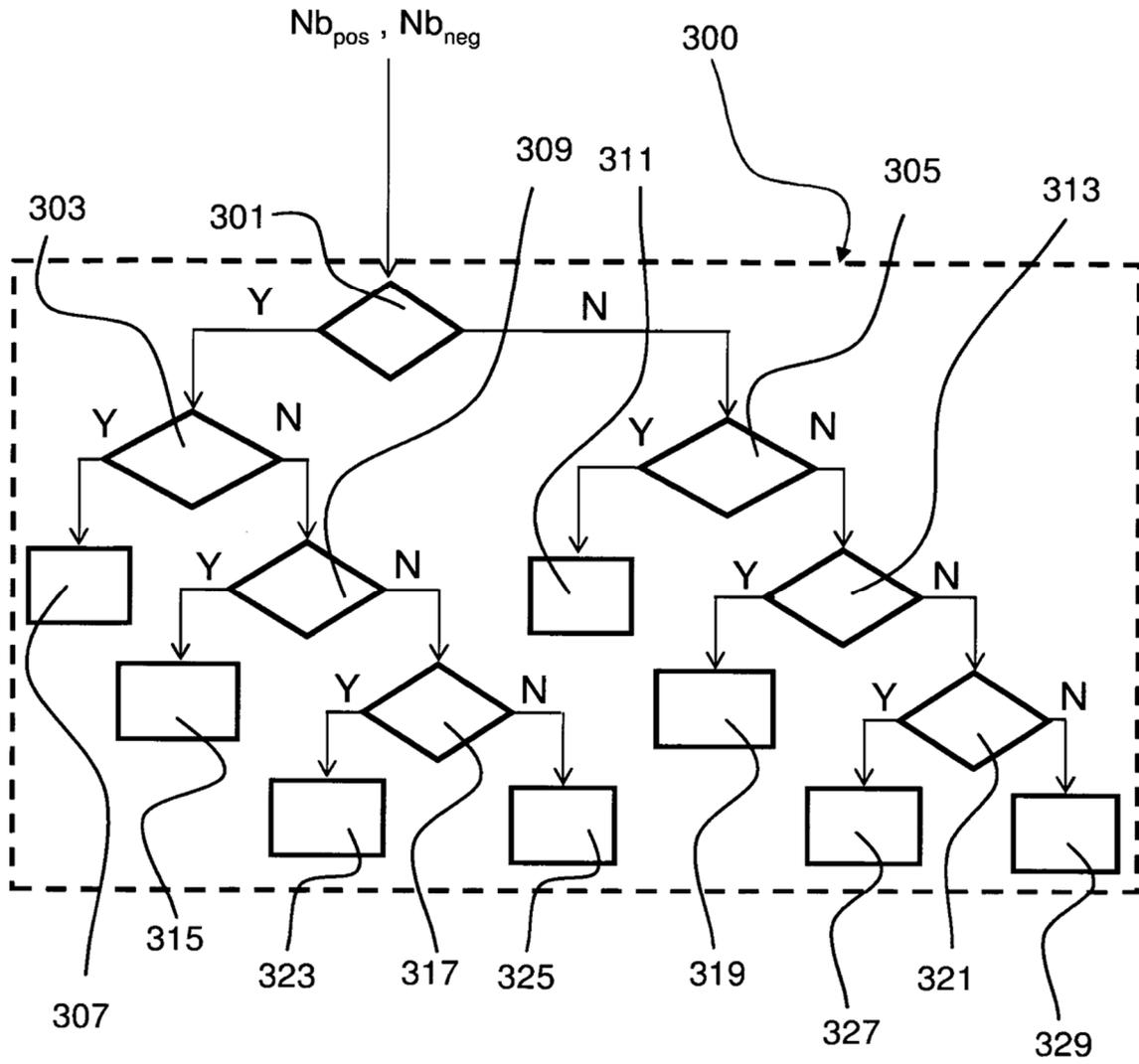


Fig. 3

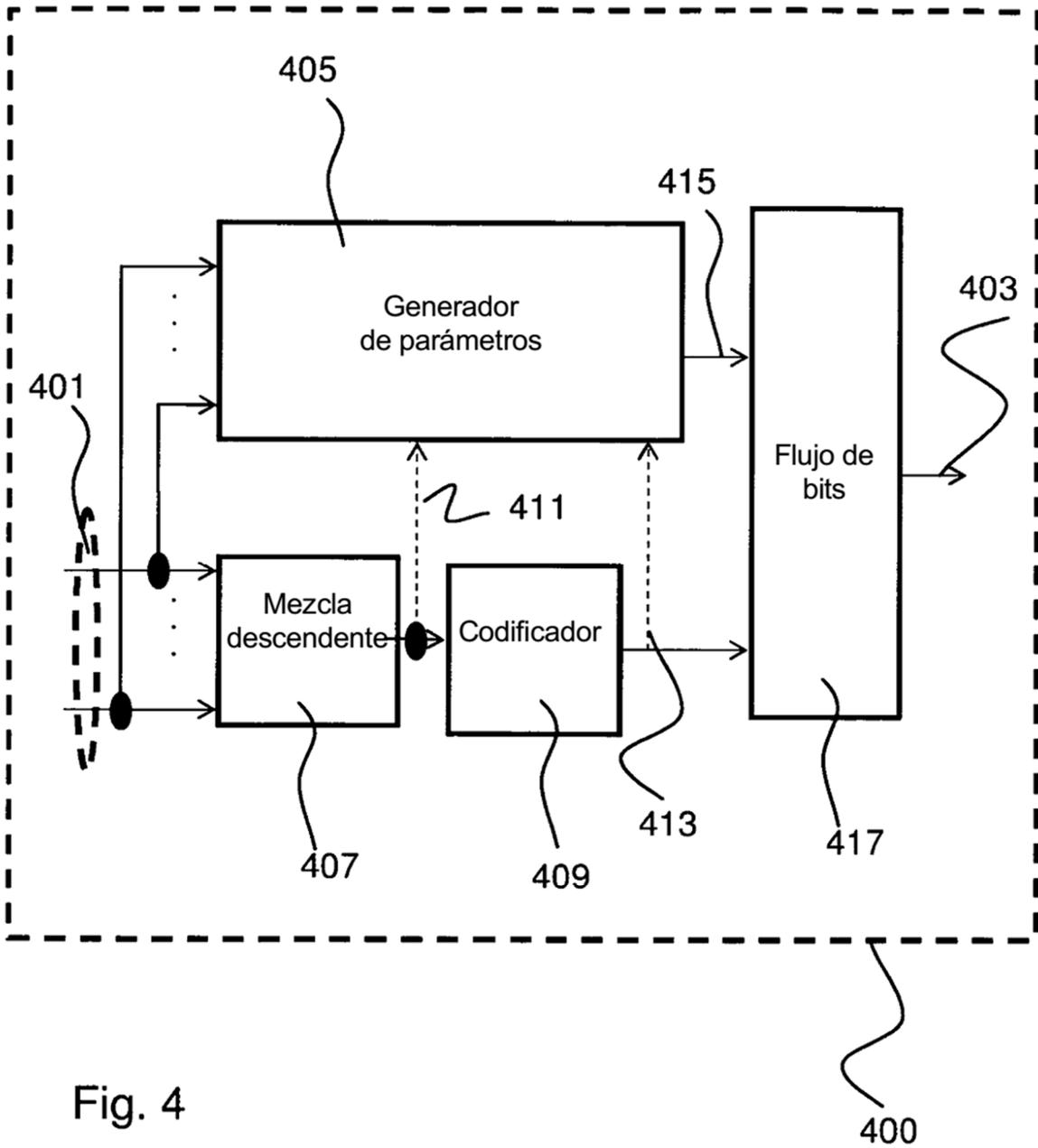


Fig. 4

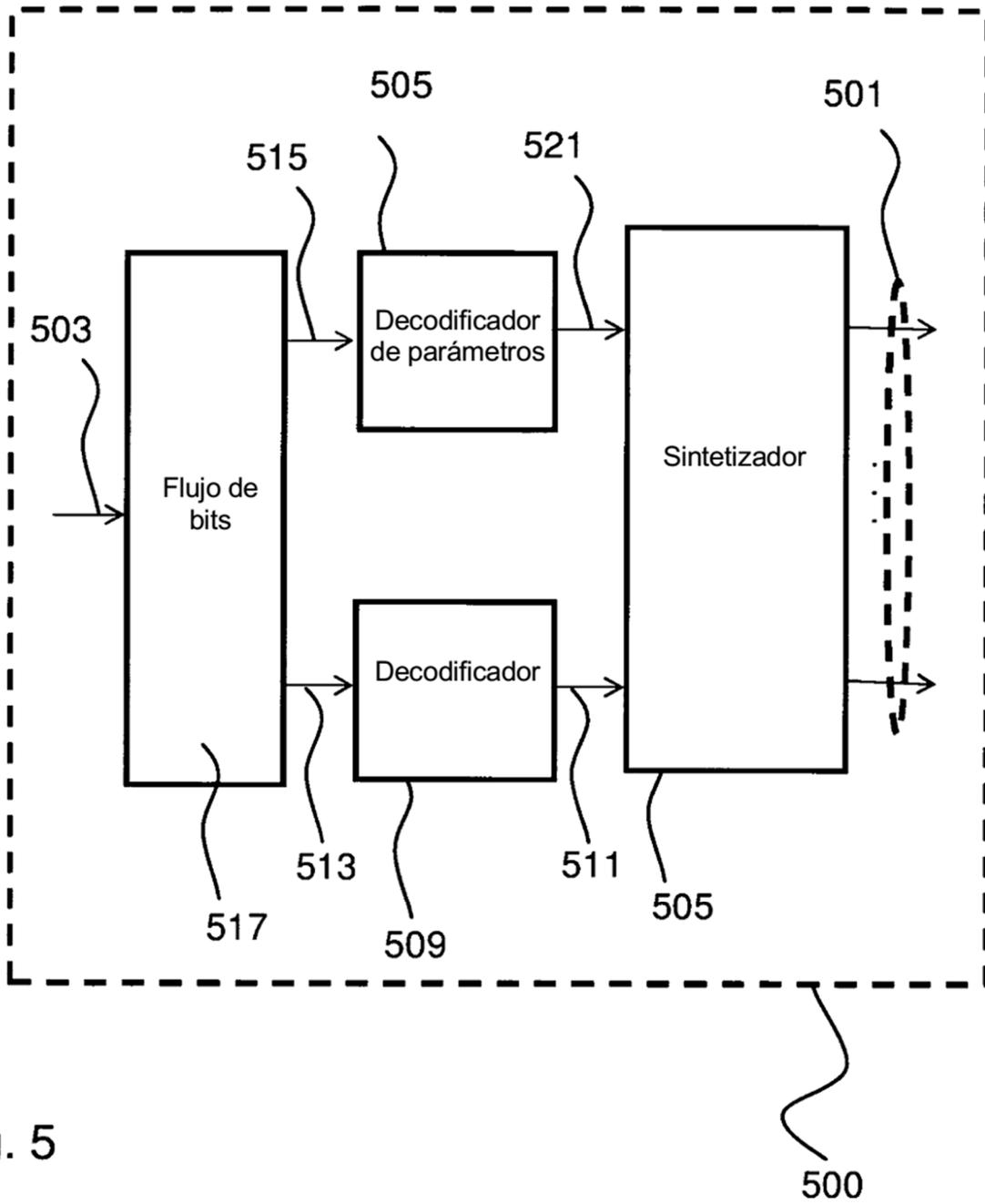


Fig. 5

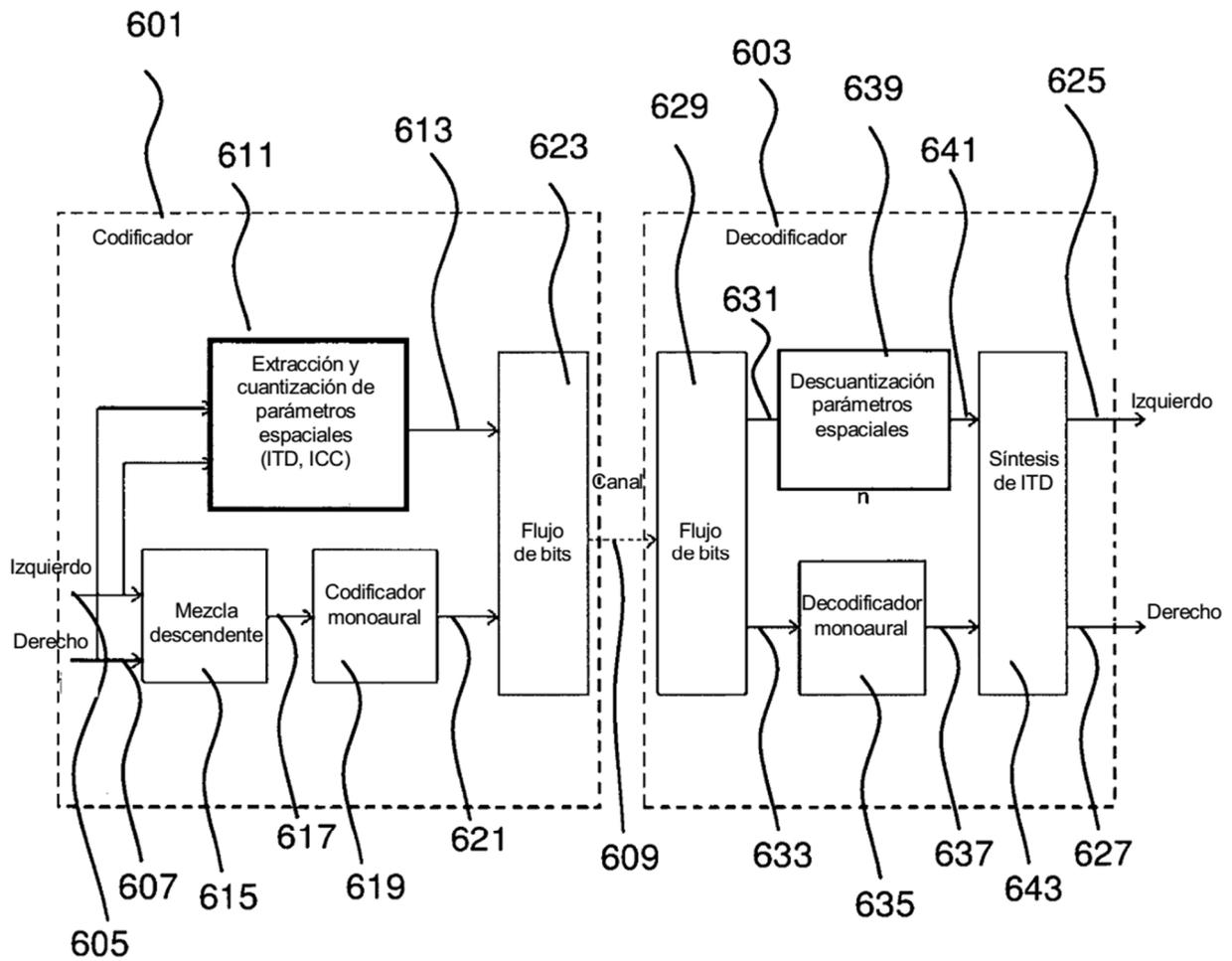


Fig. 6

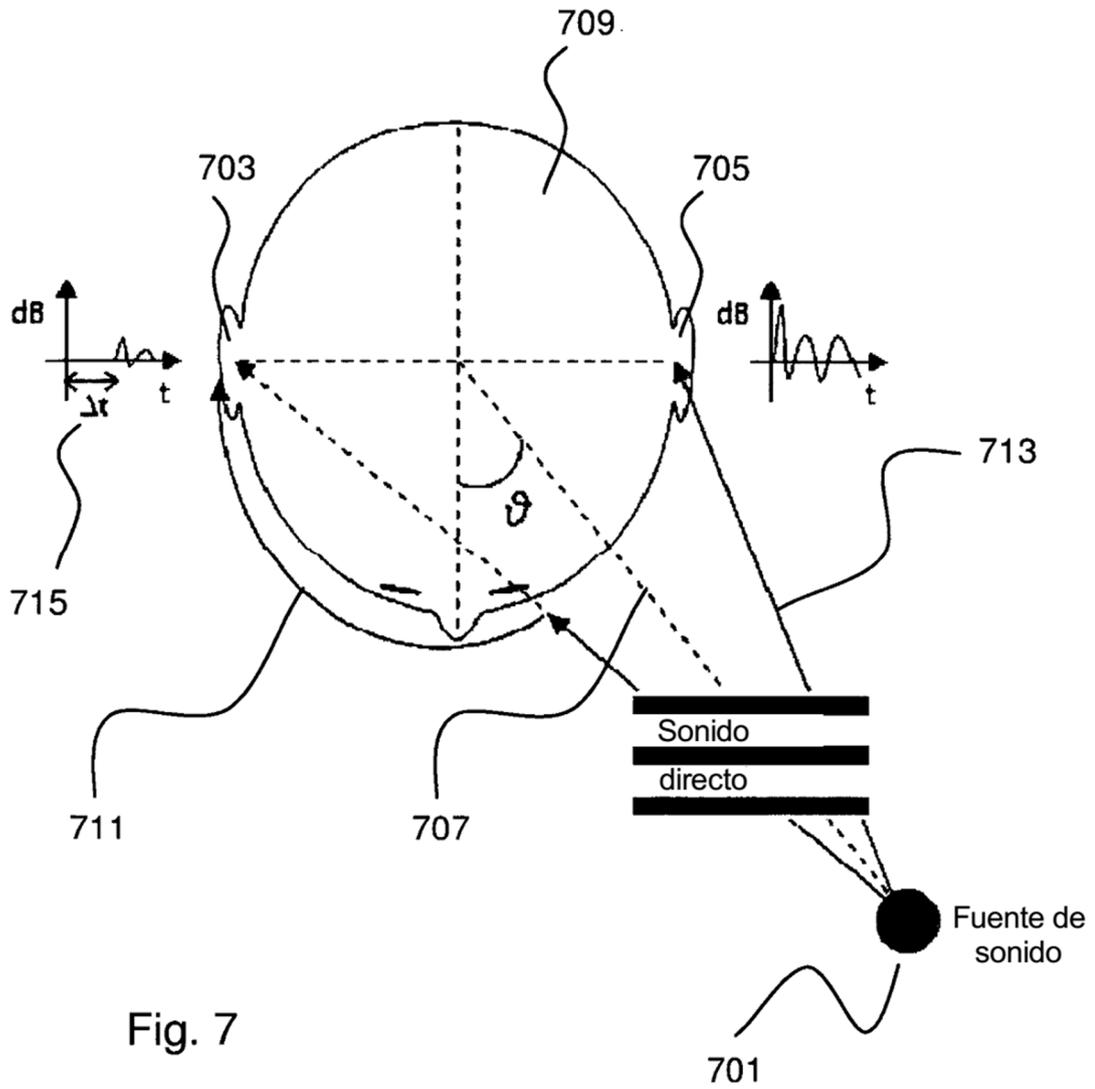


Fig. 7