

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **3 012 258**

51 Int. Cl.:

G10L 19/008 (2013.01)

G10L 19/02 (2013.01)

G06F 16/683 (2009.01)

H04S 3/00 (2006.01)

H04S 7/00 (2006.01)

G10L 25/51 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **26.03.2020** **PCT/FI2020/050194**

87 Fecha y número de publicación internacional: **01.10.2020** **WO20193865**

96 Fecha de presentación y número de la solicitud europea: **26.03.2020** **E 20778687 (2)**

97 Fecha y número de publicación de la concesión europea: **01.01.2025** **EP 3948861**

54 Título: **Determinación de la significación de parámetros de audio espacial y codificación asociada**

30 Prioridad:

28.03.2019 GB 201904303

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

09.04.2025

73 Titular/es:

NOKIA TECHNOLOGIES OY (100.00%)
Karakaari 7
02610 Espoo, FI

72 Inventor/es:

LAITINEN, MIKKO-VILLE y
VASILACHE, ADRIANA

74 Agente/Representante:

DEL VALLE VALIENTE, Sonia

ES 3 012 258 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

Determinación de la significación de parámetros de audio espacial y codificación asociada

5 **Campo**

La presente solicitud se refiere a aparatos y métodos para codificación de parámetros relacionados con el campo de sonido, pero no exclusivamente para parámetros relacionados con la dirección de dominio de tiempo-frecuencia que codifican un codificador y decodificador de audio.

10

Antecedentes

El procesamiento de audio espacial paramétrico es un campo de procesamiento de señales de audio donde el aspecto espacial del sonido se describe usando un conjunto de parámetros. Por ejemplo, en la captura de audio espacial paramétrica a partir de matrices de micrófonos, una elección típica y efectiva es estimar, a partir de las señales de matriz de micrófonos, un conjunto de parámetros tales como direcciones del sonido en bandas de frecuencia, y los coeficientes entre las partes direccionales y no direccionales del sonido capturado en las bandas de frecuencia. Se sabe que estos parámetros describen bien las propiedades espaciales perceptuales del sonido captado en la posición de la matriz de micrófonos. Estos parámetros pueden usarse en la síntesis del sonido espacial, por consiguiente, para auriculares de manera binaural, para altavoces o para otros formatos, tales como ambisónico.

20

Por tanto, las direcciones y las relaciones de energía directa con respecto a total en bandas de frecuencia son de este modo una parametrización que es particularmente eficaz para la captación de audio espacial.

25

También puede usarse un conjunto de parámetros que consiste en un parámetro de dirección en bandas de frecuencia y un parámetro de razón de energía en bandas de frecuencia (que indica la direccionalidad del sonido) como metadatos espaciales (que también pueden incluir otros parámetros como la coherencia envolvente, coherencia de propagación, número de direcciones, distancia, etc.) para un códec de audio. Por ejemplo, estos parámetros pueden estimarse a partir de señales de audio captadas en matriz de micrófonos y, por ejemplo, se puede generar una señal estéreo o mono a partir de las señales de matriz de micrófonos que se transportan con los metadatos espaciales. La señal estéreo podría codificarse, por ejemplo, con un codificador AAC y la señal mono podría codificarse con un codificador EVS. Un decodificador puede decodificar las señales de audio para dar señales de PCM, y procesar el sonido en bandas de frecuencia (usando los metadatos espaciales) para obtener la salida espacial, por ejemplo, una salida binaural.

35

La solución mencionada anteriormente es particularmente adecuada para codificar sonido espacial capturado a partir de matrices de micrófonos (por ejemplo, en teléfonos móviles, cámaras de RV, matrices de micrófono independientes). Sin embargo, puede ser deseable que dicho codificador tenga también otros tipos de entrada que las señales capturadas de matriz de micrófonos, por ejemplo, señales de altavoces, señales de objeto de audio o señales ambisónicas.

40

El análisis de entradas ambisónicas (FOA) de primer orden para la extracción de metadatos espaciales se ha documentado completamente en la literatura científica relacionada con la codificación de audio direccional (DirAC) y la expansión de onda de plano armónico (Harpex). Esto es dado que existen matrices de micrófono que proporcionan directamente una señal FOA (de forma más precisa: su variante, la señal de formato B) y el análisis de dicha entrada ha sido por lo tanto un punto de estudio en el campo.

45

Una entrada adicional para el codificador también es la entrada de altavoz multicanal, como las entradas envolventes del canal 5.1 o 7.1.

50

Sin embargo, con respecto a los componentes de los metadatos espaciales, la compresión y codificación de los parámetros de coherencia tienen considerable interés para minimizar el número total de bits requeridos para representar los parámetros de audio espacial.

55

La publicación de patente EP3451331 describe un método para determinar un parámetro de diferencia de fase entre canales (IPD) para una señal de audio multicanal, donde el parámetro se determina según una de dos maneras de extracción de parámetros de IPD preestablecidas.

60

Resumen

Según un primer aspecto, se proporciona un aparato para codificar audio espacial que comprende: medios para recibir o determinar, para una o más señales de audio, parámetros de audio espacial por subbandas para proporcionar una reproducción de audio espacial, en donde los parámetros de audio espacial comprenden un valor de coherencia para cada subbanda de una pluralidad de subbandas de una trama; medios para recibir un valor de relación de energía para cada una de la pluralidad de subbandas; medios para determinar una medida de significación para los valores de coherencia de la pluralidad de subbandas de la trama; y medios para usar la medida de significación para

65

determinar si se deben codificar los valores de coherencia de la pluralidad de subbandas de la trama, en donde el aparato se caracteriza por que los medios para determinar la medida de significación para los valores de coherencia de la pluralidad de subbandas de la trama comprenden: medios para determinar una proporción de energía no direccional coherente para cada una de la pluralidad de subbandas, en donde la proporción de energía no direccional coherente para una subbanda se determina como una relación de energía no direccional para la subbanda multiplicado por el valor de coherencia para la subbanda, en donde la relación de energía no direccional está relacionada con la relación de energía en la subbanda que queda después de haber eliminado el valor de relación de energía para la subbanda; y medios para sumar, para toda la pluralidad de subbandas, la proporción determinada de energía no direccional coherente.

Los medios para determinar la medida de significación para los valores de coherencia de la pluralidad de subbandas de la trama pueden también comprender: medios para multiplicar, para cada subbanda, la proporción de energía no direccional coherente por la relación de energía no direccional; medios para sumar, para toda la pluralidad de subbandas, el múltiplo de la proporción de energía no direccional coherente y la relación de energía no direccional para cada subbanda; y medios para normalizar el múltiplo sumado de la proporción de energía no direccional coherente y la relación de energía no direccional para cada subbanda mediante una suma de la relación de energía no direccional para cada subbanda.

La medida de significación para los valores de coherencia de la pluralidad de subbandas de la trama puede ser una primera medida de significación para los valores de coherencia de la pluralidad de subbandas de la trama.

La medida de significación para los valores de coherencia de la pluralidad de subbandas de la trama puede ser una segunda medida de significación para los valores de coherencia de la pluralidad de subbandas de la trama.

Los medios para determinar la medida de significación para los valores de coherencia de la pluralidad de subbandas de la trama pueden comprender: medios para determinar la medida de significación para los valores de coherencia de la pluralidad de subbandas de la trama como el máximo de la primera medida de significación y la segunda medida de significación.

Los medios para usar la medida de significación para determinar si se deben codificar los valores de coherencia de la pluralidad de subbandas de la trama pueden comprender: medios para comparar la medida de significación con un valor umbral; y medios para codificar los valores de coherencia de la pluralidad de subbandas de la trama cuando la medida de significación supera el valor umbral.

El valor de coherencia para cada una de la pluralidad de subbandas puede ser un valor de coherencia envolvente, en donde el valor de coherencia envolvente puede determinarse basándose en una coherencia entre canales entre dos o más señales de audio de canal. El valor de coherencia envolvente para cada una de la pluralidad de subbandas puede determinarse mediante el aparato que comprende medios para: calcular una matriz de covarianza asociada a las señales de audio de dos o más canales; monitorizar una señal de audio de canal con una energía máxima determinada en función de la matriz de covarianza y un subconjunto de otras señales de audio de canal, en donde el subconjunto es un número determinado entre 1 y uno menos que el número total de señales de audio de canal con las siguientes energías más altas; y generar el parámetro de covarianza envolvente basándose en la selección del mínimo de coherencias normalizadas determinadas entre la señal de audio de canal con la mayor energía y cada una de las señales de audio de canal con las siguientes energías más altas.

El valor de relación de energía para cada una de las subbandas puede ser un valor de relación de energía directa con respecto al total para cada una de las subbandas.

La relación de energía no direccional para una subbanda se determina como uno menos la relación entre energía directa y total para la subbanda.

Según un segundo aspecto, existe un método para codificar audio espacial que comprende: recibir o determinar, para una o más señales de audio, parámetros de audio espacial por subbandas para proporcionar una reproducción de audio espacial, en donde los parámetros de audio espacial comprenden un valor de coherencia para cada subbanda de una pluralidad de subbandas de una trama; recibir un valor de relación de energía para cada una de la pluralidad de subbandas; determinar una medida de significación para los valores de coherencia de la pluralidad de subbandas de la trama; y usar la medida de significación para determinar si se deben codificar los valores de coherencia de la pluralidad de subbandas de la trama, en donde el método se caracteriza por que determinar la medida de significación para los valores de coherencia de la pluralidad de subbandas de la trama comprende: determinar una proporción de energía no direccional coherente para cada una de la pluralidad de subbandas, en donde la proporción de energía no direccional coherente para una subbanda se determina como una relación de energía no direccional para la subbanda multiplicado por el valor de coherencia para la subbanda, en donde la relación de energía no direccional está relacionada con la relación de energía en la subbanda que queda después de haber eliminado el valor de relación de energía para la subbanda; y sumar, para toda la pluralidad de subbandas, la proporción determinada de energía no direccional coherente.

El método puede comprender además recibir un valor de relación de energía para cada una de la pluralidad de subbandas, en donde los medios para determinar la medida de significación para los valores de coherencia de la pluralidad de subbandas de la trama pueden comprender: determinar una proporción de energía no direccional coherente para cada una de la pluralidad de subbandas, en donde la proporción de energía no direccional coherente para una subbanda se determina como una relación de energía no direccional para la subbanda multiplicada por el valor de coherencia para la subbanda, en donde la relación de energía no direccional está relacionada con la relación de energía en la subbanda que queda después de haber eliminado el valor de relación de energía para la subbanda; y sumar, para toda la pluralidad de subbandas, la proporción determinada de energía no direccional coherente.

Determinar la medida de significación para los valores de coherencia de la pluralidad de subbandas de la trama puede comprender además: multiplicar, para cada subbanda, la proporción de energía no direccional coherente por la relación de energía no direccional; sumar, para toda la pluralidad de subbandas, el múltiplo de la proporción de energía no direccional coherente y la relación de energía no direccional para cada subbanda; y normalizar el múltiplo sumado de la proporción de energía no direccional coherente y la relación de energía no direccional para cada subbanda mediante una suma de la relación de energía no direccional para cada subbanda.

Un programa informático que comprende instrucciones de programa para hacer que un ordenador realice el método descrito anteriormente.

Un producto de programa informático almacenado en un medio puede hacer que un aparato realice el método tal como se describe en la presente memoria.

Un dispositivo electrónico puede comprender un aparato tal como se describe en la presente memoria.

Un conjunto de chips puede comprender un aparato tal como se describe en la presente memoria.

Las realizaciones de la presente solicitud tienen como objetivo abordar problemas asociados con el estado de la técnica.

Resumen de las figuras

Para una mejor comprensión de la presente solicitud, ahora se hará referencia a modo de ejemplo a los dibujos adjuntos, en los que:

la figura 1 muestra esquemáticamente un sistema de aparato adecuado para implementar algunas realizaciones;

la figura 2 muestra esquemáticamente un sistema de códec de audio según algunas realizaciones;

la figura 3 muestra un diagrama de flujo de la operación del procesador de análisis como se muestra en la figura 2 según algunas realizaciones;

la figura 4 muestra esquemáticamente el codificador de coherencia según muestra la figura 2 según algunas realizaciones;

la figura 5 muestra un diagrama de flujo de la operación del determinador de significación de valores de coherencia según muestra la figura 4 según algunas realizaciones;

la figura 6 muestra un diagrama de flujo de la operación del determinador de codificación de valores de coherencia según muestra la figura 4 según algunas realizaciones; y

la figura 7 muestra esquemáticamente un dispositivo ilustrativo adecuado para implementar el aparato mostrado.

Realizaciones de la solicitud

A continuación se describe con más detalle un aparato adecuado y mecanismos posibles para la provisión de parámetros de metadatos derivados de análisis espacial efectivos. En la siguiente discusión, el sistema multicanal se analiza con respecto a una implementación de micrófono multicanal. Sin embargo, como se discutió anteriormente, el formato de entrada puede ser cualquier formato de entrada adecuado, tal como un altavoz multicanal, ambisónica (FOA/HOA), etc. Se entiende que en algunas realizaciones la ubicación del canal se basa en una ubicación del micrófono o es una ubicación o dirección virtual. Además, la salida del sistema de ejemplo es una disposición de altavoces multicanal. Sin embargo, se entiende que la salida puede hacerse al usuario mediante medios que no sean altavoces. Además, las señales de altavoces multicanal pueden generalizarse para ser dos o más señales de audio de reproducción.

Los metadatos consisten en al menos una dirección (elevación, acimut), relación de energía de una dirección resultante, y coherencia envolvente independiente de la dirección, para cada bloque o panel de tiempo-frecuencia (TF)

considerado, en otras palabras una subbanda de tiempo-frecuencia. Además, se puede determinar e incluir una coherencia de propagación para cada bloque de TF. Todos estos datos se codifican y transmiten (o almacenan) por el codificador para poder reconstruir la señal espacial en el decodificador.

Las tasas de transferencia operativas generales típicas del códec pueden dejar de 2 a 10 kbps para la transmisión/almacenamiento de metadatos espaciales. Sin embargo, algunas implementaciones adicionales pueden permitir hasta 20 kps para la transmisión/almacenamiento de metadatos espaciales. La codificación de los parámetros de dirección y los componentes de relación de energía se ha examinado anteriormente. Sin embargo, la codificación de los datos de coherencia requiere una mayor investigación para minimizar las tasas de transferencia operativas generales necesarias para codificar los metadatos espaciales.

El concepto como se describe a continuación es codificar los parámetros de coherencia junto con los parámetros de dirección y relación de energía para cada bloque de tiempo-frecuencia. En situaciones particulares, puede ser posible regular la codificación de los parámetros de coherencia para una trama según su significación subjetiva para un oyente. Por ejemplo, se puede determinar, para una trama de audio, que el parámetro de coherencia para cada subbanda puede no aportar significación subjetiva alguna con respecto a las subbandas envolventes. En tales casos, se puede lograr un ahorro en la tasa de transferencia operativa al no codificar los parámetros de coherencia para la trama.

En consecuencia, la invención parte de la consideración de que, en determinadas circunstancias, se produce un ahorro global en el número de bits utilizados para codificar los parámetros de coherencia. Esto se puede lograr teniendo una medida que proporcione una indicación de la significación subjetiva del valor de coherencia de cada subbanda en relación con las subbandas envolventes, y utilizando los resultados de la medida para determinar si deben codificarse los parámetros de coherencia para la trama.

En este sentido, la figura 1 ilustra un aparato y un sistema de ejemplo para implementar realizaciones de la solicitud. El sistema 100 se muestra con una parte de "análisis" 121 y una parte de "síntesis" 131. La parte de "análisis" 121 es la parte desde la recepción de las señales de altavoces de múltiples canales hasta la codificación de los metadatos y la señal de mezcla descendente; y la parte de "síntesis" 131 es la parte de una decodificación de los metadatos codificados y la señal de mezcla descendente a la presentación de la señal regenerada (por ejemplo, en forma de altavoces de múltiples canales).

La entrada al sistema 100 y la parte de análisis 121 son las señales multicanal 102. En los siguientes ejemplos, se describe una entrada de señal de canal de micrófono, aunque puede implementarse cualquier formato de entrada (o multicanal sintético) adecuado en otras realizaciones. Por ejemplo, en algunas realizaciones, el analizador espacial y el análisis espacial pueden implementarse de forma externa al codificador. Por ejemplo, en algunas realizaciones, los metadatos espaciales asociados con las señales de audio pueden ser proporcionados a un codificador como un flujo de bits separado. En algunas realizaciones, los metadatos espaciales pueden proporcionarse como un conjunto de valores de índice espacial (dirección).

Las señales multicanal se pasan a un mezclador de señales de transporte 103 y a un procesador 105 de análisis.

En algunas realizaciones, el generador 103 de señal de transporte está configurado para recibir las señales multicanal y generar una señal de transporte adecuada formada por un número determinado de canales y emitir las señales 104 de transporte. Por ejemplo, el generador 103 de señales de transporte puede configurarse para generar una mezcla de canales de audio 2 de las señales multicanal. El número determinado de canales puede ser cualquier número adecuado de canales. El generador de señales de transporte en algunas realizaciones está configurado para seleccionar o combinar de otro modo, por ejemplo, mediante técnicas de formación de haces las señales de audio de entrada al número determinado de canales y emitir estas como señales de transporte.

En algunas realizaciones, el mezclador de señales de transporte 103 es opcional y las señales multicanal se pasan sin procesar a un codificador 107 de la misma manera que la señal de transporte en este ejemplo.

En algunas realizaciones, el procesador 105 de análisis también está configurado para recibir las señales multicanal y analizar las señales para producir metadatos 106 asociados con las señales multicanal y, por lo tanto, asociarse con las señales 104 de transporte. El procesador 105 de análisis puede configurarse para generar los metadatos que pueden comprender, para cada intervalo de análisis de tiempo-frecuencia, un parámetro de dirección 108, un parámetro 110 de relación de energía y un parámetro de coherencia 112 (y en algunas realizaciones un parámetro de difusión). La dirección, la relación de energía y los parámetros de coherencia pueden considerarse en algunas realizaciones como parámetros de audio espacial. En otras palabras, los parámetros de audio espacial comprenden parámetros que tienen como objetivo caracterizar el campo sonoro creado por las señales multicanal (o dos o más señales de audio de reproducción en general).

En algunas realizaciones, los parámetros generados pueden diferir entre las bandas de frecuencia. Por lo tanto, por ejemplo, en la banda X todos los parámetros se generan y transmiten, mientras que en la banda Y solo se genera y transmite uno de los parámetros, y además en la banda Z no se generan o transmiten parámetros. Un ejemplo práctico de esto es que para algunas bandas de frecuencia, tales como la banda más alta, no se requieren algunos de los

parámetros por motivos perceptuales. Las señales 104 de transporte y los metadatos 106 pueden pasarse a un codificador 107.

El codificador 107 puede comprender un núcleo de codificador de audio 109 que está configurado para recibir las señales de transporte (para un ejemplo de mezcla descendente) 104 y generar una codificación adecuada de estas señales de audio. En algunas realizaciones, el codificador 107 puede ser un ordenador (que ejecuta un software adecuado almacenado en la memoria y en al menos un procesador), o alternatively un dispositivo específico que utiliza, por ejemplo, FPGAs o ASIC. La codificación puede implementarse usando cualquier esquema adecuado. El codificador 107 puede comprender además un codificador de metadatos o cuantificador 111 que está configurado para recibir los metadatos y emitir una forma codificada o comprimida de la información. En algunas realizaciones, el codificador 107 puede además intercalar, multiplexar a un flujo de datos individual o incorporar los metadatos dentro de las señales de mezcla descendente codificadas antes de la transmisión o el almacenamiento mostrado en la figura 1 por la línea discontinua. La multiplexación puede implementarse usando cualquier esquema adecuado.

En el lado del decodificador, los datos recibidos o recuperados (flujo) pueden ser recibidos por un decodificador/demultiplexor 133. El decodificador/demultiplexor 133 puede demultiplexar los flujos codificados y pasar la corriente codificada de audio a un extractor de transporte 135 que está configurado para decodificar las señales de audio para obtener las señales de transporte. De manera similar, el decodificador/demultiplexor 133 puede comprender un extractor 137 de metadatos que está configurado para recibir los metadatos codificados y generar metadatos. En algunas realizaciones, el decodificador/demultiplexor 133 puede ser un ordenador (que ejecuta un software adecuado almacenado en la memoria y en al menos un procesador), o alternatively un dispositivo específico que utiliza, por ejemplo, FPGA o ASIC.

Las señales de audio de transporte y metadatos decodificados pueden pasarse a un procesador de síntesis 139.

La parte 131 de "síntesis" del sistema 100 muestra además un procesador 139 de síntesis configurado para recibir los datos de transporte y los metadatos, y para volver a crear en cualquier formato adecuado un audio espacial sintetizado en forma de señales multicanal 110 (que pueden estar en formato de altavoces multicanal o, en algunas realizaciones, en cualquier formato de salida adecuado tales como señales binaural o ambisónica, dependiendo del caso de uso) en función de las señales de transporte y los metadatos.

Por lo tanto, en resumen, el sistema (parte de análisis) está configurado para recibir señales de audio multicanal.

A continuación, el sistema (parte de análisis) está configurado para generar una señal de audio de transporte adecuada (por ejemplo, seleccionando o remezclando algunas de las canales de señal de audio).

El sistema se configura entonces para codificar para su almacenamiento/transmisión la señal de transporte y los metadatos.

Después de esto, el sistema puede almacenar/transmitir los datos de transporte codificado y los metadatos.

El sistema puede recuperar/recibir los datos de transporte codificado y los metadatos.

Posteriormente, el sistema se configura para extraer los datos de transporte y los metadatos de los parámetros codificados de transporte y metadatos, por ejemplo demultiplexar y decodificar los parámetros codificados de transporte y metadatos.

El sistema (parte de síntesis) está configurado para sintetizar una señal de audio multicanal de salida basándose en señales de audio de transporte extraídas y en metadatos.

Con respecto a la figura 2, se describe con más detalle un procesador 105 de análisis de ejemplo y un codificador/cuantificador 111 de metadatos (como puede verse en la figura 1) según algunas realizaciones más detalladas.

El procesador 105 de análisis en algunas realizaciones comprende un transformador 201 de dominio de tiempo-frecuencia.

En algunas realizaciones, el transformador 201 de dominio de tiempo-frecuencia está configurado para recibir las señales multicanal 102 y aplicar una transformación de dominio de tiempo a frecuencia adecuada tal como una Transformación de Fourier de Tiempo Corto (STFT) para convertir las señales de dominio del tiempo de entrada en una señal de tiempo-frecuencia adecuada. Estas señales de tiempo-frecuencia pueden pasar a un analizador espacial 203.

Por lo tanto, por ejemplo, las señales 202 de tiempo-frecuencia pueden representarse en la representación del dominio de tiempo-frecuencia mediante

$$s_i(b, n),$$

donde b es el índice del intervalo de frecuencia y n es el índice del bloque de frecuencia (trama) e i es el índice de canal. En otra expresión, n puede considerarse un índice de tiempo con una tasa de muestreo más baja que la de las señales originales del dominio del tiempo. Estos segmentos de frecuencia pueden agruparse en subbandas que forman uno o más de los segmentos en una subbanda de un índice de banda $K = 0, \dots, K-1$. Cada subbanda k tiene un segmento $b_{k,bajo}$ más bajo y un segmento $b_{k,alto}$ más alto, y la subbanda contiene todos los segmentos desde $b_{k,bajo}$ a $b_{k,alto}$. Las anchuras de las subbandas pueden aproximarse a cualquier distribución adecuada. Por ejemplo, la escala de ancho de banda rectangular equivalente (ERB) o la escala Bark.

En algunas realizaciones, el procesador 105 de análisis comprende un analizador espacial 203. El analizador espacial 203 puede configurarse para recibir las señales 202 de tiempo-frecuencia y basándose en estas señales estima los parámetros 108 de dirección. Los parámetros de dirección se pueden determinar basándose en cualquier determinación de “dirección” basada en audio.

Por ejemplo, en algunas realizaciones, el analizador espacial 203 está configurado para estimar la dirección con dos o más entradas de señal. Esto representa la configuración más simple para estimar una “dirección” y puede realizarse un procesamiento más complejo con incluso más señales.

El analizador espacial 203 puede configurarse para proporcionar al menos un acimut y una elevación para cada banda de frecuencia y bloque temporal-frecuencia de tiempo dentro de un fotograma de una señal de audio, denotado como acimut $\phi(k, n)$ y elevación $\theta(k, n)$. Los parámetros 108 de dirección también se pueden pasar a un codificador 205 de dirección.

El analizador espacial 203 también puede configurarse para determinar un parámetro 110 de relación de energía. Se puede considerar que la relación de energía es una determinación de la energía de la señal de audio que puede considerarse que llega desde una dirección. La relación entre energía directa y total $r(k, n)$ puede estimarse, p. ej., utilizando una medida de estabilidad de la estimación direccional, o utilizando cualquier medida de correlación, o cualquier otro método adecuado para obtener un parámetro de relación. Cada relación entre energía directa y total corresponde a una dirección espacial específica y describe qué cantidad de energía proviene de la dirección espacial específica en comparación con la energía total. Este valor también puede representarse para cada panel de tiempo-frecuencia por separado. Los parámetros de dirección espacial y la relación entre energía directa y total describen qué parte de la energía total de cada panel de tiempo-frecuencia proviene de la dirección específica. En general, un parámetro de dirección espacial también se puede considerar como la dirección de llegada (DOA).

En algunas realizaciones, el parámetro de relación entre energía directa y total se puede estimar en función del parámetro $cor'(k, n)$ de correlación cruzada normalizado entre un par de micrófonos en la banda k , estando el valor del parámetro de correlación cruzada entre -1 y 1. El parámetro $r(k, n)$ de relación entre energía directa y total puede determinarse comparando el parámetro de correlación cruzada normalizado con un parámetro de correlación cruzada

normalizado de campo difuso $cor'_D(k, n)$ como $r(k, n) = \frac{cor'(k, n) - cor'_D(k, n)}{1 - cor'_D(k, n)}$. La relación entre energía directa y total se explica con más detalle en la publicación PCT WO2017/005978.

La relación de energía se puede pasar a un codificador 207 de relación de energía.

El analizador espacial 203 puede configurarse además para determinar un número de parámetros 112 de coherencia que pueden incluir *coherencia envolvente* ($\psi(k, n)$) y *coherencia de propagación* ($\zeta(k, n)$), ambas analizadas en el dominio de tiempo-frecuencia.

A continuación se analiza cada uno de los parámetros de coherencia emitidos antes mencionados. Todo el procesamiento se realiza en el dominio de tiempo-frecuencia, por lo que los índices de tiempo-frecuencia k y n se descartan cuando es necesario por motivos de brevedad.

Consideremos primero la situación en que el sonido se reproduce de manera coherente utilizando dos altavoces espaciados (por ejemplo, los frontales izquierdo y derecho) en lugar de un solo altavoz. El analizador de coherencia puede configurarse para detectar que dicho método se ha aplicado en la mezcla envolvente.

Por lo tanto, en algunas realizaciones, el analizador espacial 203 puede configurarse para calcular la matriz C de covarianzas para el intervalo de análisis dado, que consiste en uno o más índices n de tiempo y segmentos b de frecuencia. El tamaño de la matriz es $N_L \times N_L$, y las entradas se indican como c_{ij} , donde N_L es el número de canales de altavoz e i y j son índices de canales de altavoz.

A continuación, el analizador espacial 203 puede configurarse para determinar el canal de altavoz i_c más cercano a la dirección estimada (que en este ejemplo es el acimut θ).

$$i_c = \arg(\min(|\theta - \alpha_i|))$$

donde α_i es el ángulo del altavoz i .

- 5 Además, en tales realizaciones, el analizador espacial 203 se configura para determinar los altavoces más cercanos en el lado izquierdo i_l y el lado derecho i_r del altavoz i_c .

Una coherencia normalizada entre los altavoces i y j se indica como

$$c'_{ij} = \frac{|c_{ij}|}{\sqrt{|c_{ii}c_{jj}|}},$$

usando esta ecuación, el analizador espacial 203 puede configurarse para calcular una coherencia normalizada c'_{lr} entre i_l e i_r . En otras palabras, para calcular

$$c'_{lr} = \frac{|c_{lr}|}{\sqrt{|c_{ll}c_{rr}|}}.$$

Además, el analizador espacial 203 puede configurarse para determinar la energía de los canales de altavoz i usando las entradas diagonales de la matriz de covarianzas.

$$E_i = c_{ii},$$

y para determinar una relación entre las energías de los altavoces i_l e i_r y de los altavoces i_l , i_r , e i_c como

$$\xi_{lr/lrc} = \frac{E_l + E_r}{E_l + E_r + E_c}.$$

El analizador espacial 203 puede usar después estas variables determinadas para generar un parámetro de “estereofonía”.

$$\mu = c'_{lr} \xi_{lr/lrc}.$$

Este parámetro de “estereofonía” tiene un valor entre 0 y 1. Un valor de 1 significa que hay un sonido coherente en los altavoces i_l e i_r , y que este sonido domina la energía de este sector. La razón de esto podría ser, por ejemplo, que la mezcla en los altavoces utilizó técnicas de panoramización de amplitud para crear una percepción “con aire” del sonido. Un valor de 0 significa que no se ha aplicado ninguna técnica de este tipo y, por ejemplo, el sonido puede simplemente posicionarse en el altavoz más cercano.

Además, el analizador espacial 203 puede configurarse para detectar, o al menos identificar, una situación en que el sonido se reproduzca de manera coherente usando tres (o más) altavoces para crear una percepción “cercana” (por ejemplo, usar el frontal izquierdo, el frontal derecho y el central en lugar de solo el central). Esto puede suceder si un ingeniero de mezcla de sonido genera una situación de este tipo al producir la mezcla envolvente de los altavoces multicanal.

En tales realizaciones, el analizador de coherencia usa los mismos altavoces i_l , i_r e i_c identificados anteriormente para determinar los valores de coherencia normalizados c'_{cl} y c'_{cr} usando la determinación de coherencia normalizada descrita anteriormente. En otras palabras, se calculan los siguientes valores:

$$c'_{cl} = \frac{|c_{cl}|}{\sqrt{|c_{cc}c_{ll}|}}, \quad c'_{cr} = \frac{|c_{cr}|}{\sqrt{|c_{cc}c_{rr}|}}.$$

El analizador espacial 203 puede después determinar un valor de coherencia normalizado c'_{clr} que representa la coherencia entre estos altavoces usando la siguiente expresión:

$$\xi_{clr} = \min \left(\frac{E_l}{E_c}, \frac{E_c}{E_l}, \frac{E_r}{E_c}, \frac{E_c}{E_r} \right).$$

Además, el analizador espacial 203 puede configurarse para determinar un parámetro que represente cómo de uniformemente se distribuye la energía entre los canales i_l , i_r e i_c .

$$\xi_{clr} = \min \left(\frac{E_l}{E_c}, \frac{E_c}{E_l}, \frac{E_r}{E_c}, \frac{E_c}{E_r} \right).$$

Usando estas variables, el analizador espacial 203 puede determinar un nuevo parámetro de panoramización coherente κ como,

$$\kappa = c'_{clr} \xi_{clr}.$$

Este parámetro de panoramización coherente κ tiene valores entre 0 y 1. Un valor de 1 significa que hay un sonido coherente en todos los altavoces i_l , i_r , y i_c , y que la energía de este sonido se distribuye uniformemente entre estos altavoces. La razón de esto podría deberse, por ejemplo, a que la mezcla de altavoces se generó utilizando técnicas de mezcla de estudio para crear la percepción de que una fuente de sonido está más cerca. Un valor de 0 significa que no se ha aplicado ninguna técnica de este tipo y, por ejemplo, el sonido puede simplemente posicionarse en el altavoz más cercano.

El analizador espacial 203 determinó el parámetro de “estereofonía” μ , que mide la cantidad de sonido coherente en i_l e i_r (pero no en i_c), y el parámetro de panoramización coherente κ que mide la cantidad de sonido coherente en i_l , i_r , e i_c está configurado para usarlos para determinar los parámetros de coherencia que se emitirán como metadatos.

Por lo tanto, el analizador espacial 203 está configurado para combinar el parámetro de “estereofonía” μ y el parámetro de panoramización coherente para formar un parámetro de *coherencia de propagación*, que tiene valores de 0 a 1. Un valor de *coherencia de propagación* ζ de 0 indica una fuente puntual; en otras palabras, el sonido debe reproducirse con el menor número posible de altavoces (por ejemplo, utilizando solo el altavoz i_c). A medida que aumenta el valor de la *coherencia de propagación* ζ , se propaga más energía a los altavoces que rodean el altavoz i_c ; hasta el valor 0,5, la energía se propaga uniformemente entre los altavoces i_l , i_r , e i_c . A medida que el valor de la *coherencia de propagación* ζ aumenta por encima de 0,5, la energía del altavoz i_c disminuye; hasta el valor 1, no hay energía en el altavoz i_c , y toda la energía está en los altavoces i_l e i_r .

Usando los parámetros μ y κ anteriormente mencionados, en algunas realizaciones el analizador espacial 203 se configura para determinar un parámetro de coherencia de propagación ζ , usando la siguiente expresión:

$$\zeta = \begin{cases} \max(0,5, \mu - \kappa + 0,5), & \text{si } \max(\mu, \kappa) > 0,5 \text{ \& } \kappa > \mu. \\ \max(\mu, \kappa), & \text{si no} \end{cases}$$

La expresión anterior es solo un ejemplo y debe tenerse en cuenta que el analizador espacial 203 puede estimar el parámetro de coherencia de propagación ζ de cualquier otra manera, siempre que cumpla con la definición anterior del parámetro.

Además de estar configurado para detectar las situaciones anteriores, el analizador espacial 203 puede configurarse para detectar, o al menos identificar, la situación en que el sonido se reproduce de manera coherente desde todos (o casi todos) los altavoces para crear una percepción “dentro de la cabeza” o “por encima de la cabeza”.

En algunas realizaciones, el analizador espacial 203 puede configurarse para clasificar las energías E_i y el canal de altavoz i_e con el valor más grande determinado.

El analizador espacial 203 puede configurarse después para determinar la coherencia normalizada c'_{ij} entre este canal y M_L otros canales con volumen más alto. A continuación, pueden monitorizarse estos valores de coherencia normalizados c'_{ij} entre este canal y M_L otros canales con volumen más alto. En algunas realizaciones, M_L puede ser N_{L-1} , lo que significaría monitorizar la coherencia entre los canales de altavoz con volumen más alto y todos los demás

canales de altavoz. Sin embargo, en algunas realizaciones, M_L puede ser un número menor, por ejemplo, M_L-2 . Usando estos valores de coherencia normalizados, el analizador de coherencia puede configurarse para determinar un parámetro de coherencia envolvente y usando la siguiente expresión:

$$\gamma = \min_M (c'_{iej}),$$

5

donde c'_{iej} son las coherencias normalizadas entre el canal con más volumen y los M_L siguientes canales con más volumen.

10 El parámetro de coherencia envolvente γ tiene valores de 0 a 1. Un valor de 1 significa que hay coherencia entre todos los canales de altavoces (o casi todos). Un valor de 0 significa que no hay coherencia entre todos los canales de altavoces (o incluso casi todos).

15 La expresión anterior es solo un ejemplo de una estimación de un parámetro de coherencia envolvente γ , y se puede utilizar cualquier otra forma, siempre que cumpla con la definición anterior del parámetro.

Después de esto, el procesador 203 de análisis puede aplicar un análisis direccional para determinar los parámetros de dirección y de relación de energía.

20 El procesador 203 de análisis puede configurarse después para emitir los parámetros determinados.

Aunque aquí se expresan direcciones, relaciones de energía y parámetros de coherencia para cada índice de tiempo n , en algunas realizaciones los parámetros se combinan durante varios índices de tiempo. Lo mismo se aplica para el eje de frecuencia, como se ha expresado, la dirección de varios intervalos de frecuencia b podría expresarse por un parámetro de dirección en la banda k que consiste en varios binarios de frecuencia b . Lo mismo se aplica para todos los parámetros espaciales analizados en la presente memoria.

25 En algunas realizaciones, los datos direccionales pueden representarse usando 16 bits. En tales realizaciones, el parámetro de relación de energía puede representarse en 8 bits. Para cada trama, puede haber 5 subbandas y 4 bloques de frecuencia de tiempo (TF). Por lo tanto, en este ejemplo hay $(16+8) \times 4 \times 5$ bits necesarios para almacenar los metadatos de la dirección sin comprimir y de la relación de energía para cada trama. Los datos de coherencia para cada bloque de TF pueden ser una representación entre 0 y 1, y pueden representarse originalmente en 8 bits.

30 Según muestra también la figura 2, se muestra un codificador/cuantificador 111 de metadatos de ejemplo según algunas realizaciones.

El codificador/cuantificador 111 de metadatos comprende un codificador 205 de dirección. El codificador 205 de dirección está configurado para recibir los parámetros 108 de dirección, tales como el de acimut $\phi(k, n)$ y el de elevación $\theta(k, n)$ (y en algunas realizaciones una asignación de bits esperada), y generar a partir de esto una salida codificada adecuada. En algunas realizaciones, la codificación se basa en una disposición de esferas que forman una retícula esférica dispuesta en anillos en una esfera de “superficie” que se define por una tabla de consulta definida por la resolución de cuantificación determinada. En otras palabras, la retícula esférica utiliza la idea de cubrir una esfera con esferas más pequeñas y considerar los centros de las esferas más pequeñas como puntos que definen una retícula de direcciones casi equidistantes. Por lo tanto, las esferas más pequeñas definen conos o ángulos sólidos alrededor del punto central que pueden indexarse según cualquier algoritmo de indexación adecuado. Aunque la cuantificación esférica se describe aquí cualquier cuantificación adecuada, lineal o no lineal.

35 Además, en algunas realizaciones, el codificador 205 de dirección está configurado para determinar una varianza del valor de parámetro de acimut y pasarlo al codificador 209 de coherencia.

50 Los parámetros de dirección codificados pueden pasarse entonces al combinador 211.

El codificador/cuantificador 111 de metadatos puede comprender un codificador 207 de relación de energía. El codificador 207 de relación de energía está configurado para recibir las relaciones de energía y determinar una codificación adecuada para comprimir las relaciones de energía para las subbandas y los bloques de tiempo-frecuencia. Por ejemplo, en algunas realizaciones, el codificador 207 de relación de energía está configurado para usar 3 bits para codificar cada valor de parámetro de relación de energía.

60 Además, en algunas realizaciones en lugar de transmitir o almacenar todos los valores de relación de energía para todos los bloques TF, solo se transmite o almacena un valor promedio ponderado por subbanda. El promedio puede determinarse teniendo en cuenta la energía total de cada bloque de tiempo, lo que favorece, por lo tanto, los valores de las subbandas que tienen más energía.

En tales realizaciones, el valor de relación de energía cuantificado es el mismo para todos los bloques TF de una subbanda dada.

- 5 En algunas realizaciones, el codificador 207 de relación de energía está configurado además para pasar el valor de relación de energía cuantificado (codificado) al combinador 211 y al codificador 209 de coherencia.

10 El codificador/cuantificador 111 de metadatos puede comprender un codificador 209 de coherencia. El codificador 209 de coherencia está configurado para recibir los valores de coherencia y determinar una codificación adecuada para comprimir los valores de coherencia para las subbandas y los bloques de tiempo-frecuencia. Se ha demostrado que un valor de precisión de 3 bits para los valores de parámetro de coherencia produce resultados aceptables de síntesis de audio, pero incluso esto requeriría un total de 3×20 bits para los datos de coherencia para todos los bloques TF (en el ejemplo 8 de subbanda y 5 bloques TF por fotograma).

- 15 Por lo tanto, un objetivo de las realizaciones es reducir el número de bits para representar los datos de coherencia a lo largo de varios bloques de TF.

Los valores de parámetros de coherencia codificados pueden pasarse entonces al combinador 211.

- 20 El codificador/cuantificador 111 de metadatos puede comprender un combinador 211. El combinador está configurado para recibir los parámetros direccionales codificados (o cuantificados/comprimidos), parámetros de relación de energía y parámetros de coherencia y combinar estos para generar una salida adecuada (por ejemplo, un flujo de bits de metadatos que puede combinarse con la señal de transporte o transmitirse o almacenarse por separado desde la señal de transporte).

- 25 Con respecto a la figura 3 se muestra un ejemplo de operación del codificador/cuantificador de metadatos como se muestra en la figura 2 según algunas realizaciones.

- 30 La operación inicial es obtener los metadatos (tales como valores de acimut, valores de elevación, relaciones de energía, coherencia, etc.) como se muestra en la figura 3 por la etapa 301.

Los valores direccionales (elevación, acimut) pueden ser comprimidos o codificados (por ejemplo, aplicando una cuantificación esférica o cualquier compresión adecuada) como se muestra en la figura 3 por la etapa 303.

- 35 Los valores de relación de energía se comprimen o codifican (por ejemplo, generando un promedio ponderado por subbanda y luego cuantificando estos como un valor de 3 bits) según muestra la figura 3 mediante la etapa 305.

Los valores de coherencia también se comprimen o codifican según muestra la figura 3 mediante la etapa 307.

- 40 Los valores direccionales codificados, las relaciones de energía y los valores de coherencia se combinan para después generar los metadatos codificados según muestra la figura 3 mediante la etapa 309.

Con respecto a la figura 4 se muestra un codificador 209 de coherencia de ejemplo como se muestra en la figura 2.

- 45 Como se mencionó anteriormente, se puede lograr un beneficio técnico regulando la codificación de los parámetros de coherencia en que, para algunas tramas de audio, los parámetros de coherencia están codificados y, en consecuencia, forman parte del flujo de bits codificado y, para otras tramas de audio, los parámetros de coherencia no están codificados. El beneficio técnico es la disminución general de la tasa de transferencia del flujo de bits codificado.

- 50 La decisión de codificar los parámetros de coherencia puede hacerse basándose en la significación subjetiva del valor de coherencia de cada subbanda en relación con las subbandas envolventes. En otras palabras, la decisión de codificar los parámetros de coherencia se puede tomar en función de si los parámetros de coherencia contribuyen a la experiencia auditiva perceptiva global del usuario final.

- 55 Los experimentos de escucha han demostrado que la percepción de coherencia en la señal de audio espacial puede atribuirse, al menos en parte, a la interacción entre las señales presentes en las diversas bandas de frecuencia. Esto puede parecer contrario a la premisa de que, para la percepción del sonido, la resolución de frecuencia humana puede aislarse/compartimentarse según escalas de frecuencia no lineales, tales como el ancho de banda rectangular equivalente (ERB).

- 60 Por ejemplo, se observó mediante experimentación (que comprendió pruebas de escucha perceptiva) que si la coherencia es baja en la mayoría de las bandas de frecuencia, una mayor coherencia en una sola banda de frecuencia puede tener una contribución insignificante a la percepción del audio espacial. En consecuencia, en este caso, puede ser beneficioso no codificar el parámetro de coherencia para la tasa de transferencia global del flujo de bits codificado.

- 65

Alternativamente, también se observó mediante pruebas de escucha perceptiva que si se descubre que una sola banda de frecuencia tiene una relación de energía significativamente más baja que otras bandas de frecuencia envolventes, la coherencia puede ser perceptualmente significativa en esa banda. En este caso, se descubrió que el sistema auditivo humano puede ser particularmente sensible. En consecuencia, sería prudente transmitir los parámetros de coherencia para la trama con el fin de mejorar la experiencia de escucha general del audio espacial.

Por lo tanto, los resultados de las observaciones experimentales sugieren que se puede diseñar una métrica (o medida) que represente la significación de la coherencia en la señal de audio espacial.

Estos efectos pueden ser particularmente frecuentes para la percepción de la coherencia envolvente en la señal de audio espacial. Sin embargo, medir la significación de otros tipos de parámetros de coherencia también puede resultar beneficioso a la hora de decidir si se codifican los parámetros de coherencia. Por ejemplo, este tipo de técnicas también se pueden aplicar para difundir los valores de los parámetros de coherencia.

Como se indicó anteriormente, estos efectos pueden atribuirse a la interacción entre las bandas de frecuencia cuando se trata de la percepción de la coherencia.

A este respecto, primero se hace referencia a 401 en la figura 4, donde se representa un determinador de significación de valor de coherencia que recibe los valores de coherencia de cada subbanda k para el bloque TF n . Además, 401 también se puede disponer para recibir las relaciones entre energía directa y total $r(k, n)$ para el bloque TF n (o subtrama n), es decir, las relaciones de energía asociadas a las K subbandas del bloque TF con índice de tiempo n . Del mismo modo, los valores de coherencia también se suministran al determinador 401 de significación de valor de coherencia como K valores de coherencia para las K subbandas del bloque TF con índice de tiempo n .

En una primera realización, los valores de coherencia pueden comprender la coherencia envolvente ($\gamma(k, n)$). Debe observarse que k se usa para indicar el índice de subbanda y n es el índice del bloque TF.

También se hace referencia a la figura 5, que representa ejemplos de etapas de procesamiento para determinar la significación de los valores de coherencia envolvente para un bloque TF que comprende K subbandas.

Inicialmente, el bloque 401 de procesamiento puede determinar una estimación de la relación de energía no direccional para cada subbanda k utilizando la relación entre energía directa y total $r(k, n)$. En algunos ejemplos, la relación de energía no direccional (envolvente) $a(k, n)$ puede estar dada por

$$a(k, n) = 1 - r(k, n) \text{ para } k = 0 \text{ a } K - 1$$

Cabe señalar que $r(k, n)$ es la energía en una dirección particular sobre la energía total, la dirección de cada frecuencia viene dada por el acimut y la elevación. Como se mencionó anteriormente, la dirección está representada por un índice de dirección del codificador 205 de dirección. La relación de energía $r(k, n)$ puede tener un valor entre 0 y 1, donde un 1 significaría que la energía total se concentra a lo largo de un índice de dirección para una subbanda k . Por lo tanto, $a(k, n)$ representa la relación de energía envolvente (o no direccional) para la subbanda k particular, en otras palabras, energía que no se concentra a lo largo de un índice de dirección particular.

La etapa de procesamiento para determinar la relación de energía no direccional (envolvente) se muestra como 501 en la figura 5.

La proporción de energía coherente no direccional (envolvente) para cada subbanda k puede, por lo tanto, estar dada por

$$\rho(k, n) = a(k, n)\gamma(k, n) \text{ para } k = 0 \text{ a } K - 1$$

La etapa de procesamiento para determinar la proporción de energía no direccional (envolvente) para cada subbanda se representa como la etapa 503 de procesamiento en la figura 5.

Se puede estimar una primera medida (o métrica) de la significación de los valores de coherencia envolvente para el bloque TF n (o subtrama n) sumando la proporción de energías no direccionales en todas las subbandas de la subtrama n .

$$\xi_1(n) = \frac{1}{K} \sum_k \rho(k, n)$$

Puede considerarse que la primera medida ξ_1 indica si la energía no direccional (envolvente) es prominente en múltiples subbandas de frecuencia de la subtrama n.

La etapa de procesamiento para determinar la primera medida de la significación de los valores de coherencia envolvente para la subtrama n se representa como la etapa 505 de procesamiento en la figura 5.

Se puede estimar una segunda medida (o métrica) ξ_2 para la significación de los valores de coherencia envolvente para la subtrama n tomando la proporción de energía no direccional $\rho(k, n)$ y ponderándola con la relación de la energía envolvente (o no direccional) para la subbanda k, $a(k, n)$ particular. Esto puede sumarse en todas las subbandas y normalizarse con la proporción “total” de energías envolventes (o no direccionales). La segunda medida, ξ_2 , se puede expresar como

$$\xi_2(n) = \mu \frac{\sum_k a(k, n) \rho(k, n)}{\sum_k a(k, n)},$$

donde μ es un factor de ajuste. Las observaciones experimentales han demostrado que un valor de $\mu = 0,4$ puede producir un resultado ventajoso.

Puede considerarse que la segunda medida (o métrica) ξ_2 proporciona una medida de si la energía coherente envolvente es prominente dentro de las subbandas que son predominantemente no direccionales y, por lo tanto, proporciona una indicación de si los valores de coherencia envolvente proporcionados por la subtrama pueden ser perceptualmente significativos y, por lo tanto, si los valores de coherencia deben codificarse y transmitirse.

La etapa de procesamiento para determinar la segunda medida de la significación del parámetro de coherencia envolvente para la (subtrama) TF n se representa como la etapa 507 de procesamiento en la figura 5.

En algunos ejemplos, la primera y la segunda medidas pueden fusionarse para proporcionar una medida única de la significación de los valores de coherencia envolvente para el bloque (subtrama) de TF n.

La fusión puede adoptar la forma de tomar la máxima de la primera o segunda medidas para obtener una sola medida de la significación de los valores de coherencia envolvente para el bloque (subtrama) de TF n.

$$\xi(n) = \max(\xi_1(n), \xi_2(n))$$

La salida del determinador 401 de significación de valores de coherencia puede ser la medida única ξ para la significación de los valores de coherencia envolvente. Este parámetro puede pasarse al determinador 403 de codificación de valores de coherencia. El bloque 403 de procesamiento puede disponerse para determinar si los valores de coherencia para la subtrama deben codificarse o no codificarse.

La etapa de procesamiento de fusionar la primera y segunda medidas para ofrecer una medida única que indique la significación de los valores de coherencia envolvente para el bloque de TF n se muestra mediante 509 en la figura 5.

En algunas realizaciones, la medida única ξ de la significación de los valores de coherencia envolvente puede comprender únicamente la primera medida ξ_1 de la significación de los valores envolventes. Alternativamente, en otras realizaciones, la medida única ξ de la significación de los valores de coherencia envolvente puede comprender únicamente la segunda medida ξ_2 de la significación de los valores envolventes. En otras palabras, en ambas realizaciones no hay ninguna etapa 509 de fusión. Por lo tanto, la medida ξ de la significación de los valores de coherencia envolvente puede ser solo la primera medida ξ_1 o simplemente la segunda medida ξ_2 para el funcionamiento del determinador de 401 significación de valores de coherencia.

La figura 6 muestra un ejemplo de conjunto de etapas de procesamiento que puede ser ejecutado por el determinador 403 de codificación de valores de coherencia.

En algunos ejemplos, la determinación de si deben codificarse los valores de coherencia envolvente para el bloque de TF puede consistir en comparar la medida única ξ de la significación de los valores de coherencia envolvente con un valor umbral para la subtrama n , $\tau(n)$. Esta comparación se puede expresar como

$$e(n) = 1, \quad \text{si } \xi(n) > \tau(n)$$

$$e(n) = 0, \quad \text{si } \xi(n) \leq \tau(n)$$

5

donde $e(n) = 1$ indica codificar (y transmitir) los valores de coherencia para el bloque (o subtrama) de TF n , y $e(n) = 0$ indica no codificar los valores de coherencia para la subtrama n .

10 El valor umbral puede ser una constante. Por ejemplo, puede adoptar un valor tal como $\tau(n) = \tau = 0,15$. Alternativamente, en algunas realizaciones el umbral puede cambiar con el tiempo. Por ejemplo, el valor umbral puede seleccionarse en función de un factor basado en una tasa de transferencia de codificación.

15 La etapa de comparar la medida única ξ para la significación del valor de coherencia envolvente con un valor umbral se muestra como la etapa 601 de procesamiento en la figura 6.

20 El bloque funcional 403 puede disponerse después para enviar al codificador 405 de parámetros de coherencia los valores de coherencia para el bloque (subtrama) de TF n , con la condición de que el resultado de la comparación anterior indique que los valores de coherencia son lo suficientemente significativos como para garantizar la codificación. En consecuencia, la figura 4 representa un trayecto de conexión desde el determinador 403 de codificación de parámetros de coherencia hasta el codificador 405 de parámetros de coherencia. Sin embargo, si la medida única ξ de la significación de los valores de coherencia envolvente no supera el umbral, los valores de coherencia no se codifican.

25 La etapa 603 de procesamiento en la figura 6 representa la situación donde la medida única ξ para la significación de los valores de coherencia envolvente supera el umbral, lo que resulta en que los valores de coherencia para el bloque se envían al codificador 405 de valores de coherencia para su codificación. Por el contrario, la etapa 605 de procesamiento en la figura 6 representa una situación cuando la medida única ξ para la significación de los valores de coherencia envolvente no supera el umbral, lo que da como resultado que no se codifiquen los valores de coherencia del bloque de TF.

30 El codificador 405 de parámetros de coherencia está configurado para recibir los valores de coherencia que se han determinado como perceptualmente significativos a partir del bloque funcional 403. Estos valores de coherencia pueden ser una representación de 8 bits entre 0 y 1.

35 El codificador 405 de valores de coherencia puede configurarse para codificar los valores de coherencia asociados a las subbandas para un bloque (o subtrama) de TF, bloque por bloque de TF. El codificador 405 de valores de coherencia puede adoptar un enfoque en el cual primero se cuantifique el valor de coherencia para cada subbanda de forma escalar, usando una tabla de cuantificación particular de una pluralidad de tablas de cuantificación. La selección de la tabla de cuantificación particular para cuantificar el valor de coherencia puede depender del valor de relación de energía para cada subbanda, y el número de entradas de cada tabla de cuantificación puede diferir entre sí. En esencia, cada tabla de cuantificación de la pluralidad de tablas de cuantificación puede comprender un rango de valores de coherencia cuantificados precalculados que pueden adaptarse a un valor de relación de energía particular.

45 El índice de tabla de cuantificación asociado a cada valor de coherencia puede codificarse después usando un código de tipo de producto, mediante el cual se concatenan entre sí todos los índices de cuantificación para las subbandas del bloque de TF en un índice combinado o un código de longitud de ejecución, tal como un código de Golomb Rice.

50 En el caso de usar un código Golomb Rice para codificar los índices de cuantificación para los valores de coherencia, se puede eliminar su valor medio para el bloque de TF.

55 La selección entre el índice combinado conjunto y el código de Golomb Rice para la codificación de los índices de cuantificación del valor de coherencia de subbanda para el bloque de TF se puede realizar en función de la estimación del número de bits que utilizará el codificador de Golomb Rice. Si es inferior a los bits del índice combinado conjunto, se usa después el codificador de Golomb Rice para codificar los índices de cuantificación del valor de coherencia de subbanda para el bloque de TF. La selección entre el método de índice combinado conjunto y el método de Golomb Rice se puede señalizar como un único bit dentro del flujo de bits.

60 Los valores de índice de los libros de códigos pueden después formar los valores de coherencia codificados, que se combinan después con los valores direccionales codificados y las relaciones de energía codificadas, en el combinador 211, para formar los metadatos de audio espacial.

Con respecto a la figura 7, se muestra un dispositivo electrónico de ejemplo que puede utilizarse como dispositivo de análisis o síntesis. El dispositivo puede ser cualquier dispositivo o aparato electrónico adecuado. Por ejemplo, en algunas realizaciones, el dispositivo 1400 es un dispositivo móvil, un equipo de usuario, un ordenador de tipo tableta, un ordenador, un aparato de reproducción de audio, etc.

En algunas realizaciones, el dispositivo 1400 comprende al menos un procesador o unidad 1407 de procesamiento central. El procesador 1407 puede estar configurado para ejecutar diversos códigos de programa tales como los métodos tales como los descritos en la presente memoria.

En algunas realizaciones, el dispositivo 1400 comprende una memoria 1411. En algunas realizaciones, el al menos un procesador 1407 está acoplado a la memoria 1411. La memoria 1411 puede ser cualquier medio de almacenamiento adecuado. En algunas realizaciones, la memoria 1411 comprende una sección de código de programa para almacenar códigos de programa que pueden implementarse en el procesador 1407. Además, en algunas realizaciones, la memoria 1411 puede comprender además una sección de datos almacenados para almacenar datos, por ejemplo datos que se han procesado o que van a procesarse según las realizaciones como se describe en la presente memoria. El código de programa implementado almacenado dentro de la sección de código de programa y los datos almacenados dentro de la sección de datos almacenados pueden recuperarse por el procesador 1407 siempre que sea necesario a través del acoplamiento de memoria-procesador.

En algunas realizaciones, el dispositivo 1400 comprende una interfaz 1405 de usuario. En algunas realizaciones, la interfaz 1405 de usuario puede estar acoplada al procesador 1407. En algunas realizaciones, el procesador 1407 puede controlar el funcionamiento de la interfaz 1405 de usuario y recibir entradas de la interfaz 1405 de usuario. En algunas realizaciones, la interfaz 1405 de usuario puede permitir que un usuario introduzca comandos al dispositivo 1400, por ejemplo, a través de un teclado. En algunas realizaciones, la interfaz 1405 de usuario puede permitir que el usuario obtenga información a partir del dispositivo 1400. Por ejemplo, la interfaz 1405 de usuario puede comprender un elemento de visualización configurado para visualizar información del dispositivo 1400 al usuario. En algunas realizaciones, la interfaz 1405 de usuario puede comprender una pantalla táctil o una interfaz táctil que puede tanto permitir introducir información al dispositivo 1400 como visualizar adicionalmente información al usuario del dispositivo 1400. En algunas realizaciones, la interfaz 1405 de usuario puede ser la interfaz de usuario para comunicarse con el determinador de posición como se describe en la presente memoria.

En algunas realizaciones, el dispositivo 1400 comprende un puerto 1409 de entrada/salida. En algunas realizaciones, el puerto 1409 de entrada/salida comprende un transceptor. En tales realizaciones, el transceptor puede estar acoplado al procesador 1407 y configurado para permitir una comunicación con otros aparatos o dispositivos electrónicos, por ejemplo, a través de una red de comunicación inalámbrica. En algunas realizaciones, el transceptor o cualquier transceptor o medios de transmisor y/o receptor adecuado puede estar configurado para comunicarse con otros dispositivos o aparatos electrónicos a través de un cable o acoplamiento por cable.

El transceptor puede comunicarse con un aparato adicional mediante cualquier protocolo de comunicaciones conocido adecuado. Por ejemplo, en algunas realizaciones, el transceptor puede usar un protocolo de universal mobile telecommunications system (sistema universal de telecomunicaciones móviles - UMTS), un protocolo de wireless local area network (red de área local inalámbrica - WLAN) tal como por ejemplo IEEE 802.X, un protocolo de comunicación de radiofrecuencia de corto alcance adecuado tal como Bluetooth, o infrared data communication pathway (ruta de comunicación de datos por infrarrojos - IRDA).

El puerto 1409 de entrada/salida del transceptor puede estar configurado para recibir las señales y, en algunas realizaciones, determinar los parámetros como se describe en la presente memoria mediante el uso del procesador 1407 que ejecuta un código adecuado. Además, el dispositivo puede generar una salida de señal de mezcla descendente y de parámetros adecuada que va a transmitirse al dispositivo de síntesis.

En algunas realizaciones, el dispositivo 1400 puede emplearse como al menos parte del dispositivo de síntesis. Como tal, el puerto 1409 de entrada/salida puede estar configurado para recibir las señales de mezcla descendente y, en algunas realizaciones, los parámetros determinados en el dispositivo de captación o dispositivo de procesamiento como se describe en la presente memoria, y generar una salida de formato de señal de audio adecuada mediante el uso del procesador 1407 que ejecuta un código adecuado. El puerto 1409 de entrada/salida puede acoplarse a cualquier salida de audio adecuada, por ejemplo, a un sistema de altavoces de múltiples canales y/o auriculares o similar.

En general, las diversas realizaciones de la invención pueden implementarse en hardware o circuitos de propósito especial, software, lógica o cualquier combinación de los mismos. Por ejemplo, algunos aspectos pueden implementarse en hardware, mientras que otros aspectos pueden implementarse en firmware o software que puede ejecutarse por un controlador, microprocesador u otro dispositivo informático, aunque la invención no se limita a los mismos. Aunque diversos aspectos de la invención pueden ilustrarse y describirse como diagramas de bloques, diagramas de flujo, o usando alguna otra representación gráfica, se entiende que estos bloques, aparatos, sistemas, técnicas o métodos descritos en la presente memoria pueden implementarse, como ejemplos no limitativos, en

hardware, software, firmware, circuitos o lógica de propósito especial, hardware de propósito general o controlador u otros dispositivos informáticos, o alguna combinación de los mismos.

5 Las realizaciones de esta invención pueden implementarse mediante software informático ejecutable por un procesador de datos del dispositivo móvil, tal como en la entidad de procesador, o mediante hardware, o mediante una combinación de software y hardware. Además, en este sentido, cabe señalar que cualesquiera bloques del flujo lógico como en las figuras puede representar etapas del programa, o circuitos lógicos interconectados, bloques y funciones, o una combinación de etapas de programa y circuitos lógicos, bloques y funciones. El software puede almacenarse en medios físicos tales como chips de memoria, o bloques de memoria implementados dentro del
10 procesador, medios magnéticos tales como disco duro o disquetes, y medios ópticos tales como, por ejemplo, DVD y sus variantes de datos, CD.

La memoria puede ser de cualquier tipo adecuado para el entorno técnico local y puede implementarse usando cualquier tecnología de almacenamiento de datos adecuada, tal como dispositivos de memoria basados en
15 semiconductores, dispositivos y sistemas de memoria magnéticos, dispositivos y sistemas de memoria ópticos, memoria fija y memoria extraíble. Los procesadores de datos pueden ser de cualquier tipo adecuado para el entorno técnico local, y pueden incluir uno o más de ordenadores de propósito general, ordenadores de propósito especial, microprocesadores, digital signal processors (procesadores de señales digitales - DSP), application specific integrated circuits (circuitos integrados específicos de aplicación - ASIC), circuitos de nivel de puerta y procesadores basados en
20 arquitectura de procesador de múltiples núcleos, como ejemplos no limitativos.

Las realizaciones de las invenciones pueden ponerse en práctica en diversos componentes tales como módulos de circuito integrado. El diseño de circuitos integrados es, en gran medida, un proceso altamente automatizado. Hay herramientas de software complejas y potentes disponibles para convertir un diseño de nivel lógico en un diseño de
25 circuito de semiconductores listo para grabarse y formarse en un sustrato semiconductor.

Los programas pueden enrutar conductores y ubicar componentes en un chip semiconductor, utilizando reglas de diseño bien establecidas, así como bibliotecas de módulos de diseño previamente almacenados. Una vez que se ha completado el diseño para un circuito de semiconductor, el diseño resultante, en un formato electrónico normalizado,
30 puede transmitirse a una instalación de fabricación de semiconductores o "fáb." para su fabricación.

La descripción anterior ha proporcionado, a modo de ejemplo y como ejemplos no limitativos, una descripción completa e informativa de la realización ilustrativa de esta invención. Sin embargo, diversas modificaciones y adaptaciones pueden resultar evidentes para los expertos en las técnicas relevantes a la vista de la descripción anterior, cuando se lee junto con los dibujos adjuntos y las reivindicaciones adjuntas. Sin embargo, todas estas modificaciones y similares de las enseñanzas de la presente invención seguirán estando dentro del alcance de la presente invención como se define en las reivindicaciones adjuntas.
35

REIVINDICACIONES

1. Un aparato para la codificación de audio espacial, que comprende:

medios para recibir o determinar, para una o más señales de audio, parámetros de audio espacial por subbandas para proporcionar una reproducción de audio espacial, en donde los parámetros de audio espacial comprenden un valor de coherencia para cada subbanda de una pluralidad de subbandas de una trama;

medios para recibir un valor de relación de energía para cada una de la pluralidad de subbandas;

medios para determinar una medida de significación para los valores de coherencia de la pluralidad de subbandas de la trama; y

medios para usar la medida de significación para determinar si se deben codificar los valores de coherencia de la pluralidad de subbandas de la trama, **en donde el aparato está caracterizado porque** los medios para determinar la medida de significación para los valores de coherencia de la pluralidad de subbandas de la trama comprenden:

medios para determinar una proporción de energía no direccional coherente para cada una de la pluralidad de subbandas, en donde la proporción de energía no direccional coherente para una subbanda se determina como una relación de energía no direccional para la subbanda multiplicada por el valor de coherencia para la subbanda, en donde la relación de energía no direccional está relacionada con la relación de energía en la subbanda que queda después de haber eliminado el valor de relación de energía para la subbanda; y
medios para sumar, para toda la pluralidad de subbandas, la proporción determinada de energía no direccional coherente.

2. El aparato según la reivindicación 1, en donde los medios para determinar la medida de significación para los valores de coherencia de la pluralidad de subbandas de la trama comprenden además:

medios para multiplicar, para cada subbanda, la proporción de energía no direccional coherente por la relación de energía no direccional;

medios para sumar, para toda la pluralidad de subbandas, el múltiplo de la proporción de energía no direccional coherente y la relación de energía no direccional para cada subbanda; y

medios para normalizar el múltiplo sumado de la proporción de energía no direccional coherente y la relación de energía no direccional para cada subbanda mediante una suma de la relación de energía no direccional para cada subbanda.

3. El aparato según la reivindicación 1, en donde la medida de significación para los valores de coherencia de la pluralidad de subbandas de la trama es una primera medida de significación para los valores de coherencia de la pluralidad de subbandas de la trama.

4. El aparato según la reivindicación 2, en donde la medida de significación para los valores de coherencia de la pluralidad de subbandas de la trama es una segunda medida de significación para los valores de coherencia de la pluralidad de subbandas de la trama.

5. El aparato según la reivindicación 4 en combinación con la reivindicación 3, en donde los medios para determinar la medida de significación para los valores de coherencia de la pluralidad de subbandas de la trama comprenden:

medios para determinar la medida de significación para los valores de coherencia de la pluralidad de subbandas de la trama como la máxima de la primera medida de significación y la segunda medida de significación.

6. El aparato según las reivindicaciones 1, 2 y 5, en donde los medios para usar la medida de significación para determinar si se deben codificar los valores de coherencia de la pluralidad de subbandas de la trama comprenden:

medios para comparar la medida de significación con un valor umbral; y

medios para codificar los valores de coherencia de la pluralidad de subbandas de la trama cuando la medida de significación supera el valor umbral.

7. El aparato según las reivindicaciones 1 a 6, en donde el valor de coherencia para cada una de la pluralidad de subbandas es un valor de coherencia envolvente, en donde el valor de coherencia envolvente se determina en función de una coherencia entre canales entre dos o más señales de audio de canal.

8. El aparato según la reivindicación 7, en donde el valor de coherencia envolvente para cada una de la pluralidad de subbandas lo determina el aparato que comprende medios para:

- 5 calcular una matriz de covarianzas asociada con las señales de audio de dos o más canales;
monitorizar una señal de audio de canal con una energía máxima determinada en función de la
matriz de covarianza y un subconjunto de otras señales de audio de canal, en donde el subconjunto
es un número determinado entre 1 y uno menos que el número total de señales de audio de canal
con las siguientes energías más altas; y
generar el parámetro de covarianza envolvente en función de la selección del mínimo de
coherencias normalizadas determinadas entre la señal de audio de canal con la mayor energía y
cada una de las señales de audio de canal con las siguientes energías más altas.
- 10 9. El aparato según las reivindicaciones 1 a 8, en donde el valor de relación de energía para cada una de las
subbandas es un valor de relación entre energía directa y total para cada una de las subbandas.
10. El aparato según la reivindicación 9, en donde la relación de energía no direccional para una subbanda se
determina como uno menos la relación entre energía directa y total para la subbanda.
- 15 11. Un método para la codificación de audio espacial, que comprende:
- 20 recibir o determinar, para una o más señales de audio, parámetros de audio espacial por subbandas
para proporcionar una reproducción de audio espacial, en donde los parámetros de audio espacial
comprenden un valor de coherencia para cada subbanda de una pluralidad de subbandas de una
trama;
- 25 recibir un valor de relación de energía para cada una de la pluralidad de subbandas;
determinar una medida de significación para los valores de coherencia de la pluralidad de subbandas
de la trama; y
- 30 usar la medida de significación para determinar si se deben codificar los valores de coherencia de
la pluralidad de subbandas de la trama, **en donde el método está caracterizado porque**
determinar la medida de significación para los valores de coherencia de la pluralidad de subbandas
de la trama comprende:
- 35 determinar una proporción de energía no direccional coherente para cada una de la
pluralidad de subbandas, en donde la proporción de energía no direccional coherente para
una subbanda se determina como una relación de energía no direccional para la subbanda
multiplicada por el valor de coherencia para la subbanda, en donde la relación de energía
no direccional está relacionada con la relación de energía en la subbanda que queda
después de haber eliminado el valor de relación de energía para la subbanda; y
- 40 sumar, para toda la pluralidad de subbandas, la proporción determinada de energía no
direccional coherente.
- 40 12. El método reivindicado en la reivindicación 11, en donde determinar la medida de significación para los valores
de coherencia de la pluralidad de subbandas de la trama comprende además:
- 45 multiplicar, para cada subbanda, la proporción de energía no direccional coherente por la relación
de energía no direccional;
- 45 sumar, para toda la pluralidad de subbandas, el múltiplo de la proporción de energía no direccional
coherente y la relación de energía no direccional para cada subbanda; y
- 50 normalizar el múltiplo sumado de la proporción de energía no direccional coherente y la relación de
energía no direccional para cada subbanda mediante una suma de la relación de energía no
direccional para cada subbanda.
- 50 13. Un programa informático que comprende instrucciones de programa para hacer que un dispositivo de
procesamiento realice el método según una cualquiera de las reivindicaciones 11 y 12.

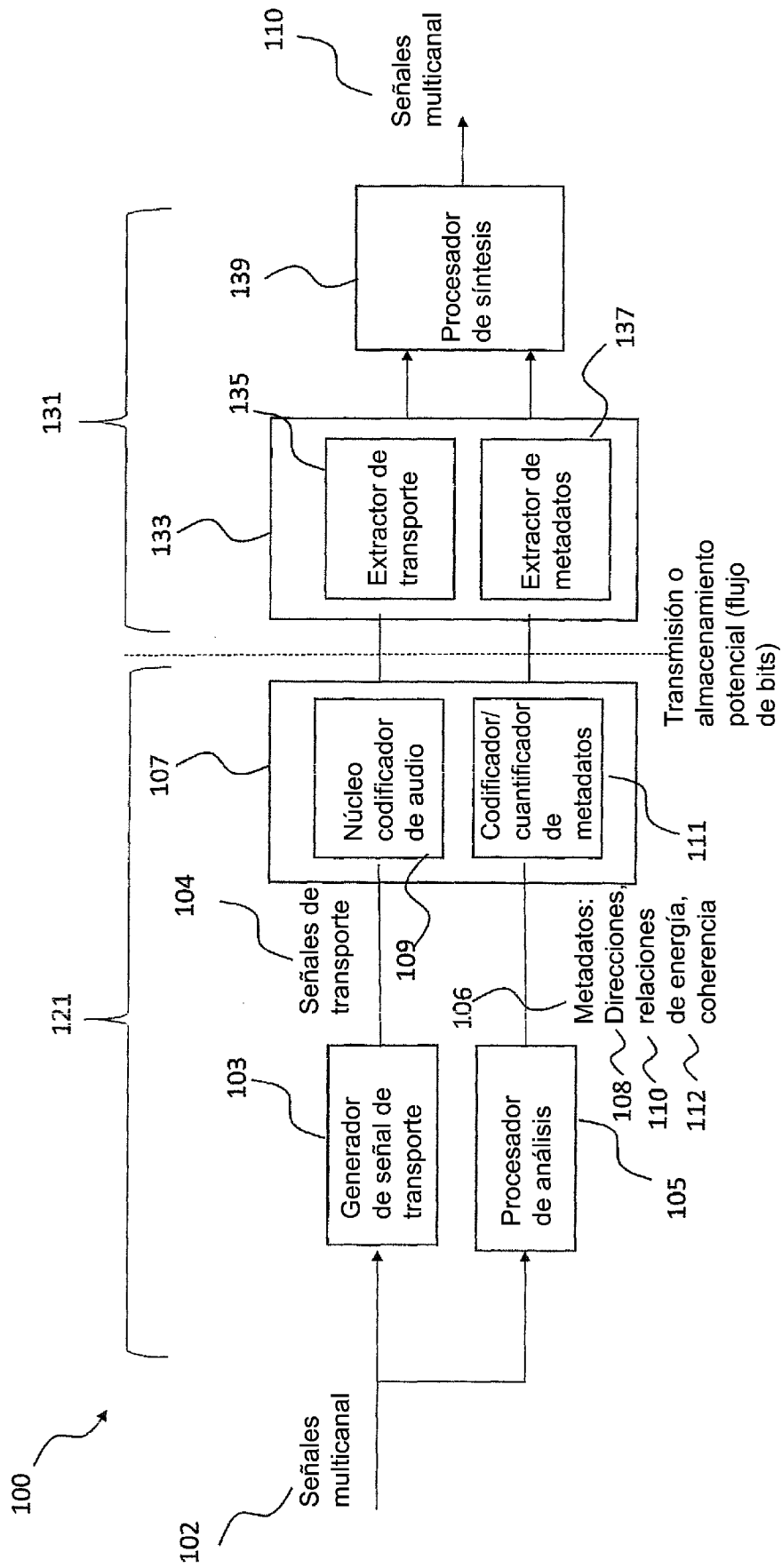


Figura 1

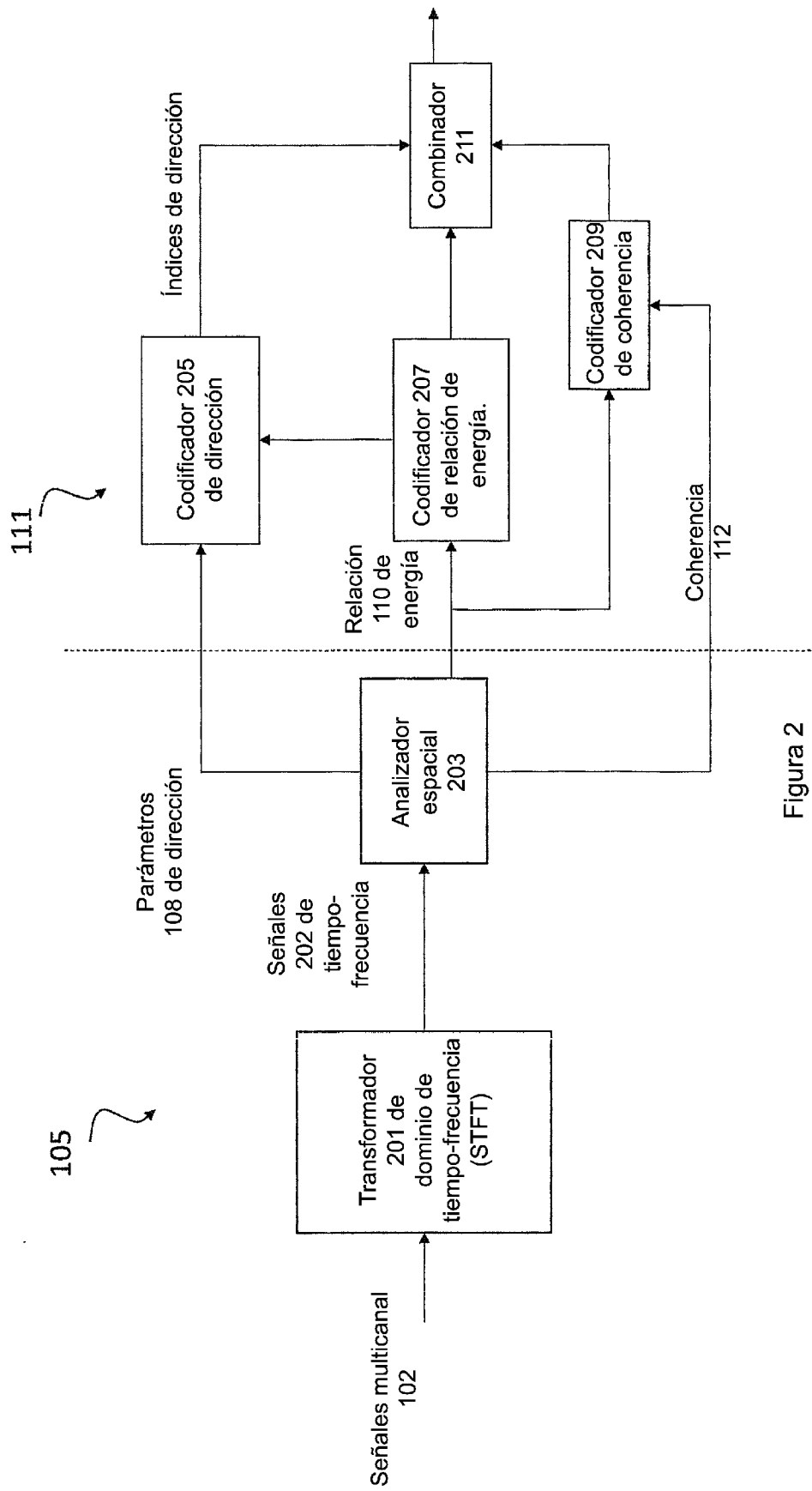


Figura 2

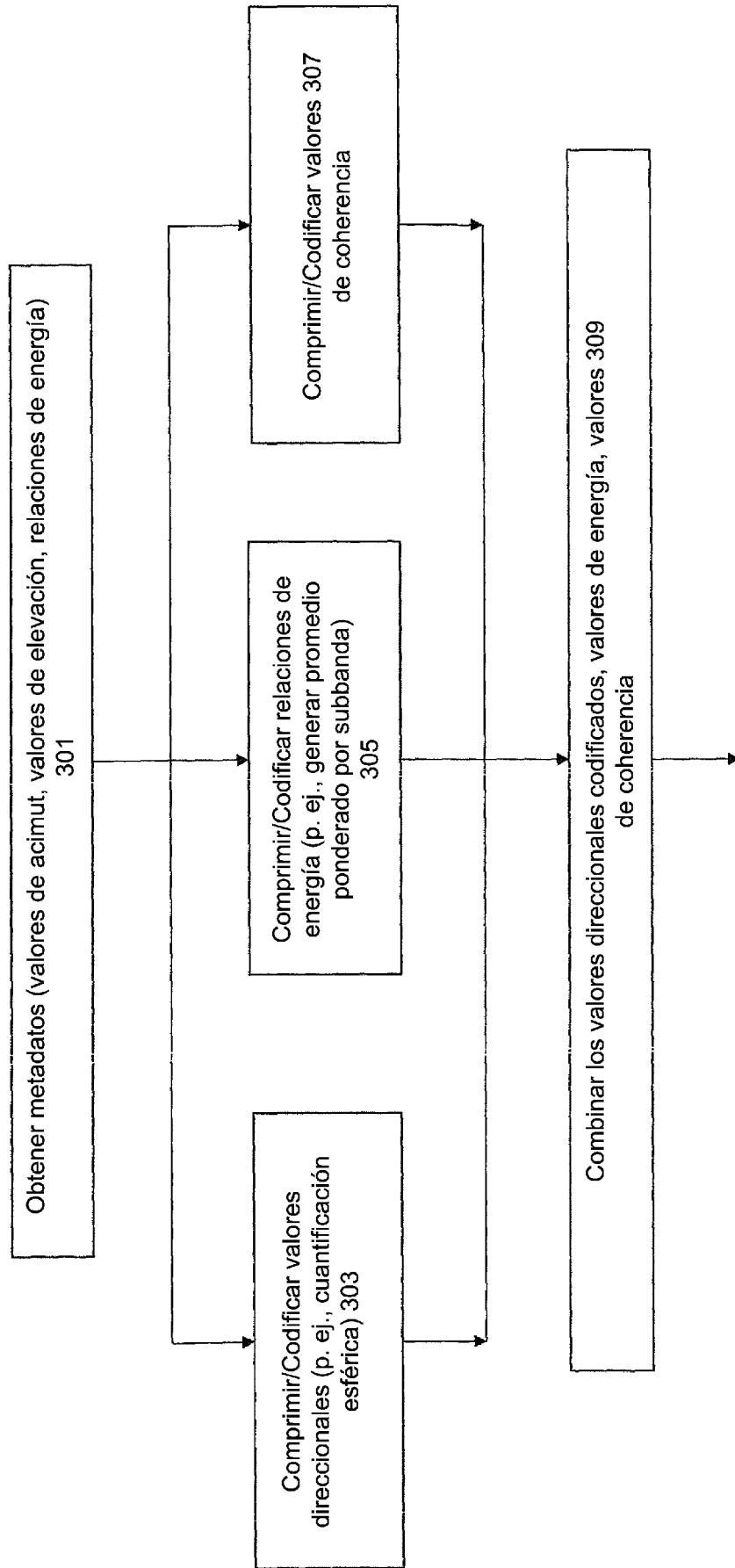


Figura 3

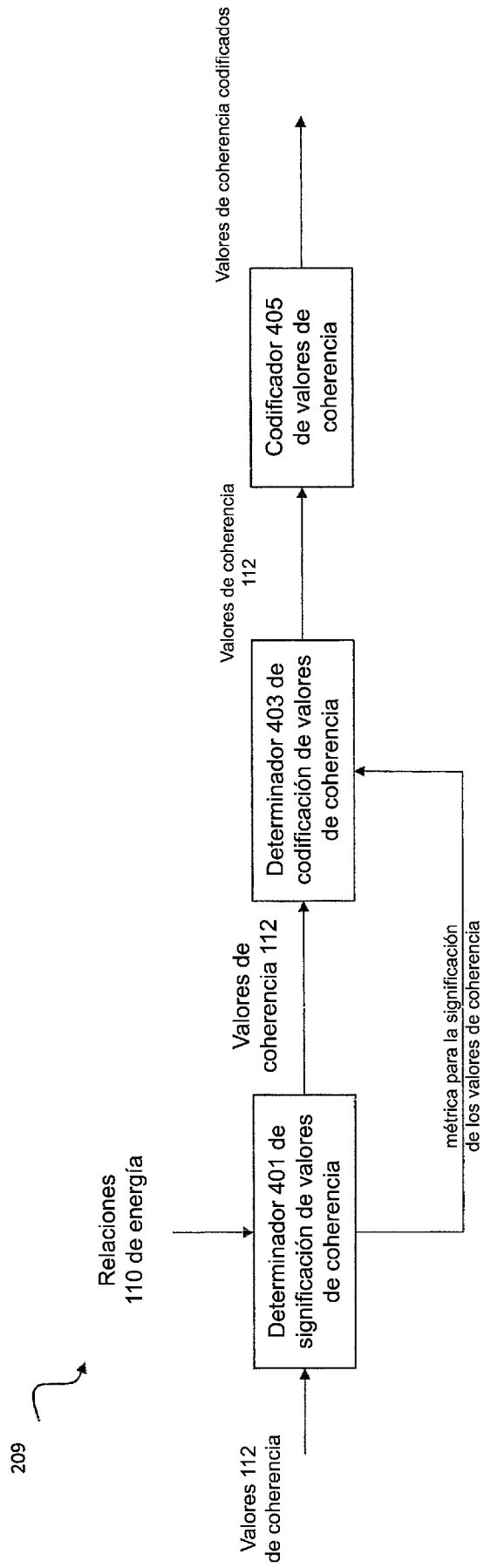


Figura 4

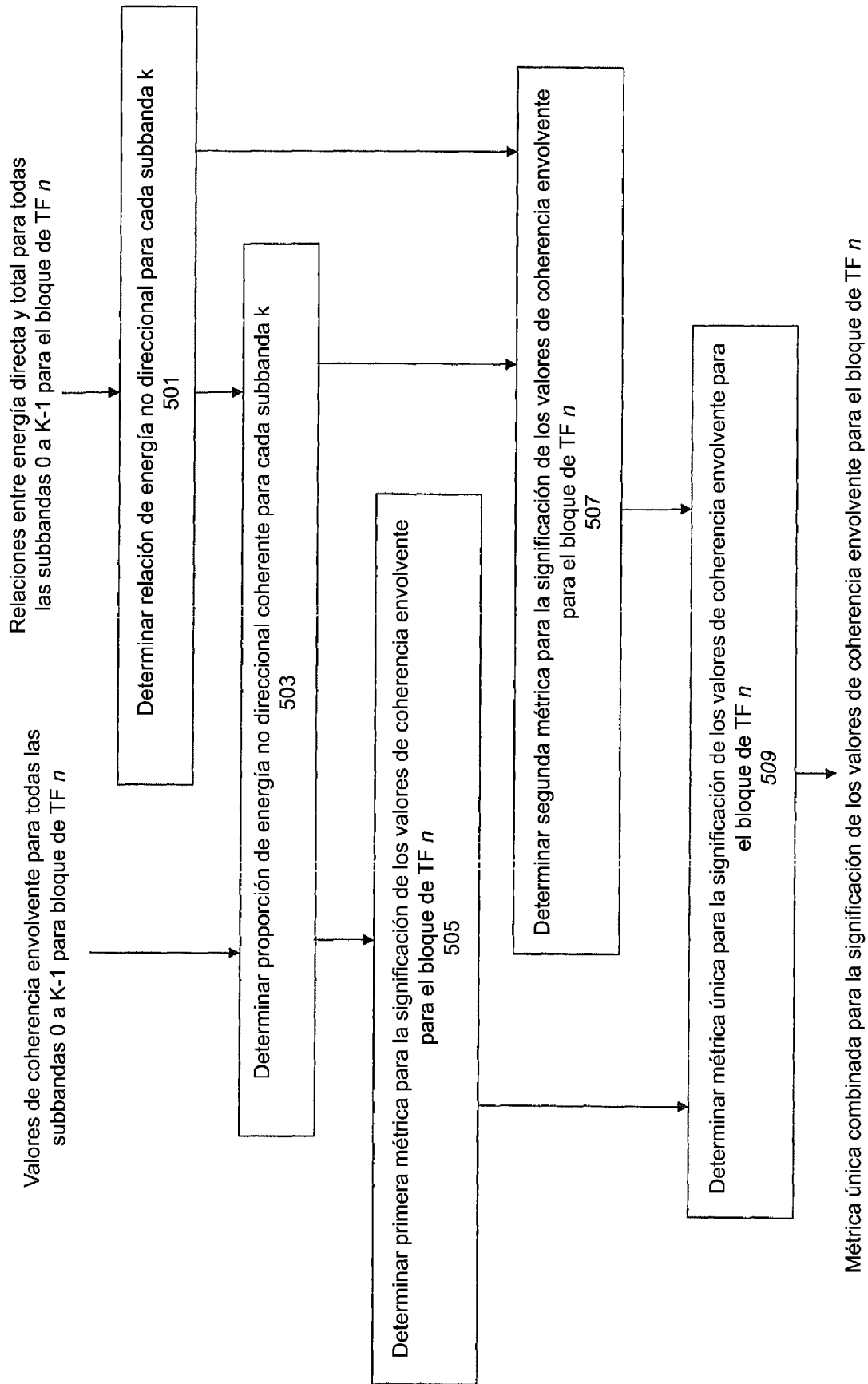


Figura 5

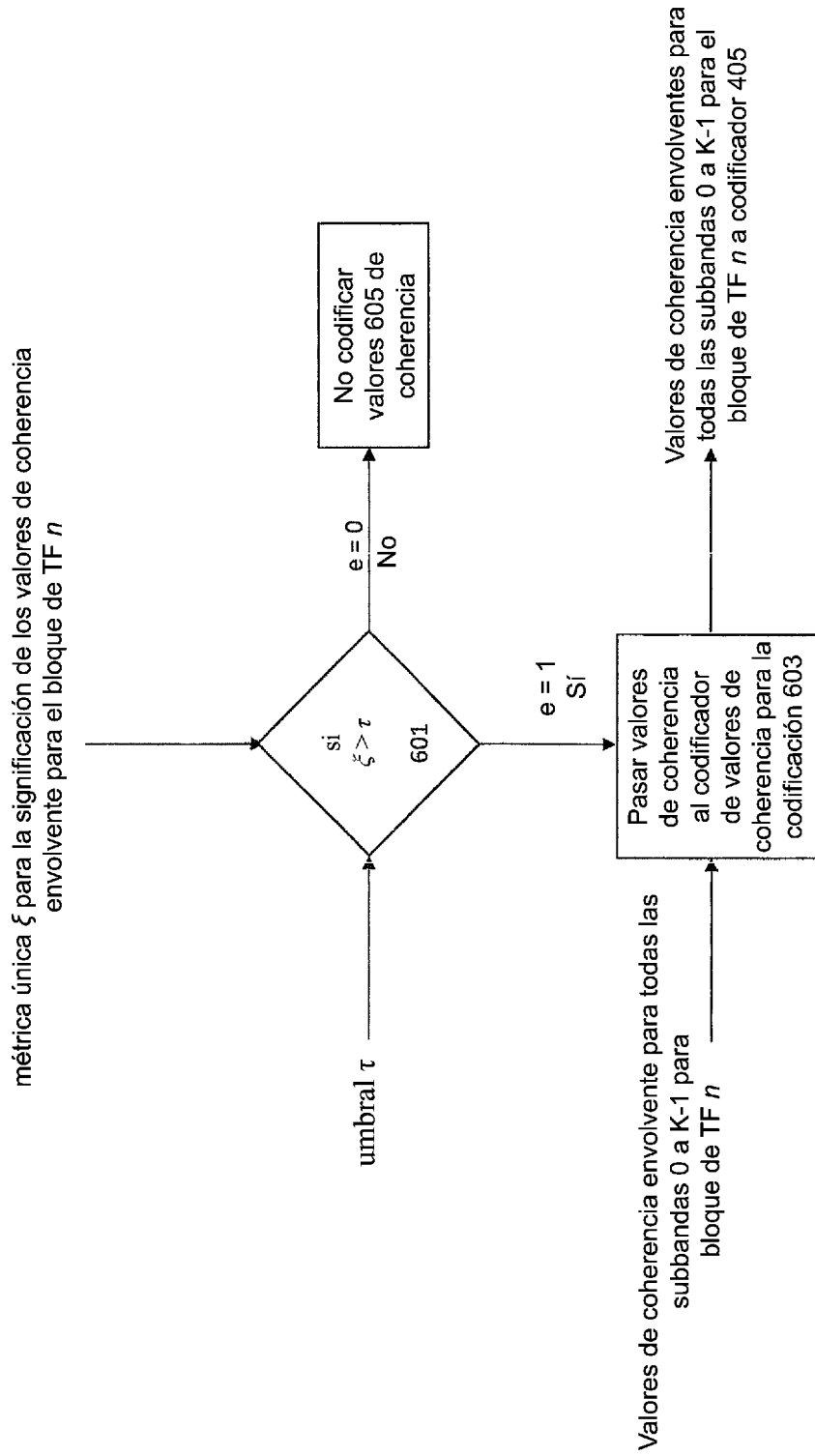


Figura 6

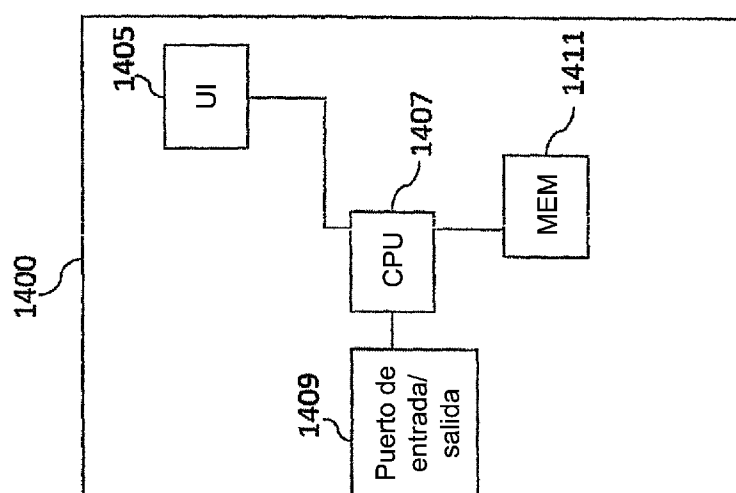


Figura 7