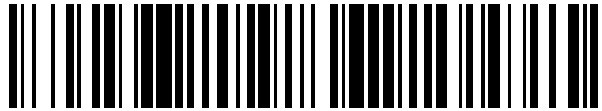


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 624 668**

51 Int. Cl.:

**G10L 19/20** (2013.01)

**G10L 19/008** (2013.01)

**H04S 3/02** (2006.01)

**H04S 5/00** (2006.01)

**H04S 7/00** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **23.05.2014 PCT/EP2014/060728**

87 Fecha y número de publicación internacional: **27.11.2014 WO14187987**

96 Fecha de presentación y número de la solicitud europea: **23.05.2014 E 14725734 (9)**

97 Fecha y número de publicación de la concesión europea: **29.03.2017 EP 3005352**

54 Título: **Codificación y descodificación de objetos de audio**

30 Prioridad:

**24.05.2013 US 201361827288 P**

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

**17.07.2017**

73 Titular/es:

**DOLBY INTERNATIONAL AB (100.0%)  
Apollo Building, 3E, Herikerbergweg 1-35  
1101 CN Amsterdam, NL**

72 Inventor/es:

**PURNHAGEN, HEIKO;  
VILLEMOES, LARS;  
SAMUELSSON, LEIF JONAS y  
HIRVONEN, TONI**

74 Agente/Representante:

**LEHMANN NOVO, María Isabel**

ES 2 624 668 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

Codificación y decodificación de objetos de audio

### Sector técnico

5 La invención de la presente memoria se refiere, en general, a codificación de audio. En particular, se refiere a la utilización y el cálculo de factores de ponderación para la descorrelación de objetos de audio en un sistema de codificación de audio.

### Antecedentes de la técnica

10 En los sistemas de audio convencionales, se utiliza un enfoque basado en canales. Cada canal puede representar, por ejemplo, el contenido de un altavoz o de un conjunto de altavoces. Los posibles esquemas de codificación para dichos sistemas incluyen codificación multicanal discreta o codificación paramétrica, tal como MPEG Surround.

15 Más recientemente se ha desarrollado un nuevo enfoque. Este enfoque está basado en objetos. En los sistemas que utilizan el enfoque basado en objetos, una escena de audio tridimensional es representada por objetos de audio con sus metadatos posicionales asociados. Estos objetos de audio se desplazan en la escena tridimensional durante la reproducción de la señal de audio. El sistema puede incluir además los denominados canales de base, que se pueden describir como objetos de audio estacionarios que están mapeados directamente a las posiciones del altavoz, por ejemplo, de un sistema de audio convencional tal como el descrito anteriormente. En el lado del descodificador de un sistema de este tipo, los canales de base/objetos se pueden reconstruir utilizando señales de mezcla descendente y una matriz de mezcla ascendente o reconstrucción, en la que en los canales de base/objetos se reconstruyen formando una combinación lineal de las señales de mezcla descendente en base al valor de los  
20 elementos correspondientes en la matriz de reconstrucción.

25 Un problema que puede surgir en un sistema de audio basado en objetos, en particular a tasas de bit objetivo bajas, es que la correlación entre los canales de base/objetos descodificados puede ser mayor de la que se tenía para los canales de base/objetos originales codificados. Un enfoque común para resolver dichos problemas, y para mejorar la reconstrucción de los objetos de audio, por ejemplo como en MPEG SAOC, es introducir elementos de descorrelación en el descodificador. En MPEG SAOC, la descorrelación introducida ayuda a restaurar una correlación correcta entre los objetos de audio dada una representación específica de los objetos de audio, es decir, en función de qué tipo de unidad de reproducción se conecta al sistema de audio.

30 El documento WO2010/149700 (Fraunhofer Ges Forschung) se refiere a un descodificador de señales de audio para proporcionar una representación de señal de mezcla ascendente que depende de una representación de señales de mezcla descendente y una información paramétrica relacionada con objetos comprende un separador de objetos configurado para descomponer la representación de señales de mezcla descendente, con el fin de proporcionar una primera información de audio que describe un primer conjunto de uno o varios objetos de audio de un primer tipo de objeto de audio y una segunda información de audio que describe un segundo conjunto de uno o varios objetos de audio de un segundo tipo de objeto de audio, dependiendo de la representación de las señales de mezcla descendente y utilizando por lo menos una parte de la información paramétrica relacionada con objetos.  
35

40 El documento WO 2008/069593 (LG Electronics Inc.) se refiere a un procedimiento para procesar una señal de audio, que comprende: recibir una señal de mezcla descendente, una primera información multicanal y una información de objetos; procesar la señal de mezcla descendente utilizando la información de objetos y una información de mezcla; y transmitir una de la primera información multicanal y una segunda información multicanal de acuerdo con la información de mezcla, en el que la segunda información multicanal es generada utilizando la información de objetos y la información de mezcla.

45 El documento "Changes for editorial consistency of SAOC FCD text" (Engdegård et al.) describe el modelo de referencia 1 (RM1, Reference Model 1) de la tecnología de codificación de objetos de audio espacial (SAOC, Spatial Audio Object Coding) que puede recrear, modificar y representar una serie de objetos de audio en base a un número menor de canales transmitidos y datos paramétricos adicionales.

50 Sin embargo, los procedimientos conocidos para sistemas de audio basados en objetos son sensibles al número de señales de mezcla descendente y al número de objetos/canales de base y pueden consistir además en una operación compleja que depende de la representación de los objetos de audio. Por lo tanto, existe la necesidad de procedimientos simples y flexibles para controlar la cantidad de descorrelación introducida en el descodificador en dichos sistemas, permitiendo de ese modo la reconstrucción mejorada de los objetos de audio.

### Breve descripción de los dibujos

A continuación se describirán realizaciones de ejemplo haciendo referencia a los dibujos adjuntos, en los cuales:

la figura 1 es un diagrama de bloques generalizado de un sistema de decodificación de audio de acuerdo con un ejemplo de realización;

la figura 2 muestra, a modo de ejemplo, un formato en el que una matriz de reconstrucción y un parámetro de ponderación son recibidos por el sistema de descodificación de audio de la figura 1;

5 la figura 3 es un diagrama de bloques generalizado de un codificador de audio para generar por lo menos un parámetro de ponderación para su utilización en un proceso de descorrelación en un sistema de descodificación de audio,

la figura 4 muestra, a modo de ejemplo, un diagrama de bloques generalizado de una parte del codificador de la figura 3 para generar dicho por lo menos un parámetro de ponderación,

las figuras 5a-5c muestran, a modo de ejemplo, funciones de mapeo utilizadas en la parte del codificador de la figura 4.

10 Todas las figuras son esquemáticas y generalmente muestran sólo las partes que son necesarias para explicar la invención, mientras que otras partes pueden estar omitidas o tan sólo sugeridas. Salvo que se indique lo contrario, los numerales de referencia similares se refieren a partes similares en las diferentes figuras.

### Descripción detallada

15 En vista de lo anterior, un objetivo es dar a conocer un codificador y un descodificador, y procedimientos asociados, que proporcionen un control menos complejo y más flexible de la descorrelación introducida, permitiendo de ese modo una reconstrucción mejorada de los objetos de audio.

#### I. Visión general - descodificador

20 Según un primer aspecto, las realizaciones de ejemplo proponen procedimientos de descodificación, descodificadores y productos de programa informático para descodificar. Los procedimientos, descodificadores y productos de programa informático propuestos pueden tener, en general, las mismas características y ventajas.

25 De acuerdo con realizaciones de ejemplo, se da a conocer un procedimiento para reconstruir una tesela de tiempo/frecuencia de N objetos de audio. El procedimiento comprende las etapas de: recibir M señales de mezcla descendente; recibir una matriz de reconstrucción que permite la reconstrucción de una aproximación de los N objetos de audio a partir de las M señales de mezcla descendente; aplicar la matriz de reconstrucción a las M  
30 señales de mezcla descendente para generar N objetos de audio aproximados; someter por lo menos un subconjunto de los N objetos de audio aproximados a un proceso de descorrelación para generar por lo menos un objeto de audio descorrelacionado, de manera que cada uno de dicho por lo menos un objeto de audio descorrelacionado corresponde a uno de los N objetos de audio aproximados; para cada uno de los N objetos de audio aproximados que no tiene un correspondiente objeto de audio descorrelacionado, reconstruir la tesela de tiempo/frecuencia del objeto de audio mediante el objeto de audio aproximado; y para cada uno de los N objetos de audio aproximados que tiene un correspondiente objeto de audio descorrelacionado, reconstruir la tesela de tiempo/frecuencia del objeto de audio mediante: recibir por lo menos un parámetro de ponderación que representa un primer factor de ponderación y un segundo factor de ponderación, ponderar el objeto de audio aproximado mediante el primer factor de ponderación, ponderar el objeto de audio descorrelacionado correspondiente al objeto  
35 de audio aproximado mediante el segundo factor de ponderación, y combinar el objeto de audio aproximado ponderado con el correspondiente objeto de audio descorrelacionado ponderado.

40 Los sistemas de codificación/descodificación de audio dividen habitualmente el espacio de tiempo-frecuencia en teselas de tiempo/frecuencia, por ejemplo aplicando bancos de filtros adecuados a las señales de audio de entrada. Una tesela de tiempo/frecuencia significa generalmente una parte de un espacio de tiempo-frecuencia correspondiente un intervalo de tiempo y a una sub-banda de frecuencia. El intervalo de tiempo puede corresponder habitualmente a la duración de una trama de tiempo utilizada en el sistema de codificación/descodificación de audio. La sub-banda de frecuencia puede corresponder habitualmente a una o varias sub-bandas de frecuencia contiguas definidas por un banco de filtros utilizado en el sistema de codificación/descodificación. En caso de que la sub-banda de frecuencia corresponda a varias sub-bandas de frecuencia contiguas definidas por el banco de filtros, esto  
45 permite tener sub-bandas de frecuencia no uniformes en el proceso de descodificación de la señal de audio, por ejemplo sub-bandas de frecuencia más anchas para frecuencias mayores de la señal de audio. En un caso de banda ancha, en el que el sistema de codificación/descodificación de audio funciona sobre todo el intervalo de frecuencias, la sub-banda de frecuencia de la tesela de tiempo/frecuencia puede corresponder a todo el intervalo de frecuencias. El procedimiento anterior da a conocer las etapas para reconstruir dicha tesela de tiempo/frecuencia de N objetos de audio. Sin embargo, se debe entender que el procedimiento se puede repetir para cada tesela de tiempo/frecuencia del sistema de descodificación de audio. Se debe entender asimismo que se pueden codificar simultáneamente varias teselas de tiempo/frecuencia. Habitualmente, las teselas de tiempo/frecuencia contiguas pueden solapar un poco en tiempo y/o frecuencia. Por ejemplo, un solape en el tiempo puede ser equivalente a una interpolación lineal de los elementos de la matriz de reconstrucción en el tiempo, es decir desde un intervalo de  
50 tiempo al siguiente. Sin embargo, esta invención se dirige a otras partes del sistema de codificación/descodificación, y cualquier solape en tiempo y/o frecuencia entre teselas de tiempo/frecuencia contiguas se deja para su implementación por el experto en la materia.

Tal como se utiliza en la presente memoria, una señal de mezcla descendente es una señal que es una combinación de uno o varios canales de base y/u objetos de audio.

El procedimiento anterior da a conocer un procedimiento flexible y simple para reconstruir una tesela de tiempo/frecuencia de N objetos de audio, donde se reduce cualquier correlación no deseada entre los N objetos de audio aproximados. Al utilizar dos factores de ponderación, uno para el objeto de audio aproximado y otro para el objeto de audio descorrelacionado, se consigue una parametrización simple que permite un control flexible de la cantidad de descorrelación que se introduce.

Además, la parametrización simple en el procedimiento no depende de a qué tipo de representación se someten los objetos de audio reconstruidos. Una ventaja de esto es que se utiliza el mismo procedimiento independientemente de qué tipo de unidad de reproducción está conectada al sistema de descodificación de audio que implementa el procedimiento, conduciendo por lo tanto un sistema de descodificación de audio menos complejo.

De acuerdo con una realización, para cada uno de los N objetos de audio aproximados que tiene un correspondiente objeto de audio descorrelacionado, dicho por lo menos un parámetro de ponderación comprende un solo parámetro de ponderación a partir del cual se puede obtener el primer factor de ponderación y el segundo factor de ponderación.

Una ventaja de esto es que se propone una parametrización simple para controlar la cantidad de descorrelación introducida en el sistema de descodificación de audio. Este enfoque utiliza un único parámetro que describe la mezcla de contribuciones "secas" (no descorrelacionadas) y "húmedas" (descorrelacionadas) por objeto y tesela de tiempo/frecuencia. Al utilizar un solo parámetro, se puede reducir la tasa de bits requerida, en comparación con utilizar varios parámetros, por ejemplo uno que describe la contribución húmeda y uno que describe la contribución seca.

De acuerdo con una realización, la suma cuadrática del primer factor de ponderación y el segundo factor de ponderación es igual a uno. En este caso, el único parámetro de ponderación comprende cualquiera del primer factor de ponderación o el segundo factor de ponderación. Ésta puede ser una manera simple de implementar un solo factor de ponderación para describir la mezcla de contribuciones seca y húmeda, por objeto y tesela de tiempo/frecuencia. Además, esto significa que el objeto reconstruido tendrá la misma energía que el objeto aproximado.

De acuerdo con una realización, la etapa de someter por lo menos un subconjunto de los N objetos de audio aproximados a un proceso de descorrelación comprende someter cada uno de los N objetos de audio aproximados a un proceso de descorrelación, de manera que cada uno de los N objetos de audio aproximados corresponde a un objeto de audio descorrelacionado. Esto puede reducir adicionalmente cualquier correlación no deseada entre los objetos de audio reconstruidos, dado que todos los objetos de audio reconstruidos se basan tanto en un objeto de audio descorrelacionado como en un objeto de audio aproximado.

De acuerdo con una realización, el primer y el segundo factores de ponderación varían con el tiempo y la frecuencia. Por consiguiente, se puede aumentar la flexibilidad del sistema de descodificación de audio dado que se puede introducir una cantidad de descorrelación diferente para diferentes teselas de tiempo/frecuencia. Esto puede asimismo reducir adicionalmente cualquier correlación no deseada entre los objetos de audio reconstruidos y mejorar la calidad de los objetos de audio reconstruidos.

De acuerdo con una realización, la matriz de reconstrucción varía con el tiempo y la frecuencia. De este modo, se incrementa la flexibilidad del sistema de descodificación de audio dado que los parámetros utilizados para reconstruir o aproximar los objetos de audio a partir de las señales de mezcla descendente pueden variar para diferentes teselas de tiempo/frecuencia.

De acuerdo con otra realización, tras su recepción, la matriz de reconstrucción y dicho por lo menos un parámetro de ponderación se disponen en una trama. La matriz de reconstrucción se dispone en un primer campo de la trama utilizando un primer formato y dicho por lo menos un parámetro de ponderación se dispone en un segundo campo de la trama utilizando un segundo formato, permitiendo de ese modo que un descodificador que soporta solamente el primer formato descodifique la matriz de reconstrucción en el primer campo y deseche dicho por lo menos un parámetro de ponderación en el segundo campo. Por lo tanto, se puede conseguir compatibilidad con un descodificador que no implemente descorrelación.

De acuerdo con una realización, el procedimiento puede comprender además recibir L señales auxiliares, donde la matriz de reconstrucción permite además la reconstrucción de la aproximación de los N objetos de audio a partir de las M señales de mezcla descendente y las L señales auxiliares, y donde el procedimiento comprende además aplicar la matriz de reconstrucción a las M señales de mezcla descendente y las L señales auxiliares para generar los N objetos de audio aproximados. Las L señales auxiliares pueden incluir, por ejemplo, por lo menos una señal auxiliar L que es igual a uno de los N objetos de audio a reconstruir. Esto puede aumentar la calidad del objeto de audio reconstruido específico. Esto puede ser ventajoso en el caso en que uno de los N objetos de audio a reconstruir representa una parte de la señal de audio que es de importancia específica, por ejemplo un objeto de audio que representa la voz del locutor en un documental. De acuerdo con una realización, por lo menos una de las

L señales auxiliares es una combinación de por lo menos dos de los N objetos de audio a reconstruir, proporcionando de ese modo un compromiso entre tasa de bits y calidad.

De acuerdo con una realización, las M señales de mezcla descendente comprenden un hiperplano, y donde por lo menos una de las L señales auxiliares no está situada en el hiperplano comprendido por las M señales de mezcla descendente. De este modo, una o varias de las L señales auxiliares pueden representar dimensiones de señal que no están incluidas en ninguna de las M señales de mezcla descendente. Por consiguiente, puede aumentar la calidad de los objetos de audio reconstruidos. En una realización, por lo menos una de las L señales auxiliares es ortogonal al hiperplano comprendido por las M señales de mezcla descendente. De este modo, toda la señal de una o varias de las L señales auxiliares representa partes de la señal de audio no incluidas en ninguna de las M señales de mezcla descendente. Esto puede aumentar la calidad de los objetos de audio reconstruidos y, al mismo tiempo, reducir la tasa de bits requerida, dado que dicha por lo menos una de las L señales auxiliares no incluye ninguna información ya presente en cualquiera de las M señales de mezcla descendente.

De acuerdo con realizaciones de ejemplo, se da a conocer un medio legible por ordenador que comprende instrucciones de código informático adaptadas para llevar a cabo cualquier procedimiento del primer aspecto cuando son ejecutadas en un dispositivo con capacidad de procesamiento.

De acuerdo con realizaciones de ejemplo, se da a conocer un aparato para reconstruir una tesela de tiempo/frecuencia de N objetos de audio, que comprende: un primer componente de recepción configurado para recibir M señales de mezcla descendente; un segundo componente de recepción configurado para recibir una matriz de reconstrucción que permite la reconstrucción de una aproximación de los N objetos de audio a partir de las M señales de mezcla descendente; un componente de aproximación de objetos de audio dispuesto más abajo del primer y del segundo componentes de recepción y configurado para aplicar la matriz de reconstrucción a las M señales de mezcla descendente con el fin de generar N objetos de audio aproximados; un componente de descorrelación dispuesto más abajo del componente de aproximación de objetos de audio y configurado para someter por lo menos un subconjunto de los N objetos de audio aproximados a un proceso de descorrelación con el fin de generar por lo menos un objeto de audio descorrelacionado, de manera que cada uno de dicho por lo menos un objeto de audio descorrelacionado corresponde a uno de los N objetos de audio aproximados; estando el segundo componente de recepción configurado además para recibir, para cada uno de los N objetos de audio aproximados que tiene un correspondiente objeto de audio descorrelacionado, por lo menos un parámetro de ponderación que representa un primer factor de ponderación y un segundo factor de ponderación; y un componente de reconstrucción de objetos de audio dispuesto más abajo del componente de aproximación de objetos de audio, del componente de descorrelación y del segundo componente de recepción, y configurado para: para cada uno de los N objetos de audio aproximados que no tiene un correspondiente objeto de audio descorrelacionado, reconstruir la tesela de tiempo/frecuencia del objeto de audio mediante el objeto de audio aproximado; y para cada uno de los N objetos de audio aproximados que tiene un correspondiente objeto de audio descorrelacionado, reconstruir la tesela de tiempo/frecuencia del objeto de audio mediante: ponderar el objeto de audio aproximado mediante el primer factor de ponderación; ponderar el objeto de audio descorrelacionado correspondiente al objeto de audio aproximado mediante el segundo factor de ponderación; y combinar el objeto de audio aproximado ponderado con el correspondiente objeto de audio descorrelacionado ponderado.

## II. Visión general - codificador

De acuerdo con un segundo aspecto, las realizaciones de ejemplo proponen procedimientos de codificación, codificadores y productos de programa informático para codificar. Los procedimientos, codificadores y productos de programa informático propuestos pueden tener, en general, las mismas características y ventajas.

De acuerdo con realizaciones de ejemplo, se da a conocer un procedimiento en un codificador para generar por lo menos un parámetro de ponderación, en el que dicho por lo menos un parámetro de ponderación se tiene que utilizar en un descodificador cuando se reconstruye una tesela de tiempo/frecuencia de un objeto de audio específico combinando una aproximación ponderada del lado del descodificador del objeto de audio específico con una correspondiente versión descorrelacionada ponderada del objeto de audio específico aproximado del lado del descodificador, comprendiendo el procedimiento las etapas de: recibir M señales de mezcla descendente que son combinaciones de por lo menos N objetos de audio que incluyen el objeto de audio específico; recibir el objeto de audio específico; calcular una primera cantidad indicativa de un nivel de energía del objeto de audio específico; calcular una segunda cantidad indicativa de un nivel de energía correspondiente a un nivel de energía de una aproximación del lado del codificador del objeto de audio específico, siendo la aproximación del lado del codificador una combinación de las M señales de mezcla descendente; calcular dicho por lo menos un parámetro de ponderación en base a la primera y la segunda cantidades.

El procedimiento anterior da a conocer las etapas para generar por lo menos un parámetro de ponderación para un objeto de audio específico durante una tesela de tiempo/frecuencia. Sin embargo, se debe entender que el procedimiento se puede repetir para cada tesela de tiempo/frecuencia del sistema de codificación/descodificación de audio y para cada objeto de audio.

Se debe entender que la teselación, es decir, dividir la señal de audio/objeto en teselas de tiempo/frecuencia, en un sistema de codificación de audio no tiene que ser igual que la teselación en un sistema de descodificación de audio.

Se debe entender asimismo que la aproximación del lado del descodificador del objeto de audio específico y la aproximación del lado del codificador del audio específico pueden ser aproximaciones diferentes o pueden ser la misma aproximación.

5 Para disminuir la tasa de bits necesaria y para reducir la complejidad, dicho por lo menos un parámetro de ponderación puede comprender un único parámetro de ponderación a partir del cual se puede obtener un primer factor de ponderación y un segundo factor de ponderación, el primer factor de ponderación para la ponderación de la aproximación del lado del descodificador del objeto de audio específico y el segundo factor de ponderación para la ponderación de la versión descorrelacionada del objeto de audio aproximado del lado del descodificador.

10 Para impedir que se añada energía a un objeto de audio reconstruido en el lado del descodificador, el objeto de audio reconstruido que comprende la aproximación del lado del descodificador del audio específico y la versión descorrelacionada del objeto de audio aproximado del lado del descodificador, la suma cuadrática del primer factor de ponderación y del segundo factor de ponderación puede ser igual a uno. En este caso, el único parámetro de ponderación puede comprender cualquiera del primer factor de ponderación o el segundo factor de ponderación.

15 De acuerdo con una realización, la etapa de calcular por lo menos un parámetro de ponderación comprende comparar la primera cantidad con la segunda cantidad. Por ejemplo, se puede comparar la energía del objeto de audio específico aproximado y la energía del objeto de audio específico.

20 De acuerdo con realizaciones de ejemplo, la comparación de la primera cantidad y la segunda cantidad comprende calcular la relación entre la segunda y la primera cantidad, elevar la relación a la potencia de  $\alpha$  y utilizar la relación elevada a la potencia de  $\alpha$  para calcular el parámetro de ponderación. Esto puede aumentar la flexibilidad del codificador. El parámetro  $\alpha$  puede ser igual a dos.

De acuerdo con realizaciones de ejemplo, la relación elevada a la potencia de  $\alpha$  es sometida a una función creciente que mapea la relación elevada a la potencia de  $\alpha$  a dicho por lo menos un parámetro de ponderación.

De acuerdo con realizaciones de ejemplo, el primer y el segundo factores de ponderación varían con el tiempo y la frecuencia.

25 De acuerdo con realizaciones de ejemplo, la segunda cantidad indicativa de un nivel de energía corresponde a un nivel de energía de una aproximación del lado del codificador del objeto de audio específico, siendo la aproximación del lado del codificador una combinación lineal de las M señales de mezcla descendente y las L señales auxiliares, estando las señales de mezcla descendente y las señales auxiliares formadas a partir de los N objetos de audio. Para mejorar la reconstrucción del objeto de audio en el lado del descodificador, se pueden incluir señales auxiliares en el sistema de codificación/descodificación de audio.

30 De acuerdo con una realización a modo de ejemplo, por lo menos una de las L señales auxiliares puede corresponder a objetos de audio particularmente importantes, tal como un objeto de audio que representa un diálogo. Por lo tanto, por lo menos una de las L señales auxiliares puede ser igual a uno de los N objetos de audio. De acuerdo con otras realizaciones, por lo menos una de las L señales auxiliares es una combinación de por lo menos dos de los N objetos de audio.

35 De acuerdo con realizaciones, las M señales de mezcla descendente comprenden un hiperplano, y en el que por lo menos una de las L señales auxiliares no está situada en el hiperplano comprendido por las M señales de mezcla descendente. Esto significa que por lo menos una de las L señales auxiliares representan dimensiones de señal de los objetos de audio que se han perdido en el proceso de generación de las M señales de mezcla descendente, lo que puede mejorar la construcción del objeto de audio en el lado del descodificador. De acuerdo con otras realizaciones, dicha por lo menos una de las L señales auxiliares es ortogonal al hiperplano comprendido por las M señales de mezcla descendente.

40 De acuerdo con realizaciones de ejemplo, se da a conocer un medio legible por ordenador que comprende instrucciones de código informático adaptadas para llevar a cabo cualquier procedimiento del segundo aspecto cuando son ejecutadas en un dispositivo con capacidad de procesamiento.

45 De acuerdo con una realización, se da a conocer un codificador para generar por lo menos un parámetro de ponderación, en el que dicho por lo menos un parámetro de ponderación se tiene que utilizar en un descodificador cuando se reconstruye una tesela de tiempo/frecuencia de un objeto de audio específico combinando una aproximación ponderada del lado del descodificador del objeto de audio específico con una correspondiente versión descorrelacionada ponderada del objeto de audio específico aproximado del lado del descodificador, comprendiendo el aparato: un componente de recepción configurado para recibir M señales de mezcla descendente que son combinaciones de por lo menos N objetos de audio que incluyen el objeto de audio específico, estando el componente de recepción configurado además para recibir el objeto de audio específico; una unidad de cálculo configurada para: calcular una primera cantidad indicativa de un nivel de energía del objeto de audio específico;

50

55 calcular una segunda cantidad indicativa de un nivel de energía correspondiente a un nivel de energía de una aproximación del lado del codificador del objeto de audio específico, siendo la aproximación del lado del codificador

una combinación de las M señales de mezcla descendente; calcular dicho por lo menos un parámetro de ponderación en base a la primera y la segunda cantidades.

### Realizaciones de ejemplo

5 La figura 1 muestra un diagrama de bloques generalizado de un sistema de descodificación de audio 100 para reconstruir N objetos de audio. El sistema de descodificación de audio 100 lleva a cabo un procesamiento resuelto en tiempo/frecuencia, lo que significa que funciona sobre teselas de tiempo/frecuencia individuales para reconstruir los N objetos de audio. A continuación, se describirá el proceso del sistema 100 para reconstruir una tesela de tiempo/frecuencia de los N objetos de audio. Los N objetos de audio pueden ser uno o varios objetos de audio.

10 El sistema 100 comprende un primer componente de recepción 102 configurado para recibir M señales de mezcla descendente 106. Las M señales de mezcla descendente pueden ser una o varias señales de mezcla descendente. Las M señales de mezcla descendente 106 pueden ser, por ejemplo, una señal envolvente 5.1 ó 7.1 que es retrocompatible con sistemas de descodificación de sonido consolidados, tales como Dolby Digital Plus, MPEG o AAC. En otras realizaciones, las M señales de mezcla descendente 106 no son retrocompatibles. La señal de entrada al primer componente de recepción 102 puede ser un flujo de bits 130 a partir del cual el componente de recepción puede extraer las M señales de mezcla descendente 106.

15 El sistema 100 comprende además un segundo componente de recepción 112 configurado para recibir una matriz de reconstrucción 104 que permite la reconstrucción de una aproximación de los N objetos de audio a partir de las M señales de mezcla descendente 106. La matriz de reconstrucción 104 se puede denominar asimismo una matriz de mezcla ascendente. La señal de entrada 126 al segundo componente de recepción 112 puede ser un flujo de bits 126 a partir del cual el componente de recepción puede extraer la matriz de reconstrucción 104 o elementos de la misma e información adicional, lo que se explicará en detalle a continuación. En algunas realizaciones del sistema de descodificación de audio 100, el primer componente de recepción 102 y el segundo componente de recepción 112 están combinados en un único componente de recepción. En algunas realizaciones, las señales de entrada 130, 126 están combinadas en una única señal de entrada que puede ser un flujo de bits con un formato que permite que los componentes de recepción 102, 112 extraigan la diferente información a partir de dicha única señal de entrada.

20 El sistema 100 puede comprender además un componente de aproximación de objetos de audio 108 dispuesto más abajo del primer 102 y el segundo 112 componentes de recepción, y configurado para aplicar la matriz de reconstrucción 104 a las M señales de mezcla descendente 106 con el fin de generar N objetos de audio aproximados 110. Más específicamente, el componente de aproximación de objetos de audio 108 puede llevar a cabo una operación matricial en la que la matriz de reconstrucción 104 se multiplica por un vector que comprende las M señales de mezcla descendente. La matriz de reconstrucción 104 puede ser variable con el tiempo y la frecuencia, es decir, el valor de los elementos en la matriz de reconstrucción 104 puede diferir para cada tesela de tiempo/frecuencia. Por lo tanto, los elementos de la matriz de reconstrucción 104 dependen de qué tesela de tiempo/frecuencia se esté procesando actualmente.

25 Un objeto de audio  $\hat{S}_n(k, l)$  aproximado n a la frecuencia k y en el intervalo de tiempo l, es decir, una tesela de tiempo/frecuencia, se calcula por ejemplo en el componente de aproximación de objetos de audio 108, por ejemplo

30 mediante  $\hat{S}_n(k, l) = \sum_{m=1}^M c_{m,b,n} Y_m(k, l)$  para todas las muestras de frecuencia k en la banda de frecuencia b,  $b = 1, \dots, B$ , donde  $c_{m,b,n}$  es el coeficiente de reconstrucción del objeto n en la banda de frecuencia b y asociado con el canal de mezcla descendente  $Y_m$ . Se debe observar que se supone que el coeficiente de reconstrucción  $c_{m,b,n}$  es constante para la tesela de tiempo/frecuencia, pero en otras realizaciones el coeficiente puede variar durante la tesela de tiempo/frecuencia.

35 El sistema 100 comprende además un componente de descorrelación de 118 dispuesto más abajo del componente de aproximación de objetos de audio 108. El componente de descorrelación 118 está configurado para someter por lo menos un subconjunto 140 de los N objetos de audio aproximados 110 a un proceso de descorrelación para generar por lo menos un objeto de audio descorrelacionado 136. En otras palabras, la totalidad o tan sólo una parte de los N objetos de audio aproximados 110 pueden ser sometidos a un proceso de descorrelación. Cada uno de dicho por lo menos un objeto de audio descorrelacionado 136 corresponde a uno de los N objetos de audio aproximados 110. Para ser más precisos, el conjunto de objetos de audio descorrelacionados 136 corresponde al conjunto 140 de objetos de audio aproximados que se introduce en el proceso de descorrelación 118. El objetivo de dicho por lo menos un objeto de audio descorrelacionado 136 es reducir la correlación no deseada entre los N objetos de audio aproximados 110. Esta correlación no deseada se puede presentar, en particular, a bajas tasas de bits objetivo de un sistema de audio que comprende el sistema de descodificación de audio 100. A bajas tasas de bits objetivo, la matriz de reconstrucción puede ser escasa. Esto significa que muchos de los elementos de la matriz de reconstrucción pueden ser cero. En este caso, un objeto de audio aproximado particular 110 puede estar basado en una única señal de mezcla descendente o en unas pocas señales de mezcla descendente a partir de las M señales de mezcla descendente 106, incrementando por lo tanto el riesgo de introducir correlación no deseada entre los objetos de audio aproximados 110. De acuerdo con algunas realizaciones, cada uno de los N objetos de audio aproximados 110 son sometidos a un proceso de descorrelación mediante el componente de descorrelación 118, de

tal modo que cada uno de los N objetos de audio aproximados 110 corresponde a un objeto de audio descorrelacionado 136.

5 Cada uno de los N objetos de audio aproximados 110 sometido al proceso de descorrelación por el componente de descorrelación 118 puede ser sometido a un proceso de descorrelación diferente, por ejemplo aplicando un filtro de ruido blanco al objeto de audio aproximado que está siendo descorrelacionado o aplicando cualquier otro proceso de descorrelación adecuado, tal como un filtro de todo paso.

10 Se pueden encontrar ejemplos de otros procesos de descorrelación en la herramienta de codificación MPEG Parametric Stereo (utilizada en HE-AAC v2, tal como se describe en ISO/IEC 14496-3 y en el documento de J. Engdegård, H. Purnhagen, J. Rödén, L. Liljeryd, "Synthetic ambience in parametric stereo coding," en la 116 Convención AES, Berlín, DE, mayo de 2004), en MPEG Surround (ISO/IEC 23003-1) y en MPEG SAOC (ISO/IEC 23003-2).

Para no introducir correlación no deseada, los diferentes procesos de descorrelación están descorrelacionados entre sí. Según otras realizaciones, varios o todos los objetos de audio aproximados 110 son sometidos al mismo proceso de descorrelación.

15 El sistema 100 comprende además un componente 128 de reconstrucción de objetos de audio. El componente 128 de reconstrucción de objetos está dispuesto más abajo del componente de aproximación de objetos de audio 108, del componente de descorrelación 118 y del segundo componente de recepción 112. El componente 128 de reconstrucción de objetos está configurado para, para cada uno de los N objetos de audio aproximados 138 que no tiene un correspondiente objeto de audio descorrelacionado 136, reconstruir la tesela de tiempo/frecuencia del objeto de audio 142 mediante el objeto de audio aproximado 138. En otras palabras, si un determinado objeto de audio aproximado 138 no ha sido sometido a un proceso de descorrelación, simplemente se reconstruye como el objeto de audio aproximado 110 proporcionado por el componente de aproximación de objetos de audio 108. El componente 128 de reconstrucción de objetos está configurado además para, para cada uno de los N objetos de audio aproximados 110 que tiene un correspondiente objeto de audio descorrelacionado 136, reconstruir la tesela de tiempo/frecuencia del objeto de audio utilizando tanto el objeto de audio descorrelacionado 136 como el correspondiente objeto de audio aproximado 110.

20 Para facilitar este proceso, el segundo componente de recepción 112 está configurado además para recibir, para cada uno de los N objetos de audio aproximados 110 que tiene un correspondiente objeto de audio descorrelacionado 136, por lo menos un parámetro de ponderación 132. Dicho por lo menos un parámetro de ponderación 132 representa un primer factor de ponderación 116 y un segundo factor de ponderación 114. El primer factor de ponderación 116, denominado asimismo factor seco, y el segundo factor de ponderación 114, denominado asimismo factor húmedo, se obtienen mediante un extractor húmedo/seco 134 a partir de dicho por lo menos un parámetro de ponderación 132. El primer y/o el segundo factores de ponderación 116, 114 pueden variar con el tiempo y la frecuencia, es decir, el valor de los factores de ponderación 116, 114 puede diferir para cada tesela de tiempo/frecuencia que se esté procesando.

25 En algunas realizaciones, dicho por lo menos un parámetro de ponderación 132 comprende el primer factor de ponderación 116 y el segundo factor de ponderación 114. En algunas realizaciones, dicho por lo menos un parámetro de ponderación 132 comprende un solo parámetro de ponderación. En este caso, el extractor húmedo/seco 134 puede obtener el primer y el segundo factores de ponderación 116, 114 a partir del único parámetro de ponderación 132. Por ejemplo, el primer y el segundo factores de ponderación 116, 114 pueden satisfacer ciertas relaciones que permiten que se obtenga uno de los factores de ponderación una vez se conoce el otro factor de ponderación. Un ejemplo de dicha relación puede ser que la suma cuadrática del primer factor de ponderación 116 y el segundo factor de ponderación 114 sea igual a uno. Por lo tanto, si el parámetro de ponderación único 132 comprende el primer factor de ponderación 116, el segundo factor de ponderación 114 se puede obtener como la raíz cuadrada de uno menos el primer factor de ponderación 116 al cuadrado, y viceversa.

30 El primer factor de ponderación 116 se utiliza para la ponderación 122, es decir para la multiplicación, con el objeto de audio aproximado 110. El segundo factor de ponderación 114 se utiliza para la ponderación 120, es decir para la multiplicación, con el correspondiente objeto de audio descorrelacionado 136. El componente 128 de reconstrucción de objetos de audio está configurado además para combinar 124, por ejemplo llevando a cabo una suma, el objeto de audio aproximado ponderado 150 con el correspondiente objeto de audio descorrelacionado ponderado 152 con el fin de reconstruir la tesela de tiempo/frecuencia del correspondiente objeto de audio 142.

35 En otras palabras, para cada objeto y cada tesela de tiempo/frecuencia, la cantidad de descorrelación se puede controlar mediante un parámetro de ponderación 132. En el extractor húmedo/seco 134, este parámetro de ponderación 132 se transforma en un factor de ponderación 116 ( $w_{seco}$ ) aplicado al objeto aproximado 110, y en un factor de ponderación 114 ( $w_{húmedo}$ ) aplicado al objeto descorrelacionado 136. La suma cuadrática de estos factores de ponderación es uno, es decir

$$w_{húmedo}^2 + w_{seco}^2 = 1,$$



lo que significa que el objeto final 142, que es el resultado de la suma 124, tiene la misma energía que el correspondiente objeto aproximado 110.

Para permitir que las señales de entrada 126, 130 sean descodificadas mediante un sistema descodificador de audio que no puede tratar la descorrelación, es decir para mantener la retrocompatibilidad con dicho descodificador de audio, la señal de entrada 126 se puede disponer en una trama 202, tal como se representa en la figura 2. Según esta realización, la matriz de reconstrucción 104 se dispone en un primer campo de la trama 202 utilizando un primer formato y dicho por lo menos un parámetro de ponderación 132 se dispone en un segundo campo de la trama 202 utilizando un segundo formato. De este modo, un descodificador que pueda leer el primer formato pero no el segundo formato, puede aún así descodificar y utilizar la matriz de reconstrucción 104 para la mezcla ascendente de la señal de mezcla descendente 106 de cualquier modo convencional. En este caso, el segundo campo de la trama 202 se puede desechar.

Según algunas realizaciones, el sistema de descodificación de audio 100 de la figura 1 puede recibir adicionalmente L señales auxiliares 144, por ejemplo en el primer componente de recepción 102. Puede haber una o varias de dichas señales auxiliares, es decir  $L \geq 1$ . Estas señales auxiliares 144 pueden estar incluidas en la señal de entrada 130. Las señales auxiliares 144 pueden estar incluidas en la señal de entrada 130, de tal modo que se mantenga la retrocompatibilidad según lo anterior, es decir, de manera que un sistema descodificador que no pueda tratar las señales auxiliares pueda seguir obteniendo las señales de mezcla descendente 106 a partir de la señal de entrada 130. La matriz de reconstrucción 104 puede permitir además la reconstrucción de la aproximación de los N objetos de audio 110 a partir de las M señales de mezcla descendente 106 y de las L señales auxiliares 144. Por lo tanto, el componente de aproximación de objetos de audio 108 puede estar configurado para aplicar la matriz de reconstrucción 104 a las M señales de mezcla descendente 106 y las L señales auxiliares 144 con el fin de generar los N objetos de audio aproximados 110.

La función de las señales auxiliares 144 es mejorar la aproximación de los N objetos de audio en el componente 108 de aproximación de objetos de audio. Según un ejemplo, por lo menos una de las señales auxiliares 144 es igual a uno de los N objetos de audio a reconstruir. En este caso, el vector en la matriz de reconstrucción 104 que se utiliza para reconstruir el objeto de audio específico contendrá solamente un único parámetro distinto de cero, por ejemplo un parámetro con el valor uno (1). Según otros ejemplos, por lo menos una de las L señales auxiliares 144 es una combinación de por lo menos dos de los N objetos de audio a reconstruir.

En algunas realizaciones, las L señales auxiliares pueden representar dimensiones de señal de los N objetos de audio para las que se perdió información en el proceso de generar las M señales de mezcla descendente 106 a partir de los N objetos de audio. Esto se puede explicar diciendo que las M señales de mezcla descendente 106 comprenden un hiperplano en un espacio de señales, y que las L señales auxiliares 144 no están situadas en este hiperplano. Por ejemplo, las L señales auxiliares 144 pueden ser ortogonales al hiperplano comprendido por las M señales de mezcla descendente 106. Basándose solamente en las M señales de mezcla descendente 106, se pueden reconstruir únicamente las señales que están situadas en el hiperplano, es decir, los objetos de audio que no estén situados en el hiperplano serán aproximados por una señal de audio en el hiperplano. Utilizando adicionalmente las L señales auxiliares 144 en la reconstrucción, se pueden reconstruir asimismo señales que no están situadas en el hiperplano. Como resultado, se puede mejorar la aproximación de los objetos de audio al utilizar asimismo las L señales auxiliares.

La figura 3 muestra, a modo de ejemplo, un diagrama de bloques generalizado de un codificador de audio 300 para generar por lo menos un parámetro de ponderación 320. Dicho por lo menos un parámetro de ponderación 320 es para su utilización en un descodificador, por ejemplo el sistema de descodificación de audio 100 descrito anteriormente, cuando se reconstruye una tesela de tiempo/frecuencia de un objeto de audio específico combinando (referencia 124 de la figura 1) una aproximación ponderada del lado del descodificador (referencia 150 de la figura 1) del objeto de audio específico con una correspondiente versión descorrelacionada ponderada (referencia 152 de la figura 1) del objeto de audio específico aproximado del lado del descodificador.

El codificador 300 comprende un componente de recepción 302 configurado para recibir M señales de mezcla descendente 312 que son combinaciones de por lo menos N objetos de audio que incluyen el objeto de audio específico. El componente de recepción 302 está configurado además para recibir el objeto de audio específico 314. En algunas realizaciones, el componente de recepción 302 está configurado además para recibir L señales auxiliares 322. Tal como se ha explicado anteriormente, por lo menos una de las L señales auxiliares 322 puede ser igual a uno de los N objetos de audio, por lo menos una de las L señales auxiliares 322 puede ser una combinación de por lo menos dos de los N objetos de audio y por lo menos una de las L señales auxiliares 322 puede contener información no presente en ninguna de las M señales de mezcla descendente.

El codificador 300 comprende además una unidad de cálculo 304. La unidad de cálculo 304 está configurada para calcular una primera cantidad 316 indicativa del nivel de energía del objeto de audio específico, por ejemplo en un primer componente de cálculo de energía 306. La primera cantidad 316 se puede calcular como una norma del objeto de audio específico. Por ejemplo, la primera cantidad 316 puede ser igual a la energía del objeto de audio específico y, por lo tanto, se puede calcular mediante la norma-dos  $Q_1 = \|S\|^2$ , donde S indica el objeto de audio

específico. La primera cantidad se puede calcular alternativamente como otra cantidad que sea indicativa de la energía del objeto de audio específico, tal como la raíz cuadrada de la energía.

5 La unidad de cálculo 304 está configurada además para calcular una segunda cantidad 318 que es indicativa de un nivel de energía correspondiente a un nivel de energía de una aproximación del lado del codificador del objeto de audio específico 314. La aproximación del lado del codificador puede ser, por ejemplo, una combinación, tal como una combinación lineal, de las M señales de mezcla descendente 312. Alternativamente, la aproximación del lado del codificador puede ser una combinación, tal como una combinación lineal, de las M señales de mezcla descendente 312 y la señal auxiliar L 322. La segunda cantidad se puede calcular en un segundo componente de cálculo de la energía 308.

10 A continuación, la aproximación del lado del codificador se puede calcular, por ejemplo, utilizando una matriz de mezcla ascendente no adaptada en energía y la señal de mezcla descendente M 312. En el contexto de la presente memoria descriptiva, mediante la expresión "no adaptada en energía" se debe entender que la aproximación del objeto de audio específico no estará adaptada en energía al propio objeto de audio específico, es decir, la aproximación tendrá un nivel de energía diferente, a menudo inferior, comparado con el objeto de audio específico 314.

15 La matriz de mezcla ascendente no adaptada en energía se puede generar utilizando diferentes enfoques. Por ejemplo, se puede utilizar un enfoque predictivo de error cuadrático medio mínimo (MMSE, Minimum Mean Squared Error) que toma como entrada por lo menos los N objetos de audio así como las M señales de mezcla descendente 312 (y posiblemente las L señales auxiliares 322). Esto se puede describir como un enfoque iterativo que ayuda a encontrar la matriz de mezcla ascendente que minimiza el error cuadrático medio de las aproximaciones de los N objetos de audio. En particular, el enfoque aproxima los N objetos de audio con una matriz de mezcla ascendente candidata, que se multiplica por las M señales de mezcla descendente 312 (y posiblemente las L señales auxiliares 322), y compara las aproximaciones con los N objetos de audio en términos del error cuadrático medio. La matriz de mezcla ascendente candidata que minimiza el error cuadrático medio se selecciona como la matriz de mezcla ascendente que se utiliza para definir la aproximación del lado del codificador del objeto de audio específico.

20 Cuando se utiliza el enfoque MMSE, el error de predicción e entre el objeto de audio específico S y el objeto de audio aproximado S' es ortogonal a S. Esto significa que:

$$\|S'\|^2 + \|e\|^2 = \|S\|^2.$$

30 En otras palabras, la energía del objeto de audio S es igual a la suma de la energía del objeto de audio aproximado y la energía del error de predicción. Debido a la relación anterior, la energía del error de predicción e proporciona por lo tanto una indicación de la energía de la aproximación del lado del codificador S'.

35 Por consiguiente, la segunda cantidad 318 se puede calcular utilizando la aproximación del objeto de audio específico S' o bien el error de predicción. La segunda cantidad se puede calcular como una norma de la aproximación del objeto de audio específico S' o una norma del error de predicción e. Por ejemplo, la segunda cantidad se puede calcular como la 2-norma, es decir  $Q_2 = \|S'\|^2$  ó  $Q_2 = \|e\|^2$ . La segunda cantidad se puede calcular alternativamente como otra cantidad, que sea indicativa de la energía del objeto de audio específico aproximado, tal como la raíz cuadrada de la energía del objeto de audio específico aproximado o la raíz cuadrada de la energía del error de predicción.

40 La unidad de cálculo está configurada además para calcular dicho por lo menos un parámetro de ponderación 320 en base a la primera 316 y la segunda 318 cantidades, por ejemplo en un componente de cálculo de parámetros 310. El componente de cálculo de parámetros 310 puede calcular, por ejemplo, dicho por lo menos un parámetro de ponderación 320 comparando la primera cantidad 316 y la segunda cantidad 318. A continuación se explicará en detalle un componente de cálculo de parámetros 310 a modo de ejemplo, junto con la figura 4 y las figuras 5a-c.

45 La figura 4 muestra a modo de ejemplo un diagrama de bloques generalizado del componente de cálculo de parámetros 310 para generar dicho por lo menos un parámetro de ponderación 320. El componente de cálculo de parámetros 310 compara la primera cantidad 316 y la segunda cantidad 318, por ejemplo en un componente de cálculo de relaciones 402, calculando la relación r entre la segunda 318 y la primera 316 cantidades. A continuación, la relación se eleva la potencia de  $\alpha$ , es decir

$$r = \left(\frac{Q_2}{Q_1}\right)^\alpha$$

50 donde  $Q_2$  es la segunda cantidad 318 y  $Q_1$  es la primera cantidad 316. Según algunas realizaciones, cuando  $Q_2 = \|S'\|$  y  $Q_1 = \|S\|$ ,  $\alpha$  es igual a 2, es decir, la relación r es la relación entre las energías del objeto de audio específico aproximado y del objeto de audio específico. La relación elevada la potencia de  $\alpha$  406 se utiliza a continuación para calcular dicho por lo menos un parámetro de ponderación 320, por ejemplo en un componente de mapeo 404. El componente de mapeo 404 somete r 406 a una función creciente que mapea r a dicho por lo menos un parámetro de ponderación 320. Dichas funciones crecientes se ejemplifican en las figuras 5a-c. En las figuras 5a-c, el eje

horizontal representa el valor de  $r$  406 y el eje vertical representa el valor del parámetro de ponderación 320. En este ejemplo, el parámetro de ponderación 320 es un solo parámetro de ponderación que corresponde al primer factor de ponderación 116 de la figura 1.

En general, el principio para la función de mapeo es:

- 5 si  $Q_2 \ll Q_1$ , el primer factor de ponderación se aproxima a 0, y si  $Q_2 \approx Q_1$ , el primer factor de ponderación se aproxima a 1.

La figura 5a muestra una función de mapeo 502 en la que, para valores de  $r$  406 entre 0 y 1, el valor de  $r$  será igual al valor del parámetro de ponderación 312. Para valores de  $r$  por encima de 1, el valor del parámetro de ponderación 320 será de 1.

- 10 La figura 5b muestra otra función de mapeo 504 en la que, para valores de  $r$  406 entre 0 y 0,5, el valor del parámetro de ponderación 320 será de 0. Para valores de  $r$  por encima de 1, el valor del parámetro de ponderación 320 será de 1. Para valores de  $r$  entre 0,5 y 1, el valor del parámetro de ponderación 320 será de  $(r-0,5)*2$ .

- 15 La figura 5c muestra una tercera función de mapeo 506 alternativa que generaliza las funciones de mapeo de las figuras 5a-b. La función de mapeo 506 está definida mediante por lo menos cuatro parámetros  $b_1$ ,  $b_2$ ,  $\beta_1$  y  $\beta_2$ , que pueden ser constantes ajustadas para una calidad perceptual óptima de los objetos de audio reconstruidos en el lado del descodificador. En general, limitar la cantidad máxima de descorrelación en la señal de audio de salida puede ser beneficioso, dado que un objeto de audio aproximado descorrelacionado tiene a menudo una calidad peor que un objeto de audio aproximado, cuando se escuchan por separado. Ajustar  $b_1$  para que sea mayor que cero controla esto directamente y, por lo tanto, puede asegurar que el parámetro de ponderación 320 (y por lo tanto el primer factor de ponderación 116 de la figura 1) será mayor que cero en todos los casos. Ajustar  $b_2$  para que sea menor que uno tiene la consecuencia de que existe siempre un nivel mínimo de energía de descorrelación en la salida desde el sistema de descodificación de audio 100. En otras palabras, el segundo factor de ponderación 114 de la figura 1 será siempre mayor que cero.  $\beta_1$  controla implícitamente la cantidad de descorrelación añadida en la salida desde el sistema de descodificación de audio 100, pero involucrando dinámicas diferentes (en comparación con  $b_1$ ). De manera similar,  $\beta_2$  controla implícitamente la cantidad de descorrelación en la salida desde el sistema de descodificación de audio 100.

En caso de que se desee una función de mapeo curva entre los valores  $\beta_1$  y  $\beta_2$  de  $r$ , es necesario por lo menos otro parámetro, que puede ser una constante.

#### Equivalentes, extensiones, alternativas y miscelánea

- 30 Tras el estudio de la descripción anterior, resultarán evidentes para un experto en la materia otras realizaciones de la presente invención. Aunque la presente descripción y los dibujos dan a conocer realizaciones y ejemplos, la invención no se limita a estos ejemplos específicos. Se pueden realizar numerosas modificaciones y variaciones sin apartarse del alcance de la presente invención, que se define mediante las reivindicaciones adjuntas. Cualesquiera signos de referencia que aparezcan en las reivindicaciones no se deben entender como limitando su alcance.

- 35 Adicionalmente, en la práctica de la invención, a partir de un estudio de los dibujos, de la invención y de las reivindicaciones adjuntas el experto en la materia puede comprender y efectuar variaciones a las realizaciones dadas a conocer. En las reivindicaciones, el término "comprende" no excluye otros elementos o etapas, y el artículo indefinido "un" o "una" no excluye una pluralidad. El mero hecho de que ciertas disposiciones se expongan en reivindicaciones dependientes diferentes entre sí, no indica que no se pueda utilizar ventajosamente una combinación de estas disposiciones.

- 40 Los sistemas y procedimientos dados a conocer en lo anterior se pueden implementar como software, software inalterable, hardware o una combinación de los mismos. En una implementación en hardware, la división de tareas entre unidades funcionales a las que se hace referencia en la descripción anterior no corresponde necesariamente a la división en unidades físicas; por el contrario, un componente físico puede tener múltiples funcionalidades, y una tarea puede ser realizada por varios componentes físicos en cooperación. Determinados componentes o todos los componentes se pueden implementar como software ejecutado por un procesador de señal digital o un microprocesador, o se pueden implementar como hardware o como un circuito integrado de aplicación específica.
- 45 Dicho software puede estar distribuido en un medio legible por ordenador, que puede comprender medios de almacenamiento informático (o medios no transitorios) y medios de comunicaciones (o medios transitorios). Tal como es bien sabido por un experto en la materia, el término medios de almacenamiento informático incluye medios volátiles y no volátiles, extraíbles y no extraíbles implementados en cualquier procedimiento o tecnología para el almacenamiento de información, tal como instrucciones legibles por ordenador, estructuras de datos, módulos de programa u otros datos. Los medios de almacenamiento informático incluyen, de forma no limitativa, RAM, ROM, EEPROM, memorias flash u otra tecnología de memoria, CD-ROM, discos versátiles digitales (DVD, digital versatile disks) u otro almacenamiento en disco óptico, casetes magnéticas, cintas magnéticas, almacenamiento en disco magnético u otros dispositivos de almacenamiento magnético, o cualquier otro medio que pueda ser utilizado para almacenar la información deseada y al que se pueda acceder mediante un ordenador. Además, es bien conocido por un experto en la materia que los medios de comunicación incorporan habitualmente instrucciones legibles por

ordenador, estructuras de datos, módulos de programa u otros datos en una señal de datos modulada, tal como una onda portadora u otro mecanismo de transporte, e incluyen cualesquiera medios de distribución de información.

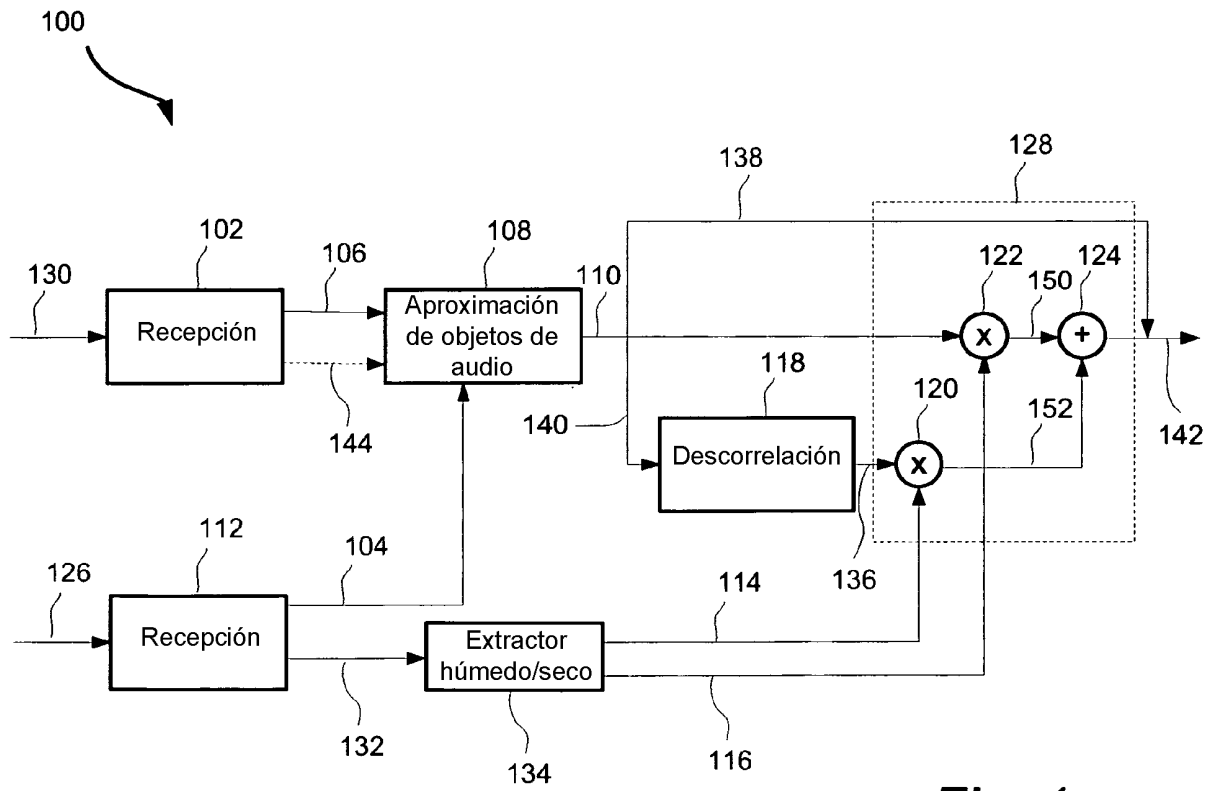
**REIVINDICACIONES**

1. Un procedimiento para reconstruir una tesela de tiempo/frecuencia de N objetos de audio, que comprende las etapas de:
- recibir M señales de mezcla descendente (106);
- 5 recibir una matriz de reconstrucción (104) que permite la reconstrucción de una aproximación de los N objetos de audio a partir de las M señales de mezcla descendente;
- aplicar la matriz de reconstrucción a las M señales de mezcla descendente para generar N objetos de audio aproximados (110);
- 10 someter por lo menos un subconjunto (140) de los N objetos de audio aproximados a un proceso de descorrelación para generar por lo menos un objeto de audio descorrelacionado (136), de tal modo que cada uno de dicho por lo menos un objeto de audio descorrelacionado corresponde a uno de los N objetos de audio aproximados;
- para cada uno de los N objetos de audio aproximados que no tiene un correspondiente objeto de audio descorrelacionado, reconstruir la tesela de tiempo/frecuencia del objeto de audio mediante el objeto de audio aproximado; y
- 15 para cada uno de los N objetos de audio aproximados que tiene un correspondiente objeto de audio descorrelacionado, reconstruir la tesela de tiempo/frecuencia del objeto de audio mediante:
- recibir un único parámetro de ponderación (132) a partir del cual se puede obtener un primer factor de ponderación (116) y un segundo factor de ponderación (114),
- ponderar (122) el objeto de audio aproximado mediante el primer factor de ponderación,
- 20 ponderar (120) el objeto de audio descorrelacionado correspondiente al objeto de audio aproximado mediante el segundo factor de ponderación, y
- combinar (124), realizando una suma, el objeto de audio aproximado ponderado (150) con el correspondiente objeto de audio descorrelacionado ponderado (152) para reconstruir la tesela de tiempo/frecuencia del objeto de audio aproximado (142),
- 25 **caracterizado por que**
- el nivel de energía de la tesela de tiempo/frecuencia reconstruida es igual al nivel de energía de una correspondiente tesela de tiempo/frecuencia del objeto de audio aproximado.
2. El procedimiento según la reivindicación 1, en el que la suma cuadrática del primer factor de ponderación y el segundo factor de ponderación es igual a uno, y en el que el parámetro de ponderación único comprende cualquiera
- 30 del primer factor de ponderación o el segundo factor de ponderación.
3. El procedimiento según cualquiera de las reivindicaciones anteriores, en el que la etapa de someter por lo menos un subconjunto de los N objetos de audio aproximados a un proceso de descorrelación comprende someter cada uno de los N objetos de audio aproximados a un proceso de descorrelación, de tal modo que cada uno de los N objetos de audio aproximados corresponde a un objeto de audio descorrelacionado.
- 35 4. El procedimiento según cualquiera de las reivindicaciones anteriores, en el que el primer y el segundo factores de ponderación varían con el tiempo y la frecuencia.
5. El procedimiento según cualquiera de las reivindicaciones anteriores, en el que, tras su recepción, la matriz de reconstrucción y dicho por lo menos un parámetro de ponderación se disponen en una trama (202), en el que la matriz de reconstrucción se dispone en un primer campo de la trama utilizando un primer formato y dicho por lo
- 40 menos un parámetro de ponderación se dispone en un segundo campo de la trama utilizando un segundo formato, permitiendo de ese modo que un descodificador que soporta solamente el primer formato descodifique la matriz de reconstrucción en el primer campo y deseche dicho por lo menos un parámetro de ponderación en el segundo campo.
6. El procedimiento según cualquiera de las reivindicaciones anteriores, que comprende además recibir L señales auxiliares, en el que la matriz de reconstrucción permite además la reconstrucción de la aproximación de los N objetos de audio a partir de las M señales de mezcla descendente y las L señales auxiliares, y en el que el procedimiento comprende además aplicar la matriz de reconstrucción a las M señales de mezcla descendente y las L señales auxiliares para generar los N objetos de audio aproximados.
- 45 7. El procedimiento según la reivindicación 6, en el que por lo menos una de las L señales auxiliares es igual a uno de los N objetos de audio a reconstruir.
- 50

8. Un aparato (100) para reconstruir una tesela de tiempo/frecuencia de N objetos de audio, que comprende:
- un primer componente de recepción (102) configurado para recibir M señales de mezcla descendente (106);
  - un segundo componente de recepción (112) configurado para recibir una matriz de reconstrucción (104) que permite la reconstrucción de una aproximación de los N objetos de audio a partir de las M señales de mezcla descendente;
- 5 un componente de aproximación de objetos de audio (108) dispuesto más abajo del primer y del segundo componentes de recepción y configurado para aplicar la matriz de reconstrucción a las M señales de mezcla descendente con el fin de generar N objetos de audio aproximados (110);
- un componente de descorrelación (118) dispuesto más abajo del componente de aproximación de objetos de audio y configurado para someter por lo menos un subconjunto (140) de los N objetos de audio aproximados a un proceso de descorrelación con el fin de generar por lo menos un objeto de audio descorrelacionado (136), de tal modo que cada uno de dicho por lo menos un objeto de audio descorrelacionado corresponde a uno de los N objetos de audio aproximados;
- 10 estando el segundo componente de recepción configurado además para recibir, para cada uno de los N objetos de audio aproximados que tiene un correspondiente objeto de audio descorrelacionado, un único parámetro de ponderación (132) a partir del cual se puede obtener un primer factor de ponderación (116) y un segundo factor de ponderación (114); y
- 15 un componente (128) de reconstrucción de objetos de audio dispuesto más abajo del componente de aproximación de objetos de audio, del componente de descorrelación y del segundo componente de recepción, y configurado para:
- 20 para cada uno de los N objetos de audio aproximados que no tiene un correspondiente objeto de audio descorrelacionado, reconstruir la tesela de tiempo/frecuencia del objeto de audio mediante el objeto de audio aproximado; y
- para cada uno de los N objetos de audio aproximados que tiene un correspondiente objeto de audio descorrelacionado, reconstruir la tesela de tiempo/frecuencia del objeto de audio mediante:
- 25 ponderar (122) el objeto de audio aproximado mediante el primer factor de ponderación;
- ponderar (120) el objeto de audio descorrelacionado correspondiente al objeto de audio aproximado mediante el segundo factor de ponderación; y
- 30 combinar (124), realizando una suma, el objeto de audio aproximado ponderado (150) con el correspondiente objeto de audio descorrelacionado ponderado (152) para reconstruir la tesela de tiempo/frecuencia del objeto de audio aproximado (142),
- caracterizado por que**
- el nivel de energía de la tesela de tiempo/frecuencia reconstruida es igual al nivel de energía de una correspondiente tesela de tiempo/frecuencia del objeto de audio aproximado.
- 35 9. Un procedimiento en un codificador (300) para generar por lo menos un parámetro de ponderación (320) para su utilización cuando se reconstruye una tesela de tiempo/frecuencia de un objeto de audio específico, comprendiendo el procedimiento las etapas de:
- recibir M señales de mezcla descendente (312) que son combinaciones de por lo menos N objetos de audio que incluyen el objeto de audio específico;
  - recibir el objeto de audio específico (314);
- 40 calcular una primera cantidad (316) indicativa de un nivel de energía del objeto de audio específico;
- caracterizado por que** el procedimiento comprende además las etapas de:
- calcular una segunda cantidad (318) indicativa de un nivel de energía correspondiente a un nivel de energía de una aproximación del lado del codificador del objeto de audio específico, siendo la aproximación del lado del codificador una combinación de las M señales de mezcla descendente;
- 45 calcular por lo menos un parámetro de ponderación en base a la primera y la segunda cantidades, en el que dicho por lo menos un parámetro de ponderación es para ponderar una aproximación del lado del decodificador del objeto de audio específico y una versión descorrelacionada de la aproximación del lado del decodificador del objeto de audio específico.
- 50 10. El procedimiento según la reivindicación 9, en el que dicho por lo menos un parámetro de ponderación comprende un único parámetro de ponderación a partir del cual se puede obtener un primer factor de ponderación y

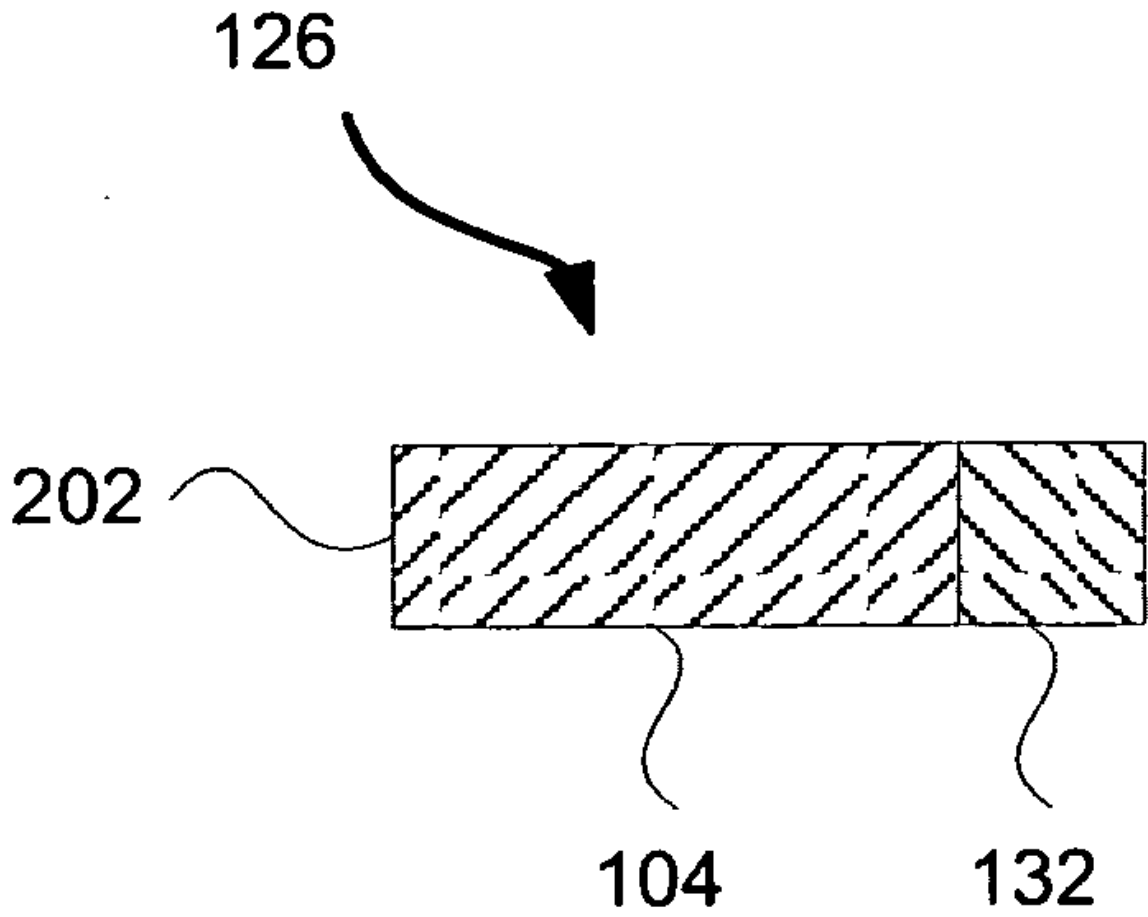
un segundo factor de ponderación, el primer factor de ponderación para ponderar la aproximación del lado del descodificador del objeto de audio específico y el segundo factor de ponderación para ponderar la versión descorrelacionada del objeto de audio aproximado del lado del descodificador.

- 5 11. El procedimiento según la reivindicación 10, en el que la suma cuadrática del primer factor de ponderación y el segundo factor de ponderación es igual a uno, y en el que el parámetro de ponderación único comprende cualquiera del primer factor de ponderación o el segundo factor de ponderación.
12. El procedimiento según cualquiera de las reivindicaciones 9 a 11, en el que el primer y el segundo factores de ponderación varían con el tiempo y la frecuencia.
- 10 13. El procedimiento según cualquiera de las reivindicaciones 9 a 12, en el que la segunda cantidad indicativa de un nivel de energía corresponde a un nivel de energía de una aproximación del lado del codificador del objeto de audio específico, siendo la aproximación del lado del codificador una combinación lineal de las M señales de mezcla descendente y las L señales auxiliares, estando las señales de mezcla descendente y las señales auxiliares formadas a partir de los N objetos de audio.
- 15 14. Un medio legible por ordenador que comprende instrucciones de código informático adaptadas para llevar a cabo el procedimiento de cualquiera de las reivindicaciones 9 a 13 o de cualquiera de las reivindicaciones 1 a 7, cuando es ejecutado en un dispositivo con capacidad de procesamiento.
15. Un codificador (300) para generar por lo menos un parámetro de ponderación (320) para su utilización cuando se reconstruye una tesela de tiempo/frecuencia de un objeto de audio específico, comprendiendo el codificador:
- 20 un componente de recepción (302) configurado para recibir M señales de mezcla descendente (312) que son combinaciones de por lo menos N objetos de audio que incluyen el objeto de audio específico, estando el componente de recepción configurado además para recibir el objeto de audio específico (314);
- una unidad de cálculo (304) configurada para:
- calcular una primera cantidad (316) indicativa de un nivel de energía del objeto de audio específico;
- estando el codificador **caracterizado por que** la unidad de cálculo está además configurada para:
- 25 calcular una segunda cantidad (318) indicativa de un nivel de energía correspondiente a un nivel de energía de una aproximación del lado del codificador del objeto de audio específico, siendo la aproximación del lado del codificador una combinación de las M señales de mezcla descendente;
- 30 en el que la unidad de cálculo calcula dicho por lo menos un parámetro de ponderación en base a la primera y la segunda cantidades, en el que dicho por lo menos un parámetro de ponderación es para ponderar una aproximación del lado del descodificador del objeto de audio específico y una versión descorrelacionada de la aproximación del lado del descodificador del objeto de audio específico.

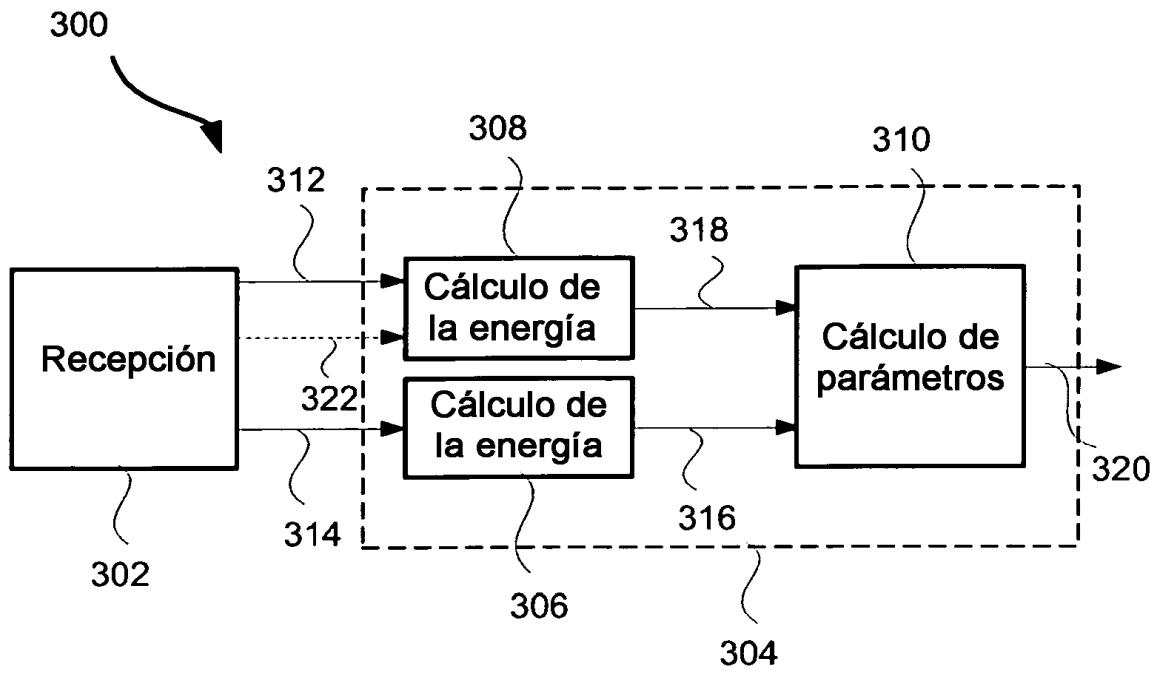


*Fig. 1*

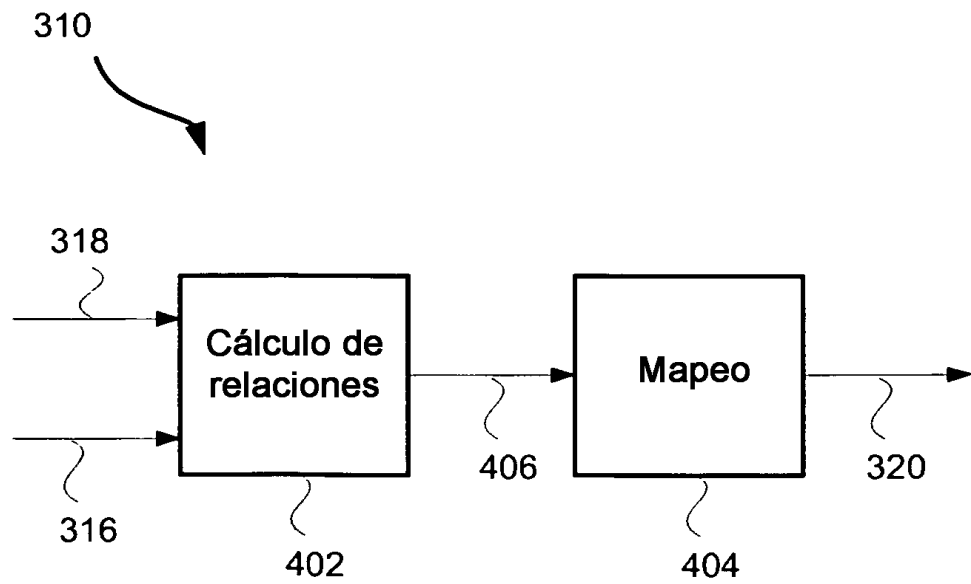




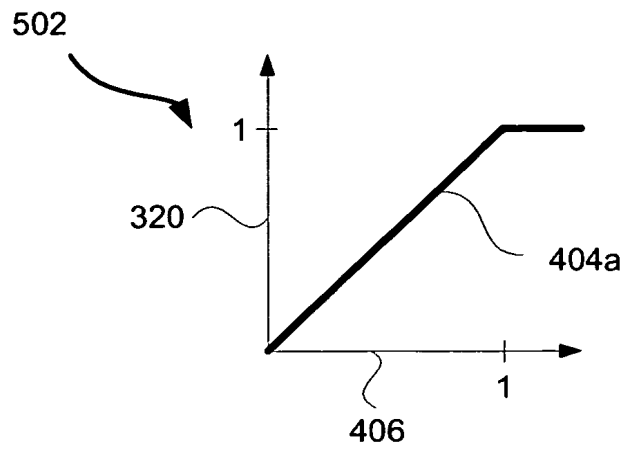
***Fig. 2***



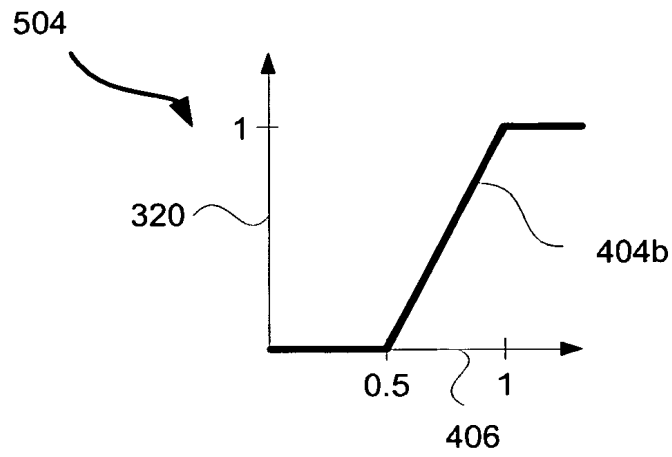
**Fig. 3**



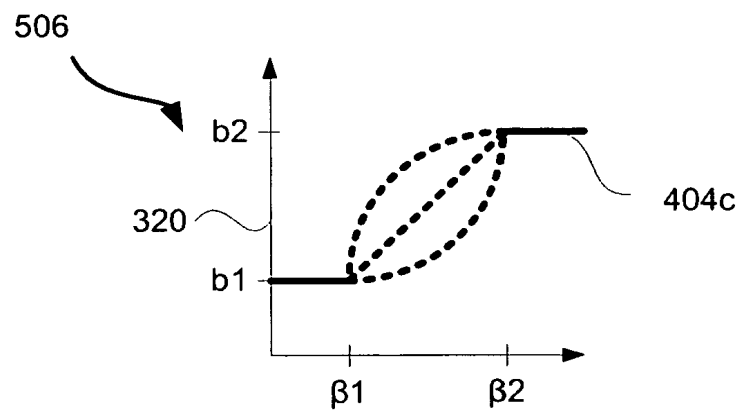
**Fig. 4**



*Fig. 5a*



*Fig. 5b*



*Fig. 5c*