

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 391 308**

51 Int. Cl.:
G10L 19/00 (2006.01)
G10L 19/02 (2006.01)
H04S 5/00 (2006.01)
H04S 7/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **07819257 .2**
96 Fecha de presentación: **23.10.2007**
97 Número de publicación de la solicitud: **1997102**
97 Fecha de publicación de la solicitud: **03.12.2008**

54 Título: **Aparato y procedimiento para generar una señal de ambiente a partir de una señal de audio, aparato y procedimiento para obtener una señal de audio multi-canal a partir de una señal de audio, y programa de ordenador**

30 Prioridad:
24.10.2006 DE 102006050068

45 Fecha de publicación de la mención BOPI:
23.11.2012

45 Fecha de la publicación del folleto de la patente:
23.11.2012

73 Titular/es:
**FRAUNHOFER-GESELLSCHAFT ZUR
FÖRDERUNG DER ANGEWANDTEN
FORSCHUNG E.V. (100.0%)
Hansastrasse 27c
80686 München, DE**

72 Inventor/es:
**UHLE, CHRISTIAN;
HERRE, JÜRGEN;
WALTHER, ANDREAS;
HELLMUTH, OLIVER y
JANSSEN, CHRISTIAAN**

74 Agente/Representante:
PONTI SALES, Adelaida

ES 2 391 308 T3

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

DESCRIPCIÓN

Aparato y procedimiento para generar una señal de ambiente a partir de una señal de audio, aparato y procedimiento para obtener una señal de audio multi-canal a partir de una señal de audio, y programa de ordenador

5 **[0001]** La presente invención se refiere en general a un aparato y a un procedimiento para generar una señal de ambiente a partir de una señal de audio, a un aparato y un procedimiento para obtener una señal de audio multi-canal a partir de una señal de audio, y a un programa de ordenador. Concretamente, la presente invención se refiere a un procedimiento y concepto para calcular una señal de ambiente a partir de una señal de audio para el *upmixing* de señales mono de audio para reproducción en sistemas multi-canal.

10 **[0002]** A continuación, se discutirá la motivación subyacente a la presente invención. En la actualidad, el material de audio multi-canal también está experimentando una creciente popularidad en entornos domésticos. La razón principal de esto es que las películas sobre medios de DVD ofrecen a menudo sonido multi-canal 5.1. Por esta razón, incluso los usuarios domésticos instalan frecuentemente sistemas de reproducción de audio capaces de reproducir señales de audio multicanal.

15 **[0003]** Una configuración correspondiente puede consistir, por ejemplo, en tres altavoces (a modo de ejemplo designados con L, C y R) dispuestos en la parte delantera, dos altavoces (designados con L3 y Rs) dispuestos detrás o hacia atrás de un oyente y un canal de efectos de baja frecuencia (también referido como LFE). Los tres altavoces dispuestos en la parte delantera (L, C, R) serán en lo sucesivo también llamados altavoces frontales. Los altavoces dispuestos detrás y en la parte posterior del oyente (Ls, Rs) son en lo sucesivo también llamados altavoces posteriores.

20 **[0004]** Además, es de señalar que, por razones de conveniencia, los detalles y las explicaciones siguientes se refieren a los sistemas 5.1. Los siguientes detalles pueden, por su puesto, también ser aplicados a otros sistemas de canales múltiples, con sólo realizar pequeñas modificaciones.

[0005] Los sistemas de canales múltiples (por ejemplo, un sistema de audio de canales múltiples 5.1) proporcionan varias ventajas conocidas sobre la reproducción estéreo de dos canales. Esto se ejemplifica por las siguientes ventajas:

25 - Ventaja 1: es tabilidad mejorada frente a la imagen, incluso de o fuera de la posición de escucha óptima (central). El "punto dulce" se amplía mediante el canal central. El término "punto dulce" se refiere a una superficie de posiciones de escucha donde se puede presentar una impresión de sonido óptima percibida (por un oyente).

- Ventaja 2: Establecer una mejor aproximación de una impresión o experiencia a una sala de conciertos. La mayor experiencia de la "envoltura" y amplitud se obtiene a través de los altavoces de canal posteriores o los altavoces de canal posteriores.

30 **[0006]** Sin embargo, todavía hay una gran cantidad de contenidos de audio legados que consta de sólo dos ("estéreo") canales de audio, tal como en los discos compactos. Incluso se venden grabaciones muy viejas y viejas películas y series de televisión en CDs y/ o DVDs que están disponibles en calidad mono y/ o por medio de solamente una señal de audio "mono" de un solo canal.

35 **[0007]** Por lo tanto, hay opciones para la reproducción de material legado de audio mono a través de una configuración multicanal 5.1:

- Opción 1: Reproducción o *playback* del canal mono a través del centro o por el altavoz central de tal manera que se obtenga una fuente mono verdadera.

40 - Opción 2: Reproducción o *playback* de la señal mono por los altavoces L y R (es decir, por el altavoz delantero izquierdo y el altavoz delantero derecho). Este enfoque produce una fuente mono fantasma que tiene una anchura de fuente percibida más amplia que una fuente de mono verdadera, pero con una tendencia hacia el altavoz más cercano al oyente cuando el oyente no se sienta en o en el punto dulce.

45 **[0008]** Este procedimiento también puede ser usado si sólo hay disponible un sistema de reproducción de dos canales, y que no hace uso de la configuración de altavoces ampliada (tal es como una configuración de altavoz con 5 o 6 altavoces). El altavoz C o altavoz central, el altavoz L3 o altavoz trasero izquierdo, el altavoz Rs o altavoz trasero derecho y el altavoz LFE o altavoz de canal de efectos de baja frecuencia permanecen sin utilizar.

50 - Opción 3: Puede emplearse un procedimiento para convertir el canal de la señal mono en una señal multicanal utilizando todos los altavoces 5.1 (es decir, todos los seis altavoces utilizados en un sistema de canales múltiples 5.1). De esta manera, la señal multi-canal se beneficia de las ventajas previamente discutidas de la configuración de varios canales. El procedimiento se puede emplear en tiempo real o "sobre la marcha" o por medio de preprocesamiento y se conoce como proceso de *upmix* o "upmixing".

[0009] En lo que respecta a la calidad del audio o la calidad del sonido, la opción 3 ofrece ventajas sobre la opción 1 y la opción 2. Si embargo, particularmente con respecto a la señal generada para la alimentación de los altavoces posteriores, el procesamiento de la señal requerida no es obvia.

[0010] En la literatura, se describen dos conceptos diferentes para un procedimiento upmix o proceso de upmix. Estos conceptos son el "Concepto Directo / Ambiente" y el "Concepto en la Banda". Los dos conceptos indicados se describen a continuación.

Concepto Directo / Ambiente

5 **[0011]** Las "fuentes de sonido directo" se reproducen a través de los tres canales frontales de forma que son percibidos en la misma posición que en la original versión de dos canales. El término "fuente de sonido directo" se utiliza aquí para describir el sonido que viene, única y directamente de una fuente de sonido discreto (por ejemplo, un instrumento), y que muestra sonido adicional escaso o nulo, por ejemplo, debido a las reflexiones de las paredes.

10 **[0012]** En este escenario, el sonido o el ruido alimentado a los altavoces posteriores debe consistir únicamente en ruido o sonido similar al ambiente (que puede o no puede estar presente en la grabación original). El sonido o ruido similar al ambiente no está asociado a una sola fuente de sonido o fuente de ruido, pero contribuye a la reproducción o playback del entorno acústico (acústica de la sala) de una grabación o a la así llamada "sensación de envolvente" del oyente. El sonido similar al ambiente o ruido similar al ambiente es ruido o sonido adicional por parte del público en las actuaciones en directo (como aplausos) o el sonido ambiental o ruido ambiental añadido con intención artística (como el ruido de la grabación, canto de los pájaros, los sonidos de grillos chirriando).

15 **[0013]** A modo de ilustración, la figura 7 representa la versión original de dos canales (de una grabación de audio). La figura 8 muestra una interpretación *upmixed* utilizando el concepto Directo / Ambiente.

Concepto en la banda

20 **[0014]** Siguiendo el concepto circundante, a menudo referido como "concepto en-la-banda", cada sonido o ruido (sonido directo, así como el ruido ambiental) puede ser completamente y/ o arbitrariamente posicionado alrededor del oyente. La posición del ruido o sonido es independiente de sus propiedades (sonido directo o ruido directo o sonido ambiental o ruido ambiental) y depende solamente del diseño específico del algoritmo y sus ajustes de parámetros.

[0015] La figura 9 representa el concepto envolvente.

25 **[0016]** En resumen, las figuras 7, 8 y 9 muestran varios conceptos de playback. Aquí, las figuras 7, 8 y 9 describen donde el oyente percibe el origen del sonido (como un área de traza oscura). La figura 7 describe la percepción acústica durante la reproducción de música. La figura 8 describe la percepción acústica y/ o localización del sonido utilizando el concepto Directo / Ambiente. La figura 9 describe la percepción de sonido y/ o la localización del sonido utilizando el concepto envolvente.

30 **[0017]** La sección siguiente proporciona una visión general sobre la técnica anterior con respecto a un upmixing la señal de un canal o de dos canales para formar una versión de múltiples canales. La literatura enseña varios procedimientos para el upmixing de señales de un solo canal y las señales de múltiples canales.

Procedimientos no señaladaptivos

35 **[0018]** La mayoría de procedimientos para la generación de una así llamada señal "pseudo estereofónica" no son señaladaptivos. Esto significa que procesan cualquier señal mono de la misma manera, independientemente del contenido de la señal. Estos sistemas funcionan a menudo con estructuras de filtro simples y/ o retrasos de tiempo a fin de descorrelacionar las señales generadas. Un repaso general de este sistema se pueden encontrar, por ejemplo, en [1].

Procedimientos Señaladaptivos

40 **[0019]** Los decodificadores de matriz (tal como el decodificador Dolby Pro Logic II, descrito en [2], el decodificador DTS NEO:6, descrito, por ejemplo, en [3] o el decodificador Harman Kardon/Lexicon Logic 7, descrito, por ejemplo, en [4]) se encuentran en casi todos los receptores de audio / vídeo que se venden actualmente. Como un subproducto de su función real o prevista, estos decodificadores de matriz son capaces de realizar upmixing ciego.

[0020] Los decodificadores mencionados utilizan las diferencias inter-canal y los mecanismos señaladaptivos de dirección a fin de crear señales de salida multicanal.

45 *Extracción de ambiente y síntesis de señales estéreo para el upmixing de audio multi-canal*

[0021] proponen una técnica en el dominio de la frecuencia con el fin de identificar y extraer la información de ambiente en señales de audio estéreo (ver [5]).

50 **[0022]** El procedimiento se basa en calcular un índice de coherencia inter-canal y una función de asignación no lineal que consiste en permitir la determinación de regiones de tiempo-frecuencia principalmente a base de componentes de ambiente o porciones de ambiente en la señal de dos canales. A continuación, se sintetizan las señales de ambiente y se utilizan para alimentar los canales de sonido envolvente de un sistema de reproducción multi-canal.

Un procedimiento para convertir sonido estéreo a sonido multicanal

[0023] Irwan y Aarts muestran un procedimiento para convertir una señal de una representación estéreo a una representación de múltiples canales (ver [6]). La señal para los canales envolventes se calcula utilizando una técnica de correlación cruzada. Un análisis de componentes principales (PCA) se usa para calcular un vector que indica la dirección de la señal dominante. Este vector es entonces asignado a partir de una representación de dos canales a una representación de tres canales con el fin de generar los tres canales frontales.

Upmixing basado en upmixing

[0024] Souldre muestra un sistema que genera una señal de múltiples canales a partir de una señal estéreo (ver [7]). La señal se descompone en los llamados "flujos de fuente individuales" y "Flujos de ambiente". Basado en estos flujos, un así llamado "Motor estético" sintetiza la salida multicanal. Sin embargo, no se dan más detalles técnicos relativos a la etapa de descomposición y la etapa de síntesis.

Pseudostereofonía basada en claves espaciales

[0025] Un proceso cuasi-señaladaptivo pseudo-estereofónico se describe en Faller en [1]. Este procedimiento utiliza una señal mono y grabaciones estéreo de la misma señal. Se extraen información adicional espacial o señales espaciales de la señal estéreo y se utilizan para convertir la señal mono a una señal estéreo.

[0026] El artículo "A Hierarchical Lossless/Lossy Coding System for High Quality Audio up to 192 kHz Sampling 24 Bit Format" de A. Jin y otros describe un sistema de codificación de audio para su uso en la transmisión continua de alta calidad de flujos de datos de audio estéreo. El sistema proporciona tanto nivel de calidad sin pérdidas y como de pérdidas variables. La calidad es seleccionable para adaptarse a una amplia gama de redes IP de banda estrecha y ancho. Se describe un codificador escalable sin pérdida / con pérdidas, que comprende una compresión con pérdidas MPEG-4. Por otra parte, el codificador escalable comprende una conversión de una señal de error, un elemento de codificación sin pérdida, descodificación sin pérdida, una conversión inversa y una reconstrucción total o parcial.

[0027] El artículo "Ambience Extraction and Synthesis from Stereo Signals for Multi-Channel Audio Up-Mix" de C. Avendano y J.M. Jot describe una técnica en el dominio de frecuencia para identificar y extraer la información ambiente de señales de audio estéreo. El método se basa en el cálculo de un índice de coherencia entre canales y una función de asignación no lineal que permite determinar regiones tiempo-frecuencia que consisten en su mayoría en los componentes de ambiente en la señal de dos canales. Las señales de ambiente son a continuación sintetizadas y utilizadas para alimentar los canales de sonido envolvente de un sistema de reproducción multi-canal.

[0028] En el artículo "Pseudostereophony Revisited" de C. Faller se describen las técnicas para la conversión de las grabaciones mono a señales estéreo de dos o más canales en un escenario de sonido y de ambiente realistas. Se describe una síntesis mono a estéreo, así como un proceso pseudostereofónico automático.

[0029] Es el objeto de la presente invención proporcionar un concepto para generar una señal de ambiente a partir de una señal de audio así como un concepto para obtener una señal de audio multi-canal a partir de una señal de audio, permitiendo los conceptos la generación de una señal de ambiente a partir de una señal de un canal sin ninguna información previa sobre la señal de un canal, de manera que la señal de ambiente transmite una impresión auditiva particularmente buena.

[0030] Este objeto se consigue mediante un aparato para generar una señal de ambiente a partir de una señal de audio según la reivindicación 1, un aparato para obtener una señal de audio multi-canal a partir de una señal de audio según la reivindicación 19, un procedimiento para generar una señal de ambiente a partir de una señal de audio según la reivindicación 24, un procedimiento para obtener una señal de audio multi-canal a partir de una señal de audio según la reivindicación 25 así como un programa de ordenador según la reivindicación 26.

[0031] La presente invención proporciona un aparato para generar una señal de ambiente a partir de una señal de audio con medios para la compresión con pérdidas de una representación espectral de la señal de audio para obtener una representación comprimida de la señal de audio que describe una señal de audio comprimida. El aparato para generar la señal de ambiente también comprende medios para calcular una diferencia entre la representación comprimida de la señal de audio y la representación espectral de la señal de audio para obtener una representación de discriminación. El aparato para generar una señal de ambiente también comprende medios para proporcionar la señal de ambiente empleando la representación de discriminación.

[0032] Es una idea esencial de la presente invención que una señal de ambiente pueda ser generada a partir de una señal de audio de un canal particularmente eficiente mediante la determinación de una diferencia entre una representación comprimida de la señal de audio, que fue generada por la compresión con pérdidas de una representación original de la señal de audio, y la representación original de la señal de audio. Esto es, se ha demostrado que en el uso de la compresión con pérdidas, la diferencia entre la señal de audio original y la señal de audio comprimida con pérdidas obtenida a partir de la señal de audio original por la compresión con pérdidas describe sustancialmente señales de ambiente, es decir, por ejemplo, similares al ruido o señales similares al ambiente o no-localizables.

5 **[0033]** En otras palabras, cuando se realiza una compresión con pérdidas, la representación comprimida de la señal de audio comprende sustancialmente los eventos de sonido localizables o eventos de sonido directo. Esto se basa en el hecho de que los eventos de sonido localizables suelen disponer en particular de alta energía específica y también formas de onda específicamente características. Por lo tanto, se prefiere procesar las señales localizables por la compresión con pérdidas a fin de que la representación comprimida comprenda sustancialmente las señales localizables de alta energía o una forma de onda característica.

10 **[0034]** Sin embargo, en la compresión con pérdidas, las señales ambientales no localizables que suelen no presentar cualquier forma de onda característica específicamente están representadas en menor medida por la representación comprimida de las señales localizables. Así, se ha reconocido que la diferencia entre la representación de la señal de audio en la forma de compresión con pérdidas y la representación original de la señal de audio describe sustancialmente la porción no localizable de la señal de audio. Además, se ha reconocido que el uso de la diferencia entre la representación en forma de compresión con pérdidas de la señal de audio y la representación original de la señal de audio como señal de ambiente resulta en una impresión auditiva particularmente buena.

15 **[0035]** En otras palabras, se ha reconocido que la compresión con pérdidas de una señal de audio típicamente no incorpora, o sólo en un grado muy pequeño, la porción de señal de ambiente de la señal de audio y que, por lo tanto, en particular la diferencia entre la representación original de la señal de audio y la representación en forma de compresión con pérdidas de la señal de audio se aproxima bien a la porción de señal de ambiente de la señal de audio. Por lo tanto, el concepto de la invención tal como se define en la reivindicación 1 es adecuado para la extracción ciega de la porción de señal de ambiente de una señal de audio.

20 **[0036]** El concepto de la invención es particularmente ventajoso en que incluso se puede extraer una señal de ambiente de una señal de un canal sin la existencia de ninguna información auxiliar adicional. Además, el concepto de la invención consiste en pasos algorítmicamente simples, es decir, realizar una compresión con pérdidas y el cálculo de una diferencia entre la representación de la señal de audio en la forma de compresión con pérdidas y la representación original de la señal de audio. Además, el procedimiento de la invención es ventajoso en que no se introducen efectos de audio sintéticos en la señal de ambiente. Por lo tanto, la señal de ambiente puede estar libre de reverberación, ya que puede producirse en el contexto de los procedimientos convencionales para la generación de una señal de ambiente. Además, es de señalar que la señal de ambiente generada según la invención típicamente deja de tener porciones de alta energía que pueden interferir con la impresión auditiva tal como ocurre en el contexto de la compresión con pérdidas, estas porciones de alta energía están contenidas en la representación de la señal de audio en la forma de compresión con pérdidas y, por lo tanto, no se producen, o sólo muy ligeramente, en la diferencia entre la representación en forma de compresión con pérdidas y la representación original de la señal de audio.

25 **[0037]** En otras palabras, según la invención, la señal de ambiente contiene exactamente aquellas partes que se consideran indispensables para la representación del contenido de información en el contexto de la compresión con pérdida. Es exactamente esta información, sin embargo, que representa el ruido de fondo.

30 **[0038]** Por lo tanto, el concepto de la invención permite la separación constante de información localizable y el ruido de fondo utilizando la compresión con pérdidas, en el que el ruido de fondo, que es el que se suprime y/o elimina por compresión con pérdidas, sirve como señal de ambiente.

35 **[0039]** La presente invención también proporciona un aparato para obtener una señal de audio multi-canal que comprende una señal de altavoz frontal y una señal de altavoz posterior a partir de una señal de audio. Aquí, el aparato para obtener la señal de audio multicanal comprende un aparato para generar una señal de ambiente de la señal de audio tal como se ha descrito más arriba. El aparato para generar la señal de ambiente está configurado para recibir la representación de la señal de audio. El aparato para obtener la señal de audio multi-canal también comprende un aparato para proporcionar la señal de audio o una señal de audio derivada de esta como señal de altavoz frontal a sí como un aparato que proporciona la señal de altavoz posterior para proporcionar la señal de ambiente proporcionada por el aparato para generar la señal de ambiente o una señal derivada de esta como señal de altavoz posterior. En otras palabras, el aparato para derivar la señal de audio multi-canal utiliza la señal de ambiente generada por el aparato para generar una señal de ambiente como señal de altavoz trasero, mientras que el aparato para obtener la señal de audio multi-canal adicional utiliza la señal de audio original como señal de altavoz frontal, o como base para la señal de altavoz frontal. Por lo tanto, el aparato para obtener una señal de audio multi-canal en su conjunto es capaz de generar, en base a una única señal de audio original, tanto la señal del altavoz frontal como la señal de altavoz trasero de una señal de audio multicanal. Por lo tanto, se utiliza la señal de audio original para proporcionar la señal de altavoz frontal (o incluso directamente representa la señal de altavoz frontal), mientras que la diferencia entre una representación de la forma de compresión con pérdidas de la señal de audio original y una representación de la señal de audio original sirve para generar la señal de altavoz posterior (o incluso se utiliza directamente como señal de altavoz posterior).

40 **[0040]** Además, la presente invención proporciona procedimientos correspondientes a los aparatos de la invención en lo que se refiere a su funcionalidad.

45 **[0041]** La presente invención proporciona además un programa de ordenador para la realización de los procedimientos de la invención.

[0042] Las realizaciones preferidas particularmente ventajosas de la presente invención se definen en las reivindicaciones adjuntas.

[0043] A continuación, las realizaciones preferidas de la presente invención se discuten con respecto a los dibujos adjuntos, en los que:

5 La figura 1 es un diagrama de bloques de un aparato de la invención para generar una señal de ambiente de una señal de audio de acuerdo con una realización de la presente invención;

La figura 2 es un diagrama de bloques de un aparato de la invención para generar una señal de ambiente de una señal de audio de acuerdo con una realización de la presente invención;

10 La figura 3 es un diagrama de bloques detallado de un aparato de la invención para generar una señal de ambiente de una señal de audio de acuerdo con una realización de la presente invención;

La figura 4a es un ejemplo de representación de una representación aproximada de una matriz por un producto de dos matrices;

La figura 4b es una representación de una matriz X;

15 La figura 5 es un diagrama de bloques de un aparato de la invención para obtener una señal de audio de múltiples canales a partir de una señal de audio de acuerdo con una realización de la presente invención;

La figura 6 es un diagrama de flujo de un procedimiento de la invención para la creación de una señal de ambiente de una señal de audio de acuerdo con una realización de la presente invención;

La figura 7 es una representación esquemática de una impresión auditiva en un concepto de reproducción estéreo;

La figura 8 es una representación esquemática de una impresión auditiva en un concepto directo / ambiente, y

20 La figura 9 es una representación esquemática de una impresión auditiva en un concepto envolvente.

[0044] La figura 1 muestra un diagrama de bloques de un aparato de la invención para generar una señal de ambiente de una señal de audio de acuerdo con una realización de la presente invención.

25 **[0045]** El aparato según la figura 1 es designado en su totalidad con 100. El aparato 100 está configurado para recibir una señal de audio en una representación que, básicamente, se puede seleccionar arbitrariamente. En otras palabras, el aparato 100 recibe una representación de una señal de audio. El aparato 100 comprende medios 110 para la compresión con pérdidas de la señal de audio o la representación de la señal de audio. Los medios 110 están configurados para recibir la representación 108 de la señal de audio. Los medios 110 generan a partir de la representación (original) 108 de la señal de audio una representación en forma de compresión con pérdidas 112 de la señal de audio.

30 **[0046]** El aparato 100 también comprende medios 120 para calcular una diferencia entre la representación 112 de la señal de audio en la forma de compresión con pérdidas de la señal de audio y la representación (original) 108. Por lo tanto, los medios 120 están configurados para recibir la representación en la forma de compresión con pérdidas 112 de la señal de audio, así como, además, la representación (original) 108 de la señal de audio. A partir de la representación (Original) 108 de la señal de audio y la representación de la forma de compresión con pérdidas 112 de la señal de audio, los medios 120 calculan una representación de discriminación 122 que describe una diferencia entre la representación (original) 108 de la señal de audio y la representación en la forma de compresión con pérdidas 112 de la señal de audio.

[0047] El aparato 100 comprende además medios 130 para proporcionar la señal de ambiente 132 usando y/o sobre la base de y/o como una función de la representación de discriminación 122.

40 **[0048]** Basándose en la descripción estructural anterior del aparato 100, a continuación se describe brevemente el funcionamiento del aparato 100. El aparato 100 recibe una representación 108 de una señal de audio. Los medios 110 generan una representación en la forma de compresión con pérdidas 112 de la señal de audio. Los medios 120 calculan una representación de discriminación 122 que describe una diferencia entre la representación 108 de la señal de audio y la representación en la forma de compresión con pérdidas 112 de la señal de audio y/o que es una función de la diferencia mencionada. Dicho de otro modo, la representación de discriminación 122 describe aquellas partes de señal de la señal de audio (original) descritas por la representación 108, que se quitan y/o que no se reproducen en la representación de la forma de compresión con pérdidas 112 de la señal de audio por los medios 110 para la compresión con pérdida. Como, típicamente, los medios 110, eliminan y/o no reproducen exactamente aquellas partes de la señal que exhiben una curva irregular en la representación de la forma de compresión con pérdidas 112 de la señal de audio, la representación de discriminación 122 describe exactamente a aquellas porciones de señal que tienen una curva irregular o una distribución de energía irregular, es decir, por ejemplo, similar a partes de señal de ruido. Como, normalmente, las porciones directas y/o "partes de señal localizables", que son de particular importancia para el oyente, son reproducidas por los altavoces frontales (y no por los altavoces posteriores), la representación de discriminación

122 está, en relación con esta materia, adaptada a los requisitos de la reproducción de audio. Por lo tanto, las porciones directas y/ o porciones localizables de la señal de audio original están contenidas en la representación en la forma de compresión con pérdidas 112 de la señal de audio de una manera sustancialmente no corrompida, y por lo tanto sustancialmente su primida en la representación de discriminación 122 tal como se busca. Por otra parte, en la representación de la forma de compresión con pérdidas 112 de la señal de audio, se reducen las porciones de información que tienen energía distribuida irregularmente y/ o poca localizabilidad. La razón es que en la compresión con pérdidas, tal como se realiza por los medios 110 para la compresión con pérdidas, la información de energía distribuida regularmente y/ o que tiene alta energía se transfiere a la representación en la forma de compresión con pérdidas 112 de la señal de audio, mientras que las porciones de la señal (original) de audio que tienen energía de distribución irregular y/ o una energía más baja se transfieren a la representación en la forma de compresión con pérdidas 112 de la señal de audio en una forma atenuada o solamente muy poco. Como resultado de ello, por medio de la atenuación de las porciones de señal que tienen una distribución de energía irregular y/ o de las partes de señal de baja energía de la señal de audio que se produce en el contexto de la compresión con pérdidas, la representación de discriminación 112 todavía comprende comparativamente una gran porción de las partes de señal de baja energía y/ o porciones de señal que tienen la energía de distribución irregular. Exactamente estas partes de señal no muy ricas en porciones de energía y/ o señal con energía distribuida irregularmente, tal como se describen por la representación de discriminación 122, representan la información que resulta en una impresión auditiva particularmente buena y agradable en la reproducción (por medio de los altavoces posteriores).

[0049] Para resumir, se puede afirmar que en la representación de discriminación 122, las porciones de señal que tienen la energía distribuida regularmente (es decir, por ejemplo, señales localizables) son suprimidas o atenuadas. En contraste con esto, en la representación de discriminación 122, las porciones de señal que tienen la energía de distribución irregular (tal como señales no localizables) no se suprimen y no se atenúan. Por lo tanto, en la representación de la discriminación, las porciones de señal que tienen distribución irregular de energía se destacan o acentúan en comparación con las porciones de señal que tienen energía distribuida regularmente. Por lo tanto, la representación de la discriminación es particularmente adecuada como señal de ambiente.

[0050] Dicho de otro modo, en una realización preferida, todo lo que aparece en varias ocasiones en la representación tiempo-frecuencia está bien aproximado por la compresión con pérdida.

[0051] La distribución de energía regular está aquí destinada a ser, por ejemplo, la distribución de energía que presenta un patrón recurrente en una representación tiempo-frecuencia o que proporciona una concentración local de energía en la representación tiempo-frecuencia.

[0052] Dicho de otro modo, en una realización preferida, la señal de ambiente comprende sustancialmente porciones de señal que tienen una distribución de energía no estructurada (por ejemplo no estructurada en la distribución tiempo-frecuencia), mientras que la representación en la forma de compresión con pérdidas de la señal de audio comprende sustancialmente partes de señal que tienen una distribución de energía estructurada (por ejemplo, estructura en la representación tiempo-frecuencia tal como se describió anteriormente).

[0053] Por lo tanto, los medios 130 para proporcionar la señal de ambiente sobre la base de la representación de discriminación 122 proporcionan una señal de ambiente que está particularmente bien adaptada a las expectativas de un oyente humano.

[0054] Los medios 110 para la compresión con pérdidas pueden ser también, por ejemplo, un compresor de audio MP3, un compresor de audio MP4, un compresor de audio ELP o un compresor de audio SPR.

[0055] A continuación y con respecto a las figuras 2 y 3, se describe con mayor detalle una realización de la presente invención. Para este fin, la figura 2 muestra un diagrama de bloques de un aparato de la invención para generar una señal de ambiente de una señal de audio de acuerdo con una realización de la presente invención. Además, la figura 3 muestra un diagrama de bloques detallado de un aparato de la invención para generar una señal de ambiente de una señal de audio de acuerdo con una realización de la presente invención. En su totalidad, el aparato según la figura 2 se designa con 200, y, en su totalidad, el aparato según la figura 3 se designa con 300.

[0056] El aparato 200 está configurado para recibir una señal de entrada presente, por ejemplo, en forma de una representación en el tiempo $x[n]$. La señal de entrada 208 describe típicamente una señal de audio.

[0057] Los medios 200 comprenden un proveedor de distribución de tiempo-frecuencia 210. El proveedor de distribución de tiempo-frecuencia 210 está configurado para generar una distribución tiempo-frecuencia (TFD) de la señal de entrada 208 presente en una representación de tiempo $x[n]$. Es de señalar que el proveedor de distribución de tiempo-frecuencia 210 es opcional. Es decir, una representación 212 de una representación tiempo-frecuencia también puede servir como señal de entrada del aparato 200 de modo que en este caso la conversión de la señal de entrada 208 ($x[n]$) , que está presente como una señal de tiempo, 212 a la representación de la distribución tiempo-frecuencia puede omitirse.

[0058] Se ha de señalar además que la representación 212 de la distribución tiempo-frecuencia puede, por ejemplo, estar presente en forma de una matriz de distribución tiempo-frecuencia. Además, es de señalar que, por ejemplo, la

matriz $\mathbf{X}(\omega, k)$, que se explicará con mayor detalle en la siguiente, o bien la matriz $|\mathbf{X}(\omega, k)|$ puede servir como la representación 212 de la distribución tiempo-frecuencia.

5 **[0059]** Los medios 200 comprenden, además, medios de aproximación 220, configurados para recibir la representación 212 de la distribución tiempo-frecuencia, y para generar una representación aproximada 222 de la representación tiempo-frecuencia 212 que es típicamente comprimida con pérdidas en comparación con la representación 212. En otras palabras, la aproximación o la representación aproximada 222 de la distribución tiempo-frecuencia 212 se realiza mediante los medios de aproximación 220, por ejemplo usando un procedimiento de optimización numérica, tal como se describirá con más detalle a continuación. Se asume, sin embargo, que la aproximación produce una desviación entre la representación (original) 212 de la distribución tiempo-frecuencia (que es una representación original de la señal de audio) y la representación aproximada 222 de la distribución tiempo-frecuencia. En una realización de la presente invención, la diferencia entre la representación original 212 y la representación aproximada 222 de la distribución tiempo-frecuencia se basa en el hecho de que los medios 220 de aproximación están configurados preferentemente para realizar una aproximación con pérdidas, en la que se prefiere pasar partes de la señal que presentan una distribución regular de la energía y/o llevan una gran energía de la señal los que se deberá pasar a la representación aproximada, mientras que las porciones de señal que exhiben energía distribuida irregularmente comparativamente y/o energía señal comparativamente menor son atenuadas o confundidas en la representación aproximada 222 en comparación con las porciones de señal que tienen la energía regularmente distribuida y/o una señal de energía grande.

20 **[0060]** El aparato 200 comprende además un determinador de diferencia 230 configurado para recibir la representación original 212 de la distribución tiempo-frecuencia, así como la representación aproximada 222 de la representación tiempo-frecuencia para generar, sobre la base de una diferencia entre la representación original 212 y la representación aproximada 222, una representación de discriminación 232 que esencialmente describe la diferencia entre la representación original 212 y la representación aproximada 222 y/o que es una función de la diferencia entre la representación original 212 y la representación aproximada 222. Los detalles sobre el cálculo de la representación de discriminación 232 se explicarán a continuación.

25 **[0061]** El aparato 200 comprende además medios de re-síntesis 240. Los medios de re-síntesis 240 están configurados para recibir la representación de discriminación 232 a fin de generar una nueva señal sintetizada 242 basándose en esta. Los medios de re-síntesis 240 pueden por ejemplo ser configurados para convertir la representación de discriminación 232, que está presente en la forma de una distribución tiempo-frecuencia, en una señal temporal 242.

30 **[0062]** Se ha de señalar además que los medios de re-síntesis 240 son opcionales y pueden ser omitidos si hay reprocesamiento directo de la representación de discriminación 232, que puede, por ejemplo, estar presente en la forma de una distribución tiempo-frecuencia, si se desea.

35 **[0063]** Los medios 200 comprenden además medios 250 opcionales para el montaje de una señal de audio de múltiples canales y/o para el procesamiento. Los medios 250 están, por ejemplo, configurados para recibir la señal re-sintetizada 242 desde los medios 240 para la re-síntesis y para generar una pluralidad de señales de ambiente 252, 254 (también indicadas con $a_i[n], \dots, a_k[n]$) de la señal re-sintetizada 242.

[0064] La generación de la pluralidad de las señales ambientales 252, 254 se explicará con mayor detalle a continuación.

40 **[0065]** En resumen, se muestra que la presente invención se refiere sustancialmente al cálculo de una señal de ambiente. El diagrama de bloques de la figura 2 ha servido para proporcionar una breve descripción del concepto inventivo y el aparato de la invención y el procedimiento de la invención de acuerdo con una realización de la presente invención. El concepto de la invención se puede resumir en breve tal como sigue:

Una distribución tiempo-frecuencia 212 (TFD) de la señal de entrada 208 ($x[n]$) es (opcionalmente) calculada en un os medios (opcionales) 210 para la determinación de la distribución tiempo-frecuencia. El cálculo se explicará con mayor detalle a continuación. Una aproximación 220 de la distribución tiempo-frecuencia 212 (TFD) de la señal de entrada 208 ($x[n]$) se calcula, por ejemplo, utilizando un procedimiento de aproximación numérica que se describirá con mayor detalle a continuación. Este cálculo puede, por ejemplo, llevarse a cabo en los medios 220 para la aproximación. Mediante el cálculo de una distinción o diferencia entre la distribución tiempo-frecuencia 212 (TFD) de la señal de entrada 208 ($x[n]$) y su aproximación 212 (por ejemplo, en los medios 230 para el cálculo de una diferencia), se obtiene una estimación 232 de una distribución de tiempo-frecuencias (TFD) de la señal de ambiente. Acto seguido, se lleva a cabo una re-síntesis de una señal temporal 242 de la señal de ambiente (por ejemplo, en los medios de re-síntesis opcionales 240). La re-síntesis se explicará con mayor detalle a continuación. Además, el uso opcional está hecha de postprocesado (realizada por ejemplo en los medios opcionales 250 para la síntesis de una señal de audio multicanal y/o para el postprocesamiento) a fin de mejorar la impresión auditiva de la señal multicanal obtenida (que consiste en, por ejemplo, en señales de ambiente 252, 254). El procesamiento posterior opcional también se explicará con mayor detalle a continuación.

[0066] Los detalles relativos a las etapas de procesamiento individuales que se muestran en el contexto de la figura 2 se explicarán a continuación. De este modo, también se hace referencia a la figura 3, que muestra un diagrama de bloques más detallado de un aparato de la invención para generar una señal de ambiente de una señal de audio.

5 **[0067]** El aparato 300 de acuerdo con la figura 3 está configurado para recibir una señal de entrada 308 presente, por ejemplo, en la forma de una señal de entrada $x(t)$ continua en el tiempo o en la forma de una señal de entrada de tiempo discreto $x[n]$. De lo contrario, la señal de entrada 308 corresponde a la señal de entrada 208 del aparato 200.

10 **[0068]** El aparato 300 comprende además un convertidor de señal de tiempo a distribución de tiempo-frecuencia de 310. El convertidor de señal de tiempo a distribución de tiempo-frecuencia de 310 está configurado para recibir la señal de entrada 308 y para proporcionar una representación de una distribución tiempo-frecuencia (TFD) 312. La representación de la distribución tiempo-frecuencia 312 corresponde de otro modo sustancialmente a la representación 212 de la distribución tiempo-frecuencia en el aparato 200. Se ha de señalar además que en lo que sigue, la distribución tiempo-frecuencia también se denota con $\mathbf{X}(\omega, k)$.

15 **[0069]** Se ha de señalar además que la distribución tiempo-frecuencia $|\mathbf{X}(\omega, k)|$ también puede ser la señal de entrada del aparato 300, es decir, puede omitirse el aparato 310. El aparato 300 comprende además (opcionalmente) un divisor de magnitud fase 314. El divisor de magnitud fase 314 se utiliza preferentemente cuando la distribución tiempo-frecuencia 312 puede adoptar valores complejos (no puramente reales). En este caso, el divisor de magnitud fase 314 está configurado preferentemente para proporcionar una representación de la magnitud 316 de la distribución tiempo-frecuencia 312, así como una representación de fase 318 de la distribución tiempo-frecuencia 312, basada en la distribución tiempo-frecuencia 312. De lo contrario, la magnitud de la representación de distribución tiempo-frecuencia 312 también se designa con $|\mathbf{X}(\omega, k)|$. Es de señalar que la representación de la magnitud 316 de la distribución tiempo-frecuencia 312 puede ser sustituida por la representación 212 en el aparato 200.

20 **[0070]** Además, es de señalar que el uso de la representación de la fase 318 de la distribución tiempo-frecuencia 312 es opcional. También es de señalar que la representación de fase 318 de la distribución tiempo-frecuencia 312 se designa en algunos casos, también con $\phi(\omega, k)$.

25 **[0071]** Se su pone además que la representación de la magnitud 316 de la distribución tiempo-frecuencia 312 está presente en forma de matriz.

30 **[0072]** El aparato 300 comprende además un aproximador de matriz 320 configurado para aproximar la representación de magnitud 316 de la distribución tiempo-frecuencia 312 por un producto de dos matrices $\mathbf{W}\cdot\mathbf{H}$, tal como se describe a continuación. El aproximador de matriz 320 corresponde sustancialmente a los medios 220 para la aproximación como se utilizan en el aparato 200. Por lo tanto, el aproximador de matriz 320 recibe la representación de la magnitud 316 de la distribución tiempo-frecuencia 312 y proporciona una aproximación 322 de la representación de magnitud 316. La aproximación 322 también se designa en algunos casos con $\mathbf{X}(\omega, k)$. En otros casos la aproximación 322 corresponde a la representación aproximada 222 de la figura 2.

35 **[0073]** El aparato 300 comprende además un formador de diferencia 330 que recibe tanto la representación de magnitud 316 como la aproximación 322. Además, el formador de diferencia 330 proporciona una representación de discriminación 332 que corresponde sustancialmente a la representación $|\mathbf{A}(\omega, k)|$ descrita a continuación. De lo contrario, hay que señalar que la representación de discriminación 332 también corresponde sustancialmente a la representación de discriminación 232 en el aparato 200.

40 **[0074]** El aparato 300 comprende además un sumador de fase 334. El sumador de fase 334 recibe la representación de discriminación 332, así como la representación de fase 318 y por lo tanto añade una fase a los elementos de la representación de discriminación 332 tal como se describe por la representación de fase 318. Por lo tanto, el sumador de fase 334 proporciona una representación de discriminación 336 provista de una fase, que también se designa con $\mathbf{A}(\omega, k)$. Es de señalar que el sumador de fase 334 puede ser considerado como opcional, de modo que, si se omite el sumador de fase 334, la representación de discriminación 332 puede ser, por ejemplo, sustituida por la representación de discriminación 336 provista de una fase. Se ha de señalar además que, en función de cada caso particular, tanto la representación de discriminación 332 y la representación de discriminación 336 provista de una fase pueden corresponder a la representación de discriminación 232.

45 **[0075]** El aparato 300 comprende además un convertidor (opcional) de distribución tiempo-frecuencia a señal de tiempo de 340. El convertidor (opcional) de distribución tiempo-frecuencia a señal de tiempo de 340 está configurada para recibir la representación de discriminación 336 provista de una fase (como alternativa: la representación de discriminación 332) y proporcionar una señal temporal 342 también de signada con $a(t)$ o $a[n]$ formando así una representación en el dominio de tiempo (o representación en señal de tiempo) de la señal de ambiente.

50 **[0076]** Se tiene que observar además que el convertidor (opcional) de distribución tiempo-frecuencia a señal de tiempo de 340 corresponde sustancialmente a los medios de re-síntesis 240 según la figura 2. Además, la señal 342 proporcionada por el convertidor de distribución tiempo-frecuencia a señal de tiempo de 340 corresponde sustancialmente a la señal 242, tal como se muestra en el aparato 200.

Distribución tiempo-frecuencia de la señal de entrada

[0077] A continuación se describe la manera en que se puede calcular una distribución tiempo-frecuencia (TFD) de la señal de entrada, es decir, por ejemplo, una representación 212, 312. Las distribuciones tiempo-frecuencia (TFD) son representaciones y/o ilustraciones de una señal temporal (es decir, por ejemplo, de la señal de entrada 208 o la señal de entrada 308) tanto frente al tiempo como a la frecuencia. Entre las formulaciones múltiples de una distribución tiempo-frecuencia (por ejemplo, utilizando un banco de filtros o una transformada discreta del coseno (DCT)), la transformada de Fourier de corto tiempo (STFT) es un procedimiento flexible y eficiente computacionalmente para el cálculo de la distribución tiempo-frecuencia. La transformada de Fourier de corto tiempo (STFT) $X(\omega, k)$ con el contenedor de frecuencia o índice de frecuencia ω y el índice de tiempo k se calcula como una secuencia de transformadas de Fourier de los segmentos de datos de ventana de la señal temporal discreta $x[n]$ (es decir, por ejemplo, de la señal de entrada 208, 308). Por lo tanto, se cumple lo siguiente:

$$X(\omega, k) = \sum_{n=-\infty}^{\infty} x[n]w[n-m]e^{-j\omega n} \quad (1)$$

[0078] Aquí, $w[n]$ representa la función de ventana. La relación entre el índice m y el índice de trama (o índice de tiempo) k es una función de la longitud de la ventana y la cantidad de superposición de ventanas adyacentes.

[0079] Si la distribución tiempo-frecuencia (TFD) es de valores complejos (por ejemplo, en el caso de la utilización de una transformada de Fourier corto tiempo de (STFT)), en una realización preferida, el cálculo adicional se puede efectuar usando valores absolutos de los coeficientes de la distribución tiempo-frecuencia (TFD). Los valores y/o magnitudes absolutos de los coeficientes de la distribución tiempo-frecuencia (TFD) también se designan con $|X(\omega, k)|$. En este caso, una información de fase $\phi(\omega, k) = \angle X(\omega, k)$ se almacena en la etapa de re-síntesis para su uso posterior. Es de señalar que en el aparato 300 la representación de magnitud $X(\omega, k)$ se designa con 316. La información de fase $\phi(\omega, k)$ se designa con 318.

[0080] Es de señalar que $X(\omega, k)$ denota coeficientes de Fourier individuales (en general: coeficientes individuales de una distribución tiempo-frecuencia), ya que pueden obtenerse, por ejemplo, por la STFT. En contraste, $\mathbf{X}(\omega, k)$ denota una matriz que contiene una pluralidad de coeficientes (ω, k) . Por ejemplo, la matriz $\mathbf{X}(\omega, k_1)$ por ejemplo, la matriz $\mathbf{X}(\omega', k')$ for $\omega' = 1, 2, \dots, n$ y $k' = k_1, k_1 + 1, \dots, k_1 + m - 1$. Aquí, n es una primera dimensión de la matriz $\mathbf{X}(\omega, k_1)$, por ejemplo un número de filas, y m es una segunda dimensión de la matriz $\mathbf{X}(\omega, k_1)$. Por lo tanto, para un elemento $X_{i,j}$ de la matriz $\mathbf{X}(\omega, k_1)$ se cumple lo siguiente:

$$X_{i,j} = X(\omega = \omega_i, k = k_{1+j-1}).$$

[0081] Aquí, se cumple lo siguiente:

$$1 \leq j \leq n$$

y

$$1 \leq i \leq m.$$

[0082] El contexto descrito también se muestra en la figura 4b.

[0083] En otras palabras, la matriz $\mathbf{X}(\omega, k)$ comprende una pluralidad de valores de distribución de tiempo frecuencia $X(\omega, k)$.

[0084] Se ha de señalar además que en el siguiente, el cálculo de una magnitud de una matriz, designada con $|\mathbf{X}|$, denota una formación de magnitud a nivel de elemento a menos que se represente de otro modo.

Aproximación de la distribución tiempo-frecuencia (TFD)

[0085] En el contexto de la presente invención, de acuerdo con una forma de realización, se calcula una aproximación de la distribución tiempo-frecuencia de la señal de entrada utilizando un procedimiento de optimización numérica. La aproximación de la distribución tiempo-frecuencia, así como el procedimiento de optimización numérica se describen a continuación.

[0086] Una aproximación $\hat{X}(\omega, k)$ de la matriz $\mathbf{X}(\omega, k)$ se obtiene con la ayuda de un procedimiento de optimización numérica minimizando el error de la aproximación. Aquí, la minimización significa una minimización con un error relativo de no más de 50%, preferentemente no más de 20%. De lo contrario, una minimización puede ser una determinación de un mínimo absoluto o local.

[0087] De lo contrario, el error de aproximación se mide con la ayuda de una función de distancia o una función de la divergencia. La diferencia entre una distancia y una divergencia es de naturaleza matemática y se basa en el hecho de que una distancia es simétrica en el sentido de que para una distancia entre dos matrices **A**, **B** se cumple lo siguiente:

$$d(\mathbf{A}, \mathbf{B}) = d(\mathbf{B}, \mathbf{A}).$$

5 [0088] En contraste con esto, la divergencia puede ser asimétrica.

[0089] Es de señalar que la aproximación de la distribución tiempo-frecuencia o la matriz de tiempo-frecuencia de distribución de $\mathbf{X}(\omega, k)$ que se describe a continuación puede, por ejemplo, ser efectuada por medio de los métodos de aproximación 220 o el aproximador de matriz 320.

10 [0090] Se ha de señalar además que la matriz de factorización no negativa (NMF) es un procedimiento adecuado para el cálculo de la aproximación.

Matriz de factorización no negativa (NMF)

[0091] A continuación, se describe la matriz de factorización no negativa. Una matriz de factorización no negativa (NMF) es una aproximación de una matriz $\mathbf{V} \in \mathbf{R}^{n \times m}$ con elementos no negativos, como un producto de dos matrices $\mathbf{W} \in \mathbf{R}^{n \times r}$ y $\mathbf{H} \in \mathbf{R}^{r \times m}$. En este caso, para los elementos $W_{i,k}$ de la matriz **W** y $H_{i,k}$ de la matriz **H**, se cumple lo siguiente:

15
$$W_{i,k} \geq 0;$$

y

$$H_{i,k} \geq 0.$$

[0092] Dicho de otro modo, las matrices **W** y **H** se determinan de tal manera que se cumple lo siguiente:

20
$$\mathbf{V} \approx \mathbf{WH}$$

[0093] Expresando este elemento de forma amplia, se cumple lo siguiente:

$$V_{i,k} \approx (\mathbf{WH})_{i,k} = \sum_{a=1}^r W_{i,a} H_{a,k} \quad (2)$$

[0094] Si el rango r de la factorización satisface la condición

$$(n + m) r < nm$$

25 entonces el producto **WH** es una representación de datos comprimida de **V** (see[8]). Una explicación intuitiva de la ecuación (2) es la siguiente: la matriz $\mathbf{V} \in \mathbf{R}^{n \times m}$ se aproxima como la suma de los de r productos externos de un vector columna w_i y un vector fila h_i , en el que se cumple lo siguiente: $i \in [1, r]$, $w_i \in \mathbf{R}^{n \times 1}$ y $h_i \in \mathbf{R}^{1 \times m}$. El objeto descrito está representado por un simple ejemplo en la figura 4a. Dicho de otro modo, la figura 4a muestra un ejemplo ilustrativo de una matriz de factorización no negativa (NMF), con un rango r de factorización $r = 2$.

30 [0095] Los factores **W** y **H** se calculan resolviendo el problema de optimización de la minimización de una función de coste $c = f(\mathbf{V}, \mathbf{WH})$ que mide el error de la aproximación. Dicho de otro modo, la función de coste c mide el error de la aproximación, es decir, la distancia (la divergencia) entre las matrices **V** y **WH**. Una medida de la distancia adecuada entre las dos matrices **A** y **B** es la norma de Frobenius $D_F(\mathbf{A}, \mathbf{B})$ en su diferencia por elemento (ecuación 3):

$$D_F(\mathbf{A}, \mathbf{B}) = \|\mathbf{A} - \mathbf{B}\|_F^2 = \sum_{i,k} (A_{i,k} - B_{i,k})^2 \quad (3)$$

35 [0096] La norma de Frobenius es ideal para, datos distribuidos según Gauss no correlacionados (ver [9]). Dicho de otro modo, una función de coste c se calcula en la forma de realización preferida, en la que se cumple lo siguiente:

$$c = D_F(\mathbf{X}(\omega, k), \hat{\mathbf{X}}(\omega, k)).$$

[0097] En otras palabras, la aproximación $\hat{\mathbf{X}}(\omega, k)$ se calcula como el producto de dos matrices, **W** y **H**, en el que:

$$\hat{\mathbf{X}}(\omega, k) = \mathbf{W}\mathbf{H}.$$

[0098] Una función más conocida es el error de divergencia generalizada de Kullback-Leibler (GKLD) (ecuación 4). La divergencia generalizada de Kullback-Leibler (GKLD) está más relacionada con una distribución de Poisson (véase [9]) o una distribución exponencial y por lo tanto incluso más adecuada para una aproximación de los espectros de cantidad o magnitud de las señales de audio musicales. La definición de la divergencia generalizada Kullback-Leibler entre dos matrices A y B es la siguiente:

$$D_{GKL}(\mathbf{A}, \mathbf{B}) = \sum_{i,j} \left(A_{ij} \log \frac{A_{ij}}{B_{ij}} - A_{ij} + B_{ij} \right) \quad (4)$$

[0099] De lo contrario, A_{ij} y B_{ij} son las entradas o elementos de matriz de las matrices \mathbf{A} y \mathbf{B} , respectivamente.

[0100] Dicho de otro modo, la función de coste c se puede seleccionar de la siguiente manera:

$$c = D_{GKL}(\mathbf{X}, \hat{\mathbf{X}} = \mathbf{W}\mathbf{H}).$$

[0101] Lo que sigue es una descripción de cómo se pueden determinar las entradas de las matrices de aproximación \mathbf{W} y \mathbf{H} . Una técnica simple de optimización numérica conocida como descenso de gradiente iterativamente se acerca a un mínimo local (o global) de la función de coste $f(\mathbf{x})$ mediante la aplicación de la regla de actualización y/o regla de iteración

$$\mathbf{x} \leftarrow \mathbf{x} + \alpha \cdot \nabla f(\mathbf{x}) \quad (5)$$

con el tamaño de paso α y el gradiente $\nabla f(\mathbf{X})$ de la función de coste.

[0102] Para el problema de optimización de acuerdo con la ecuación (2) con la función de coste de acuerdo con la ecuación (3), la regla de actualización aditiva o regla de iteración está dada por las siguientes ecuaciones:

$$H_{ik} \leftarrow H_{ik} + \alpha \cdot \left[(\mathbf{W}^T \mathbf{V})_{ik} - (\mathbf{W}^T \mathbf{W}\mathbf{H})_{ik} \right] \quad (6)$$

$$W_{ik} \leftarrow W_{ik} + \alpha \cdot \left[(\mathbf{V}\mathbf{H}^T)_{ik} - (\mathbf{W}\mathbf{H}\mathbf{H}^T)_{ik} \right] \quad (7)$$

[0103] En el contexto del algoritmo de la invención, en una realización, se cumple lo siguiente:

$$\mathbf{V} = \mathbf{X}(\omega, k).$$

[0104] Se ha de señalar además que Lee y Seung han encontrado o identificado una regla de actualización multiplicativa o regla de iteración de acuerdo con las ecuaciones (8) y (9) (ver [10]). Además, Lee y Seung han demostrado la relación de la regla de actualización multiplicativa para el procedimiento de descenso de gradiente y la convergencia de los mismos. Las reglas de actualización multiplicativas son como sigue:

$$H_{ik} \leftarrow H_{ik} \frac{(\mathbf{W}^T \mathbf{V})_{ik}}{(\mathbf{W}^T \mathbf{W}\mathbf{H})_{ik}} \quad (8)$$

$$W_{ik} \leftarrow W_{ik} \frac{(\mathbf{V}\mathbf{H}^T)_{ik}}{(\mathbf{W}\mathbf{H}\mathbf{H}^T)_{ik}} \quad (9)$$

[0105] Una vez más, en una realización preferida, se cumple lo siguiente:

$$\mathbf{V} = \mathbf{X}(\omega, k).$$

[0106] La velocidad y la eficacia del procedimiento de descenso de gradiente depende fuertemente de la elección correcta del tamaño del paso o anchura de paso [alfa]. Una ventaja principal de la regla de actualización multiplicativa sobre el procedimiento de descenso de gradiente es la independencia de la elección del tamaño del paso o la anchura de paso. El procedimiento y el procedimiento es fácil de implementar, computacionalmente eficiente y garantiza la búsqueda de un mínimo local de la función de coste.

Matriz de factorización no negativa (NMF) en el contexto de la separación ambiente

[0107] En el contexto del procedimiento presentado, se utiliza una matriz de factorización no negativa (NMF) para calcular una aproximación del espectrograma de cantidad o magnitud $|X(\omega, k)|$ de la señal de audio de entrada $x[n]$. En relación con esto, es de señalarse que el espectrograma de magnitud $X(\omega, k)$ se deriva de la matriz $X(\omega, k)$ realizando una formación de magnitud que abarca a todo el elemento. Dicho de otro modo, para el elemento que tiene los índices i, j de $|X(\omega, k)|$, designados con $|X(\omega, k)|_{ij}$, se cumple lo siguiente:

$$|X(\omega, k)|_{ij} = |X(\omega, k)_{ij}|.$$

[0108] $X(\omega, k)_{ij}$ designa aquí un elemento de la matriz $X(\omega, k)$ con los índices i y j . $| \cdot |$ de lo contrario designa la operación de formación de magnitud.

[0109] La matriz de factorización no negativa (NMF) de $|X|$ resultados en los factores W y H . En una realización preferida, se requiere un amplio rango de factorización r entre 40 y 100, dependiendo de la longitud de la señal y el contenido de la señal, para representar una cantidad suficiente de sonido directo o ruido partecular mediante la aproximación.

[0110] En resumen, se muestra que la matriz de factorización no negativa d escrita anteriormente se consigue sustancialmente una representación aproximada de la distribución tiempo-frecuencia, que se designa con 222, por ejemplo, en el aparato 200 de acuerdo con la figura 2, y tal como se designó además con 322 o $|X(\omega, k)|$ en el aparato 300 de acuerdo con la figura 3. Un espectrograma de cantidad o magnitud $|A|$ de la señal de ambiente se obtiene básicamente mediante el cálculo de la diferencia entre la representación de cantidad o magnitud $|X|$ de la distribución tiempo-frecuencia X y su aproximación WH , tal como se representa en la ecuación (10):

$$|A| = |X| - WH \tag{10}$$

[0111] Sin embargo, en una realización preferida, el resultado de acuerdo con la ecuación 10 no se considera directamente tal como se explicará a continuación. Esto es, para aproximaciones que minimizan las funciones de coste descritas anteriormente, la aplicación de la ecuación (10) resulta en un espectrograma de cantidad o magnitud A con elementos tanto negativos como positivos. Sin embargo, en una realización preferida en la que el espectrograma de cantidad o magnitud $|A|$ incluye solamente elementos de valor positivo, se prefiere emplear un procedimiento que controla los elementos de valor negativo de la diferencia $|X| - WH$.

[0112] Pueden emplearse varios procedimientos para la manipulación de los elementos negativos. Un procedimiento sencillo para la manipulación de los elementos negativos consiste en multiplicar los valores negativos con un factor β entre 0 y -1 ($\beta = 0, \dots, -1$). Dicho de otro modo: $-1 \leq \beta \leq 0$. Aquí, $\beta = 0$ corresponde a una rectificación de media onda, y $\beta = -1$ corresponde a una rectificación de onda completa.

[0113] Una formulación general para el cálculo del espectrograma de magnitud o el espectrograma amplitud $|A|$ de la señal de ambiente está dada por las siguientes ecuaciones:

$$|A|_{ik} = \beta_{ik} \cdot (|X| - WH)_{ik} \tag{11}$$

con

$$\beta_{ik} = \begin{cases} \gamma, & \text{if } (WH)_{ik} > |X|_{ik} \\ +1, & \text{otherwise} \end{cases} \tag{12}$$

en la que $\gamma \in [-1, 0]$ es una constante..

[0114] Es de señalarse que en la ecuación anterior, $|A|_{ik}$ designa un elemento de matriz con los índices i y k del espectrograma de magnitud o espectrograma de amplitud $|A|$. Además, $(|X| - WH)_{ik}$ designa un elemento de la matriz de una diferencia entre el espectrograma de magnitud o el espectrograma de amplitud $|X|$ de la distribución tiempo-frecuencia y la aproximación asociada $WH = X^{\wedge}$, que tiene los índices i y k .

[0115] Además, $(WH)_{ik}$ denota un elemento de matriz de la aproximación $WH = X^{\wedge}$ con los índices i y k . $|X|_{ik}$ es un elemento de matriz del espectrograma de cantidad o magnitud $|X|$ con los índices i y k . Por lo tanto, se puede ver a partir de las ecuaciones (11) y (12) que el factor β_{ik} y/o la rectificación de las entradas de la diferencia $(|X| - WH)$ se determina elemento por elemento en la realización preferida.

[0116] En lo siguiente, se describe un procedimiento alternativo para la determinación del espectrograma de cantidad o magnitud $|A|$ de la señal de ambiente. Una alternativa simple se obtiene determinando en primer lugar el espectrograma de cantidad o magnitud $|A|$ de la señal ambiental de acuerdo

$$|A| = |X| - \zeta \cdot |WB|,$$

5 con $0 \leq \zeta \leq 1$ y efectuando, después de esto, una rectificación de onda completa de elementos negativos en la así determinada matriz $|A|$. Aquí, el parámetro ζ facilita el ajuste y/o control de la cantidad de ambiente en comparación con la señal directa contenida en la señal de ambiente.

[0117] Es de señalar que el último procedimiento descrito, en contraste con el procedimiento descrito con respecto a las ecuaciones (11) y (12) implica el efecto, en el cálculo de la matriz $|A|$, de que aparece una mayor cantidad de sonido o ruido directo en la señal de ambiente. Por lo tanto, típicamente, se prefiere el procedimiento descrito en el contexto de las ecuaciones (11) y (12).

[0118] Además, hay una tercera alternativa de procedimiento para determinar la matriz $|A|$, tal como se describe a continuación. La tercera alternativa de procedimiento consiste en añadir una restricción de límite o condición límite para la función de coste con el fin de influir en la cantidad o el valor de los elementos de valor negativo en el término

$$15 \quad |A| = |X| - |WB|$$

[0119] Dicho de otro modo, la elección apropiada de la restricción de límite o condición límite respecto a la función de coste puede servir para lograr que haya tan pocos valores negativos como sea posible (como alternativa: tan pocos valores positivos como sea posible) puede, por ejemplo, producirse en la diferencia $|A| = |X| - |WB|$.

[0120] Dicho de otro modo, el procedimiento de optimización para determinar las entradas de las matrices W y H está adaptado de tal manera que la diferencia mencionada comprende preferentemente valores positivos y/o valores comparativamente menos negativos (o viceversa).

[0121] Se puede formular una nueva función de coste

$$C = f(|X|, |WB|)$$

de la manera siguiente:

$$25 \quad C = \sum_{i,k} (|X|_{i,k} - (|WB|)_{i,k})^2 - \epsilon \sum_{i,k} (|X|_{i,k} - (|WB|)_{i,k}) \quad (13)$$

[0122] Aquí, ϵ es una constante para determinar la influencia de la restricción de límite o condición de frontera en el coste total (o sobre el valor total de la función de coste C). La regla de actualización y/o regla de iteración para el descenso de gradiente se obtiene insertando el operador de derivación $\partial C / \partial H$ (Según la ecuación 14) y el operador de derivación $\partial C / \partial W$ en la ecuación (5). Para los operadores de derivación $\partial C / \partial H$ y $\partial C / \partial W$, se cumple lo siguiente:

30

$$\frac{\partial C}{\partial H} = \left[(W^T |X|)_{i,k} - (W^T |WB|)_{i,k} - \epsilon \sum_i W_{i,k} \right] \quad (14)$$

$$\frac{\partial C}{\partial W} = \left[(|X| H^T)_{i,k} - (|WB| H^T)_{i,k} - \epsilon \sum_k H_{i,k} \right] \quad (15)$$

[0123] De lo contrario, hay que señalar que el procedimiento como se ha descrito con respecto a las ecuaciones (11) y (12) se prefiere ya que es fácil de implementar y proporciona buenos resultados.

[0124] En resumen, se muestra que la determinación de la matriz $|A|$ descrita anteriormente, para la que se describen tres diferentes procedimientos, puede ser ejecutada, por ejemplo, mediante los medios de determinación de la diferencia 230 o de formación de diferencia 330 en realizaciones preferidas de la presente invención.

Reconstrucción de la señal de tiempo

[0125] Sigue una descripción de cómo se pueden obtener la representación $\mathbf{A}(\omega, k)$ provista de una información de fase (también designada con 336) a partir de la representación de magnitud $|\mathbf{A}(\omega, k)|$ (también designada con 332) de la señal de ambiente.

5 [0126] El espectrograma complejo $\mathbf{A}(\omega, k)$ de la señal de ambiente se calcula utilizando la fase $\phi = \mathbf{X}$ de la distribución tiempo-frecuencia (TFD) \mathbf{X} de la señal de entrada 308 (también designada con $x(t)$, $x[n]$) se calcula según la ecuación (16):

$$\mathbf{A}(\omega, k) = |\mathbf{A}(\omega, k)| \cdot [\cos(\phi(\omega, k)) + j \cdot \sin(\phi(\omega, k))] \quad (16)$$

10 [0127] Aquí, ϕ es, por ejemplo, una matriz de valores de ángulo. Dicho de otro modo, la información de fase o información de ángulo de la distribución tiempo-frecuencia (TFD) \mathbf{X} se añade para todo el elemento a la representación de cantidad o magnitud $|\mathbf{A}|$. Dicho de otro modo, a una entrada o elemento de matriz $A_{i,j}$ con un índice de fila i y un índice de columna j , se añade la información de fase de una entrada o elemento de matriz $X_{i,j}$ con un índice de fila i y un índice de columna j , por ejemplo por multiplicación con un respectivo número complejo de magnitud 1. El resultado general es una representación $\mathbf{A}(\omega, k)$ de la señal de ambiente provista de una información de fase (designada con 336).

15 [0128] La señal de ambiente $a[n]$ (o una representación de tiempo discreto de la señal de ambiente o bien una representación en tiempo continuo de la señal de ambiente) se obtiene a continuación (opcionalmente) de la representación $\mathbf{A}(\omega, k)$ provista de una información de fase, sometiendo $\mathbf{A}(\omega, k)$ a un proceso inverso de cálculo de la distribución tiempo-frecuencia (TFD). Es decir, una representación $\mathbf{A}(\omega, k)$ provista de una información de fase es, por ejemplo, procesada por una transformada inversa de Fourier de tiempo corto con un esquema de superposición y suma que da como resultado la señal temporal de $x[n]$ cuando se aplica a $\mathbf{X}(\omega, k)$.

20 [0129] El procedimiento descrito se aplica de lo contrario a la superposición de segmentos de una longitud de unos pocos segundos cada uno. Los segmentos son vistos utilizando una ventana de Hann para garantizar una transición suave entre los segmentos adyacentes.

25 [0130] Es de señalar que los últimos procedimientos descritos para la obtención de la representación de la hora a $[n]$ de la señal de ambiente pueden, por ejemplo, ser efectuados en los medios 240 para la re-síntesis o en el Convertidor de distribución tiempo-frecuencia a señal tiempo 340.

Montaje de una señal de audio multicanal

30 [0131] Una señal 5.0 o una señal de audio 5.0 (es decir, por ejemplo, una señal de audio que comprende un canal trasero izquierdo, un canal delantero central, así como un canal delantero derecho, un canal posterior izquierdo y un canal posterior derecho) se obtiene mediante la alimentación a los canales posteriores (es decir, por ejemplo, al menos el canal trasero izquierdo o el canal trasero derecho, o de ambos el canal posterior izquierdo y el canal posterior derecho) con señal de ambiente. Los canales delanteros (es decir, por ejemplo, el canal delantero izquierdo, el canal central y/ o el canal delantero derecho) reproducen la señal original en una realización preferida. Aquí, por ejemplo, los parámetros de ganancia y/ o parámetros de sonoridad aseguran que se obtiene una energía total (o permanece sustancialmente sin cambios) cuando se utiliza el canal central adicional.

35 [0132] Además, es de señalar que el concepto descrito para la generación de una señal de ambiente se puede emplear en cualquier sistema multicanal y sistemas de canales múltiples de reproducción de audio. Por ejemplo, el concepto de la invención se puede emplear en un sistema 7.0 (por ejemplo, en un sistema que tiene tres altavoces frontales, dos altavoces laterales y dos altavoces posteriores). Por lo tanto, la señal de ambiente puede ser, por ejemplo, suministrada a uno o ambos altavoces laterales y/ o a uno o ambos altavoces posteriores.

40 [0133] Después de la separación del ambiente (o después de la generación de la señal de ambiente), el procesamiento adicional puede ser opcionalmente llevado a cabo con el fin de obtener una señal de audio multicanal de alta calidad perceptual. En el montaje de una señal de audio de múltiples canales a partir de un solo canal, se desea que la imagen frontal sea preservada mientras se añade impresión de amplitud. Esto se logra, por ejemplo, mediante la introducción o adición de retardo de unos pocos milisegundos a la señal de ambiente y/ o por la supresión de porciones de transitorios en la señal de ambiente. Además, es ventajosa la descorrelación de la señal de alimentación de los altavoces posteriores o altavoces posteriores entre sí y/ o en relación con las señales que alimentan a los altavoces delanteros.

Supresión de transitorios y/ o supresión de picos u operaciones de ajuste

[0134] Los algoritmos para la detección de transitorios (y/o picos u operaciones de ajuste) y para manipular transitorios se utilizan en varias aplicaciones de procesamiento de señales de audio, como por ejemplo para los efectos de audio digitales (ver [11, 12]) y para upmixing (ver [13]).

[0135] La supresión de los transitorios en el contexto de upmixing tiene como objetivo mantener la imagen frontal. Cuando el ruido transitorio o sonido transitorio aparecen en la señal de ambiente, las fuentes que generan de estos transitorios (por ejemplo por medio de un oyente) no están localizadas en la parte frontal. Este es un efecto no deseado: la "fuente de sonido directo" o aparece más ancha (o más extendida) que en los originales o, peor aún, se percibe como una "fuente de sonido directo" independiente en la parte posterior del oyente.

Descorrelación de las señales de los canales posteriores o posteriores

[0136] En la literatura, el término "descorrelación" describe un proceso que manipula una señal de entrada de tal manera que (2 o más) señales de salida presentan diferentes formas de onda pero suenan igual que la señal de entrada (ver [14])., por ejemplo, dos señales de ruido similares y coherentes de banda ancha se reproducen simultáneamente o se presentan por un par de altavoces, se percibirá un acontecimiento auditivo compacto (ver [15]). La disminución de la correlación de las dos señales de canal aumenta el ancho percibido o la extensión de la fuente de sonido o fuente de ruido hasta que se perciben dos fuentes separadas. Una correlación de dos señales centradas x e y (es decir, señales que tienen un valor medio de cero) se expresan a menudo por medio del coeficiente de correlación R_{xy} , tal como se describe por la ecuación (17):

$$R_{xy} = \lim_{l \rightarrow \infty} \frac{\sum_{k=-l}^l x(k)y^*(k)}{\sqrt{\sum_{k=-l}^l |x(k)|^2} \sqrt{\sum_{k=-l}^l |y(k)|^2}} \quad (17)$$

[0137] Aquí, $y^*(k)$ indica el número complejo conjugado de $y(k)$. Como el coeficiente de correlación no es independiente de pequeños retrasos entre la señales x e y, otra medida para el grado de la similitud entre dos señales centrada en x e y se define por el uso de la correlación Γ entre canales (ver [15]) o por la coherencia entre canales (ver [16]) ecuación ((18). En la ecuación (18), la correlación entre canales coherencia o inter-canal Γ se define como sigue:

$$\Gamma = \max_{\tau} |r_{xy}(\tau)| \quad (18)$$

[0138] Aquí, la correlación cruzada normalizada r_{xy} se define de acuerdo a la ecuación (19):

$$r_{xy}(\tau) = \lim_{l \rightarrow \infty} \frac{\sum_{k=-l}^l x(k)y^*(k+\tau)}{\sqrt{\sum_{k=-l}^l |x(k)|^2} \sqrt{\sum_{k=-l}^l |y(k)|^2}} \quad (19)$$

[0139] Ejemplos de procesos de descorrelación son la reverberación natural y varios procesadores de señal (*flanger*, *chorus*, *phaser*, reverberación sintética).

[0140] Un anterior procedimiento de descorrelación en el campo del procesamiento de señal de audio se describe en [17]. Aquí, dos canales de salida de las señales son generadas por la suma de la señal de entrada y una versión retardada de la señal de entrada, en el que en un canal, se invierte la fase del canal retardada.

[0141] Otros procedimientos generan señales descorrelacionadas por medio de la convolución.

Un par de señales de salida con una medida de correlación determinada o especificada se generan mediante convolución de la señal de entrada con un par de respuestas de impulsos que están correlacionadas entre sí de acuerdo con el valor dado (ver [14]).

[0142] Una descorrelación dinámica (es decir, variable en el tiempo) se obtiene mediante el uso de filtros Allpass variables en el tiempo, es decir filtros Allpass en que las nuevas respuestas de fase aleatorias se calculan para marcos de tiempo adyacentes (ver [18], [11]).

[0143] En [18], se ha descrito un procedimiento de subbanda, en el que la correlación en las bandas de frecuencia individuales se cambian de forma variable.

[0144] En el contexto del procedimiento de la invención de escrito aquí, se aplica una descorrelación a la señal de ambiente. En una configuración de 5,1 (es decir, en una configuración con, por ejemplo, seis altavoces) (pero también

en otra configuración con al menos dos altavoces) se desea que las señales ambientales que finalmente se alimentan a los dos canales posteriores o posteriores estén correlacionadas entre sí por lo menos en cierta medida.

[0145] Las propiedades deseadas del procedimiento de la invención son la difusión de campo de sonido (o difusión de campo de ruido de o de ampliación de campo de) y envolvente.

5 **[0146]** En lo que sigue y con referencia a la figura 5, se describe un aparato para derivar una señal de audio multi-canal que comprende una señal de altavoz frontal y una señal de altavoz posterior de una señal de audio. El aparato para obtener la señal de audio canal multi de acuerdo con la figura 5 se designa en su totalidad con 500. El aparato 500 recibe la señal de audio 508 o una representación 508 de la señal de audio. El aparato 500 comprende un aparato 510 para generar una señal de ambiente, en el que el aparato 510 recibe la señal de audio 508 o la representación 508 de la señal de audio. El aparato 510 proporciona una señal de ambiente 512. Es de señalar que, en una realización preferida, el aparato 510 es el aparato 100 de acuerdo con la figura 1. En una realización preferida adicional, el aparato 510 es el aparato 200 de acuerdo con la figura 2. En una realización preferida adicional, el aparato 510 es el aparato 300 de acuerdo con la figura 3.

15 **[0147]** La señal de ambiente 512, que puede estar presente en forma de una representación de dominio en el tiempo (o representación tiempo señal) y/o en una representación tiempo-frecuencia se alimenta también a los medios de postprocesado 520. Los medios de postprocesado 520 son opcionales y pueden, por ejemplo, comprender un reductor de pulso configurado para reducir o eliminar los transitorios presentes en la señal de ambiente 512. Aquí, los transitorios son partes de la señal de alta energía que pueden exhibir una pendiente de flanco mayor que una determinada inclinación de borde máxima admisible. Por otra parte, los eventos transitorios también podrían de lo contrario ser picos de señal en la señal de ambiente 512, cuyas amplitudes exceden una cierta amplitud dada máxima.

20 **[0148]** Además, los medios de postprocesado 520 podrían (opcionalmente) comprender un retardador o medios de retardo de la señal de ambiente 512. Por lo tanto, los medios de postprocesado 520 proporcionan una señal de postprocesamiento ambiente 522 en la que, por ejemplo, los transitorios se reducen o eliminan en comparación con la señal de ambiente(original) 512 y/o que está por ejemplo retrasada en comparación con la señal de ambiente (original) 512.

25 **[0149]** Si se omiten los medios de postprocesado 520, entonces la señal 522 puede ser idéntica a la señal 512.

30 **[0150]** El aparato 500 comprende además (opcionalmente) un combinador 530. Si el combinador está incluido, el combinador 520, por ejemplo, proporciona una señal de altavoz posterior 532, que está formada por una combinación de la señal de ambiente postprocesada 522 y una versión (opcionalmente postprocesada) de la señal de audio original 508.

35 **[0151]** Si se omite el combinador 530 opcional, entonces la señal 532 puede ser idéntica a la señal 522. El aparato 500 comprende además (opcionalmente) un correlador 540, que recibe la señal del altavoz posterior 532 y a partir de esta suministra basados al menos dos señales descorrelacionadas de altavoces posteriores 542, 544. La primera señal de altavoz posterior 542 puede, por ejemplo, representar una señal de altavoz posterior para el altavoz posterior izquierdo. La segunda señal de altavoz posterior 544 puede, por ejemplo, representar una señal de altavoz posterior para el altavoz posterior derecho.

40 **[0152]** En el caso más simple (por ejemplo, si se omiten los medios de postprocesado 520, el combinador 530 y el correlador 540), por ejemplo la señal de ambiente 512 generada por el aparato 510 se utiliza como la primera señal de altavoz posterior 542 y/o como la segunda señal de altavoz posterior 544. En general, se puede decir que, con respecto a los medios de postprocesado 520, el combinador 530 y/o el correlador 540, se considera la señal de ambiente 512 generada por el aparato 510 para la generación de la primera señal de altavoz posterior 542 y/o para la generación de la segunda señal de altavoz posterior 544.

45 **[0153]** Por lo tanto, la presente invención comprende explícitamente la utilización de la señal de ambiente 512 generada por el aparato 510 como una primera señal de altavoz posterior 542 y/o como una segunda señal de altavoz posterior 544.

[0154] Del mismo modo, la presente invención también comprende explícitamente generar la primera señal de altavoz posterior 542 y/o la segunda señal de altavoz posterior 544 utilizando la señal de ambiente 512 generada por el aparato 510.

50 **[0155]** El aparato puede estar, opcionalmente, configurado además para generar una primera señal de altavoz frontal, una segunda señal de altavoz delantero y/o una tercera señal de altavoz frontal. Para este propósito, por ejemplo, la señal de audio (original) 508 se alimenta a los medios de postprocesado 550. Los medios de postprocesado 550 están configurados para recibir y procesar la señal de audio 508 y generar una señal de audio postprocesada 552, que es, por ejemplo, (opcionalmente) alimentada al combinador 530. Si se omiten los medios de postprocesado, la señal 542 puede ser idéntica a la señal 508. Si no, la señal 552 constituye una señal de altavoz frontal.

55 **[0156]** En una realización preferida, el aparato 500 comprende un divisor de señal 560 configurado para recibir la señal de altavoz frontal 552 y para generar, basándose en esta, una primera señal de altavoz frontal 562, una segunda señal

de altavoz frontal 564 y/ o una tercera señal de altavoz frontal 566. La primera señal de altavoz frontal 562 puede, por ejemplo, ser una señal de altavoz para un altavoz frontal izquierdo. La segunda señal de altavoz frontal 564 puede, por ejemplo, ser una señal de altavoz para un altavoz frontal derecho. La tercera señal de altavoz frontal 566 puede, por ejemplo, ser una señal de altavoz para un altavoz frontal central.

5 **[0157]** La figura 6 muestra otra forma de un diagrama de flujo de un procedimiento de la invención de acuerdo con una realización de la presente invención. El procedimiento de acuerdo con la figura 6 se designa en su totalidad con 600. El procedimiento 600 comprende una primera etapa 610. La primera etapa 610 comprende la compresión con pérdidas de la señal de audio (o de una representación de la señal de audio) a fin de obtener una representación de la señal de audio en la forma de compresión con pérdida. Una segunda etapa 620 del procedimiento 600 comprende calcular una diferencia entre la representación comprimida de la señal de audio y la representación de la señal de audio a fin de obtener una representación de discriminación.

[0158] Una tercera etapa 630 comprende proporcionar una señal de ambiente mediante la representación de discriminación. Por lo tanto, como un todo, el procedimiento 600 permite la generación de una señal de ambiente a partir de una señal de audio.

15 **[0159]** Es de señalar aquí que el procedimiento de la invención 600 según la figura 6 puede ser completado con las etapas que son ejecutados por los aparatos de la invención descritos más arriba. Por lo tanto, el procedimiento puede, por ejemplo, ser modificado y/ o complementado de forma que cumpla la función del aparato 100 de acuerdo con la figura 2, la función del aparato 200 de acuerdo con la figura 2, la función del aparato 300 de acuerdo con la figura 3 y/ o la función del aparato 500 de acuerdo con la figura 5.

20 **[0160]** En otras palabras, el aparato de la invención y el procedimiento de la invención pueden implementarse en hardware o en software. La aplicación se puede efectuar en un medio de almacenamiento digital tal como un disco flexible, un CD, un DVD o una memoria flash con señales de control legibles electrónicamente que cooperan de tal modo con un sistema informático programable que se ejecuta el procedimiento respectivo. Por lo tanto, en general la presente invención también está constituida por un producto de programa informático con un código de programa para realizar el procedimiento de la invención almacenado en un soporte legible por ordenador, cuando el producto de programa de ordenador se ejecuta en un ordenador. Por lo tanto, en otras palabras, la invención puede ser realizada como un programa informático con un código de programa para realizar el procedimiento, cuando el programa informático se ejecuta en un ordenador.

Descripción general del procedimiento

30 **[0161]** En resumen, se puede decir que se genera una señal de ambiente a partir de la señal de entrada y se alimenta a los canales posteriores. Aquí, puede utilizarse un concepto tal como se describe bajo el título "Concepto Directo / Ambiente". La quinta esencia de la invención se refiere al cálculo de la señal de ambiente, en el que la figura 2 muestra un diagrama de bloques de un tratamiento, que puede ser utilizado para la obtención de la señal de ambiente.

[0162] En resumen, se muestra lo siguiente:

35 Una distribución tiempo-frecuencia (TFD) de la señal de entrada se calcula tal como se explica en el epígrafe "La distribución tiempo-frecuencia de la señal de entrada". Una aproximación de la distribución tiempo-frecuencia (TFD) de la señal de entrada se calcula utilizando el procedimiento de optimización numérica, tal como se describe en la sección "aproximación de la distribución tiempo-frecuencia". Mediante el cálculo de una distinción o diferencia entre la distribución tiempo-frecuencia (TFD) de la señal de entrada y su aproximación, se obtiene una estimación de la distribución tiempo-frecuencia (TFD) de la señal de ambiente. La estimación se designa también con $|A|$ y/o A . La estimación se designa también con $[\lambda]$ y/ o A . Una re-síntesis de la señal temporal de la señal ambiental se explica también en la sección con título "Reconstrucción de la señal temporal". Además, el postprocesado puede ser utilizado (opcionalmente) para mejorar la impresión auditiva de la señal multicanal obtenida, tal como se describe bajo el título "Montaje de una señal de audio multicanal".

Conclusión

45 **[0163]** En resumen, se puede decir que la presente invención describe un procedimiento y un concepto para separar una señal de ambiente a partir de señales de un solo canal de audio (o de una señal de audio de un canal). La señal de ambiente obtenida presenta una alta calidad de audio. Se compone de elementos de sonido o elementos de ruido procedentes de ambiente, es decir, reverberación, el ruido público, así como el ruido ambiente o del entorno. La cantidad o el volumen de sonido directo o ruido directo en la señal de ambiente es muy baja o incluso evanescente.

[0164] Las razones para el éxito de l procedimiento descrito puede describirse tal como sigue de una manera simplificada:

55 Las distribuciones de frecuencias de tiempo-(TFD) de sonido directo o ruido directo son generalmente más escasas o menos densas que las distribuciones de tiempo-frecuencia (TFD) de ruido ambiente o sonido ambiente. Es decir, la energía del ruido directo o sonido directo está más concentrado en menos celdas o entradas de la matriz que la energía de ruido ambiente o sonido ambiente. Por lo tanto, la aproximación directa

detecta ruido o sonido directo, pero no (o sólo en muy poca medida) el ruido ambiente o sonido ambiente. Como alternativa, se puede decir que la aproximación directa detecta ruido o sonido directo en mayor grado que el ruido ambiente o sonido ambiente. La distinción o diferencia entre la distribución de tiempo-frecuencia (TFD) de la señal de entrada y su aproximación es por lo tanto una buena representación de la distribución tiempo-frecuencia (TFD) de todo el ruido ambiental y/ o sonido ambiental presente en la señal de entrada.

[0165] Sin embargo, la presente invención comprende un procedimiento de cálculo de señales multicanal (o una señal multicanal) a partir de una señal de un canal o una señal de dos canales (o de las señales de un canal o señales de dos canales). Por lo tanto, el uso del procedimiento descrito y concepto permite el suministro de grabaciones convencionales en un sistema de múltiples canales (o sistemas de canales múltiples) de una manera en la que se mantienen las ventajas de la representación multi-señal.

[0166] Además, es deseñalar que en el procedimiento de la invención, en una realización preferida, no se utilizan efectos de audio artificiales y que la manipulación de los sonidos y/ o señales de audio sólo implican solamente envolvemento y amplitud. No hay ninguna coloración de tono del sonido original o del ruido original. La impresión auditiva prevista por el autor de la señal de audio se mantiene.

[0167] Por lo tanto, se puede decir que el procedimiento y concepto de la invención descritos superan inconvenientes sustanciales de los procedimientos o conceptos conocidos. Es de señalar que los procedimientos de adaptación de señal descritos en la introducción calculan la señal de canal posterior (es decir, la señal para los altavoces posteriores) mediante el cálculo de las diferencias inter-canal de la señal de entrada de dos canales. Por lo tanto estos procedimientos no son capaces de generar una señal multicanal a partir de una señal de entrada de acuerdo con la opción 3 cuando ambos canales de la señal de entrada son idénticos (es decir, cuando la señal de entrada es una señal dual-mono) o cuando las señales de los dos canales son casi idénticas.

[0168] El procedimiento descrito en el epígrafe "Pseudostereofonía basada en un entorno espacial" siempre requerirá una versión multi-canal con el mismo contenido o un operador que genere las claves espaciales manualmente. Por lo tanto, el procedimiento conocido mencionado no puede ser empleado ni de una manera en tiempo real o automáticamente cuando no hay una versión multi-canal de la misma señal de entrada.

[0169] En contraste, el procedimiento y el concepto de la invención descritos en la presente memoria son capaces de generar una señal de ambiente de una señal de un canal sin ninguna información previa sobre la señal. Además, no se utilizan objetos de audio sintético o efectos de audio (tales como reverberación).

[0170] A continuación, se describe un ejemplo particularmente ventajoso de parámetros para la aplicación del concepto de la invención de acuerdo con una realización de la presente invención.

[0171] En otras palabras, a continuación, se describen los parámetros óptimos para el procedimiento de separación de ambiente para aplicaciones mono-upmix. Además, se darán valores mínimos y máximos de los parámetros, los cuales, si bien pueden funcionar, no producen resultados óptimos con respecto a la calidad de audio y/ o la carga de procesamiento requerida. Además, se darán valores mínimos y máximos de los parámetros, los cuales, si bien pueden funcionar, no producen resultados óptimos con respecto a la calidad de audio y/ o la carga de procesamiento requerida.

[0172] Aquí, el tamaño de parámetro FFT (nfft) describe cómo se procesan muchas bandas de frecuencia. Dicho de otro modo, el tamaño de parámetro FFT indica, cuantas frecuencias discriminables ω_1 a ω_n existen. Por lo tanto, el tamaño de parámetro FFT es también una medida de cuán grande es una primera dimensión (por ejemplo, un número de filas de la matriz) de la matriz $\mathbf{x}(\omega, k)$ is. En otras palabras, en una realización preferida, el tamaño del parámetro IFFT describe el número de filas (o columnas) de la matriz $\mathbf{x}(\omega, k)$. Por lo tanto, el tamaño de parámetro FFT, por ejemplo, se corresponde con el valor n. Además, el valor de tamaño de FFT también describe cuántas muestras se utilizan para el cálculo de una sola entrada $X_{i,j}$ de la matriz \mathbf{x} . Dicho de otro modo, se utilizan las muestras nfft de una representación temporal de la señal de entrada para calcular a partir de estas nfft coeficientes espectrales para nfft diferentes frecuencias ω_1 a ω_{nfft} . Por lo tanto, a partir de las muestras nfft, se calcula una columna de la matriz $\mathbf{x}(\omega, k)$.

[0173] La ventana que define las muestras contempladas de la señal de entrada es entonces desplazada en un número de muestras definido por el parámetro de salto. Las muestras nfft de la señal de entrada definidas por la ventana desplazada se asignan a nfft coeficientes espectrales por una transformada de Fourier, definiendo el coeficiente espectral una columna de la matriz \mathbf{x} siguiente.

[0174] A modo de ejemplo se puede decir que la primera columna de la matriz \mathbf{X} puede estar formada por una transformada de Fourier de las muestras de la señal de entrada con los índices 1 a nfft. La segunda columna de la matriz \mathbf{X} puede estar formada por una transformada de Fourier de las muestras de la señal de entrada con los índices 1 + salto a nfft + salto.

[0175] La longitud del segmento de parámetro indica cuán largo es un segmento de una trama de señal, cuyo espectrograma se factoriza. Dicho de otro modo, la longitud del segmento de parámetro describe cuán larga es una duración de tiempo de la señal de audio de entrada considerada para el cálculo de las entradas de la matriz \mathbf{X} . Por lo tanto, se puede decir que la matriz \mathbf{X} describe la señal temporal de entrada durante un período igual a la longitud del segmento de parámetro (SegLen).

[0176] El rango de factorización de parámetro describe el rango de factorización de la matriz de factorización no negativa, es decir, el parámetro r . Dicho de otro modo, el rango factorización de parámetro indica la extensión de una dimensión de la primera matriz de aproximación \mathbf{W} y una dimensión de la segunda matriz de aproximación \mathbf{H} are.

[0177] Los valores preferidos para los parámetros se indican en la siguiente tabla:

5

Parámetro	Descripción	Unidad	Mín.	Máx.	Valor óptimo
Tamaño FFT (nfft)	Tamaño de la trama de señal para FFT	Muestras	1024	4096	2048 o 4096
Tamaño de salto (saltos)	Tamaño de salto para FFT	Muestras	1	nfft	$0,125*nfft$ o $0,20,25*nfft$
Longitud de segmento	Tamaño de trama de señal cuyo espectrograma está siendo factorizado	Segundos	1	Longitud de la señal de entrada	2-4
Rango de factorización	Rango de factorización para NMF		10	Número de columnas del espectrograma	40...100

[0178] Como parámetro adicional, se determina además que el error de medida c se utiliza para el cálculo de la NMF. Se prefiere el uso de la divergencia de Kullback-Leibler cuando se procesan espectrogramas de cantidad o magnitud. Se pueden utilizar otras medidas de distancia cuando se procesan los valores de espectrograma con el logaritmo (SPL) o valores de espectrograma de energía.

10

[0179] Además, es de señalar que los intervalos preferidos de valor se han descrito anteriormente. Es de señalar que, utilizando el procedimiento de la invención, el tamaño de la FFT puede estar en un rango de 128 a 65.536. El tamaño de salto puede ser de entre $1/6$ a 4 del tamaño de la FFT y una unidad del tamaño de la FFT. La longitud del segmento asciende normalmente a por lo menos 0,1 segundos.

15

[0180] Para resumir brevemente, se puede decir que la presente invención comprende un nuevo concepto o procedimiento de cálculo de una señal de ambiente a partir de una señal de audio. La señal de ambiente obtenida es de particular beneficio para el u pmixing de señales de audio de música para su reproducción en sistemas de canales múltiples. Un aspecto de las ventajas del concepto o procedimiento inventivo descrito, en comparación con otros procedimientos, es su capacidad para procesar señales de un canal sin usar efectos de audio sintéticos.

20

[0181] Además, es de señalar que la presente invención también puede utilizarse en un sistema simple. Puede contemplarse un sistema, en el que sólo uno de los altavoces delantero y un altavoz posterior están presentes y/o activos. En este caso, por ejemplo, la señal de audio original puede ser reproducida en el altavoz delantero. La señal ambiental derivada de la señal de audio original se puede reproducir en el altavoz posterior. Dicho de otro modo, la señal mono de audio original se puede reproducir como una señal mono en un altavoz delantero único, mientras que la señal de ambiente derivada de la señal de audio original se reproduce como un canal de posterior único.

25

[0182] Sin embargo, si hay varios canales, pueden ser procesados individualmente en una realización de la presente invención. Dicho de otro modo, se considera un primer canal de la señal de audio original para la generación de una primera señal de ambiente, y se utiliza un segundo canal de la señal de audio original para generar una segunda señal de ambiente. El primer canal de la señal de audio original se reproduce, por ejemplo, en un altavoz frontal primero (por ejemplo, izquierdo delantero), y el segundo canal de la señal de audio original se reproduce, por ejemplo, en un segundo altavoz frontal (por ejemplo, frontal derecho). Además, por ejemplo, la primera señal de ambiente se reproduce en un primer altavoz posterior (trasero izquierdo, por ejemplo), mientras que la segunda señal de ambiente se reproduce, por ejemplo, en un segundo altavoz posterior (por ejemplo, derecho trasero).

30

[0183] Por lo tanto, la presente invención también comprende la generación de dos señales de altavoces posteriores a partir de dos señales de altavoz frontal en la forma descrita.

35

[0184] En una realización adicional, la señal de audio original se compone de tres canales, por ejemplo, un canal delantero izquierdo, un canal central frontal y un canal delantero derecho. Por lo tanto, se obtiene una primera señal de ambiente a partir del primer canal (por ejemplo canal delantero izquierdo) de la señal de audio original. A partir del segundo canal (por ejemplo, canal central frontal) de la señal de audio original, se obtiene una segunda señal de ambiente. A partir del tercer canal (por ejemplo canal delantero derecho) de la señal de audio original, se obtiene una señal de ambiente tercero (opcionalmente).

40

[0185] Dos de las señales ambientales (por ejemplo la primera señal de ambiente y la segunda señal de ambiente) se combinan (por ejemplo, mezcladas o combinadas por una suma ponderada o no ponderada) de modo que se obtenga

una primera señal de altavoz de ambiente, que se alimenta a un primer altavoz de ambiente (por ejemplo, un altavoz trasero izquierdo).

5 **[0186]** Opcionalmente, además, dos señales adicionales de ambiente (por ejemplo la segunda señal de ambiente y la tercera señal de ambiente) se combinan para obtener una segunda señal de altavoz de ambiente que se alimenta a un segundo altavoz de ambiente (por ejemplo, un altavoz trasero derecho).

10 **[0187]** Por lo tanto, una primera señal de altavoz de ambiente está formada por una primera combinación de señales ambientales, cada una formada a partir de un canal de la señal de audio multicanal original, mientras que una segunda señal de altavoz de ambiente se forma mediante una segunda combinación de las señales ambientales. La primera combinación comprende preferentemente al menos dos señales ambientales, y la segunda combinación comprende preferentemente al menos dos señales ambientales. Además, se prefiere que la primera combinación sea diferente de la segunda combinación, en la que, sin embargo, se prefiere que las combinaciones primera y segunda utilicen una señal de ambiente común.

15 **[0188]** Además, es de señalar que una señal de ambiente generada en la forma de la invención puede, por ejemplo, también alimentarse a un altavoz lateral si, por ejemplo, una disposición de altavoces que se utiliza comprende altavoces laterales. Por lo tanto, una señal de ambiente puede ser alimentada a un altavoz lateral izquierdo en un uso de una disposición de altavoces 7.1. Además, una señal de ambiente también se puede alimentar al altavoz lateral derecho, en el que la señal de ambiente suministrada al altavoz lateral izquierdo difiere preferentemente de la señal de ambiente suministrada al altavoz derecho.

20 **[0189]** Por lo tanto, la presente invención en su conjunto proporciona una sobre extracción particularmente buena de una señal de ambiente de una señal de un canal.

Referencias:

[0190]

[1] C.Faller, Pseudostereophony Revisited, Presented at the 118th Convention of the AES, Barcelona, Spain, 2005.

25 [2] R. Dressler, "Dolby Surround Pro Logic2Decoder - Principles of operation", Tech. Rep., Dolby Laboratories, http://www.dolby.com/assets/pdf/tech_library/208_Dolby_Surround_Pro_Logic_Decoder.pdf, 2000.

[3] DTS Technical Staff, "An overview of DTS Neo:6 multi-channel", DTS, <http://www.dts.com/media/uploads/pdfs/DTS%20Neo6%20overview.pdf>.

[4] Harman Kardon Technical Staff, "Logic 7 explained", Harman Kardon, <http://manuals.harman.-com/HK/Technology%20Articles/logic7-TechSheet.pdf>.

30 [5] C.Avendano and J. M. Jot, "Ambience Extraction and Synthesis from Stereo Signals for Multi-channel Audio Upmix", Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Orlando, FL, May 2002.

[6] R. Irwan and R. M. Aarts, "A method to convert stereo to multi-channel sound". Presented at the AES 19th International Conference, Schloss Elmau, -Germany, June 21-24, pages 139-143, 2001.

35 [7] G. Soulodre, "Ambience-Based Upmixing", Workshop Spatial Coding of Surround Sound: A Progress Report, Presented at the 117th AES convention, San Francisco, CA, USA, 2004.

[8] D. Lee and H. Seung, "Learning the parts of objects by Non-negative Matrix Factorization", In Letters To Nature 1999.

[9] A. Cichocki, R. Zdunek, and S. Amari, "New algorithms for Non-negative Matrix Factorization in applications to Blind Source Separation". In Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2006.

40 [10] D. Lee and H. Seung, "Algorithms for Non-negative Matrix Factorization". In Proceedings of NIPS, 2001.

[11] C. Duxbury, M. Davies, M. Sandler, "Separation of transient information in musical audio using multiresolution analysis techniques", in Proceedings of DAFX-01, 2001.

[12] M. Goodwin and C.Avendano, "Enhancement of Audio Signals Using Transient Detection and Modification", Presented at the 117th AES convention, San Francisco, CA, USA, 2004.

45 [13] A. Walther, C. Janssen, J. Herre, and O. Hellmuth "Transient suppression in ambience-like signals." Patent FPL-Fallnummer: 06F47292-IIS.

[14] G. Kendall, "The decorrelation of audio signals and its impact on spatial imagery", Computer Music Journal, 19:4, 1995.

[15] C. Faller and F. Baumgarte, "Binaural Cue Coding - Part 2: Schemes and Applications", IEEE-TSAP, 2003.

[16] C. Faller, "Parametric coding of spatial audio", Phd thesis, 2004.

[17] M. Schroeder, "An artificial stereophonic effect obtained from using a single signal,", JAES 6:74, 1957.

[18] G. Potard and I. Burnett, "Decorrelation techniques for the rendering of apparent sound source width in 3D EP 1 997 102 B1 23 5 10 15 20 25 30 35 40 45 50 55 audio displays", in Proceedings of DAFX-04, 2004.

REIVINDICACIONES

1. Aparato (100; 200; 300; 510) para generar una señal de ambiente (132; 252, 254; 342; 512) a partir de una señal de audio (108; 208; 308; 508), que comprende:
- 5 medios (110; 220; 320) para una compresión con pérdidas de una representación espectral (108; 212; 316) de la señal de audio para obtener una representación comprimida (112; 222; 322) de la señal de audio;
- medios (122; 230; 330) para calcular una diferencia entre la representación comprimida (112; 222; 322) de la señal de audio y la representación espectral (108; 212; 316) de la señal de audio para obtener una representación de discriminación (122; 232; 332); y
- 10 medios (130, 240, 340) para proporcionar la señal de ambiente (132; 252, 254; 342) empleando la representación de discriminación;
- en el que los medios (110; 220; 320) para la compresión con pérdidas están configurados para comprimir una representación espectral (108; 212; 316), que describe un espectrograma de la señal de audio (108; 208; 308; 508) para obtener como representación comprimida (112; 222; 322) una representación espectral comprimida de la señal de audio.
- 15 2. Aparato (100; 200; 300; 510) según la reivindicación 1, en el que los medios (110; 220; 320) para la compresión con pérdidas están configurados para emplear, como representación espectral (108; 212; 316; $|X|$) de la señal de audio (108; 208; 308; $x(t)$; $x[n]$), una matriz de distribución tiempo-frecuencia ($|X|$) que describe un espectrograma de la señal de audio, y para aproximar la matriz de distribución tiempo-frecuencia ($|X|$) por un producto (WH) de una primera matriz de aproximación (W) y una segunda matriz de aproximación (H).
- 20 3. Aparato (100; 200; 300; 510) según la reivindicación 2, en el que los medios (110; 220; 320) para la compresión con pérdidas están configurados para emplear, como representación espectral (108; 212; 316; $|X|$) de la señal de audio (108; 208; 308; $x(t)$; $x[n]$), una matriz de distribución tiempo-frecuencia de valores reales ($|X|$) que describe un espectrograma de la señal de audio.
- 25 4. Aparato (100; 200; 300; 510) según la reivindicación 3, en el que los medios (110; 220; 320) para la compresión con pérdidas están configurados para emplear, como representación espectral (108; 212; 316; $|X|$) de la señal de audio (108; 208; 308; $x(t)$; $x[n]$), una matriz de distribución tiempo-frecuencia ($|X|$), cuyas entradas (X) describen amplitudes o energías en la pluralidad de dominios de frecuencia (ω) de la señal de audio para una pluralidad de intervalos de tiempo.
- 30 5. Aparato (100; 200; 300; 510) según cualquiera de las reivindicaciones 2 a 4, en el que los medios (110; 220; 320) para la compresión con pérdidas están configurados para emplear, como representación espectral (108; 212; 316) de la señal de audio (108; 208; 308; $x(t)$; $x[n]$), una matriz de distribución tiempo-frecuencia ($|X|$) que comprende entradas exclusivamente no-negativas o exclusivamente no-positivas.
- 35 6. Aparato (100; 200; 300; 510) según cualquiera de las reivindicaciones 2 a 5, en el que los medios (110; 220; 320) para la compresión con pérdidas están configurados para aproximar la matriz de distribución tiempo-frecuencia ($|X|$) por un producto (WH) de la primera matriz de aproximación (W) y la segunda matriz de aproximación (H), de modo que la primera matriz de aproximación (W) y la segunda matriz de aproximación (H) tienen entradas exclusivamente no-negativas o entradas exclusivamente no-positivas, o de modo que la primera matriz de aproximación (W) tiene entradas exclusivamente no-negativas y la segunda matriz de aproximación (H) tiene entradas exclusivamente no-positivas, o de modo que la primera matriz de aproximación (W) tiene entradas exclusivamente no-positivas y la segunda matriz de aproximación (H) tiene entradas exclusivamente no-negativas.
- 40 7. Aparato (100; 200; 300; 510) según cualquiera de las reivindicaciones 2 a 6, en el que los medios (110; 220; 320) para la compresión con pérdidas están configurados para determinar entradas de la primera matriz de aproximación (W) y entradas de la segunda matriz de aproximación (H) mediante la evaluación de una función de coste (c) que comprende una descripción cuantitativa de una diferencia entre la matriz de distribución tiempo-frecuencia ($|X|$) por un lado y el producto (WH) de la primera matriz de aproximación (W) y la segunda matriz de aproximación (H) por otra parte.
- 45 8. Aparato (100; 200; 300; 510) según la reivindicación 7, en el que los medios (110; 220; 320) para la compresión con pérdidas están configurados para determinar las entradas de la primera matriz de aproximación (W) y la segunda matriz de aproximación (H) empleando un procedimiento para determinar un valor extremo de la función de coste (c) o empleando un procedimiento para una aproximación al valor extremo de la función de coste (c).
- 50 9. Aparato (100; 200; 300; 510) según las reivindicaciones 7 o 8, en el que la función de coste (c) se selecciona de modo que la función de coste (c) comprende una parte que depende de un signo de una diferencia entre una entrada ($|X|_{ij}$) de la matriz de distribución tiempo-frecuencia por un lado y una entrada ($(WH)_{ij}$) del producto (WH) de la primera matriz de aproximación (W) y la segunda matriz de aproximación (H) por otra parte.

5 **10.** Aparato (100; 200; 300; 510) según las reivindicaciones 7, 8 o 9, en el que la función de coste (c) o una condición límite de los medios para la compresión con pérdidas se selecciona de modo que en las diferencias entre una entrada ($|X|_{ij}$) de la matriz de distribución tiempo-frecuencia por un lado y una entrada ($(WH)_{ij}$) del producto (WH) de la primera matriz de aproximación (W) y la segunda matriz de aproximación (H) por otra parte, valores de un primer signo se prefiere que ocurran comparados con valores de un signo inverso de este último.

11. Aparato (100; 200; 300; 510) según cualquiera de las reivindicaciones 7 a 10, en el que la función de coste (c) está configurada para determinar una norma de Frobenius de una diferencia por elemento entre la matriz de distribución tiempo-frecuencia (X) por un lado y el producto (WH) de la primera matriz de aproximación (W) y la segunda matriz de aproximación (H) por otra parte.

10 **12.** Aparato (100; 200; 300; 510) según cualquiera de las reivindicaciones 7 a 10, en el que la función de coste (c) está configurada para determinar la divergencia de Kullback-Leibler generalizada de una diferencia por elemento entre la matriz de distribución tiempo-frecuencia (X) por un lado y el producto (WH) de la primera matriz de aproximación (W) y la segunda matriz de aproximación (H) por otra parte.

15 **13.** Aparato (100; 200; 300; 510) según cualquiera de las reivindicaciones 2 a 12, en el que la matriz de distribución tiempo-frecuencia (X) comprende una primera dimensión de matriz n asociada y una segunda dimensión de matriz m asociada; en el que la primera matriz de aproximación (W) comprende una primera dimensión de matriz n asociada y una segunda dimensión de matriz r asociada; en el que la segunda matriz de aproximación (H) comprende una primera dimensión de matriz r asociada y una segunda dimensión de matriz m asociada; y en el que es verdad lo siguiente:

$$(n+m) \cdot r < nm.$$

20 **14.** Aparato (100; 200; 300; 510) según cualquiera de las reivindicaciones 2 a 13, en el que los medios (120; 230; 330) para calcular una diferencia están configurados para derivar una matriz de error de aproximación ($|A|$) de modo que los elementos (A) de la matriz de error de aproximación ($|A|$) son una función de una diferencia entre elementos de la matriz de distribución tiempo-frecuencia (X) por un lado y elementos ($(WH)_{ij}$) del producto (WH) de la primera matriz de aproximación (W) y la segunda matriz de aproximación (H) por otra parte; en el que la matriz de error de aproximación ($|A|$) constituye la representación de discriminación (122; 232; 332).

25 **15.** Aparato (100; 200; 300; 510) según la reivindicación 14, en el que los medios (120; 230; 330) para calcular una diferencia está configurada para determinar, en el cálculo de una entrada determinada ($|A|_{ij}$) de la matriz de error de aproximación ($|A|$), una diferencia entre una entrada ($|X|_{ij}$) de la matriz tiempo-frecuencia (X) asociada a la entrada determinada ($|A|_{ij}$) por un lado y una entrada ($(WH)_{ij}$) del producto (WH) de la primera matriz de aproximación (W) y la segunda matriz de aproximación (H) asociada a la entrada determinada ($|X|_{ij}$) por otra parte, y para calcular la entrada determinada ($|A|_{ij}$) de la matriz de error de aproximación ($|A|$) como una función de la diferencia por ponderación de la diferencia en función de el signo de la diferencia.

30 **16.** Aparato (100; 200; 300; 510) según la reivindicación 14, en el que los medios (120; 230; 330) para calcular están configurados para determinar, en el cálculo de una entrada determinada ($|A|_{ij}$) de la matriz de error de aproximación ($|A|$), una diferencia entre una entrada ($|X|_{ij}$) de la matriz tiempo-frecuencia (X) asociada a la entrada determinada ($|A|_{ij}$) por un lado y una entrada ($(WH)_{ij}$) del producto (WH) de la primera matriz de aproximación (W) y la segunda matriz de aproximación (H), que se pondera por un factor de ponderación (ξ) no igual a uno asociado con la entrada determinada ($|A|_{ij}$) por otra parte, y para determinar la entrada determinada ($|A|_{ij}$) de la matriz de error de aproximación ($|A|$) para que sea una magnitud de la diferencia.

35 **17.** Aparato (100; 200; 300; 510) según cualquiera de las reivindicaciones 2 a 16, en el que los medios (120; 230; 330) para calcular la diferencia entre la representación comprimida ($112; 222; 322; X^{\wedge}$) de la señal de audio (108; 208; 308) y la representación (108; 212; 316; X) de la señal de audio está configurada para describir la diferencia por una medida de cantidad de valores reales ($|A|$); y en el que los medios (130; 240; 334) para proporcionar la señal de ambiente (132; 242; 336; 352, 254; 342) están configurados para asignar un valor de fase (ϕ) derivado de una representación (108; 212; 312) de la señal de audio a la diferencia, descrita por la medida de cantidad de valores reales ($|A|$), entre la representación comprimida de la señal de audio y la representación de la señal de audio, para obtener la señal de ambiente.

40 **18.** Aparato (100; 200; 300; 510) según la reivindicación 17, en el que los medios (130; 240; 334) para proporcionar están configurados para asignar un valor de fase (ϕ) obtenido en la matriz de distribución tiempo-frecuencia (X) a la diferencia descrita por la medida de cantidad de valores reales ($|A|$).

45 **19.** Aparato (500) para obtener una señal de audio multi-canal que comprende una señal de altavoz frontal (562, 564, 566) y una señal de altavoz posterior (542, 544) a partir de una señal de audio (508), que comprende:

un aparato (100; 200; 300; 510) para generar una señal de ambiente (512) a partir de una señal de audio (508) según cualquiera de las reivindicaciones 1 a 18,

en el que el aparato (510) para generar la señal de ambiente (512) está configurado para recibir la señal de audio (508);

un aparato (550, 560) para proporcionar la señal de audio (508) o una señal derivada de esta como señal de altavoz frontal (563, 564, 566); y

5 un aparato que proporciona la señal de altavoz posterior (520, 530, 540) para proporcionar la señal de ambiente (512) proporcionada por el aparato (510) para generar la señal de ambiente (512) o una señal derivada de esta como señal de altavoz posterior (542, 544).

10 **20.** Aparato (500) según la reivindicación 19, en el que el aparato que proporciona la señal de altavoz posterior (520, 530, 540) está configurado para generar la señal de altavoz posterior (542, 544) de modo que la señal de altavoz posterior está retardada comparado con la señal de altavoz frontal (562, 564, 566) en un intervalo entre un milisegundo y 50 milisegundos.

21. Aparato (500) según las reivindicaciones 19 o 20, en el que el aparato que proporciona la señal de altavoz posterior (520, 530, 540) está configurado para atenuar partes de señal de tipo impulso en la señal de altavoz posterior (542, 544) o para eliminar las partes de señal de tipo impulso de la señal de altavoz posterior (542, 544).

15 **22.** Aparato (500) según cualquiera de las reivindicaciones 19 a 21, en el que el aparato que proporciona la señal de altavoz posterior (520, 530, 540) está configurado para proporcionar, a partir de la señal de ambiente proporcionada por el aparato (510) para generar la señal de ambiente (512), una primera señal de altavoz posterior (542) para un primer altavoz posterior y una segunda señal de altavoz posterior (544) para un segundo altavoz posterior.

20 **23.** Aparato (500) según la reivindicación 22, en el que el aparato que proporciona la señal de altavoz posterior (520, 530, 540) está configurado para proporcionar la primera señal de altavoz posterior (542) y la segunda señal de altavoz posterior (544) a partir de la señal de ambiente (512) de modo que la primera señal de altavoz posterior y la segunda señal de altavoz posterior están al menos parcialmente descorrelacionadas entre sí.

24. Procedimiento (600) para generar una señal de ambiente (132; 242; 252; 254; 336; 342) a partir de una señal de audio (108; 208; 308), que comprende:

25 la compresión con pérdidas (610) de una representación espectral (108; 212; 316) de la señal de audio, que describe un espectrograma de la señal de audio, para obtener una representación espectral comprimida (112; 222; 322) de la señal de audio;

30 calcular (620) una diferencia (122; 232; 332) entre la representación espectral comprimida de la señal de audio y la representación espectral de la señal de audio para obtener una representación de discriminación (122; 232; 332); y

proporcionar (630) la señal de ambiente empleando la representación de discriminación.

25. Procedimiento para obtener una señal de audio multi-canal que comprende una señal de altavoz frontal y una señal de altavoz posterior a partir de una señal de audio, que comprende:

generar una señal de ambiente de la señal de audio según la reivindicación 24;

35 proporcionar la señal de audio o una señal derivada de esta como señal de altavoz frontal; y

proporcionar la señal de ambiente o una señal derivada de esta como señal de altavoz posterior.

26. Programa de ordenador para realizar el procedimiento según las reivindicaciones 24 o 25 cuando el programa de ordenador se ejecuta en un ordenador.

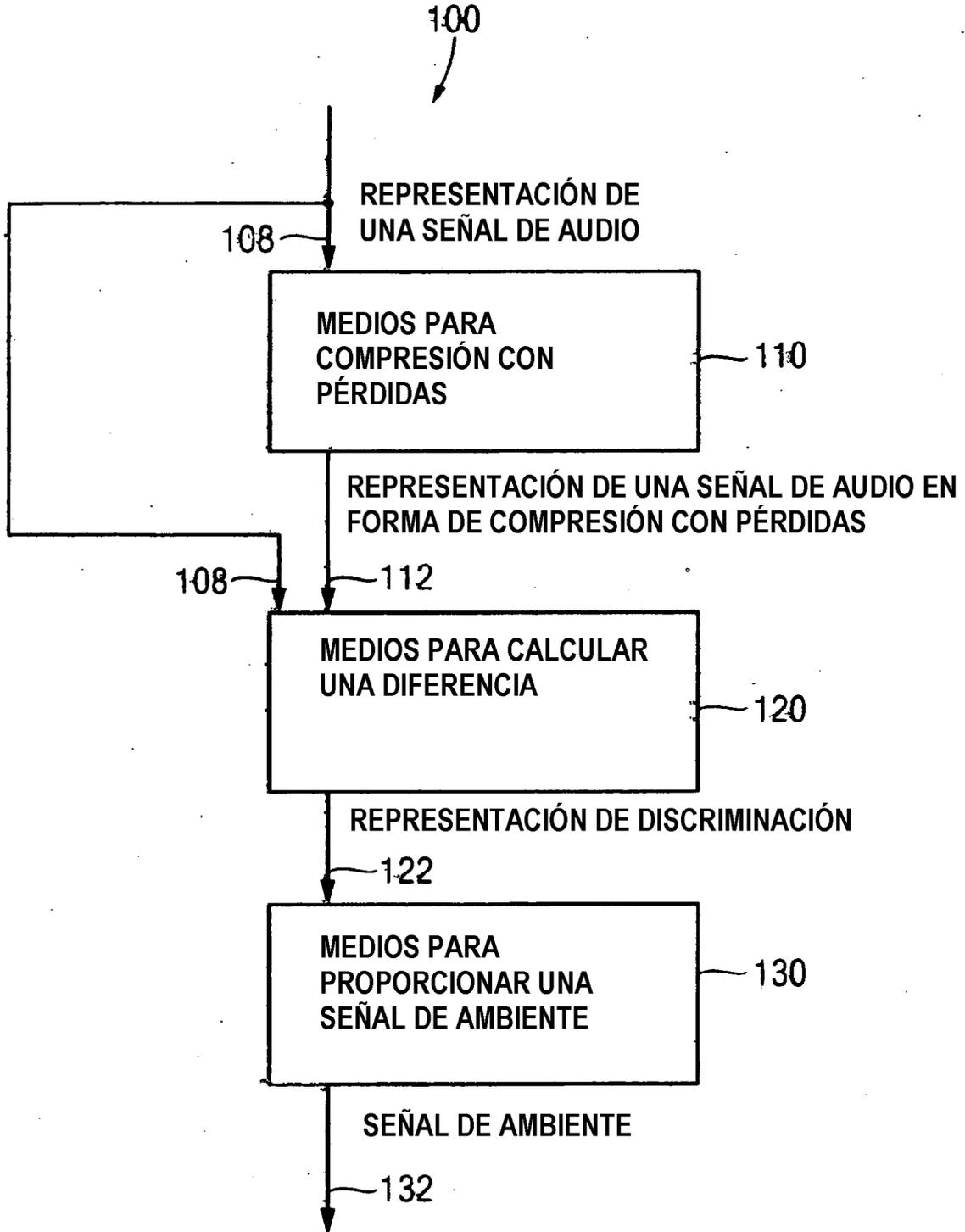


FIGURA 1

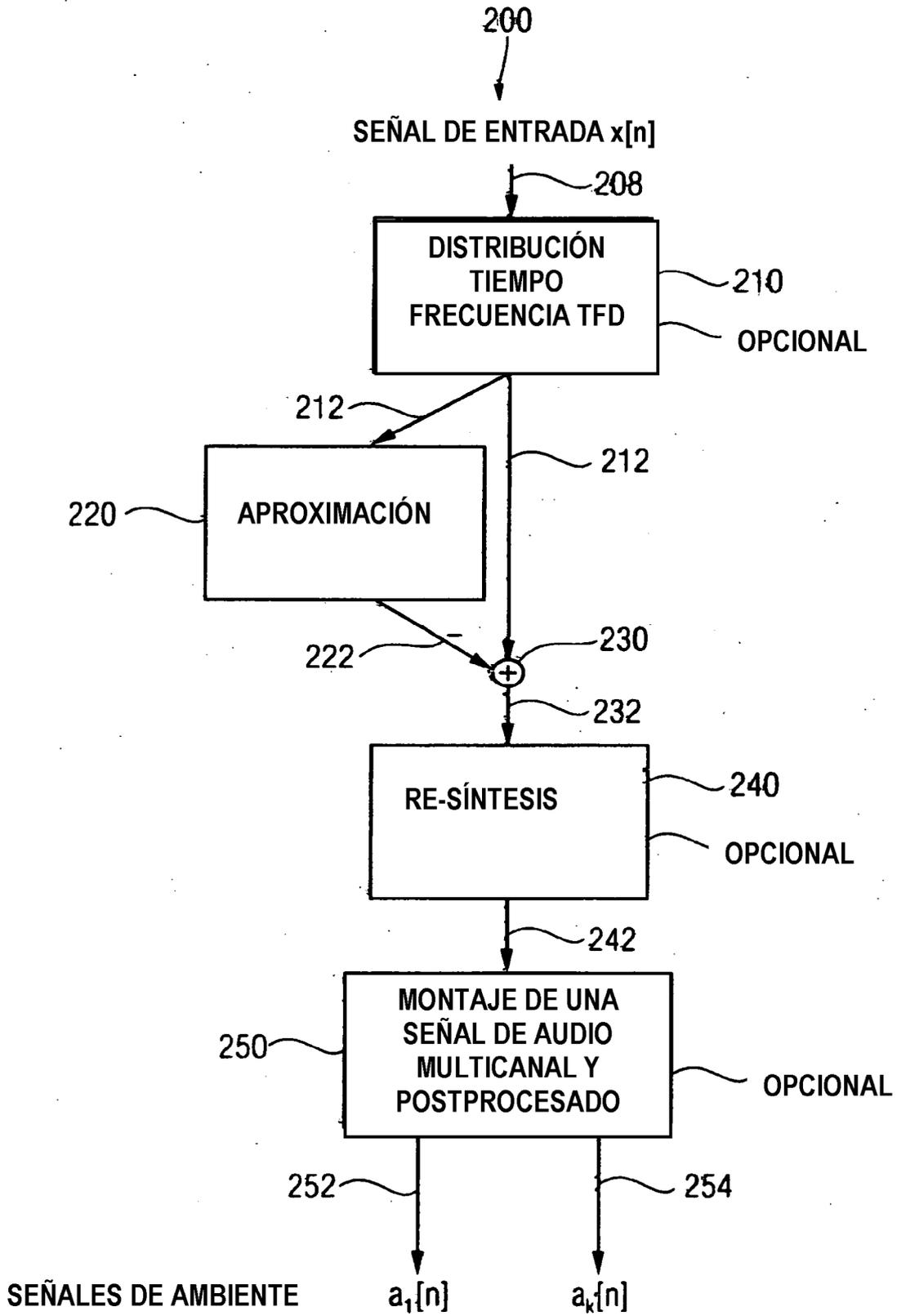


FIGURA 2

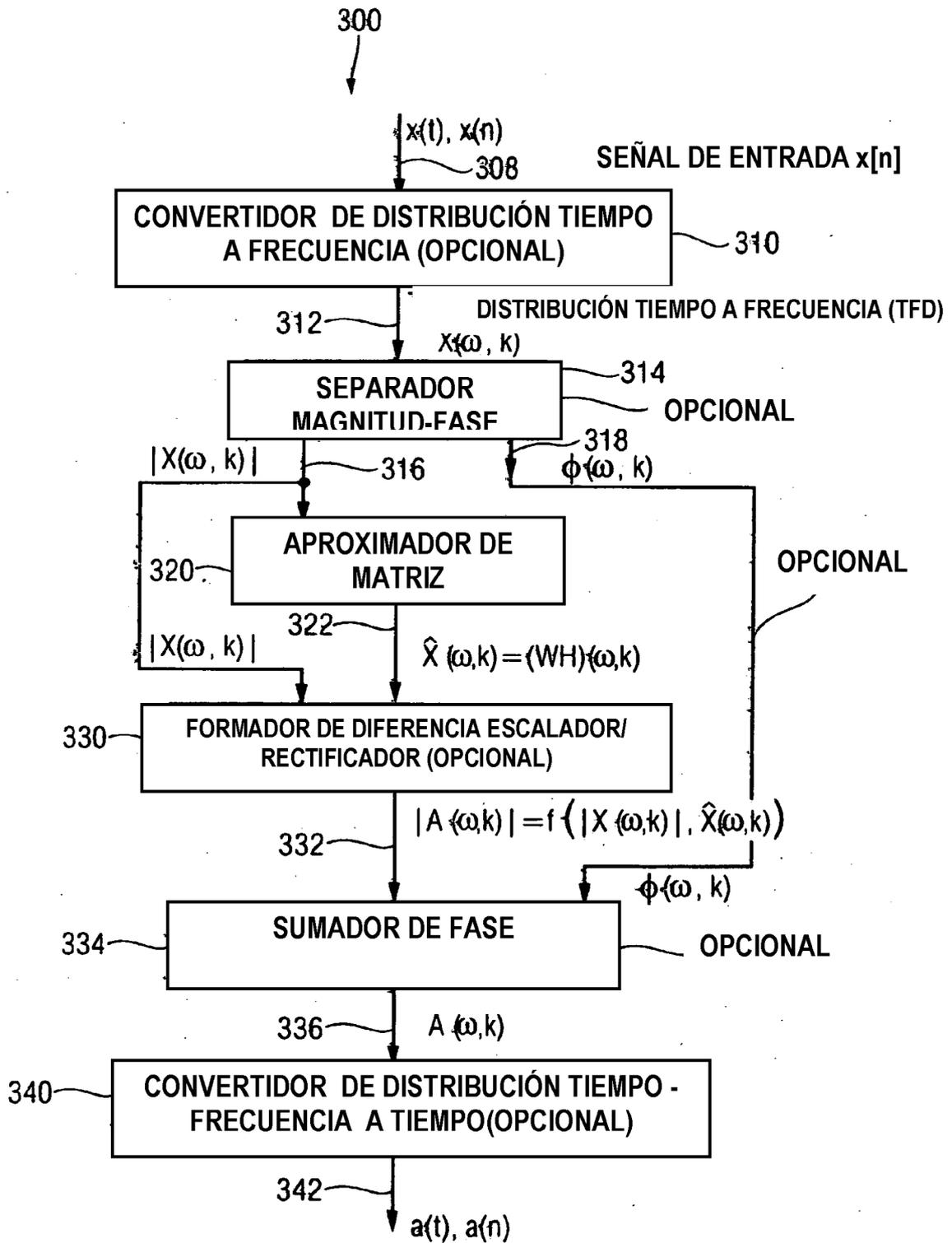


FIGURA 3

$$\begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \approx \begin{pmatrix} 3 & 0 \\ 0 & 2 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 3 \\ 0 \\ 0 \end{pmatrix} (100) + \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix} (010)$$

FIGURA 4A

	k_1	k_2	k_3	k_4	...	k_i	k_{i+1}	k_{i+2}
w_1	$x(w_1, k_1)$	$x(w_1, k_2)$	$x(w_1, k_3)$	x	...	$x(w_1, k_i)$...	x
w_2	x	x	x	x	...	x	...	x
w_3	x	x	x	x	...	x	...	x
\vdots	\vdots	\vdots	\vdots	\vdots		\vdots	\vdots	\vdots
w_n	$x(w_n, k_1)$	x	x	x		$x(w_n, k_i)$	x	x

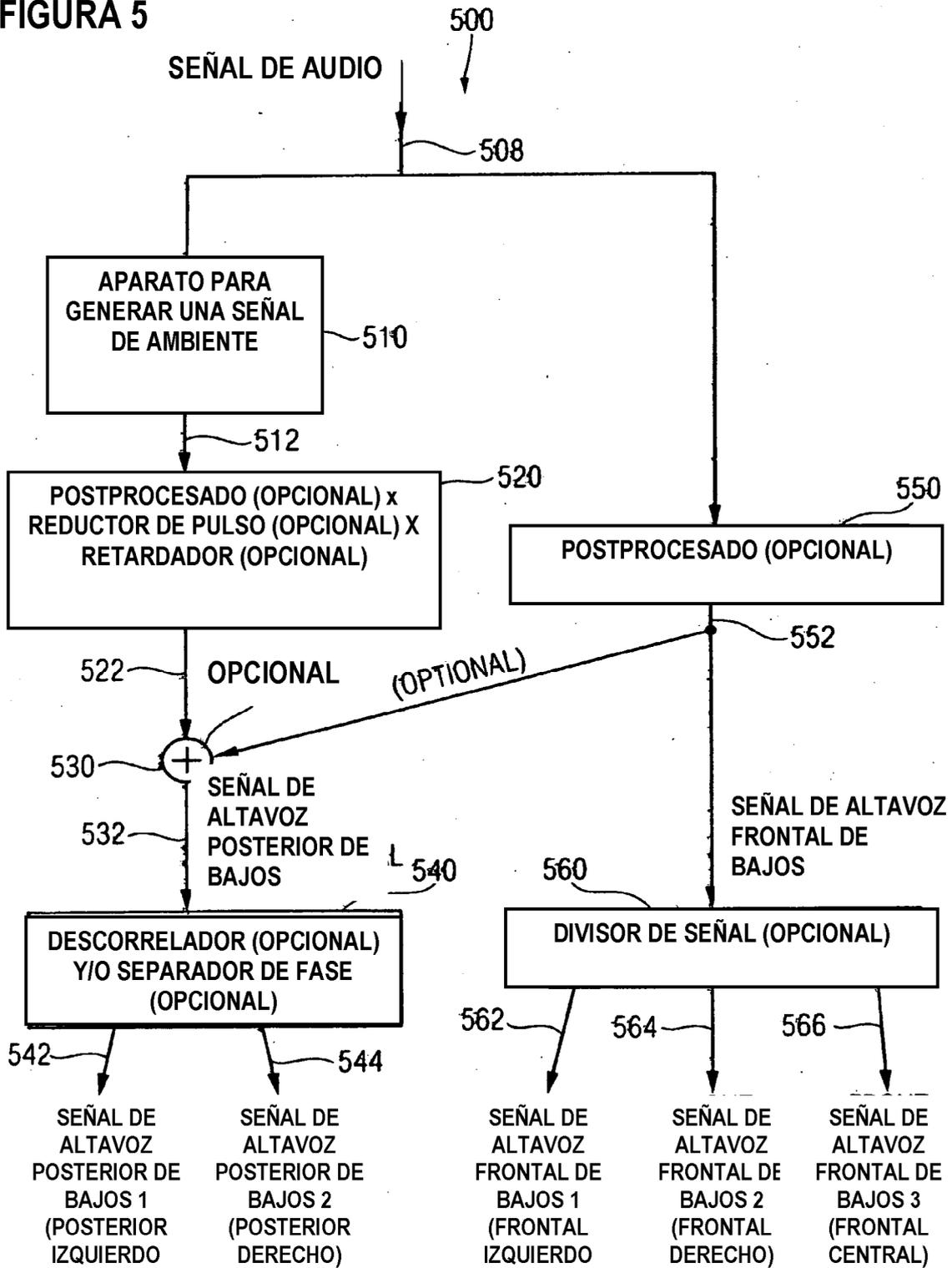
$x(w, k_1)$

	k_1	k_2	k_3	k_4	...	k_i	k_{i+1}	k_{i+2}
w_1	x	$x(w_1, k_2)$	x	x	...	x	$x(w_1, k_{i+1})$	x
w_2	x	x	x	x	...	x		x
w_3	x	x	x	x	...	x		x
\vdots	\vdots	\vdots	\vdots	\vdots		\vdots	\vdots	\vdots
w_n	x	$x(w_n, k_2)$	x	x	...	x	$x(w_n, k_{i+1})$	x

$x(w, k_2)$

FIGURA 4B

FIGURA 5



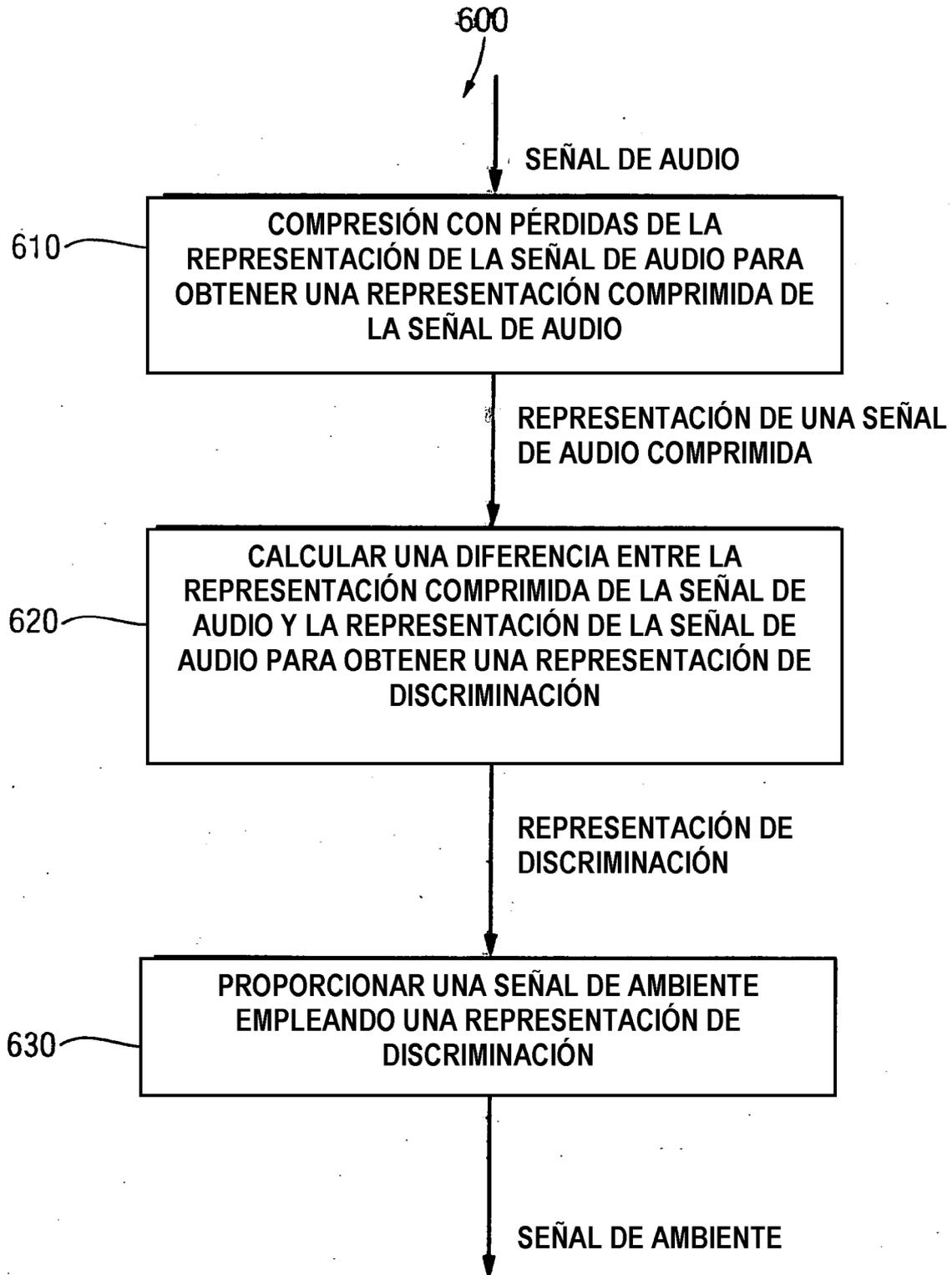


FIGURA 6



FIGURA 7



FIGURA 8

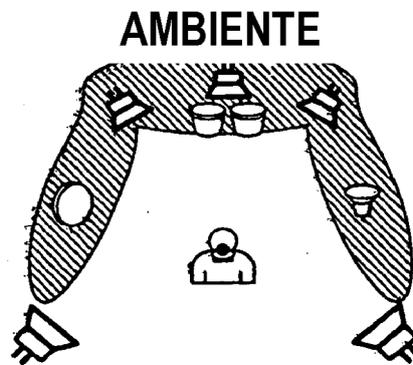


FIGURA 9