

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 461 191**

51 Int. Cl.:

G10L 19/008 (2013.01)

G10L 19/02 (2013.01)

H04S 5/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **23.10.2007** **E 12168768 (5)**

97 Fecha y número de publicación de la concesión europea: **02.04.2014** **EP 2500900**

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

30 Prioridad:

24.10.2006 DE 102006050068

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

19.05.2014

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 2 461 191 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, procedimiento y programa de ordenador para obtener una señal de audio multi-canal a partir de una señal de audio

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 mezclar

10

[0002] A continuación, se analizará 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 en medios de DVD ofrecen a menudo sonido multi-canal 5.1. Por esta razón,

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 L_S y R_S) dispuestos detrás o hacia atrás de un oyente y un canal de efectos de baja frecuencia (también denominado como LFE). Los tres altavoces dispuestos en la parte delantera (L, C, R) se denominarán también en lo sucesivo como altavoces frontales. Los altavoces dispuestos detrás y en la parte posterior del oyente (L_S , R_S) se denominan también en lo sucesivo altavoces posteriores.

20

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

25

[0005] Los sistemas de canales múltiples (tal como un sistema de audio multi-canal 5.1) proporcionan varias ventajas ya conocidas sobre la reproducción estéreo de dos canales. Esto se ejemplifica mediante las siguientes ventajas:

30

- Ventaja 1: estabilidad 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 un área de posiciones de escucha donde se puede percibir una impresión de sonido óptima (por un oyente).

35

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

[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 antiguas películas y series de televisión en CD y/o DVD que están disponibles en calidad mono y/o por medio de solamente una señal de audio "mono" de un único canal.

40

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

45

- Opción 1: Reproducción o *playback* del canal mono a través del centro o a través del altavoz central para obtener una fuente mono verdadera.

50

- 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 dentro o en el punto dulce.

55

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 (tales como un a configuración de altavoz con 5 ó 6

altavoces). El altavoz C o altavoz central, el altavoz L_S, o altavoz trasero izquierdo, el altavoz R_S, o altavoz trasero derecho y el altavoz LFE, o altavoz de canal de efectos de baja frecuencia, permanecen sin utilizar.

- 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, los seis altavoces usados en un sistema de canales múltiples 5.1). De esta manera, la señal multi-canal se beneficia de las ventajas que se han analizado previamente de la configuración de varios canales. El procedimiento se puede emplear en tiempo real o "sobre la marcha" o por medio de procesamiento y se conoce como proceso de mezcla o "mezcla".

10 **[0008]** 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. Sin embargo, particularmente con respecto a la señal generada para la alimentación de los altavoces posteriores, el procesamiento de la señal requerida no es obvio.

15 **[0009]** En la bibliografía, se describen dos conceptos diferentes para un procedimiento de mezcla o proceso de mezcla. Estos conceptos son el "Concepto Directo/Ambiente" y el "Concepto en la banda". A continuación, se describirán los dos conceptos indicados.

Concepto Directo/Ambiente

20 **[0010]** Las "fuentes de sonido directo" se reproducen a través de los tres canales frontales de forma que se perciben en la misma posición que en la versión original de dos canales. La expresión "fuente de sonido directo" se usa aquí para describir el sonido que viene, única y directamente de una fuente de sonido discreto (por ejemplo, un instrumento), y que muestra poco o ningún sonido adicional, por ejemplo, debido a reflejos de las paredes.

25 **[0011]** 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 estar presente en la grabación original). El sonido o ruido similar al ambiente no está asociado a una única 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 denominada "sensación de envolvente" del oyente. El sonido similar al ambiente o ruido similar al ambiente es ruido o sonido adicional por parte del público
30 en las actuaciones en directo (tal como aplausos) o el sonido ambiental o ruido ambiental añadido con intención artística (tal como el ruido de la grabación, canto de los pájaros, los sonidos de grillos chirriando).

[0012] A modo de ilustración, la figura 7 representa la versión original de dos canales (de una grabación de audio). La figura 8 muestra una entrega mezclada usando el concepto Directo/Ambiente.

35

Concepto en la banda

[0013] Siguiendo el concepto circundante, a menudo denominado como "concepto en la banda", cada sonido o ruido (sonido directo, así como el ruido ambiental) puede posicionarse completa y/o arbitrariamente alrededor del
40 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.

[0014] La figura 9 representa el concepto envolvente.

45 **[0015]** 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 usando el concepto Directo/Ambiente. La figura 9 describe la percepción de sonido y/o la localización del sonido usando el concepto envolvente.

50

[0016] La siguiente sección proporciona una visión general sobre la técnica anterior con respecto a la mezcla de una señal de un canal o de dos canales para formar una versión de múltiples canales. La bibliografía indica varios procedimientos para mezclar señales de un único canal y señales de múltiples canales.

55 Procedimientos no señal adaptivos

[0017] La mayoría de procedimientos para la generación de una señal denominada "pseudo estereofónica" no son señal adaptivos. 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. Puede encontrarse un análisis general de dicho sistema, por ejemplo, en [1].

Procedimientos Señal adaptativos

5

[0018] 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 una mezcla ciega.

10

[0019] Los decodificadores mencionados usan las diferencias entre canales y los mecanismos señaladaptivos de dirección a fin de crear señales de salida multicanal.

15 Extracción de ambiente y síntesis de señales estéreo para la mezcla de audio multi-canal

[0020] Avendano y Jot proponen una técnica del dominio de la frecuencia con el fin de identificar y extraer la información de ambiente en señales de audio estéreo (véase [5]).

20 **[0021]** El procedimiento se basa en calcular un índice de coherencia inter-canal y una función de mapeo 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 usan para alimentar los canales de sonido envolvente de un sistema de reproducción multi-canal.

25 Un procedimiento para convertir sonido estéreo en sonido multicanal

[0022] Irwan y Aarts muestran un procedimiento para convertir una señal de una representación estéreo en una representación de múltiples canales (véase [6]). La señal para los canales envolventes se calcula usando una técnica de correlación cruzada. Se usa un análisis de componentes principales (PCA) para calcular un vector que indica la dirección de la señal dominante. Entonces, este vector se mapea a partir de una representación de dos canales con respecto a una representación de tres canales con el fin de generar los tres canales frontales.

30

Mezclado basado en ambiente

35 **[0023]** Souloudre muestra un sistema que genera una señal de múltiples canales a partir de una señal estéreo (véase [7]). La señal se descompone en los llamados "flujos de fuente individuales" y "flujos de ambiente". En base a estos flujos, un denominado "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.

40 Pseudostereofonía basada en claves espaciales

[0024] Se describe un proceso cuasi-señaladaptivo pseudo-estereofónico por Faller en [1]. Este procedimiento usa una señal mono y grabaciones estéreo de la misma señal. Se extrae información adicional espacial o señales espaciales de la señal estéreo y se usan para convertir la señal mono en una señal estéreo.

45

[0025] 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 transmita una impresión auditiva particularmente buena.

50

[0026] Este objeto se consigue mediante un aparato para obtener una señal de ambiente a partir de una señal de audio de acuerdo con la reivindicación 1 ó 2, un procedimiento para obtener una señal de audio multi-canal a partir de una señal de audio según la reivindicación 3 ó 4, así como un programa de ordenador según la reivindicación 5.

55

[0027] Es una idea esencial de la presente invención que una señal de ambiente pueda ser generada a partir de una señal de audio de una manera particularmente eficiente mediante la determinación de una diferencia entre una representación comprimida de la señal de audio, que se generó por la compresión con pérdidas de una representación original de la señal de audio, y la representación original de la señal de audio. Es decir, 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

[0028] 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

[0029] 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

20

[0030] 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 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.

25

[0031] 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 etapas algorítmicamente sencillas, es decir, realizar una compresión con pérdidas, así como 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, cabe señalar que la señal de ambiente generada en la forma de la invención típicamente deja de tener porciones de alta energía que pueden interferir con la impresión auditiva como en el contexto de la compresión con pérdidas, dichas 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.

30

35

40

[0032] 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, la que representa el ruido de fondo.

45

[0033] Por lo tanto, el concepto de la invención permite la separación constante de información localizable y el ruido de fondo usando 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.

50

[0034] Además, la presente invención proporciona procedimientos que corresponden a los aparatos de la invención en lo que se refiere a su funcionalidad.

[0035] La presente invención proporciona adicionalmente un programa de ordenador para la realización de los procedimientos de la invención.

55

[0036] Las realizaciones preferidas particularmente ventajosas de la presente invención se definen de otro modo por las reivindicaciones adjuntas.

[0037] A continuación, las realizaciones preferidas de la presente invención se analizan con respecto a los dibujos adjuntos, en los que:

La figura 1 es un diagrama de bloques de un aparato para generar una señal de ambiente a partir de una señal de audio;

la figura 2 es un diagrama de bloques de un aparato para generar una señal de ambiente a partir de una señal de audio;

10 la figura 3 es un diagrama de bloques detallado de un aparato de la invención para generar una señal de ambiente a partir de una señal de audio;

la figura 4a es una representación ejemplar de una representación aproximada de una matriz por un producto de dos matrices;

15 la figura 4b es una representación esquemática de una matriz x;

la figura 5 es un diagrama de bloques de un aparato para obtener una señal de audio multi-canal a partir de una señal de audio;

20 la figura 6 es un diagrama de flujo de un procedimiento de la invención para la creación de una señal de ambiente a partir de una señal de audio;

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

25 la figura 8 es una representación esquemática de una impresión auditiva en un concepto directo/ambiente; y

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

30 **[0038]** La figura 1 muestra un diagrama de bloques de un aparato de la invención para generar una señal de ambiente a partir de una señal de audio.

[0039] El aparato de acuerdo con la figura 1 se designa 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. El medio 110 se configura 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.

40 **[0040]** El aparato 100 comprende adicionalmente 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. En base a 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.

50 **[0041]** El aparato 100 comprende adicionalmente medios 130 para proporcionar la señal de ambiente 132 usando y/o en base a y/o en función de la representación de discriminación 122.

[0042] En base a 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. En otras palabras, 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, por los medios 110, exactamente aquellas partes de la señal que exhiben una curva irregular se eliminan y/o no reproducen en la representación en la forma de compresión con pérdidas 112 de la señal de audio, la representación de discriminación 122 describe exactamente aquellas 5 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, con respecto a 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 10 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 suprimida en la representación discriminación 122 como se desea. 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, como se realiza por los medios 110 para la 15 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, por medio de la atenuación de las porciones de señal que tienen una 20 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 discriminación 112 todavía comprende una gran porción comparable 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, como se describen por la representación de discriminación 122, representan la información que da 25 como resultado una impresión auditiva particularmente buena y agradable en la reproducción (por medio de los altavoces posteriores).

[0043] 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) se suprimen o se atenúan. En 30 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 ni 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. 35

[0044] En otras palabras, todo lo que aparece repetidamente en la representación tiempo-frecuencia está bien aproximado por la compresión con pérdida.

[0045] La distribución de energía regular está aquí destinada a ser, por ejemplo, la distribución de energía que 40 presenta un patrón recurrente en una representación tiempo-frecuencia o que proporciona una concentración local de energía en la representación tiempo-frecuencia. Por ejemplo, la distribución de energía irregular es distribución de energía que no produce ningún patrón recurrente ni una concentración local de energía en una representación tiempo-frecuencia.

[0046] Dicho de otro modo, la señal de ambiente comprende sustancialmente porciones de señal que tienen una 45 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, estructurada en la representación tiempo-frecuencia como se ha descrito anteriormente). 50

[0047] Por lo tanto, los medios 130 para proporcionar la señal de ambiente en base a la representación de discriminación 122 proporcionan una señal de ambiente que está particularmente bien adaptada a las expectativas de un oyente humano.

[0048] 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. 55

[0049] La figura 2 muestra un diagrama de bloques de un aparato para generar una señal de ambiente a partir de una señal de audio. Además, la figura 3 muestra un diagrama de bloques detallado de un aparato para generar una

señal de ambiente a partir de una señal de audio. En su totalidad, el aparato de acuerdo con la figura 2 se designa con 200, y, en su totalidad, el aparato de acuerdo con la figura 3 se designa con 300.

[0050] 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.

[0051] 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 del tiempo $x[n]$. Ha de apreciarse 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, en la representación 212 de la distribución tiempo-frecuencia puede omitirse.

[0052] 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, se ha de señalar que, por ejemplo, la matriz $\mathbf{X}(\omega, k)$, que se explicará con mayor detalle a continuación, o bien la matriz $|\mathbf{X}(\omega, k)|$ puede servir como la representación 212 de la distribución tiempo-frecuencia.

[0053] Los medios 200 comprenden adicionalmente 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, 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. 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 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 a la representación aproximada, mientras que las porciones de señal que muestran energía distribuida irregularmente comparable y/o energía de señal comparativamente menor se atenúan o neutralizan 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.

[0054] 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, en base a 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.

[0055] 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 en base a la misma. Los medios de re-síntesis 240 pueden, por ejemplo, configurarse para convertir la representación de discriminación 232, que está presente en la forma de una distribución tiempo-frecuencia, en una señal temporal 242.

[0056] Se ha de señalar además que los medios de re-síntesis 240 son opcionales y pueden omitirse si hay reprocesamiento directo de la representación de discriminación 232, que puede, por ejemplo, estar presente en forma de una distribución tiempo-frecuencia, si se desea.

[0057] Los medios 200 comprenden adicionalmente medios 250 opcionales para el montaje de una señal de audio de múltiples canales y/o para el posprocesamiento. 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 $a_1[n]$, ..., $a_k[n]$) de la señal re-sintetizada 242.

[0058] La generación de la pluralidad de las señales ambientales 252, 254 se explicara con mayor detalle a

continuación.

[0059] 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 en el que se basa la presente invención. Este concepto de la invención se puede resumir en breve como se indica a continuación:

[0060] Una distribución tiempo-frecuencia 212 (TFD) de la señal de entrada 208 ($x[n]$) se calcula (opcionalmente) en unos 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 realizarse, por ejemplo, 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. Después de esto, se realiza 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 posprocesado (realizada por ejemplo en los medios opcionales 250 para la síntesis de una señal de audio multicanal y/o para el posprocesamiento) 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.

[0061] 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.

[0062] 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.

[0063] 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 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 a continuación, la distribución tiempo-frecuencia también se representa con $X(\omega, k)$.

[0064] Se ha de señalar además que la distribución tiempo-frecuencia $X(\omega, k)$ 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 preferiblemente cuando la distribución tiempo-frecuencia 312 puede adoptar valores complejos (no puramente reales). En este caso, el divisor de magnitud-fase 314 está configurado preferiblemente 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, en base a la distribución tiempo-frecuencia 312. Por el contrario, la representación de la magnitud de la distribución tiempo-frecuencia 312 también se designa con $|X(\omega, k)|$. Se ha de apreciar que la representación de la magnitud 316 de la distribución tiempo-frecuencia 312 puede sustituirse por la representación 212 en el aparato 200.

[0065] Además, se ha de apreciar que el uso de la representación de la fase 318 de la distribución tiempo-frecuencia 312 es opcional. También se ha de apreciar que la representación de fase 318 de la distribución tiempo-frecuencia 312 se designa en algunos casos, también con $\phi(\omega, k)$.

[0066] Se supone adicionalmente que la representación de la magnitud 316 de la distribución tiempo-frecuencia 312 está presente en forma de matriz.

[0067] El aparato 300 comprende adicionalmente 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 W , H , como se describirá a continuación. El aproximador de matriz 320 corresponde sustancialmente a los medios 220

para la aproximación como usa 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 $\hat{\mathbf{X}}(\omega, k)$. De otro modo, la aproximación 322 corresponde a la representación aproximada 222 de la figura 2.

5

[0068] El aparato 300 comprende adicionalmente 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 $|A(\omega, k)|$ 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.

[0069] El aparato 300 comprende adicionalmente 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, como se describe por la representación de fase 318. Por lo tanto, el sumador de fase 334 proporciona una representación de discriminación 336 dotada de una fase, que también se designa con $A(\omega, k)$. Ha de apreciarse 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 pueda, por ejemplo, estar sustituida por la representación de discriminación 336 dotada de una fase. Se ha de señalar además que, en función de cada caso particular, tanto la representación de discriminación 332 como la representación de discriminación 336 provista de una fase pueden corresponder a la representación de discriminación 232.

[0070] El aparato 300 comprende además un convertidor (opcional) de distribución tiempo-frecuencia a señal de tiempo 340. El convertidor (opcional) de distribución tiempo-frecuencia a señal de tiempo 340 está configurado 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 una representación de dominio de tiempo (o representación de señal de tiempo) de la señal de ambiente.

[0071] Se ha de observar adicionalmente que el convertidor (opcional) de distribución tiempo-frecuencia a señal de tiempo 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 340 corresponde sustancialmente a la señal 242, como se muestra en el aparato 200.

Distribución tiempo-frecuencia de la señal de entrada

[0072] 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, usando 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:

45

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

[0073] 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.

50

[0074] Si la distribución tiempo-frecuencia (TFD) es de valores complejos (por ejemplo, en el caso de la utilización de una transformada de Fourier de corto tiempo (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. Ha de apreciarse que en el aparato 300, la representación de magnitud $|X(\omega, k)|$ se designa con 316.

55

La información de fase $\phi(\omega, k)$ se designa con 318.

[0075] Ha de apreciarse 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 5 contraste, $\mathbf{X}(\omega, k)$ denota una matriz que contiene una pluralidad de coeficientes (ω, k) . Por ejemplo, la matriz $\mathbf{X}(\omega, k_1)$ contiene coeficientes $X(\omega', k')$ para $\omega' = 1, 2, \dots, n$ y $k' = k_1, k_1+1, k_1+ m-1$. 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:

$$10 \quad X_{i,j} = X(\omega = \omega_i, k = k_{1+j-1}).$$

[0076] Aquí, se cumple lo siguiente:

$$15 \quad 1 \leq j \leq n$$

y

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

20 **[0077]** El contexto descrito también se muestra en la figura 4b.

[0078] En otras palabras, la matriz $\mathbf{X}(\omega, k)$ comprende una pluralidad de valores de distribución de tiempo frecuencia $X(\omega, k)$.

25 **[0079]** Se ha de señalar adicionalmente que a continuación, 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)

30 **[0080]** 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.

35 **[0081]** Una aproximación $\hat{\mathbf{X}}(\omega, k)$ 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 del 50%, preferiblemente no más del 20%. De lo contrario, una minimización puede ser una determinación de un mínimo absoluto o local.

40 **[0082]** 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 \mathbf{A}, \mathbf{B} se cumple lo siguiente:

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

45

[0083] En contraste con esto, la divergencia puede ser asimétrica.

[0084] Ha de apreciarse 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, efectuarse por medio de los medios de 50 aproximación 220 o el aproximador de matriz 320.

[0085] Se ha de señalar adicionalmente que la factorización de matriz no negativa (NMF) es un procedimiento adecuado para el cálculo de la aproximación.

Factorización de matriz no negativa (NMF)

[0086] A continuación, se describe la factorización de matriz no negativa. Una factorización de matriz 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 $\mathbf{H} \in \mathbf{R}^{r \times m}$. Aquí, para los elementos $W_{i,k}$ de la matriz \mathbf{W} y $H_{i,k}$ de la matriz \mathbf{H} , se cumple lo siguiente:

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

y

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

10

[0087] En otras palabras, las matrices \mathbf{W} y \mathbf{H} se determinan de tal manera que se cumple lo siguiente:

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

15

[0088] Expresando este elemento de forma amplia, se cumple lo siguiente:

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

20 [0089] Si el rango r de la factorización satisface la condición

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

entonces el producto \mathbf{WH} es una representación de datos comprimidos de \mathbf{V} (véase [8]). Una explicación intuitiva de la ecuación (2) es como se indica a continuación: la matriz $\mathbf{V} \in \mathbf{R}^{n \times m}$ se aproxima como la suma de los r productos externos de un vector columna \bar{w}_i y un vector fila \bar{h}_i , en el que se cumple lo siguiente: $i \in [1, r]$, $\bar{w}_i \in \mathbf{R}^{n \times 1}$ y $\bar{h}_i \in \mathbf{R}^{1 \times m}$. La materia objeto descrita se representa por un simple ejemplo en la figura 4a. En otras palabras, la figura 4a muestra un ejemplo ilustrativo de una factorización de matriz no negativa (NMF) con un rango de factorización $r=2$.

30

[0090] Los factores \mathbf{W} y \mathbf{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. En otras palabras, la función de coste c mide el error de la aproximación, es decir, la distancia (y/o la divergencia) entre las matrices \mathbf{V} y \mathbf{WH} . Una medida de la distancia adecuada entre las dos matrices \mathbf{A} y \mathbf{B} es la norma de Frobenius $D_F(\mathbf{A}, \mathbf{B})$ en su diferencia por elemento (ecuación 3):

35

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

[0091] La norma de Frobenius es ideal para datos distribuidos según Gauss no correlacionados (véase [9]). En otras palabras, una función de coste c se calcula en la realización preferida, en la que se cumple lo siguiente:

40

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

[0092] En otras palabras, la aproximación $\hat{\mathbf{X}}(\omega, k)$ se calcula como el producto de dos matrices, \mathbf{W} y \mathbf{H} , en el que:

45

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

[0093] Una función de error más conocida es la 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 de Kullback-Leibler entre dos matrices **A** y **B** es la siguiente:

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

10 **[0094]** De otro modo, A_{ij} y B_{ij} son las entradas o elementos de matriz de las matrices **A** y **B**, respectivamente.

[0095] En otras palabras, la función de coste c puede seleccionarse como se indica a continuación:

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

15

[0096] Lo que sigue es una descripción de cómo se pueden determinar las entradas de las matrices de aproximación **W** y **H**. Una técnica sencilla 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

20

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

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

25 **[0097]** 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 se da por las siguientes ecuaciones:

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

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

30 **[0098]** En el presente contexto se cumple lo siguiente:

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

35 **[0099]** Debe apreciarse adicionalmente que Lee y Seung han descubierto o identificado una regla de actualización multiplicativa o regla de iteración de acuerdo con las ecuaciones (8) y (9) (véase [10]). Además, Lee y Seung han demostrado la relación de la regla de actualización multiplicativa con respecto al procedimiento de descenso de gradiente y la convergencia de los mismos. Las reglas de actualización multiplicativas son como se indican a continuación:

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

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

[0100] De nuevo, se cumple lo siguiente:

$$V = X(\omega, k) .$$

5

[0101] La velocidad y la solidez del procedimiento de descenso de gradiente dependen en gran medida de la elección correcta del tamaño del paso o anchura de paso α . 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 método es fácil de implementar, computacionalmente eficiente y garantiza la búsqueda de un mínimo local de la función de coste.

10

Factorización de matriz no negativa (NMF) en el contexto de la separación ambiente

[0102] En el contexto del procedimiento presentado, se usa una factorización de matriz no negativa (NMF) para calcular una aproximación del espectrograma de cantidad o magnitud $|X(\omega, k)|$ de la señal de audio de entrada $x[n]$. Con respecto a esto, cabe señalar que el espectrograma de magnitud $|X(\omega, k)|$ se obtiene a partir de la matriz $X(\omega, k)$ realizando una formación de magnitud que incluye todo el elemento. Dicho de otro modo, para el elemento que tiene los índices i, j de $|X(\omega, k)|$, designado con $|X(\omega, k)|_{ij}$, se cumple lo siguiente:

20

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

[0103] Aquí, $X(\omega, k)_{ij}$ designa un elemento de la matriz $X(\omega, k)$ con los índices i y j . Por otro lado, $|\cdot|$ designa la operación de formación de magnitud.

25

[0104] La factorización de matriz no negativa (NMF) de $|X|$ da como resultado 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 particular mediante la aproximación.

30

[0105] En resumen, se muestra que mediante la factorización de matriz no negativa que se ha descrito 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 como se designa adicionalmente con 322 o $\tilde{X}(\omega, k)$ 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 , como se representa en la ecuación (10):

35

$$|A| = |X| - WH \quad (10)$$

[0106] Sin embargo, el resultado de acuerdo con la ecuación 10 no se considera directamente como se explicará a continuación. Es decir, para aproximaciones que minimizan las funciones de coste que se han descrito anteriormente, la aplicación de la ecuación (10) da como resultado un espectrograma de cantidad o magnitud $|A|$ con elementos de valor tanto negativo como positivo. Sin embargo, puesto que en una realización se prefiere que el espectrograma de cantidad o magnitud $|A|$ incluya únicamente elementos de valor positivo, se prefiere emplear un procedimiento que maneje los elementos de valor negativo de la diferencia $|X| - WH$.

45

[0107] Pueden emplearse varios procedimientos para la manipulación de los elementos negativos. Un sencillo enfoque para la manipulación de los elementos negativos consiste en multiplicar los valores negativos con un factor β entre 0 y -1 ($\beta = 0, \dots, -1$). En otras palabras: $-1 \leq \beta \leq 0$. Aquí, $\beta = 0$ corresponde a una rectificación de media onda, y $\beta = -1$ corresponde a una rectificación de onda completa.

5

[0108] Una formulación general para el cálculo del espectrograma de magnitud o el espectrograma amplitud $|\mathbf{A}|$ de la señal de ambiente se determina por las siguientes ecuaciones:

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

10

con

$$\beta_{ik} = \begin{cases} \gamma, & \text{if } (\mathbf{WH})_{ik} > |\mathbf{X}|_{ik} \\ +1, & \text{de otro modo} \end{cases} \quad (12)$$

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

[0109] Debe apreciarse que en la ecuación anterior, $|\mathbf{A}|_{ik}$ designa un elemento de matriz con los índices i y k del espectrograma de magnitud o espectrograma de amplitud $|\mathbf{A}|$. Además, $(|\mathbf{X}| - \mathbf{WH})_{ik}$ designa un elemento de matriz de una diferencia entre el espectrograma de magnitud o el espectrograma de amplitud $|\mathbf{X}|$ de la distribución tiempo-frecuencia y la aproximación asociada $\mathbf{WH} = \hat{\mathbf{X}}$, que tiene los índices i y k .

20

[0110] Además, $(\mathbf{WH})_{ik}$ representa un elemento de matriz de la aproximación $\mathbf{WH} = \hat{\mathbf{X}}$ con los índices i y k . $|\mathbf{X}|_{ik}$ es un elemento de matriz del espectrograma de cantidad o magnitud $|\mathbf{X}|$ con los índices i y k . Por lo tanto, se puede apreciar a partir de las ecuaciones (11) y (12) que el factor β_{ik} y/o la rectificación de las entradas de la diferencia $(|\mathbf{X}| - \mathbf{WH})$ se determina preferiblemente elemento por elemento.

25

[0111] A continuación, se describe un procedimiento alternativo para la determinación del espectrograma de cantidad o magnitud $|\mathbf{A}|$ de la señal de ambiente. Se obtiene una alternativa sencilla determinando en primer lugar el espectrograma de cantidad o magnitud $|\mathbf{A}|$ de la señal ambiental de acuerdo con

30

$$|\mathbf{A}| = |\mathbf{X}| - \zeta \cdot \mathbf{WH},$$

en la que $0 \leq \zeta \leq 1$ y efectuando, después de esto, una rectificación de onda completa de elementos negativos en la matriz determinada de este modo $|\mathbf{A}|$. 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.

35

[0112] Debe apreciarse 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 $|\mathbf{A}|$, 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).

40

[0113] Adicionalmente, hay un tercer procedimiento alternativo adicional para determinar la matriz $|\mathbf{A}|$, como se describe a continuación. El tercer procedimiento alternativo consiste en añadir una restricción límite o condición límite con respecto a la función de coste con el fin de influir en la cantidad o el valor de los elementos de valor negativo en el término

45

$$|\mathbf{A}| = |\mathbf{X}| - \mathbf{WH}$$

[0114] En otras palabras, la elección apropiada de la restricción límite o condición límite respecto a la función de coste puede servir para conseguir que puedan producirse tan pocos valores negativos como sea posible (como alternativa: tan pocos valores positivos como sea posible), por ejemplo, en la diferencia $|\mathbf{A}| = |\mathbf{X}| - \mathbf{WH}$.

5 [0115] Dicho de otro modo, el procedimiento de optimización para determinar las entradas de las matrices \mathbf{W} y \mathbf{H} está adaptado de tal manera que la diferencia mencionada comprende preferiblemente valores positivos y/o valores comparativamente menos negativos (o viceversa).

[0116] Puede formularse una nueva función de coste
10

$$c = f(\mathbf{X}, \mathbf{WH})$$

como se indica a continuación:

$$c = \sum_{i,k} (|\mathbf{x}|_{i,k} - (\mathbf{WH})_{i,k})^2 - \epsilon \sum_{i,k} (|\mathbf{x}|_{i,k} - (\mathbf{WH})_{i,k}) \quad (13)$$

[0117] Aquí, ϵ es una constante para determinar la influencia de la restricción límite o condición límite 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 $\partial c / \partial \mathbf{H}$ (de acuerdo con la ecuación 14) y el operador de derivación $\partial c / \partial \mathbf{W}$ en la ecuación (5). Para los operadores de derivación $\partial c / \partial \mathbf{H}$ y $\partial c / \partial \mathbf{W}$, se cumple lo siguiente:
20

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

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

[0118] De otro modo, debe apreciarse que se prefiere el procedimiento como se ha descrito con respecto a las
25 ecuaciones (11) y (12) ya que es fácil de poner en práctica y proporciona buenos resultados.

[0119] En resumen, se muestra que la determinación de la matriz $|\mathbf{A}|$ que se ha descrito anteriormente, para lo cual se describieron tres procedimientos diferentes, puede ejecutarse, por ejemplo, mediante los medios de determinación de la diferencia 230 o el formador de diferencia 330.
30

Reconstrucción de la señal de tiempo

[0120] A continuación se muestra una descripción de cómo puede obtenerse la representación $\mathbf{A}(\omega, k)$ proporcionada con una información de fase (también designada con 336) a partir de la representación de magnitud
35 $|\mathbf{A}(\omega, k)|$ (también designada con 332) de la señal de ambiente.

[0121] El espectrograma complejo $\mathbf{A}(\omega, k)$ de la señal de ambiente se calcula usando la fase $\phi = \angle \mathbf{X}$ de la distribución tiempo-frecuencia (TFD) \mathbf{X} de la señal de entrada 308 (también designada con $x(t)$, $x[n]$) se calcula de acuerdo con la ecuación (16):
40

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

[0122] Aquí, ϕ es, por ejemplo, una matriz de valores de ángulo. En otras palabras, 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 $|A|$. En otras palabras, 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 total es una representación $A(\omega,k)$ de la señal de ambiente proporcionada con una información de fase (designada con 336).

[0123] 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 entonces (opcionalmente) a partir de la representación $A(\omega,k)$ dotada de una información de fase, sometiendo $A(\omega,k)$ a un proceso inverso de cálculo de la distribución tiempo-frecuencia (TFD). Es decir, una representación $A(\omega,k)$ dotada de una información de fase se procesa, por ejemplo, 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 $X(\omega,k)$.

[0124] 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 usando una ventana de Hann para garantizar una transición suave entre los segmentos adyacentes.

[0125] Debe apreciarse que los procedimientos para la obtención de la representación de tiempo $a[n]$ de la señal de ambiente descritos por último pueden realizarse, por ejemplo, en los medios 240 para la re-síntesis o en el convertidor de distribución tiempo-frecuencia a señal de tiempo 340.

Montaje de una señal de audio multicanal

[0126] Se obtiene 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) mediante la alimentación de los canales posteriores (es decir, por ejemplo, al menos el canal trasero izquierdo o el canal trasero derecho, o tanto el canal posterior izquierdo como el canal posterior derecho) con la 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. Aquí, por ejemplo, los parámetros de ganancia y/o los parámetros de sonoridad aseguran que se obtiene una energía total (o permanece sustancialmente sin cambios) cuando se usa el canal central adicional.

[0127] Además, debe apreciarse 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, 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 suministrarse, por ejemplo, a uno o ambos altavoces laterales y/o a uno o ambos altavoces posteriores.

[0128] Después de la separación del ambiente (o después de la generación de la señal de ambiente), el procesamiento adicional puede realizarse opcionalmente con el fin de obtener una señal de audio multicanal de alta calidad perceptual. Al montar una señal de audio multicanal a partir de un único canal, se desea que la imagen frontal se conserve mientras que se añade la impresión de amplitud. Por ejemplo, esto se consigue mediante la introducción o adición de retardo de unos pocos milisegundos a la señal de ambiente y/o por la supresión de porciones transitorias en la señal de ambiente. Además, es ventajosa la descorrelación de la señales 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

[0129] Los algoritmos para la detección de transitorios (y/o picos u operaciones de ajuste) y para manipular transitorios se usan en diversas aplicaciones de procesamiento de señales de audio, tal como para los efectos de audio digitales (véanse [11, 12]) y para el mezclado (véase [13]).

[0130] La supresión de los transitorios en el contexto de la mezcla tiene como objetivo mantener la imagen frontal. Cuando el ruido transitorio o el 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. Éste es un efecto no deseado: la "fuente de sonido directo" o parece 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 traseros

[0131] En la bibliografía, el término "descorrelación" describe un proceso que manipula una señal de entrada de tal manera que (2 o más) señales de salida muestran diferentes formas de onda pero suenan igual que la señal de entrada (véase [14]). Si, 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 (véase [15]). La disminución de la correlación de las dos señales de canal aumenta el ancho o la extensión percibida de la fuente de sonido o la 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} , como se describe por la ecuación (17):

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

[0132] Aquí, $y^*(k)$ representa el número complejo conjugado de $y(k)$. Ya que 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 centradas x e y se define por el uso o la correlación entre canales Γ (véase [15]) o por la coherencia entre canales (véase [16]) (ecuación (18)). En la ecuación (18), la correlación entre canales o la coherencia entre canales Γ se define como se indica a continuación:

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

[0133] Aquí, la correlación cruzada normalizada r_{xy} se define de acuerdo con la ecuación (19):

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

[0134] Los ejemplos de procesos de descorrelación son la reverberación natural y varios procesadores de señal (*flanger, chorus, phaser, reverberación sintética*).

[0135] Un anterior procedimiento de descorrelación en el campo del procesamiento de señal de audio se describe en [17]. Aquí, se generan dos señales del canal de salida por la suma de la señal de entrada y una versión retardada de la señal de entrada, donde en un canal, se invierte la fase del canal retardado.

[0136] 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 (véase [14]).

[0137] Se obtiene una descorrelación dinámica (es decir, variable en el tiempo) mediante el uso de filtros Allpass variables en el tiempo, es decir, filtros Allpass en los que las nuevas respuestas de fase aleatorias se calculan para lapsos de tiempo adyacentes (véanse [18], [11]).

[0138] En [18], se describe un procedimiento de subbanda, en el que la correlación en las bandas de frecuencia individuales se cambia de forma variable.

[0139] En el contexto del procedimiento descrito aquí, se aplica una descorrelación a la señal de ambiente. En una configuración 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 de ambiente que finalmente se alimentan a los dos canales posteriores o traseros estén descorrelacionadas entre sí al menos en cierta medida.

[0140] Las propiedades deseadas del presente procedimiento son la difusión de campo de sonido (o difusión de

campo de ruido de o ampliación de campo de sonido o de ruido) y envolvente.

[0141] En lo que sigue y con referencia a la figura 5, se describe un aparato para obtener 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
5 para obtener la señal de audio multicanal 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. Debe apreciarse que, preferiblemente, el aparato 510 es el aparato 100 de acuerdo con la figura 1, el aparato 200 de acuerdo con la
10 figura 2 o aparato 300 de acuerdo con la figura 3.

[0142] La señal de ambiente 512, que puede estar presente en forma de una representación de dominio en el tiempo (o representación de tiempo-sígnal) y/o en una representación tiempo-frecuencia se alimenta adicionalmente a los medios de posprocesado 520. Los medios de posprocesado 520 son opcionales y pueden, por ejemplo,
15 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 mostrar una pendiente de flanco mayor que una inclinación de flanco máxima admisible determinada. Además, los eventos transitorios también podrían de lo contrario ser picos de señal en la señal de ambiente 512, cuyas amplitudes exceden una cierta amplitud máxima dada.
20

[0143] Además, los medios de posprocesado 520 podrían (opcionalmente) comprender un retardador o medios de retardo que retrasan la señal de ambiente 512. Por lo tanto, los medios de posprocesado 520 proporcionan una señal de ambiente posprocesada 522 en la que, por ejemplo, los transitorios se reducen o se eliminan en comparación con la señal de ambiente (original) 512 y/o que, por ejemplo, se retrasa en comparación con la señal
25 de ambiente (original) 512.

[0144] Si se omiten los medios de posprocesado 520, entonces la señal 522 puede ser idéntica a la señal 512.

[0145] El aparato 500 comprende adicionalmente (de forma opcional) un combinador 530. Si el combinador está
30 incluido, el combinador 520, por ejemplo, proporciona una señal de altavoz posterior 532, que está formada por una combinación de la señal de ambiente posprocesada 522 y una versión (opcionalmente posprocesada) de la señal de audio original 508.

[0146] Si se omite el combinador opcional 530, entonces la señal 532 puede ser idéntica a la señal 522. El aparato
35 500 comprende adicionalmente (opcionalmente) un descorrelador 540, que recibe la señal del altavoz posterior 532 y en base al mismo suministra al menos dos señales de los altavoces posteriores descorrelacionadas 542, 544. La primera señal de altavoz posterior 542 puede, por ejemplo, representar una señal de altavoz posterior para un 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

[0147] En el caso más simple (por ejemplo, si se omiten los medios de posprocesado 520, el combinador 530 y el descorrelador 540), por ejemplo la señal de ambiente 512 generada por el aparato 510 se usa 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 posprocesado 520, el combinador 530 y/o el descorrelador 540, se considera la señal de
45 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.

[0148] Por lo tanto, la presente invención comprende explícitamente el uso 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
50 posterior 544.

[0149] Asimismo, la presente invención también comprende explícitamente generar la primera señal de altavoz posterior 542 y/o la segunda señal de altavoz posterior 544 usando la señal de ambiente 512 generada por el aparato 510.
55

[0150] El aparato puede configurarse, opcionalmente, para generar una primera señal de altavoz frontal, una segunda señal de altavoz delantero y/o una tercera señal de altavoz frontal. Para este fin, por ejemplo, la señal de audio (original) 508 se suministra a los medios de posprocesado 550. Los medios de posprocesado 550 se configuran para recibir y procesar la señal de audio 508 y generar una señal de audio posprocesada 552, que, por

ejemplo, (opcionalmente) se suministra al combinador 530. Si se omiten los medios de posprocesado, la señal 542 puede ser idéntica a la señal 508. De otro modo, la señal 552 forma una señal de altavoz frontal.

[0151] Preferiblemente, el aparato 500 comprende un divisor de señal 560 configurado para recibir la señal de altavoz frontal 552 y generar, en base a la misma, 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.

10

[0152] La figura 6 muestra de otro modo un diagrama de flujo de un procedimiento para crear una señal de ambiente a partir de una señal de audio. 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.

[0153] Una tercera etapa 630 comprende proporcionar una señal de ambiente usando la representación de discriminación. Por lo tanto, en su conjunto, el procedimiento 600 permite la generación de una señal de ambiente a partir de una señal de audio.

[0154] Debe apreciarse aquí que el procedimiento de la invención 600 de acuerdo con la figura 6 puede completarse por las etapas que se ejecutan por los aparatos anteriores. Por lo tanto, el procedimiento puede, por ejemplo, modificarse y/o complementarse 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.

[0155] En otras palabras, el aparato y el procedimiento pueden implementarse en hardware o en software. La implementación puede realizarse 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 que con un sistema informático programable se ejecuta el procedimiento respectivo. De acuerdo con la presente invención, por lo tanto, se proporciona un programa informático, como se expone en la reivindicación independiente 5.

35 *Descripción general del procedimiento*

[0156] En resumen, se puede decir que se genera una señal de ambiente a partir de la señal de entrada y se suministra a los canales posteriores. Aquí, puede usarse un concepto 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 la que la figura 2 muestra un diagrama de bloques de un procesamiento, ya que puede usarse para la obtención de la señal de ambiente.

[0157] En resumen, se muestra lo siguiente:

[0158] Una distribución tiempo-frecuencia (TFD) de la señal de entrada se calcula como se analiza en el epígrafe "Distribución tiempo-frecuencia de la señal de entrada". Una aproximación de la distribución tiempo-frecuencia (TFD) de la señal de entrada se calcula usando el procedimiento de optimización numérica, 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**. Una re-síntesis de la señal temporal de la señal ambiental se explica también en la sección con el título "Reconstrucción de la señal temporal". Además, el posprocesado puede usarse (opcionalmente) para mejorar la impresión auditiva de la señal multicanal obtenida, como se describe en el título "Montaje de una señal de audio multicanal".

55 *Conclusión*

[0159] En resumen, se puede decir que lo anterior describe un procedimiento y un concepto para separar una señal de ambiente a partir de señales de un único canal de audio (o de una señal de audio de un canal). La señal de ambiente obtenida muestra una alta calidad de audio. Comprende elementos de sonido o elementos de ruido

procedentes del 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.

5 **[0160]** Las razones para el éxito del procedimiento descrito pueden describirse como se indica a continuación de forma simplificada:

10 **[0161]** Las distribuciones tiempo-frecuencia (TFD) de sonido directo o ruido directo son generalmente más escasas o menos densas que las distribuciones de frecuencia-tiempo (TFD) de ruido ambiente o sonido ambiente. Es decir, la energía del ruido directo o sonido directo se concentra más en menos celdas o entradas de matriz que la energía de ruido ambiente o sonido ambiente. Por lo tanto, la aproximación directa detecta el ruido o el sonido directo, pero no (o sólo en muy poca medida) el ruido ambiente o el sonido ambiente. Como alternativa, se puede decir que la aproximación directa detecta el ruido o el sonido directo en mayor grado que el ruido ambiente o el sonido ambiente. Por lo tanto, la distinción o diferencia entre la distribución tiempo-frecuencia (TFD) de la señal de entrada y su aproximación es una buena representación de la distribución tiempo-frecuencia (TFD) de todo el ruido ambiental y/o sonido ambiental presente en la señal de entrada.

20 **[0162]** Sin embargo, lo anterior 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.

25 **[0163]** Además, debe apreciarse que preferiblemente, no se usa ningún efecto de audio artificial y que la manipulación del sonido y/o señales de audio únicamente implican envolvimiento y amplitud. No hay ninguna coloración del tono del sonido original o del ruido original. Se mantiene la impresión auditiva prevista por el autor de la señal de audio.

30 **[0164]** Por lo tanto, se puede decir que el procedimiento descrito y el concepto superan inconvenientes sustanciales de los procedimientos o conceptos conocidos. Debe apreciarse 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 entre canales 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.

35 **[0165]** El procedimiento descrito en el título "Pseudostereofonía basada en un entorno espacial" siempre requerirá una versión multi-canal del mismo contenido o un operador que genera las pistas espaciales manualmente. Por lo tanto, el procedimiento conocido mencionado no puede emplearse ni de manera en tiempo real ni automáticamente cuando no hay una versión multi-canal de la misma señal de entrada.

40 **[0166]** En contraste, el procedimiento y el concepto descritos en este documento 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).

45 **[0167]** A continuación, se describe una elección particularmente ventajosa de parámetros para la aplicación del concepto anterior.

50 **[0168]** En otras palabras, a continuación, se describen los ajustes de parámetros óptimos para el procedimiento de separación de ambiente para aplicaciones mono-upmix. Además, se darán valores mínimos y máximos de los parámetros, que, aunque pueden funcionar, no producen resultados óptimos con respecto a la calidad de audio y/o la carga de procesamiento requerida.

55 **[0169]** Aquí, el tamaño del parámetro FFT (n_{fft}) describe cómo se procesan muchas bandas de frecuencia. En otras palabras, el tamaño del parámetro FFT indica, cuántas frecuencias discriminables existen ω_1 a ω_n . Por lo tanto, el tamaño del 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 $\mathbf{X}(\omega, k)$. En otras palabras, el tamaño del parámetro FFT describe el número de filas (o columnas) de la matriz $\mathbf{X}(\omega, k)$. Por lo tanto, el tamaño del parámetro FFT, por ejemplo, corresponde al valor n . Además, el tamaño del valor FFT también describe cuántas muestras se usan para el cálculo de una única entrada $X_{i,j}$ de la \mathbf{X} . Dicho de otro modo, se usan las muestras n_{fft} de una representación de tiempo de la señal de

entrada para calcular en base a estas los coeficientes espectrales nfft para frecuencias nfft diferentes ω_1 a ω_{nfft} . Por lo tanto, en base a las muestras nfft, se calcula una columna de la matriz $\mathbf{X}(\omega, k)$.

5 **[0170]** La ventana que define las muestras contempladas de la señal de entrada se desplaza entonces varias muestras definidas por el salto de parámetros. Las muestras nfft de la señal de entrada definida por la ventana desplazada se mapean entonces con respecto a los coeficientes espectrales nfft por una transformada de Fourier, definiendo el coeficiente espectral una columna siguiente de la matriz \mathbf{X} .

10 **[0171]** Se puede decir a modo de ejemplo 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.

15 **[0172]** La longitud del segmento de parámetros indica cuán largo es un segmento de una trama de señal, cuyo espectrograma se factoriza. En otras palabras, la longitud del segmento de parámetros describe cuán larga es una duración de tiempo de la señal de audio de entrada que se considera 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 de tiempo de entrada durante un período igual a la longitud del segmento de parámetros (SegLen).

20 **[0173]** El rango de factorización de parámetros describe el rango de factorización de la factorización de matriz no negativa, es decir, el parámetro r . Dicho de otro modo, el rango de factorización de parámetros indica cuán grande es una dimensión de la primera matriz de aproximación \mathbf{W} y una dimensión de la segunda matriz de aproximación \mathbf{H} .

25 **[0174]** Los valores preferidos para los parámetros se dan en la siguiente tabla:

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

30 **[0175]** Como parámetro adicional, se determina adicionalmente qué medida de error c se usa 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 usar otras medidas de distancia cuando se procesan los valores de espectrograma con los valores de espectrograma de energía o del logaritmo tomado (SPL).

35 **[0176]** Además, debe apreciarse que los intervalos de valor preferidos se han descrito anteriormente. Cabe señalar que, usando el procedimiento anterior, el tamaño de la FFT puede estar en un intervalo de 128 a 65.536. El tamaño de salto puede estar entre 1/6 4 del tamaño de la FFT y una unidad del tamaño de la FFT. La longitud del segmento asciende típicamente a al menos 0,1 segundos.

40 **[0177]** 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 la mezcla de señales de audio de música para su reproducción en sistemas de canales múltiples. Una ventaja 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.

45 **[0178]** Además, debe apreciarse que la presente invención también puede usarse en un sistema sencillo. 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 reproducirse en el altavoz delantero. La señal ambiental obtenida a partir de la señal de audio original se puede reproducir en el altavoz posterior. En otras palabras, la señal mono de audio original se puede reproducir como una señal mono en un altavoz delantero único, mientras que la señal de ambiente obtenida de la señal de audio original se reproduce como un canal de posterior único.

- 5 [0179] Sin embargo, si hay varios canales, pueden procesarse individualmente. En otras palabras, se considera un primer canal de la señal de audio original para la generación de una primera señal de ambiente, y se usa un segundo canal de la señal de audio original para generar una segunda señal de ambiente. Después, 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 (por ejemplo, trasero izquierdo), mientras que la segunda señal de ambiente se reproduce, por ejemplo, en un segundo altavoz posterior (por ejemplo, derecho trasero).
- 10 [0180] Por lo tanto, también es posible generar dos señales de altavoces posteriores a partir de dos señales de altavoz frontal en la forma descrita.
- 15 [0181] En una realización adicional, la señal de audio original comprende 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 tercera señal de ambiente (opcionalmente).
- 20 [0182] Después, 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) con el fin de obtener una primera señal de altavoz de ambiente, que se suministra a un primer altavoz de ambiente (por ejemplo, un altavoz trasero izquierdo).
- 25 [0183] Opcionalmente, además, dos señales de ambiente adicionales (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 suministra a un segundo altavoz de ambiente (por ejemplo, un altavoz trasero derecho).
- 30 [0184] Por lo tanto, se forma una primera señal de altavoz de ambiente por una primera combinación de señales de ambiente, 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 de ambiente. La primera combinación comprende preferiblemente al menos dos señales de ambiente, y la segunda combinación comprende preferiblemente al menos dos señales de ambiente. Además, se prefiere que la primera combinación sea diferente de la segunda combinación, en la que, sin embargo, se prefiere que la primera combinación y la segunda combinación usen una señal de ambiente común.
- 35 [0185] Además, debe apreciarse que una señal de ambiente generada en la forma de la invención puede también suministrar, por ejemplo, a un altavoz lateral si se usa, por ejemplo, una disposición de altavoces que comprende altavoces laterales. Por lo tanto, puede suministrarse una señal de ambiente a un altavoz lateral izquierdo en un uso de una disposición de altavoces 7.1. Además, también puede suministrarse una señal de ambiente al altavoz lateral derecho, en el que la señal de ambiente suministrada al altavoz lateral izquierdo difiere preferiblemente de la señal de ambiente suministrada al altavoz derecho.
- 40 [0186] Por lo tanto, la presente invención en su conjunto proporciona una extracción particularmente buena de una señal de ambiente de una señal de un canal.
- 45 [0187]

Referencias:

50 [0187]

- [1] C. Faller, Pseudostereophony Revisited, Presentado en la 118ª Convención de AES, Barcelona, España, 2005.
 [2] R. Dressler, "Dolby Surround Prologic2Decoder -Principles of operation", Tech. Rep., Dolby Laboratories, http://www.dolby.com/assets/pdf/tech_library/208_Dolby_Surround_Pro_Logic_Decoder.pdf, 2000.
 55 [3] DTS Technical Stuff, "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>.
 [5] C. Avendano y 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, mayo de 2002.
- [6] R. Irwan y R. M. Aarts, "A method to convert stereo to multi-channel sound", Presentado en la 19ª Conferencia Internacional de AES, Schloss Elmau, Alemania, 21-24 de junio, páginas 139-143, 2001.
- 5 [7] G. Soulodre, "Ambience-Based Upmixing", Workshop Spatial Coding of Surround Sound: A Progress Report, Presentado en la 117ª Convención de AES, San Francisco, CA, Estados Unidos, 2004.
- [8] D. Lee y H. Seung, "Learning the parts of objects by Non-negative Matrix Factorization", en Letters To Nature, 1999.
- [9] A. Cichocki, R. Zdunek y S. Amari, "New algorithms for Non-negative Matrix Factorization in applications to Blind
10 Source Separation". En Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2006.
- [10] D. Lee y 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", en Proceedings of DAFX-01, 2001.
- 15 [12] M. Goodwin y C. Avendano, "Enhancement of Audio Signals Using Transient Detection and Modification", Presentado en la 117ª Convención de AES, San Francisco, CA, Estados Unidos, 2004.
- [13] A. Walther, C. Janssen, J. Herre y O. Hellmuth "Transient suppression in ambience-like signals "Patent FPL-Fallnummer: 06F47242-IIS.
- [14] G. Kendall, "The decorrelation of audio signals and its impact on spatial imagery", Computer Music Journal, 19:4,
20 1995.
- [15] C. Faller y 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 y I. Burnett, "Decorrelation techniques for the rendering of apparent sound source width in 3D audio
25 displays", en Proceedings of DAFX-04, 2004.

REIVINDICACIONES

1. 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 (108; 208; 308; 508), que
5 comprende:
- un aparato (100; 200; 300; 510) para generar una señal de ambiente (132; 232, 242, 252, 254; 332, 336, 342; 512) a partir de la señal de audio (108; 208; 308; 508),
en el que el aparato (100; 200; 300; 510) para generar la señal de ambiente (132; 232, 242, 252, 254; 332, 336, 342;
10 512) a partir de la señal de audio (108; 208; 308; 508) comprende:
- medios (110; 220; 320) para una compresión con pérdidas de una primera representación (108; 212; 316; $\mathbf{X}(\omega, k)$) de la señal de audio con el fin de obtener una representación comprimida (112; 222; 322; $\hat{\mathbf{X}}(\omega, k)$) de la señal de audio;
y
- 15 medios (120; 230; 330) para calcular una diferencia entre la representación comprimida (112; 222; 322; $\hat{\mathbf{X}}(\omega, k)$) de la señal de audio y la primera representación (108; 212; 316; $\mathbf{X}(\omega, k)$) de la señal de audio con el fin de obtener una representación de discriminación (122; 232; 332, 336; $|\mathbf{A}(\omega, k)|$), describiendo la representación de discriminación la diferencia entre la primera representación de la señal de audio y la representación comprimida de la señal de audio, y
- 20 describiendo la representación de discriminación las porciones de la señal de audio no incorporadas en la representación comprimida con pérdidas o eliminadas en la representación comprimida con pérdidas, y en el que los medios para la compresión con pérdidas se configuran de tal forma que se prefiere que las porciones de señal que muestran una distribución regular de la energía o que llevan una gran energía de señal se incluyan en la representación comprimida, mientras que las porciones de señal que tienen una energía distribuida de forma
25 irregular o una energía inferior se transfieren a la representación comprimida de forma atenuada o sólo en escasa medida;
en el que la representación de discriminación forma la señal de ambiente;
un aparato (550, 560) para proporcionar la señal de audio (108; 208; 308; 508) o una señal obtenida a partir del mismo como la señal de altavoz frontal (562, 564, 566); y
- 30 un aparato que proporciona la señal del altavoz posterior (520, 530, 540) para proporcionar la señal de ambiente (132; 232, 242, 252, 254; 332, 336, 342; 512) proporcionada por el aparato (510) para generar la señal de ambiente (132; 232, 242, 252, 254; 332, 336, 342; 512) o una señal obtenida a partir del mismo como la señal de altavoz posterior (542, 544).
- 35 2. 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 (108; 208; 308; 508), que comprende:
- un aparato (100; 200; 300; 510) para generar una señal de ambiente (132; 232, 242, 252, 254; 332, 336, 342; 512) a partir de la señal de audio (108; 208; 308; 508),
40 en el que el aparato (100; 200; 300; 510) para generar la señal de ambiente (132; 232, 242, 252, 254; 332, 336, 342; 512) a partir de la señal de audio (108; 208; 308; 508) comprende:
- 45 medios (110; 220; 320) para una compresión con pérdidas de una primera representación (108; 212; 316; $\mathbf{X}(\omega, k)$) de la señal de audio con el fin de obtener una representación comprimida (112; 222; 322; $\hat{\mathbf{X}}(\omega, k)$) de la señal de audio,
- medios (120; 230; 330) para calcular una diferencia entre la representación comprimida (112; 222; 322; $\hat{\mathbf{X}}(\omega, k)$) de la señal de audio y la primera representación (108; 212; 316; $\mathbf{X}(\omega, k)$) de la señal de audio con el fin de obtener una
50 representación de discriminación (122; 232; 332, 336; $|\mathbf{A}(\omega, k)|$),
describiendo la representación de discriminación la diferencia entre la primera representación de la señal de audio y la representación comprimida de la señal de audio, y
- 55 describiendo la representación de discriminación las porciones de la señal de audio no incorporadas en la representación comprimida con pérdidas o eliminadas en la representación comprimida con pérdidas, y

medios (130, 240, 340) para proporcionar la señal de ambiente (132; 242, 252, 254; 336, 342; $a(t)$, $a[n]$) usando la representación de discriminación,

5 en el que los medios para la compresión con pérdidas se configuran de tal forma que se prefiere que las porciones de señal que muestran una distribución regular de la energía o que llevan una gran energía de señal se incluyan en la representación comprimida, mientras que las porciones de señal que tienen una energía distribuida de forma irregular o una energía inferior se transfieren a la representación comprimida de forma atenuada o sólo en escasa medida;

10 en el que el aparato (510) para generar la señal de ambiente (132; 230, 242, 252, 254; 332, 336, 342; 512) se configura para recibir la señal de audio (108; 208; 308; 508);

un aparato (550, 560) para proporcionar la señal de audio (108; 208; 308; 508) o una señal obtenida a partir del mismo como la señal de altavoz frontal (562, 564, 566); y

15

un aparato que proporciona la señal del altavoz posterior (520, 530, 540) para proporcionar la señal de ambiente (132; 230, 242, 252, 254; 332, 336, 342; 512) proporcionada por el aparato (510) para generar la señal de ambiente (132; 230, 242, 252, 254; 332, 336, 342; 512) o una señal obtenida a partir del mismo como la señal de altavoz posterior (542, 544).

20

3. 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:

25 generar una señal de ambiente a partir de la señal de audio, en el que la generación de la señal de ambiente de la señal de audio comprende

la compresión con pérdidas (610) de una primera representación (108; 212; 316) de la señal de audio con el fin de obtener una representación comprimida (112; 222; 322) de la señal de audio; y

30 calcular (620) una diferencia (122; 232; 332, 336) entre la representación comprimida de la señal de audio y la primera representación de la señal de audio con el fin de obtener una representación de discriminación (122; 232; 332, 336) que forme la señal de ambiente,

35 en el que la representación de discriminación describe la diferencia entre la primera representación de la señal de audio y la representación comprimida de la señal de audio, y

en el que la representación de discriminación describe las porciones de la señal de audio no incorporadas en la representación comprimida con pérdidas o eliminadas en la representación comprimida con pérdidas, y

40 en el que la compresión con pérdidas se realiza de tal forma que se prefiere que las porciones de señal que muestran una distribución regular de la energía o que llevan una gran energía de señal se incluyan en la representación comprimida, mientras que las porciones de señal que tienen una energía distribuida de forma irregular o una energía inferior se transfieren a la representación comprimida de forma atenuada o sólo en escasa medida;

45

en el que la representación de discriminación forma la señal de ambiente;

proporcionar la señal de audio o una señal obtenida a partir del mismo como la señal de altavoz frontal; y

50 proporcionar la señal de ambiente o una señal obtenida a partir del mismo como la señal de altavoz posterior.

4. 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:

55 generar una señal de ambiente a partir de la señal de audio, en el que la generación de la señal de ambiente de la señal de audio comprende

la compresión con pérdidas (610) de una primera representación (108; 212; 316) de la señal de audio con el fin de obtener una representación comprimida (112; 222; 322) de la señal de audio;

calcular (620) una diferencia (122; 232; 332, 336) entre la representación comprimida de la señal de audio y la primera representación de la señal de audio con el fin de obtener una representación de discriminación (122; 232; 332, 336), y

5

proporcionar (630) la señal de ambiente usando la representación de discriminación,

en el que la representación de discriminación describe la diferencia entre la primera representación de la señal de audio y la representación comprimida de la señal de audio, y

10

en el que la representación de discriminación describe las porciones de la señal de audio no incorporadas en la representación comprimida con pérdidas o eliminadas en la representación comprimida con pérdidas, y

15 en el que la compresión con pérdidas se realiza de tal forma que se prefiere que las porciones de señal que muestran una distribución regular de la energía o que llevan una gran energía de señal se incluyan en la representación comprimida, mientras que las porciones de señal que tienen una energía distribuida de forma irregular o una energía inferior se transfieren a la representación comprimida de forma atenuada o sólo en escasa medida;

20 proporcionar la señal de audio o una señal obtenida a partir del mismo como la señal de altavoz frontal; y

proporcionar la señal de ambiente o una señal obtenida a partir del mismo como la señal de altavoz posterior.

5. Programa de ordenador para realizar el procedimiento de acuerdo con las reivindicaciones 3 ó 4
25 cuando el programa de ordenador se ejecuta en un ordenador.

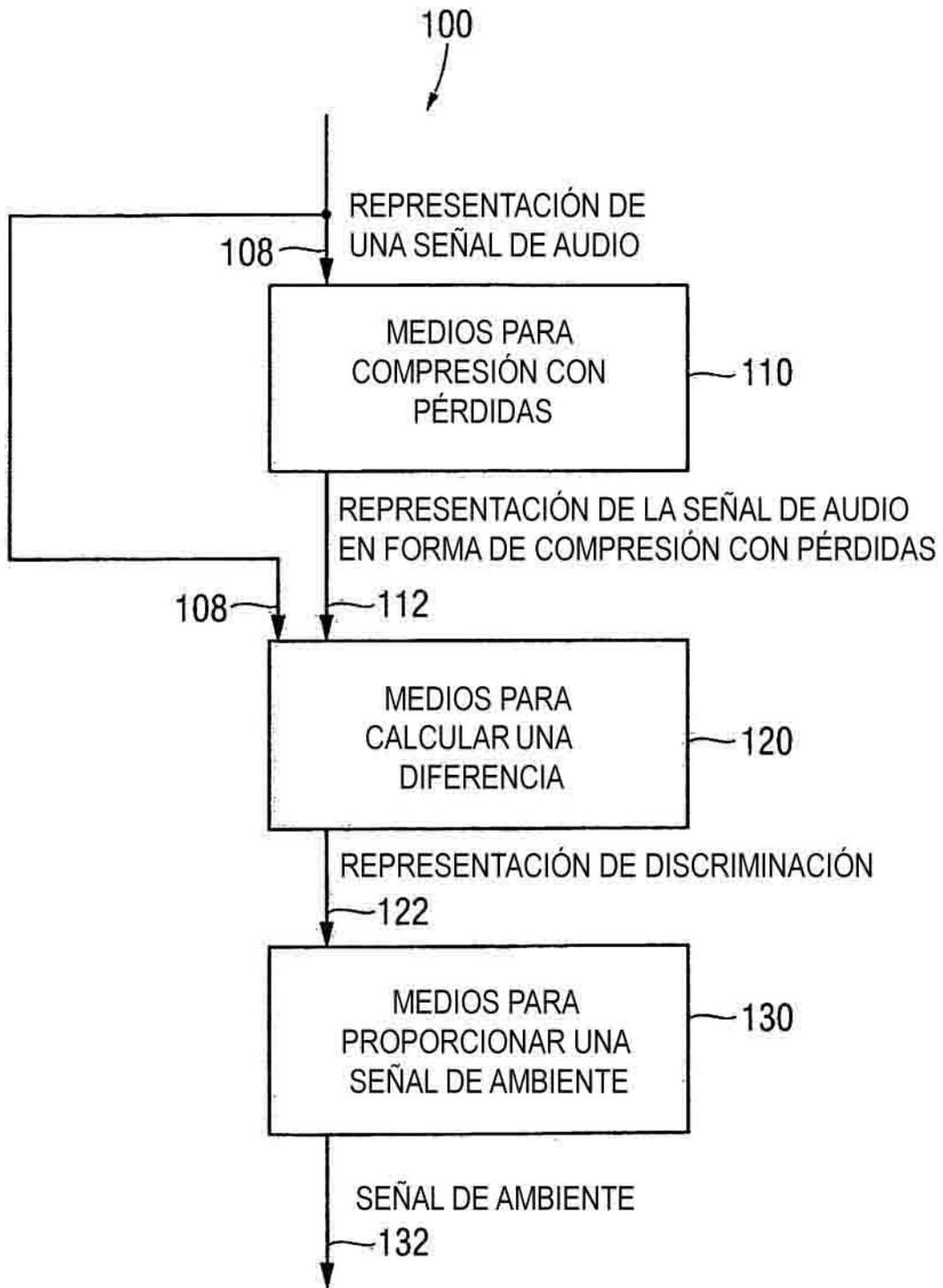


FIGURA 1

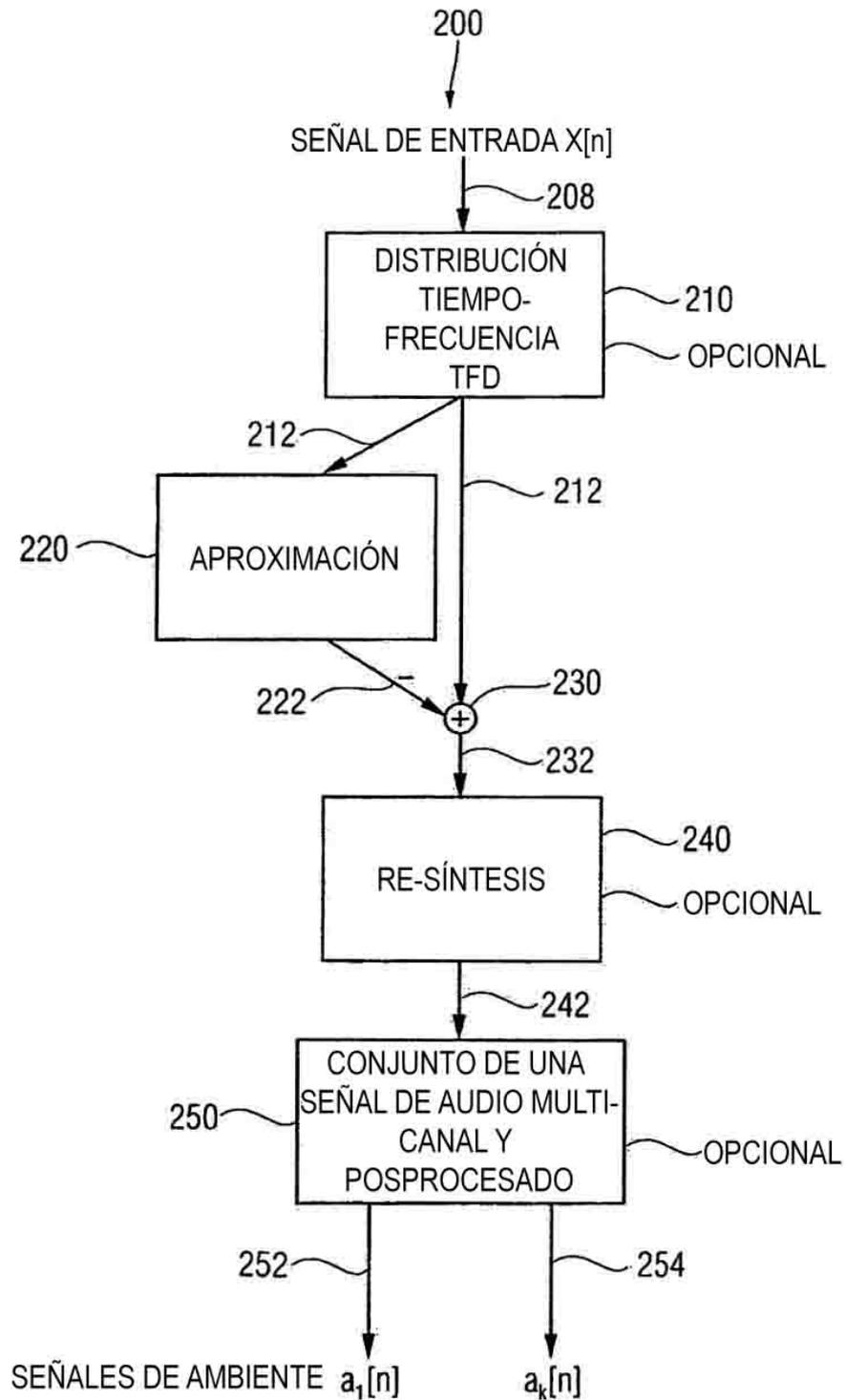


FIGURA 2

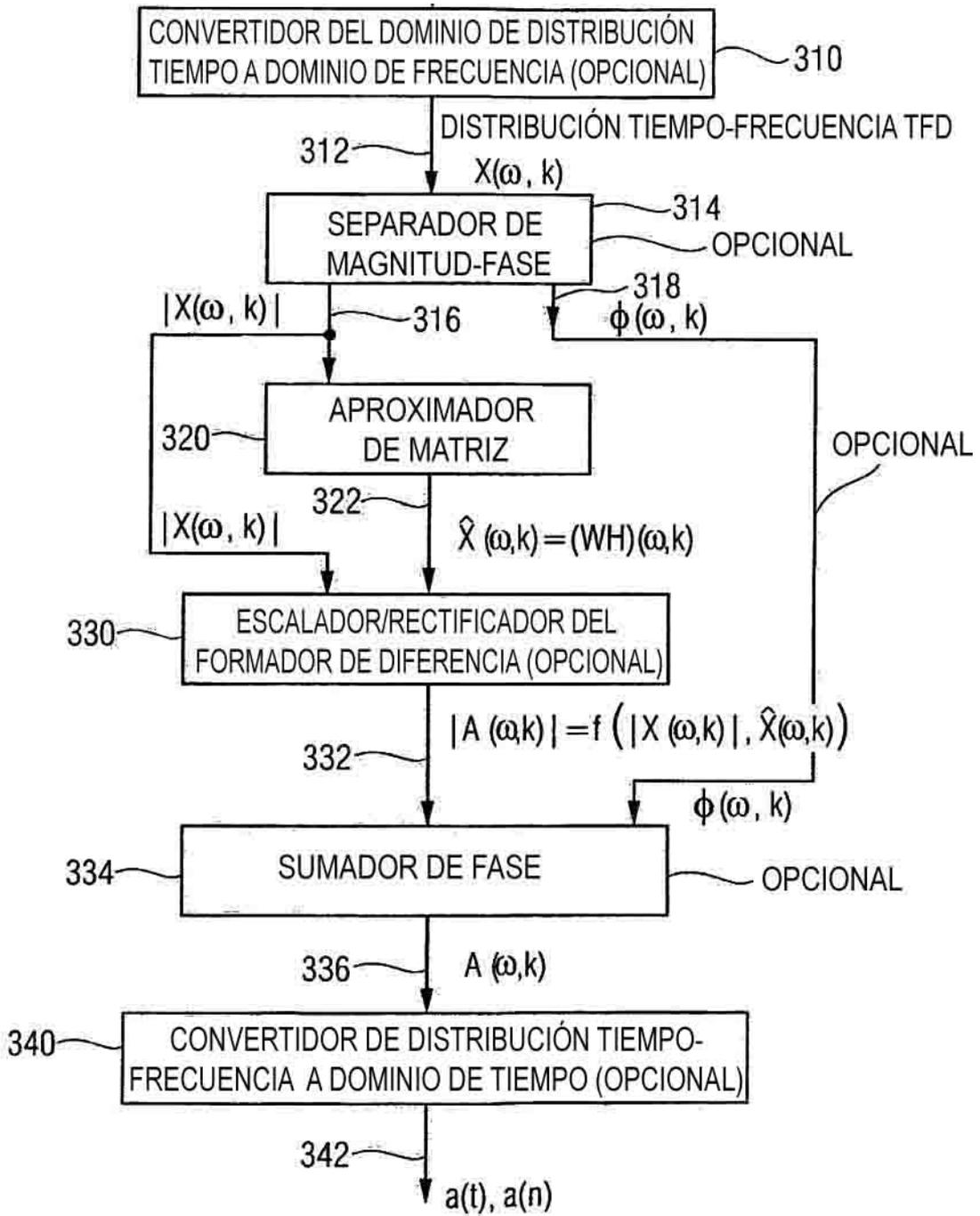


FIGURA 3

$$\begin{pmatrix} 300 \\ 020 \\ 001 \end{pmatrix} \approx \begin{pmatrix} 30 \\ 02 \\ 00 \end{pmatrix} \begin{pmatrix} 100 \\ 010 \end{pmatrix} = \begin{pmatrix} 3 \\ 0 \\ 0 \end{pmatrix} (100) + \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix} (010)$$

FIGURA 4A

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

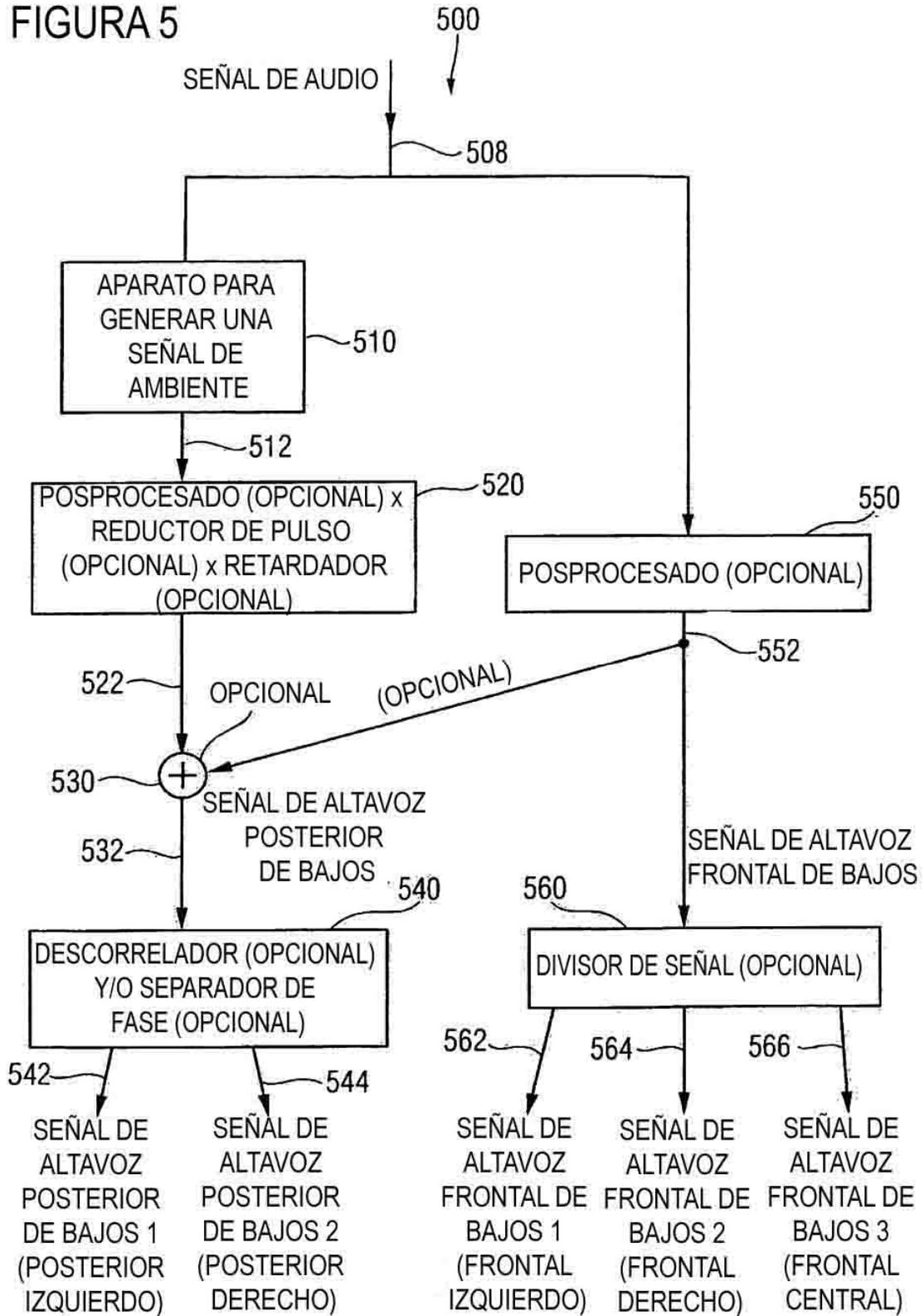
$x(w, k_1)$

	k_1	k_2	k_3	k_4	...	k_l	k_{l+1}	k_{l+2}
w_1	x	$x(w_1, k_2)$	x	x	...	x	$x(w_1, k_{l+1})$	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	$x(w_n, k_{l+1})$	x

$x(w, k_2)$

FIGURA 4B

FIGURA 5



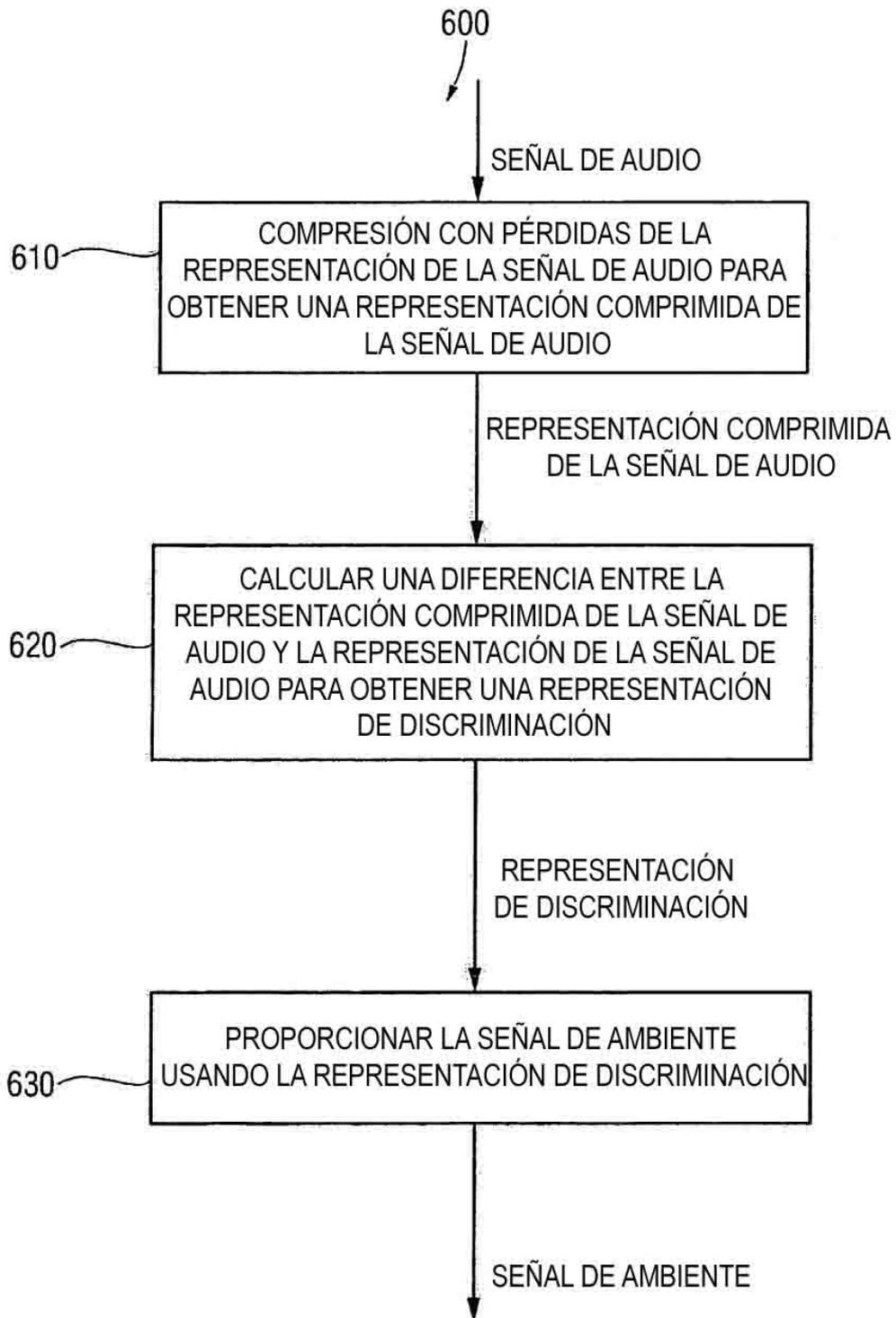


FIGURA 6



FIGURA 7



FIGURA 8



FIGURA 9