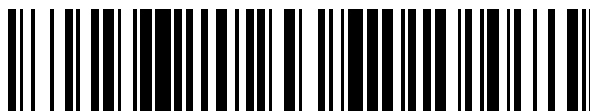


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 649 739**

51 Int. Cl.:

G10L 19/008

(2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **05.08.2013 PCT/EP2013/066405**

87 Fecha y número de publicación internacional: **06.02.2014 WO14020182**

96 Fecha de presentación y número de la solicitud europea: **05.08.2013 E 13759676 (3)**

97 Fecha y número de publicación de la concesión europea: **13.09.2017 EP 2880654**

54 Título: **Procedimiento y decodificador para un concepto paramétrico de codificación de objetos de audio espacial generalizado para casos de mezcla descendente/mezcla ascendente de multicanal**

30 Prioridad:

03.08.2012 US 201261679404 P

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

15.01.2018

73 Titular/es:

**FRAUNHOFER-GESELLSCHAFT ZUR
FÖRDERUNG DER ANGEWANDTEN
FORSCHUNG E.V. (100.0%)
Hansastraße 27c
80636 Munich, DE**

72 Inventor/es:

**KASTNER, THORSTEN;
HERRE, JÜRGEN;
TERENTIV, LEON y
HELLMUTH, OLIVER**

74 Agente/Representante:

SALVA FERRER, Joan

ES 2 649 739 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

Procedimiento y decodificador para un concepto paramétrico de codificación de objetos de audio espacial generalizado para casos de mezcla descendente/mezcla ascendente de multicanal

[0001] La presente invención se refiere a un aparato y un procedimiento para un concepto paramétrico de codificación de objetos de audio espacial generalizado para casos de mezcla descendente/mezcla ascendente de multicanal.

[0002] En los sistemas de audio digital modernos, es una tendencia mayor permitir modificaciones relacionadas con el objeto de audio del contenido transmitido en el lado del receptor. Estas modificaciones incluyen modificaciones de ganancia de partes seleccionadas de la señal de audio y/o re-posicionamiento espacial de objetos de audio dedicados en caso de reproducción de multicanal vía altavoces distribuidos espacialmente. Esto se puede obtener al alimentar individualmente diferentes partes del contenido de audio a los diferentes altavoces

[0003] En otras palabras, en la técnica de procesamiento de audio, transmisión de audio y el almacenamiento de audio existe un creciente deseo de permitir la interacción del usuario sobre la reproducción del contenido de audio orientado al objeto y también una demanda de utilizar las posibilidades extendidas de reproducción de multicanal para renderizar individualmente contenido de audio o parte del mismo con el fin de mejorar la impresión de audición. Mediante esto, el uso del contenido de audio de multi-canal trae consigo mejoras significativas para el usuario. Por ejemplo, se puede obtener una impresión de audición tridimensional, lo que trae consigo una satisfacción del usuario mejorada en aplicaciones de entretenimiento. Sin embargo, el contenido de audio de multi-canal es también útil en entornos profesionales, por ejemplo, en aplicaciones de conferencia telefónica, debido a que la inteligibilidad del hablante puede ser mejorada mediante el uso de reproducción de audio de multi-canal. Otra aplicación posible es ofrecer al oyente una pieza musical con la opción de ajustar individualmente el nivel de reproducción y/o posición espacial de diferentes partes (también denominados como "objetos de audio") o pistas, tales como una parte vocal o diferentes instrumentos. El usuario puede efectuar tal ajuste por razones de gusto personal, para facilitar la transcripción de una o más partes componentes de la pieza musical, con fines educativos, karaoke, ensayo, etc.

[0004] La transmisión discreta directa de todo el contenido de multicanal digital o contenido de audio de multi-objetos, por ejemplo, en forma de datos de modulación por código de impulsos (PCM) o incluso formatos de audio comprimido demanda velocidades de bits muy altas. Sin embargo, también es deseable transmitir y almacenar datos de audio de manera eficiente de velocidad de bits. Por consiguiente, es deseable aceptar una solución intermedia razonable entre la calidad del audio y requerimientos de velocidad de bits con el fin de evitar una carga excesiva de recursos provocada por aplicaciones de multi-canal/multi-objeto.

[0005] Recientemente, en el campo de codificación de audio, se han introducido técnicas paramétricas para la transmisión/almacenamiento eficiente en velocidad de bits de las señales de audio de multi-canal/multi-objeto mediante, por ejemplo, el grupo de expertos en películas (MPEG) y otros. Un ejemplo es MPEG Surround (MPS) como una estrategia orientada al canal [MPS, BCC] o codificación de objetos de audio espacial (SAOC) de MPEG, como una estrategia orientada al objeto [JSC, SAOC, SAOC1, SAOC2]. Otra estrategia orientada al objeto es denominado como "separación de fuentes informada" [ISS1, ISS2, ISS3, ISS4, ISS5, ISS6]. Estas técnicas tienen como objetivo la reconstrucción de una escena de audio de salida deseada o un objeto fuente de audio deseada en base a una mezcla descendente de canales/objetos e información lateral adicional que describe la escena de audio transmitida/almacenada y/o los objetos fuente de audio en la escena de audio.

[0006] La estimación y la aplicación de información lateral relacionada con canal/objetos en tales sistemas se hace de manera selectiva de tiempo-frecuencia. Por consiguiente, tales sistemas emplean transformadas de tiempo-frecuencia, tales como la Transformada Discreta de Fourier (DFT), la transformada de Fourier de tiempo corto (STFT) o bancos de filtros, tales como bancos de filtros de espejo en cuadratura (QMF), etc. El principio básico de tales sistemas es ilustrado en la figura 2, utilizando el ejemplo de SAOC de MPEG.

[0007] En el caso de STFT, la dimensión temporal es representada por el número de tiempo-bloque y la dimensión espectral es capturada por el número de coeficiente espectral ("bin"). En el caso de QMF, la dimensión temporal es representada por el número de segmento de tiempo y la dimensión espectral es capturada por el número de sub-banda. Si la resolución espectral del QMF es mejorada mediante la aplicación subsecuente de una segunda etapa de filtro, todo el banco de filtros es denominado QMF híbrido y las sub-bandas de resolución fina son denominadas sub-bandas híbridas.

[0008] Como ya se mencionó anteriormente, en SAOC el procesamiento general se lleva a cabo de manera tiempo-frecuencia selectiva y puede ser descrito como sigue, dentro de cada banda de frecuencia, como se ilustra en la figura 2:

- N señales de objeto de audio de entrada $S_1 \dots S_N$ son mezclados descendientemente a P canales $X_1 \dots X_P$ como parte del procesamiento del codificador, usando una matriz de mezcla descendente que consiste de los elementos $d_{1,1} \dots d_{N,P}$. Además, el codificador extrae información lateral que describe las características de los objetos de audio de entrada (módulo estimador de información lateral (SIE)). Para SAOC de MPEG, las relaciones de las potencias de objetos w.r.t. entre sí son la forma más básica de tal información lateral.

- La(s) señal(es) de mezcla descendente e información lateral son transmitidas/almacenadas. Para este fin, la(s) señal(es) de audio de mezcla descendente puede(n) ser comprimida(s), por ejemplo usando codificadores de audio perceptual bien conocidos, tales como MPEG-1/2 Capa II o III (conocido como mp3), codificación de audio avanzada MPEG-2/4 (AAC), etc.

- En el extremo receptor, el descodificador trata de restaurar conceptualmente las señales del objeto original ("separación del objeto") de las señales de mezcla descendente (descodificadas) usando la información lateral transmitida. Estas señales de objeto aproximadas $\hat{S}_1 \dots \hat{S}_N$ son luego mezcladas a una escena objetivo representada por M canales de salida de audio $\hat{Y}_1 \dots \hat{Y}_M$ usando una matriz de renderización descrita por los coeficientes $r_{1,1} \dots r_{N,M}$ en la figura 2. La escena objetivo deseada puede ser, en el caso extremo, la renderización de solamente una señal fuente de la mezcla (escenario de separación de fuente), pero también cualquier otra escena acústica arbitraria que consiste en los objetos transmitidos. Por ejemplo, la salida puede ser una escena objetivo de un solo canal, una escena objetivo estéreo de 2 canales o una escena objetivo de multicanal 5.1.

[0009] El incremento del ancho de banda/almacenamiento disponible y mejoras en marcha en el campo de codificación de audio permiten al usuario seleccionar una elección incrementada uniformemente de producciones de audio de multicanal. Los formatos de audio 5.1 de multicanal ya son estándar en producciones de DVD y Blue-Ray. Los nuevos formatos de audio como audio 3D de MPEG-H con aún más canales de transporte de audio aparecen en el horizonte, que proveerán a los usuarios finales una experiencia de audio altamente inmersiva.

[0010] Los esquemas de codificación de objeto de audio paramétricos están limitados actualmente a un máximo de dos canales de mezcla descendente. Solamente pueden ser aplicados a alguna extensión en mezclas de multicanal, por ejemplo en solamente dos canales de mezcla descendente seleccionados. La flexibilidad que estos esquemas de codificación ofrecen al usuario para ajustar la escena de audio a sus propias preferencias es así severamente limitada, por ejemplo, con respecto al cambio de nivel de audio del comentarista deportivo y la atmósfera en difusión de deportes.

[0011] Además, los esquemas de codificación de objeto de audio actuales ofrecen solamente una variabilidad limitada en el procedimiento de mezcla en el lado del codificador. El procedimiento de mezcla es limitado a la mezcla variable en el tiempo de los objetos de audio y no es posible la mezcla variable en frecuencia.

[0012] Por consiguiente, sería altamente apreciado si se proveen conceptos mejorados para codificación de objetos de audio.

[0013] El objeto de la presente invención es proveer conceptos mejorados para la codificación de objeto de audio. El objeto de la presente invención es resuelto por el descodificador según la reivindicación 1, por el procedimiento según la reivindicación 10 y por el programa informático según la reivindicación 11. Se provee un descodificador para generar una señal de salida de audio que comprende uno o más canales de salida de audio de una señal de mezcla descendente que comprende uno o más canales de mezcla descendente.

[0014] La señal de mezcla descendente codifica una o más señales de objeto de audio. El descodificador comprende un determinador de umbral para determinar un valor de umbral dependiendo de la energía de señal y/o una energía de ruido de al menos una de las señales de objeto de audio o más y/o dependiendo de la energía de señal y/o energía de ruido de al menos uno del uno o más canales de mezcla descendente. Además, el descodificador comprende una unidad de procesamiento para generar el uno o más canales de salida de audio a partir de uno o más canales de mezcla descendente dependiendo del valor de umbral.

[0015] Según una realización, la señal de mezcla descendente puede comprender dos o más canales de mezcla descendente y el determinador de umbral puede ser configurado para determinar el valor de umbral dependiendo de la energía de ruido de cada uno de los dos o más canales de mezcla descendente.

[0016] En una realización, el determinador de umbral puede ser configurado para determinar el valor de umbral dependiendo de la suma de toda la energía de ruido en los dos o más canales de mezcla descendente.

[0017] Según una realización, la señal de mezcla descendente puede codificar dos o más señales de objeto de audio y el determinador de umbral puede ser configurado para determinar el valor de umbral dependiendo de la energía de señal de la señal de objeto de audio de las dos o más señales de objeto de audio que tiene la mayor

energía de señal de las dos o más señales de objeto de audio.

[0018] En una realización, la señal de mezcla descendente puede comprender dos o más canales de mezcla descendente y el determinador de umbral puede ser configurado para determinar el valor de umbral dependiendo de la suma de toda la energía de ruido en los dos o más canales de mezcla descendente.

[0019] Según una realización, la señal de mezcla descendente puede codificar una o más señales de objeto de audio para cada mosaico de tiempo-frecuencia de una pluralidad de mosaicos de tiempo-frecuencia. El determinador de umbral puede ser configurado para determinar el valor de umbral para cada mosaico de tiempo-frecuencia de la pluralidad de mosaicos de tiempo-frecuencia dependiendo de la energía de señal o de la energía de ruido de al menos una de las señales de objeto de audio o más o dependiendo de la energía de señal o de la energía de ruido de al menos uno del uno o más canales de mezcla descendente, en el que un primer valor de umbral de un primer mosaico de tiempo-frecuencia de la pluralidad de mosaicos de tiempo-frecuencia puede diferir de un segundo mosaico de tiempo-frecuencia de la pluralidad de mosaicos de tiempo-frecuencia. La unidad de procesamiento puede ser configurada para generar para cada mosaico de tiempo-frecuencia de la pluralidad de mosaicos de tiempo-frecuencia un valor de canal de cada uno de uno o más canales de salida de audio de uno o más canales de mezcla descendente dependiendo del valor de umbral del mosaico de tiempo-frecuencia.

[0020] En una realización, el descodificador puede estar configurado para determinar el valor de umbral T en decibelios, según la fórmula

$$T [dB] = E_{\text{ruido}} [dB] - E_{\text{ref}} [dB] - Z$$

o según la fórmula

$$T [dB] = E_{\text{ruido}} [dB] - E_{\text{ref}} [dB],$$

en la que $T [dB]$ indica el valor de umbral en decibelios, en la que $E_{\text{ruido}} [dB]$ indica la suma de toda la energía de ruido en los dos o más canales de mezcla descendente en decibelios, en la que $E_{\text{ref}} [dB]$ indica la energía de señal de una de las señales de objeto en decibelios y en la que Z indica un parámetro adicional que es un número. En una realización alternativa, $E_{\text{ruido}} [dB]$ indica la suma de toda la energía de ruido en los dos o más canales de mezcla descendente en decibelios dividida por el número de los canales de mezcla descendente.

[0021] Según una realización, el descodificador puede estar configurado para determinar el valor de umbral T según la fórmula

$$T = \frac{E_{\text{ruido}}}{E_{\text{ref}} \cdot Z} \quad \text{o según la fórmula}$$

$$T = \frac{E_{\text{ruido}}}{E_{\text{ref}}},$$

en la que T indica el valor de umbral, en la que E_{ruido} indica la suma de toda la energía de ruido en los dos o más canales de mezcla descendente, en la que E_{ref} indica la energía de señal de una de las señales de objeto de audio y en la que Z indica un parámetro adicional que es un número. En una realización alternativa, $E_{\text{ruido}} [dB]$ indica la suma de toda la energía de ruido en los dos o más canales de mezcla descendente dividida por el número de los canales de mezcla descendente.

[0022] Según una realización, la unidad de procesamiento puede estar configurada para generar uno o más canales de salida de audio de uno o más canales de mezcla descendente dependiendo de una matriz de covarianza de objeto (E) de una o más señales de objeto de audio, dependiendo de una matriz de mezcla descendente (D) para la mezcla descendente de las dos o más señales de objeto de audio para obtener los dos o más canales de mezcla descendente y dependiendo del valor de umbral.

[0023] En una realización, la unidad de procesamiento está configurada para generar uno o más canales de salida de audio de uno o más canales de mezcla descendente mediante la aplicación del valor de umbral en una función para invertir una matriz de correlación cruzada de canal de mezcla descendente Q , en el que Q es definido como $Q = DED^*$, en el que D es la matriz de mezcla descendente para la mezcla descendente de las dos o más señales de objeto de audio para obtener los dos o más canales de mezcla descendente y en el que E es la matriz de

covarianza de objeto de una o más señales de objeto de audio.

[0024] Por ejemplo, la unidad de procesamiento puede estar configurada para generar uno o más canales de salida de audio de uno o más canales de mezcla descendente mediante el cálculo de los valores propios de la matriz de correlación cruzada de canal de mezcla descendente $\mathbf{Q}$ o al calcular los valores singulares de la matriz de correlación cruzada de canal de mezcla descendente $\mathbf{Q}$.

[0025] Por ejemplo, la unidad de procesamiento puede estar configurada para generar uno o más canales de salida de audio de uno o más canales de mezcla descendente al multiplicar el valor propio más grande de los valores propios de la matriz de correlación cruzada de canal de mezcla descendente $\mathbf{Q}$ con el valor de umbral para obtener un umbral relativo.

[0026] Por ejemplo, la unidad de procesamiento puede estar configurada para generar uno o más canales de salida de audio de uno o más canales de mezcla descendente al generar una matriz modificada. La unidad de procesamiento puede estar configurada para generar la matriz modificada dependiendo solamente de los vectores propios de la matriz de correlación cruzada de canal de mezcla descendente $\mathbf{Q}$, que tienen un valor propio de los valores propios de la matriz de correlación cruzada de canal de mezcla descendente $\mathbf{Q}$, que es mayor o igual al umbral modificado. Además, la unidad de procesamiento puede estar configurada para llevar a cabo una inversión de matriz de la matriz modificada para obtener una matriz invertida. Además, la unidad de procesamiento puede estar configurada para aplicar la matriz invertida sobre uno o más de los canales de mezcla descendente para generar uno o más canales de salida de audio.

[0027] Además, se provee un procedimiento para generar una señal de salida de audio que comprende uno o más canales de salida de audio de una señal de mezcla descendente que comprende uno o más canales de mezcla descendente. La señal de mezcla descendente codifica una o más señales de objeto de audio. El descodificador comprende:

- Determinar un valor de umbral dependiendo de la energía de señal o la energía de ruido de al menos una de las señales de objeto de audio o dependiendo de una energía de señal o energía de ruido de al menos uno de uno o más canales de mezcla descendente y;

- Generar uno o más canales de salida de audio a partir de uno o más canales de mezcla descendente dependiendo del valor de umbral.

[0028] Además, se provee un programa informático para implementar el procedimiento descrito anteriormente cuando es ejecutado en un ordenador o procesador de señales.

[0029] El documento "Corrections of the parameter processor for MPEG SAOC" (ISO/IEC JTC1/SC29/WG11, enero de 2011 de Engdegård y otros) señala los problemas que los umbrales de valores propios particulares podrían plantear para ciertos tipos de señal, pero propone modificar el umbral fijo en lugar de introducir un umbral variable. En lo siguiente, se describen en más detalle realizaciones de la presente invención con referencia a las figuras, en las cuales:

La figura 1 ilustra un descodificador para generar una señal de salida de audio que comprende uno o más canales de salida de audio según una realización,

La figura 2 es una vista general del sistema SAOC que ilustra el principio de tales sistemas utilizando el ejemplo de SAOC de MPEG,

La figura 3 ilustra una vista general del concepto de mezcla ascendente paramétrico de G-SAOC y

La figura 4 ilustra un concepto de mezcla descendente/mezcla ascendente general.

[0030] Antes de describir las realizaciones de la presente invención, se proveen más antecedentes con respecto a los sistemas de SAOC del estado de la técnica.

[0031] La figura 2 muestra una disposición general de un codificador de SAOC 10 y un descodificador de SAOC 12. El codificador de SAOC 10 recibe como entrada N objetos, es decir, las señales de audio s_1 a s_N . En particular, el codificador 10 comprende un elemento de mezcla descendente 16 que recibe las señales de audio s_1 a s_N y mezcla descendentemente las mismas a una señal de mezcla descendente 18. Alternativamente, el elemento de mezcla descendente puede ser provisto externamente ("elemento de mezcla descendente artístico") y el sistema estima información lateral adicional para hacer que la mezcla descendente provista coincida con la mezcla descendente calculada. En la figura 2, se muestra que la señal de mezcla descendente es una señal de canal P . Así,

es concebible cualquier configuración de señal de mezcla descendente mono ($P = 1$), estéreo ($P = 2$) o multi-canal ($P > 2$).

[0032] En caso de una mezcla descendente estéreo, los canales de la señal de mezcla descendente 18 son denotados $L0$ y $R0$, en caso de una mezcla descendente mono, el mismo es denotado simplemente como $L0$. Con el fin de permitir que el descodificador de SAOC 12 recupere los objetos individuales S_1 a S_N , el estimador de información lateral 17 provee al descodificador de SAOC 12 con información lateral que incluye parámetros de SAOC. Por ejemplo, en caso de una mezcla descendente estéreo, los parámetros de SAOC comprenden diferencias a nivel de objeto (OLD), correlaciones de inter-objeto (IOC) (parámetros de correlación cruzada de inter-objeto), valores de ganancia de mezcla descendente (DMG) y diferencias de nivel de canal de mezcla descendente (DCLD). La información lateral 20, que incluye los parámetros de SAOC, junto con la señal de mezcla descendente 18, forma la corriente de datos de salida de SAOC recibida por el descodificador de SAOC 12.

[0033] El descodificador de SAOC 12 comprende un mezclador ascendente 18 que recibe la señal de mezcla descendente 18, también como la información lateral 20 con el fin de recuperar y renderizar las señales de audio $\hat{s}_1$ y $\hat{s}_N$ sobre cualquier conjunto de canales seleccionados por el usuario $\hat{y}_1$ a $\hat{y}_M$, siendo prescrita la renderización por la información de renderización 26 introducida al descodificador de SAOC 12.

[0034] Las señales de audio S_1 y S_N pueden ser introducidas al codificador 10 en cualquier dominio de codificación, tal como en el dominio de tiempo o dominio espectral. En caso de que las señales de audio S_1 y S_N sean introducidas al codificador 10 en el dominio del tiempo, tales como PCM codificadas, el codificador 10 puede utilizar un banco de filtros, tal como el banco QMF híbrido, con el fin de transferir las señales al dominio espectral, en el que las señales de audio son representadas en varias sub-bandas asociadas con diferentes porciones espectrales, a una resolución de banco de filtros específica. Si las señales de audio S_1 y S_N ya están en la representación esperada por el codificador 10, el mismo no tiene que realizar la descomposición espectral.

[0035] Más flexibilidad en el procedimiento de mezcla permite un aprovechamiento óptimo de las características de objeto de señal. Se puede producir una mezcla descendente que es optimizada para la separación paramétrica en el lado del descodificador respecto a la calidad percibida.

[0036] Las realizaciones extienden la parte paramétrica del esquema de SAOC a un número arbitrario de canales de mezcla descendente/mezcla ascendente. La siguiente figura provee una vista general del concepto de mezcla ascendente paramétrico de codificación de objetos de audio espacial generalizado (G-SAOC):

La figura 3 ilustra una vista general del concepto de mezcla ascendente paramétrico de G-SAOC. Se puede realizar una post-mezcla plenamente flexible (renderización) de los objetos de audio reconstruidos paraméricamente.

[0037] La figura 3 ilustra, entre otras cosas, un descodificador de audio 310, un separador de objetos 320 y un renderizador 330.

[0038] Considérese la siguiente notación común:

x señal de objeto de audio de entrada (de tamaño N_{obj})
y señal de audio de mezcla descendente (de tamaño N_{dmx})
z señal de escena de salida presentada (de tamaño $N_{mezcla\ ascendente}$)
D matriz de mezcla descendente (de tamaño $N_{obj} \times N_{dmx}$)
R matriz de renderización (de tamaño $N_{obj} \times N_{mezcla\ ascendente}$)
G matriz de mezcla ascendente paramétrica (de tamaño $N_{dmx} \times N_{mezcla\ ascendente}$)
E matriz de covarianza de objeto (de tamaño $N_{obj} \times N_{obj}$)

[0039] Todas las matrices introducidas son (en general) variantes en tiempo y frecuencia.

[0040] En lo siguiente, se provee la relación constitutiva para mezcla ascendente paramétrica.

[0041] En primer lugar, se proveen conceptos de mezcla descendente/mezcla ascendente con referencia a la figura 4. En particular, la figura 4 ilustra concepto de mezcla descendente/mezcla ascendente general, en el que la figura 4 ilustra sistemas modelados (izquierda) y de mezcla ascendente paramétricos (derecha).

[0042] Más en particular, la figura 4 ilustra una unidad de renderización 410, una unidad de mezcla descendente 421 y una unidad de mezcla ascendente paramétrica 422.

[0043] La señal de escena de salida renderizada ideal (modelada) z es definida como, véase la figura (izquierda):

$$R_x = z. \quad (1)$$

[0044] La señal de audio de mezcla descendente y es determinada como, véase la figura 4 (derecha):

$$D_x = y. \quad (2)$$

[0045] La relación constitutiva (aplicada a la señal de audio de mezcla descendente) para la reconstrucción de señal de escena de salida paramétrica puede ser representada como, véase la figura 4 (derecha):

$$G_y = z. \quad (3)$$

[0046] La matriz de mezcla ascendente paramétrica puede ser definida de (1) y (2) como la siguiente función de las matrices de mezcla descendente y renderización $G = G(D, R)$:

$$G = RED, (DED^*)^{-1} \quad (4)$$

[0047] En lo siguiente, se considera la mejora de la estabilidad de la estimación de fuente paramétrica según las realizaciones.

[0048] El esquema de separación paramétrica en SAOC DE MPEG está basado en una estimación mínima cuadrática media (LMS) de las fuentes en la mezcla. La estimación de LMS involucra la inversión de la matriz de covarianza de canal de mezcla descendente descrita paramétricamente $Q = DED^*$. Los algoritmos para la inversión de matriz son en general sensibles a las matrices mal acondicionadas. La inversión de tal matriz puede causar sonidos no naturales, llamados artefactos, en la escena de salida renderizada. Un umbral fijo determinado heurísticamente T en SAOC de MPEG evita actualmente esto. Aunque los artefactos son evitados mediante este procedimiento, no se puede lograr un desempeño de separación posible suficiente en el lado del descodificador.

[0049] La figura 1 ilustra un descodificador para generar una señal de salida de audio que comprende uno o más canales de salida de audio de una señal de mezcla descendente que comprende uno o más canales de mezcla descendente según una realización. La señal de mezcla descendente codifica una o más señales de objeto de audio.

[0050] El descodificador comprende un determinador de umbral 110 para determinar un valor de umbral dependiendo de la energía de señal y/o energía de ruido de al menos una de las señales de objeto de audio y/o dependiendo de la energía de señal y/o energía de ruido de al menos uno de los canales de mezcla descendente.

[0051] Además, el descodificador comprende una unidad de procesamiento 120 para generar uno o más canales de salida de audio de uno o más canales de mezcla descendente dependiendo del valor de umbral.

[0052] En contraste con el estado de la técnica, el valor de umbral determinado por el determinador de umbral 110 depende de la energía de señal o la energía del ruido de uno o más canales de mezcla descendente o de una o más señales de objeto de audio codificadas. En realizaciones, a medida que las energías de señal y ruido de uno o más canales de mezcla descendente y/o de uno o más valores de señal de objeto de audio varían, así varía el valor de umbral, por ejemplo, de instancia de tiempo a instancia de tiempo o de mosaico de tiempo-frecuencia a mosaico de tiempo-frecuencia.

[0053] Las realizaciones proveen un procedimiento de umbral adaptable para que la inversión de matriz obtenga una separación paramétrica mejorada de los objetos de audio en el lado del descodificador. El desempeño de separación es en promedio mejor, pero nunca menor que el esquema de umbral fijo utilizado actualmente en SAOC de MPEG en el algoritmo para invertir la matriz Q .

[0054] El umbral T es adaptado dinámicamente a la precisión de los datos para cada mosaico de tiempo-frecuencia procesado. El desempeño de separación es así mejorado y se evitan los artefactos en la escena de salida presentada provocados por la inversión de matrices mal acondicionadas.

[0055] Según una realización, la señal de mezcla descendente puede codificar dos o más canales de mezcla descendente y el determinador de umbral 110 puede estar configurado para determinar el valor de umbral dependiendo de la energía de ruido de cada uno de los dos o más canales de mezcla descendente.

[0056] En una realización, el determinador de umbral 110 puede estar configurado para determinar el valor de umbral dependiendo de la suma de toda la energía de ruido en los dos o más canales de mezcla descendente.

[0057] Según una realización, la señal de mezcla descendente puede comprender dos o más señales de objeto de audio y el determinador de umbral 110 puede estar configurado para determinar el valor de umbral dependiendo de la energía de señal de la señal de objeto de audio de las dos o más señales de objeto de audio que tiene la mayor energía de señal de las dos o más señales de objeto de audio.

[0058] En una realización, la señal de mezcla descendente puede comprender dos o más canales de mezcla descendente y el determinador de umbral 110 puede estar configurado para determinar el valor de umbral dependiendo de la suma de toda la energía de ruido en los dos o más canales de mezcla descendente.

[0059] Según una realización, la señal de mezcla descendente puede codificar una o más señales de objeto de audio por cada mosaico de tiempo-frecuencia de una pluralidad de mosaicos de tiempo-frecuencia. El determinador de umbral 110 puede estar configurado para determinar un valor de umbral para cada mosaico de tiempo-frecuencia de la pluralidad de mosaicos de tiempo-frecuencia dependiendo de la energía de señal o la energía de ruido de al menos uno de las señales de objeto de audio o dependiendo de la energía de señal o energía de ruido de al menos uno del uno o más canales de mezcla descendente, en el que un primer valor de umbral de un primer mosaico de tiempo-frecuencia de la pluralidad de mosaicos de tiempo-frecuencia puede diferir de un segundo mosaico de tiempo-frecuencia de la pluralidad de mosaicos de tiempo-frecuencia. La unidad de procesamiento 120 puede estar configurada para generar para cada mosaico de tiempo-frecuencia de la pluralidad de mosaicos de tiempo-frecuencia un valor de canal de cada uno del uno o más canales de salida de audio del uno o más canales de mezcla descendente, dependiendo del valor de umbral del mosaico de tiempo-frecuencia.

[0060] Según una realización, el descodificador puede estar configurado para determinar el valor de umbral T según la fórmula

$$T = \frac{E_{\text{ruido}}}{E_{\text{ref}} \cdot Z} \quad \text{o según la fórmula}$$

$$T = \frac{E_{\text{ruido}}}{E_{\text{ref}}},$$

en la que T indica el valor de umbral, en la que E_{ruido} indica la suma de toda la energía de ruido en los dos o más canales de mezcla descendente, en la que E_{ref} indica la energía de señal de una de las señales de objeto de audio y en la que Z indica un parámetro adicional que es un número. En una realización alternativa, E_{ruido} indica la suma de toda la energía de ruido en los dos o más canales de mezcla descendente, dividido por el número de los canales de mezcla descendente.

[0061] En una realización, el descodificador puede estar configurado para determinar el valor de umbral T en decibelios, según la fórmula

$$T \text{ [dB]} = E_{\text{ruido}} \text{ [dB]} - E_{\text{ref}} \text{ [dB]} - Z$$

o según la fórmula

$$T \text{ [dB]} = E_{\text{ruido}} \text{ [dB]} - E_{\text{ref}} \text{ [dB]},$$

en la que $T \text{ [dB]}$ indica el valor de umbral en decibelios, en la que $E_{\text{ruido}} \text{ [dB]}$ indica la suma de toda la energía de ruido en los dos o más canales de mezcla descendente en decibelios, en la que $E_{\text{ref}} \text{ [dB]}$ indica la energía de señal de una de las señales de objeto de audio en decibelios y en la que Z indica un parámetro adicional que es un número. En una realización alternativa, $E_{\text{ruido}} \text{ [dB]}$ indica la suma de toda la energía de ruido en los dos o más canales de mezcla descendente en decibelios, dividido por el número de los canales de mezcla descendente.

[0062] En particular, una estimación aproximada del umbral se puede dar para cada mosaico de tiempo-frecuencia por:

$$T \text{ [dB]} = E_{\text{ruido}} \text{ [dB]} - E_{\text{ref}} \text{ [dB]} - Z. \quad (5)$$

[0063] E_{ruido} puede indicar el nivel de ruido de fondo, por ejemplo, la suma de toda la energía de ruido en los canales de mezcla descendente. El ruido de fondo puede ser definido por la resolución de los datos de audio, por ejemplo, un ruido de fondo provocado por la PCM-codificación de los canales. Otra posibilidad es tener en cuenta el

ruido de codificación si la mezcla descendente es comprimida. Para tal caso, el ruido de fondo provocado por el algoritmo de codificación puede ser agregado. En una realización alternativa, E_{ruido} [dB] indica la suma de toda la energía de ruido en los dos o más canales de mezcla descendente en decibelios dividida, por el número de los canales de mezcla descendente.

[0064] E_{ref} puede indicar una energía de señal de referencia. En la forma más simple, esta puede ser la energía del objeto de audio más fuerte:

$$E_{\text{ref}} = \max(E). \quad (6)$$

[0065] Z puede indicar un factor de penalización para hacer frente a los parámetros adicionales que afectan la resolución de separación, por ejemplo, la diferencia del número de canales de mezcla descendente y número de objetos fuente. El desempeño de separación disminuye con el número incrementado de objetos de audio. Además, los efectos de la cuantificación de la información lateral paramétrica en la separación también pueden ser incluidos.

[0066] En una realización, la unidad de procesamiento 120 está configurada para generar uno o más canales de salida de audio de uno o más canales de mezcla descendente dependiendo de la matriz de mezcla descendente D para la mezcla descendente de las dos o más señales de objeto de audio para obtener los dos o más canales de mezcla descendente y dependiendo del valor de umbral.

[0067] Según una realización, para generar uno o más canales de salida de audio de uno o más canales de mezcla descendente dependiendo del valor de umbral, la unidad de procesamiento 120 puede estar configurada para proceder como sigue:

El umbral (que puede ser referido como un "umbral de separación-resolución") es aplicado al lado del descodificador en la función de invertir la matriz de correlación cruzada de canal de mezcla descendente estimada paraméricamente Q .

[0068] Los valores singulares de Q o los valores propios de Q son calculados.

[0069] El valor propio más grande es tomado y multiplicado con el umbral T .

[0070] Todos excepto el valor propio más grande son comparados con este umbral relativo y omitidos si son más pequeños.

[0071] Luego se lleva a cabo la inversión de matriz en una matriz modificada, en la que la matriz modificada puede, por ejemplo, ser la matriz definida por el conjunto reducido de vectores. Cabe señalar que para el caso en que todos excepto el valor propio más alto son omitidos, el valor propio más alto debe ser establecido al nivel del ruido de fondo si el valor propio es inferior.

[0072] Por ejemplo, la unidad de procesamiento 120 puede estar configurada para generar uno o más canales de salida de audio a partir de uno o más canales de mezcla descendente al generar la matriz modificada. La matriz modificada puede ser generada dependiendo solamente de aquellos vectores propios de la matriz de correlación cruzada de canal de mezcla descendente Q , que tienen un valor propio de los valores propios de la matriz de correlación cruzada de canal de mezcla descendente Q , que es mayor o igual al umbral modificado. La unidad de procesamiento 120 puede estar configurada para llevar a cabo una inversión de matriz de la matriz modificada para obtener una matriz invertida. Luego, la unidad de procesamiento 120 puede estar configurada para aplicar la matriz invertida sobre uno o más de los canales de mezcla descendente para generar uno o más canales de salida de audio. Por ejemplo, la matriz invertida puede ser aplicada en uno o más de los canales de mezcla descendente de una de las maneras como la matriz invertida del producto de matriz DED^* es aplicado sobre los canales de mezcla descendente (véase, por ejemplo, [SAOC], véase, en particular, por ejemplo: ISO/IEC, "Tecnologías de audio MPEG - Parte 2: Codificación de objeto de audio espacial (SAOC)", ISO/IEC JTC 1/SC29/GT1 1 (MPEG) Norma Internacional 23003-2: 2010, en particular, véase, capítulo "Procesamiento de SAOC", más en particular, véase la sección "Modos de transcodificación" y subcapítulo "Modos de decodificación").

[0073] Los parámetros que pueden ser empleados para estimar el umbral T pueden ya sea ser determinados en el codificador y embebidos en la información lateral paramétrica o estimados directamente en el lado del descodificador.

[0074] Una versión simplificada del estimador de umbral puede ser usada en el lado del codificador para indicar posibles inestabilidades en la estimación fuente en el lado del descodificador. En su forma más simple, despreciando todos los términos de ruido, la norma de la matriz de mezcla descendente puede ser calculada, indicando que no se puede aprovechar el pleno potencial de los canales de mezcla descendente disponibles para

estimar paramétricamente las señales fuente en el lado del descodificador. Tal indicador puede ser usado durante el procedimiento de mezcla para evitar mezclar matrices que son críticas para estimar las señales fuente.

[0075] En cuanto a la parametrización de la matriz de covarianza de objeto, se puede ver que el procedimiento de mezcla ascendente paramétrico descrito en base a la relación constitutiva (4) es invariante al signo de las entidades fuera de la diagonal de la matriz de covarianza de objeto E . Esto da como resultado la posibilidad de una parametrización más eficiente (en comparación con SAOC) (de cuantificación y codificación) de los valores que representan correlaciones de inter-objeto.

[0076] Con respecto al transporte de información que representa la matriz de mezcla descendente, en general, las señales de entrada y de mezcla descendente de audio x , y , junto con la matriz de covarianza E son determinadas en el lado del codificador. La representación codificada de la señal de mezcla descendente de audio y y la información que describe la matriz de covarianza E son transmitidas al lado del descodificador (vía carga de corriente de bits). La matriz de renderización R es ajustada y está disponible en el lado del descodificador.

[0077] La información que representa la matriz de mezcla descendente D (aplicada en el codificador y utilizada como el descodificador) puede ser determinada (en el codificador) y obtenida (en el descodificador) utilizando los siguientes procedimientos principales.

[0078] La matriz de mezcla descendente D puede ser:

- establecida y aplicada (en el codificador) y su representación cuantificada y codificada transmitida explícitamente (al descodificador) vía la carga de corriente de bits.

- asignada y aplicada (en el codificador) y restaurada (en el descodificador) utilizando tablas de consulta almacenadas (es decir, conjunto de matrices mezcla descendente predeterminadas).

- asignada y aplicada (en el codificador) y restaurada (en el descodificador) según el algoritmo o procedimiento específico (por ejemplo, colocación ponderada especialmente y equidistante ordenada de los objetos de audio a los canales de mezcla descendente disponibles).

- estimada y aplicada (en el codificador) y restaurada (en el descodificador) usando el criterio de optimización particular que permite la "mezcla flexible" de los objetos de audio de entrada (es decir, la generación de la matriz de mezcla descendente que es optimizada para la estimación paramétrica de los objetos de audio en el lado del descodificador). Por ejemplo, el codificador genera la matriz de mezcla descendente de una manera de hacer la mezcla paramétrica ascendente más eficiente, en términos de reconstrucción de propiedad de señal especial, como covarianza, correlación inter-senal o mejorar/asegurar la estabilidad numérica del algoritmo de mezcla ascendente paramétrico.

[0079] Las realizaciones provistas pueden ser aplicadas en un número arbitrario de canales de mezcla descendente/mezcla ascendente. Pueden ser combinadas con cualquier formato de audio actual y también del futuro.

[0080] La flexibilidad del procedimiento de la invención permite la desviación de los canales no alterados para reducir la complejidad computacional, reducir la carga de flujo de bits/cantidad de datos reducida.

[0081] Se provee un codificador de audio, procedimiento o programa informático para codificación. Además, se provee un descodificador de audio, procedimiento o programa informático para decodificación. Además, se provee una señal codificada.

[0082] Aunque se han descrito algunos aspectos en el contexto de un aparato, es claro que estos aspectos también representan una descripción del procedimiento correspondiente, en el que un bloque o dispositivo corresponde a una etapa de procedimiento o un elemento de una etapa de procedimiento. Análogamente, aspectos descritos en el contexto de una etapa de procedimiento también representan una descripción de un bloque o ítem o característica correspondiente de un aparato correspondiente.

[0083] La señal descompuesta de la invención puede ser almacenada en un medio de almacenamiento digital o puede ser transmitida en un medio de transmisión tal como un medio de transmisión inalámbrico o un medio de transmisión por cable, tal como Internet.

[0084] Dependiendo de ciertos requisitos de implementación, las realizaciones de la invención pueden ser implementadas en hardware o en software. La implementación puede ser efectuada utilizando un medio de almacenamiento digital, por ejemplo, un disquete, un DVD, un CD, una ROM, una PROM, una EPROM, una EEPROM o una memoria FLASH, que tiene señales de control que se pueden leer electrónicamente almacenadas en el mismo, que cooperan (o son aptas de cooperar) con un sistema de ordenador programable, de tal manera que

se puede efectuar el procedimiento respectivo.

[0085] Algunas realizaciones según la invención comprenden un portador de datos no transitorio que tiene señales de control que se pueden leer electrónicamente, que son capaces de cooperar con un sistema informático programable, de tal manera que uno de los procedimientos descritos en esta invención es efectuado.

[0086] En general, las realizaciones de la presente invención pueden ser implementadas como un producto de programa informático con un código de programa, siendo operativo el código de programa para efectuar uno de los procedimientos, cuando el producto de programa informático se ejecuta en un ordenador. El código de programa puede por ejemplo ser almacenado en un portador que se puede leer por la máquina.

[0087] Otras realizaciones comprenden el programa informático para efectuar uno de los procedimientos descritos en esta invención, almacenado en un portador que se puede leer por la máquina.

[0088] En otras palabras, una realización del procedimiento de la invención es, por consiguiente, un programa informático que tiene un código de programa para efectuar uno de los procedimientos descritos en esta invención, cuando el programa informático se ejecuta en un ordenador.

[0089] Una realización adicional de los procedimientos de la invención es, por consiguiente, un portador de datos (o medio de almacenamiento digital o medio que se puede leer por ordenador) que comprende, grabado en el mismo, el programa informático para efectuar uno de los procedimientos descritos en esta invención.

[0090] Una realización adicional del procedimiento de la invención es, por consiguiente, una corriente de datos o una secuencia de señales que representan el programa informático para efectuar uno de los procedimientos descritos en esta invención. La corriente de datos o la secuencia de señales pueden estar por ejemplo configurada para ser transferido vía una conexión de comunicación de datos, por ejemplo vía Internet.

[0091] Una realización adicional comprende un medio de procesamiento, por ejemplo un ordenador o un dispositivo lógico programable, configurado o apto para efectuar uno de los procedimientos descritos en esta invención.

[0092] Una realización adicional comprende un ordenador que tiene instalado en el mismo el programa informático para efectuar uno de los procedimientos descritos en esta invención.

[0093] En algunas realizaciones, un dispositivo lógico programable (por ejemplo, arreglo de compuertas programable en campo) puede ser usado para efectuar algunas o todas las funcionalidades de los procedimientos descritos en esta invención. En algunas realizaciones, un arreglo de compuerta programable en campo puede cooperar con un microprocesador para efectuar uno de los procedimientos descritos en esta invención. En general, los procedimientos son efectuados preferentemente por cualquier aparato de hardware.

[0094] Las realizaciones descritas anteriormente son meramente ilustrativas de los principios de la presente invención. Se entiende que modificaciones y variaciones de las disposiciones y los detalles descritos en esta invención serán evidentes para otros expertos en la materia. Es la intención, por consiguiente, estar limitados sólo por el alcance de las reivindicaciones de patente inminentes y no por los detalles específicos presentados a modo de descripción y explicación de las realizaciones de esta invención.

Referencias

[0095]

[MPS] ISO/IEC 23003-1:2007, MPEG-D (MPEG audio technologies), Part 1: MPEG Surround, 2007.

[BCC] C. Faller and F. Baumgarte, "Binaural Cue Coding - Part II: Schemes and applications," IEEE Trans. on Speech and Audio Proc., vol. 11, no. 6, Nov. 2003

[JSC] C. Faller, "Parametric Joint-Coding of Audio Sources", 120th AES Convention, Paris, 2006

[SAOC1] J. Herre, S. Disch, J. Hilpert, O. Hellmuth: "From SAC To SAOC - Recent Developments in Parametric Coding of Spatial Audio", 22nd Regional UK AES Conference, Cambridge, UK, April 2007

[SAOC2] J. Engdegård, B. Resch, C. Falch, O. Hellmuth, J. Hilpert, A. Hölzer, L. Terentiev, J. Breebaart, J. Koppens, E. Schuijers y W. Oomen: " Spatial Audio Object Coding (SAOC) – The Upcoming MPEG Standard on Parametric Object Based Audio Coding", 124th AES Convention, Amsterdam 2008

[SAOC] ISO/IEC, "MPEG audio technologies – Part 2: Spatial Audio Object Coding (SAOC)," ISO/IEC JTC1/SC29/WG11 (MPEG) International Standard 23003-2.

- 5 [ISS1] M. Parvaix and L. Girin: "Informed Source Separation of underdetermined instantaneous Stereo Mixtures using Source Index Embedding", IEEE ICASSP, 2010
- [ISS2] M. Parvaix, L. Girin, J.-M. Brossier: "A watermarking-based method for informed source separation of audio signals with a single sensor", IEEE Transactions on Audio, Speech and Language Processing, 2010
- 10 [ISS3] A. Liutkus y J. Pinel and R. Badeau and L. Girin and G. Richard: "Informed source separation through spectrogram coding and data embedding", Signal Processing Journal, 2011
- [ISS4] A. Ozerov, A. Liutkus, R. Badeau, G. Richard: "Informed source separation: source coding meets source separation", IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2011
- 15 [ISS5] Shuhua Zhang and Laurent Girin: "An Informed Source Separation System for Speech Signals", INTERSPEECH, 2011
- 20 [ISS6] L. Girin y J. Pinel: "Informed Audio Source Separation from Compressed Linear Stereo Mixtures", AES 42nd International Conference: Semantic Audio, 2011

REIVINDICACIONES

1. Un decodificador para generar una señal de salida de audio que comprende uno o más canales de salida de audio de una señal de mezcla descendente que comprende dos o más canales de mezcla descendente, en el que la señal de mezcla descendente codifica dos o más señales de objeto de audio, en el que el decodificador comprende:

un determinador de umbral (110) para determinar un valor de umbral dependiendo de la energía de señal o energía de ruido de al menos una de las señales de objeto de audio o dependiendo de la energía de señal o energía de ruido de al menos uno de uno o más canales de mezcla descendente y

una unidad de procesamiento (120) para generar uno o más canales de salida de audio de uno o más canales de mezcla descendente dependiendo del valor de umbral, en el que una unidad de procesamiento (120) está configurada para generar uno o más canales de salida de audio de uno o más canales de mezcla descendente dependiendo de una matriz de covarianza de objetos (**E**) de uno o más señales de objeto de audio, dependiendo de una matriz descendente (**D**) para la mezcla descendente de dos o más señales de objeto de audio para obtener uno o más canales de mezcla descendente, y dependiendo del valor de umbral,

en el que la unidad de procesamiento (120) está configurada para generar uno o más canales de salida de audio de uno o más canales de mezcla descendente mediante la aplicación del valor de umbral en una función para invertir una matriz de correlación cruzada de canal de mezcla descendente **Q**,

en el que **Q** es definido como $Q = DED^*$,

en el que **D** es la matriz de mezcla descendente para la mezcla descendente de las dos o más señales de objeto de audio para obtener los dos o más canales de mezcla descendente,

en el que **E** es la matriz de covarianza de objeto de una o más señales de objeto de audio, y

en el que la unidad de procesamiento (120) está configurada para generar uno o más canales de salida de audio de uno o más canales de mezcla descendente mediante el cálculo de los valores propios de la matriz de correlación cruzada de canal de mezcla descendente **Q**.

2. Un decodificador según la reivindicación 1,

en el que la señal de mezcla descendente comprende dos o más canales de mezcla descendente, y en el que el determinador de umbral (110) está configurado para determinar el valor de umbral dependiendo de la energía de ruido de cada uno de los dos o más canales de mezcla descendente.

3. Un decodificador según la reivindicación 2, en el que el determinador de umbral (110) está configurado para determinar el valor de umbral dependiendo de la suma de toda la energía de ruido en los dos o más canales de mezcla descendente.

4. Un decodificador según una de las reivindicaciones anteriores,

en el que el determinador de umbral (110) está configurado para determinar el valor de umbral dependiendo de la energía de señal de la señal de objeto de audio de los dos o más señales de objeto de audio que tiene la mayor energía de señal de las dos o más señales de objeto de audio.

5. Un decodificador según una de las reivindicaciones anteriores,

en el que la señal de mezcla descendente codifica las dos o más señales de objeto de audio para cada mosaico de tiempo-frecuencia de una pluralidad de mosaicos de tiempo-frecuencia,

en el que el determinador de umbral (110) está configurado para determinar un valor de umbral para cada mosaico de tiempo-frecuencia de la pluralidad de mosaicos de tiempo-frecuencia dependiendo de la energía de señal o de la energía de ruido de al menos una de las dos o más señales de objeto de audio o dependiendo de la energía de señal o de la energía de ruido de al menos uno de uno o más canales de mezcla descendente, en el que un primer valor de umbral de un primer mosaico de tiempo-frecuencia de la pluralidad de mosaicos de tiempo-frecuencia difiere de un segundo mosaico de tiempo-frecuencia de la pluralidad de mosaicos de tiempo-frecuencia y en el que la unidad de procesamiento (120) está configurada para generar para, cada mosaico de tiempo-frecuencia de la pluralidad de mosaicos de tiempo-frecuencia, un valor de canal de cada uno de uno o más canales de salida de audio de uno o más canales de mezcla descendente dependiendo del valor de umbral de dicho mosaico de tiempo-frecuencia.

6. Un decodificador según una de las reivindicaciones anteriores,

en el que la señal de mezcla descendentes comprende dos o más canales de mezcla descendente,

en el que el descodificador está configurado para determinar el valor de umbral T en decibelios, según la fórmula

$$T [dB] = E_{ruido}[dB] - E_{ref}[dB] - Z$$

5 o según la fórmula

$$T [dB] = E_{ruido}[dB] - E_{ref}[dB],$$

en la que T[dB] indica el valor de umbral en decibelios,

10 en la que $E_{ruido}[dB]$ indica la suma de toda la energía de ruido en los dos o más canales de mezcla descendente en decibelios o $E_{ruido}[dB]$ indica la suma de toda la energía de ruido en los dos o más canales de mezcla descendente en decibelios, dividido por el número de los dos o más canales de mezcla descendente,

en la que $E_{ref}[dB]$ indica la energía de señal de una de las señales de objeto de audio en decibelios y

en la que Z indica un parámetro adicional que es un número.

15

7. Un descodificador según una de las reivindicaciones 1 a 5,

en el que la señal de mezcla descendente comprende dos o más canales de mezcla descendente,

en el que el descodificador está configurado para determinar el valor de umbral T según la fórmula

20

$$T = \frac{E_{ruido}}{E_{ref} \cdot Z} \quad \text{o según la fórmula}$$

$$T = \frac{E_{ruido}}{E_{ref}},$$

en la que T indica el valor de umbral,

en la que E_{ruido} indica la suma de toda la energía de ruido en los dos o más canales de mezcla descendente o

25 $E_{ruido}[dB]$ indica en decibelios la suma de toda la energía de ruido en los dos o más canales de mezcla descendente en decibelios, dividido por el número de los dos o más canales de mezcla descendente,

en la que E_{ref} indica la energía de señal de una de las señales de objeto de audio y

en la que Z indica un parámetro adicional que es un número.

30

8. Un descodificador según una de las reivindicaciones anteriores, en el que la unidad de procesamiento (120) está configurada para generar uno o más canales de salida de audio de uno o más canales de mezcla descendente al multiplicar el valor propio más grande de los valores propios de la matriz de correlación cruzada de canal de mezcla descendente Q con el valor de umbral para obtener un umbral relativo.

35

9. Un descodificador según la reivindicación 8,

en el que la unidad de procesamiento (120) está configurada para generar uno o más canales de salida de audio de uno o más canales de mezcla descendente al generar una matriz modificada,

40

en el que la unidad de procesamiento (120) está configurada para generar la matriz modificada dependiendo solamente de aquellos vectores propios de la matriz de correlación cruzada de canal de mezcla descendente Q, que tienen un valor propio de los valores propios de la matriz de correlación cruzada de canal de mezcla descendente Q, que es mayor o igual al umbral modificado,

en el que la unidad de procesamiento (120) está configurada para llevar a cabo una inversión de matriz de la matriz modificada para obtener una matriz invertida y

45

en el que la unidad de procesamiento (120) está configurada para aplicar la matriz invertida sobre uno o más de los canales de mezcla descendente para generar uno o más canales de salida de audio.

10. Un procedimiento para generar una señal de salida de audio que comprende uno o más canales de

50

salida de audio de una señal de mezcla descendente que comprende dos o más canales de mezcla descendente, en el que la señal de mezcla descendente codifica dos o más señales de objeto de audio, en el que el procedimiento comprende:

determinar un valor de umbral dependiendo de la energía de señal o energía de ruido de al menos una de las dos o más señales de objeto de audio o dependiendo de la energía de señal o energía de ruido de al menos uno de uno o más canales de mezcla descendente, y

55

generar uno o más canales de salida de audio de uno o más canales de mezcla descendente dependiendo del valor

de umbral,

en el que generar uno o más canales de salida de audio de uno o más canales de mezcla descendente dependiendo de una matriz de covarianza de objeto (**E**) de una o más señales de objeto de audio se lleva a cabo dependiendo de una matriz descendente (**D**) para la mezcla descendente de dos o más señales de objeto de audio para obtener uno o más canales de mezcla descendente, y dependiendo del valor de umbral,

5 en el que generar uno o más canales de salida de audio de uno o más canales de mezcla descendente se lleva a cabo mediante la aplicación del valor de umbral en una función para invertir una matriz de correlación cruzada de canal de mezcla descendente **Q**,

en el que **Q** es definido como $Q = DED^*$,

10 en el que **D** es la matriz de mezcla descendente para la mezcla descendente de las dos o más señales de objeto de audio para obtener los dos o más canales de mezcla descendente,

en el que **E** es la matriz de covarianza de objeto de una o más señales de objeto de audio, y

en el que generar uno o más canales de salida de audio de uno o más canales de mezcla descendente se lleva a cabo mediante el cálculo de los valores propios de la matriz de correlación cruzada de canal de mezcla descendente

15 **Q**.

11. Un programa informático para implementar el procedimiento de la reivindicación 10 cuando es ejecutado en un ordenador o procesador de señales.

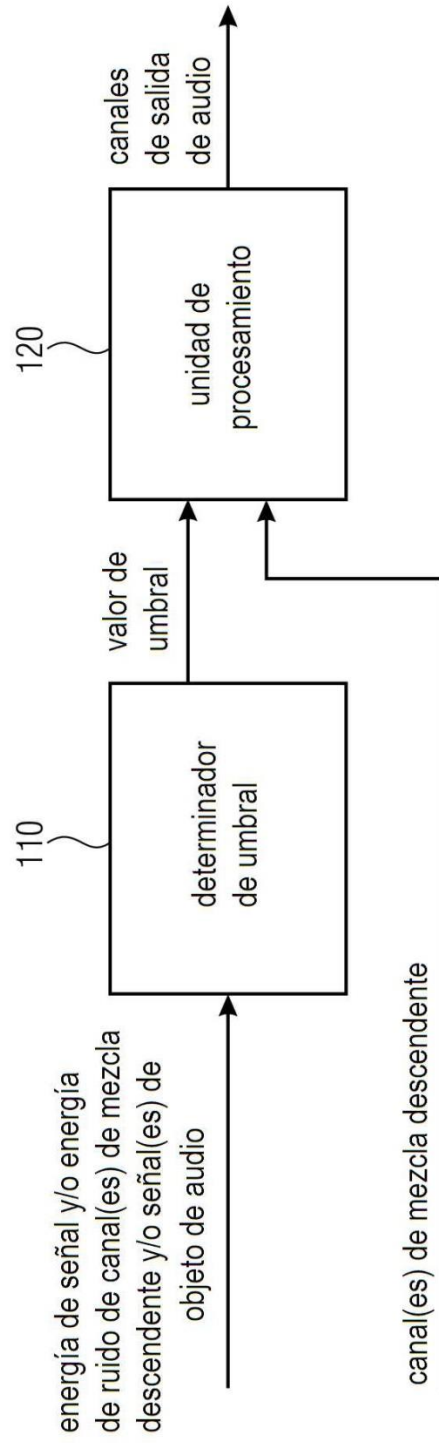


FIG 1

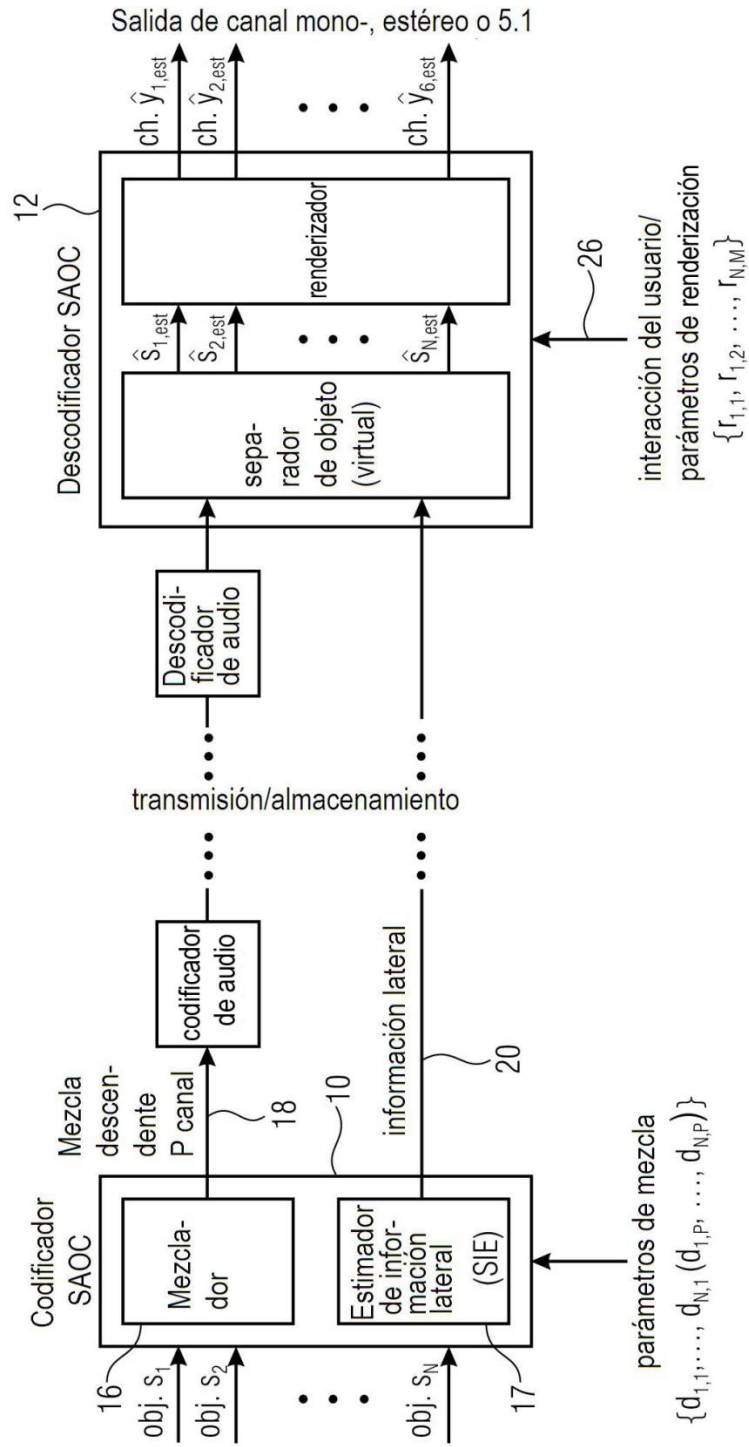


FIG 2

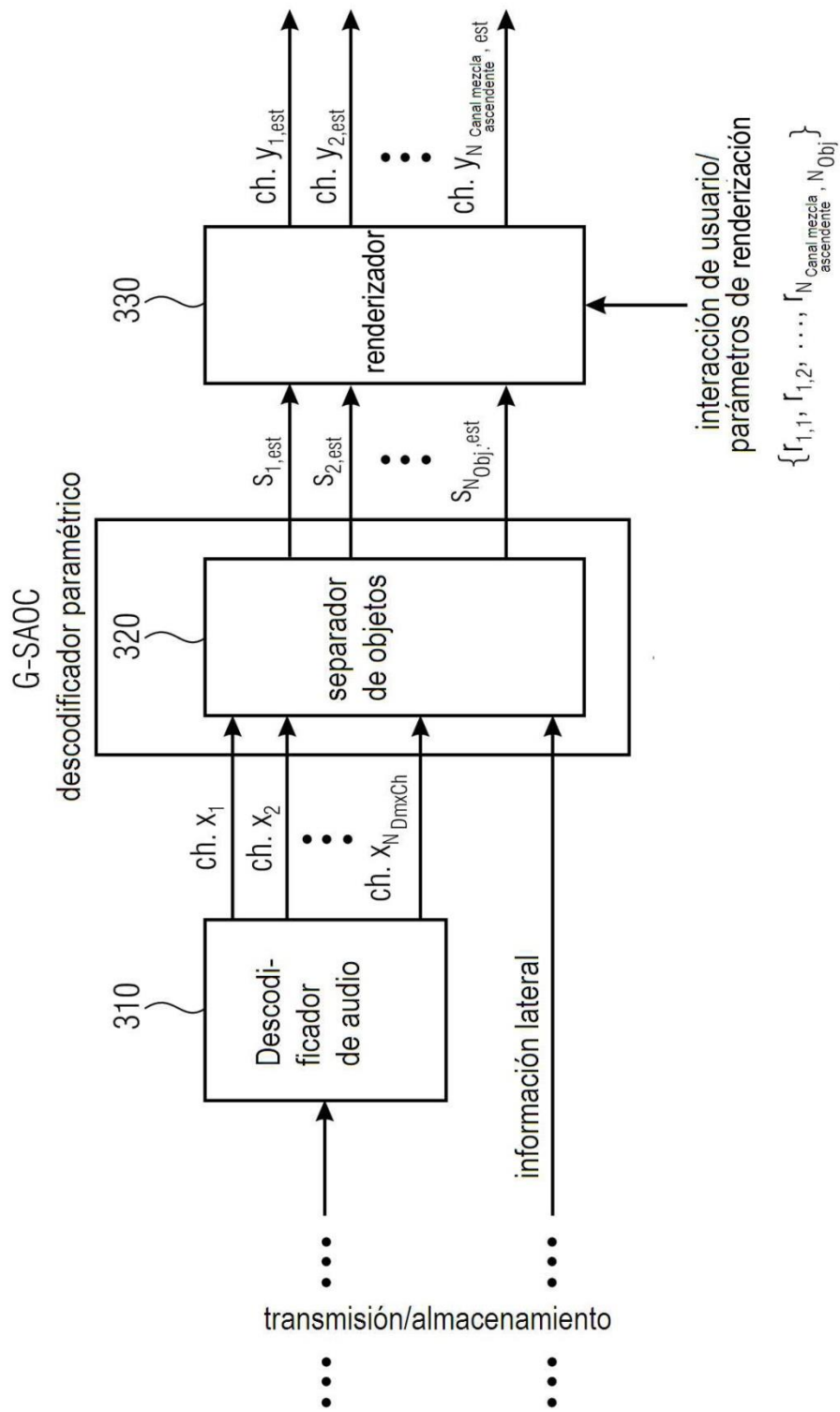


FIG 3

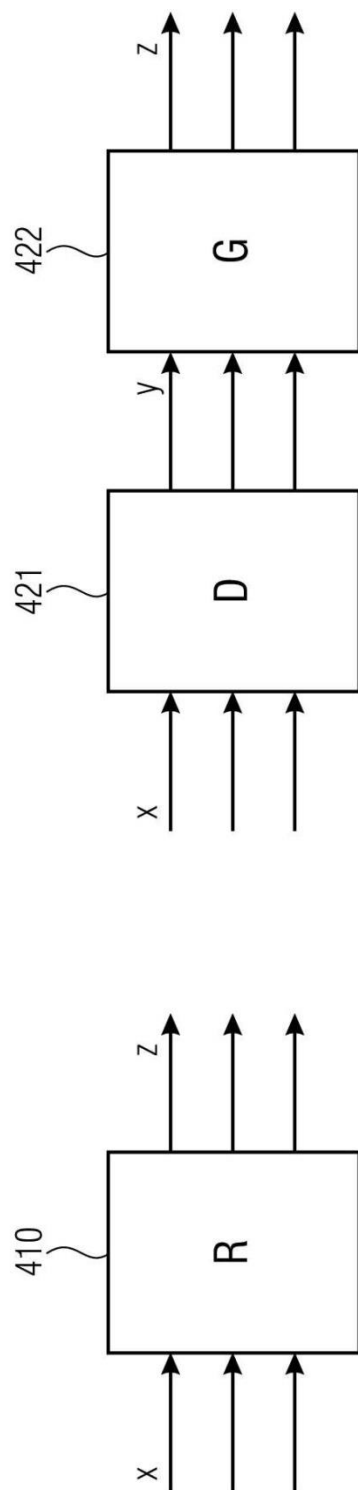


FIG 4