

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 978 713**

51 Int. Cl.:

G10L 19/008 (2013.01)

G10L 19/16 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **01.02.2016** **PCT/EP2016/052037**

87 Fecha y número de publicación internacional: **11.08.2016** **WO16124524**

96 Fecha de presentación y número de la solicitud europea: **01.02.2016** **E 16702413 (2)**

97 Fecha y número de publicación de la concesión europea: **27.03.2024** **EP 3254280**

54 Título: **Aparato y método para procesar una señal de audio codificada**

30 Prioridad:

02.02.2015 EP 15153486

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
18.09.2024

73 Titular/es:

**FRAUNHOFER-GESELLSCHAFT ZUR
FÖRDERUNG DER ANGEWANDTEN FORSCHUNG
E.V. (100.0%)
Hansastr. 27c
80686 München, DE**

72 Inventor/es:

**MURTAZA, ADRIAN;
PAULUS, JOUNI;
FUCHS, HARALD;
CAMILLERI, ROBERTA;
TERENTIV, LEON;
DISCH, SASCHA;
HERRE, JÜRGEN y
HELLMUTH, OLIVER**

74 Agente/Representante:

ARIZTI ACHA, Monica

ES 2 978 713 T3

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

DESCRIPCIÓN

Aparato y método para procesar una señal de audio codificada

5 La invención se refiere a un aparato y un método para procesar una señal de audio codificada.

Recientemente, se han propuesto técnicas paramétricas para la transmisión/almacenamiento eficiente en tasa de bits de escenas de audio que contienen múltiples objetos de audio en el campo de la codificación de audio (ver las siguientes referencias [BCC, JSC, SAOC, SAOC1, SAOC2]) y separación de origen informada (ver por ej. las siguientes referencias [ISS1, ISS2, ISS3, ISS4, ISS5, ISS6]).

Estas técnicas apuntan a reconstruir una escena de audio de salida deseada u objetos de audio de origen sobre la base de información lateral adicional que describe las señales de audio transmitidas/almacenadas y/o objetos de origen en la escena de audio. Esta reconstrucción tiene lugar en el decodificador empleando un esquema paramétrico informado de separación de origen.

Lamentablemente, se ha encontrado que, en algunos casos, los esquemas paramétricos de separación pueden llevar a severas distorsiones audibles, causando una experiencia auditiva insatisfactoria.

El documento WO 2014/021588 A1 divulga un método y un dispositivo para procesar una señal de audio de objeto y, más concretamente, un método y un dispositivo para codificar o decodificar una señal de audio de objeto o renderizar la señal de audio de objeto en un espacio tridimensional. El método para procesar una señal de audio comprende las etapas de: generar un primer grupo de señales de objeto y un segundo grupo de señales de objeto obtenidos clasificando una pluralidad de señales de objeto según un método determinado; generar una primera señal de mezcla descendente para el primer grupo de señales de objeto; generar una segunda señal de mezcla descendente para el segundo grupo de señales de objeto; generar una primera información de extracción de objeto en correspondencia con la primera señal de mezcla descendente con respecto a las señales de objeto incluidas en el primer grupo de señales de objeto; y generar una segunda información de extracción de objeto en correspondencia con la segunda señal de mezcla descendente con respecto a las señales de objeto incluidas en el segundo grupo de señales de objeto.

El documento ISO/IEC 23008-3/CD, 3D audio No. N14459, 15 de abril de 2014 divulga sistemas de sonido 3D que son capaces de realizar una experiencia de sonido significativamente mejorada en relación con los actuales programas y sistemas de reproducción de audio de 5.1 canales generalizados. Estos sistemas exigen una codificación de audio de alta calidad y una transmisión sin errores para mantener el timbre, la localización del sonido y la envolvente sonora del programa de audio original. También se contempla la presentación a través de auriculares con una espacialización adecuada.

Por lo tanto, un objetivo de la invención es mejorar la calidad del audio de señales de audio decodificadas utilizando técnicas de codificación paramétrica.

El objeto se consigue mediante un decodificador SAOC 3D según la reivindicación 1, mediante un método correspondiente según la reivindicación 6, o programa informático de según la reivindicación 7.

El objeto se consigue mediante un decodificador SAOC 3D para procesar una señal de audio codificada. La señal de audio codificada comprende una pluralidad de señales de mezcla descendente asociadas a una pluralidad de objetos de audio de entrada y parámetros de objeto (E). El aparato comprende un agrupador, un procesador y un combinador.

El agrupador está configurado para agrupar la pluralidad de señales de mezcla descendente para formar una pluralidad de grupos de señales de mezcla descendente. Cada grupo de señales de mezcla descendente está asociado a una serie de objetos de audio de entrada (o señales de audio de entrada) de la pluralidad de objetos de audio de entrada. En otras palabras: los grupos cubren sub-series de la serie de señales de audio de entrada representada por la señal de audio codificada. Cada grupo de señales de mezcla descendente está asociado asimismo a algunos de los parámetros de objeto E que describen los objetos de audio de entrada. En lo sucesivo, los grupos individuales G_k se identifican con un índice k donde $1 \leq k \leq K$ donde K representa el número de grupos de señales de mezcla descendente.

Además, el procesador – después del agrupamiento – está configurado para ejecutar por lo menos una etapa de procesamiento individual de los parámetros de objeto de cada serie de objetos de audio de entrada. Por ende, por lo menos una etapa de procesamiento se ejecuta en forma no simultánea en todos los parámetros de objeto sino individualmente en los parámetros de objeto que pertenecen al grupo respectivo de señales de mezcla descendente. En una realización se ejecuta sólo una etapa en forma individual. En una realización diferente se ejecuta más de una etapa, en tanto que, en una realización alternativa, la totalidad del procesamiento se realiza en forma individual en los grupos de señales de mezcla descendente. El procesador emite los resultados correspondientes a los grupos individuales.

En una realización diferente, el procesador – después del agrupamiento – está configurado para ejecutar por lo menos una etapa de procesamiento individualmente en cada grupo de la pluralidad de grupos de señales de mezcla descendente. Por ende, por lo menos una etapa de procesamiento se ejecuta de manera no simultánea en todas las

5 señales de mezcla descendente sino en forma individual en los respectivos grupos de señales de mezcla descendente.

Eventualmente, el combinador está configurado para combinar los resultados de los grupos o los resultados de los grupos procesados a fin de producir una señal de audio decodificada. A continuación, los resultados del grupo se combinan para proporcionar una señal de audio decodificada. La señal de audio decodificada corresponde a la

10 pluralidad de objetos de audio de entrada que son codificados por la señal de audio codificada.

El agrupamiento efectuado por el agrupador se realiza bajo por lo menos la restricción de que cada objeto de audio de entrada de la pluralidad de objetos de audio de entrada pertenece sólo o exactamente a una serie de objetos de audio de entrada. Esto implica que cada objeto de audio de entrada pertenece sólo a un grupo de señales de mezcla descendente. Esto implica asimismo que cada señal de mezcla descendente pertenece a sólo un grupo de señales de mezcla descendente.

15

De acuerdo con una realización, el agrupador está configurado para agrupar la pluralidad de señales de mezcla descendente en la pluralidad de grupos de señales de mezcla descendente de tal manera que cada objeto de audio de entrada de cada serie de objetos de audio de entrada carezca de una relación señalizada en la señal de audio codificada con otros objetos de audio de entrada o tenga una relación señalizada en la señal de audio codificada sólo con por lo menos un objeto de audio de entrada que pertenece a la misma serie de objetos de audio de entrada. Esto implica que ningún objeto de audio de entrada tiene una relación señalizada con un objeto de audio de entrada perteneciente a un grupo diferente de señales de mezcla descendente. Esa relación señalizada consiste, en una realización, que dos objetos de audio de entrada son las señales estéreo que se originan en una única fuente.

20 25

El decodificador SAOC 3D inventivo procesa una señal de audio codificada que comprende señales de mezcla descendente.

La ejecución de la mezcla en forma descendente es una parte del proceso de codificación de un número dado de señales de audio individuales e implica que un determinado número de objetos de audio de entrada se combina en una señal de mezcla en forma descendente. De esa manera, el número de objetos de audio de entrada se reduce a un menor número de señales de mezcla descendente. Debido a esto, las señales de mezcla descendente se asocian a una pluralidad de objetos de audio de entrada.

30 35

Las señales de mezcla descendente son agrupadas en grupos de señales de mezcla descendente son sometidas individualmente – es decir en forma de grupos individuales – a por lo menos una etapa de procesamiento. Por ende, el aparato ejecuta por lo menos una etapa de procesamiento de manera no conjunta en todas las señales de mezcla descendente sino en forma individual en los grupos individuales de señales de mezcla descendente. En una realización diferente, los parámetros de objeto de los grupos son tratados por separado para obtener las matrices que se han de aplicar a la señal de audio codificada.

40

En una realización, cada señal de mezcla descendente es atribuida a un grupo de señales de mezcla descendente y, en consecuencia, es procesada en forma individual con respecto a por lo menos una etapa de procesamiento. En esta realización el número de grupos de señales de mezcla descendente es igual al número de señales de mezcla descendente. Esto implica que el agrupamiento el procesamiento individual coinciden.

45

En un ejemplo la combinación es una de las etapas finales del procesamiento de la señal de audio codificada. En una realización que no forma parte de la invención, los resultados de los grupos son sometidos a su vez a etapas de procesamiento diferentes, que se ejecutan de manera individual o conjunta sobre los resultados de los grupos.

50

Se ha demostrado que el agrupamiento (o la detección de grupos) y el tratamiento individual de los grupos conduce a una mejora de la calidad del audio. Esto es así especialmente, por ejemplo, en el caso de las técnicas de codificación paramétrica.

55

De acuerdo con una realización, el agrupador del aparato está configurado para agrupar la pluralidad de señales de mezcla descendente para formar la pluralidad de grupos de señales de mezcla descendente minimizando a la vez un número de señales de mezcla descendente dentro de cada grupo de señales de mezcla descendente. En esta realización, el aparato trata de reducir el número de señales de mezcla descendente perteneciente a cada grupo. En un caso, a por lo menos un grupo de señales de mezcla descendente pertenece sólo una señal de mezcla descendente.

60

De acuerdo con una realización, el agrupador está configurado para agrupar dicha pluralidad de señales de mezcla descendente para formar dicha pluralidad de grupos de señales de mezcla descendente por lo que sólo una señal

individual de mezcla descendente pertenece a un grupo de señales de mezcla descendente. En otras palabras: El agrupamiento da lugar a diversos grupos de señales de mezcla descendente en los cuales se da por lo menos un grupo de señales de mezcla descendente al cual pertenece sólo una señal de mezcla descendente. Por consiguiente, por lo menos un grupo de señales de mezcla descendente se refiere a sólo una señal individual de mezcla descendente. En otra realización adicional, se maximiza el número de grupos de señales de mezcla descendente al cual sólo pertenece una señal de mezcla descendente.

En una realización, el agrupador del aparato está configurado para agrupar la pluralidad de señales de mezcla descendente en la pluralidad de grupos de señales de mezcla descendente sobre la base de información dentro de la señal de audio codificada. En otra realización, el aparato utiliza sólo información dentro de la señal de audio codificada para agrupar las señales de mezcla descendente. El uso de la información presente dentro del flujo de bits de la señal de audio codificada comprende – en una realización – tomar en cuenta la información de correlación o covarianza. El agrupador, especialmente, extrae de la señal de audio codificada la información sobre la relación entre diferentes objetos de audio de entrada.

En una realización, el agrupador está configurado para agrupar dicha pluralidad de señales de mezcla descendente para formar dicha pluralidad de grupos de señales de mezcla descendente sobre la base de valores bsRelatedTo dentro de dicha señal de audio codificada. Con respecto a estos valores remitirse, por ejemplo, a el documento WO 2011/039195 A1.

De acuerdo con una realización, el agrupador está configurado para agrupar la pluralidad de señales de mezcla descendente para formar la pluralidad de grupos de señales de mezcla descendente aplicando por lo menos las siguientes etapas (a cada grupo de señales de mezcla descendente):

- detectar si una señal de mezcla descendente está asignada a un grupo existente de señales de mezcla descendente;
- detectar si por lo menos un objeto de audio de entrada de la pluralidad de objetos de audio de entrada asociados a la señal de mezcla descendente es parte de una serie de objetos de audio de entrada asociados a un grupo existente de señales de mezcla descendente;
- asignar la señal de mezcla descendente a un nuevo grupo de señales de mezcla descendente en caso de que la señal de mezcla descendente esté excluida de la asignación a un grupo existente de señales de mezcla descendente (por ende, la señal de mezcla descendente ya no está asignada a un grupo) y en caso de que todos los objetos de audio de entrada de la pluralidad de objetos de audio de entrada asociados a la señal de mezcla descendente estén libres de una asociación con un grupo existente de señales de mezcla descendente (por ende, los objetos de audio de entrada de la señal de mezcla descendente no están ya – por medio de una señal de mezcla descendente diferente – asignada a un grupo) y
- combinar la señal de mezcla descendente con un grupo existente de señales de mezcla descendente ya sea en caso de que la señal de mezcla descendente esté asignada al grupo existente de señales de mezcla descendente o en caso de que por lo menos un objeto de audio de entrada de la pluralidad de objetos de audio de entrada asociados a la señal de mezcla descendente esté asociado al grupo existente de señales de mezcla descendente.

Si se toma en cuenta también una relación señalizada en la señal de audio codificada se ha de agregar otro paso de detección, lo que origina una necesidad adicional de asignar y combinar las señales de mezcla descendente.

De acuerdo con una realización, el procesador está configurado para ejecutar en forma individual diversas etapas de procesamiento de los parámetros de objeto (E_k) de cada serie de objetos de audio de entrada (o de cada grupo de señales de mezcla descendente) para producir matrices individuales como resultados de los grupos. El combinador está configurado para combinar las matrices individuales a fin de emitir dicha señal de audio decodificada. Los parámetros de objeto (E_k) pertenecen a los objetos de audio de entrada del respectivo grupo de señales de mezcla descendente con el índice k y son procesados para obtener matrices individuales para este grupo que tiene el índice k .

De acuerdo con una realización diferente, el procesador está configurado para ejecutar diversas etapas de procesamiento en forma individual sobre cada grupo de dicha pluralidad de grupos de señales de mezcla descendente a fin de emitir señales de audio de salida como resultados de los grupos. El combinador está configurado para combinar las señales de audio de salida a fin de emitir dicha señal de audio decodificada.

En este ejemplo, los grupos de señales de mezcla descendente se procesan de tal manera que se obtienen señales de audio de salida que corresponden a los objetos de audio de entrada que pertenecen al respectivo grupo de señales de mezcla descendente. Por ende, la combinación de las señales de audio de salida con las señales de audio decodificadas está cerca de las etapas finales de los procesos de decodificación ejecutados sobre la señal de audio codificada. En este ejemplo, por consiguiente, cada grupo de señales de mezcla descendente es sometida

individualmente a todos las etapas de procesamiento después de la detección de los grupos de señales de mezcla descendente.

En un ejemplo, el procesador está configurado para ejecutar por lo menos una etapa de procesamiento individualmente en cada grupo de dicha pluralidad de grupos de señales de mezcla descendente a fin de emitir señales procesadas como resultados de los grupos. El aparato comprende además un pos-procesador configurado para procesar conjuntamente dichas señales procesadas a fin de emitir señales de audio de salida. El combinador está configurado para combinar las señales de audio de salida en forma de resultados de los grupos procesados a fin de emitir dicha señal de audio decodificada.

En este ejemplo, los grupos de señal de mezcla descendente se someten a por lo menos una etapa de procesamiento individual y a por lo menos una etapa de procesamiento conjunta con otros grupos. El procesamiento individual da lugar a señales procesadas que, en un ejemplo, se procesan conjuntamente.

En lo que respecta a las matrices, en una realización, el procesador está configurado para ejecutar por lo menos una etapa de procesamiento individualmente sobre los parámetros de objeto (E_k) de cada serie de objetos de audio de entrada a fin de emitir matrices individuales. Un pos-procesador comprendido por el aparato está configurado para procesar conjuntamente los parámetros de objeto a fin de emitir por lo menos una matriz general. El combinador está configurado para combinar dichas matrices individuales y dicha por lo menos una matriz general. En una realización el pos-procesador ejecuta por lo menos una etapa de procesamiento conjunto de las matrices individuales a fin de obtener por lo menos una matriz general.

Las siguientes realización se refieren a etapas de procesamiento ejecutadas por el procesador. Algunas de estas etapas también son adecuadas para el pos-procesador mencionado en la realización precedente.

En una realización, el procesador comprende un desmezclador configurado para desmezclar las señales de mezcla descendente de los respectivos grupos de dicha pluralidad de grupos de señales de mezcla descendente. Al desmezclar las señales de mezcla descendente el procesador obtiene representaciones de los objetos de audio de entrada originales que fueron mezclados en forma descendente introduciéndolos en la señal de mezcla descendente.

De acuerdo con una realización, el desmezclador está configurado para desmezclar las señales de mezcla descendente de los respectivos grupos de dicha pluralidad de grupos de señales de mezcla descendente sobre la base de un algoritmo de Error Cuadrático Medio Mínimo (MMSE, por sus siglas en inglés). Dicho algoritmo será explicado en la descripción siguiente.

En una realización diferente, en la cual el procesador comprende un desmezclador configurado para procesar los parámetros de objeto de cada serie de objetos de audio de entrada individualmente a fin de emitir matrices individuales de desmezcla.

En una realización, el procesador comprende una calculadora configurada para computar, individualmente por cada grupo de señales de mezcla descendente, matrices con tamaños que dependen de por lo menos uno de un número de objetos de audio de entrada de la serie de objetos de audio de entrada asociados al respectivo grupo de señales de mezcla descendente y un número de señales de mezcla descendente que pertenecen al respectivo grupo de señales de mezcla descendente. Como los grupos de señales de mezcla descendente son de menor tamaño que todo el conjunto de señales de mezcla descendente y como los grupos de señales de mezcla descendente se refieren a números menores de señales de audio de entrada, las matrices utilizadas para el procesamiento de los grupos de señales de mezcla descendente son de menor tamaño que las empleadas en el estado actual de la técnica. Esto facilita el cómputo.

De acuerdo con un ejemplo, la calculadora está configurada para computar, respecto de las matrices de desmezcla individuales, un umbral individual sobre la base de un valor máximo de energía dentro del respectivo grupo de señales de mezcla descendente.

De acuerdo con una realización, el procesador está configurado para computar un umbral individual sobre la base de un valor máximo de energía dentro del respectivo grupo de señales de mezcla descendente por cada grupo de señales de mezcla descendente individualmente.

En una realización, la calculadora está configurada para computar para una etapa de regularización para desmezclar las señales de mezcla descendente de cada grupo de señales de mezcla descendente, un umbral individual sobre la base de un valor máximo de energía dentro del respectivo grupo de señales de mezcla descendente. Los umbrales para los grupos de señales de mezcla descendente son computados, en una realización diferente, por el desmezclador en sí.

La siguiente explicación demostrará el efecto interesante del cómputo del umbral para los grupos (un umbral por cada

grupo) y no para todas las señales de mezcla descendente.

De acuerdo con una realización, el procesador comprende un renderizador configurado para renderizar las señales de mezcla descendente desmezcladas de los grupos respectivos para una situación de salida de dicha señal de audio decodificada a fin de emitir señales renderizadas. La renderización tiene lugar sobre la base de una entrada provista por el oyente o sobre la base de datos acerca de la situación de salida real.

En una realización, el procesador comprende un renderizador configurado para procesar los parámetros de objeto a fin de emitir por lo menos una matriz de renderización.

El procesador comprende, en una realización, un pos-mezclador configurado para procesar los parámetros de objeto a fin de emitir por lo menos una matriz de descorrelación.

De acuerdo con una realización, el procesador comprende un pos-mezclador configurado para ejecutar por lo menos una etapa de descorrelación en dichas señales renderizadas y configurado para combinar los resultados ($Y_{húm}$) de la etapa de descorrelación ejecutado con dichas señales renderizadas respectivas (Y_{seca}).

De acuerdo con la invención, el procesador está configurado para determinar una matriz individual de mezcla en forma descendente (D_k) por cada grupo de señales de mezcla descendente (donde k es el índice del grupo respectivo), el procesador está configurado para determinar una matriz individual de covarianza de grupos (E_k) para cada grupo de señales de mezcla descendente, el procesador está configurado para determinar una matriz individual de covarianza de mezcla descendente de grupos (Δ_k) para cada grupo de señales de mezcla descendente sobre la base de la matriz individual de mezcla en forma descendente (D_k) y la matriz individual de grupos (E_k), y el procesador está configurado para determinar una matriz individual de grupos inversa regularizada (J_k) para cada grupo de señales de mezcla descendente.

De acuerdo con la invención, el combinador está configurado para combinar las matrices inversas regularizadas de grupos individuales (J_k) para obtener una matriz inversa regularizada general de grupos (J).

De acuerdo a un ejemplo no conforme con la invención reivindicada, el procesador está configurado para determinar una matriz paramétrica individual de desmezcla de grupos (U_k) por cada grupo de señales de mezcla descendente sobre la base de la matriz de mezcla descendente individual (D_k), la matriz individual de covarianza de grupos (E_k), y la matriz individual inversa regularizada de grupos (J_k), y el combinador está configurado para combinar una matriz paramétrica individual de desmezcla de grupos (U_k) para obtener una matriz paramétrica general de desmezcla de grupos (U).

De acuerdo con un ejemplo, el procesador está configurado para determinar una matriz individual de renderización de grupo (R_k) por cada grupo de señales de mezcla descendente.

De acuerdo a un ejemplo no conforme con la invención reivindicada, el procesador está configurado para determinar una matriz individual de mezcla ascendente ($R_k U_k$) por cada grupo de señales de mezcla descendente sobre la base de la matriz individual de renderización de grupo (R_k) y la matriz paramétrica individual de desmezcla de grupos (U_k), y el combinador está configurado para combinar las matrices individuales de mezcla ascendente ($R_k U_k$) para obtener una matriz general de mezcla ascendente (RU).

De acuerdo con un ejemplo, el procesador está configurado para determinar una matriz individual de covarianza de grupos (C_k) por cada grupo de señales de mezcla descendente sobre la base de la matriz individual de renderización de grupo (R_k) y la matriz individual de covarianza de grupos (E_k), y el combinador está configurado para combinar las matrices individuales de covarianza de grupos (C_k) para obtener una matriz general de covarianza de grupos (C).

De acuerdo a un ejemplo no conforme a la invención reivindicada, el procesador está configurado para determinar una matriz individual de covarianza de grupos de la señal estimada en forma paramétrica ($E_{y^{seca}})_k$ sobre la base de la matriz individual de renderización de grupo (R_k), la matriz paramétrica individual de desmezcla de grupos (U_k), la matriz de mezcla descendente individual (D_k), y la matriz individual de covarianza de grupos (E_k), y el combinador está configurado para combinar las matrices individuales de covarianza de grupos de la señal estimada en forma paramétrica ($E_{y^{seca}})_k$ para obtener una señal general estimada en forma paramétrica $E_{y^{seca}}$.

De acuerdo con un ejemplo no conforme a la invención reivindicada, el procesador está configurado para determinar una matriz inversa regularizada (J) sobre la base de una descomposición en valores singulares de una matriz de covarianza de mezcla descendente (E_{DMX}).

De acuerdo con una realización, el procesador está configurado para determinar una sub-matriz (Δ_k) para la determinación de una matriz paramétrica de desmezcla (U), mediante la selección de elementos (Δ (m , n)) que corresponden a las señales de mezcla descendente (m , n) asignadas al grupo respectivo (que tiene el índice k) de

señales de mezcla descendente. Cada grupo de señales de mezcla descendente cubre un número estipulado de señales de mezcla descendente y una serie de objetos de audio de entrada asociada y en la presente se lo indica con el índice k.

- 5 De acuerdo con esta realización, las sub-matrices individuales (Δ_k) se obtienen seleccionando o eligiendo los elementos de la matriz de covarianza de mezcla descendente Δ que pertenecen al grupo respectivo k.

De acuerdo con la invención, las sub-matrices individuales (Δ_k) se invierten individualmente y los resultados se combinan en la matriz inversa regularizada (J).

- 10 En una realización diferente, las sub-matrices (Δ_k) se obtienen utilizando su definición, que es $\Delta_k = D_k E_k D_k^*$ con la matriz individual de mezcla descendente individual (D_k)

- 15 De acuerdo con un ejemplo, el combinador está configurado para determinar una matriz de pos-mezcla (P) sobre la base de las matrices individualmente determinadas por cada grupo de señales de mezcla descendente y el combinador está configurado para aplicar la matriz de pos-mezcla (P) a la pluralidad de señales de mezcla descendente a fin de obtener la señal de audio decodificada. En este ejemplo, de los parámetros de objeto se computa una matriz de pos-mezcla que se aplica a la señal de audio codificada a fin de obtener la señal de audio decodificada.

- 20 De acuerdo con una realización, el aparato y sus componentes respectivos están configurados para ejecutar, por cada grupo de señales de mezcla descendente individualmente, por lo menos uno de los siguientes cálculos:

- el cálculo de matriz de covarianza de grupos E_k de un tamaño N_k veces N_k con los elementos:

25
$$e_{i,j}^k = \sqrt{OLD_i^k OLD_j^k} IOC_{i,j}^k,$$

- el cálculo de la matriz de covarianza de mezcla descendente de grupos Δ_k de un tamaño M_k veces M_k : $\Delta_k = D_k E_k D_k^*$,

- 30 $\Delta_k = D_k E_k D_k^*$: $\Delta_k = V_k \Lambda_k V_k^*$,

- el cálculo de la matriz de grupo inversa regularizada J_k que aproxima $J_k \approx \Delta_k^{-1}$: $J_k = V_k \Lambda_k^{inv} V_k^*$ que incluye el cálculo de la matriz individual Λ_k^{inv} (los detalles se presentan más adelante).

- 35 De acuerdo con un ejemplo, el aparato y sus componentes respectivos están configurados para ejecutar, por cada grupo de señales de mezcla descendente individualmente, por lo menos uno de los siguientes cálculos:

- el cálculo de la matriz de desmezcla paramétrica de grupos U_k de un tamaño N_k veces M_k : $U_k = E_k D_k^* J_k$,

- 40 • la multiplicación de la matriz de renderización de grupos R_k de un tamaño N_{upmix} veces N_k con la matriz de desmezcla U_k de un tamaño N_k veces M_k : $R_k U_k$,

- el cálculo de la matriz de covarianza de grupo C_k de un tamaño N_{out} veces N_{out} : $C_k = R_k E_k R_k^*$,

- 45 • el cálculo de la covarianza de grupo de la señal estimada en forma paramétrica $(E_y^{seca})_k$ de un tamaño N_{out} veces N_{out} :
$$(E_y^{dry})_k = R_k U_k (D_k E_k D_k^*) U_k^* R_k^*.$$

En este aspecto, k denota un índice de grupo del respectivo grupo de señales de mezcla descendente, N_k denota el número de objetos de audio de entrada de la serie de objetos de audio de entrada asociada, M_k denota el número de señales de mezcla descendente que pertenece al respectivo grupo de señales de mezcla descendente y N_{out} denota el número de canales de salida mezclados en forma ascendente o de salida.

- 55 Las matrices computadas tienen un tamaño menor que las utilizadas en el estado de la técnica. En consecuencia, en una realización se ejecuta la mayor cantidad de etapas de procesamiento posibles individualmente en los grupos de señales de mezcla descendente.

También se obtiene el objetivo de la invención mediante un correspondiente método para procesar una señal de audio codificada. La señal de audio codificada comprende una pluralidad de señales de mezcla descendente asociadas a una pluralidad de objetos de audio de entrada y parámetros de objeto. El método comprende las siguientes etapas:

60

- agrupar las señales de mezcla descendente para obtener una pluralidad de grupos de señales de mezcla descendente asociadas a una serie de objetos de audio de entrada de la pluralidad de objetos de audio de entrada,
- ejecutar por lo menos una etapa de procesamiento individualmente de los parámetros de objeto de cada serie de objetos de audio de entrada a fin de emitir resultados de grupos, y
- combinar dichos resultados de grupos a fin de emitir una señal de audio decodificada.

El agrupamiento se lleva a cabo con por lo menos la restricción de que cada objeto de audio de entrada de la pluralidad de objetos de audio de entrada pertenece a sólo una serie de objetos de audio de entrada.

Las realizaciones antes citadas del aparato también se pueden poner en práctica mediante las etapas del método y las correspondientes realizaciones del método. Por lo tanto, las explicaciones presentadas respecto de las realizaciones del aparato también se aplican al método.

La invención se explicará a continuación en relación con los dibujos adjuntos y las realizaciones representadas en los dibujos adjuntos, en los que:

La Fig. 1 muestra una reseña general de un concepto de mezcla descendente/mezcla ascendente paramétrico basado en MMSE,

La Fig. 2 muestra un sistema de reconstrucción paramétrica con descorrelación aplicada a la salida renderizada,

La Fig. 3 muestra una estructura de un procesador de mezcla descendente,

La Fig. 4 muestra espectrogramas de cinco objetos de audio de entrada (columna de la izquierda) y espectrogramas de los canales de mezcla descendente correspondientes (columna de la derecha),

La Fig. 5 muestra espectrogramas de señales de salida de referencia (columna de la izquierda) y espectrogramas de las correspondientes señales de salida 3D de SAOC decodificadas y renderizadas (columna de la derecha),

La Fig. 6 muestra espectrogramas de las señales 3D SAOC de salida que utilizan la invención,

La Fig. 7 muestra un procesamiento de parámetros de trama de acuerdo con el estado de la técnica,

La Fig. 8 muestra un procesamiento de parámetros de trama de acuerdo con la invención,

La Fig. 9 muestra un ejemplo de una implementación de una función de detección de grupos,

La Fig. 10 muestra esquemáticamente un aparato para codificar objetos de audio de entrada,

La Fig. 11 muestra esquemáticamente un ejemplo de un aparato de la invención para el procesamiento de una señal de audio codificada,

La Fig. 12 muestra esquemáticamente un ejemplo diferente de un aparato de la invención para el procesamiento de una señal de audio codificada,

La Fig. 13 muestra una secuencia de etapas de una realización del método inventivo,

La Fig. 14 muestra esquemáticamente un ejemplo de un aparato de la invención,

La Fig. 15 muestra esquemáticamente un ejemplo adicional de un aparato,

La Fig. 16 muestra esquemáticamente un procesador de un aparato de la invención y

La Fig. 17 muestra esquemáticamente la aplicación de un aparato de la invención.

A continuación se presenta una reseña general de esquemas de separación paramétrica, utilizando el ejemplo de la tecnología MPEG Spatial Audio Object Coding (SAOC, por sus siglas en inglés, Codificación de Objeto de Audio Espacial ([SAOC]) y la parte de procesamiento 3D SAOC de Audio 3D MPEG-H ([SAOC3D, SAOC3D2]). Se consideran las propiedades matemáticas de estos métodos.

Se utiliza la siguiente notación matemática:

| | | |
|----|--------------------------------------|---|
| | N | número de objetos de audio de entrada (o de otro modo: objetos de entrada) |
| | N_{dmx} | número de canales de mezcla descendente (transporte) |
| 5 | N_{out} | número de canales de mezcla ascendente (renderizados) |
| | N_{samples} | número de muestras por señal de audio |
| 10 | D | matriz de mezcla en forma descendente, tamaño N_{dmx} veces N |
| | S | señal de objeto de audio de entrada, tamaño N veces N_{samples} |
| | E | matriz de covarianza de objetos, tamaño N veces N , que aproxima $\mathbf{E} \approx \mathbf{S}\mathbf{S}^*$ |
| 15 | X | señales de audio de mezcla descendente, tamaño N_{dmx} veces N_{samples} , que se define como $\mathbf{X} = \mathbf{D}\mathbf{S}$ |
| | E_{DMX} | matriz de covarianza de las señales de mezcla descendente, tamaño N_{dmx} veces N_{dmx} , que se define como $\mathbf{E}_{DMX} = \mathbf{D}\mathbf{E}\mathbf{D}^*$ |
| 20 | U | matriz paramétrica de estimación de origen, tamaño N veces N_{dmx} , que se aproxima a $\mathbf{U} \approx \mathbf{E}\mathbf{D}^* (\mathbf{D}\mathbf{E}\mathbf{D}^*)^{-1}$ |
| | R | matriz de renderización (especificada del lado del decodificador), tamaño N_{out} veces N |
| 25 | $\hat{\mathbf{S}}$ | señales de objeto paramétricamente reconstruidas, tamaño N veces N_{samples} , que se aproxima a S y se define como $\hat{\mathbf{S}} = \mathbf{U}\mathbf{X}$, |
| | Y_{seca} | señales de objeto paramétricamente reconstruidas y renderizadas, tamaño N_{out} veces N_{samples} , que se define como $\mathbf{Y}_{seca} = \mathbf{R}\mathbf{U}\mathbf{X}$ |
| 30 | Y_{húm} | salidas del descorrelacionador, tamaño N_{out} veces N_{samples} |
| | Y | salida final, tamaño N_{out} veces N_{samples} |
| 35 | (·)* | operador autoadjunto (Hermítico), que representa la transpuesta conjugada de (·) |
| | F_{decorr} (·) | función del descorrelacionador |

Sin perder el carácter de generalidad, para mejorar la facilidad de lectura de las ecuaciones, se omiten los índices que indican dependencia del tiempo y la frecuencia para todas las variables introducidas.

Sistemas de separación paramétrica de objetos:

Los esquemas generales de separación paramétrica tienen por fin estimar un número de orígenes de audio a partir de una mezcla de señales (mezcla en forma descendente) utilizando información paramétrica auxiliar. La solución típica para esta tarea se basa en la aplicación de algoritmos de estimación de Error Cuadrático Medio Mínimo (Minimum Mean Squared Error (MMSE por sus siglas en inglés)). La tecnología SAOC es un ejemplo de esos sistemas de codificación paramétrica de audio.

La Fig. 1 muestra el principio general de la arquitectura del codificador/decodificador SAOC.

El procesamiento paramétrico general de mezcla descendente/ascendente se lleva a cabo de manera selectiva en tiempo/frecuencia y se puede describir como secuencia de las siguientes etapas:

- Se provee al "codificador" de "objetos de audio" de entrada **S** y "parámetros de mezcla" **D**. El "mezclador" mezcla en forma descendente los "objetos de audio" **S** para dar origen a un número de "señales de mezcla descendente" **X** usando los "parámetros de mezcla" **D** (por ej., ganancias de mezcla en forma descendente).

- El "estimador de información lateral" extrae la información que describe las características de los "objetos de audio" de entrada **S** (por ej., propiedades de covarianza).

- Las "señales de mezcla descendente" **X** y la información lateral son transmitidas o almacenadas. Estas señales de audio de mezcla descendente pueden ser comprimidas a su vez utilizando codificadores de audio (tales como la Capa II o III de MPEG-1/2, Advanced Audio Coding, Codificación Avanzada de Audio (AAC, por sus siglas en inglés) MPEG-2/4, Unified Speech y Audio Coding, Codificación Unificada de Voz y Audio (USAC, por sus siglas en inglés) MPEG,

etc.). La información lateral también puede ser representada y codificada de manera eficiente (por ej., en forma de relaciones codificadas de las potencias de objeto y los coeficientes de correlación de objetos).

El "decodificador" restablece los "objetos de audio" originales de las "señales de mezcla descendente" decodificadas empleando la información lateral transmitida (esta información suministra los parámetros de objeto). El "procesador de información lateral" estima los coeficientes de desmezcla que se han de aplicar a las "señales de mezcla descendente" dentro del "separador paramétrico de objetos" para obtener la reconstrucción paramétrica de los objetos de **S**. Los "objetos de audio" reconstruidos son renderizados a una escena objetivo (multicanal), representada por los canales de salida **Y**, aplicando un "parámetro de renderización" **R**.

Se aplica el mismo principio y las etapas sucesivas que en el procesamiento por SAOC 3D, que incorpora una ruta de descorrelación adicional.

La Fig. 2 presenta una reseña general del concepto paramétrico de mezcla descendente/ascendente con ruta de descorrelación integrada.

Usando el ejemplo de la técnica SAOC 3D, parte del Audio 3D de MPEG-H, las principales etapas de procesamiento de ese sistema de separación paramétrica se pueden resumir de la siguiente manera:

El decodificador SAOC 3D produce la salida renderizada modificada **Y** en forma de mezcla de una señal paraméricamente reconstruida y renderizada (señal seca) **Y_{seca}** y su versión descorrelacionada (señal húmeda) **Y_{húm}**.

Las etapas de procesamiento – en lo que respecta a lo relevante a la invención – se pueden diferenciar como se muestra en la Fig. 3:

- Desmezcla, lo que reconstruye paraméricamente los objetos de audio de entrada usando la matriz **U**,
- Renderización utilizando la información de renderización (matriz **R**),
- Descorrelación,
- Pos-mezcla usando la matriz **P**, computado sobre la base de información contenida en el flujo de bits.

La separación paramétrica de objetos se obtiene de la señal de mezcla descendente **X** utilizando la matriz de desmezcla **U** sobre la base de la información lateral adicional: $\hat{\mathbf{S}} = \mathbf{U}\mathbf{X}$.

La información de renderización **R** se utiliza para obtener la señal seca en forma de: $\mathbf{Y}_{seca} = \mathbf{R} \hat{\mathbf{S}} = \mathbf{R}\mathbf{U}\mathbf{X}$.

La señal final de salida **Y** se computa a partir de las señales **Y_{seca}** y **Y_{húm}** como $\mathbf{Y} = \mathbf{P} \begin{bmatrix} \mathbf{Y}_{seca} \\ \mathbf{Y}_{húm} \end{bmatrix}$.

La matriz de mezcla **P** se computa, por ejemplo, sobre la base de información de renderización, información de correlación, información de energía, información de covarianza, etc.

En la invención, esto ha de ser la matriz de pos-mezcla aplicada a la señal de audio codificada para obtener la señal de audio decodificada.

A continuación se explica la operación común de separación paramétrica de objetos utilizando MMSE.

La matriz de desmezcla **U** se obtiene sobre la base de información derivada de las variables contenidas en el flujo de bits (por ejemplo, la matriz de mezcla en forma descendente **D** y la información de covarianza **E**), usando el algoritmo de estimación por Error Cuadrático Medio Mínimo (Minimum Mean Squared Error (MMSE, por sus siglas en inglés)): $\mathbf{U} = \mathbf{E}\mathbf{D}^*\mathbf{J}$.

La matriz **J** de un tamaño N_{dmx} veces N_{dmx} representa una aproximación de la pseudo-inversa de la matriz de covarianza de mezcla descendente $\mathbf{E}_{DMX} = \mathbf{D}\mathbf{E}\mathbf{D}^*$ como: $\mathbf{J} \approx \mathbf{E}_{DMX}^{-1}$.

El cómputo de la matriz **J** se deriva de acuerdo con: $\mathbf{J} = \mathbf{V}\mathbf{\Lambda}^{inv}\mathbf{V}^*$, en el que las matrices **V** y **Λ** se determinan utilizando la descomposición en valores singulares (SVD) de la matriz \mathbf{E}_{DMX} como: $\mathbf{E}_{DMX} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^*$.

Se debe tener en cuenta que se pueden obtener resultados similares utilizando métodos de descomposición diferentes tales como: descomposición en eigenvalores, descomposición de Schur, etc.

La operación inversa regularizada $(\cdot)^{inv}$, usada para la matriz diagonal en valores singulares **Λ**, se puede determinar,

por ejemplo, como se hace en SAOC 3D, usando un truncamiento de los valores singulares con respecto al valor singular más alto:

$$\mathbf{\Lambda}^{inv} = \lambda_{i,j}^{-1} = \begin{cases} \frac{1}{\lambda_{i,i}} & i = j \text{ y } \lambda_{i,i} \geq T_{reg}^{\Lambda}, \\ 0 & \text{o bien.} \end{cases}$$

5

En una realización diferente, se utiliza la siguiente fórmula:

$$\mathbf{\Lambda}^{inv} = \lambda_{i,j}^{-1} = \begin{cases} \frac{1}{\lambda_{i,i}} & i = j \text{ y } \text{abs}(\lambda_{i,i}) \geq T_{reg}^{\Lambda}, \\ 0 & \text{o bien} \end{cases}$$

- 10 El escalar de regularización respectivo T_{reg}^{Λ} se determina utilizando el umbral absoluto T_{reg} y el valor máximo de $\mathbf{\Lambda}$ como: $T_{reg}^{\Lambda} = \max_i (\lambda_{i,i}) T_{reg}$, donde $T_{reg} = 10^{-2}$, por ejemplo.

Dependiendo de la definición valores singulares, $\lambda_{i,i}$ se puede limitar sólo a valores positivos (si $\lambda_{i,i} < 0$ luego $\lambda_{i,i} = \text{abs}(\lambda_{i,i})$ y $\text{sign}(\lambda_{i,i})$ se multiplica por el correspondiente vector singular izquierdo o derecho) o se pueden admitir valores negativos.

15

En el segundo caso con valores negativos de $\lambda_{i,i}$ el escalar de regularización relativo T_{reg}^{Λ} se computa como:

$$T_{reg}^{\Lambda} = \max_i (\text{abs}(\lambda_{i,i})) T_{reg}$$

- 20 Para simplificar, en lo sucesivo se utiliza la segunda definición de T_{reg}^{Λ} .

Se pueden obtener resultados similares utilizando el truncamiento de los valores singulares con respecto a un valor absoluto u otros métodos de regularización empleados para la inversión de la matriz.

- 25 La inversión de valores singulares muy pequeños puede llevar a coeficientes de desmezcla muy elevados y, en consecuencia, a altas amplificaciones de los canales de mezcla descendente correspondientes. En ese caso, se pueden amplificar los canales con niveles de energía muy bajos utilizando altas ganancias y esto puede llevar a distorsiones audibles. Para reducir este efecto desfavorable, los valores singulares menores que el umbral relativo T_{reg}^{Λ} se truncan a cero.

30

Se explican ahora las desventajas descubiertas de la técnica de separación paramétrica de objetos del estado de la técnica.

- 35 Los métodos de separación paramétrica de objetos descritos en el estado actual de la técnica especifican el uso de la inversión regularizada de la matriz de covarianza de mezcla descendente a fin de evitar distorsiones producidas por la separación. Sin embargo, en algunas situaciones de casos de uso real de mezcla, se identificaron distorsiones nocivas causadas por una regularización demasiado agresiva en la salida del sistema.

A continuación se construye y analiza un ejemplo de esa situación.

40

Se codifica un número $N = 5$ de objetos de audio de entrada (**S**) usando la técnica descrita (más precisamente, el método de la parte de procesamiento SAOC 3D del Audio MPEG-H 3D) para obtener un número $N_{dmx} = 3$ de canales de mezcla descendente (**X**).

45

Los objetos de audio de entrada del ejemplo pueden consistir en:

- un grupo de dos objetos de audio correlacionados que contienen señales de acompañamiento musical (izquierdo y derecho de un par estéreo),

50

- un grupo de un objeto de audio independiente que contiene una señal de voz y

- un grupo de dos objetos de audio correlacionados que contienen una grabación de piano (izquierdo y derecho de un par estéreo).

Las señales de entrada se mezclan en forma descendente en tres grupos de canales de transporte:

- el grupo G_1 con $M_1 = 1$ canales de mezcla descendente, que contiene el primer grupo de objetos,
- el grupo G_2 con $M_2 = 1$ canales de mezcla descendente, que contiene el segundo grupo de objeto, y
- el grupo G_3 con $M_3 = 1$ canales de mezcla descendente, que contiene el tercer grupo de objetos,

de tal manera que $N_{dmx} = M_1 + M_2 + M_3$.

Las matrices de mezcla descendente \mathbf{D}_k que corresponden a cada grupo G_k , donde $k = 1, 2, 3$, se construyen utilizando ganancias de mezcla unitarias y la matriz completa de mezcla descendente \mathbf{D} está dada por:

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}_1 & 0 & 0 \\ 0 & \mathbf{D}_2 & 0 \\ 0 & 0 & \mathbf{D}_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}, \text{ donde } \begin{cases} \mathbf{D}_1 = \begin{bmatrix} 1 & 1 \end{bmatrix} \\ \mathbf{D}_2 = \begin{bmatrix} 1 \end{bmatrix} \\ \mathbf{D}_3 = \begin{bmatrix} 1 & 1 \end{bmatrix} \end{cases}$$

Se puede notar la ausencia de mezcla cruzada entre el grupo de las primeras dos señales de objeto, la tercera señal de objeto y el grupo de las últimos dos señales de objeto. Nótese asimismo que la tercera señal de objeto que contiene la voz se mezcla sola en un canal de mezcla descendente. Por lo tanto, es de esperar una buena reconstrucción de este objeto y, en consecuencia, también una buena renderización. Los espectrogramas de las señales de entrada y la señal de mezcla descendente obtenida están ilustrados en la Fig. 4.

Se omite aquí la codificación de núcleo de la posible señal de mezcla descendente en un sistema real para una mejor descripción somera del efecto adverso. Del lado del decodificador se utiliza la decodificación paramétrica de SAOC 3D para reconstruir y para renderizar las señales de objeto de audio a una configuración de 3 canales ($N_{out} = 3$): los canales Izquierdo (L), Central (C) y Derecho (R).

A continuación se utiliza una remezcla sencilla de los objetos de audio de entrada del ejemplo:

- los dos primeros objetos de audio (el acompañamiento musical) se silencian (es decir, se renderizan con una ganancia 0),

- se renderiza el tercer objeto de entrada (la voz) hacia el canal central y
- se renderiza el objeto 4 hacia el canal izquierdo y el objeto 5 hacia el canal derecho.

En consecuencia, la matriz de renderización utilizada está dada por:

$$\mathbf{R} = [\mathbf{R}_1 \quad \mathbf{R}_2 \quad \mathbf{R}_3] = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

donde:

$$\mathbf{R}_1 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{R}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \text{ y } \mathbf{R}_3 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}$$

La salida de referencia se puede computar aplicando la matriz de renderización especificada directamente a las señales de entrada: $\mathbf{Y}_{ref} = \mathbf{R}\mathbf{S}$.

Los espectrogramas de la salida de referencia y las señales de salida de la decodificación y renderización de SAOC 3D están representados por las dos columnas de la Fig. 5.

De los espectrogramas expuestos de la salida del decodificador SAOC 3D, se pueden señalar las siguientes

observaciones:

• El canal central que contiene sólo la señal de voz está severamente dañado en comparación con la señal de referencia. Se pueden notar grandes agujeros espectrales. Estos agujeros espectrales (que son regiones de tiempo-frecuencia con energía faltante) llevan a severas distorsiones audibles.

• También se presentan pequeñas brechas espectrales en los canales izquierdo y derecho, especialmente en las regiones de baja frecuencia, donde se concentra la mayor parte de la energía de la señal. También, estas brechas espectrales llevan a distorsiones audibles.

• No hay mezcla cruzada de grupos de objetos en los canales de mezcla descendente; es decir que los objetos mezclados en un canal de mezcla descendente no están presentes en ningún otro canal de mezcla descendente. El segundo canal de mezcla descendente contiene sólo un objeto (la voz); por lo tanto, sólo se pueden generar las brechas espectrales en el sistema porque se lo procesa juntos con los demás canales de mezcla descendente.

Basándose en las observaciones mencionadas, se puede concluir que:

• El sistema SAOC 3D no es un sistema de "transferencia" ("pass-through"), es decir que si se mezcla una señal de entrada sola hacia un canal de mezcla descendente, la calidad del audio de esta señal original debe ser preservada mediante la decodificación y renderización.

• El sistema SAOC 3D puede introducir distorsiones audibles debido al procesamiento de señales de mezcla descendente multicanales. La calidad de salida de los objetos contenidos en un grupo de canales de mezcla descendente depende del procesamiento del resto de los canales de mezcla descendente.

Las brechas espectrales, especialmente las que se presentan en el canal central, indican que cierta información útil contenida en los canales de mezcla descendente ha sido descartada en el procesamiento. Esta pérdida de información puede ser rastreada hasta volver a la etapa de separación paramétrica de objetos, más precisamente a la etapa de regularización de inversión de la matriz de covarianza de mezcla descendente.

Por definición, la matriz de mezcla descendente del ejemplo tiene una estructura diagonal en bloques:

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}_1 & 0 & 0 \\ 0 & \mathbf{D}_2 & 0 \\ 0 & 0 & \mathbf{D}_3 \end{bmatrix}$$

Además, debido a la relación especificada entre los objetos de entrada (por ej., la señalización de las correlaciones paramétricas) también la matriz de covarianza de señales de objeto de entrada tiene una estructura diagonal en bloques:

$$\mathbf{E} = \begin{bmatrix} \mathbf{E}_1 & 0 & 0 \\ 0 & \mathbf{E}_2 & 0 \\ 0 & 0 & \mathbf{E}_3 \end{bmatrix}$$

Como resultado, la matriz de covarianza de mezcla descendente puede ser representada en forma diagonal en bloques:

$$\mathbf{E}_{\text{DMX}} = \begin{bmatrix} \mathbf{E}_1^{\text{DMX}} & 0 & 0 \\ 0 & \mathbf{E}_2^{\text{DMX}} & 0 \\ 0 & 0 & \mathbf{E}_3^{\text{DMX}} \end{bmatrix} = \begin{bmatrix} \mathbf{D}_1 \mathbf{E}_1 \mathbf{D}_1^* & 0 & 0 \\ 0 & \mathbf{D}_2 \mathbf{E}_2 \mathbf{D}_2^* & 0 \\ 0 & 0 & \mathbf{D}_3 \mathbf{E}_3 \mathbf{D}_3^* \end{bmatrix} = \mathbf{D} \mathbf{E} \mathbf{D}^*$$

En este caso, \mathbf{E}_{DMX} ya es diagonal en bloques, aunque en la generalidad de los casos se puede obtener su forma diagonal en bloques, después de la permutación de filas/ columnas utilizando el operador de permutación Φ : $\mathbf{E}_{\text{DMX}} = \Phi \mathbf{E}_{\text{DMX}} \Phi^*$.

Un operador de permutación Φ se define como una matriz obtenida por la permutación de las filas de una matriz de identidad. Si se puede representar una matriz simétrica \mathbf{A} en forma diagonal en bloques mediante la permutación de filas y columnas, se puede utilizar el operador de permutación para expresar la matriz obtenida como resultado $\tilde{\mathbf{A}}$ como: $\mathbf{A} = \Phi \tilde{\mathbf{A}} \Phi^*$.

Si Φ es un operador de permutación, luego prevalecen las siguientes propiedades:

- en principio, si V es una matriz unitaria, luego $T = \Phi V$ también es una matriz unitaria y
- en segundo lugar, $\Phi \Phi^* = \Phi^* \Phi = I$ con la matriz de identidad I .

Por lo tanto, los operadores de permutación son transparentes a los algoritmos de descomposición en valores singulares. Esto significa que la matriz original A y la matriz permutada \bar{A} comparten los mismos valores singulares y los vectores singulares permutados:

$$V \Lambda V^* = A \Rightarrow \begin{cases} (\Phi V) \Lambda (\Phi V)^* = \Phi A \Phi^* \\ (\Phi V) \Lambda (\Phi V)^* = \bar{A} \end{cases} \Rightarrow T \Lambda T^* = \bar{A}, \quad \text{donde } T = \Phi V$$

Debido a la representación diagonal en bloques, los valores singulares de la matriz E_{DMX} se pueden computar aplicando la SVD a la matriz E_{DMX} o aplicando la SVD a las sub-matrices diagonales en bloques E_{DMX_k} y combinando los resultados:

$$E_{DMX} = V \Lambda V^* = \begin{bmatrix} V_1 \Lambda_1 V_1^* & 0 & 0 \\ 0 & V_2 \Lambda_2 V_2^* & 0 \\ 0 & 0 & V_3 \Lambda_3 V_3^* \end{bmatrix}$$

con

$$\Lambda = \begin{bmatrix} \lambda_{1,1} & 0 & 0 \\ 0 & \lambda_{2,2} & 0 \\ 0 & 0 & \lambda_{3,3} \end{bmatrix}, \quad \Lambda_1 = [\lambda_{1,1}], \quad \Lambda_2 = [\lambda_{2,2}] \quad \text{y} \quad \Lambda_3 = [\lambda_{3,3}].$$

Dado que los valores singulares de la matriz de covarianza de mezcla descendente están directamente relacionados con los niveles de energía de los canales de mezcla descendente (que están descritos por la diagonal principal de la matriz E_{DMX}):

$$\sum_{k=1}^{N_{dmx}} \lambda_{k,k} = \sum_{k=1}^{N_{dmx}} E_{DMX}(k, k)$$

y los objetos contenidos en un canal no están contenidos en ningún otro canal de mezcla descendente, se puede concluir que cada valor singular corresponde a un canal de mezcla descendente.

Por lo tanto, si uno de los canales de mezcla descendente tiene un nivel de energía mucho menor que el resto de los canales de mezcla descendente, el valor singular que corresponde a este canal ha de ser mucho menor que el resto de los valores singulares.

La etapa de truncamiento utilizado en la inversión de la matriz que contiene los valores singulares de la matriz E_{DMX} .

$$\Lambda^{inv} = \lambda_{i,j}^{-1} = \begin{cases} \frac{1}{\lambda_{i,j}} & i = j \text{ y } \lambda_{i,i} \geq T_{reg}^{\Lambda}, \\ 0 & \text{o bien} \end{cases}$$

o

$$\Lambda^{inv} = \lambda_{i,j}^{-1} = \begin{cases} \frac{1}{\lambda_{i,j}} & i = j \text{ y } \text{abs}(\lambda_{i,i}) \geq T_{reg}^{\Lambda}, \\ 0 & \text{o bien} \end{cases}$$

puede llevar al truncamiento de los valores singulares que corresponden al canal de mezcla descendente con el nivel de energía bajo (con respecto al canal de mezcla descendente con la energía más elevada). Por ello, la información presente en este canal de mezcla descendente con baja energía relativa se descarta y se generan las brechas espectrales observadas en las figuras de espectrograma y en la salida de audio.

Para una mejor comprensión, se debe tener en cuenta que la mezcla en forma descendente de los objetos de audio

de entrada se produce por cada muestra y por cada banda de frecuencia por separado. Especialmente, la separación en diferentes bandas ayuda a comprender por qué se pueden encontrar brechas en los espectrogramas de las señales de salida a diferentes frecuencias.

- 5 Se puede aislar el problema identificado debido al hecho de que se computa el umbral de regularización relativa correspondiente valores singulares sin tener en cuenta que la matriz que se ha de invertir es diagonal en bloques:

$$T_{reg}^{\Lambda} = \max_i \left(\text{abs}(\lambda_{i,i}) \right) T_{reg}$$

- 10 Cada matriz diagonal en bloques corresponde a un grupo independiente de canales de mezcla descendente. El truncamiento se realiza con respecto al valor singular más alto, aunque este valor describe sólo un grupo de canales. Por consiguiente, la reconstrucción de los objetos contenidos en todos los grupos independientes de canales de mezcla descendente pasa a depender del grupo que contiene este valor singular más elevado.

- 15 A continuación, la invención se explicará basándose en la realización descrita anteriormente en relación con el estado de la técnica:

- 20 Considerando el ejemplo antes descrito, las tres matrices de covarianza pueden estar asociadas a tres grupos diferentes de canales de mezcla descendente \mathbf{G}_k con $1 \leq k \leq 3$. Los objetos de audio u objetos de audio de entrada contenidos en los canales de mezcla descendente de cada grupo no están contenidos en ningún otro grupo. Además, no se señala ninguna relación (por ej., correlación) entre los objetos contenidos en los canales de mezcla descendente de diferentes grupos.

- 25 Para solucionar el problema identificado del sistema de reconstrucción paramétrica, el método inventivo propone la aplicación del paso de regularización de manera independiente por cada grupo. Esto implica que se computan tres umbrales diferentes para la inversión de las tres matrices independientes de covarianza de mezcla descendente:

$$T_{reg}^{\Lambda_k} = \max_{i \in \mathbf{G}_k} \left(\text{abs}(\lambda_{i,i}) \right) T_{reg}$$

, donde $1 \leq k \leq 3$. Por lo tanto, en la invención en una realización tal umbral se calcula para cada grupo por separado y no como en el estado de la técnica un umbral global para las respectivas bandas de frecuencia y muestras.

- 30 La inversión de los valores singulares se obtiene de manera acorde aplicando la regularización en forma independiente para las sub-matrices $\mathbf{E}^{\text{DMX}_k}$, con $1 \leq k \leq 3$:

$$\Lambda_k^{inv} = \left(\lambda_{i,j}^{-1} \right)_{i,j \in \mathbf{G}_k} = \begin{cases} \frac{1}{\lambda_{i,i}} & i = j \text{ y } \lambda_{i,i} \geq T_{reg}^{\Lambda, \mathbf{G}_k}, \\ 0 & \text{si no} \end{cases}$$

- 35 En una realización diferente, se utiliza la siguiente fórmula:

$$\Lambda_k^{inv} = \left(\lambda_{i,j}^{-1} \right)_{i,j \in \mathbf{G}_k} = \begin{cases} \frac{1}{\lambda_{i,i}} & i = j \text{ y } \text{abs}(\lambda_{i,i}) \geq T_{reg}^{\Lambda, \mathbf{G}_k}, \\ 0 & \text{de lo contrario.} \end{cases}$$

- 40 Usando el método inventivo propuesto en un sistema SAOC 3D de lo contrario idéntico para el ejemplo descrito en la sección anterior, se mejora la calidad de salida de audio de la salida decodificada y renderizada. Las señales así obtenidas están mostradas en la Fig. 6.

- 45 Comparando los espectrogramas de la columna derecha de la Fig. 5 y de la Fig. 6, puede observarse que el método inventivo resuelve los problemas identificados en el sistema de separación paramétrica existente en la técnica anterior. El método inventivo garantiza la característica de "transferencia" ("pass-through") del sistema y, lo que es más importante, se eliminan las brechas espectrales.

- 50 La solución descrita para el procesamiento de tres grupos independientes de canales de mezcla descendente se puede extender fácilmente a cualquier número de grupos.

El método inventivo propone modificar la técnica de separación paramétrica de objetos haciendo uso del agrupamiento de la información en la inversión de la matriz de covarianza de señales de mezcla descendente. Esto supone una mejora significativa de la calidad de salida de audio.

El agrupamiento se puede obtener, por ej., de la información de mezcla y/o correlación ya existente en el decodificador sin señalización adicional.

Más precisamente, un grupo está definido, en una realización, por la serie más pequeña de señales de mezcla descendente con las dos propiedades siguientes en este ejemplo:

- en primer lugar, los objetos de audio de entrada contenidos en estos canales de mezcla descendente no están contenidos en ningún otro canal de mezcla descendente.
 - en segundo lugar, ninguna de las señales de entrada contenidas en los canales de mezcla descendente de un grupo se relaciona (por ej., no se señala una intercorrelación dentro de la señal de audio codificada) con alguna otra señal contenida en los canales de mezcla descendente de ningún otro grupo. Esa inter-correlación implica un manejo combinado de los respectivos objetos de audio durante la decodificación.
- Basándose en la definición de grupo presentada, se puede definir un número de K ($1 \leq K \leq N_{dmx}$) grupos: \mathbf{G}_k ($1 \leq k \leq K$) y la matriz de covarianza de mezcla descendente \mathbf{E}_{DMX} se puede expresar utilizando una forma diagonal en bloques aplicando un operador de permutación Φ :

$$\bar{\mathbf{E}}_{DMX} = \Phi \mathbf{E}_{DMX} \Phi^* = \begin{bmatrix} \mathbf{E}_1^{DMX} & 0 & \dots & 0 \\ 0 & \mathbf{E}_2^{DMX} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{E}_K^{DMX} \end{bmatrix}$$

Las sub-matrices \mathbf{E}_k^{DMX} se construyen mediante la selección de elementos de la matriz de covarianza de mezcla descendente correspondientes a los grupos independientes \mathbf{G}_k . Por cada grupo \mathbf{G}_k , la matriz \mathbf{E}_k^{DMX} de un tamaño M_k veces M_k se expresa usando la SVD de la siguiente manera: $\mathbf{E}_k^{DMX} = \mathbf{V}_k \mathbf{\Lambda}_k \mathbf{V}_k^*$

$$\text{donde: } \mathbf{\Lambda}_k = \begin{bmatrix} \lambda_{1,1}^k & 0 & \dots & 0 \\ 0 & \lambda_{2,2}^k & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_{M_k, M_k}^k \end{bmatrix} \text{ y } \sum_{k=1}^K M_k = N_{dmx}$$

La pseudo-inversa de la matriz \mathbf{E}_k^{DMX} se computa como $(\mathbf{E}_k^{DMX})^{-1} = \mathbf{V}_k \mathbf{\Lambda}_k^{inv} \mathbf{V}_k^*$ donde la matriz inversa regularizada $\mathbf{\Lambda}_k^{inv}$ está dada, en una realización, por:

$$\mathbf{\Lambda}_k^{inv} = (\lambda_{i,j}^{-1})_{i,j \in \mathbf{G}_k} = \begin{cases} \frac{1}{\lambda_{i,i}} & i = j \text{ y } \lambda_{i,i} \geq T_{reg}^{\Lambda_k}, \\ 0 & \text{de lo contrario.} \end{cases}$$

y en una realización diferente por:

$$\mathbf{\Lambda}_k^{inv} = (\lambda_{i,j}^{-1})_{i,j \in \mathbf{G}_k} = \begin{cases} \frac{1}{\lambda_{i,i}} & i = j \text{ y } abs(\lambda_{i,i}) \geq T_{reg}^{\Lambda_k}, \\ 0 & \text{de lo contrario.} \end{cases}$$

El escalar de regularización relativo $T_{reg}^{\Lambda_k}$ se determina empleando el umbral absoluto T_{reg} y el valor máximo de $\mathbf{\Lambda}_k$ como: $T_{reg}^{\Lambda_k} = \max_{i \in \mathbf{G}_k} (\lambda_{i,i}) T_{reg}$ con $T_{reg} = 10^{-2}$ por ejemplo.

La inversa de la matriz de covarianza de mezcla descendente permutada $\bar{\mathbf{E}}_{DMX}$ se obtiene de la siguiente manera:

$$\tilde{\mathbf{E}}_{\text{DMX}}^{-i} = \begin{bmatrix} (\mathbf{E}_1^{\text{DMX}})^{-1} & 0 & \dots & 0 \\ 0 & (\mathbf{E}_2^{\text{DMX}})^{-1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & (\mathbf{E}_K^{\text{DMX}})^{-1} \end{bmatrix}$$

y la inversa de la matriz de covarianza de mezcla descendente se computa aplicando la operación de permutación

$$\mathbf{E}_{\text{DMX}}^{*-1} = \Phi \tilde{\mathbf{E}}_{\text{DMX}}^{-i} \Phi$$

inversa:

Además, el método inventivo propone, en una realización, determinar los grupos basándose por completo en la información contenida en el flujo de bits. Por ejemplo, esta información puede ser provista por la información de mezcla en forma descendente y la información de correlación.

Más precisamente, un grupo \mathbf{G}_k está definido por la serie más pequeña de canales de mezcla descendente con las siguientes propiedades:

- Los objetos de audio de entrada contenidos en los canales de mezcla descendente del grupo \mathbf{G}_k no están contenidos en ningún otro canal de mezcla descendente. Un objeto de audio de entrada no está contenido en un canal de mezcla descendente, por ejemplo, si la correspondiente ganancia de mezcla descendente está dada por el índice de cuantificación más pequeño, o si es igual a cero.

- Todas las señales de entrada i contenidas en los canales de mezcla descendente del grupo \mathbf{G}_k no están relacionadas con ninguna señal de entrada j contenidas en ningún canal de mezcla descendente de ningún otro grupo. Por ejemplo (comparar, por ej., WO 2011/039195 A1) se puede usar el flujo de bits variable $\text{bsRelatedTo}[i][j]$ para señalar si dos objetos están relacionados ($\text{bsRelatedTo}[i][j] == 1$) o si no están relacionados ($\text{bsRelatedTo}[i][j] == 0$). También se pueden utilizar diferentes métodos para señalar dos objetos que están relacionados sobre la base de información de correlación o covarianza, por ejemplo.

Los grupos pueden ser determinados una vez por trama o una vez por serie de parámetros con respecto a todas las bandas de procesamiento, o una vez por trama o una vez por serie de parámetros por cada banda de procesamiento.

El método inventivo también permite, en una realización, reducir significativamente la complejidad informática del sistema de separación paramétrica (por ej., el decodificador SAOC 3D) haciendo uso de la información de agrupamiento en los componentes de procesamiento paramétrico más costosos desde el punto de vista informático.

Por lo tanto, el método inventivo propone eliminar los cálculos que no aportan ninguna contribución a la calidad del audio de salida final. Estos cálculos se pueden seleccionar sobre la base de la información de agrupamiento.

Más precisamente, el método inventivo propone computar todas las etapas de procesamiento paramétricas de manera independiente por cada grupo predeterminado y combinar los resultados al final.

Utilizando el ejemplo de la parte de procesamiento SAOC 3D de Audio MPEG-H 3D las operaciones informáticamente complejas están dadas por:

- el cómputo de la matriz de covarianza \mathbf{E} de un tamaño N veces N con los elementos:

$$e_{i,j} = \sqrt{\text{OLD}_i \text{OLD}_j} \text{IOC}_{i,j},$$

- el cómputo de la matriz de covarianza de señales de mezcla descendente Δ de un tamaño N_{dmx} veces N_{dmx} : $\Delta = \mathbf{DED}^*$,

- el cómputo de la descomposición en valores singulares de la matriz $\Delta = \mathbf{DED}^*$: $\Delta = \mathbf{V} \Lambda \mathbf{V}^*$,

- el cómputo de la matriz inversa regularizada \mathbf{J} que aproxima $\mathbf{J} \approx \Delta^{-1}$: $\mathbf{J} = \mathbf{V} \Lambda^{\text{inv}} \mathbf{V}^*$,

- el cómputo de la matriz paramétrica de desmezcla \mathbf{U} de un tamaño N veces N_{dmx} : $\mathbf{U} = \mathbf{ED}^* \mathbf{J}$,

- la multiplicación de la matriz de renderización \mathbf{R} de un tamaño N_{out} veces N por la matriz de desmezcla \mathbf{U} de un tamaño N veces N_{dmx} : \mathbf{RU} ,

• el cómputo de la matriz de covarianza \mathbf{C} de un tamaño N_{out} veces N_{out} : $\mathbf{C} = \mathbf{RER}^*$,

• el cómputo de la covarianza de la señal estimada en forma paramétrica \mathbf{E}_y^{seca} de un tamaño N_{out} veces N_{out} :
 $\mathbf{E}_y^{dry} = \mathbf{RU}(\mathbf{DED}^*)\mathbf{U}^*\mathbf{R}^*$.

5 Las Diferencias de Nivel de los Objetos (OLD) se refieren a la energía relativa de un objeto con respecto al objeto con la mayor cantidad de energía en una determinada banda de tiempo y frecuencia y la Coherencia Cruzada entre Objetos (IOC) describe la cantidad de similitud, o correlación cruzada de dos objetos en una determinada banda de tiempo y frecuencia.

10 El método inventivo está proponiendo reducir la complejidad información mediante el cómputo de todas las etapas de procesamiento paramétrico correspondientes a todos los K grupos predeterminados \mathbf{G}_k donde $1 \leq k \leq K$ de modo independiente y combinar los resultados al final del procesamiento de los parámetros.

15 Un grupo \mathbf{G}_k contiene M_k canales de mezcla descendente y N_k objetos de audio de entrada de tal manera que:

$$\sum_{k=1}^K M_k = N_{dmx} \text{ y } \sum_{k=1}^K N_k = N.$$

20 Por cada grupo \mathbf{G}_k , una matriz de mezcla descendente de grupo se define como \mathbf{D}_k mediante la selección de elementos de la matriz de mezcla descendente \mathbf{D} que corresponde a los canales de mezcla descendente y objetos de audio de entrada contenidos por el grupo \mathbf{G}_k .

25 De manera similar, se obtiene una matriz de renderización de grupos \mathbf{R}_k de la matriz de renderización \mathbf{R} mediante la selección de las filas correspondientes a los objetos de audio de entrada contenidos en el grupo \mathbf{G}_k .

De manera similar, se obtiene una OLD^k de vectores de grupo y una IOC^k de matriz de grupos a partir de la OLD de vector y la IOC de matriz mediante la selección de los elementos correspondientes a los objetos de audio de entrada contenidos en el grupo \mathbf{G}_k .

30 Por cada grupo \mathbf{G}_k , se reemplazan los pasos de procesamiento descritos por etapas de procesamiento con menos complejidad informática de la siguiente manera:

• el cómputo de matriz de covarianza de grupos \mathbf{E}_k de un tamaño N_k veces N_k con los elementos:

35
$$e_{i,j}^k = \sqrt{OLD_i^k OLD_j^k} IOC_{i,j}^k,$$

• el cómputo de la matriz de covarianza de mezcla descendente de grupos $\mathbf{\Delta}_k$ de un tamaño M_k veces M_k : $\mathbf{\Delta}_k = \mathbf{D}_k \mathbf{E}_k \mathbf{D}_k^*$,

40 • el cómputo de la descomposición en valores singulares de la matriz de covarianza de mezcla descendente de grupo $\mathbf{\Delta}_k = \mathbf{D}_k \mathbf{E}_k \mathbf{D}_k^*$: $\mathbf{\Delta}_k = \mathbf{V}_k \mathbf{\Lambda}_k \mathbf{V}_k^*$,

• el cómputo de la matriz de grupo inversa regularizada \mathbf{J}_k que aproxima $\mathbf{J}_k \approx \mathbf{\Lambda}_k^{-1}$: $\mathbf{J}_k = \mathbf{V}_k \mathbf{\Lambda}_k^{inv} \mathbf{V}_k^*$,

45 • el cómputo de la matriz de desmezcla paramétrica de grupos \mathbf{U}_k de un tamaño N_k veces M_k : $\mathbf{U}_k = \mathbf{E}_k \mathbf{D}_k^* \mathbf{J}_k$,

• la multiplicación de la matriz de renderización de grupos \mathbf{R}_k de un tamaño N_{upmix} veces N_k por la matriz de desmezcla \mathbf{U}_k de un tamaño N_k veces M_k : $\mathbf{R}_k \mathbf{U}_k$,

50 • el cómputo de la matriz de covarianza de grupo \mathbf{C}_k de un tamaño N_{out} veces N_{out} : $\mathbf{C}_k = \mathbf{R}_k \mathbf{E}_k \mathbf{R}_k^*$,

• el cómputo de la covarianza de grupo de la señal estimada en forma paramétrica $(\mathbf{E}_y^{seca})_k$ de un tamaño N_{out} veces N_{out} :

$$(\mathbf{E}_y^{dry})_k = \mathbf{R}_k \mathbf{U}_k (\mathbf{D}_k \mathbf{E}_k \mathbf{D}_k^*) \mathbf{U}_k^* \mathbf{R}_k^*.$$

55 Y se combinan los resultados de las etapas de procesamiento de los grupos individuales al final:

• la matriz de mezcla ascendente \mathbf{RU} de un tamaño N_{out} veces N_{dmx} se obtiene fusionando las matrices de grupo $\mathbf{R}_k \mathbf{U}_k$:
 $\mathbf{RU} = [\mathbf{R}_1 \mathbf{U}_1 \mathbf{R}_2 \mathbf{U}_2 \cdots \mathbf{R}_K \mathbf{U}_K],$

$$\mathbf{C}_k: \mathbf{C} = \sum_{k=1}^K \mathbf{C}_k,$$

• la matriz de covarianza \mathbf{C} de un tamaño N_{out} veces N_{out} se obtiene sumando las matrices de grupo

• la covarianza de la señal estimada en forma paramétrica \mathbf{E}_y^{seca} de un tamaño N_{out} veces N_{out} se obtiene sumando las

$$(\mathbf{E}_y^{dry})_k: \mathbf{E}_y^{dry} = \sum_{k=1}^K (\mathbf{E}_y^{dry})_k$$

matrices de grupo

Resumiendo las etapas de procesamiento de acuerdo con la estructura del procesador de mezcla en forma descendente ilustrado en la Fig. 3, y omitiendo la etapa de descorrelación, se puede describir el procesamiento de trama existente en la técnica anterior como en la Fig. 7.

Usando el método inventivo propuesto se reduce la complejidad informática utilizando la detección de grupos mostrados en la Fig. 8.

Un ejemplo de una implementación de una función de detección de grupos, denominada: $[K, \mathbf{G}_k] = \text{groupDetect}(\mathbf{D}, \text{RelatedTo})$ es el expuesto en la Fig. 9 que utiliza el código ANSI C la función estática "getSaocCoreGroups()".

El método inventivo propuesto demuestra ser significativamente más eficiente desde el punto de vista informático que la ejecución de las operaciones sin agrupamiento. También da lugar a una mejor asignación y uso de la memoria, admite la paralelización de los cálculos, reduce la acumulación de errores numéricos, etc.

El método inventivo propuesto y el aparato propuesto por la invención solucionan un problema existente de los sistemas de separación paramétrica de objetos del estado actual de la técnica y ofrecen una calidad de audio de salida significativamente más elevada.

El método inventivo propuesto describe un método de detección de grupos que se ejecuta en su totalidad sobre la base de la información del flujo de bits existente.

El método inventivo propuesto da origen a una significativa reducción de la complejidad informática. En general, la descomposición en valores singulares es costosa desde el punto de vista informático y su complejidad crece

exponencialmente con el tamaño de la matriz que se debe invertir: $O(N_{dmx}^3)$.

Para un gran número de canales de mezcla descendente, el computo de K veces una operación SVD para una matriz

de menor tamaño es mucho más eficiente desde el punto de vista computacional: $\sum_{k=1}^K O(M_k^3)$.

Usando las mismas consideraciones, todos las etapas de procesamiento paramétrico realizados en el decodificador pueden ser eficientemente implementados computando todas las multiplicaciones de matrices descritas en el sistema sólo para los grupos independientes y combinando los resultados.

En la siguiente tabla se presenta una estimación de la reducción de la complejidad correspondiente a diferentes números de objetos de audio de entrada, es decir objetos de audio de entrada, canales de mezcla descendente, y un número fijo de 24 canales de salida:

| | | | | | | | |
|---|-------|-------|-------|-------|-------|-------|-------|
| Número de objetos de audio de entrada | 8 | 16 | 32 | 60 | 96 | 128 | 256 |
| Número de canales de mezcla descendente, N_{dmx} | 4 | 8 | 16 | 24 | 24 | 32 | 64 |
| Número de grupos, K | 2 | 4 | 4 | 6 | 6 | 8 | 8 |
| Procesamiento de parámetros SAOC 3D [MOPS] | 7,5 | 28 | 56 | 464 | 1000 | 2022 | 12000 |
| Procesamiento de parámetros del método inventivo [MOPS] | 3 | 3 | 7,5 | 10 | 20 | 20 | 81 |
| Reducción de la complejidad [%] | 60,00 | 89,29 | 86,61 | 97,84 | 98,00 | 99,01 | 99,33 |

La invención presenta las siguientes ventajas adicionales:

• en el caso de situaciones en que sólo se puede crear un grupo, la salida es idéntica en bits a la del sistema del estado actual de la técnica.

• El agrupamiento preserva la característica de “transferencia” del sistema. Esto implica que, si un objeto de audio de entrada se mezcla solo en un canal de mezcla descendente, el decodificador puede reconstruirlo perfectamente.

La invención da lugar a las modificaciones propuestas a modo de ejemplo del texto estándar.

5

Se agrega “operación inversa regularizada 9.5.4.2.4”:

La matriz inversa regularizada \mathbf{J} que aproxima $\mathbf{J} \approx \Delta^{-1}$ se calcula como $\mathbf{J} = \mathbf{V} \mathbf{\Lambda}^{inv} \mathbf{V}^*$.

10

Las matrices \mathbf{V} y $\mathbf{\Lambda}$ se determinan como la descomposición en valores singulares de la matriz Δ como: $\Delta = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^*$.

La inversa regularizada $\mathbf{\Lambda}^{inv}$ de la matriz diagonal de valores singulares $\mathbf{\Lambda}$ se computa de acuerdo con 9.5.4.2.5.

15

En el caso de uso de la matriz Δ en el cálculo de la matriz paramétrica de desmezcla \mathbf{U} , se aplican las operaciones descritas para todas las sub-matrices Δ_k . Se obtiene una sub-matriz Δ_k seleccionando los elementos $\Delta(m, n)$ que corresponden a los canales de mezcla descendente m y n asignados al grupo k .

El grupo k está definido por la serie más pequeña de canales de mezcla descendente con las siguientes propiedades:

20

• Las señales de entrada contenidas en los canales de mezcla descendente del grupo k no están contenidos en ningún otro canal de mezcla descendente. Una señal de entrada no está contenida en un canal de mezcla descendente si la correspondiente ganancia de mezcla descendente está dada por el índice de cuantificación más pequeño (Tabla 49 de ISO/IEC 23003-2:2010).

25

• Todas las señales de entrada i contenidas en los canales de mezcla descendente del grupo k no están relacionadas con ninguna señal de entrada contenida en ningún canal de mezcla descendente de ningún otro grupo (es decir, $\text{bsRelatedTo}[i][j] == 0$).

De acuerdo con la presente invención, los resultados de las operaciones de inversión regularizadas independientes

30

$\mathbf{J}_k \approx \Delta_k^{-1}$ se combinan para obtener la matriz \mathbf{J} .

La invención da lugar asimismo a las siguientes modificaciones propuestas a manera de ejemplo del texto estándar.

Operación inversa regularizada 9.5.4.2.5

35

La matriz inversa regularizada \mathbf{J} que aproxima $\mathbf{J} \approx \Delta^{-1}$ se calcula como:

$$\mathbf{J} = \mathbf{V} \mathbf{\Lambda}^{inv} \mathbf{V}^*.$$

40

Las matrices \mathbf{V} y $\mathbf{\Lambda}$ se determinan como la descomposición en valores singulares de la matriz Δ como:

$$\mathbf{V} \mathbf{\Lambda} \mathbf{V}^* = \Delta.$$

45

La inversa regularizada $\mathbf{\Lambda}^{inv}$ de la matriz diagonal de valores singulares $\mathbf{\Lambda}$ se computa de acuerdo con 9.5.4.2.6.

En el caso de utilizarse la matriz en el cálculo de la matriz paramétrica de desmezcla \mathbf{U} , se aplican las operaciones descritas para todas las sub-matrices Δ_q . Una submatriz Δ_q de tamaño $N_g^q \times N_g^q$, con elementos $\Delta_q(id_{x1}, id_{x2})$, se obtiene seleccionando los elementos $\Delta(ch_1, ch_2)$ correspondientes a los canales de mezcla descendente ch_1 y ch_2 asignados al grupo g_q (es decir, $g_q(id_{x1}) = ch_1$ y $g_q(id_{x2}) = ch_2$).

50

El grupo g_q de un tamaño $1 \times N_g^q$ está definido por la serie más pequeña de canales de mezcla descendente con las siguientes propiedades:

55

• Las señales de entrada contenidas en los canales de mezcla descendente del grupo g_q no están contenidas en ningún otro canal de mezcla descendente. Una señal de entrada no está contenida en un canal de mezcla descendente si la correspondiente ganancia de mezcla descendente está dada por el índice de cuantificación más pequeño (Tabla 49 de ISO/IEC 23003-2:2010).

60

• Todas las señales de entrada i contenidas en los canales de mezcla descendente del grupo g_q no están relacionadas con ninguna señal de entrada j contenidas en ningún canal de mezcla descendente de ningún otro grupo (es decir, $\text{bsRelatedTo}[i][j] == 0$).

Los resultados de las operaciones de inversión regularizada $\mathbf{J}_q \approx \mathbf{A}_q^{-1}$ se combinan para obtener \mathbf{J} según:

$$\mathbf{J}(ch_1, ch_2) = \begin{cases} \mathbf{J}_q(id_{x_1}, id_{x_2}), & \text{si } \mathbf{g}_q(id_{x_1}) = ch_1 \text{ y } \mathbf{g}_q(id_{x_2}) = ch_2, \\ 0, & \text{de lo contrario.} \end{cases}$$

5 Regularización de valores singulares 9.5.4.2.6

La operación inversa regularizada $(\cdot)^{inv}$ usada para la matriz diagonal de valores singulares \mathbf{A} se determina de la siguiente manera:

$$\Lambda^{inv} = \lambda_{i,j}^{-1} = \begin{cases} \frac{1}{\lambda_{i,i}}, & \text{si } i = j \text{ y } abs(\lambda_{i,i}) \geq T_{reg}^{\Lambda}, \\ 0, & \text{de lo contrario.} \end{cases}$$

El escalar de regularización relativa T_{reg}^{Λ} se determina utilizando el umbral absoluto T_{reg} y el valor máximo de \mathbf{A} de la siguiente manera:

$$T_{reg}^{\Lambda} = \max_i(abs(\lambda_{i,i}))T_{reg}, \text{ donde } T_{reg} = 10^{-2}.$$

En algunas de las siguientes figuras se muestran señales individuales como obtenidas de diferentes etapas de procesamiento. Esto se hace para una mejor comprensión de la invención y es una posibilidad de realizar la invención, es decir, extraer señales individuales y realizar etapas de procesamiento en estas señales o señales procesadas.

La otra realización consiste en calcular todas las matrices necesarias y aplicar las mismas como último etapa a la señal de audio codificada a fin de obtener la señal de audio decodificada. Esto incluye el cálculo de las diferentes matrices y sus respectivas combinaciones.

Una realización combina ambas maneras.

La Fig. 10 muestra esquemáticamente un aparato 10 para el procesamiento de una pluralidad (en este ejemplo cinco) de objetos de audio de entrada 111 a fin de emitir una representación de los objetos de audio de entrada 111 mediante una señal de audio codificada 100.

Los objetos de audio de entrada 111 son asignados o mezclados en forma descendente con las señales de mezcla descendente 101. En la realización expuesta, cuatro de los cinco objetos de audio de entrada 111 son asignados a dos señales de mezcla descendente 101. Un objeto de audio de entrada 111 solo es asignado a una tercera señal de mezcla descendente 101. De esa manera, cinco objetos de audio de entrada 111 están representados por tres señales de mezcla descendente 101.

A continuación - posiblemente después de algunas etapas de procesamiento no mostradas - estas señales de mezcla descendente 101 se combinan para obtener la señal de audio codificada 100.

Esa señal de audio codificada 100 es alimentada a un aparato de la invención 1, del cual se muestra una realización en la Fig. 11.

De la señal de audio codificada 100 se extraen las tres señales de mezcla descendente 101 (comparar con la Fig. 10).

Las señales de mezcla descendente 101 se agrupan- en el ejemplo expuesto - en dos grupos de señales de mezcla descendente 102.

Como cada señal de mezcla descendente 101 está asociada a un número dado de objetos de audio de entrada, cada grupo de señales de mezcla descendente 102 se refiere a un número dado de objetos de audio de entrada (una expresión correspondiente es el objeto de entrada). Por ende, cada grupo de señales de mezcla descendente 102 está asociado a una serie de objetos de audio de entrada de la pluralidad de objetos de audio de entrada que son codificados por la señal de audio codificada 100 (comparar con la Fig. 10).

El agrupamiento tiene lugar, en la realización mostrada en las siguientes restricciones:

1. Cada objeto de audio de entrada 111 pertenece a sólo una serie de objetos de audio de entrada y, por consiguiente,

a un grupo de señales de mezcla descendente 102.

2. Cada objeto de audio de entrada 111 carece de relación señalizada en la señal de audio codificada con un objeto de audio de entrada 111 perteneciente a una serie diferente asociada a un grupo diferente de señales de mezcla descendente. Esto significa que la señal de audio codificada no tiene esa información, lo que debido a la norma daría como resultado un cómputo combinado de los respectivos objetos de audio de entrada.

3. Se minimiza el número de señales de mezcla descendente 101 dentro de los respectivos grupos 102.

Los grupos de señales de mezcla descendente 102 (en este caso dos) son procesados individualmente a continuación para obtener cinco señales de audio de salida 103 correspondientes a los cinco objetos de audio de entrada 111.

Un grupo de señales de mezcla descendente 102 que está asociado a las dos señales de mezcla descendente 101 que cubren dos pares de objetos de audio de entrada 111 (comparar con la Fig. 10) permite obtener cuatro señales de audio de salida 103.

El otro grupo de señales de mezcla descendente 102 da origen a una señal de salida 103 como señal de mezcla descendente única 101 o este grupo de señales de mezcla descendente 102 (o más precisamente: el grupo de una señal de mezcla descendente) se refiere a un objeto de audio de entrada 111 (comparar con la Fig. 10).

Las cinco señales de audio de salida 103 se combinan para formar una señal de audio decodificada 110 como salida del aparato 1.

En la realización de la Fig. 11 todas las etapas de procesamiento se llevan a cabo individualmente en los grupos de señales de mezcla descendente 102.

La realización del aparato 1 mostrado en la Fig. 12 puede recibir aquí la misma señal de audio codificada 100 que el aparato 1 mostrado en la Fig. 11 y obtenida por un aparato 10 mostrado en la Fig. 10.

De la señal de audio codificada 100 se obtienen las tres señales de mezcla descendente 101 (correspondientes a tres canales de transporte) y se las agrupa en dos grupos de señales de mezcla descendente 102. Estos grupos 102 se procesan individualmente para obtener cinco señales procesadas 104 que corresponden a los cinco objetos de audio de entrada mostrados en la Fig. 10.

En las siguientes etapas, de las cinco señales procesadas 104 se obtienen conjuntamente ocho señales de audio de salida 103, por ej., se las renderiza para ser utilizadas para ocho canales de salida. Las señales de audio de salida 103 se combinan en la señal de audio decodificada 110 que es emitida por el aparato 1. En esta realización, se realiza un procesamiento individual, como así también conjunto, de los grupos de las señales de mezcla descendente 102.

La Fig. 13 muestra algunas etapas de una realización del método inventivo en la cual se decodifica una señal de audio codificada.

En la etapa 200 se extraen las señales de mezcla descendente de la señal de audio codificada. En la etapa siguiente 201, las señales de mezcla descendente son asignadas a grupos de señales de mezcla descendente.

En la etapa 202 se procesa cada grupo de señales de mezcla descendente individualmente a fin de emitir resultados de grupos individuales. El manejo individual de los grupos comprende por lo menos la desmezcla para obtener representaciones de las señales de audio que se combinan mediante la mezcla en forma descendente de los objetos de audio de entrada en el proceso de codificación. En una realización - que no se muestra aquí - al procesamiento individual le sigue un procesamiento conjunto.

En la etapa 203 estos resultados de grupo se combinan en una señal de audio decodificada para constituir la salida.

La Fig. 14 muestra una vez más una realización del aparato 1 en la cual todos las etapas de procesamiento posteriores al agrupamiento de las señales de mezcla descendente 101 de la señal de audio codificada 100 para formar grupos de señales de mezcla descendente 102 se realizan individualmente. El aparato 1 que recibe la señal de audio codificada 100 con las señales de mezcla descendente 101 comprende un agrupador 2 que agrupa las señales de mezcla descendente 101 a fin de emitir los grupos de señales de mezcla descendente 102. Los grupos de señales de mezcla descendente 102 son procesados por un procesador 3 que ejecuta todas las etapas necesarias en forma individual sobre cada grupo de señales de mezcla descendente 102. Los resultados de los grupos individuales del procesamiento de los grupos de señales de mezcla descendente 102 son señales de audio de salida 103 que son combinadas por el combinador 4 a fin de obtener la señal de audio decodificada 110 que ha de constituir la salida del aparato 1.

El aparato 1 mostrado en la Fig. 15 difiere de la realización mostrada en la Fig. 14 después del agrupamiento de las señales de mezcla descendente 101. En el ejemplo, no todas las etapas de procesamiento son ejecutadas individualmente en los grupos de señales de mezcla descendente 102, sino que algunas etapas son ejecutadas en forma conjunta, tomando así en cuenta más de un grupo de señales de mezcla descendente 102.

Debido a esto, el procesador 3 en esta realización está configurado para ejecutar sólo algunos o por lo menos una etapa de procesamiento individualmente. El resultado del procesamiento consiste en señales procesadas 104 que son procesadas conjuntamente por el pos-procesador 5. Las señales de audio de salida obtenidas 103 son combinadas en última instancia por el combinador 4 para dar origen a la señal de audio decodificada 110.

En la Fig. 16 se muestra esquemáticamente un procesador 3 que recibe los grupos de señales de mezcla descendente 102 y produce las señales de audio de salida 103.

El procesador 3 comprende un desmezclador 300 configurado para desmezclar las señales de mezcla descendente 101 de los respectivos grupos de señales de mezcla descendente 102. El desmezclador 300, reconstruye, de esa manera, los objetos de audio de entrada individuales que fueran combinados por el codificador para obtener las respectivas señales de mezcla descendente 101.

Los objetos de audio de entrada reconstruidos o separados son enviados a un renderizador 302. El renderizador 302 está configurado para renderizar las señales de mezcla descendente desmezcladas de los grupos respectivos para una situación de salida de dicha señal de audio decodificada 110 a fin de emitir señales renderizadas 112. Las señales renderizadas 112, por consiguiente, se adaptan al tipo de configuración de reproducción de la señal de audio decodificada. La renderización depende, por ej., del número de parlantes utilizado, de su disposición o del tipo de efectos que se han de obtener mediante la reproducción de la señal de audio decodificada.

Las señales renderizadas 112, Y_{seca} , se someten, además, a un pos-mezclador 303 configurado para ejecutar por lo menos una etapa de descorrelación de dichas señales renderizadas 112 y configurado para combinar los resultados $Y_{húm}$ de la etapa de descorrelación ejecutado con dichas señales renderizadas respectivas 112, Y_{seca} . El pos-mezclador 303 ejecuta, por consiguiente, las etapas para descorrelacionar las señales que se combinarán en una señal de mezcla descendente.

Las señales de audio de salida 103 así obtenidas son sometidas, por último, a un combinador, como se mostró anteriormente.

Para las etapas, el procesador 3 se basa en una calculadora 301 que en este caso es separada de las diferentes unidades del procesador 3 pero que, en una realización alternativa - que no se muestra - es una característica del agrupador 300, el renderizador 302 y el pos-mezclador 303, respectivamente.

Es relevante el hecho de que las matrices, valores, etc, necesarios se calculan individualmente en relación con los respectivos grupos de señales de mezcla descendente 102. Esto implica que, por ej., las matrices a computar son más pequeñas que las matrices empleadas en el estado actual de la técnica. Las matrices tienen tamaños que dependen de un número de objetos de audio de entrada de la respectiva serie de objetos de audio de entrada asociada a los grupos de señales de mezcla descendente y/o de un número de señales de mezcla descendente que pertenecen al respectivo grupo de señales de mezcla descendente.

En el estado de la técnica, la matriz a utilizar para la desmezcla tiene un tamaño del número de objetos de audio de entrada o señales de audio de entrada multiplicado por este número. La invención permite computar una pequeña matriz con un tamaño que depende del número de señales de audio de entrada que pertenecen al respectivo grupo de señales de mezcla descendente.

En la Fig. 17 se explica el propósito de la renderización.

El aparato 1 recibe una señal de audio codificada 100 y la decodifica produciendo una señal de audio decodificada 110.

Esta señal de audio decodificada 110 es reproducida en una situación de salida o configuración de salida 400 específica. En el ejemplo, la señal de audio decodificada 110 debe ser emitida por cinco altavoces 401: Izquierdo, Derecho, Central, Izquierdo Envolvente y Derecho Envolvente. El oyente 402 se encuentra en el centro del escenario 400 frente al altavoz central.

El renderizador incluido en el aparato 1 distribuye las señales de audio reconstruidas que han de ser transferidas a los altavoces individuales 401 y, de esa manera, para distribuir una representación reconstruida de los objetos de audio originales como orígenes de las señales de audio en la situación de salida dada 400.

La renderización, por lo tanto, depende del tipo de situación de salida 400 y del gusto y preferencias individuales del oyente 402.

5 Si bien se han descrito algunos aspectos en el contexto de un aparato, es obvio que estos aspectos también representan una descripción del método correspondiente, en el cual un bloque o dispositivo corresponde a una etapa del método o a una característica de una etapa del método. De manera análoga, los aspectos descritos en el contexto de una etapa del método también representan una descripción de un bloque o elemento correspondiente o de una característica de un aparato correspondiente. Algunos o todos las etapas del método pueden ser ejecutados por medio de (o utilizando) un aparato de hardware, como por ejemplo, un microprocesador, un ordenador programable o un
10 circuito electrónico. En algunas realizaciones, uno o más de las etapas más importantes del método pueden ser ejecutadas por dicho aparato.

Dependiendo de ciertos requisitos de implementación, las formas pueden ser implementadas en hardware o en software o por lo menos parcialmente en software, o por lo menos parcialmente en hardware. La implementación se puede realizar empleando un medio de almacenamiento digital, por ejemplo un disco blando, un DVD, un Blu-Ray, un CD, una ROM, una PROM, una EPROM, una EEPROM o una memoria FLASH, que tiene almacenadas en la misma señales control legibles electrónicamente, que cooperan (o tienen capacidad para cooperar) con un sistema de informático programable de tal manera que se ejecute el método respectivo. Por lo tanto, el medio de almacenamiento digital puede ser legible por ordenador.
15

20 Algunas realizaciones de acuerdo con la invención comprenden un transportador no transitorio de datos que comprende señales de control legibles electrónicamente, con capacidad para cooperar con un sistema informático programable de tal manera que se ejecute uno de los métodos descritos en la presente.

25 En general, las realizaciones de la presente invención pueden ser implementadas en forma de producto de programa informático con un código de programa, donde el código de programa cumple la función de ejecutar uno de los métodos al ejecutarse el programa informático en un ordenador. El código de programa puede ser almacenado, por ejemplo, en un portador legible por una máquina.

30 Otras realizaciones comprenden el programa informático para ejecutar uno de los métodos aquí descritos, almacenado en un portador legible por una máquina.

En otras palabras, una realización del método inventivo es, por lo tanto, un programa informático que tiene un código de programa para realizar uno de los métodos descritos en la presente, cuando el programa informático se ejecuta en un ordenador.
35

Otra realización de los métodos inventivos es, por lo tanto, un soporte de datos (o un medio de almacenamiento digital, o un medio legible por ordenador) que comprende, grabado en el mismo, el programa informático para realizar uno de los métodos descritos en la presente. El portador de datos, el medio de almacenamiento digital o el medio grabado son por lo general tangibles y/o no transitorios.
40

Otra realización del método inventivo es, por lo tanto, un flujo de datos o una secuencia de señales que representan el programa informático para realizar uno de los métodos descritos en la presente. El flujo de datos o la secuencia de señales pueden estar configurados, por ejemplo, para ser transferida a través de una conexión de comunicación de datos, por ejemplo a través de Internet.
45

Otra realización comprende un medio de procesamiento, por ejemplo un ordenador o un dispositivo lógico programable, configurado o adaptado para realizar uno de los métodos descritos en el presente documento.

50 Otra realización comprende un ordenador que tiene instalado en él el programa informático para llevar a cabo uno de los métodos descritos en la presente.

Otra realización según la invención comprende un aparato o un sistema configurado para transferir (por ejemplo, electrónica u ópticamente) un programa informático para realizar uno de los métodos descritos en la presente a un receptor. El receptor puede ser, por ejemplo, un ordenador, un dispositivo móvil, un dispositivo de memoria o similar. El aparato o sistema puede, por ejemplo, comprender un servidor de archivos para transferir el programa informático al receptor.
55

En algunas realizaciones, se puede utilizar un dispositivo lógico programable (por ejemplo, una matriz de puertas programable en campo) para llevar a cabo algunas o todas las funcionalidades de los métodos descritos en la presente. En algunas realizaciones, una matriz de puertas programables en campo puede cooperar con un microprocesador para llevar a cabo uno de los métodos descritos en la presente. Por lo general, los métodos son ejecutados preferentemente por cualquier aparato de hardware.
60

El aparato aquí descrito puede implementarse utilizando un aparato de hardware, o utilizando un ordenador, o utilizando una combinación de un aparato de hardware y un ordenador.

5 Los métodos aquí descritos pueden llevarse a cabo utilizando un aparato de hardware, o utilizando un ordenador, o utilizando una combinación de un aparato de hardware y un ordenador.

Referencias

- [BCC] C. Faller y F. Baumgarte, "Binaural Cue Coding - Part II: Schemes y applications," IEEE Trans. on Speech y Audio Proc., vol. 11, no. 6, nov. 2003.
- [ISS1] M. Parvaix y 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 y Language Processing, 2010.
- [ISS3] A. Liutkus, J. Pinel, R. Badeau, L. Girin, G. Richard: "Informed source separation through spectrogram coding y 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 y Acoustics, 2011.
- [ISS5] S. Zhang y L. Girin: "An Informed Source Separation System for Speech Signals", INTERSPEECH, 2011.
- [ISS6] L. Girin y J. Pinel: "Informed Audio Source Separation from Compressed Linear Stereo Mixtures", AES 42nd International Conference: Semantic Audio, 2011.
- [JSC] C. Faller, "Parametric Joint-Coding of Orígenes de audio", 120a Convención de AES, París, 2006.
- [SAOC] ISO/IEC, "MPEG audio technologies – Part 2: Spatial Audio Object Coding (SAOC)," ISO/IEC JTC1/SC29/WG11 (MPEG) International Standard 23003-2.
- [SAOC1] J. Herre, S. Disch, J. Hilpert, O. Hellmuth: "From SAC To SAOC - Recent Developments in Parametric Coding of Spatial Audio", 22a Conferencia Regional de AES del Reino Unido, Cambridge, UK, abril 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", 124a Convención de AES, Ámsterdam 2008.
- [SAOC3D] ISO/IEC, JTC1/SC29/WG11 N14747, Text of ISO/MPEG 23008-3/DIS 3D Audio, Sapporo, julio de 2014.
- [SAOC3D2] J. Herre, J. Hilpert, A. Kuntz, y J. Plogsties, "MPEG-H Audio - The new standard for universal spatial / 3D audio coding," 137a Convención de AES, Los Angeles, 2011.

REIVINDICACIONES

1. El decodificador SAOC 3D (1) para procesar una señal de audio codificada (100) que comprende una pluralidad de señales de mezcla descendente (101) asociadas a una pluralidad de objetos de audio de entrada (111) y parámetros de objeto (E), que comprende:
un agrupador (2) configurado para agrupar dicha pluralidad de señales de mezcla descendente (101) en una pluralidad de grupos de señales de mezcla descendente (102) basándose en la información contenida en dicha señal de audio codificada (100), en la que cada grupo de señales de mezcla descendente (102) está asociado a un conjunto de objetos de audio de entrada (111) de dicha pluralidad de objetos de audio de entrada (111), en la que dicho agrupador (2) está configurado para
agrupar dicha pluralidad de señales de mezcla descendente (101) en dicha pluralidad de grupos de señales de mezcla descendente (102) de modo que cada objeto de audio de entrada (111) de dicha pluralidad de objetos de audio de entrada (111) pertenezca a un solo grupo de objetos de audio de entrada (111), y
agrupar dicha pluralidad de señales de mezcla descendente (101) en dicha pluralidad de grupos de señales de mezcla descendente (102) de modo que cada objeto de audio de entrada (111) de cada conjunto de objetos de audio de entrada (111) esté libre de una relación señalada en la señal de audio codificada (100) con otros objetos de audio de entrada (111) o tenga una relación señalada en la señal de audio codificada (100) sólo con al menos un objeto de audio de entrada (111) perteneciente al mismo conjunto de objetos de audio de entrada (111),
un procesador (3) configurado para realizar al menos una etapa de procesamiento individual sobre los parámetros de objeto (E_k) de cada conjunto de objetos de audio de entrada (111) con el fin de proporcionar resultados de grupo (103, 104), en el que dicho procesador (3) está configurado
para determinar una matriz de mezcla descendente (D_k) individual para cada grupo de señales de mezcla descendente (102),
para determinar una matriz de covarianza de grupo individual (E_k) para cada grupo de señales de mezcla descendente (102),
determinar una matriz de covarianza de mezcla descendente de grupo individual (Δ_k) para cada grupo de señales de mezcla descendente (102) basándose en la matriz de mezcla descendente individual (D_k) y la matriz de covarianza de grupo individual (E_k), y
para determinar una matriz de grupo inversa regularizada individual (J_k) para cada grupo de señales de mezcla descendente (102), y
un combinador (4) configurado para combinar dichos resultados de grupo (103, 104) a fin de proporcionar una señal de audio descodificada (110), en el que dicho combinador (4) está configurado para combinar las matrices de grupo inversas regularizadas individuales (J_k) para obtener una matriz de grupo inversa regularizada general (J),
en el que el decodificador SAOC (1) está configurado para
calcular una señal seca (Y_{seca}) a partir de la pluralidad de señales de mezcla descendente (101), una matriz de renderización (R) y una matriz de desmezcla paramétrica general (U), derivándose la matriz de desmezcla paramétrica general (U) de la matriz de grupo inversa regularizada (J), una matriz de desmezcla general (D) y una matriz de covarianza general (E) derivada de los parámetros del objeto,
para calcular una señal húmeda ($Y_{húm}$) utilizando una descorrelación de la señal seca (Y_{seca}), y
para calcular la señal de audio decodificada (110) a partir de la señal seca (Y_{seca}) y la señal húmeda ($Y_{húm}$) utilizando una matriz de mezcla (P).
2. El decodificador 3D SAOC de la reivindicación 1, en el que dicho agrupador (2) está configurado para agrupar dicha pluralidad de señales de mezcla descendente (101) en dicha pluralidad de grupos de señales de mezcla descendente (102) mientras se minimiza un número de señales de mezcla descendente (101) dentro de cada grupo de señales de mezcla descendente (102).
3. El decodificador 3D SAOC de la reivindicación 1 o 2, en el que dicho agrupador (2) está configurado para agrupar dicha pluralidad de señales de mezcla descendente (101) en dicha pluralidad de grupos de señales

de mezcla descendente (102) de modo que solo una única señal de mezcla descendente (101) pertenezca a un grupo de señales de mezcla descendente (102).

4. El decodificador SAOC 3D de una cualquiera de las reivindicaciones 1 a 3, en el que dicho agrupador (2) está configurado para agrupar dicha pluralidad de señales de mezcla descendente (101) en dicha pluralidad de grupos de señales de mezcla descendente (102) aplicando al menos las siguientes etapas:

detectar si una señal de mezcla descendente (101) está asignada a un grupo existente de señales de mezcla descendente(102);

detectar si por lo menos un objeto de audio de entrada (111) de la pluralidad de objetos de audio de entrada (111) asociados a la señal de mezcla descendente (101) es parte de una serie de objetos de audio de entrada (111) asociados a un grupo existente de señales de mezcla descendente (102);

asignar la señal de mezcla descendente (101) a un nuevo grupo de señales de mezcla descendente (102) en caso de que la señal de mezcla descendente (101) esté libre de una asignación a un grupo existente de señales de mezcla descendente (102) y en caso de que todos los objetos de audio de entrada (111) de la pluralidad de objetos de audio de entrada asociados con la señal de mezcla descendente (101) estén libres de una asociación con un grupo existente de señales de mezcla descendente (102); y

combinar la señal de mezcla descendente (101) con un grupo existente de señales de mezcla descendente (102) ya sea en caso de que la señal de mezcla descendente (101) esté asignada al grupo existente de señales de mezcla descendente (102) o en caso de que por lo menos un objeto de audio de entrada (111) de la pluralidad de objetos de audio de entrada asociados a la señal de mezcla descendente (101) esté asociado al grupo existente de señales de mezcla descendente(102).

5. El decodificador SAOC 3D de cualquiera de las reivindicaciones 1 a 4, en el que dicho procesador (3) está configurado para computar para cada grupo de señales de mezcla descendente un umbral individual basado en un valor de energía máximo dentro del respectivo grupo de señales de mezcla descendente (102) al determinar la matriz de grupo inversa regularizada individual (J_k) para cada grupo de señales de mezcla descendente (102).

6. El método del decodificador SAOC 3D para procesar una señal de audio codificada (100) que comprende una pluralidad de señales de mezcla descendente (101) asociadas a una pluralidad de objetos de audio de entrada (111) y parámetros del objeto, dicho método comprende:

agrupar dicha pluralidad de señales de mezcla descendente (101) en una pluralidad de grupos de señales de mezcla descendente (102) basándose en la información contenida en dicha señal de audio codificada (100), en donde cada grupo de señales de mezcla descendente (102) está asociado a un conjunto de objetos de audio de entrada (111) de dicha pluralidad de objetos de audio de entrada (111), en donde se lleva a cabo la agrupación de dicha pluralidad de señales de mezcla descendente (101) en dicha pluralidad de grupos de señales de mezcla descendente (102)

de modo que cada objeto de audio de entrada (111) de dicha pluralidad de objetos de audio de entrada (111) pertenezca a un solo conjunto de objetos de audio de entrada (111), y

de modo que cada objeto de audio de entrada (111) de cada conjunto de objetos de audio de entrada esté libre de una relación señalada en la señal de audio codificada (100) con otros objetos de audio de entrada (111) o tenga una relación señalada en la señal de audio codificada (100) sólo con al menos un objeto de audio de entrada (111) perteneciente al mismo conjunto de objetos de audio de entrada (111);

realizar al menos un paso de procesamiento individualmente en los parámetros de objeto (E_k) de cada conjunto de objetos de audio de entrada (111) para proporcionar resultados de grupo (103, 104), en el que el realizar al menos una etapa de procesamiento individualmente en los parámetros de objeto (E_k) de cada conjunto de objetos de audio de entrada (111) comprende:

determinar una matriz de mezcla descendente (D_k) individual para cada grupo de señales de mezcla descendente (102),

determinar una matriz de covarianza de grupo individual (E_k) para cada grupo de señales de mezcla descendente (102),

determinar una matriz de covarianza de mezcla descendente de grupo individual (Δ_k) para cada grupo de señales de mezcla descendente (102) basándose en la matriz de mezcla descendente individual (D_k) y la

- matriz de covarianza de grupo individual (E_k), y
- determinar una matriz de grupo inversa regularizada individual (J_k) para cada grupo de señales de mezcla descendente (102), y
- 5 combinar dichos resultados de grupo (103, 104) a fin de proporcionar una señal de audio decodificada (110), en el que la combinación comprende combinar las matrices inversas de grupo regularizadas individuales (J_k) para obtener una matriz inversa de grupo regularizada general (J),
- 10 en el que el método comprende:
- calcular una señal seca (Y_{seca}) a partir de la pluralidad de señales de mezcla descendente (101), una matriz de renderización (R) y una matriz de desmezcla paramétrica general (U), derivándose la matriz de desmezcla paramétrica general (U) de la matriz de grupo inversa general regularizada (J), una matriz de desmezcla general (D) y una matriz de covarianza general (E) derivada de los parámetros del objeto,
- 15 calcular una señal húmeda ($Y_{húm}$) utilizando una descorrelación de la señal seca (Y_{seca}), y
- 20 calcular la señal de audio decodificada (110) a partir de la señal seca (Y_{seca}) y la señal húmeda ($Y_{húm}$) utilizando una matriz de mezcla (P).
7. El programa informático que comprende instrucciones que, cuando el programa es ejecutado por un ordenador o un procesador, hacen que el ordenador o el procesador lleven a cabo el método de decodificación SAOC 3D de la reivindicación 6.
- 25

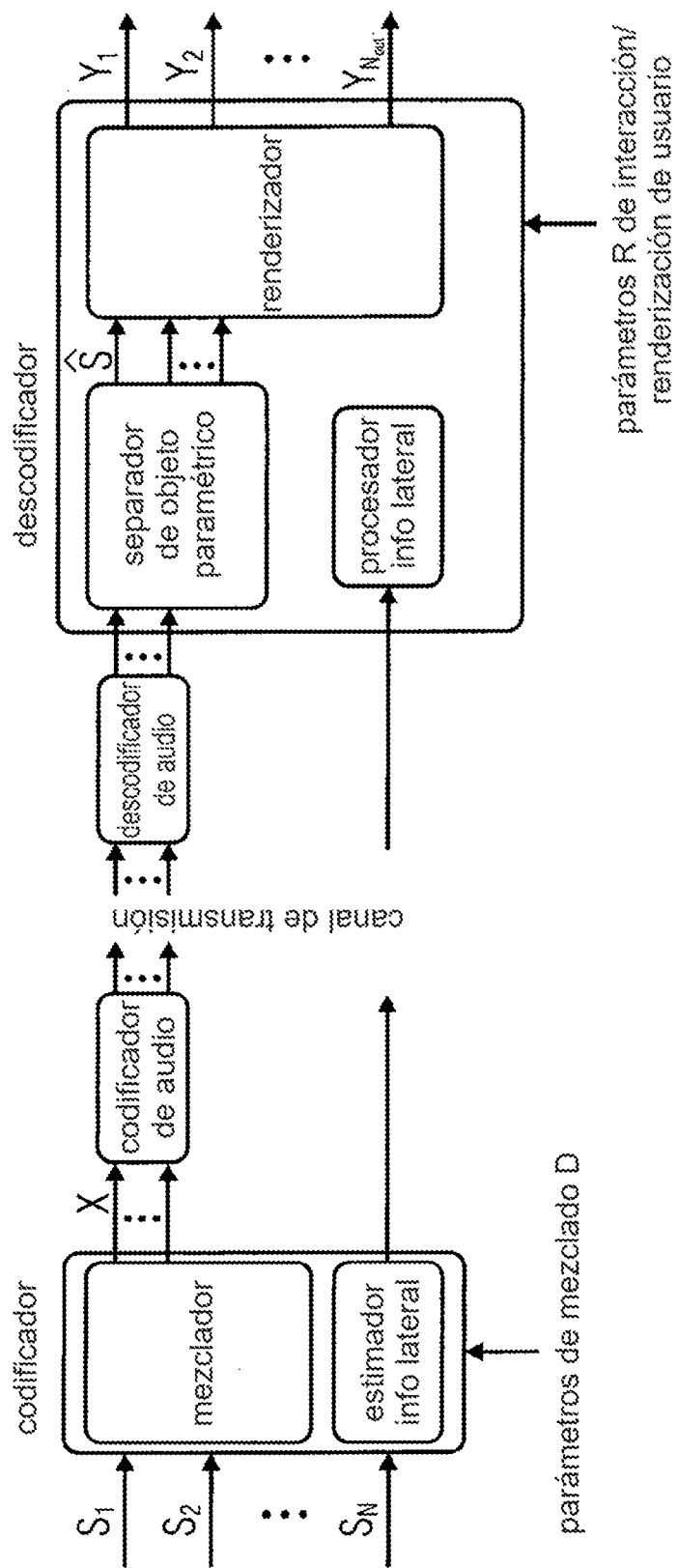


Fig. 1

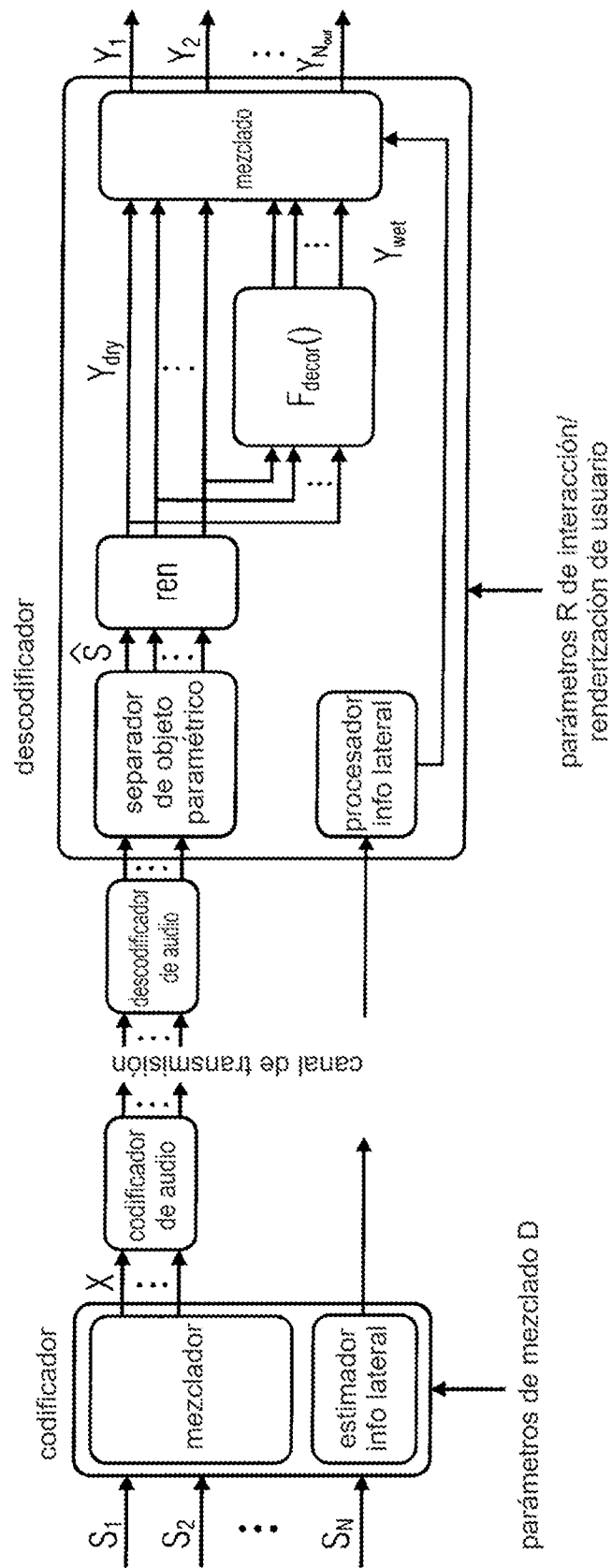


Fig. 2

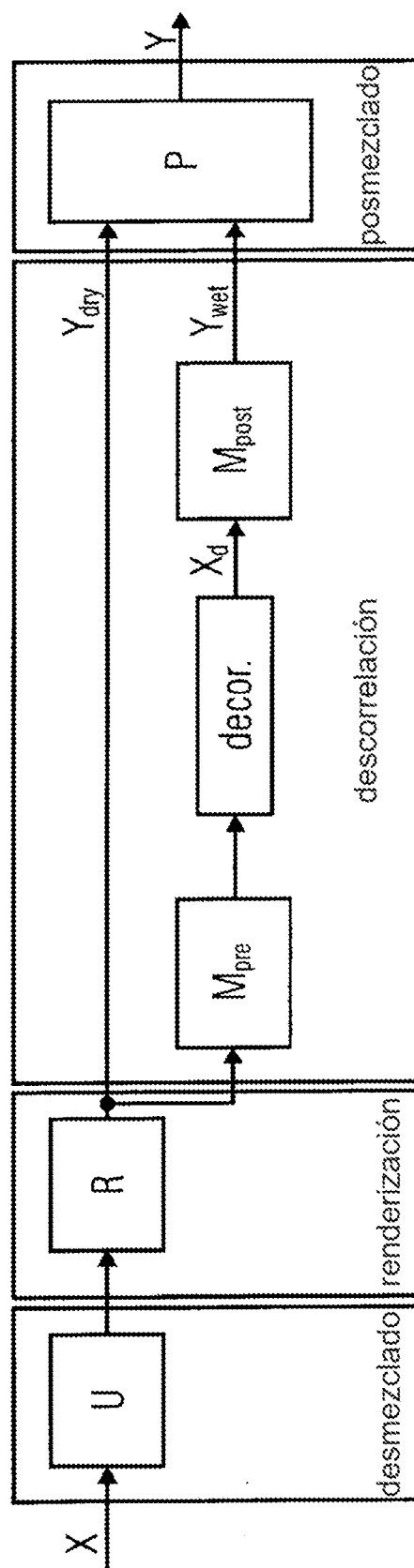


Fig. 3

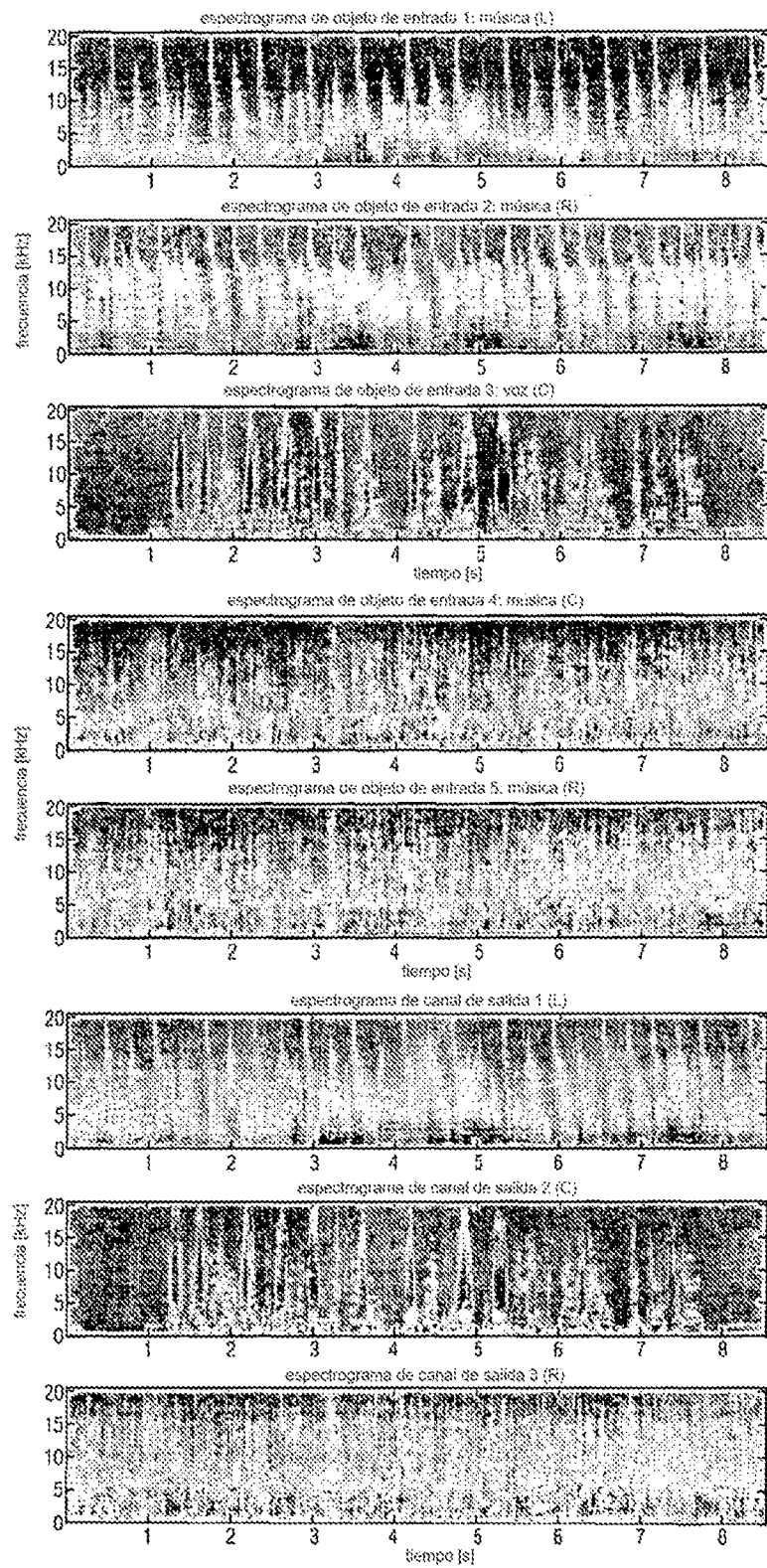


Fig. 4

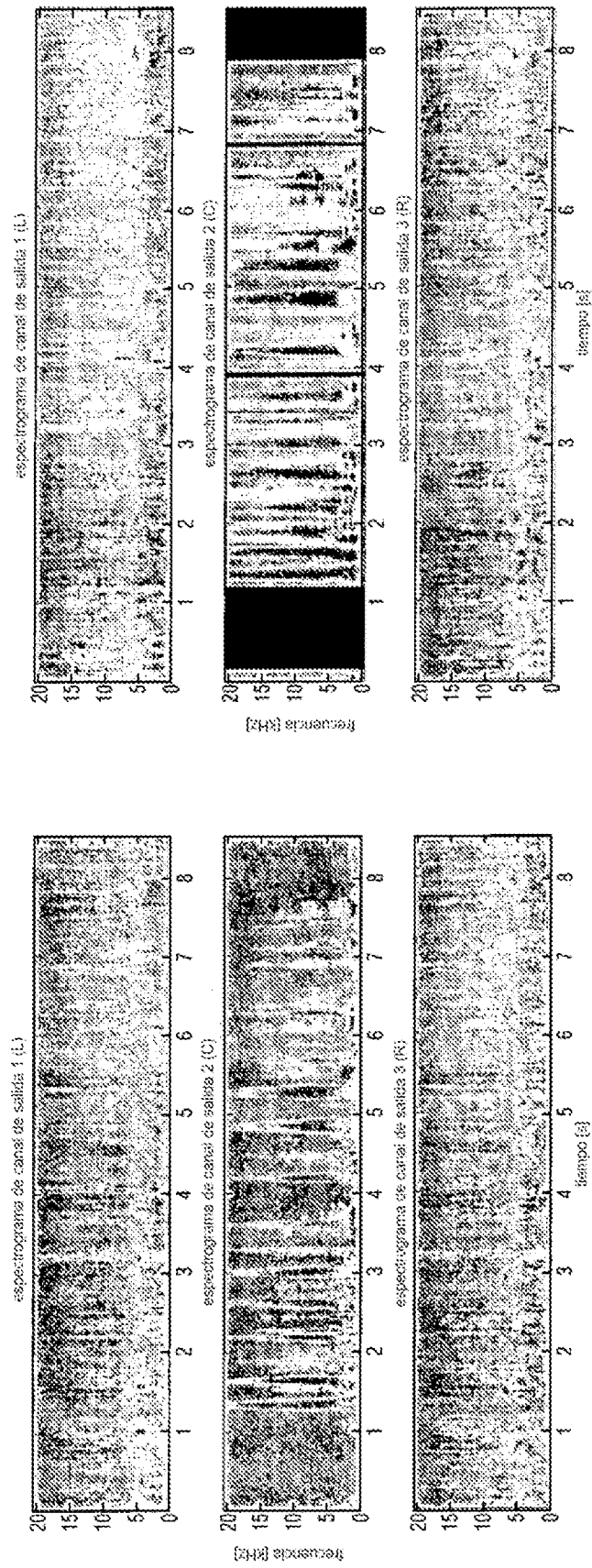


Fig. 5

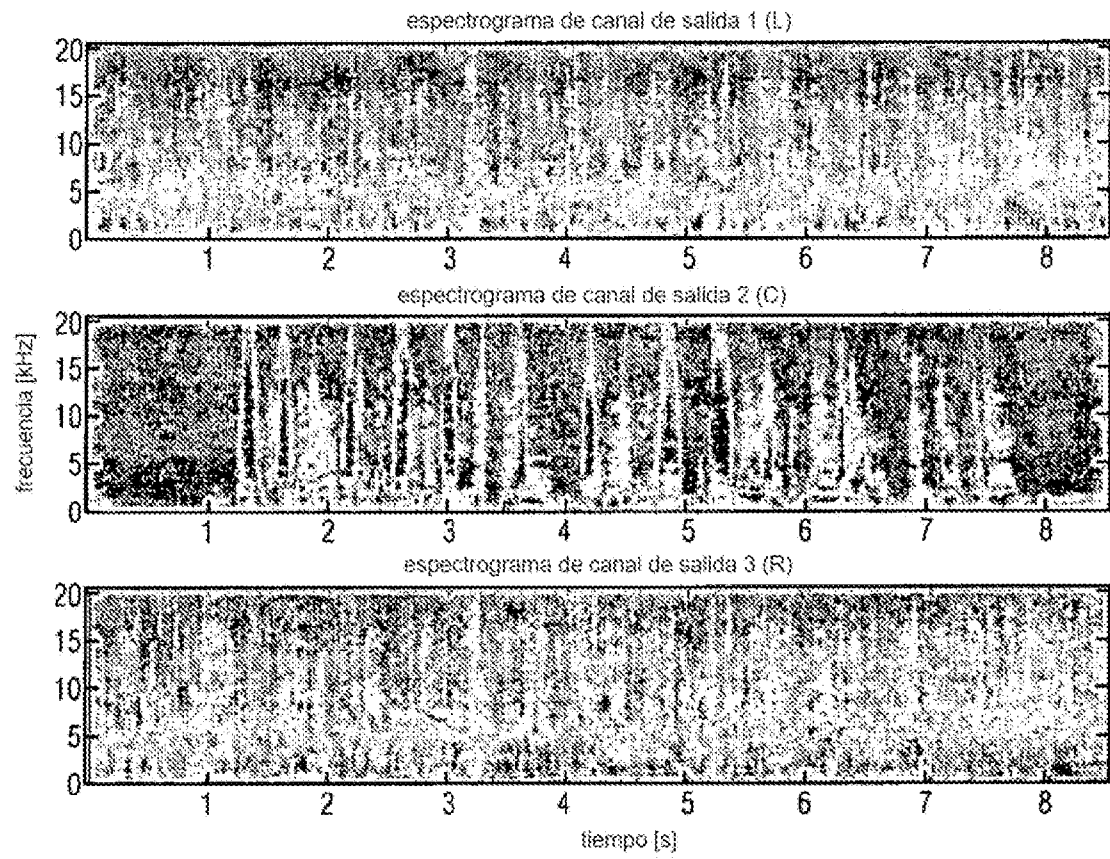


Fig. 6

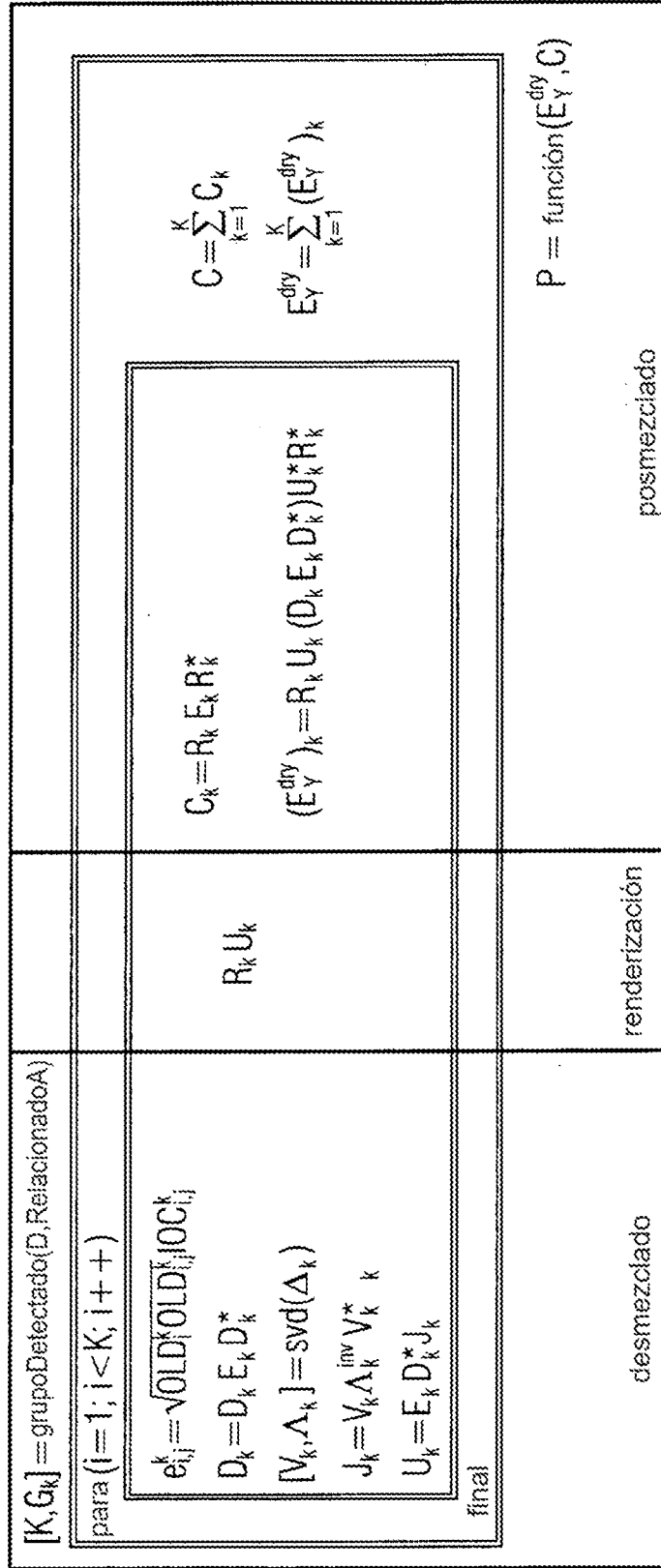


Fig. 8

```

#define SAOC_MAX_OBJECTS          128
#define SAOC_MAX_DMx_CHANNELS    32

#define SAOC_MAX_GROUP_OBJECTS    32 /* número máximo de objetos en un grupo */
#define SAOC_MAX_GROUP_DMx_CHANNELS 16 /* número máximo de canales de mezcla descendente en un grupo */

typedef struct saocCoreGroupList {
    int numDmx;
    int numObj;
    int dmxList[SAOC_MAX_DMx_CHANNELS];
    int objList[SAOC_MAX_OBJECTS];
} saocCoreGroupList;

typedef struct saocCoreGroup {
    int numGroups;
    saocCoreGroupList saocGroup[SAOC_MAX_DMx_CHANNELS];
} saocCoreGroup;

static int getSaocCoreGroups(int CrossCov[SAOC_MAX_DMx_CHANNELS][SAOC_MAX_OBJECTS],
                             int numObjects,
                             int numDmxChannels,
                             saocCoreGroup *groupInfo)
{
    int iGr, iObj, jObj, iDmx, jDmx, indexObj, indexChan, groupIndex, foundGroup;

    int channelsFlag[SAOC_MAX_DMx_CHANNELS];
    int objectsFlag[SAOC_MAX_OBJECTS];

    /* Init empty grouping info */
    for (iGr = 0; iGr < SAOC_MAX_DMx_CHANNELS; iGr++) {
        for (iObj = 0; iObj < numObjects; iObj++) {
            groupInfo->saocGroup[iGr].objList[iObj] = -1;
        }
        for (iDmx = 0; iDmx < SAOC_MAX_DMx_CHANNELS; iDmx++) {
            groupInfo->saocGroup[iGr].dmxList[iDmx] = -1;
        }
        groupInfo->saocGroup[iGr].numDmx = 0;
        groupInfo->saocGroup[iGr].numObj = 0;
    }
    groupInfo->numGroups = 0;

    for (iObj = 0; iObj < numObjects; iObj++) {
        objectsFlag[iObj] = -1;
    }
    for (jDmx = 0; jDmx < numDmxChannels; jDmx++) {
        channelsFlag[jDmx] = jDmx;
    }
}

```

Fig. 9 (1)

```

for (iDmx = 0; iDmx < numDmxChannels; iDmx++) {
    for (iObj = 0; iObj < numObjects; iObj++) {
        if (CrossCov[iDmx][iObj] > 0) {
            if (objectsFlag[iObj] == -1) {
                objectsFlag[iObj] = iDmx;
            } else {
                foundGroup = objectsFlag[iObj];
                for (jObj = 0; jObj < numObjects; jObj++) {
                    if (objectsFlag[jObj] == foundGroup) {
                        objectsFlag[jObj] = iDmx;
                    }
                }
            }
        }
        for (jDmx = 0; jDmx < iDmx; jDmx++) {
            if (channelsFlag[jDmx] == foundGroup) {
                channelsFlag[jDmx] = iDmx;
            }
        }
    }
}

for (iObj = 0; iObj < numObjects; iObj++) {
    iGr = groupInfo->numGroups;
    if (objectsFlag[iObj] > -1) {
        groupIndex = objectsFlag[iObj];
        for (jObj = 0; jObj < numObjects; jObj++) {
            if (objectsFlag[jObj] == groupIndex) {
                indexObj = groupInfo->saocGroup[iGr].numObj++;
                groupInfo->saocGroup[iGr].objList[indexObj] = jObj;
                objectsFlag[jObj] = -1;
            }
        }
        for (iDmx = 0; iDmx < numDmxChannels; iDmx++) {
            if (channelsFlag[iDmx] == groupIndex) {
                indexChan = groupInfo->saocGroup[iGr].numDmx++;
                groupInfo->saocGroup[iGr].dmxList[indexChan] = iDmx;
                channelsFlag[iDmx] = -1;
            }
        }
        groupInfo->numGroups++;
    }
}

```

/* buscar para objetos no asignados */

/* iobjeto no asignado a ningún grupo */

/* asignar iobjeto a grupo iDmx */

/* iobjeto ya asignado a un grupo */

/* asignar todos los objetos de este grupo existente a uno nuevo */

/* asignar todos los canales de este grupo existente a uno nuevo */

/* objeto no asignado a ningún grupo */

/* asignar todos los objetos de este grupo en la lista de objetos */

/* marcar objeto como asignado a una lista de objetos */

/* asignar todos los canales de este grupo en la lista */

/* marcar canal como asignado a una lista de canales */

Fig. 9 (2)

```

for (iGr = 0; iGr < groupInfo->numGroups; iGr++) {
    assert(groupInfo->saocGroup[iGr].numDmx < SAOC_MAX_GROUP_DMX_CHANNELS);
    assert(groupInfo->saocGroup[iGr].numObj < SAOC_MAX_GROUP_OBJECTS);
}

return 0;
}

static void getSaocDmxPositions(float Dmx[SAOC_MAX_DMX_CHANNELS][SAOC_MAX_OBJECTS],
                                int RelatedToMatrix[SAOC_MAX_OBJECTS][SAOC_MAX_OBJECTS],
                                int CrossCov[SAOC_MAX_DMX_CHANNELS][SAOC_MAX_OBJECTS],
                                int numObjects,
                                int firstObject,
                                int numDmxChannels,
                                int firstDmxChannel)
{
    int ch, dCh, obj, obj2, objIdx, objIdx2;
    int D[SAOC_MAX_DMX_CHANNELS][SAOC_MAX_OBJECTS] = {{0}};
    float minVal = (float)pow(10.0f, -150.0f/20.0f) + SAOC_EPSILON;

    for (ch=0; ch<numDmxChannels; ch++) {
        dCh = firstDmxChannel + ch;
        for (obj=0; obj<numObjects; obj++) {
            D[ch][obj] = (minVal < Dmx[dCh][firstObject + obj] ? 1 : 0);
        }
    }

    for (ch=0; ch<numDmxChannels; ch++) {
        dCh = firstDmxChannel + ch;
        for (obj=0; obj<numObjects; obj++) {
            objIdx = firstObject + obj;
            CrossCov[ch][obj] = 0.0f;
            for (obj2=0; obj2<numObjects; obj2++) {
                objIdx2 = firstObject + obj2;
                CrossCov[ch][obj] += D[ch][obj] * RelatedToMatrix[objIdx][objIdx2];
            }
        }
    }
}

static int getSaocCoreGrouping(float Dmx[][SAOC_MAX_OBJECTS],
                                int RelatedToMatrix[SAOC_MAX_OBJECTS][SAOC_MAX_OBJECTS],
                                int numObjects,
                                int numDmxChannels,
                                saocCoreGroup *groupInfo)
{
    int CrossCov[SAOC_MAX_DMX_CHANNELS][SAOC_MAX_OBJECTS] = {{0}};

    getSaocDmxPositions(Dmx, RelatedToMatrix, CrossCov, numObjects, 0, numDmxChannels, 0);
    getSaocCoreGroups(CrossCov, numObjects, numDmxChannels, groupInfo);

    return 0;
}

```

Fig. 9 (3)

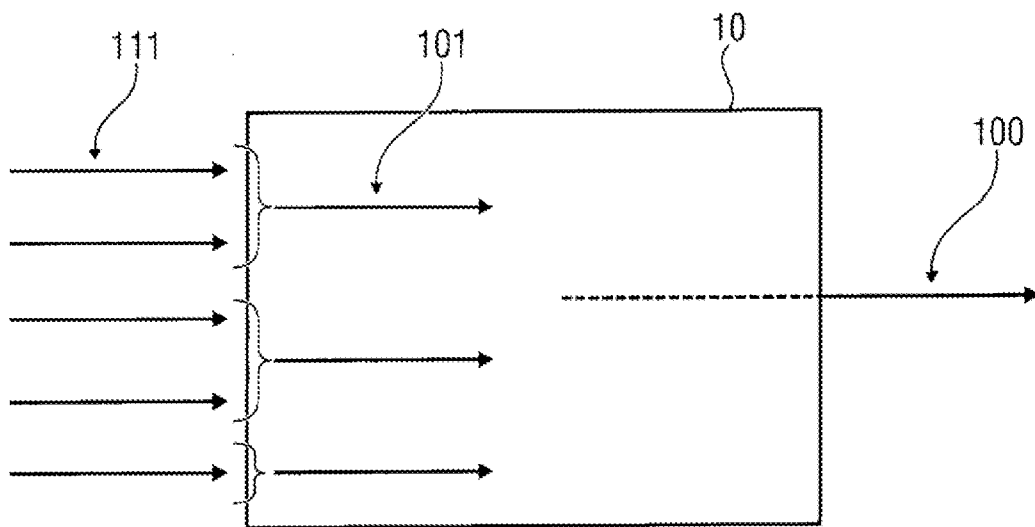


Fig. 10

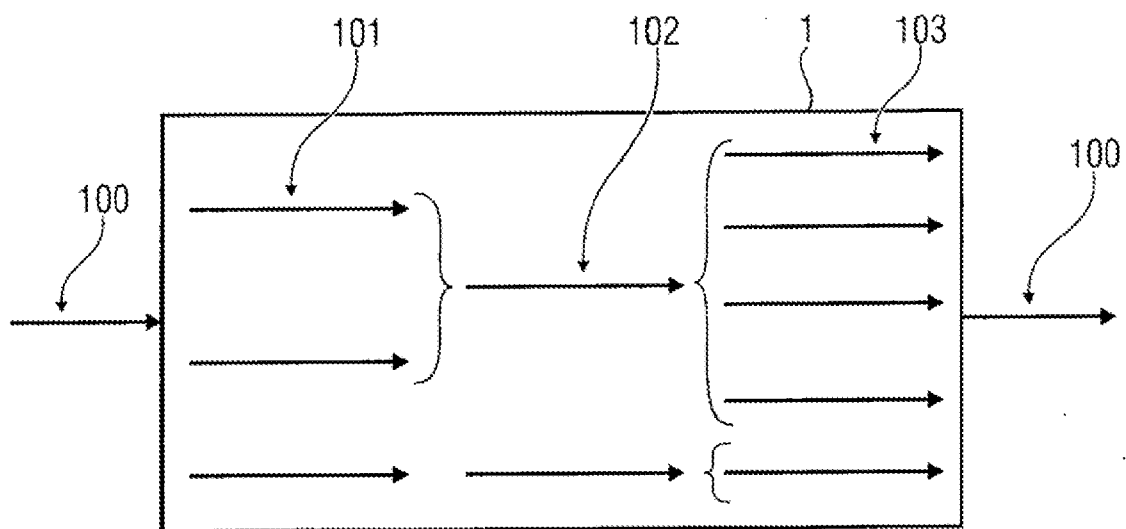


Fig. 11

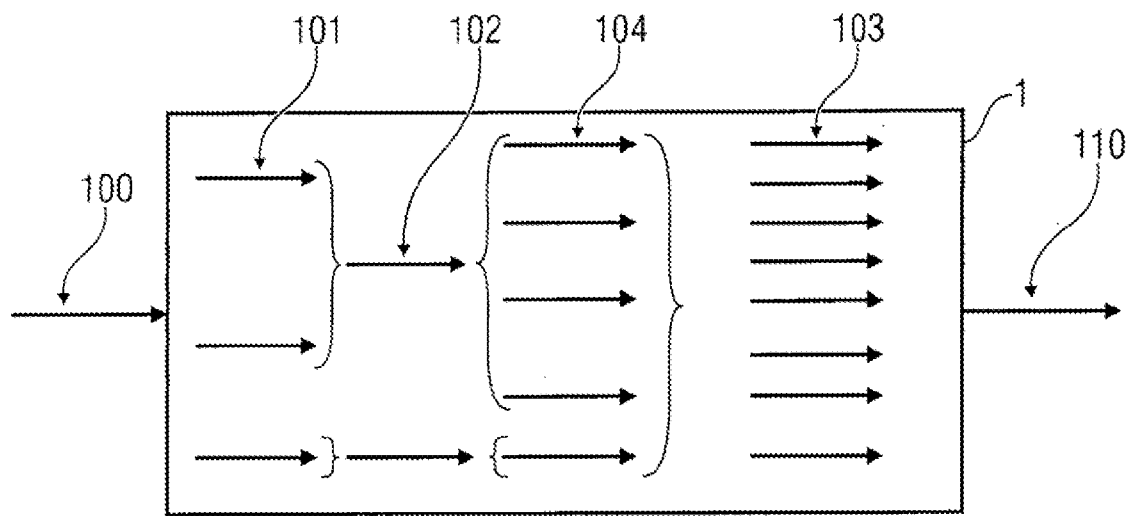


Fig. 12

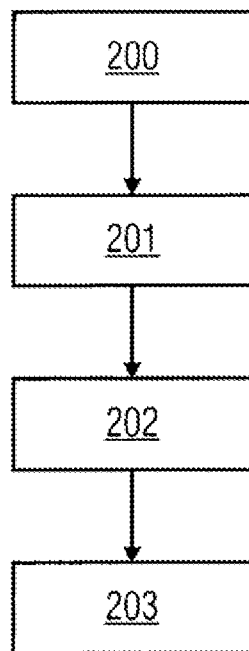


Fig. 13

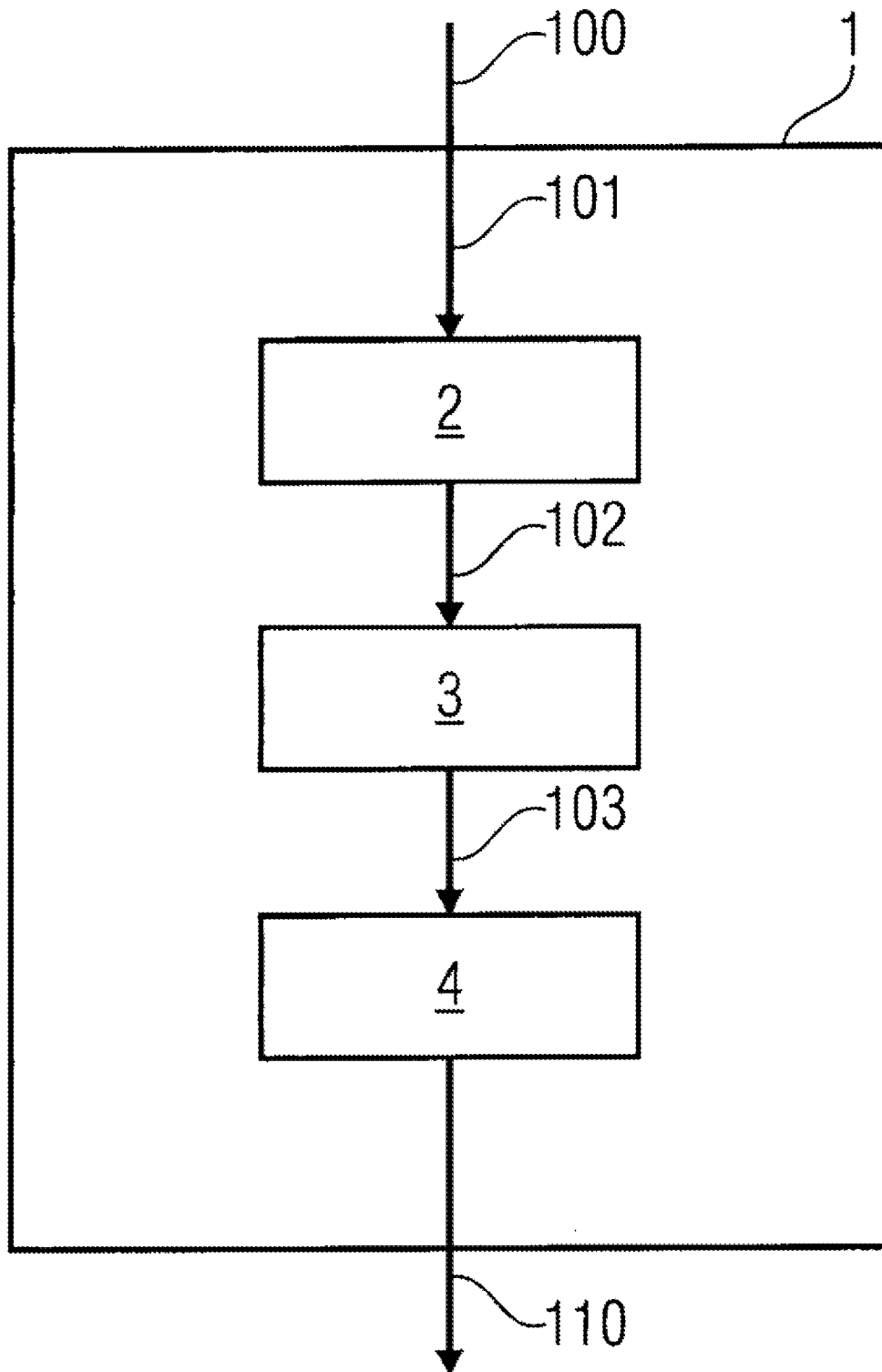


Fig. 14

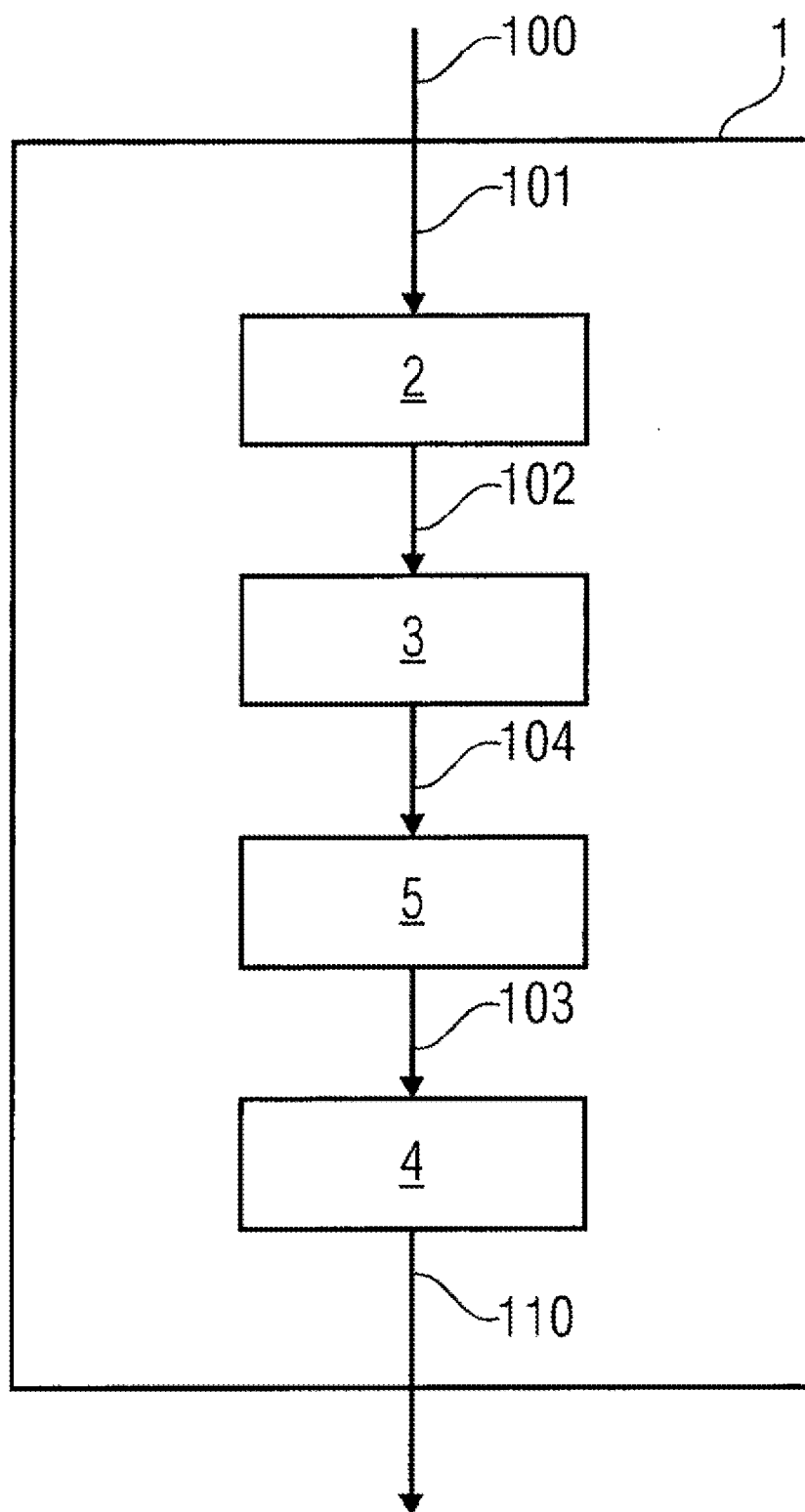


Fig. 15

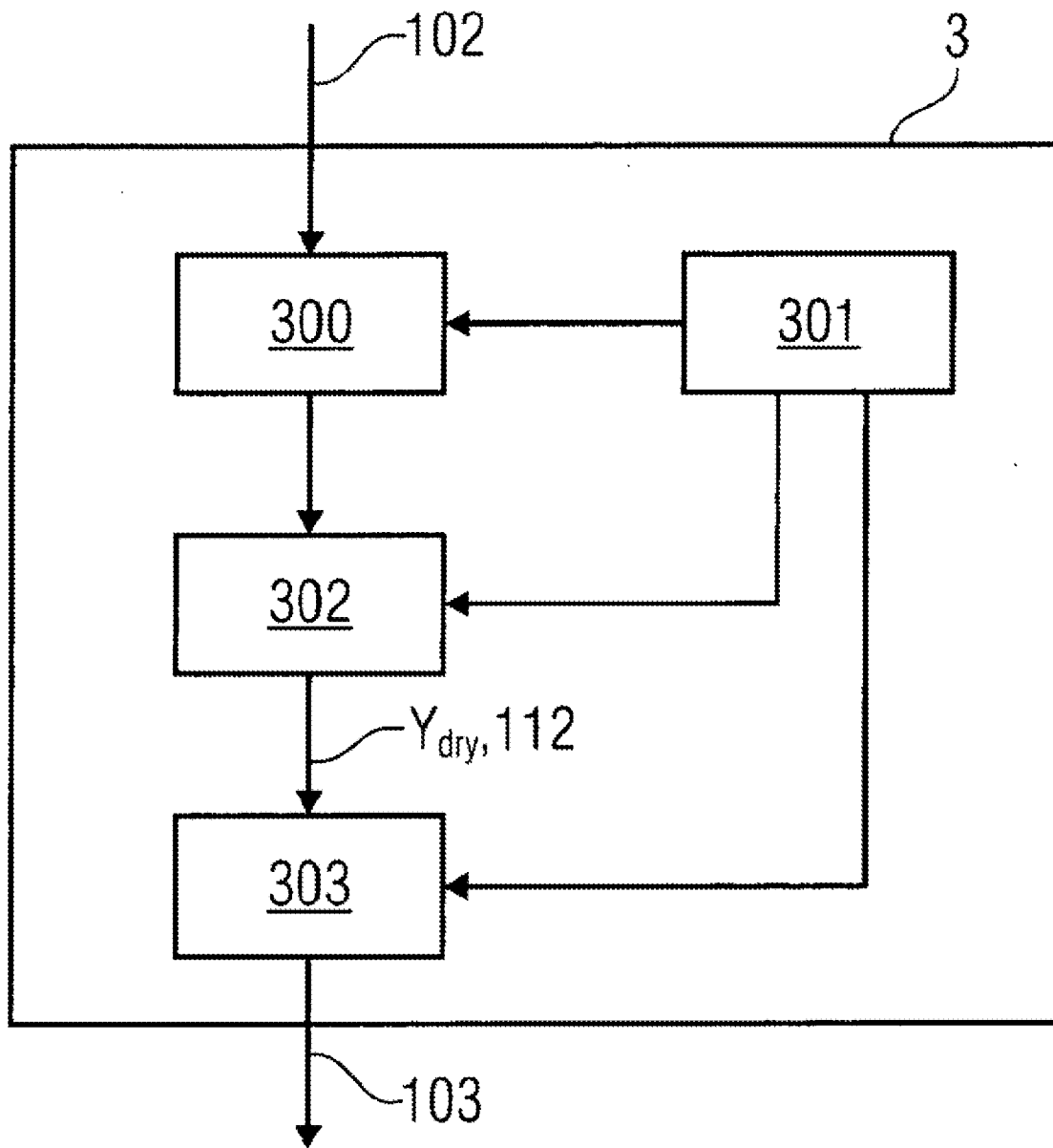


Fig. 16

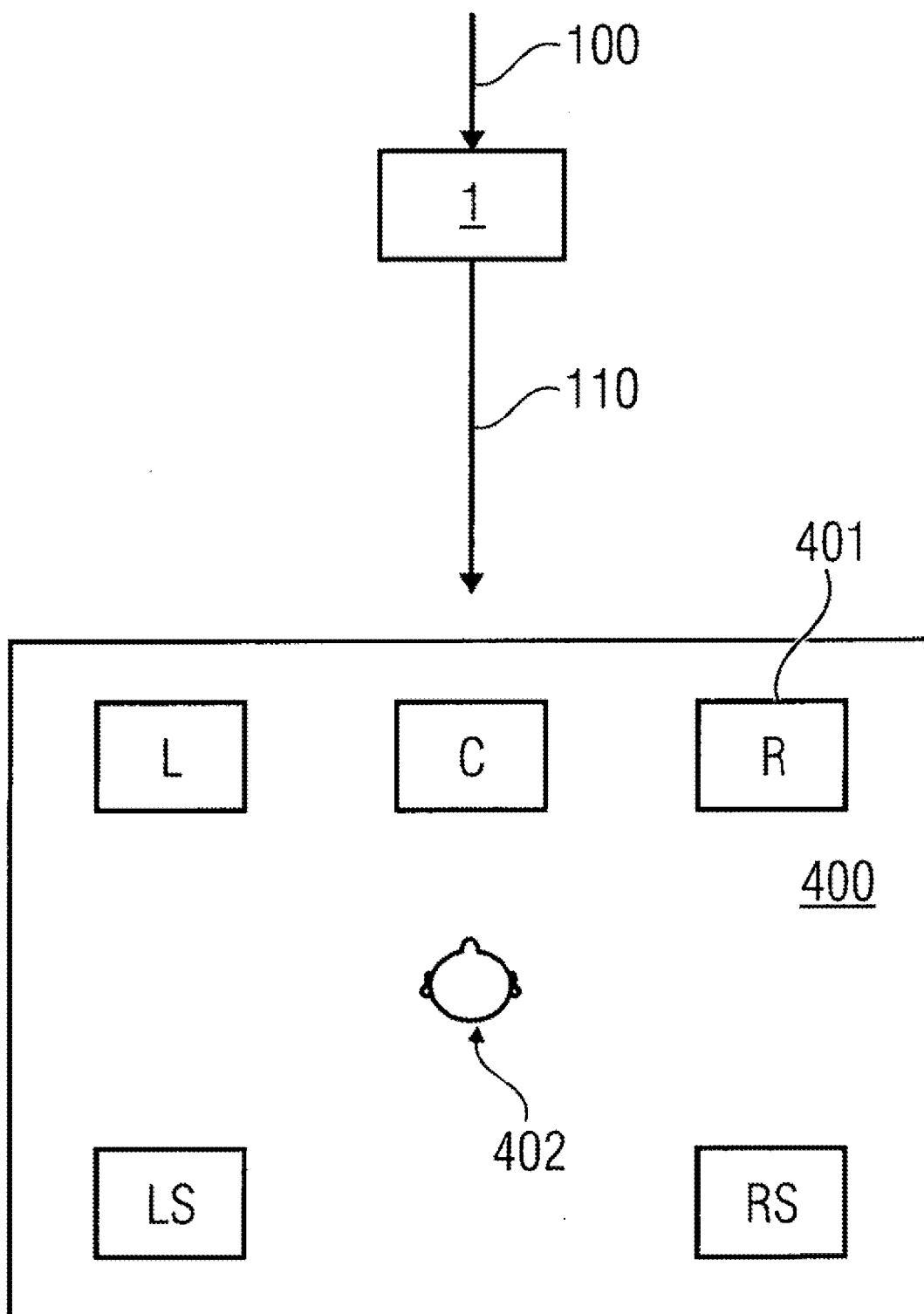


Fig. 17