

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 378 734**

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

12

### TRADUCCIÓN DE PATENTE EUROPEA

T3

- 96 Número de solicitud europea: **09004406 .6**
- 96 Fecha de presentación: **05.10.2007**
- 97 Número de publicación de la solicitud: **2068307**
- 97 Fecha de publicación de la solicitud: **10.06.2009**

54 Título: **Codificación mejorada y representación de parámetros de codificación de objetos de mezcla descendente multicanal**

30 Prioridad:  
**16.10.2006 US 829649 P**

45 Fecha de publicación de la mención BOPI:  
**17.04.2012**

45 Fecha de la publicación del folleto de la patente:  
**17.04.2012**

73 Titular/es:  
**Dolby International AB  
Apollo Building, 3E Herikerbergweg 1-35  
1101 CN Amsterdam Zuid-oost , NL**

72 Inventor/es:  
**Engdegard, Jonas;  
Villemoes, Lars;  
Purnhagen, Heiko y  
Resch, Barbara**

74 Agente/Representante:  
**Arizti Acha, Monica**

**ES 2 378 734 T3**

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

## DESCRIPCIÓN

Codificación mejorada y representación de parámetros de codificación de objetos de mezcla descendente multicanal

## CAMPO TÉCNICO

5 La presente invención se refiere a la decodificación de múltiples objetos a partir de una señal multiobjeto codificada basándose en una mezcla descendente multicanal disponible y datos de control adicionales.

## ANTECEDENTES DE LA INVENCION

10 El desarrollo reciente en audio facilita la recreación de una representación multicanal de una señal de audio basándose en una señal estéreo (o mono) y datos de control correspondientes. Estos métodos de codificación envolvente paramétrica habitualmente comprenden una parametrización. Un decodificador de audio multicanal paramétrico, (por ejemplo, el decodificador de *MPEG Surround* definido en la norma ISO/IEC 23003-1 [1], [2]), reconstruye  $M$  canales basándose en  $K$  canales transmitidos, donde  $M > K$ , mediante el uso de los datos de control adicionales. Los datos de control consisten en una parametrización de la señal multicanal basada en IID (*Inter channel Intensity Difference*; diferencia de intensidad entre canales) e ICC (*Inter Channel Coherence*; coherencia entre canales). Estos parámetros se extraen normalmente en la fase de codificación y describen relaciones de potencia y correlación entre pares de canales usados en el proceso de mezcla ascendente. Usar un esquema de codificación de este tipo permite la codificación a una tasa de transmisión de datos significativamente inferior a la de la transmisión de todos los  $M$  canales, haciendo la codificación muy eficaz mientras al mismo tiempo garantiza una compatibilidad tanto con dispositivos de  $K$  canales como con dispositivos de  $M$  canales.

20 Un sistema de codificación muy relacionado es el codificador de objetos de audio correspondiente [3], [4] en el que varios objetos de audio se mezclan de manera descendente en el codificador y más tarde se mezclan de manera ascendente de manera guiada mediante los datos de control. El proceso de mezcla ascendente también puede considerarse como una separación de los objetos que se mezclan en la mezcla descendente. La señal de mezcla ascendente resultante puede reproducirse en uno o más canales de reproducción. De manera más precisa, [3,4] presenta un método para sintetizar canales de audio a partir de una mezcla descendente (denominada señal suma), información estadística acerca de los objetos fuente y datos que describen el formato de salida deseado. En casos de que se usen varias señales de mezcla descendente, estas señales de mezcla descendente consisten en diferentes subconjuntos de los objetos, y la mezcla ascendente se realiza para cada canal de mezcla descendente de manera individual.

30 En el nuevo método introducimos un método en el que la mezcla ascendente se realiza conjuntamente para todos los canales de mezcla descendente. Los métodos de codificación de objetos, antes de la presente invención, no presentaban una solución para la decodificación conjunta de una mezcla descendente con más de un canal.

**Referencias:**

35 [1] L. Villemoes, J. Herre, J. Breebaart, G. Hotho, S. Disch, H. Pumhagen y K. Kjörling, "MPEG Surround: The Forthcoming ISO Standard for Spatial Audio Coding", en 28th International AES Conference, The Future of Audio Technology Surround and Beyond, Piteå, Suecia, 30 de junio - 2 de julio, 2006.

[2] J. Breebaart, J. Herre, L. Villemoes, C. Jin, K. Kjörling, J. Plogsties y J. Koppens, "Multi-Channels goes Mobile: MPEG Surround Binaural Rendering", en 29th International AES Conference, Audio for Mobile and Handheld Devices, Seúl, 2-4 de septiembre de 2006.

40 [3] C. Faller, "Parametric Joint-Coding of Audio Sources", Convention Paper 6752 presentado en 120th AES Convention, París, Francia, 20-23 de mayo de 2006.

[4] C. Faller, "Parametric Joint-Coding of Audio Sources", solicitud de patente PCT/EP2006/050904, 2006.

45 El documento WO 2006/048203 A2 da a conocer conceptos para el rendimiento mejorado de la reconstrucción multicanal basada en predicción. Particularmente, una pérdida de energía introducida mediante un proceso de mezcla ascendente predictivo se tiene en cuenta en una reconstrucción multicanal. Particularmente, un canal original izquierdo, un canal original central y un canal original derecho se mezclan de manera descendente en un canal de mezcla descendente izquierdo y un canal de mezcla descendente derecho, en los que el canal de mezcla descendente izquierdo sólo contiene el canal original izquierdo y una parte del canal central original, y el canal de mezcla descendente derecho sólo contiene el canal original derecho y una parte del canal central original. Esto está definido en una matriz de mezcla descendente. Los dos canales base se transmiten junto con dos parámetros de mezcla ascendente diferentes a un mezclador ascendente que cumple una regla de mezcla ascendente de no conservación de energía. Se generan los canales izquierdo, derecho y central originales reconstruidos y estos canales se someten a una corrección de energía para obtener canales izquierdo, derecho y central corregidos.

50 Es un objeto de la presente invención proporcionar un esquema de codificación/decodificación de objetos de audio mejorado.

Este objeto se logra mediante un sintetizador de audio según la reivindicación 1, un método de sintetización de audio según la reivindicación 6, un codificador de objetos de audio según la reivindicación 7, un método de codificación de objetos de audio según la reivindicación 10, una señal de objeto de audio codificada según la reivindicación 11 o un programa informático según la reivindicación 13.

5 **SUMARIO DE LA INVENCION**

10 Un primer aspecto de la invención se refiere a un codificador de objetos de audio para generar una señal de objeto de audio codificada usando una pluralidad de objetos de audio, que comprende: un generador de información de mezcla descendente para generar información de mezcla descendente que indica una distribución de la pluralidad de objetos de audio en al menos dos canales de mezcla descendente; un generador de parámetro de objeto para generar parámetros de objeto para los objetos de audio; y una interfaz de salida para generar la señal de objeto de audio codificada usando la información de mezcla descendente y los parámetros de objeto.

15 Un segundo aspecto de la invención se refiere a un método de codificación de objetos de audio para generar una señal de objeto de audio codificada usando una pluralidad de objetos de audio, que comprende: generar información de mezcla descendente que indica una distribución de la pluralidad de objetos de audio en al menos dos canales de mezcla descendente; generar parámetros de objeto para los objetos de audio; y generar la señal de objeto de audio codificada usando la información de mezcla descendente y los parámetros de objeto.

20 Un tercer aspecto de la invención se refiere a un sintetizador de audio para generar datos de salida usando una señal de objeto de audio codificada, que comprende: un sintetizador de datos de salida para generar los datos de salida que pueden usarse para crear una pluralidad de canales de salida de una configuración de salida de audio predefinida que representa la pluralidad de objetos de audio, siendo el sintetizador de datos de salida operativo para usar información de mezcla descendente que indica una distribución de la pluralidad de objetos de audio en al menos dos canales de mezcla descendente, y parámetros de objeto de audio para los objetos de audio.

25 Un cuarto aspecto de la invención se refiere a un método de sintetización de audio para generar datos de salida usando una señal de objeto de audio codificada, que comprende: generar los datos de salida que pueden usarse para crear una pluralidad de canales de salida de una configuración de salida de audio predefinida que representa la pluralidad de objetos de audio, siendo el sintetizador de datos de salida operativo para usar información de mezcla descendente que indica una distribución de la pluralidad de objetos de audio en al menos dos canales de mezcla descendente, y parámetros de objeto de audio para los objetos de audio.

30 Un quinto aspecto de la invención se refiere a una señal de objeto de audio codificada que incluye una información de mezcla descendente que indica una distribución de una pluralidad de objetos de audio en al menos dos canales de mezcla descendente y parámetros de objeto, siendo los parámetros de objeto de manera que es posible la reconstrucción de los objetos de audio usando los parámetros de objeto y los al menos dos canales de mezcla descendente. Un sexto aspecto de la invención se refiere a un programa informático para realizar, cuando se ejecuta en un ordenador, el método de codificación de objetos de audio o el método de decodificación de objetos de audio.

35 **BREVE DESCRIPCION DE LOS DIBUJOS**

A continuación se describirá la presente invención a modo de ejemplos ilustrativos, que no limitan el alcance o espíritu de la invención, con referencia a los dibujos adjuntos, en los que:

- 40 la figura 1a ilustra la operación de codificación de objetos de audio espacial que comprende la codificación y la decodificación;
- la figura 1b ilustra la operación de codificación de objetos de audio espacial reutilizando un decodificador de *MPEG Surround*;
- la figura 2 ilustra la operación de un codificador de objetos de audio espacial;
- la figura 3 ilustra un extractor de parámetros de objeto de audio que opera en un modo basado en energía;
- la figura 4 ilustra un extractor de parámetros de objeto de audio que opera en un modo basado en predicción;
- 45 la figura 5 ilustra la estructura de un transcodificador de SAOC a *MPEG Surround*;
- la figura 6 ilustra diferentes modos de operación de un convertidor de mezcla descendente;
- la figura 7 ilustra la estructura de un decodificador de *MPEG Surround* para una mezcla descendente estéreo;
- la figura 8 ilustra un caso de uso práctico que incluye un codificador SAOC;
- la figura 9 ilustra una realización de codificador;
- 50 la figura 10 ilustra una realización de decodificador;

la figura 11 ilustra una tabla para mostrar diferentes modos de decodificador/sintetizador preferidos;  
 la figura 12 ilustra un método para calcular determinados parámetros de mezcla ascendente espacial;  
 la figura 13a ilustra un método para calcular parámetros de mezcla ascendente espacial adicionales;  
 la figura 13b ilustra un método para calcular el uso de parámetros de predicción;  
 5 la figura 14 ilustra una visión general de un sistema de codificador/decodificador;  
 la figura 15 ilustra un método para calcular parámetros de objeto de predicción; y  
 la figura 16 ilustra un método de reproducción estéreo.

**DESCRIPCIÓN DE REALIZACIONES PREFERIDAS**

10 Las realizaciones descritas a continuación son meramente ilustrativas de los principios de la presente invención para la CODIFICACIÓN MEJORADA Y REPRESENTACIÓN DE PARÁMETROS DE CODIFICACIÓN DE OBJETOS DE MEZCLA DESCENDENTE MULTICANAL. Se entiende que serán evidentes para otros expertos en la técnica modificaciones y variaciones de las disposiciones y los detalles descritos en el presente documento. Por tanto, sólo se pretende que quede limitada por el alcance de las reivindicaciones de patente adjuntas y no por los detalles específicos presentados a modo de descripción y explicación de las realizaciones en el presente documento.

15 Las realizaciones preferidas proporcionan un esquema de codificación que combina la funcionalidad de un esquema de codificación de objetos con las capacidades de reproducción de un decodificador multicanal. Los datos de control transmitidos se refieren a los objetos individuales y permiten, por tanto, una manipulación en la reproducción en cuanto a nivel y posición espacial. Por tanto, los datos de control están directamente relacionados con la denominada descripción de escena, dando información sobre el posicionamiento de los objetos. La descripción de escena puede controlarse o bien en el lado de decodificador de manera interactiva por el oyente o bien también en el lado de  
 20 codificador por el productor. Una fase de transcodificador, tal como se enseña por la invención, se usa para convertir los datos de control relacionados con el objeto y la señal de mezcla descendente en datos de control y una señal de mezcla descendente que esté relacionada con el sistema de reproducción, como por ejemplo el decodificador de *MPEG Surround*.

25 En el esquema de codificación presentado, los objetos pueden distribuirse arbitrariamente en los canales de mezcla descendente disponibles en el codificador. El transcodificador hace uso explícito de la información de mezcla descendente multicanal, proporcionando una señal de mezcla descendente transcodificada y datos de control relacionados con el objeto. De esta forma, la mezcla ascendente en el decodificador no se realiza para todos los canales de manera individual tal como se propone en [3], sino que todos los canales de mezcla descendente se tratan al  
 30 mismo tiempo en un único proceso de mezcla ascendente. En el nuevo esquema la información de mezcla descendente multicanal debe ser parte de los datos de control y se codifica por el codificador de objetos.

La distribución de los objetos en los canales de mezcla descendente puede realizarse de una manera automática o puede ser una elección de diseño en el lado de codificador. En el último caso, puede diseñarse la mezcla descendente para que sea adecuada para la reproducción por un esquema de reproducción multicanal existente (por ejemplo, un sistema de reproducción estéreo), que ofrece una reproducción y que omita la fase de decodificación multicanal y transcodificación. Esto es una ventaja adicional sobre los esquemas de codificación anteriores, que consisten en un  
 35 único canal de mezcla descendente, o múltiples canales de mezcla descendente que contienen subconjuntos de los objetos fuente.

Mientras que los esquemas de codificación de objetos de la técnica anterior describen únicamente el proceso de decodificación usando un único canal de mezcla descendente, la presente invención no sufre esta limitación ya que proporciona un método para decodificar conjuntamente mezclas descendentes que contienen mezcla descendente de más de un canal. La calidad que puede obtenerse en la separación de objetos aumenta con un mayor número de canales de mezcla descendente. Por tanto, la invención llena satisfactoriamente el vacío entre un esquema de codificación de objetos con un único canal mono de mezcla descendente y un esquema de codificación multicanal en el  
 40 que cada objeto se transmite en un canal separado. Por tanto, el esquema propuesto permite un ajuste a escala flexible de calidad para la separación de objetos según requisitos de la aplicación y las propiedades del sistema de transmisión (tal como la capacidad de canal).

Además, usar más de un canal de mezcla descendente es ventajoso puesto que permite considerar adicionalmente una correlación entre los objetos individuales en lugar de restringir la descripción a diferencias de intensidad tal como en los esquemas de codificación de objetos de la técnica anterior. Los esquemas de la técnica anterior se basan en la suposición que todos los objetos son independientes y no correlacionados entre sí (cero correlación cruzada), aunque en objetos reales no es poco probable que estén correlacionados, como por ejemplo el canal izquierdo y derecho de una señal estéreo. Incorporar la correlación en la descripción (datos de control) tal como enseña la invención, la hace más completa y facilita así adicionalmente la capacidad para separar los objetos.  
 50

Las realizaciones preferidas comprenden al menos una de las siguientes características:

Un sistema para transmitir y crear una pluralidad de objetos de audio individuales usando una mezcla descendente multicanal y datos de control adicionales que describen los objetos, que comprende: un codificador de objetos de audio espacial para codificar una pluralidad de objetos de audio en una mezcla descendente multicanal, información acerca de la mezcla descendente multicanal, y parámetros de objeto; o un decodificador de objetos de audio espacial para decodificar una mezcla descendente multicanal, información acerca de la mezcla descendente multicanal, parámetros de objeto, y una matriz de reproducción de objeto en una segunda señal de audio multicanal adecuada para la reproducción de audio.

La figura 1a ilustra la operación de codificación de objetos de audio espacial (SAOC), que comprende un codificador 101 SAOC y un decodificador 104 SAOC. El codificador 101 de objetos de audio espacial codifica  $N$  objetos en una mezcla descendente de objetos que consiste en  $K > 1$  canales de audio, según parámetros de codificador. La información acerca de la matriz de pesos de mezcla descendente  $D$  aplicada la emite el codificador SAOC junto con datos opcionales concernientes a la potencia y correlación de la mezcla descendente. La matriz  $D$  es con frecuencia, pero no necesariamente siempre, constante a lo largo del tiempo y la frecuencia, y por tanto representa una cantidad relativamente baja de información. Finalmente, el codificador SAOC extrae parámetros de objeto para cada objeto en función tanto del tiempo como de la frecuencia a una resolución definida por consideraciones de percepción. El decodificador 104 de objetos de audio espacial toma los canales de mezcla descendente de objetos, la información de mezcla descendente y los parámetros de objeto (generados por el codificador) como entradas y genera una salida con  $M$  canales de audio para la presentación al usuario. La reproducción de  $N$  objetos en  $M$  canales de audio hace uso de una matriz de reproducción proporcionada como entrada de usuario para el decodificador SAOC.

La figura 1b ilustra la operación de codificación de objetos de audio espacial reutilizando un decodificador de *MPEG Surround*. Un decodificador 104 SAOC enseñado por la invención actual puede realizarse como un transcodificador 102 de SAOC a *MPEG Surround* y un decodificador 103 de *MPEG Surround* basado en mezcla descendente estéreo. Una matriz de reproducción  $A$  controlada por el usuario de tamaño  $M \times N$  define la reproducción objetivo de los  $N$  objetos a  $M$  canales de audio. Esta matriz puede depender tanto del tiempo como de la frecuencia y es la salida final de una interfaz más sencilla para el usuario para la manipulación de objetos de audio (que también puede hacer uso de una descripción de escena externamente proporcionada). En el caso de una configuración de altavoces 5.1 el número de canales de audio de salida es  $M = 6$ . La tarea del decodificador SAOC es recrear perceptivamente la reproducción objetivo de los objetos de audio originales. El transcodificador 102 de SAOC a *MPEG Surround* toma como entrada la matriz de reproducción  $A$ , la mezcla descendente de objetos, la información secundaria de mezcla descendente que incluye la matriz de pesos de mezcla descendente  $D$ , y la información secundaria de objeto, y genera una mezcla descendente estéreo e información secundaria de *MPEG Surround*. Cuando se construye el transcodificador según la invención actual, un decodificador 103 de *MPEG Surround* posterior alimentado con estos datos producirá una salida de audio de  $M$  canales con las propiedades deseadas.

Un decodificador SAOC enseñado por la invención actual consiste en un transcodificador 102 de SAOC a *MPEG Surround* y un decodificador 103 de *MPEG Surround* basado en mezcla descendente estéreo. Una matriz de reproducción  $A$  controlada por el usuario de tamaño  $M \times N$  define la reproducción objetivo de los  $N$  objetos a  $M$  canales de audio. Esta matriz puede depender tanto del tiempo como de la frecuencia y es la salida final de una interfaz más sencilla para el usuario para la manipulación de objetos de audio. En el caso de una configuración de altavoces 5.1 el número de canales de audio de salida es  $M = 6$ . La tarea del decodificador SAOC es recrear perceptivamente la reproducción objetivo de los objetos de audio originales. El transcodificador 102 de SAOC a *MPEG Surround* toma como entrada la matriz de reproducción  $A$ , la mezcla descendente de objetos, la información secundaria de mezcla descendente que incluye la matriz de pesos de mezcla descendente  $D$ , y la información secundaria de objeto, y genera una mezcla descendente estéreo e información secundaria de *MPEG Surround*. Cuando se construye el transcodificador según la invención actual, un decodificador 103 de *MPEG Surround* posterior alimentado con estos datos producirá una salida de audio de  $M$  canales con las propiedades deseadas.

La figura 2 ilustra la operación de un codificador 101 de objetos de audio espacial (SAOC) enseñada por la invención actual. Los  $N$  objetos de audio se alimentan tanto en un mezclador 201 descendente como un extractor 202 de parámetros de objeto de audio. El mezclador 201 descendente mezcla los objetos en una mezcla descendente de objetos que consiste en  $K > 1$  canales de audio, según los parámetros de codificador y también emite información de mezcla descendente. Esta información incluye una descripción de la matriz de pesos de mezcla descendente  $D$  aplicada y, opcionalmente, si el extractor de parámetros de objeto de audio posterior opera en modo de predicción, describiendo los parámetros la potencia y correlación de la mezcla descendente de objetos. Como se comentará en un párrafo posterior, el papel de tales parámetros adicionales es dar acceso a la energía y correlación de subconjuntos de canales de audio reproducidos en el caso en el que los parámetros de objeto se expresan sólo en relación con la mezcla descendente, siendo el ejemplo principal las indicaciones frontales/traseras de una configuración de altavoces 5.1. El extractor 202 de parámetros de objeto de audio extrae parámetros de objeto según los parámetros de codificador. El control de codificador determina, basándose en la variación de tiempo y frecuencia, cuál de dos modos de codificador se aplica, el *modo basado energía* o el *basado en predicción*. En el modo basado en energía, los parámetros de codificador contienen además información sobre una agrupación de los  $N$  objetos de audio en  $P$  objetos estéreo y  $N-2P$  objetos mono. Cada modo se describirá adicionalmente mediante las figuras 3 y 4.

La figura 3 ilustra un extractor 202 de parámetros de objeto de audio que opera en un modo basado en energía. Una agrupación 301 en  $P$  objetos estéreo y  $N-2P$  objetos mono se realiza según la información de agrupación contenida en los parámetros de codificador. Para cada intervalo de tiempo-frecuencia considerado se realizan entonces las siguientes operaciones. Dos potencias de objeto y una correlación normalizada se extraen para cada uno de los  $P$  objetos estéreo por el extractor 302 de parámetros estéreo. Un parámetro de potencia se extrae para cada uno de los  $N-2P$  objetos mono por el extractor 303 de parámetros mono. El conjunto total de  $N$  parámetros de potencia y  $P$  parámetros de correlación normalizada se codifica entonces en 304 junto con los datos de agrupación para formar los parámetros de objeto. La codificación puede contener una etapa de normalización con respecto a la potencia de objeto más grande o con respecto a la suma de potencias de objeto extraídas.

La figura 4 ilustra un extractor 202 de parámetros de objeto de audio que opera en un modo basado en predicción. Para cada intervalo de tiempo-frecuencia considerado, se realizan las siguientes operaciones. Para cada uno de los  $N$  objetos, se deriva una combinación lineal de los  $K$  canales de mezcla descendente de objetos que coincide con el objeto dado en un sentido de mínimos cuadrados. Los  $K$  pesos de esta combinación lineal se denominan coeficientes de predicción de objeto (OPC) y se calculan por el extractor 401 de OPC. El conjunto total de los  $N \cdot K$  OPC se codifican en 402 para formar los parámetros de objeto. La codificación puede incorporar una reducción del número total de OPC basándose en interdependencias lineales. Tal como se enseña por la presente invención, este número total puede reducirse a  $\max\{K \cdot (N-K), 0\}$  si la matriz de pesos de mezcla descendente  $D$  tiene rango completo.

La figura 5 ilustra la estructura de un transcodificador 102 de SAOC a *MPEG Surround* tal como se enseña por la invención actual. Para cada intervalo de tiempo-frecuencia, la información secundaria de mezcla descendente y los parámetros de objeto se combinan con la matriz de reproducción mediante el calculador 502 de parámetros para formar parámetros de *MPEG Surround* de tipo CLD, CPC e ICC, y una matriz de convertidor de mezcla descendente  $G$  de tamaño  $2 \times K$ . El convertidor 501 de mezcla descendente convierte la mezcla descendente de objetos en una mezcla descendente estéreo aplicando una operación de matriz según las matrices  $G$ . En un modo simplificado del transcodificador para  $K = 2$ , esta matriz es la matriz de identidad y la mezcla descendente de objetos se pasa a través de la misma sin alterarse como mezcla descendente estéreo. Este modo se ilustra en el dibujo con el conmutador 503 selector en la posición A, mientras que el modo de operación normal tiene el conmutador en la posición B. Una ventaja adicional del transcodificador es su posibilidad de uso como una aplicación autónoma en la que los parámetros de *MPEG Surround* se ignoran y la salida del convertidor de mezcla descendente se usa directamente como una reproducción estéreo.

La figura 6 ilustra diferentes modos de operación de un convertidor 501 de mezcla descendente tal como se enseña por la presente invención. Dada la mezcla descendente de objetos transmitida en el formato de una salida de flujo de bits desde un codificador de audio de  $K$  canales, este flujo de bits se decodifica en primer lugar por el decodificador 601 de audio en  $K$  señales de audio en el dominio del tiempo. Estas señales entonces se transforman todas al dominio de la frecuencia por un banco de filtros QMF híbrido de *MPEG Surround* en la unidad 602 de T/F. La operación de matriz de variación en el tiempo y la frecuencia definida por los datos de matriz de convertidor se realiza en las señales de dominio QMF híbrido resultantes mediante la unidad 603 de matrización que emite una señal estéreo en el dominio QMF híbrido. La unidad 604 de síntesis híbrida convierte la señal en el dominio QMF híbrido estéreo en una señal en el dominio QMF estéreo. El dominio QMF híbrido se define con el fin de obtener mejor resolución de frecuencia hacia frecuencias más bajas por medio de un filtrado posterior de las subbandas QMF. Cuando este filtrado posterior se define por bancos de filtros de Nyquist, la conversión del dominio QMF híbrido al convencional consiste en simplemente la suma de grupos de señales de subbanda híbridadas, véase [E. Schuijers, J. Breebart y H. Purnhagen "Low complexity parametric stereo coding" Proc 116th AES convention Berlín, Alemania 2004, Preprint 6073]. Esta señal constituye el primer formato de salida posible del convertidor de mezcla descendente según se define por el conmutador 607 selector en la posición A. Una señal en el dominio QMF de este tipo puede alimentarse directamente a la interfaz en el dominio QMF correspondiente de un decodificador de *MPEG Surround*, y este es el modo de operación más ventajoso en cuanto a retardo, complejidad y calidad. La siguiente posibilidad se obtiene realizando una síntesis 605 de banco de filtros QMF con el fin de obtener una señal en el dominio del tiempo estéreo. Con el conmutador 607 selector en la posición B, el convertidor emite una señal estéreo de audio digital que también puede alimentarse a la interfaz en el dominio del tiempo de un decodificador de *MPEG Surround* posterior, o se reproduce directamente en un dispositivo de reproducción estéreo. La tercera posibilidad con el conmutador 607 selector en la posición C se obtiene codificando la señal estéreo en el dominio del tiempo con un codificador 606 de audio estéreo. El formato de salida del convertidor de mezcla descendente es entonces un flujo de bits de audio estéreo que es compatible con un decodificador de núcleo contenido en el decodificador de MPEG. Este tercer modo de operación es adecuado para el caso en el que el transcodificador de SAOC a *MPEG Surround* está separado por el decodificador de MPEG mediante una conexión que impone restricciones sobre la tasa de bits, o en el caso en el que el usuario desea almacenar una reproducción de objeto particular para su reproducción futura.

La figura 7 ilustra la estructura de un decodificador de *MPEG Surround* para una mezcla descendente estéreo. La mezcla descendente estéreo se convierte en tres canales intermedios mediante la casilla de dos a tres (TTT). Estos canales intermedios se dividen además en dos por las tres casillas de uno a dos (OTT) para conseguir los seis canales de una configuración de canales 5.1.

La figura 8 ilustra un caso de uso práctico que incluye un codificador SAOC. Un mezclador 802 de audio emite una señal estéreo (L y R) que normalmente se compone combinando señales de entrada de mezclador (en este caso los

canales de entrada 1-6) y opcionalmente entradas adicionales de retornos de efectos tales como reverberación, etc. El mezclador también emite un canal individual (en este caso el canal 5) desde el mezclador. Esto puede realizarse, por ejemplo, por medio de funcionalidades de mezclador comúnmente usadas tales como "salidas directas" o "envío auxiliar" con el fin de emitir un canal individual posterior a cualquiera de los procesos de inserción (tal como procesamiento dinámico y EQ). La señal estéreo (L y R) y la salida de canal individual (obj5) se introducen al codificador 801 SAOC, que no es sino un caso especial del codificador 101 SAOC en la figura 1. Sin embargo, ilustra claramente una aplicación típica en la que el objeto de audio obj5 (que contiene, por ejemplo, voz) debe someterse a modificaciones de nivel controladas por el usuario en el lado de decodificador mientras que todavía es parte de la mezcla estéreo (L y R). A partir del concepto, también es obvio que pueden conectarse dos o más objetos de audio al panel de "entrada de objeto" en 801, y además la mezcla estéreo puede extenderse mediante una mezcla multicanal tal como una mezcla 5.1.

En el texto que sigue, se expondrá la descripción matemática de la presente invención. Para señales complejas discretas  $x, y$ , el producto interior complejo y norma cuadrada (energía) se define por

$$\left\{ \begin{array}{l} \langle x, y \rangle = \sum_k x(k) \bar{y}(k), \\ \|x\|^2 = \langle x, x \rangle = \sum_k |x(k)|^2, \end{array} \right\} \quad (1)$$

donde  $\bar{y}(k)$  indica la señal conjugada compleja de  $y(k)$ . Todas las señales consideradas en este caso son muestras de subbanda de un banco de filtros modulados o análisis FFT con función ventana de señales de tiempo discretas. Se entiende que estas subbandas deben transformarse de nuevo al dominio de tiempo discreto mediante correspondientes operaciones de banco de filtros de síntesis. Un bloque de señal de  $L$  muestras representa la señal en un intervalo de tiempo y frecuencia que es parte del enlosado motivado perceptivamente del plano tiempo-frecuencia que se aplica para la descripción de propiedades de señal. En esta situación, los objetos de audio dados pueden representarse como  $N$  filas de longitud  $L$  en una matriz,

$$\mathbf{S} = \begin{bmatrix} s_1(0) & s_1(1) & \dots & s_1(L-1) \\ s_2(0) & s_2(1) & \dots & s_2(L-1) \\ \vdots & \vdots & & \vdots \\ s_N(0) & s_N(1) & \dots & s_N(L-1) \end{bmatrix}. \quad (2)$$

La matriz de pesos de mezcla descendente  $D$  de tamaño  $K \times N$ , donde  $K > 1$  determina la señal de mezcla descendente de  $K$  canales en forma de una matriz con  $K$  filas a través de la multiplicación de matrices

$$\mathbf{X} = \mathbf{D}\mathbf{S}. \quad (3)$$

La matriz de reproducción de objeto  $A$  controlada por el usuario de tamaño  $M \times N$  determina la reproducción objetivo de  $M$  canales de los objetos de audio en forma de una matriz con  $M$  filas a través de la multiplicación de matrices

$$\mathbf{Y} = \mathbf{A}\mathbf{S}. \quad (4)$$

Ignorando de momento los efectos de la codificación de audio núcleo, la tarea del decodificador SAOC es generar una aproximación en el sentido perceptivo de la reproducción objetivo  $Y$  de los objetos de audio originales, dada la matriz de reproducción  $A$ , la mezcla descendente  $X$ , la matriz de mezcla descendente  $D$  y parámetros de objeto.

Los parámetros de objeto en el modo de energía enseñado por la presente invención llevan información acerca de la covarianza de los objetos originales. En una versión determinística conveniente para la posterior derivación y también descriptiva de las operaciones de codificador típicas, esta covarianza viene dada en forma no normalizada por el producto de matrices  $\mathbf{S}\mathbf{S}^*$  donde el asterisco indica la operación de matriz transpuesta conjugada compleja. Por tanto, los parámetros de objeto del modo de energía proporcionan una matriz positiva semidefinida  $E$   $N \times N$  de manera que, posiblemente hasta un factor de escala,

$$\mathbf{S}\mathbf{S}^* \approx \mathbf{E}. \quad (5)$$

5 La codificación de objetos de audio de la técnica anterior suele considerar un modelo de objeto en el que todos los objetos están no correlacionados. En este caso, la matriz E es diagonal y sólo contiene una aproximación a las energías de objeto  $S_n = \|s_n\|^2$  para  $n = 1, 2, \dots, N$ . El extractor de parámetros de objeto según la figura 3, permite un importante refinado de esta idea, particularmente relevante en casos en los que los objetos se proporcionan como señales estéreo para las que no se sostienen las suposiciones sobre ausencia de correlación. Una agrupación de  $P$  pares estéreo seleccionados de objetos se describe mediante los conjuntos de índices  $\{(n_p, m_p), p = 1, 2, \dots, P\}$ . Para estos pares estéreo se calcula la correlación  $\langle s_n, s_m \rangle$  y el valor complejo, real, o absoluto de la correlación normalizada (ICC)

$$\rho_{n,m} = \frac{\langle s_n, s_m \rangle}{\|s_n\| \|s_m\|} \quad (6)$$

10 se extrae por el extractor 302 de parámetros estéreo. En el decodificador, los datos ICC pueden entonces combinarse con las energías con el fin de formar una matriz E con  $2P$  entradas fuera de la diagonal. Por ejemplo, para un total de  $N=3$  objetos de los que los primeros dos consisten en un único par (1,2), la energía transmitida y los datos de correlación son  $(S_1, S_2, S_3)$  y  $\rho_{1,2}$ . En este caso, la combinación en la matriz E da

$$\mathbf{E} = \begin{bmatrix} S_1 & \rho_{1,2} \sqrt{S_1 S_2} & 0 \\ \rho_{1,2}^* \sqrt{S_1 S_2} & S_2 & 0 \\ 0 & 0 & S_3 \end{bmatrix}$$

15 Los parámetros de objeto en el modo de predicción enseñado por la presente invención pretenden poner una matriz de coeficientes de predicción de objeto (OPC) C de  $N \times K$  a disposición del decodificador de manera que

$$\mathbf{S} \approx \mathbf{C}\mathbf{X} = \mathbf{C}\mathbf{D}\mathbf{S}. \quad (7)$$

En otras palabras, para cada objeto hay una combinación lineal de los canales de mezcla descendente de manera que el objeto puede recuperarse aproximadamente mediante

$$s_n(k) \approx c_{n,1}x_1(k) + \dots + c_{n,K}x_K(k). \quad (8)$$

20 En una realización preferida, el extractor 401 de OPC resuelve las ecuaciones normales

$$\mathbf{C}\mathbf{X}\mathbf{X}^* = \mathbf{S}\mathbf{X}^*, \quad (9)$$

o, para el caso de OPC de valor real más atractivo, resuelve

$$\mathbf{C}\text{Re}\{\mathbf{X}\mathbf{X}^*\} = \text{Re}\{\mathbf{S}\mathbf{X}^*\}. \quad (10)$$

25 En ambos casos, suponiendo una matriz de pesos de mezcla descendente D de valor real, y una covarianza de mezcla descendente no singular, se desprende mediante la multiplicación desde la izquierda con D que

$$\mathbf{D}\mathbf{C} = \mathbf{I}, \quad (11)$$

30 donde I es la matriz de identidad de tamaño K. Si D tiene rango completo, se desprende mediante álgebra lineal elemental que el conjunto de soluciones de (9) pueden parametrizarse mediante parámetros  $\max\{K \cdot (N-K), 0\}$ . Esto se aprovecha en la codificación conjunta en 402 de los datos OPC. La matriz de predicción C completa puede recrearse en el decodificador a partir del conjunto reducido de parámetros y la matriz de mezcla descendente.

Por ejemplo, considérese para una mezcla descendente estéreo ( $K=2$ ) el caso de tres objetos ( $N=3$ ) que comprenden una pista de música estéreo ( $s_1, s_2$ ) y una pista de voz o de instrumento individual con panorámica central  $s_3$ . La matriz de mezcla descendente es

$$\mathbf{D} = \begin{bmatrix} 1 & 0 & 1/\sqrt{2} \\ 0 & 1 & 1/\sqrt{2} \end{bmatrix}, \quad (12)$$

Esto es, el canal izquierdo de mezcla descendente es  $x_1 = s_1 + s_3/\sqrt{2}$  y el canal derecho es  $x_2 = s_2 + s_3/\sqrt{2}$ . Los OPC para la pista individual pretenden aproximar  $s_3 \approx c_{31}x_1 + c_{32}x_2$  y la ecuación (11) puede resolverse en este caso para

conseguir  $c_{11} = 1 - c_{31}/\sqrt{2}$ ,  $c_{12} = -c_{32}/\sqrt{2}$ ,  $c_{21} = -c_{31}/\sqrt{2}$ ,  $c_{22} = 1 - c_{32}/\sqrt{2}$ . Por tanto, el número de OPC que bastan viene dado por  $K(N - K) = 2 \cdot (3 - 2) = 2$ .

Los OPC  $c_{31}, c_{32}$  pueden hallarse a partir de las ecuaciones normales

$$\begin{bmatrix} c_{31} & c_{32} \end{bmatrix} \begin{bmatrix} \|x_1\| & \langle x_1, x_2 \rangle \\ \langle x_2, x_1 \rangle & \|x_2\| \end{bmatrix} = \begin{bmatrix} \langle s_3, x_1 \rangle & \langle s_3, x_2 \rangle \end{bmatrix}$$

Transcodificador de SAOC a MPEG Surround

En referencia a la figura 7, los  $M = 6$  canales de salida de la configuración 5.1 son  $(y_1, y_2, \dots, y_6) = (l_f, l_s, r_f, r_s, c, lfe)$ . El transcodificador debe emitir una mezcla descendente estéreo  $(l, r)$  y parámetros para las casillas TTT y OTT. Como la atención se centra ahora en la mezcla descendente estéreo, se supondrá a continuación que  $K=2$ . Puesto que tanto los parámetros de objeto como los parámetros MPS TTT existen tanto en modo de energía como en un modo de predicción, las cuatro combinaciones deben considerarse. El modo de energía es una opción adecuada, por ejemplo, en caso de que el codificador de audio de mezcla descendente no sea de codificador de forma de onda en el intervalo de frecuencia considerado. Se entiende que los parámetros de MPEG Surround derivados en el siguiente texto deben cuantificarse y codificarse adecuadamente antes de su transmisión. Para aclarar mejor las cuatro combinaciones mencionadas anteriormente, éstas comprenden

1. Parámetros de objeto en modo de energía y transcodificador en modo de predicción
2. Parámetros de objeto en modo de energía y transcodificador en modo de energía
3. Parámetros de objeto en modo de predicción (OPC) y transcodificador en modo de predicción
4. Parámetros de objeto en modo de predicción (OPC) y transcodificador en modo de energía

Si el codificador de audio de mezcla descendente es un codificador de forma de onda en el intervalo de frecuencia considerado, los parámetros de objeto pueden estar tanto en modo de energía como de predicción, pero el transcodificador debe operar preferiblemente en modo de predicción. Si el codificador de audio de mezcla descendente no es un codificador de forma de onda en el intervalo de frecuencia considerado, el codificador de objetos y el transcodificador deben operar ambos en modo de energía. La cuarta combinación es la que tiene menos relevancias de modo que la siguiente descripción se dirigirá a las primeras tres combinaciones únicamente.

Parámetros de objeto dados en modo de energía

En modo de energía, los datos disponibles para el transcodificador se describen por el triplete de matrices  $(D, E, A)$ . Los parámetros OTT de MPEG Surround se obtienen realizando estimaciones de energía y correlación sobre una reproducción virtual derivada de los parámetros transmitidos y de la matriz de reproducción  $A$   $6 \times N$ . La covarianza objetivo de seis canales viene dada por

$$\mathbf{Y}\mathbf{Y}^* = \mathbf{A}\mathbf{S}(\mathbf{A}\mathbf{S})^* = \mathbf{A}(\mathbf{S}\mathbf{S}^*)\mathbf{A}^*, \quad (13)$$

Insertando (5) en (13) se obtiene la aproximación

$$\mathbf{Y}\mathbf{Y}^* \approx \mathbf{F} = \mathbf{A}\mathbf{E}\mathbf{A}^*, \quad (14)$$

que se define completamente por los datos disponibles. Digamos que  $f_u$  son los elementos de  $F$ . Entonces, los parámetros CLD y ICC se leen a partir de

$$CLD_0 = 10 \log_{10} \left( \frac{f_{55}}{f_{66}} \right), \quad (15)$$

$$CLD_1 = 10 \log_{10} \left( \frac{f_{33}}{f_{44}} \right), \quad (16)$$

$$CLD_2 = 10 \log_{10} \left( \frac{f_{11}}{f_{22}} \right), \quad (17)$$

$$ICC_1 = \frac{\varphi(f_{34})}{\sqrt{f_{33}f_{44}}}, \quad (18)$$

$$ICC_2 = \frac{\varphi(f_{12})}{\sqrt{f_{11}f_{22}}}, \quad (19)$$

5

donde  $\varphi$  es el operador o bien de valor absoluto  $\varphi(z) = |z|$  o bien de valor real  $\varphi(z) = \text{Re}\{z\}$ .

Como un ejemplo ilustrativo, considérese el caso de tres objetos previamente descritos en relación con la ecuación (12). Digamos que la matriz de reproducción viene dada por

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

10 La reproducción objetivo consiste por tanto en colocar el objeto 1 entre frontal derecho y envolvente derecho, el objeto 2 entre frontal izquierdo y envolvente izquierdo, y el objeto 3 en frontal derecho, centro y lfe. Supóngase también por motivos de simplicidad que los tres objetos no están correlacionados y todos tienen la misma energía de manera que

$$\mathbf{E} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

En este caso, el lado derecho de la fórmula (14) pasa a ser

$$\mathbf{F} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}.$$

15

Insertando los valores apropiados en las fórmulas (15)-(19) se obtiene entonces

$$CLD_0 = 10 \log_{10} \left( \frac{f_{35}}{f_{66}} \right) = 10 \log_{10} \left( \frac{1}{1} \right) = 0 \text{ dB},$$

$$CLD_1 = 10 \log_{10} \left( \frac{f_{33}}{f_{44}} \right) = 10 \log_{10} \left( \frac{2}{1} \right) = 3 \text{ dB},$$

$$CLD_2 = 10 \log_{10} \left( \frac{f_{11}}{f_{22}} \right) = 10 \log_{10} \left( \frac{1}{1} \right) = 0 \text{ dB},$$

$$ICC_1 = \frac{\varphi(f_{34})}{\sqrt{f_{33}f_{44}}} = \frac{\varphi(1)}{\sqrt{2 \cdot 1}} = \frac{1}{\sqrt{2}},$$

$$ICC_2 = \frac{\varphi(f_{12})}{\sqrt{f_{11}f_{22}}} = \frac{\varphi(1)}{\sqrt{1 \cdot 1}} = 1,$$

5

Como consecuencia, se indicará al decodificador de *MPEG Surround* que use alguna descorrelación entre frontal derecho y envolvente derecho, pero no descorrelación entre frontal derecho y envolvente izquierdo.

Para los parámetros TTT de *MPEG Surround* en modo de predicción, la primera etapa es formar una matriz de

reproducción reducida  $A_3$  de tamaño  $3 \times N$  para los canales combinados  $(l, r, qc)$  donde  $q = 1/\sqrt{2}$ . Se cumple que  $A_3 = D_{36}A$  donde la matriz de mezcla descendente parcial de 6 a 3 se define por

10

$$D_{36} = \begin{bmatrix} w_1 & w_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & w_2 & w_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & qw_3 & qw_3 \end{bmatrix}. \quad (20)$$

Los pesos de mezcla descendente parcial  $w_p$ ,  $p=1,2,3$  se ajustan de manera que la energía de  $w_p(y_{2p-1} + y_{2p})$  es igual a la suma de energías  $\|y_{2p-1}\|^2 + \|y_{2p}\|^2$  hasta un factor límite. Todos los datos que se requieren para derivar la matriz de mezcla descendente parcial  $D_{36}$  están disponibles en  $F$ . A continuación, se produce una matriz de predicción  $C_3$  de tamaño  $3 \times 2$  de manera que

15

$$C_3 X \approx A_3 S, \quad (21)$$

Una matriz de este tipo se deriva preferiblemente considerando en primer lugar las ecuaciones normales

$$C_3 (DED') = A_3 ED',$$

20

La solución a las ecuaciones normales da la mejor coincidencia de forma de onda posible para (21) dado el modelo de covarianza de objeto  $E$ . Es preferible algún postprocesamiento de la matriz  $C_3$ , incluyendo factores fila para un canal total o individual basándose en compensación de pérdida de predicción.

Para ilustrar y aclarar las etapas anteriores, considérese una continuación del ejemplo de reproducción de seis canales específico dado anteriormente. En cuanto a los elementos de matriz de  $F$ , los pesos de mezcla descendente son soluciones a las ecuaciones

25

$$w_p^2 (f_{2p-1,2p-1} + f_{2p,2p} + 2f_{2p-1,2p}) \approx f_{2p-1,2p-1} + f_{2p,2p}, \quad p=1,2,3,$$

que en el ejemplo específico pasa a ser

$$\left\{ \begin{array}{l} w_1^2(1+1+2 \cdot 1) = 1+1 \\ w_2^2(2+1+2 \cdot 1) = 2+1 \\ w_3^2(1+1+2 \cdot 1) = 1+1 \end{array} \right\},$$

De manera que,  $(w_1, w_2, w_3) = (1/\sqrt{2}, \sqrt{3/5}, 1/\sqrt{2})$ . La inserción en (20) da

$$\mathbf{A}_3 = \mathbf{D}_{36} \mathbf{A} = \begin{bmatrix} 0 & \sqrt{2} & 0 \\ 2\sqrt{\frac{3}{5}} & 0 & \sqrt{\frac{3}{5}} \\ 0 & 0 & 1 \end{bmatrix}.$$

Resolviendo el sistema de ecuaciones  $\mathbf{C}_3(\text{DED}) = \mathbf{A}_3 \mathbf{E} \mathbf{D}$  se halla entonces, (conmutando ahora a precisión finita),

$$\mathbf{C}_3 = \begin{bmatrix} -0.3536 & 1.0607 \\ 1.4358 & -0.1134 \\ 0.3536 & 0.3536 \end{bmatrix}.$$

5

La matriz  $\mathbf{C}_3$  contiene los mejores pesos para obtener una aproximación a la reproducción de objeto deseada a los canales combinados  $(l, r, qc)$  a partir de la mezcla descendente de objetos. Este tipo general de operación de matriz no puede implementarse por el decodificador de *MPEG Surround*, que está restringido a un espacio limitado de matrices TTT mediante el uso de sólo dos parámetros. El objeto del convertidor de mezcla descendente de la invención es procesar previamente la mezcla descendente de objetos de manera que el efecto combinado del procesamiento previo y la matriz TTT de *MPEG Surround* sea idéntico a la mezcla ascendente deseada descrita por  $\mathbf{C}_3$ .

10

En *MPEG Surround*, la matriz TTT para la predicción de  $(l, r, qc)$  a partir de  $(l_0, r_0)$  se parametriza mediante tres parámetros  $(\alpha, \beta, \gamma)$  mediante

$$\mathbf{C}_{\text{TTT}} = \frac{\gamma}{3} \begin{bmatrix} \alpha + 2 & \beta - 1 \\ \alpha - 1 & \beta + 2 \\ 1 - \alpha & 1 - \beta \end{bmatrix}. \quad (22)$$

15

La matriz de convertidor de mezcla descendente  $\mathbf{G}$  enseñada por la presente invención se obtiene eligiendo  $\gamma=1$  y resolviendo el sistema de ecuaciones

$$\mathbf{C}_{\text{TTT}} \mathbf{G} = \mathbf{C}_3. \quad (23)$$

Tal como puede verificarse fácilmente, se cumple que  $\mathbf{D}_{\text{TTT}} \mathbf{C}_{\text{TTT}} = \mathbf{I}$ , donde  $\mathbf{I}$  es la matriz de identidad de dos por dos y

$$\mathbf{D}_{\text{TTT}} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}. \quad (24)$$

20

Por tanto, una multiplicación de matrices desde la izquierda por  $\mathbf{D}_{\text{TTT}}$  de ambos lados de (23) conduce a

$$\mathbf{G} = \mathbf{D}_{\text{TTT}} \mathbf{C}_3. \quad (25)$$

En el caso genérico,  $\mathbf{G}$  podrá invertirse y (23) tiene una solución única para  $\mathbf{C}_{\text{TTT}}$  que cumple  $\mathbf{D}_{\text{TTT}} \mathbf{C}_{\text{TTT}} = \mathbf{I}$ . Los parámetros TTT  $(\alpha, \beta)$  se determinan mediante esta solución.

Para el ejemplo específico considerado anteriormente, puede verificarse fácilmente que las soluciones vienen dadas por

$$\mathbf{G} = \begin{bmatrix} 0 & 1.4142 \\ 1.7893 & 0.2401 \end{bmatrix} \quad \text{y} \quad (\alpha, \beta) = (0.3506, 0.4072).$$

5

Obsérvese que una parte principal de la mezcla descendente estéreo se intercambia entre izquierda y derecha para esta matriz de convertidor, lo que refleja el hecho de que el ejemplo de reproducción pone objetos que están en el canal de mezcla descendente de objetos izquierdo en la parte derecha de la escena de sonido y viceversa. Tal comportamiento es imposible de obtener de un decodificador de *MPEG Surround* en modo estéreo.

10

Si es imposible aplicar un convertidor de mezcla descendente, puede desarrollarse un procedimiento que no sea del todo óptimo como sigue. Para los parámetros TTT de *MPEG Surround* en modo de energía, lo que se requiere es la distribución de energía de los canales combinados (*l,r,c*). Por tanto los parámetros CLD relevantes pueden derivarse directamente a partir de los elementos de *F* a través de

$$CLD_{TTT}^0 = 10 \log_{10} \left( \frac{\|l\|^2 + \|r\|^2}{\|c\|^2} \right) = 10 \log_{10} \left( \frac{f_{11} + f_{22} + f_{33} + f_{44}}{f_{55} + f_{66}} \right), \quad (26)$$

$$CLD_{TTT}^1 = 10 \log_{10} \left( \frac{\|l\|^2}{\|r\|^2} \right) = 10 \log_{10} \left( \frac{f_{11} + f_{22}}{f_{33} + f_{44}} \right). \quad (27)$$

15

En este caso, es adecuado usar sólo una matriz diagonal *G* con entradas positivas para el convertidor de mezcla descendente. Es operacional conseguir la distribución de energía correcta de los canales de mezcla descendente antes de la mezcla ascendente TTT. Con la matriz de mezcla descendente de seis a dos canales  $D_{26} = D_{TTT} D_{36}$  y las definiciones a partir de

$$\mathbf{Z} = \mathbf{D} \mathbf{E} \mathbf{D}' , \quad (28)$$

$$\mathbf{W} = \mathbf{D}_{26} \mathbf{E} \mathbf{D}'_{26} , \quad (29)$$

se elige simplemente

$$\mathbf{G} = \begin{bmatrix} \sqrt{w_{11}/z_{11}} & 0 \\ 0 & \sqrt{w_{22}/z_{22}} \end{bmatrix}. \quad (30)$$

20

Una observación adicional es que tal convertidor de mezcla descendente de forma diagonal puede omitirse del transcodificador de objetos a *MPEG Surround* e implementarse por medio de la activación de los parámetros de ganancia de mezcla descendente arbitrarios (ADG) del decodificador de *MPEG Surround*. Estas ganancias se darán entonces en el dominio logarítmico mediante  $ADG_i = 10 \log_{10} (w_{ii}/z_{ii})$  para  $i=1,2$ .

25

Parámetros de objeto dados en modo de predicción (OPC)

En el modo de predicción de objeto, los datos disponibles están representados por el triplete de matrices (D,C,A) donde C es la matriz  $N \times 2$  que contiene los N pares de OPC. Debido a la naturaleza relativa de los coeficientes de predicción, será necesario además para la estimación de parámetros de *MPEG Surround* basados en energía tener acceso a una aproximación a la matriz de covarianza de  $2 \times 2$  de la mezcla descendente de objetos,

$$\mathbf{X} \mathbf{X}' \approx \mathbf{Z}. \quad (31)$$

30

Esta información se transmite preferiblemente desde el codificador de objetos como parte de la información secundaria de mezcla descendente, pero también podría estimarse en el transcodificador a partir de mediciones realizadas en la mezcla descendente recibida, o indirectamente derivarse de (D, C) mediante consideraciones de modelo de objeto aproximado. Dada Z, la covarianza de objeto puede estimarse insertando el modelo predictivo  $Y = CX$ , dando

$$5 \quad \mathbf{E} = \mathbf{CZC}^*, \quad (32)$$

y todos los parámetros TTT de modo de energía y OTT de *MPEG Surround* pueden estimarse a partir de E como en el caso de parámetros de objeto basados en energía. Sin embargo, la gran ventaja de usar OPC surge en combinación con parámetros TTT de *MPEG Surround* en modo de predicción. En este caso, la aproximación de forma de onda  $D_{36} Y \approx A_3 CX$  da inmediatamente la matriz de predicción reducida

$$10 \quad \mathbf{C}_3 = \mathbf{A}_3 \mathbf{C}, \quad (32)$$

a partir de cual las etapas restantes para lograr los parámetros TTT ( $\alpha, \beta$ ) y el convertidor de mezcla descendente son similares al caso de parámetros de objeto proporcionados en modo de energía. De hecho, las etapas de las fórmulas (22) a (25) son completamente idénticas. Se alimenta la matriz resultante G al convertidor de mezcla descendente y los parámetros TTT ( $\alpha, \beta$ ) se transmiten al decodificador de *MPEG Surround*.

15 Aplicación autónoma del convertidor de mezcla descendente para reproducción estéreo

En todos los casos descritos anteriormente, el convertidor 501 de mezcla descendente de objetos a estéreo emite una aproximación a una mezcla descendente estéreo de la reproducción de canales 5.1 de los objetos de audio. Esta reproducción estéreo puede expresarse mediante una matriz  $A_2$   $2 \times N$  definida por  $A_2 = D_{26} A$ . En muchas aplicaciones esta mezcla descendente es interesante en sí misma y una manipulación directa de la reproducción estéreo  $A_2$  es atractiva. Considérese como ejemplo ilustrativo de nuevo el caso de una pista estéreo con una pista de voz mono con panorámica central superpuesta codificada siguiendo un caso especial del método expuesto en la figura 8 y comentado en la sección en torno a la fórmula (12). Un control de usuario del volumen de voz puede realizarse mediante la reproducción

$$20 \quad \mathbf{A}_2 = \frac{1}{\sqrt{1+v^2}} \begin{bmatrix} 1 & 0 & v/\sqrt{2} \\ 0 & 1 & v/\sqrt{2} \end{bmatrix}, \quad (33)$$

25 donde v es el control de cociente de voz a música. El diseño de la matriz de convertidor de mezcla descendente se basa en

$$\mathbf{GDS} \approx \mathbf{A}_2 \mathbf{S}. \quad (34)$$

Para los parámetros de objeto basados en predicción, simplemente se inserta la aproximación  $S \approx CDS$  y se obtiene la matriz de convertidor  $G \approx A_2 C$ . Para parámetros de objeto basados en energía, se resuelven las ecuaciones normales

$$30 \quad \mathbf{G}(\mathbf{DED}^*) = \mathbf{A}_2 \mathbf{ED}^*. \quad (35)$$

La figura 9 ilustra una realización preferida de un codificador de objetos de audio según un aspecto de la presente invención. El codificador 101 de objetos de audio ya se ha descrito generalmente en conexión con las figuras anteriores. El codificador de objetos de audio para generar la señal de objeto codificada usa la pluralidad de objetos 90 de audio que se han indicado en la figura 9 cuando entran en un mezclador 92 descendente y un generador 94 de parámetros de objeto. Además, el codificador 101 de objetos de audio incluye el generador 96 de información de mezcla descendente para generar información 97 de mezcla descendente que indica una distribución de la pluralidad de objetos de audio en al menos dos canales de mezcla descendente indicados en 93 cuando salen del mezclador 92 descendente.

El generador de parámetros de objeto es para generar parámetros 95 de objeto para los objetos de audio, en el que los parámetros de objeto se calculan de manera que es posible la reconstrucción del objeto de audio usando los parámetros de objeto y al menos dos canales 93 de mezcla descendente. Notablemente, sin embargo, esta reconstrucción no tiene lugar en el lado de codificador, sino que tiene lugar en el lado de decodificador. No obstante, el generador de parámetros de objeto en el lado de codificador calcula los parámetros de objeto para los objetos 95 de modo que esta reconstrucción total pueda realizarse en el lado de decodificador.

Además, el codificador 101 de objetos de audio incluye una interfaz 98 de salida para generar la señal 99 de objeto de audio codificada usando la información 97 de mezcla descendente y los parámetros 95 de objeto. Dependiendo de la aplicación, los canales 93 de mezcla descendente también pueden usarse y codificarse en la señal de objeto de audio

codificada. Sin embargo, también puede haber situaciones en las que la interfaz 98 de salida genera una señal 99 de objeto de audio codificada que no incluya los canales de mezcla descendente. Esta situación puede surgir cuando cualquier canal de mezcla descendente que va a usarse en el lado de decodificador ya está en el lado de decodificador, de modo que la información de mezcla descendente y los parámetros de objeto para los objetos de audio se transmitan por separado de los canales de mezcla descendente. Tal situación es útil cuando los canales 93 de mezcla descendente de objetos pueden adquirirse por separado a partir de los parámetros de objeto y la información de mezcla descendente por una cantidad menor de dinero, y los parámetros de objeto y la información de mezcla descendente pueden adquirirse por una cantidad adicional de dinero con el fin de proporcionar al usuario en el lado de decodificador un valor añadido.

Sin los parámetros de objeto y la información de mezcla descendente, un usuario puede reproducir los canales de mezcla descendente como una señal estéreo o multicanal dependiendo del número de canales incluidos en la mezcla descendente. Naturalmente, el usuario también podría reproducir una señal mono añadiendo simplemente los al menos dos canales de mezcla descendente de objetos transmitidos. Para aumentar la flexibilidad de reproducir y escuchar calidad y utilidad, los parámetros de objeto y la información de mezcla descendente permiten que el usuario forme una reproducción flexible de los objetos de audio en cualquier configuración de reproducción de audio prevista, tal como un sistema estéreo, un sistema multicanal o incluso un sistema de síntesis de campo de onda. Mientras que los sistemas de síntesis de campo de onda aún no son muy populares, los sistemas multicanal tales como sistemas 5.1 o sistemas 7.1 están volviéndose cada vez más populares en el mercado de consumo.

La figura 10 ilustra un sintetizador de audio para generar datos de salida. Para este fin, el sintetizador de audio incluye un sintetizador 100 de datos de salida. El sintetizador de datos de salida recibe, como entrada, la información 97 de mezcla descendente y los parámetros 95 de objeto de audio y, probablemente, los datos de fuente de audio previstos tales como un posicionamiento de las fuentes de audio o un volumen de especificado por el usuario de una fuente específica, al que debería estar la fuente cuando se reproduce, tal como se indica en 101.

El sintetizador 100 de datos de salida es para generar datos de salida que pueden usarse para crear una pluralidad de canales de salida de una configuración de salida de audio predefinida que representa una pluralidad de objetos de audio. Particularmente, el sintetizador 100 de datos de salida es operativo para el uso de la información 97 de mezcla descendente, y los parámetros 95 de objeto de audio. Tal como se comenta en conexión con la figura 11 más adelante, los datos de salida pueden ser datos de una gran variedad de diferentes aplicaciones útiles, que incluyen la reproducción específica de canales de salida o que incluyen sólo una reconstrucción de las señales fuente o que incluyen una transcodificación de parámetros en parámetros de reproducción espacial para una configuración de mezcla ascendente espacial sin ninguna reproducción específica de canales de salida, sino por ejemplo para almacenar o transmitir tales parámetros espaciales.

El escenario de aplicación general de la presente invención se resume en la figura 14. Hay un lado 140 de codificador que incluye el codificador 101 de objetos de audio que recibe, como entrada, N objetos de audio. La salida del codificador de objetos de audio preferido comprende, además de la información de mezcla descendente y los parámetros de objeto que no se muestran en la figura 14, los K canales de mezcla descendente. El número de canales de mezcla descendente según la presente invención es mayor que o igual a dos.

Los canales de mezcla descendente se transmiten a un lado 142 de decodificador, que incluye un mezclador 143 ascendente espacial. El mezclador 143 ascendente espacial puede incluir el sintetizador de audio de la invención, cuando el sintetizador de audio se opera en un modo de transcodificador. Sin embargo, cuando el sintetizador 101 de audio tal como se ilustra en la figura 10 trabaja en un modo de mezclador ascendente espacial, entonces el mezclador 143 ascendente espacial y el sintetizador de audio son el mismo dispositivo en esta realización. El mezclador ascendente espacial genera M canales de salida que van a reproducirse a través de M altavoces. Estos altavoces se colocan en ubicaciones espaciales predefinidas y conjuntamente representan la configuración de salida de audio predefinida. Un canal de salida de la configuración de salida de audio predefinida puede considerarse como una señal de altavoz digital o analógica que va a enviarse desde una salida del mezclador 143 ascendente espacial hasta la entrada de un altavoz en una posición predefinida entre la pluralidad de posiciones predefinidas de la configuración de salida de audio predefinida. Dependiendo de la situación, el número de M canales de salida puede ser igual a dos cuando se realiza la reproducción estéreo. Sin embargo, cuando se realiza una reproducción multicanal, entonces el número de M canales de salida es mayor que dos. Normalmente, habrá una situación en la que el número de canales de mezcla descendente sea más pequeño que el número de canales de salida debido a un requisito de un enlace de transmisión. En este caso, M es mayor que K y puede ser incluso mucho mayor que K, tal como doblar el tamaño o incluso más.

La figura 14 incluye además varias notaciones de matriz con el fin de ilustrar la funcionalidad del lado de codificador de la invención y el lado de decodificador de la invención. Generalmente, se procesan bloques de valores de muestreo. Por tanto, tal como se indica en la ecuación (2), se representa un objeto de audio como una línea de L valores de muestreo. La matriz S tiene N líneas que corresponden al número de objetos y L columnas que corresponden al número de muestras. Se calcula la matriz E tal como se indica en la ecuación (5) y tiene N columnas y N líneas. La matriz E incluye los parámetros de objeto cuando se proporcionan los parámetros de objeto en el modo de energía. Para objetos no correlacionados, la matriz E tiene, tal como se indicó anteriormente en conexión con la ecuación (6), sólo elementos en la diagonal principal, donde un elemento en la diagonal principal da la energía de un objeto de audio. Todos los

elementos fuera de la diagonal representan, tal como se indicó anteriormente, una correlación de dos objetos de audio, que es específicamente útil cuando algunos objetos son dos canales de la señal estéreo.

Dependiendo de la realización específica, la ecuación (2) es una señal en el dominio del tiempo. Entonces, se genera un único valor de energía para la banda entera de objetos de audio. Preferiblemente, sin embargo, los objetos de audio se procesan por un convertidor de tiempo/frecuencia que incluye, por ejemplo, un tipo de transformada o un algoritmo de banco de filtros. En este último caso, la ecuación (2) es válida para cada subbanda de modo que se obtiene una matriz E para cada subbanda y, naturalmente, cada trama de tiempo.

La matriz de canal de mezcla descendente X tiene K líneas y L columnas y se calcula tal como se indica en la ecuación (3). Tal como se indica en la ecuación (4), los M canales de salida se calculan usando los N objetos aplicando la denominada matriz de reproducción A a los N objetos. Dependiendo de la situación, los N objetos pueden regenerarse en el lado de decodificador usando la mezcla descendente y los parámetros de objeto y la reproducción puede aplicarse a las señales de objeto reconstruidas directamente.

Alternativamente, la mezcla descendente puede transformarse directamente a los canales de salida sin un cálculo explícito de las señales fuente. Generalmente, la matriz de reproducción A indica el posicionamiento de las fuentes individuales con respecto a la configuración de salida de audio predefinida. Si se tuvieran seis objetos y seis canales de salida, entonces podría situarse cada objeto en cada canal de salida y la matriz de reproducción reflejaría este esquema. Sin embargo, si se desea situar todos los objetos entre dos ubicaciones de altavoz de salida, entonces la matriz de reproducción A parecería diferente y reflejaría esta situación diferente.

La matriz de reproducción o, expresado de manera más general, el posicionamiento previsto de los objetos y también un volumen relativo previsto de las fuentes de audio pueden calcularse en general por un codificador y transmitirse al decodificador como una denominada descripción de escena. En otras realizaciones, sin embargo, esta descripción de escena puede generarse por el propio usuario para generar la mezcla ascendente específica del usuario para la configuración de salida de audio específica del usuario. Por tanto, no se requiere necesariamente una transmisión de la descripción de escena, sino que la descripción de escena también puede generarse por el usuario con el fin de cumplir los deseos del usuario. El usuario podría desear situar, por ejemplo, determinados objetos de audio en lugares que son diferentes de los lugares en los que estaban estos objetos cuando se generaron estos objetos. También hay casos en los que los objetos de audio se diseñan en sí mismos y no tienen ninguna ubicación "original" con respecto a los otros objetos. En esta situación, la ubicación relativa de las fuentes de audio se genera por el usuario por primera vez.

Volviendo a la figura 9, se ilustra un mezclador 92 descendente. El mezclador descendente es para la mezcla descendente de la pluralidad de objetos de audio en la pluralidad de canales de mezcla descendente, en el que el número de objetos de audio es mayor que el número de canales de mezcla descendente, y en el que el mezclador descendente se acopla al generador de información de mezcla descendente de modo que la distribución de la pluralidad de objetos de audio en la pluralidad de canales de mezcla descendente se lleva a cabo según se indique en la información de mezcla descendente. La información de mezcla descendente generada por el generador 96 de información de mezcla descendente en la figura 9 puede crearse automáticamente o ajustarse manualmente. Se prefiere proporcionar la información de mezcla descendente con una resolución menor que la resolución de los parámetros de objeto. Por tanto, pueden ahorrarse bits de información secundaria sin mayores pérdidas de calidad, puesto que se ha demostrado que basta con información de mezcla descendente fija para una determinada pieza de audio o una situación de mezcla descendente que sólo cambia lentamente, que no necesariamente tiene que ser selectiva en frecuencia. En una realización, la información de mezcla descendente representa una matriz de mezcla descendente que tiene K líneas y N columnas.

El valor en una línea de la matriz de mezcla descendente tiene un determinado valor cuando el objeto de audio correspondiente a este valor en la matriz de mezcla descendente está en el canal de mezcla descendente representado por la fila de la matriz de mezcla descendente. Cuando un objeto de audio se incluye en más de un canal de mezcla descendente, los valores de más de una fila de la matriz de mezcla descendente tienen un determinado valor. Sin embargo, se prefiere que los valores cuadrados cuando se suman entre sí para un único objeto de audio sumen 1.0. Sin embargo, otros valores son posibles también. Adicionalmente, pueden introducirse objetos de audio en uno o más canales de mezcla descendente con diversos niveles, y estos niveles pueden indicarse por pesos en la matriz de mezcla descendente que son diferentes de uno y que no suman 1.0 para un determinado objeto de audio.

Cuando los canales de mezcla descendente se incluyen en la señal de objeto de audio codificada generada por la interfaz 98 de salida, la señal de objeto de audio codificada puede ser por ejemplo una señal de multiplexación en el tiempo en un determinado formato. Alternativamente, la señal de objeto de audio codificada puede ser cualquier señal que permita la separación de los parámetros 95 de objeto, la información 97 de mezcla descendente y los canales 93 de mezcla descendente en un lado de decodificador. Además, la interfaz 98 de salida puede incluir codificadores para los parámetros de objeto, la información de mezcla descendente o los canales de mezcla descendente. Los codificadores para los parámetros de objeto y la información de mezcla descendente pueden ser codificadores diferenciales y/o codificadores de entropía, y los codificadores para los canales de mezcla descendente pueden ser codificadores de audio mono o estéreo tales como codificadores de MP3 o codificadores de AAC. Todas estas operaciones de codificación dan como resultado una compresión de datos adicional con el fin de disminuir además la tasa de transmisión de datos requerida para la señal 99 de objeto de audio codificada.

Dependiendo de la aplicación específica, el mezclador 92 descendente es operativo para incluir la representación estéreo de música de fondo en los al menos dos canales de mezcla descendente y además introduce la pista de voz en los al menos dos canales de mezcla descendente en una relación predefinida. En esta realización, un primer canal de la música de fondo está dentro del primer canal de mezcla descendente y el segundo canal de la música de fondo está dentro del segundo canal de mezcla descendente. Esto da como resultado una reproducción óptima de la música de fondo estéreo en un dispositivo de reproducción estéreo. El usuario aún puede modificar, sin embargo, la posición de la pista de voz entre el altavoz estéreo izquierdo y el altavoz estéreo derecho. Alternativamente, los canales de música de fondo primero y segundo pueden incluirse en un canal de mezcla descendente y la pista de voz puede incluirse en el otro canal de mezcla descendente. Por tanto, al eliminar un canal de mezcla descendente, puede separarse totalmente la pista de voz de la música de fondo, lo que es particularmente adecuado para aplicaciones de karaoke. Sin embargo, la calidad de reproducción estéreo de los canales de música de fondo se verá afectada debido a la parametrización de objeto que, naturalmente, es un método de comprensión con pérdidas.

Se adapta un mezclador 92 descendente para realizar una suma muestra a muestra en el dominio del tiempo. Esta adición usa muestras a partir de objetos de audio que van a mezclarse de manera descendente en un único canal de mezcla descendente. Cuando va a introducirse un objeto de audio en un canal de mezcla descendente con un determinado porcentaje, tiene lugar una ponderación previa antes del proceso de suma con por muestras. Alternativamente, la suma también puede tener lugar en el dominio de la frecuencia, o un dominio de subbanda, es decir, en un dominio posterior a la conversión de tiempo/frecuencia. Por tanto, incluso podría realizarse la mezcla descendente en el dominio de banco de filtros cuando la conversión de tiempo/frecuencia es un banco de filtros o en el dominio de transformada cuando la conversión de tiempo/frecuencia es un tipo de FFT, MDCT o cualquier otra transformada.

En un aspecto de la presente invención, el generador 94 de parámetros de objeto genera parámetros de energía y, adicionalmente, los parámetros de correlación entre dos objetos cuando dos objetos de audio conjuntamente representan la señal estéreo, tal como queda claro por la ecuación posterior (6). Alternativamente, los parámetros de objeto son parámetros de modo de predicción. La figura 15 ilustra etapas de algoritmo o medios de un dispositivo de cálculo para calcular estos parámetros de predicción de objeto de audio. Tal como se ha comentado en conexión con las ecuaciones (7) a (12), tiene que se calcularse cierta información estadística en los canales de mezcla descendente en la matriz  $X$  y los objetos de audio en la matriz  $S$ . Particularmente, el bloque 150 ilustra la primera etapa de cálculo de la parte real de  $S \cdot X^*$  y la parte real de  $X \cdot X^*$ . Estas partes reales no son sólo números sino matrices, y estas matrices se determinan en una realización a través de las notaciones en la ecuación (1) cuando se considera la realización posterior a la ecuación (12). Generalmente, los valores de la etapa 150 pueden calcularse usando datos disponibles en el codificador 101 de objeto de audio. Entonces, la matriz de predicción  $C$  se calcula tal como se ilustra en la etapa 152. Particularmente, el sistema de ecuaciones se resuelve tal como se conoce en la técnica de modo que se obtienen todos los valores de la matriz de predicción  $C$  que tiene  $N$  líneas y  $K$  columnas. Generalmente, los factores de ponderación  $c_{n,i}$  tal como se proporcionan en la ecuación (8) se calculan de manera que la adición lineal ponderada de todos los canales de mezcla descendente reconstruye un objeto de audio correspondiente lo mejor posible. Esta matriz de predicción da como resultado una mejor reconstrucción de objetos de audio cuando el número de canales de mezcla descendente aumenta.

A continuación se comentará la figura 11 en más detalle. Particularmente, la figura 7 ilustra varias clases de datos de salida que pueden usarse para crear una pluralidad de canales de salida de una configuración de salida de audio predefinida. La línea 111 ilustra una situación en la que los datos de salida del sintetizador 100 de datos de salida son fuentes de audio reconstruidas. Los datos de entrada requeridos por el sintetizador 100 de datos de salida para reproducir las fuentes de audio reconstruidas incluyen información de mezcla descendente, los canales de mezcla descendente y los parámetros de objeto de audio. Para reproducir las fuentes reconstruidas, sin embargo, una configuración de salida y un posicionamiento previsto de las propias fuentes de audio en la configuración de salida de audio espacial no se requieren necesariamente. En este primer modo indicado por el número de modo 1 en la figura 11, el sintetizador 100 de datos de salida emitirá fuentes de audio reconstruidas. En el caso de parámetros de predicción como parámetros de objeto de audio, el sintetizador 100 de datos de salida funciona tal como se define por la ecuación (7). Cuando los parámetros de objeto están en el modo de energía, entonces el sintetizador de datos de salida usa una inversa de la matriz de mezcla descendente y la matriz de energía para reconstruir las señales fuente.

Alternativamente, el sintetizador 100 de datos de salida opera como un transcodificador tal como se ilustra por ejemplo en el bloque 102 en la figura 1b. Cuando el sintetizador de salida es un tipo de transcodificador para generar parámetros de mezclador espaciales, se requiere la información de mezcla descendente, los parámetros de objeto de audio, la configuración de salida y el posicionamiento previsto de las fuentes. Particularmente, la configuración de salida y el posicionamiento previsto se proporcionan a través de la matriz de reproducción  $A$ . Sin embargo, no se requieren los canales de mezcla descendente para generar los parámetros de mezclador espaciales tal como se comentará en más detalle en conexión con la figura 12. Dependiendo de la situación, los parámetros de mezclador espaciales generados por el sintetizador 100 de datos de salida pueden entonces usarse por un mezclador espacial directo tal como un mezclador de *MPEG-surround* para mezclar de manera ascendente los canales de mezcla descendente. Esta realización no necesita necesariamente modificar los canales de mezcla descendente de objetos, pero puede proporcionar una matriz de conversión simple que sólo tiene elementos diagonales tal como se comentó en la ecuación (13). En el modo 2 tal como se indica por 112 en la figura 11, el sintetizador 100 de datos de salida emitirá, por tanto, parámetros de mezclador espaciales y, preferiblemente, la matriz de conversión  $G$  tal como se indica en la ecuación

(13), que incluye ganancias que pueden usarse como parámetros de ganancia de mezcla descendente arbitraria (ADG) del decodificador de *MPEG-surround*.

En el número de modo 3 tal como se indica por 113 de la figura 11, los datos de salida incluyen parámetros de mezclador espaciales en una matriz de conversión tal como la matriz de conversión ilustrada en conexión con la ecuación (25). En esta situación, el sintetizador 100 de datos de salida no necesariamente tiene que realizar la conversión de mezcla descendente real para convertir la mezcla descendente de objetos en una mezcla descendente estéreo.

Un modo diferente de operación indicado por el número de modo 4 en la línea 114 en la figura 11 ilustra el sintetizador 100 de datos de salida de la figura 10. En esta situación, el transcodificador se opera tal como se indica por 102 en la figura 1b y emite no sólo parámetros de mezclador espaciales sino que emite adicionalmente una mezcla descendente convertida. Sin embargo, ya no es necesario emitir la matriz de conversión G además de la mezcla descendente convertida. Emitir la mezcla descendente convertida y los parámetros de mezclador espaciales es suficiente tal como se indica por la figura 1b.

El número de modo 5 indica otro uso del sintetizador 100 de datos de salida ilustrado en la figura 10. En esta situación indicada por la línea 115 en la figura 11, los datos de salida generados por el sintetizador de datos de salida no incluyen ningún parámetro de mezclador espacial sino que sólo incluyen una matriz de conversión G tal como se indica por la ecuación (35) por ejemplo o en realidad incluyen la salida de las propias señales estéreo tal como se indica en 115. En esta realización, sólo una reproducción estéreo es de interés y no se requiere ningún parámetro de mezclador espacial. Para generar la salida estéreo, sin embargo, se requiere toda la información de entrada disponible tal como se indica en la figura 11.

Se indica otro modo de sintetizador de datos de salida mediante el número de modo 6 en la línea 116. En este caso, el sintetizador 100 de datos de salida genera una salida multicanal, y el sintetizador 100 de datos de salida sería similar al elemento 104 en la figura 1b. Para este fin, el sintetizador 100 de datos de salida requiere toda la información de entrada disponible y emite una señal de salida multicanal que tiene más de dos canales de salida que van a producirse mediante un número correspondiente de altavoces que van a colocarse en posiciones de altavoz previstas según la configuración de salida de audio predefinida. Tal salida multicanal es una salida 5.1, una salida 7.1 o sólo una salida 3.0 que tiene un altavoz izquierdo, un altavoz central y un altavoz derecho.

A continuación se hace referencia a la figura 11 para ilustrar un ejemplo para calcular varios parámetros a partir del concepto de parametrización de la figura 7 conocido del decodificador de *MPEG-surround*. Tal como se indica, la figura 7 ilustra una parametrización de lado de decodificador de *MPEG-surround* que parte de la mezcla 70 descendente estéreo que tiene un canal de mezcla descendente izquierdo  $l_0$  y un canal de mezcla descendente derecho  $r_0$ . Conceptualmente, ambos canales de mezcla descendente se introducen en una denominada casilla 71 de dos a tres. La casilla de dos a tres se controla por varios parámetros 72 de entrada. La casilla 71 genera tres canales 73a, 73b, 73c de salida. Cada canal de salida se introduce en una casilla de uno a dos. Esto significa que el canal 73a se introduce en la casilla 74a, el canal 73b se introduce en la casilla 74b, y el canal 73c se introduce en la casilla 74c. Cada casilla emite dos canales de salida. La casilla 74a emite un canal frontal izquierdo  $l_f$  y un canal envolvente izquierdo  $l_s$ . Además, la casilla 74b emite un canal frontal derecho  $r_f$  y un canal envolvente derecho  $r_s$ . Además, la casilla 74c emite un canal central  $c$  y un canal de mejora de baja frecuencia  $l_{fe}$ . Notablemente, se realiza la mezcla ascendente entera desde los canales 70 de mezcla descendente hasta los canales de salida usando una operación de matriz, y la estructura en árbol tal como se muestra en la figura 7 no se implementa necesariamente etapa por etapa sino que puede implementarse a través de una única o varias operaciones de matriz. Además, las señales intermedias indicadas por 73a, 73b y 73c no se calculan explícitamente mediante una determinada realización, pero se ilustran en la figura 7 sólo por motivos de ilustración. Además, las casillas 74a, 74b reciben algunas señales residuales  $res_1^{OTT}$ ,  $res_2^{OTT}$  que pueden usarse para introducir una determinada aleatoriedad en las señales de salida.

Tal como se muestra a partir del decodificador de *MPEG-surround*, la casilla 71 se controla mediante o bien parámetros de predicción CPC o bien parámetros de energía  $CLD_{TTT}$ . Para la mezcla ascendente desde dos canales hasta tres canales, se requieren al menos dos parámetros de predicción CPC1, CPC2 o al menos dos parámetros de energía  $CLD^1_{TTT}$  y  $CLD^2_{TTT}$ . Además, la medida de correlación  $ICC_{TTT}$  puede ponerse en la casilla 71 que es, sin embargo, sólo una característica opcional que no se usa en una realización de la invención. Las figuras 12 y 13 ilustran las etapas necesarias y/o medios para calcular todos los parámetros CPC/CLD<sub>TTT</sub>, CLD0, CLD1, ICC1, CLD2, ICC2 a partir de los parámetros 95 de objeto de la figura 9, la información 97 de mezcla descendente de la figura 9 y el posicionamiento previsto de las fuentes de audio, por ejemplo la descripción 101 de escena tal como se ilustra en la figura 10. Estos parámetros son para el formato de salida de audio predefinido de un sistema envolvente 5.1.

Naturalmente, el cálculo específico de parámetros para esta implementación específica puede adaptarse para otros formatos o parametrizaciones de salida en vista de las enseñanzas de este documento. Además, la secuencia de las etapas o la disposición de medios en las figuras 12 y 13a,b sólo es a modo de ejemplo y puede cambiarse dentro del sentido lógico de las ecuaciones matemáticas.

En la etapa 120, se proporciona una matriz de reproducción A. La matriz de reproducción indica dónde va a situarse la fuente de la pluralidad de fuentes en el contexto de la configuración de salida predefinida. La etapa 121 ilustra la

5 derivación de la matriz de mezcla descendente parcial  $D_{36}$  tal como se indicó en la ecuación (20). Esta matriz refleja la situación de una mezcla descendente desde seis canales de salida hasta tres canales y tiene un tamaño de  $3 \times N$ . Cuando se pretende generar más canales de salida que la configuración 5.1, tal como una configuración de salida de 8 canales (7.1), entonces la matriz determinada en el bloque 121 sería una matriz  $D_{38}$ . En la etapa 122, se genera una matriz de reproducción reducida  $A_3$  multiplicando la matriz  $D_{36}$  y la matriz de reproducción total tal como se define en la etapa 120. En la etapa 123, se introduce la matriz de mezcla descendente  $D$ . Esta matriz de mezcla descendente  $D$  puede recuperarse de la señal de objeto de audio codificada cuando la matriz se incluye totalmente en esta señal. Alternativamente, la matriz de mezcla descendente podría parametrizarse por ejemplo para el ejemplo específico de la información de mezcla descendente y la matriz de mezcla descendente  $G$ .

10 Además, se proporciona la matriz de energía de objeto en la etapa 124. Esta matriz de energía de objeto se refleja mediante los parámetros de objeto para los  $N$  objetos y puede extraerse de los objetos de audio importados o reconstruidos usando una determinada regla de reconstrucción. Esta regla de reconstrucción puede incluir una decodificación de entropía, etc.

15 En la etapa 125, se define la matriz de predicción "reducida"  $C_3$ . Los valores de esta matriz pueden calcularse resolviendo el sistema de ecuaciones lineales tal como se indica en la etapa 125. Específicamente, los elementos de matriz  $C_3$  pueden calcularse multiplicando la ecuación en ambos lados mediante una inversa de  $(DED^*)$ .

20 En la etapa 126, se calcula la matriz de conversión  $G$ . La matriz de conversión  $G$  tiene un tamaño de  $K \times K$  y se genera tal como se define por la ecuación (25). Para resolver la ecuación en la etapa 126, va a proporcionarse la matriz específica  $D_{TTT}$  tal como se indica por la etapa 127. Un ejemplo para esta matriz viene dado por la ecuación (24) y la definición puede derivarse de la ecuación correspondiente para  $C_{TTT}$  tal como se definió en la ecuación (22). La ecuación (22), por tanto, define lo que se realizará en la etapa 128. La etapa 129 define las ecuaciones para calcular la matriz  $C_{TTT}$ . Tan pronto como se determine la matriz  $C_{TTT}$  según la ecuación en el bloque 129, pueden producirse los parámetros  $\alpha$ ,  $\beta$  y  $\gamma$ , que son los parámetros CPC. Preferiblemente,  $\gamma$  se ajusta a 1 de modo los únicos parámetros CPC restantes introducidos en el bloque 71 son  $\alpha$  y  $\beta$ .

25 Los parámetros restantes necesarios para el esquema en la figura 7 son los parámetros introducidos en los bloques 74a, 74b y 74c. El cálculo de estos parámetros se comenta en conexión con la figura 13a. En la etapa 130, se proporciona la matriz de reproducción  $A$ . El tamaño de la matriz de reproducción  $A$  es  $N$  líneas para el número de objetos de audio y  $M$  columnas para el número de canales de salida. Esta matriz de reproducción incluye la información del vector de escena, cuando se usa un vector de escena. Generalmente, la matriz de reproducción incluye la información de situar una fuente de audio en una determinada posición en una configuración de salida. Cuando se considera, por ejemplo, la matriz de reproducción  $A$  bajo la ecuación (19), resulta evidente cómo una determinada colocación de objetos de audio puede codificarse dentro de la matriz de reproducción. Naturalmente, pueden usarse otras maneras de indicar una determinada posición, tal como mediante valores no iguales a 1. Además, cuando se usan valores que son menores que 1 por un lado y son mayores que 1 por otro lado, la sonoridad de determinados objetos de audio puede verse influida también.

30 En una realización, la matriz de reproducción se genera en el lado de decodificador sin ninguna información desde el lado de codificador. Esto permite que un usuario sitúe los objetos de audio en cualquier lugar que el usuario desee sin prestar atención a una relación espacial de los objetos de audio en la configuración de codificador. En otra realización, la ubicación relativa o absoluta de fuentes de audio puede codificarse en el lado de codificador y transmitirse al decodificador como una clase de un vector de escena. Entonces, en el lado de decodificador, esta información sobre ubicaciones de fuentes de audio que es preferiblemente independiente de una configuración de reproducción de audio prevista se procesa para dar como resultado una matriz de reproducción que refleja las ubicaciones de las fuentes de audio personalizadas a la configuración de salida de audio específica.

35 En la etapa 131, se proporciona la matriz de energía de objeto  $E$  que ya se ha comentado en conexión con la etapa 124 de la figura 12. Esta matriz tiene el tamaño de  $N \times N$  e incluye los parámetros de objeto de audio. En una realización, tal matriz de energía de objeto se proporciona para cada subbanda y cada bloque muestras en el dominio del tiempo o muestras en el dominio de subbanda.

40 En la etapa 132, se calcula la matriz de energía de salida  $F$ .  $F$  es la matriz de covarianza de los canales de salida. Puesto que los canales de salida son, sin embargo, aún desconocidos, la matriz de energía de salida  $F$  se calcula usando la matriz de reproducción y la matriz de energía. Estas matrices se proporcionan en las etapas 130 y 131 y están disponibles fácilmente en el lado de decodificador. Entonces, las ecuaciones específicas (15), (16), (17), (18) y (19) se aplican para calcular los parámetros de diferencia de nivel de canales  $CLD_0$ ,  $CLD_1$ ,  $CLD_2$  y los parámetros de coherencia entre canales  $ICC_1$  e  $ICC_2$  de modo que están disponibles los parámetros para las casillas 74a, 74b, 74c. Notablemente, los parámetros espaciales se calculan combinando los elementos específicos de la matriz de energía de salida  $F$ .

45 Después de la etapa 133, están disponibles todos los parámetros para un mezclador ascendente espacial, tal como el mezclador ascendente espacial tal como se ilustra esquemáticamente en la figura 7.

En las realizaciones anteriores, se proporcionaron los parámetros de objeto como parámetros de energía. Sin embargo, cuando los parámetros de objeto se proporcionan como parámetros de predicción, es decir como una matriz de predicción de objeto C tal como se indica por el elemento 124a en la figura 12, el cálculo de la matriz de predicción reducida  $C_3$  es sólo una multiplicación de matrices tal como se ilustra en el bloque 125a y se comentó en conexión con la ecuación (32). La matriz  $A_3$  tal como se usa en el bloque 125a es la misma matriz  $A_3$  que se mencionó en el bloque 122 de la figura 12.

Cuando la matriz de predicción de objeto C se genera por un codificador de objetos de audio y se transmite al decodificador, entonces se requieren algunos cálculos adicionales para generar los parámetros para las casillas 74a, 74b, 74c. Estas etapas adicionales se indican en la figura 13b. De nuevo, se proporciona la matriz de predicción de objeto C tal como se indica por 124a en la figura 13b, que es la misma que se comentó en conexión con el bloque 124a de la figura 12. Entonces, tal como se comentó en conexión con la ecuación (31), se calcula la matriz de covarianza de la mezcla descendente de objetos Z usando la mezcla descendente transmitida o se genera y se transmite como información secundaria adicional. Cuando se transmite la información en la matriz Z, entonces el decodificador no necesariamente tiene que realizar ningún cálculo de energía que de manera inherente introduce algún procesamiento retardado y aumenta la carga de procesamiento en el lado de decodificador. Sin embargo, cuando estos temas no son decisivos para una determinada aplicación, entonces puede ahorrarse ancho de banda de transmisión y la matriz de covarianza Z de la mezcla descendente de objetos también puede calcularse usando las muestras de mezcla descendente que, naturalmente, están disponibles en el lado de decodificador. Tan pronto como se complete la etapa 134 y la matriz de covarianza de la mezcla descendente de objetos esté lista, la matriz de energía de objeto E puede calcularse tal como se indica por la etapa 135 usando la matriz de predicción C y la matriz de covarianza de mezcla descendente o de "energía de mezcla descendente" Z. Tan pronto como se complete la etapa 135, pueden realizarse todas las etapas comentadas en conexión con la figura 13a, tal como las etapas 132, 133, para generar todos los parámetros para los bloques 74a, 74b, 74c de la figura 7.

La figura 16 ilustra una realización adicional, en la que sólo se requiere una reproducción estéreo. La reproducción estéreo es la salida tal como se proporcionó por el número de modo 5 o línea 115 de la figura 11. En este caso, el sintetizador 100 de datos de salida de la figura 10 no es interesante en ningún parámetro de mezcla ascendente espacial pero es interesante principalmente en una matriz de conversión específica G para convertir la mezcla descendente de objetos en una mezcla descendente estéreo útil y, naturalmente, fácilmente influenciada y fácilmente controlable.

En la etapa 160 de la figura 16, se calcula una matriz de mezcla descendente parcial de M a 2. En el caso de seis canales de salida, la matriz de mezcla descendente parcial sería una matriz de mezcla descendente de seis a dos canales, pero otras matrices de mezcla descendente están disponibles también. El cálculo de esta matriz de mezcla descendente parcial puede derivarse, por ejemplo, de la matriz de mezcla descendente parcial  $D_{36}$  tal como se generó en la etapa 121 y la matriz  $D_{TTT}$  tal como se usó en la etapa 127 de la figura 12.

Además, una matriz de reproducción estéreo  $A_2$  se genera usando el resultado de la etapa 160 y la matriz de reproducción "grande" A se ilustra en la etapa 161. La matriz de reproducción A es la misma matriz que se ha comentado en conexión con el bloque 120 en la figura 12.

Posteriormente, en la etapa 162, la matriz de reproducción estéreo puede parametrizarse mediante parámetros de colocación  $\mu$  y  $\kappa$ . Cuando  $\mu$  se ajusta a 1 y  $\kappa$  se ajusta a 1 también, entonces se obtiene la ecuación (33), que permite una variación del volumen de voz en el ejemplo descrito en conexión con la ecuación (33). Sin embargo, cuando se usan otros parámetros tales como  $\mu$  y  $\kappa$ , entonces la colocación de las fuentes pueden variarse también.

Entonces, tal como se indica en la etapa 163, se calcula la matriz de conversión G usando la ecuación (33). Particularmente, puede calcularse la matriz  $(DED^*)$ , invertirse y la matriz invertida puede multiplicarse en el lado derecho de la ecuación en el bloque 163. Naturalmente, pueden aplicarse otros métodos para resolver la ecuación en el bloque 163. Entonces, se tiene la matriz de conversión G, y la mezcla descendente de objetos X puede convertirse multiplicando la matriz de conversión y la mezcla descendente de objetos tal como se indica en el bloque 164. Entonces, la mezcla descendente convertida X' puede reproducirse en estéreo usando dos altavoces estéreo. Dependiendo de la implementación, determinados valores para  $\mu$ ,  $\nu$  y  $\kappa$  pueden ajustarse para calcular la matriz de conversión G. Alternativamente, la matriz de conversión G puede calcularse usando estos tres parámetros como variables de modo que los parámetros puedan ajustarse después de la etapa 163 según se requiera por el usuario.

Realizaciones preferidas resuelven el problema de transmitir un número de objetos de audio individuales (usando una mezcla descendente multicanal y datos de control adicionales que describen los objetos) y reproducir los objetos a un sistema de reproducción dado (configuración de altavoces). Se introduce una técnica de cómo modificar los datos de control relacionados con el objeto en datos de control que sean compatibles con el sistema de reproducción. Propone además métodos de codificación adecuados basados en el esquema de codificación de *MPEG Surround*.

Dependiendo de determinados requisitos de implementación de los métodos de la invención, los métodos y señales de la invención pueden implementarse en hardware o en software. La implementación puede realizarse usando un medio de almacenamiento digital, en particular un disco o un CD que tiene señales de control electrónicamente legibles almacenadas en el mismo, que puede cooperar con un sistema informático programable de manera que se realicen los

métodos de la invención. Generalmente, la presente invención es, por tanto, un producto de programa informático con un código de programa almacenado en un soporte legible por máquina, configurándose el código de programa para realizar al menos uno de los métodos de la invención, cuando el producto de programa informático se ejecuta en un ordenador. En otras palabras, los métodos de la invención son, por tanto, un programa informático que tiene un código de programa para realizar los métodos de la invención, cuando el programa informático se ejecuta en un ordenador.

En otras palabras, según una realización del presente caso, un codificador de objetos de audio para generar una señal de objeto de audio codificada usando una pluralidad de objetos de audio, comprende un generador de información de mezcla descendente para generar información de mezcla descendente que indica una distribución de la pluralidad de objetos de audio en al menos dos canales de mezcla descendente; un generador de parámetros de objeto para generar parámetros de objeto para los objetos de audio; y una interfaz de salida para generar la señal de objeto de audio codificada usando la información de mezcla descendente y los parámetros de objeto.

Opcionalmente, la interfaz de salida puede operarse para generar la señal de audio codificada usando además la pluralidad de canales de mezcla descendente.

Además o alternativamente, el generador de parámetros puede estar operativo para generar los parámetros de objeto con una primera resolución de tiempo y de frecuencia, y en el que el generador de información de mezcla descendente es operativo para generar la información de mezcla descendente con una segunda resolución de tiempo y de frecuencia, siendo la segunda resolución de tiempo y de frecuencia más pequeña que la primera resolución de tiempo y de frecuencia.

Además, el generador de información de mezcla descendente puede ser operativo para generar la información de mezcla descendente de manera que la información de mezcla descendente sea igual para toda la banda de frecuencia de los objetos de audio.

Además, el generador de información de mezcla descendente puede ser operativo para generar la información de mezcla descendente de manera que la información de mezcla descendente represente una matriz de mezcla descendente definida tal como se sigue:

$$X = DS$$

donde D es la matriz de mezcla descendente, y donde X es una matriz y representa la pluralidad de canales de mezcla descendente y tiene un número de líneas que es igual al número de canales de mezcla descendente.

Además, la información sobre una parte puede ser un factor menor a 1 y superior a 0.

Además, el mezclador descendente puede ser operativo para incluir la representación estéreo de música de fondo en los al menos dos canales de mezcla descendente, y para introducir una pista de voz en los al menos dos canales de mezcla descendente en una relación predefinida.

Además, el mezclador descendente puede ser operativo para realizar una suma por muestras de señales que van a introducirse en un canal de mezcla descendente según se indique por la información de mezcla descendente.

Además, la interfaz de salida puede ser operativa para realizar una compresión de datos de la información de mezcla descendente y los parámetros de objeto antes de generar la señal de objeto de audio codificada.

Además, la pluralidad de objetos de audio puede incluir un objeto estéreo representado por dos objetos de audio que tienen una determinada correlación distinta de cero, y en el que el generador de información de mezcla descendente genera una información de agrupación que indica los dos objetos de audio que forman el objeto estéreo.

Además, el generador de parámetros de objeto puede ser operativo para generar parámetros de predicción de objeto para los objetos de audio, calculándose los parámetros de predicción de manera que la suma ponderada de los canales de mezcla descendente para un objeto fuente controlado por los parámetros de predicción o el objeto fuente da como resultado una aproximación del objeto fuente.

Además, los parámetros de predicción pueden generarse por banda de frecuencia, y en los que los objetos de audio cubren una pluralidad de bandas de frecuencia.

Además, el número de objetos de audio puede ser igual a N, el número de canales de mezcla descendente es igual a K, y el número de parámetros de predicción de objeto calculados por el generador de parámetros de objeto es igual a o menor que N-K.

Además, el generador de parámetros de objeto puede ser operativo para calcular como máximo  $K \cdot (N-K)$  parámetros de predicción de objeto.

Además, el generador de parámetros de objeto puede incluir un mezclador ascendente para mezclar de manera ascendente la pluralidad de canales de mezcla descendente usando diferentes conjuntos de parámetros de predicción

de objeto de prueba; y en el que el codificador de objetos de audio comprende además un controlador de iteración para hallar los parámetros de predicción de objeto de prueba que dan como resultado la desviación más pequeña entre una señal fuente reconstruida por el mezclador ascendente y la señal de fuente original correspondiente entre los diferentes conjuntos de parámetros de predicción de objeto de prueba.

5 Además, el sintetizador de datos de salida puede ser operativo para determinar la matriz de conversión usando la información de mezcla descendente, en el que se calcula la matriz de conversión de modo que al menos partes de los canales de mezcla descendente se intercambian cuando un objeto de audio incluido en un primer canal de mezcla descendente que representa la primera mitad de un plano estéreo va a reproducirse en la segunda mitad del plano estéreo.

10 Además, el sintetizador de audio, puede comprender un reproductor de canales para reproducir canales de salida de audio para la configuración de salida de audio predefinida usando los parámetros espaciales y los al menos dos canales de mezcla descendente o los canales de mezcla descendente convertidos.

Además, el sintetizador de datos de salida puede ser operativo para emitir los canales de salida de la configuración de salida de audio predefinida usando además los al menos dos canales de mezcla descendente.

15 Además, el sintetizador de datos de salida puede ser operativo para calcular pesos de mezcla descendente reales para la matriz de mezcla descendente parcial de manera que una energía de una suma ponderada de dos canales es igual a las energías de los canales dentro de un factor límite.

Además, los pesos de mezcla descendente para la matriz de mezcla descendente parcial pueden determinarse tal como sigue:

20 
$$w_p^2(f_{2p-1,2p-1} + f_{2p,2p} + 2f_{2p-1,2p}) = f_{2p-1,2p-1} + f_{2p,2p}, p = 1,2,3,$$

donde  $w_p$  es un peso de mezcla descendente,  $p$  es una variable de índice entero,  $f_{ji}$  es un elemento de matriz de una matriz de energía que representa una aproximación de una matriz de covarianza de los canales de salida de la configuración de salida predefinida.

25 Además, el sintetizador de datos de salida puede ser operativo para calcular coeficientes separados de la matriz de predicción resolviendo un sistema de ecuaciones lineales.

Además, el sintetizador de datos de salida puede ser operativo para resolver el sistema de ecuaciones lineales basándose en:

$$C_3(DED^*) = A_3ED^*,$$

30 donde  $C_3$  es la matriz de predicción de dos a tres,  $D$  es la matriz de mezcla descendente derivada de la información de mezcla descendente,  $E$  es una matriz de energía derivada de los objetos de fuente de audio, y  $A_3$  es la matriz de mezcla descendente reducida, y donde "\*" indica la operación conjugada compleja.

35 Además, los parámetros de predicción para la mezcla ascendente de dos a tres pueden derivarse de una parametrización de la matriz de predicción de modo que la matriz de predicción se define usando sólo dos parámetros, y siendo el sintetizador de datos de salida operativo para procesar previamente los al menos dos canales de mezcla descendente de modo que el efecto del procesamiento previo y la matriz de predicción parametrizada corresponda a una matriz de mezcla ascendente deseada.

Además, la parametrización de la matriz de predicción puede ser tal como sigue:

$$C_{TTT} = \frac{\gamma}{3} \begin{bmatrix} \alpha + 2 & \beta - 1 \\ \alpha - 1 & \beta + 2 \\ 1 - \alpha & 1 - \beta \end{bmatrix},$$

donde el índice TTT es la matriz de predicción parametrizada, y donde  $\alpha$ ,  $\beta$  y  $\gamma$  son factores.

40 Además, puede calcularse una matriz de conversión de mezcla descendente  $G$  tal como sigue:

$$G = D_{TTT}C_3,$$

donde  $C_3$  es una matriz de predicción de dos a tres, donde  $D_{TTT}$  y  $C_{TTT}$  es igual a 1, donde  $I$  es una matriz de identidad de dos por dos, y donde  $C_{TTT}$  se basa en:

$$C_{TTT} = \frac{\gamma}{3} \begin{bmatrix} \alpha + 2 & \beta - 1 \\ \alpha - 1 & \beta + 2 \\ 1 - \alpha & 1 - \beta \end{bmatrix},$$

donde  $\alpha$ ,  $\beta$  y  $\gamma$  son factores constantes.

- 5 Además, los parámetros de predicción para la mezcla ascendente de dos a tres pueden determinarse como  $\alpha$  y  $\beta$ , donde  $\gamma$  se ajusta a 1.

Además, el sintetizador de datos de salida puede ser operativo para calcular los parámetros de energía para la mezcla ascendente de tres a seis usando una matriz de energía  $F$  basándose en:

$$YY^* \approx F = AEA^*,$$

- 10 donde  $A$  es la matriz de reproducción,  $E$  es la matriz de energía derivada de los objetos de fuente de audio,  $Y$  es una matriz de canal de salida y "\*" indica la operación conjugada compleja.

Además, el sintetizador de datos de salida puede ser operativo para calcular los parámetros de energía combinando elementos de la matriz de energía.

- 15 Además, el sintetizador de datos de salida puede ser operativo para calcular los parámetros de energía basándose en las siguientes ecuaciones:

$$CLD_0 = 10 \log_{10} \left( \frac{f_{55}}{f_{66}} \right),$$

$$CLD_1 = 10 \log_{10} \left( \frac{f_{33}}{f_{44}} \right),$$

$$CLD_2 = 10 \log_{10} \left( \frac{f_{11}}{f_{22}} \right),$$

$$ICC_1 = \frac{\varphi(f_{34})}{\sqrt{f_{33}f_{44}}},$$

$$ICC_2 = \frac{\varphi(f_{12})}{\sqrt{f_{11}f_{22}}},$$

- 20 donde  $\varphi$  es un operador de valor absoluto  $\varphi(z)=|z|$  o de valor real  $\varphi(z)=\text{Re}\{z\}$ , donde  $CLD_0$  es un primer parámetro de energía de diferencia de nivel de canal, donde  $CLD_1$  es un segundo parámetro de energía de diferencia de nivel de canal, donde  $CLD_2$  es un tercer parámetro de energía de diferencia de nivel de canal, donde  $ICC_1$  es un primer parámetro de energía de coherencia entre canales, y  $ICC_2$  es un segundo parámetro de energía de coherencia entre canales, y donde  $f_{ij}$  son elementos de una matriz de energía  $F$  en las posiciones  $ij$  en esta matriz.

Además, el primer grupo de parámetros puede incluir parámetros de energía, y siendo el sintetizador de datos de salida operativo para derivar los parámetros de energía combinando elementos de la matriz de energía  $F$ .

Además, los parámetros de energía pueden derivarse basándose en:

$$CLD_{TTT}^0 = 10 \log_{10} \left( \frac{\|p\|^2 + \|r\|^2}{\|c\|^2} \right) = 10 \log_{10} \left( \frac{f_{11} + f_{22} + f_{33} + f_{44}}{f_{55} + f_{66}} \right),$$

$$CLD_{TTT}^1 = 10 \log_{10} \left( \frac{\|p\|^2}{\|r\|^2} \right) = 10 \log_{10} \left( \frac{f_{11} + f_{22}}{f_{33} + f_{44}} \right),$$

donde  $CLD_{TTT}^0$  es un primer parámetro de energía del primer grupo y donde  $CLD_{TTT}^1$  es un segundo parámetro de energía del primer grupo de parámetros.

5 Además, el sintetizador de datos de salida puede ser operativo para calcular factores de peso para ponderar los canales de mezcla descendente, usándose los factores de peso para controlar factores de ganancia de mezcla descendente arbitraria del decodificador espacial.

Además, el sintetizador de datos de salida puede ser operativo para calcular los factores de peso basándose en:

$$Z = DED^*,$$

$$W = D_{26}ED_{26}^*,$$

$$G = \begin{bmatrix} \sqrt{w_{11}/z_{11}} & 0 \\ 0 & \sqrt{w_{22}/z_{22}} \end{bmatrix},$$

10 donde D es la matriz de mezcla descendente, E es una matriz de energía derivada de los objetos de fuente de audio, donde W es una matriz intermedia, donde  $D_{26}$  es la matriz de mezcla descendente parcial para mezcla descendente de 6 a 2 canales de la configuración de salida predeterminada, y donde G es la matriz de conversión que incluye los factores de ganancia de mezcla descendente arbitraria del decodificador espacial.

Además, el sintetizador de datos de salida puede ser operativo para calcular la matriz de energía basándose en:

$$E = CZC^*,$$

15 donde E es la matriz de energía, C es la matriz de parámetro de predicción, y Z es una matriz de covarianza de los al menos dos canales de mezcla descendente.

Además, el sintetizador de datos de salida puede ser operativo para calcular la matriz de conversión basándose en:

$$G = A_2 \cdot C,$$

20 donde G es la matriz de conversión,  $A_2$  es la matriz de reproducción parcial, y C es la matriz de parámetros de predicción.

Además, el sintetizador de datos de salida puede ser operativo para calcular la matriz de conversión basándose en:

$$G(DED^*) = A_2 ED^*,$$

25 donde G es una matriz de energía derivada de la fuente de audio de las pistas, D es una matriz de mezcla descendente derivada de la información de mezcla descendente,  $A_2$  es una matriz de reproducción reducida, y "\*" indica la operación conjugada completa.

Además, la matriz de reproducción estéreo parametrizada  $A_2$  puede determinarse tal como sigue:

$$\begin{bmatrix} \mu & 1-\mu & \nu \\ 1-\kappa & \kappa & \nu \end{bmatrix}$$

donde  $\mu$ ,  $\nu$ , y  $\kappa$  son parámetros de valor reales que van a ajustarse según la posición y volumen de uno o más objetos de audio fuente.

**REIVINDICACIONES**

1. Sintetizador (104) de audio para generar datos de salida usando una señal (95, 97) de objeto de audio codificada, que comprende:

un sintetizador (100) de datos de salida para generar los datos de salida que pueden usarse para reproducir una pluralidad de canales de salida de una configuración de salida de audio predefinida que representa la pluralidad de objetos de audio, siendo el sintetizador de datos de salida operativo para usar información de mezcla descendente que indica una distribución de la pluralidad de objetos de audio en al menos dos canales de mezcla descendente, información de potencia, información de correlación que indican una característica de potencia y una característica de correlación de los al menos dos canales (93) de mezcla descendente, y parámetros de objeto de audio para los objetos de audio, en el que el sintetizador (100) de datos de salida es operativo para transcodificar (502) los parámetros de objeto de audio en parámetros espaciales para la configuración de salida de audio predefinida usando además un posicionamiento previsto de los objetos (90) de audio en la configuración de salida de audio.

2. Sintetizador de audio según la reivindicación 1, en el que el sintetizador (100) de datos de salida es operativo para convertir una pluralidad de canales de mezcla descendente en la mezcla descendente estéreo para la configuración de salida de audio predefinida usando una matriz de conversión derivada del posicionamiento previsto de los objetos de audio.

3. Sintetizador de audio según la reivindicación 1, en el que los parámetros espaciales incluyen el primer grupo de parámetros para una mezcla ascendente de dos a tres y un segundo grupo de parámetros de energía para una mezcla ascendente de tres a seis, y

en el que el sintetizador (100) de datos de salida es operativo para calcular los parámetros de predicción para la matriz de predicción de dos a tres usando una matriz de reproducción determinada mediante un posicionamiento previsto de los objetos (90) de audio, describiendo una matriz de mezcla descendente parcial la mezcla descendente de los canales de salida a tres canales generados por un hipotético proceso de mezcla ascendente de dos a tres, y la matriz de mezcla descendente.

4. Sintetizador de audio según la reivindicación 3, en el que los parámetros de objeto son parámetros de objeto de predicción, y en el que el sintetizador (100) de datos de salida es operativo para calcular previamente una matriz de energía basándose en los parámetros de objeto de predicción, la información de mezcla descendente y la información de energía correspondiente a los canales de mezcla descendente.

5. Sintetizador de audio según la reivindicación 1, en el que el sintetizador (100) de datos de salida es operativo para generar (165) dos canales estéreo para una configuración de salida estéreo calculando una matriz de reproducción estéreo parametrizada y una matriz de conversión dependiendo de la matriz de reproducción estéreo parametrizada.

6. Método de sintetización de audio para generar datos de salida usando una señal (95, 97) de objeto de audio codificada, que comprende:

generar los datos de salida que pueden usarse para crear una pluralidad de canales de salida de una configuración de salida de audio predefinida que representa la pluralidad de objetos (90) de audio, en el que se usan información de mezcla descendente que indica una distribución de la pluralidad de objetos de audio en al menos dos canales de mezcla descendente, información de potencia, información de correlación que indican una característica de potencia y una característica de correlación de los al menos dos canales (93) de mezcla descendente, y parámetros de objeto de audio para los objetos de audio, y en el que los parámetros de objeto de audio se transcodifican (502) en parámetros espaciales para la configuración de salida de audio predefinida usando además un posicionamiento previsto de los objetos (90) de audio en la configuración de salida de audio.

7. Codificador (101) de objetos de audio para generar una señal de objeto de audio codificada usando una pluralidad de objetos (90) de audio, que comprende:

un generador (96) de información de mezcla descendente para generar información (97) de mezcla descendente que indica una distribución de la pluralidad de objetos de audio en al menos dos canales de mezcla descendente, en el que el generador (96) de información de mezcla descendente está configurado para generar (150) una información de potencia y una información de correlación que indican una característica de potencia y una característica de correlación de los al menos dos canales (93) de mezcla descendente;

un generador (94) de parámetro de objeto para generar parámetros (95) de objeto para los objetos de audio; y

una interfaz (98) de salida para generar la señal (99) de objeto de audio codificada, comprendiendo la señal de objeto codificada la información de mezcla descendente, la información de potencia, la información de correlación y los parámetros de objeto.

8. Codificador de objetos de audio según la reivindicación 7, que comprende además:

5 un mezclador (92) descendente para la mezcla descendente de la pluralidad de objetos de audio en la pluralidad de canales de mezcla descendente, en el que el número de objetos de audio es mayor que el número de canales de mezcla descendente, y en el que el mezclador (92) descendente se acopla al  
 10 generador de información de mezcla descendente de modo que la distribución de la pluralidad de objetos de audio en la pluralidad de canales de mezcla descendente se lleva a cabo según se indique en la información de mezcla descendente.

9. Codificador de objetos de audio según la reivindicación 7, en el que el generador (96) de información de mezcla descendente es operativo para calcular la información de mezcla descendente de modo que la información de mezcla descendente indica,

15 qué objeto de audio se incluye total o parcialmente en uno o más de la pluralidad de canales de mezcla descendente, y

cuando se incluye un objeto de audio en más de un canal de mezcla descendente, una información sobre una parte de los objetos de audio incluidos en un canal de mezcla descendente de los más de un canales de mezcla descendente.

20 10. Método (101) de codificación de objetos de audio para generar una señal de objeto de audio codificada usando una pluralidad de objetos de audio, que comprende:

generar información (97) de mezcla descendente que indica una distribución de la pluralidad de objetos (90) de audio en al menos dos canales de mezcla descendente,

generar (150) una información de potencia y una información de correlación que indican una característica de potencia y una característica de correlación de los al menos dos canales de mezcla descendente;

25 generar parámetros (94) de objeto para los objetos de audio; y

generar la señal (99) de objeto de audio codificada, comprendiendo la señal de objeto de audio codificada la información de potencia, la información de correlación, la información de mezcla descendente y los parámetros de objeto.

30 11. Señal de objeto de audio codificada que incluye una información de mezcla descendente que indica una distribución de una pluralidad de objetos de audio en al menos dos canales de mezcla descendente, una información de potencia y una información de correlación que indican una característica de potencia y una característica de correlación de los al menos dos canales de mezcla descendente, y parámetros de objeto, siendo los parámetros de objeto de manera que es posible la reconstrucción de los objetos de audio usando los parámetros de objeto y los al menos dos canales de mezcla descendente.

35 12. Señal de objeto de audio codificada según la reivindicación 11, almacenada en un medio de almacenamiento legible por ordenador.

13. Programa informático para realizar, cuando se ejecuta en un ordenador, un método según uno cualquiera de los métodos de las reivindicaciones 6 ó 10.

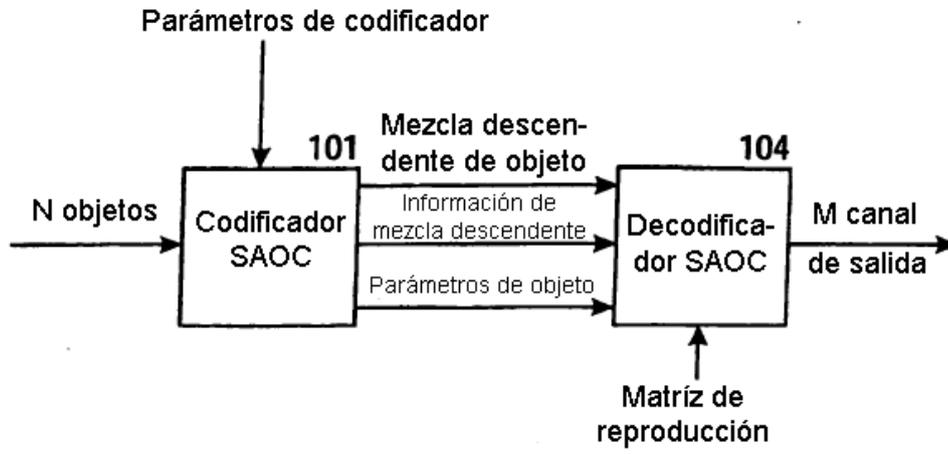


FIG 1A

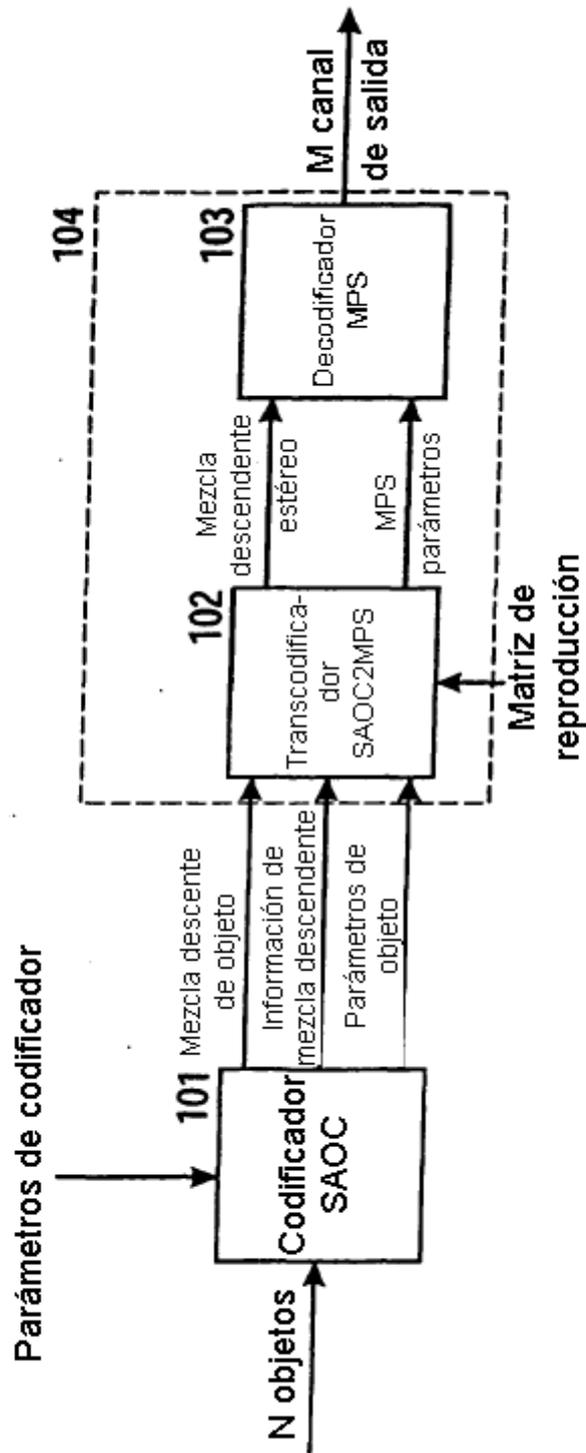


FIG 1B

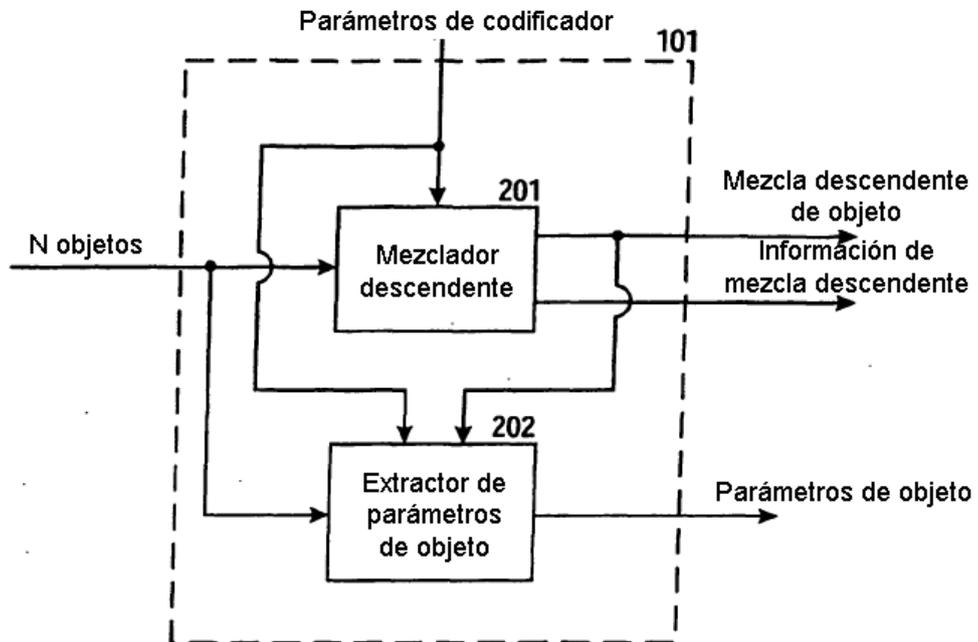


FIG 2

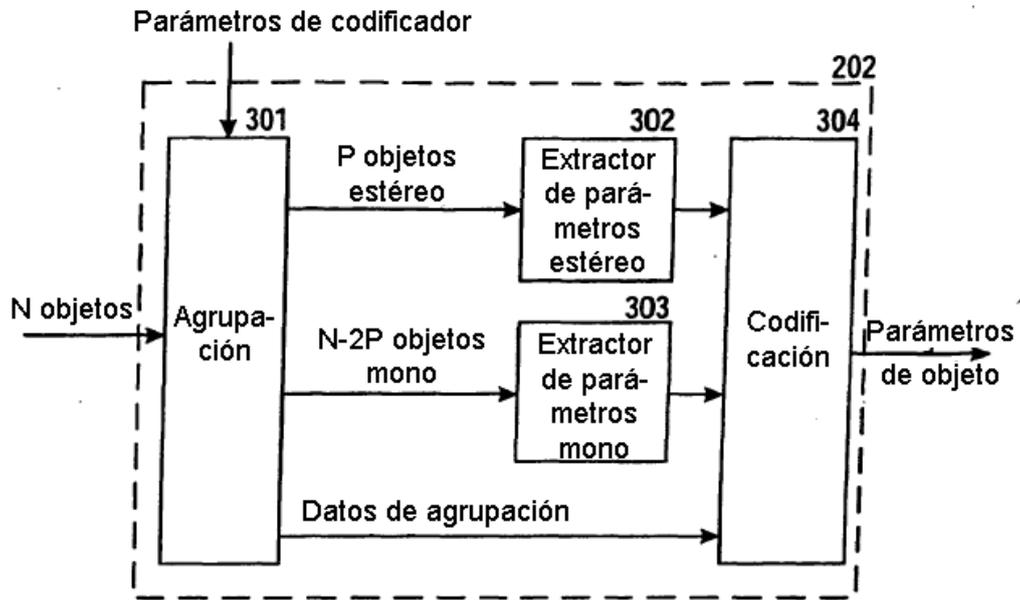


FIG 3

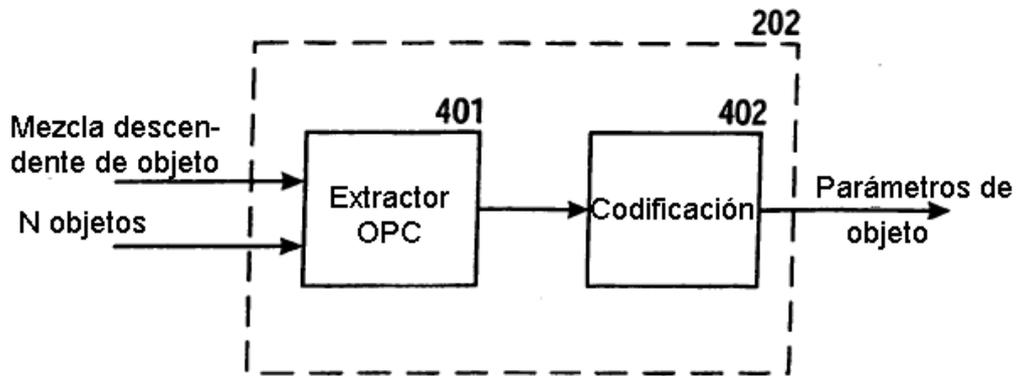


FIG 4

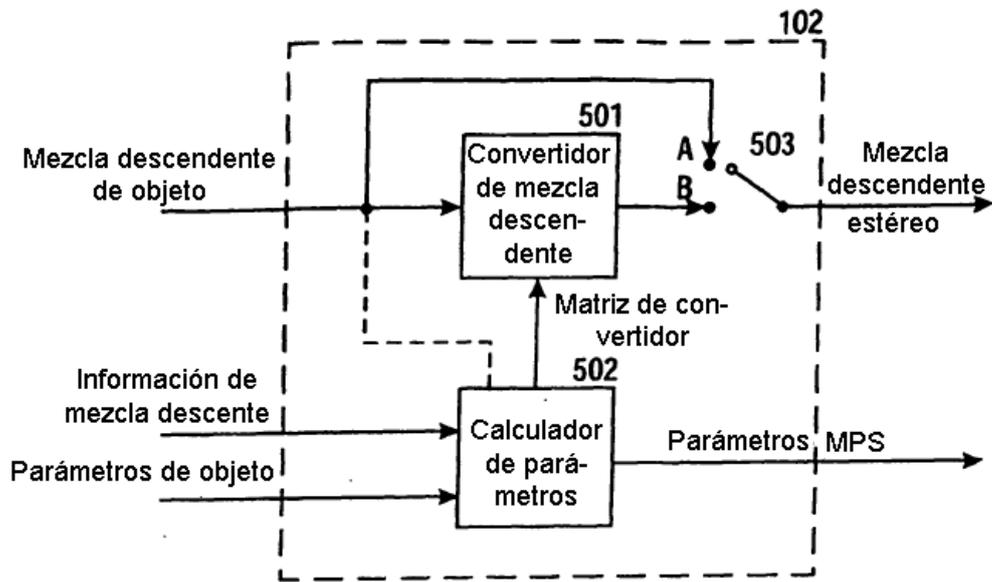


FIG 5

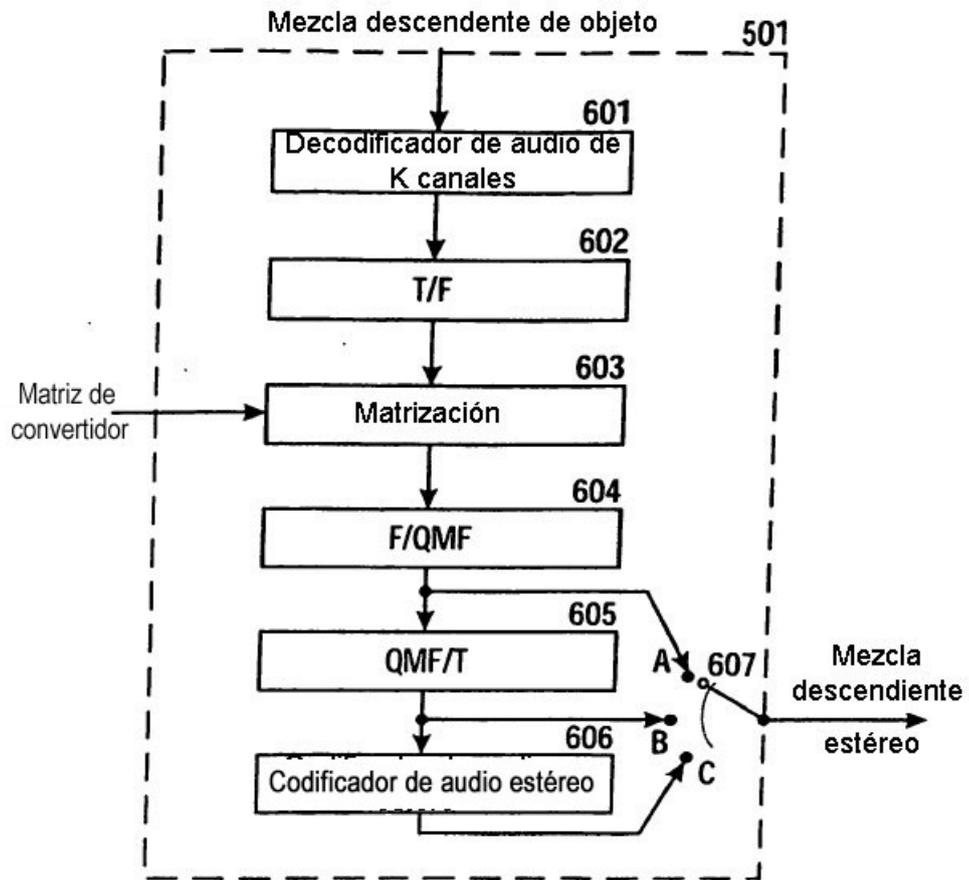


FIG 6

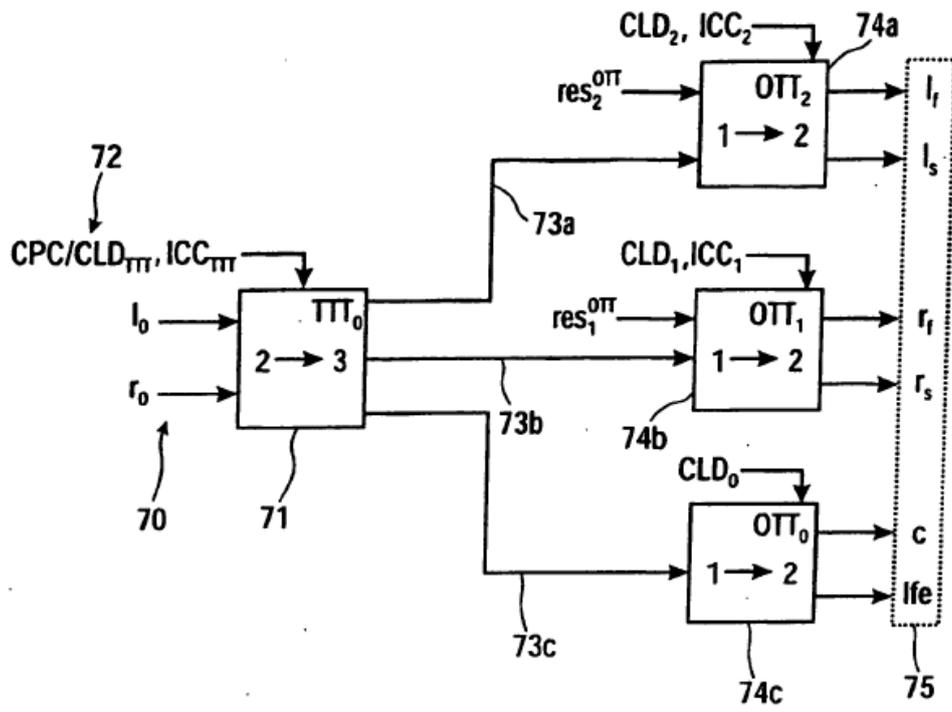


FIG 7

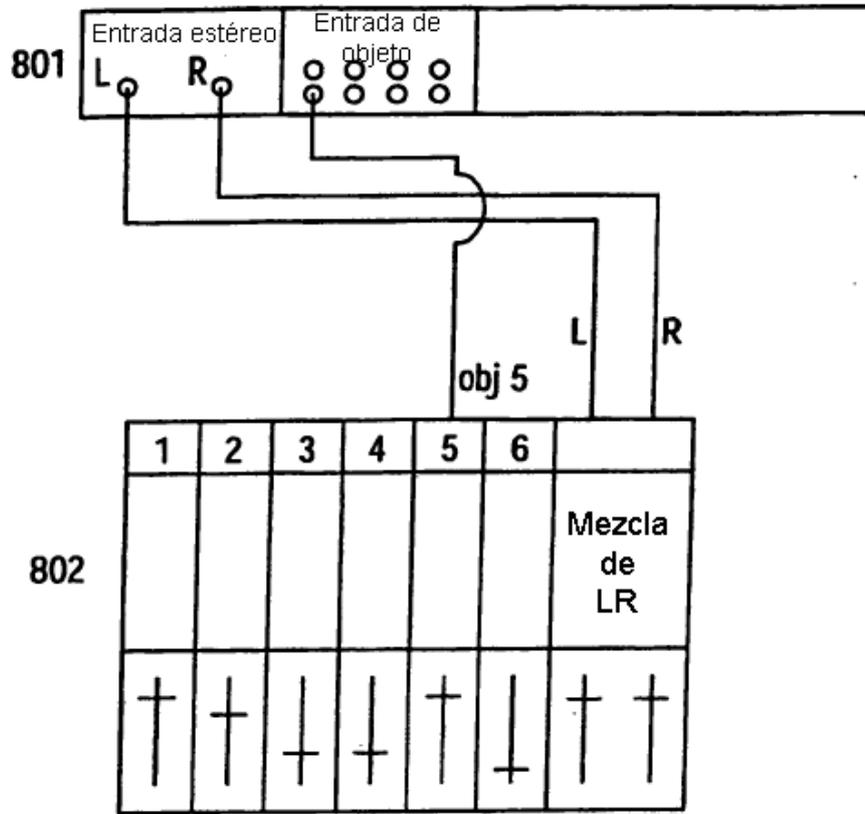


FIG 8

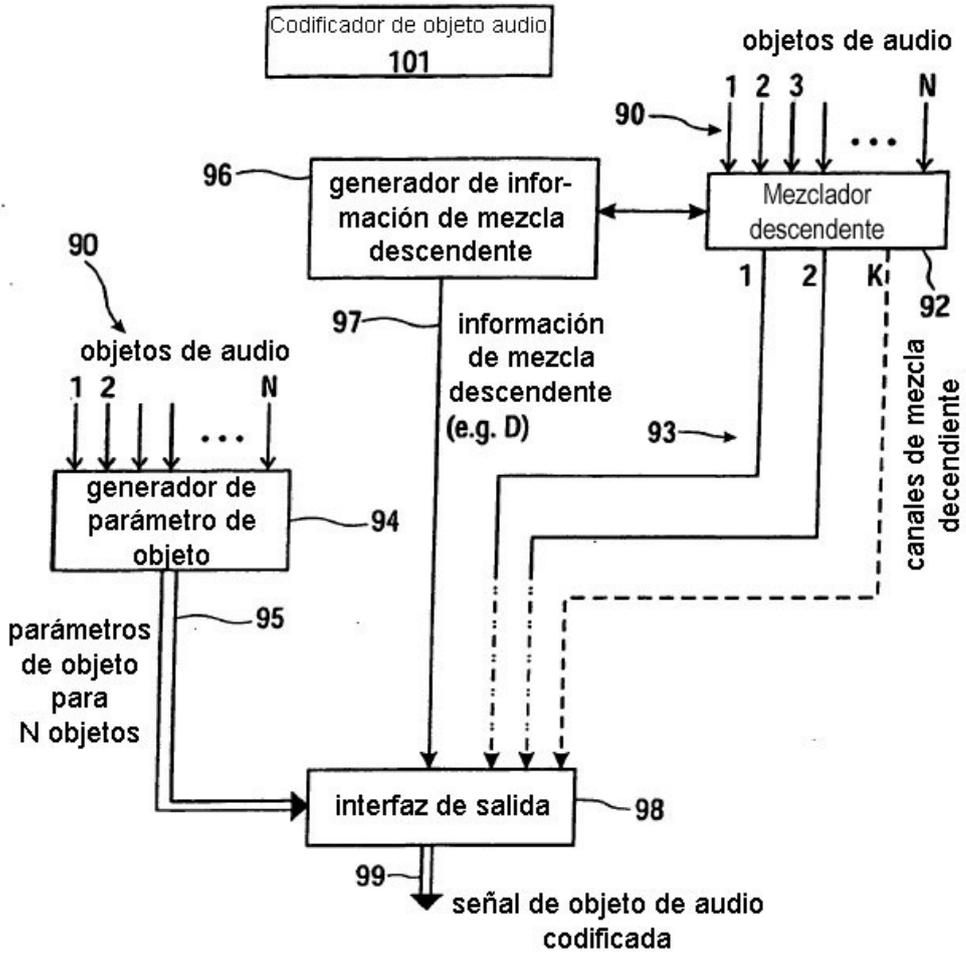


FIG 9

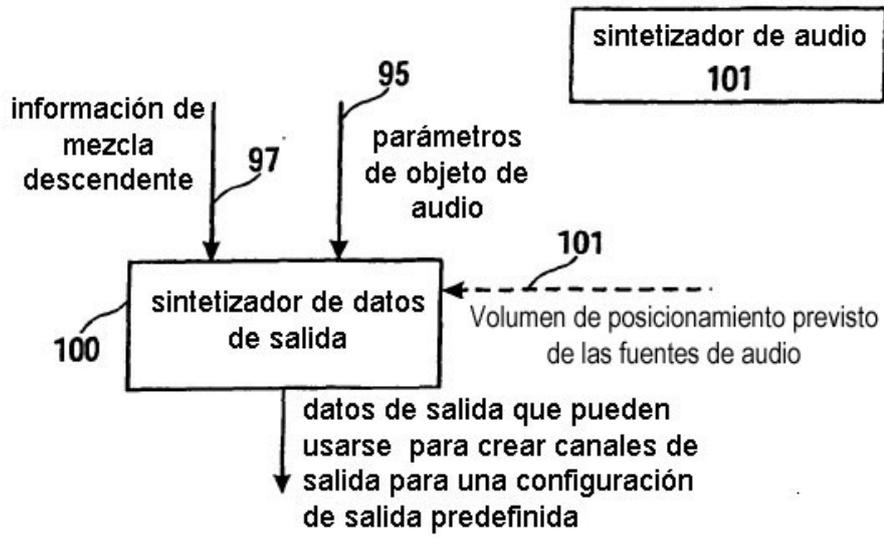


FIG 10

NÚMERO DE MODO	INFORMACIÓN DE MEZCLA DESCENDENTE	CANALES DE MEZCLA DESCENDENTE	PARÁMETROS DE OBJETO DE AUDIO	CONFIGURACIÓN DE SALIDA	POSICIONAMIENTO PREVISTO	DATOS DE SALIDA QUE PUEDEN USARSE PARA CREAR
1	X	X	X			FUENTES RECONSTRUIDAS <span style="float: right;">111</span>
2	X		X	X	X	PARÁMETROS DE MEZCLADOR ESPACIALES <span style="float: right;">112</span>
3	X		X	X	X	PARÁMETROS DE MEZCLADOR ESPACIALES Y MATRIZ DE CONVERSIÓN <span style="float: right;">113</span>
4	X	X	X	X	X	PARÁMETROS DE MEZCLADOR ESPACIALES Y MEZCLA DESCENDENTE CONVERTIDA <span style="float: right;">114</span>
5	X	X	X	X	X	SALIDA ESTÉREO <span style="float: right;">115</span>
6	X	X	X	X	X	SALIDA MULTICANAL (M>2) e.g. 5.1, ... <span style="float: right;">116</span>

FIG 11

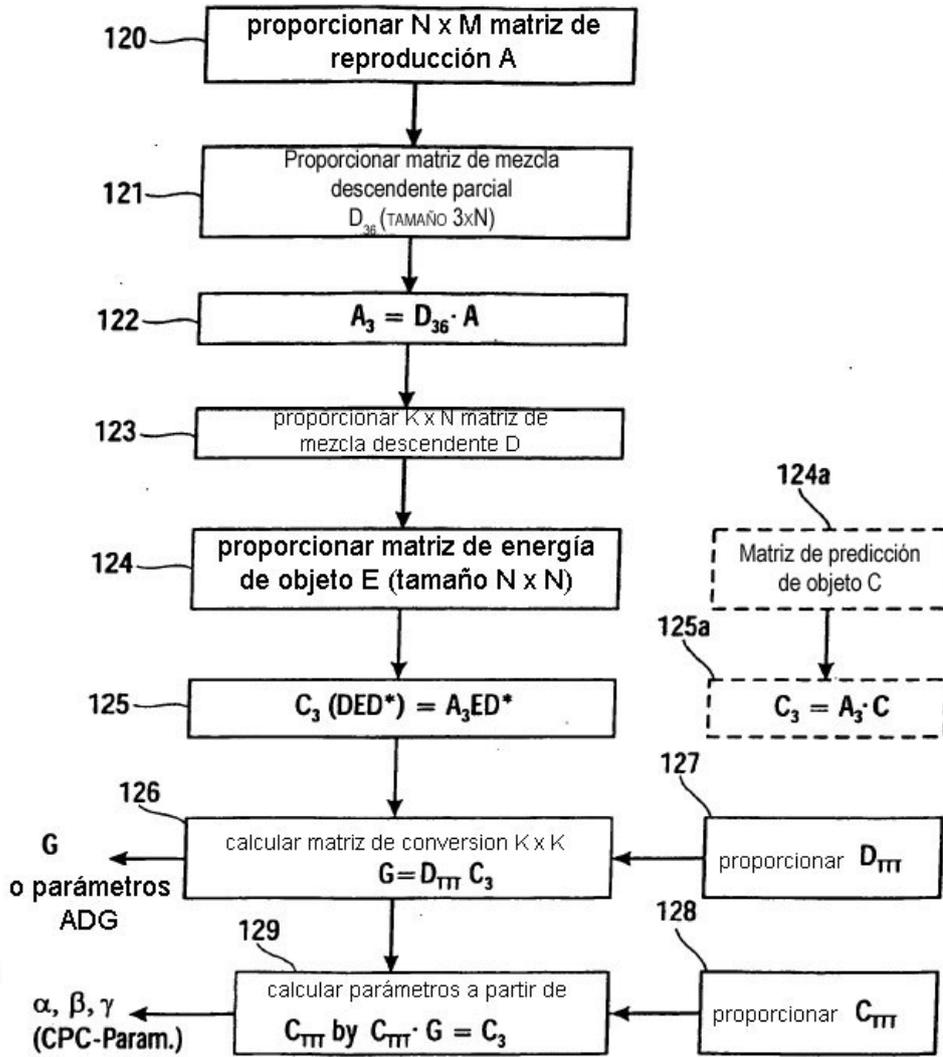


FIG 12

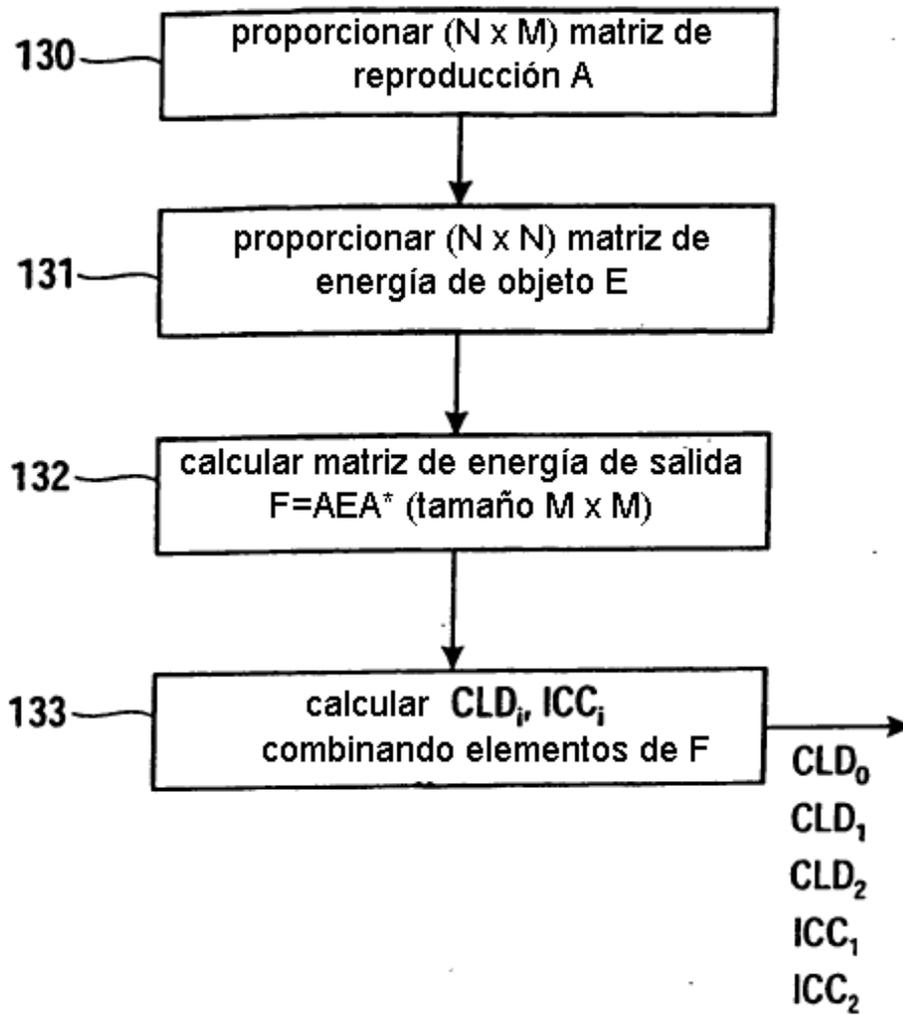


FIG 13A

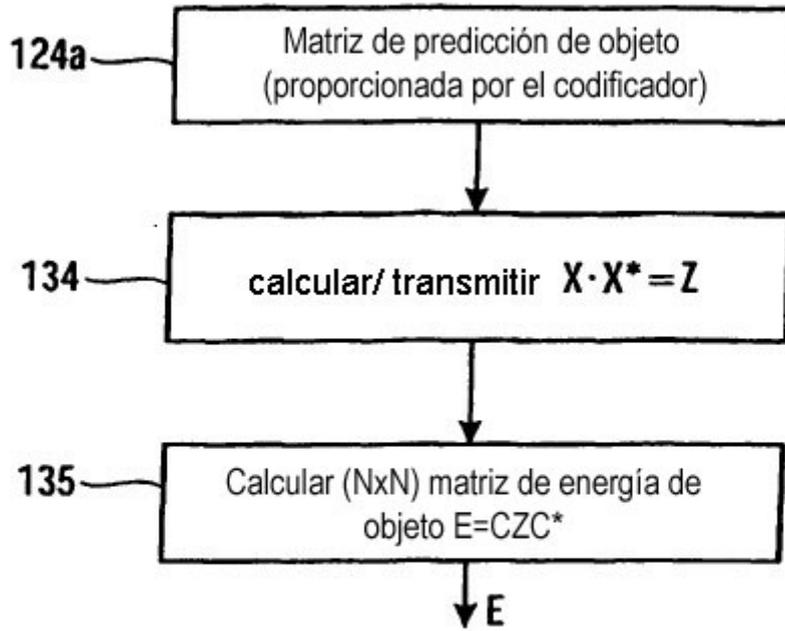


FIG 13B

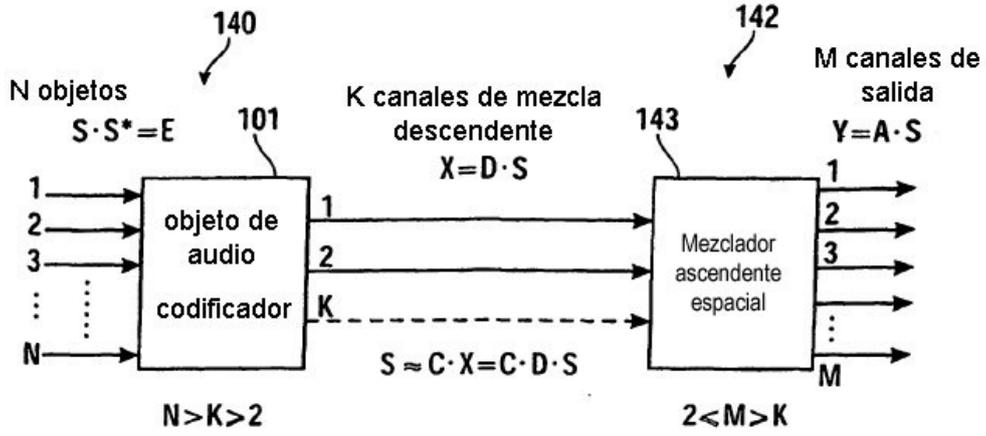


FIG 14

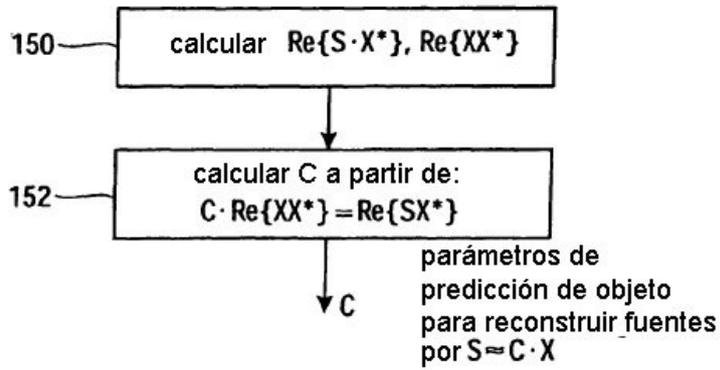


FIG 15

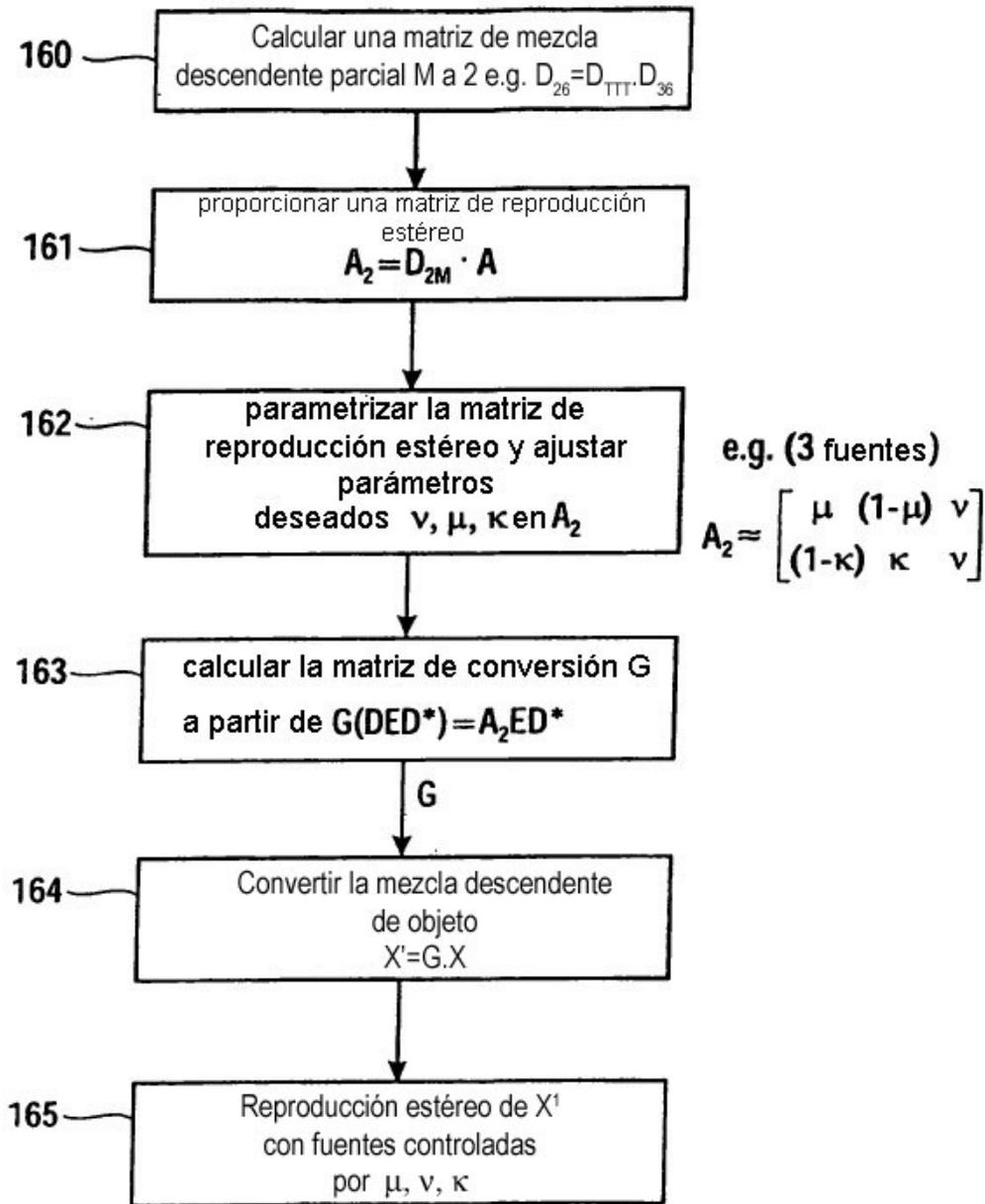


FIG 16