



19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 360 740**

51 Int. Cl.:  
**G10L 19/00** (2006.01)  
**H04S 3/00** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **07826410 .8**  
96 Fecha de presentación : **17.09.2007**  
97 Número de publicación de la solicitud: **2067138**  
97 Fecha de publicación de la solicitud: **10.06.2009**

54 Título: **Codificación y decodificación de objetos de audio.**

30 Prioridad: **18.09.2006 EP 06120819**  
**10.11.2006 EP 06123799**

45 Fecha de publicación de la mención BOPI:  
**08.06.2011**

45 Fecha de la publicación del folleto de la patente:  
**08.06.2011**

73 Titular/es:  
**KONINKLIJKE PHILIPS ELECTRONICS N.V.**  
**Groenewoudseweg 1**  
**5621 BA Eindhoven, NL**

72 Inventor/es: **Breebaart, Dirk J.**

74 Agente: **Zuazo Araluze, Alexander**

ES 2 360 740 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

## CAMPO DE LA INVENCION

La invención se refiere a la codificación y decodificación de objetos de audio y en particular, pero no exclusivamente a la manipulación de objetos de audio de una señal espacial de mezcla descendente.

## 5 ANTECEDENTES DE LA INVENCION

La codificación digital de diversas señales de audio ha pasado a ser cada vez más importante durante las últimas décadas porque la representación y comunicación de señal digital ha sustituido cada vez más a la representación y comunicación analógica.

10 En la última década ha habido una tendencia al audio multicanal y específicamente al audio espacial que se extiende más allá de las señales estéreo convencionales. Por ejemplo, las grabaciones estéreo tradicionales sólo comprenden dos canales mientras que los sistemas de audio avanzados modernos normalmente usan cinco o seis canales, como en los sistemas de sonido envolvente 5.1 populares. Esto proporciona una experiencia sonora más complicada en la que el usuario puede estar rodeado por fuentes de sonido.

15 Se han desarrollado diversas técnicas y normas para la comunicación de tales señales multicanal. Por ejemplo, pueden transmitirse seis canales discretos que representan un sistema envolvente 5.1 según normas tales como la norma de codificación de audio avanzada (AAC) o las normas de Dolby Digital.

20 Sin embargo, para proporcionar retrocompatibilidad, se conoce mezclar de manera descendente el mayor número de canales para dar un menor número de canales y específicamente se usa con frecuencia para mezclar de manera descendente una señal de sonido envolvente 5.1 para dar una señal estéreo permitiendo reproducir una señal estéreo por decodificadores legados (estéreo) y una señal 5.1 por decodificadores de sonido envolvente.

25 Un ejemplo es el procedimiento de codificación retrocompatible envolvente MPEG normalizado por el grupo de expertos en imágenes en movimiento (MPEG). En tal sistema, una señal multicanal se mezcla de manera descendente para dar una señal estéreo y las señales adicionales se codifican mediante datos paramétricos en la parte de datos auxiliares permitiendo que un decodificador multicanal envolvente MPEG genere una representación de la señal multicanal. Un decodificador mono o estéreo legado ignorará los datos auxiliares y por tanto sólo decodificará la mezcla descendente mono o estéreo.

30 Por tanto, en (de)codificadores de audio espacial (paramétricos), los parámetros se extraen de la señal de audio original para producir una señal de audio que tiene un número reducido de canales, por ejemplo sólo un canal único, más un conjunto de parámetros que describen las propiedades espaciales de la señal de audio original. En los decodificadores de audio espacial (paramétricos), las propiedades espaciales descritas por los parámetros espaciales transmitidos se usan para recrear la señal multicanal espacial original. Un codificador y decodificador de audio espacial paramétricos a modo de ejemplo se da a conocer en el documento US 2003/0035553 A1.

35 Recientemente, las técnicas para la distribución de objetos de audio individuales que pueden procesarse y manipularse en el extremo receptor han atraído un interés significativo. Por ejemplo, dentro del marco MPEG, un elemento de trabajo se inicia en codificación de audio espacial basada en objetos. El objetivo de este elemento de trabajo es explorar nueva tecnología y reutilizar los componentes y tecnologías envolventes MPEG actuales para la codificación eficaz de tasa de transmisión de bits de múltiples fuentes u objetos de sonido a un número de canales de mezcla descendente y parámetros espaciales correspondientes. Por tanto, la intención es usar técnicas similares a las que se usan para la mezcla descendente de canales (envolventes) espaciales a menos canales para mezclar de manera descendente objetos de audio independientes a un menor número de canales.

40 En sistemas de audio orientados a objetos, el decodificador puede proporcionar un posicionamiento discreto de estas fuentes/objetos y una adaptación a diversas configuraciones de altavoz así como presentación binaural. Adicionalmente, puede usarse la interacción del usuario para controlar el reposicionamiento/desplazamiento panorámico de las fuentes individuales en el lado de reproducción.

45 En otras palabras, el objetivo de la investigación es codificar múltiples objetos de audio en un conjunto limitado de canales de mezcla descendente acompañados por parámetros. En el lado del decodificador, los usuarios pueden interactuar con el contenido por ejemplo reposicionando los objetos individuales. Como ejemplo específico, un número de instrumentos individuales puede codificarse y distribirse como objetos de audio permitiendo de ese modo a un usuario que recibe los datos codificados posicionar independientemente los instrumentos individuales en la imagen de sonido.

50 La figura 1 ilustra un ejemplo de un codificador y decodificador de audio orientado a objetos según la técnica anterior. En el ejemplo, un conjunto de objetos de audio ( $O_1$  a  $O_4$ ) se codifican en un codificador 101 orientado a objetos que genera una señal de mezcla descendente y parámetros de objeto. Éstos se transmiten al decodificador 103 orientado a objetos que genera copias aproximadas de las señales de objeto de audio usando los parámetros de objeto transmitidos.

55 Posteriormente, un elemento 105 de presentación genera la señal de salida que tiene las características deseadas. Por ejemplo, el elemento 105 de presentación puede posicionar los objetos en posiciones de fuente de

sonido indicadas por el usuario, por ejemplo usando una ley de desplazamiento panorámico. La configuración de la señal de salida es flexible. Por ejemplo, si la señal de salida es mono, el usuario aún puede manipular la intensidad/volumen relativo de cada objeto. En una configuración de la señal de salida estéreo, puede aplicarse una ley simple de desplazamiento panorámico para posicionar cada objeto en una posición deseada. Obviamente, para una configuración de salida multicanal, la flexibilidad es incluso mayor.

Sin embargo, aunque el sistema puede proporcionar un rendimiento ventajoso, también tiene un número de desventajas. Por ejemplo, en muchos casos la calidad reproducida es subóptima y no es posible una manipulación completamente libre e independiente de los objetos de audio individuales. Específicamente, en general la mezcla descendente del codificador no es completamente reversible en el decodificador que por consiguiente sólo puede generar aproximaciones de los objetos de audio originales. Por tanto, el decodificador no puede reconstruir completamente las señales de objeto individuales sino que sólo puede estimarlas según criterios de percepción. Esto da como resultado específicamente una interferencia cruzada (diafonía) entre objetos de audio dando así como resultado que los objetos de audio ya no sean completamente independientes. Como resultado las manipulaciones en un objeto de audio afectan a las características y la percepción de otro objeto.

Por ejemplo, uno de los parámetros más importantes que los usuarios normalmente desearían ajustar es el volumen relativo de cada objeto de audio. Sin embargo, si se realizan grandes ajustes de volumen esto dará como resultado artefactos considerables y una diafonía no deseada dando como resultado una degradación de calidad perceptible.

Por tanto, sería ventajoso un sistema mejorado para la codificación/decodificación de objetos de audio y en particular sería ventajoso un sistema que permitiera una flexibilidad aumentada, calidad mejorada, implementación facilitada y/o rendimiento mejorado.

### SUMARIO DE LA INVENCION

Por consiguiente, la invención busca preferiblemente mitigar, aliviar o eliminar una o más de las desventajas anteriormente mencionadas de manera individual o en cualquier combinación.

Según un primer aspecto de la invención se proporciona un codificador para codificar objetos de audio, comprendiendo el codificador: medios para recibir una pluralidad de objetos de audio; medios de codificación para codificar la pluralidad de objetos de audio en un número de señales de audio y datos paramétricos que representan la pluralidad de objetos de audio en relación con el número de señales de audio, comprendiendo los datos paramétricos un conjunto de parámetros de objeto para al menos uno de los diferentes objetos de audio; medios para recibir datos de modificación de codificación desde una unidad remota; y medios de parámetro para determinar los datos paramétricos en respuesta a los datos de modificación de codificación.

La invención puede permitir la codificación mejorada de objetos de audio y en particular puede permitir un sistema de distribución de audio en el que puede lograrse una experiencia de usuario mejorada por ejemplo con control de usuario individual mejorado de objetos de audio individuales. La invención puede permitir un control mejorado de características de objetos de audio individuales y en particular puede reducir la degradación por interferencia de objetos de audio cruzada cuando se manipulan objetos de audio. El codificador puede permitir una manipulación controlada remota eficaz mientras se modifican los datos de modificación de codificación de manera que un decodificador orientado a objetos decodificará los objetos de audio manipulados correctamente. La invención puede permitir una distribución mejorada de manipulación de objetos de audio entre un codificador y un decodificador dando así como resultado una flexibilidad, rendimiento y/o calidad mejorados.

Los medios de codificación pueden además generar el número de señales de audio en respuesta a los datos de modificación de codificación. Los parámetros de objeto pueden ser parámetros de intensidad por ejemplo que indican una diferencia de intensidad relativa entre diferentes objetos de audio y/o un factor de conversión de energía entre una o más de las señales de audio y los objetos de audio. Los parámetros de objeto pueden comprender parámetros para bloques de frecuencia-tiempo individuales.

Según una característica opcional de la invención, los medios de codificación están dispuestos para generar el número de señales de audio mediante una mezcla descendente de los objetos de audio y los medios de parámetro están dispuestos para modificar un peso de mezcla descendente de al menos uno de los objetos de audio en respuesta a los datos de modificación de codificación.

Esto puede proporcionar un control de alta calidad y/o altamente eficaz del volumen relativo de un objeto de audio por parte de un oyente mientras se reduce o elimina el efecto en otros objetos de audio. Puede lograrse un control de volumen de objetos de audio individuales de alto rendimiento.

Según una característica opcional de la invención, los medios de parámetro están dispuestos para ajustar a escala al menos un primer objeto de audio en respuesta a los datos de modificación de codificación y para modificar parámetros de objeto para el primer objeto de audio en respuesta al ajuste a escala.

Esto puede proporcionar un control de alta calidad y/o altamente eficaz del volumen relativo de un objeto de audio por parte de un oyente mientras se reduce o elimina el efecto en otros objetos de audio. Puede lograrse un control de volumen de objetos de audio individuales de alto rendimiento.

Según una característica opcional de la invención, al menos algunos de los datos de modificación de codificación son específicos de la frecuencia y los medios de parámetro están dispuestos para determinar al menos un parámetro de objeto en respuesta a una característica de frecuencia del parámetro de objeto.

5 Esto puede permitir un control mejorado de la experiencia sonora y en particular puede permitir manipular la respuesta de frecuencia del audio por parte de un oyente. Las características de frecuencia de los objetos individuales pueden modificarse individualmente o por separado con efecto reducido o eliminado en otros objetos de audio. En particular, puede lograrse una ecualización eficaz y/o de alta calidad de objetos de audio individuales.

10 Según una característica opcional de la invención, los medios de codificación están dispuestos para modificar al menos un objeto de audio en respuesta a los datos de modificación de codificación antes de una mezcla descendente de los objetos de audio al número de señales de audio.

Los medios de parámetro pueden disponerse para determinar los datos paramétricos en respuesta a características del/de los objeto(s) de audio modificado(s). Esto puede permitir una implementación de alto rendimiento y/o facilitada.

15 Según una característica opcional de la invención, los medios de codificación están dispuestos para generar el número de señales de audio como una mezcla descendente espacial.

Esto puede permitir un rendimiento mejorado en muchas realizaciones y en particular puede permitir una operación mejorada en relación con codificadores que no tienen capacidad de presentación o la tienen limitada. El codificador puede disponerse por ejemplo para presentar una señal multicanal espacial que comprende los objetos de audio y puede disponerse específicamente para generar una señal binaural espacial.

20 Según una característica opcional de la invención, los medios de codificación están dispuestos para modificar en respuesta a los datos de modificación de codificación al menos una característica seleccionada del grupo que consiste en: una ubicación espacial de al menos uno de los objetos de audio; una característica de distancia de al menos uno de los objetos de audio; un modo de presentación espacial del codificador, y una característica de frecuencia de al menos uno de los objetos de audio.

25 Esto puede permitir un rendimiento mejorado y los parámetros pueden permitir en particular a un oyente modificar parámetros significativos desde el punto de vista de la percepción de una señal espacial presentada.

Según una característica opcional de la invención, cada objeto de audio se asocia con un conjunto de fuentes de audio que son independientes de fuentes de audio de otros objetos de audio.

30 Los objetos de audio pueden ser independientes uno de otro. Los objetos de audio pueden corresponder a fuentes de sonido diferentes e independientes. Específicamente, los objetos de audio pueden ser objetos de audio diferentes que se generan individualmente y por separado de los otros objetos de audio y sin ninguna relación específica. Por ejemplo, los objetos de audio pueden ser voces o instrumentos musicales de captura/grabados individualmente.

35 Los objetos de audio pueden ser objetos de audio no espaciales. Los objetos de audio pueden ser simples fuentes de sonido sin ninguna característica o información espacial asociada y en particular puede no haber ninguna asociación, conocimiento o relación espacial relativa entre los objetos de audio.

40 Según una característica opcional de la invención, el codificador está dispuesto para recibir un primer objeto de audio desde la unidad remota y los medios para recibir los datos de modificación de codificación están dispuestos para extraer los datos de modificación de codificación de los datos de codificación recibidos para el primer objeto de audio.

45 Por ejemplo, los datos de modificación de codificación pueden estar incrustados en una señal de voz, música u otra señal de audio. Los datos de modificación de codificación pueden estar incrustados específicamente en campos de datos de usuario o auxiliares de una señal de audio codificada recibida desde la unidad remota, tal como por ejemplo un flujo de bits MPEG 4. Esto puede permitir una comunicación eficaz, de baja complejidad y retrocompatible de datos de control y en particular puede ser útil en sistemas que emplean comunicaciones bidireccionales entre un aparato que comprende el codificador y la unidad remota.

50 Según una característica opcional de la invención, el codificador está dispuesto para recibir datos de modificación de codificación desde una pluralidad de unidades remotas y para generar diferentes datos paramétricos para las diferentes unidades remotas en respuesta a recibir diferentes datos de modificación de codificación desde las diferentes unidades remotas.

Esto puede permitir una operación mejorada y/o servicios adicionales en muchas realizaciones. Los medios de codificación pueden además disponerse para generar diferentes señales de audio para las diferentes unidades remotas. Por tanto, el enfoque puede permitir por ejemplo a un codificador de objetos de audio centralizado adaptar los datos transmitidos a los requisitos y preferencias de los usuarios individuales de las unidades remotas.

55 Según otro aspecto de la invención, se proporciona un decodificador para decodificar objetos de audio, comprendiendo el decodificador: un receptor para recibir desde un codificador de objetos de audio remoto un

número de señales de audio que son una mezcla descendente de una pluralidad de objetos de audio y datos paramétricos que representan la pluralidad de objetos de audio en relación con el número de señales de audio, comprendiendo los datos paramétricos un conjunto de parámetros de objeto para al menos uno de los diferentes objetos de audio; medios de decodificación para decodificar los objetos de audio a partir del número de señales de audio en respuesta a los datos paramétricos; medios de presentación para generar una señal de salida multicanal espacial a partir de los objetos de audio; medios para generar datos de modificación de codificación para el codificador de objetos de audio remoto; y medios para transmitir los datos de modificación de codificación al codificador de objetos de audio remoto.

Los medios de decodificación y los medios de presentación pueden combinarse en algunas realizaciones y la señal de salida multicanal espacial puede generarse directamente a partir de las señales de audio sin generar explícitamente el objeto de audio. Por ejemplo, una multiplicación de matriz puede aplicarse a valores de señal de las señales de audio para generar valores de señal de objeto de audio. Una segunda multiplicación de matriz puede aplicarse entonces a los valores de señal de objeto de audio para generar los valores de señal de audio multicanal espacial. Alternativamente, la primera y segunda multiplicación de matriz pueden combinarse en una sola multiplicación de matriz. Por tanto, puede aplicarse una sola multiplicación de matriz a los valores de señal de las señales de audio para generar directamente los valores de señal de audio multicanal espacial. Por tanto, la decodificación de los objetos de audio puede ser implícita en la presentación/multiplicación de matriz y no es necesaria una generación explícita/directa de valores de objeto audio.

Según otros aspectos de la invención, se proporcionan un centro de teleconferencia tal como se establece en la reivindicación 12, un transmisor tal como se establece en la reivindicación 13, un receptor tal como se establece en la reivindicación 14, un sistema de comunicación tal como se establece en la reivindicación 15, un procedimiento de codificación de objetos de audio tal como se establece en la reivindicación 16, un procedimiento de decodificación de objetos de audio tal como se establece en la reivindicación 17, un producto de programa informático tal como se establece en la reivindicación 18, y un dispositivo de reproducción de audio tal como se establece en la reivindicación 19.

Estos y otros aspectos, características y ventajas de la invención serán evidentes a partir de y se aclararán con referencia a la(s) realización(es) descrita(s) a continuación en el presente documento.

#### **BREVE DESCRIPCIÓN DE LOS DIBUJOS**

Las realizaciones de la invención se describirán, sólo a modo de ejemplo, con referencia a los dibujos, en los que

la figura 1 es una ilustración de un sistema de audio según la técnica anterior;

la figura 2 ilustra un ejemplo de un sistema de comunicación para la comunicación de una señal de audio según algunas realizaciones de la invención;

la figura 3 ilustra una interacción entre un codificador y un decodificador según algunas realizaciones de la invención;

la figura 4 ilustra un ejemplo de un codificador según algunas realizaciones de la invención;

la figura 5 ilustra un ejemplo de un decodificador según algunas realizaciones de la invención;

la figura 6 ilustra un ejemplo de un procedimiento de codificación de señales de audio según algunas realizaciones de la invención; y

la figura 7 ilustra un ejemplo de un procedimiento de decodificación de objetos de audio según algunas realizaciones de la invención.

#### **DESCRIPCIÓN DETALLADA DE ALGUNAS REALIZACIONES DE LA INVENCION**

La siguiente descripción se centra en realizaciones de la invención aplicables a la codificación y/o decodificación de objetos de audio para una aplicación de teleconferencia. Sin embargo, se apreciará que la invención no está limitada a esta aplicación pero puede aplicarse en muchas otras aplicaciones que incluyen por ejemplo aplicaciones de distribución de audio de música.

La figura 2 ilustra un sistema 200 de comunicación para la comunicación de una señal de audio según algunas realizaciones de la invención. El sistema 200 de transmisión comprende un transmisor 201 que está acoplado a un receptor 203 a través de una red 205 que específicamente puede ser Internet.

En el ejemplo específico, el transmisor 201 forma parte de un centro de teleconferencia. En una aplicación de teleconferencia, las señales de voz de varios hablantes de extremo lejano se mezclan en un centro de teleconferencia. Entonces para cada persona en la teleconferencia, se transmite una mezcla de todas las señales excepto la suya a todos los receptores. Por tanto, el transmisor 201 puede recibir señales de voz desde una pluralidad de unidades de comunicación remotas que participan en la teleconferencia y puede generar y distribuir señales de voz a las unidades de comunicación remotas. En el ejemplo, el receptor 203 es un dispositivo reproductor de señales que puede generar una salida de voz para un participante de la llamada de conferencia.

Específicamente, el receptor 203 forma parte de una unidad de comunicación remota tal como un teléfono.

Se apreciará que en otras realizaciones un transmisor y un receptor pueden usarse en otras aplicaciones y para otros propósitos. Por ejemplo, el transmisor 201 y/o el receptor 203 pueden formar parte de una funcionalidad de transcodificación y por ejemplo pueden proporcionar una interfaz a otros destinos o fuentes de señales.

5 En el ejemplo específico, el transmisor 201 comprende un receptor 207 que recibe señales de voz desde las unidades de comunicación remotas implicadas en la llamada de teleconferencia. Cada una de las señales de voz se trata como un objeto de audio separado e independiente.

10 El receptor 207 se acopla al codificador 209 de la figura 2 que se alimenta a los objetos de audio de voz individuales y que codifica los objetos de audio según un algoritmo de codificación. El codificador 209 se acopla a un transmisor 211 de red que recibe la señal codificada y establece una interfaz con Internet 205. El transmisor de red puede transmitir la señal codificada al receptor 203 a través de Internet 205.

El receptor 203 comprende un receptor 213 de red que establece una interfaz con Internet 205 y que se dispone para recibir la señal codificada desde el transmisor 201.

15 El receptor 213 de red se acopla a un decodificador 215. El decodificador 215 recibe la señal codificada y la decodifica según un algoritmo de decodificación. Específicamente, el decodificador 215 es un decodificador orientado a objetos que puede decodificar los objetos de audio individuales y presentar una señal de salida de audio basándose en los objetos de audio decodificados.

20 En el ejemplo específico en el que se soporta una función de reproducción de señales, el receptor 203 comprende además un reproductor 217 de señales que recibe la señal de audio decodificada desde el decodificador 215 y la presenta al usuario. Específicamente, el reproductor 217 de señales puede comprender un convertidor de digital a analógico, amplificadores y altavoces tal como se requiere para emitir la señal de audio decodificada.

La figura 3 ilustra la interacción entre el codificador 209 y el decodificador 215 en más detalle.

25 Tal como se ilustra, el codificador 209 orientado a objetos recibe una pluralidad de objetos de audio desde el receptor 207. Los objetos de audio son señales de sonido individuales que son independientes entre sí y que corresponden específicamente a fuentes de sonido individuales e independientes. En algunas realizaciones, los objetos de audio pueden ser fuentes de sonido grabadas individualmente. Además, los objetos de audio no tienen ninguna asociación espacial y específicamente no hay ninguna relación espacial entre los diferentes objetos de audio.

30 Por tanto, a diferencia de por ejemplo una grabación de sonido envolvente en la que la misma imagen de sonido (y fuentes de sonido) se graban en diferentes posiciones para generar diferentes canales de la misma señal espacial, los objetos de audio del presente ejemplo son fuentes de sonido individuales y aisladas.

35 En la aplicación de teleconferencia, cada objeto de audio corresponde a una señal de voz recibida desde un participante en la llamada de teleconferencia. Por tanto, el codificador 209 recibe objetos de audio en forma de señales de voz recibidas desde una pluralidad de unidades de comunicación remotas que participan en la llamada de conferencia.

40 El codificador 209 orientado a objetos codifica los objetos de audio en un número limitado de canales y adicionalmente genera datos paramétricos que permiten y facilitan una regeneración de los objetos de audio originales a partir de los canales de audio generados en el lado del decodificador. Específicamente, el codificador 209 de audio puede generar una mezcla descendente de los objetos de audio de una manera similar a la generación de una mezcla descendente de una señal de sonido envolvente espacial para dar por ejemplo una señal estéreo. Por ejemplo, el codificador 209 puede generar una mezcla descendente multiplicando los valores de muestra de objeto de audio por una matriz de mezcla descendente para generar valores de muestra de la mezcla descendente.

El codificador 209 genera un flujo de bits que comprende tanto los datos de codificación para el número limitado de canales como los datos paramétricos asociados. Estos datos se transmiten al decodificador 215.

45 El decodificador 215 comprende una unidad 303 de decodificador orientada a objetos que genera réplicas aproximadas locales de los objetos de audio originales basándose en los canales de audio recibidos y los datos paramétricos recibidos. Específicamente, la unidad 303 de decodificador orientada a objetos puede generar los objetos de audio aplicando una matriz de mezcla ascendente a las muestras de audio recibidas. Los coeficientes de la matriz de mezcla ascendente se determinan en respuesta a los datos paramétricos recibidos desde el codificador 209.

50 El decodificador 215 comprende además una unidad 305 de presentación que está dispuesta para generar una señal de salida basándose en las entradas de audio. La unidad 305 de presentación puede manipular libremente y mezclar los objetos de audio recibidos para generar una señal de salida deseada. Por ejemplo, la unidad 305 de presentación puede generar una señal de sonido envolvente de cinco canales y puede colocar libremente cada objeto de audio individual en la imagen de sonido generada. Como otro ejemplo, la unidad 305 de presentación puede generar una señal estéreo binaural que puede proporcionar una experiencia espacial a través de por ejemplo un conjunto de auriculares.

En muchos sistemas prácticos, la funcionalidad de la unidad 303 de decodificación y la unidad 305 de presentación se combina en una sola etapa de procesamiento. Por ejemplo, la operación de la unidad 303 de decodificación normalmente corresponde a una multiplicación de matriz por una matriz de mezcla ascendente y la operación de la unidad 305 de presentación corresponde de manera similar a una multiplicación de matriz realizada en la salida de la multiplicación de matriz de mezcla ascendente. Por tanto, combinando las matrices de mezcla ascendente y de presentación en una sola matriz, la multiplicación de matriz en cascada puede combinarse en una sola multiplicación de matriz.

En el ejemplo específico, la unidad 305 de presentación puede colocar cada altavoz individual de la llamada de conferencia en una ubicación diferente en la imagen de sonido pudiendo seleccionarse libremente la ubicación específica para cada altavoz por ejemplo por un usuario que controla la unidad 305 de presentación. Como otro ejemplo, si el objeto de audio corresponde a diferentes instrumentos musicales de una pieza de música, el usuario puede mezclar, ecualizar, etc. libremente los instrumentos individuales así como colocarlos libremente en la imagen de sonido. Por tanto, el enfoque descrito permite un alto grado de libertad para que el usuario individual manipule los diferentes objetos de audio para generar una salida de audio personalizada que puede ser independiente de la salida de audio generada para otros usuarios y receptores de la señal codificada desde el codificador 209.

Sin embargo, a pesar de proporcionar un grado amplio de flexibilidad manipulando los objetos de audio en la unidad 305 de presentación, tal manipulación también puede dar como resultado una degradación en la calidad de la señal de audio generada. En particular, para generar réplicas exactas de los objetos de audio en el decodificador 215, es necesario aplicar una matriz de mezcla ascendente que es la inversa de la matriz de mezcla descendente usada en el codificador 209. Sin embargo, esto generalmente no es posible (por ejemplo no es posible cuando el número de señales de audio generadas es menor que el número de objetos de audio porque no existe una matriz inversa para la matriz de mezcla descendente en este caso) y por consiguiente sólo pueden generarse aproximaciones de las señales de audio originales. Específicamente, los objetos de audio generados en el decodificador contendrán una cantidad de interferencia cruzada de otros objetos de audio. Como resultado, la manipulación de un objeto de audio afectará a la percepción y características de otro objeto de audio que puede dar como resultado un rendimiento degradado y artefactos perceptibles.

En el sistema de la figura 3, el codificador 215 puede además generar datos de control en forma de datos de modificación de codificación que se transmiten al codificador 209. Los datos de modificación de codificación se evalúan entonces por el codificador 209 que modifica el proceso de codificación dependiendo de la información de control recibida. Específicamente, el codificador 209 puede modificar la mezcla descendente de los objetos de audio y los parámetros espaciales que se generan para la mezcla descendente. Como un ejemplo específico, los datos de modificación de codificación pueden especificar que el volumen de un objeto de audio específico debe reducirse. Por consiguiente el codificador 209 reduce el nivel de este objeto de audio (por ejemplo antes de o como parte de la operación de mezcla descendente) y modifica (directa o indirectamente) los datos paramétricos para el objeto de audio de manera que cuando los objetos de audio se decodifican en el decodificador, el nivel se reducirá de manera apropiada y preferiblemente de manera que los datos paramétricos modificados representan correctamente el cambio en el nivel para el/los objeto(s) de audio respectivo(s).

El enfoque permite por tanto la realización de toda o parte de la manipulación de objetos en el lado de codificación. Como el codificador tiene acceso a los objetos de audio independientes originales en vez de sólo a las réplicas aproximadas, puede lograrse un rendimiento mejorado y en particular puede ser posible proporcionar una calidad mejorada. Por ejemplo, se reduce la interferencia cruzada y por tanto el impacto en los otros objetos de audio de aumentar o disminuir el volumen de un objeto de audio puede reducirse sustancialmente o incluso eliminarse completamente.

La figura 4 ilustra el codificador 209 en más detalle. A continuación se describirá en más detalle la operación del codificador 209 con referencia al ejemplo específico en el que el lado del decodificador genera datos de modificación de codificación que se transmiten al codificador y se usan para controlar los niveles relativos de los objetos de audio individuales.

El codificador 209 comprende una unidad 401 receptora que recibe los objetos de audio que en este caso son las señales de voz recibidas desde unidades de comunicación remotas, tales como teléfonos, que participan en la llamada de teleconferencia. Los objetos de voz se alimentan a una unidad 403 de codificación que mezcla de manera descendente los objetos para dar un número de señales de audio que es menor que el número de objetos de audio de voz. Específicamente, la unidad 403 de codificación realiza la multiplicación de matriz dada por:

$$Y = D \times X$$

en la que X indica un vector dimensional N que comprende las muestras de objeto de voz (donde N es el número de objetos de voz), Y es un vector dimensional M que comprende las muestras de salida de mezcla descendente (donde M es el número de canales de salida) y D es una matriz de mezcla descendente N,M. M puede ser significativamente menor que N. Por ejemplo, para una teleconferencia en seis direcciones, pueden mezclarse de manera descendente cinco señales de voz para dar una sola señal mono que se transmite a la sexta unidad de comunicación.

El codificador 209 comprende además una unidad 405 de parámetro que genera datos paramétricos que pueden usarse para recrear los objetos de audio a partir de la señal de mezcla descendente. Específicamente, la

unidad 405 de parámetro genera un conjunto de parámetros de objeto para cada objeto de voz que puede usarse por el decodificador 215 para recrear los objetos de voz. En el caso ideal, los parámetros de objeto se determinarían de manera que podría determinarse una matriz de mezcla ascendente correspondiente a la inversa de la matriz de mezcla descendente, es decir, la matriz de mezcla ascendente  $U=D^{-1}$ . Sin embargo, no existe una matriz inversa para una matriz de mezcla descendente (en la que  $N>M$ ) y por tanto sólo pueden generarse los datos de parámetro que permiten una regeneración no ideal de los objetos de voz originales.

Por consiguiente, la unidad 405 de parámetro genera parámetros que representan características de los objetos de voz individuales en relación con la señal de mezcla descendente. En el ejemplo específico, la unidad de parámetro primero transforma el objeto de voz en el dominio de frecuencia en bloques de tiempo (por ejemplo mediante el uso de una FFT) y luego realiza la multiplicación de matriz de mezcla descendente para cada bloque de tiempo-frecuencia (o losa de tiempo-frecuencia). Además, para los bloques de tiempo-frecuencia, se determina la amplitud relativa de cada objeto de voz en relación con el resultado de mezcla descendente. Por tanto, la unidad 405 de parámetro genera información de nivel relativa descrita en losas de tiempo/frecuencia separadas para los diversos objetos de voz. De ese modo, se genera un vector de nivel para las losas de tiempo/frecuencia representando cada elemento del vector la cantidad de energía en la losa de tiempo/frecuencia del objeto de ese elemento. Este proceso puede dar como resultado un conjunto de parámetros de energía  $\sigma_{b,t}^n$  para la banda de frecuencia  $b$ , segmento de tiempo  $t$ , y señal  $n$ . Estos parámetros pueden entonces transmitirse (preferiblemente en un dominio cuantificado, logarítmico) al extremo receptor. Por tanto, el enfoque para generar los datos de parámetro puede ser similar al enfoque usado para codificación espacial envolvente MPEG y puede lograrse una reutilización de funcionalidad en muchas realizaciones.

La unidad 405 de parámetro y la unidad 403 de codificación se acoplan a un procesador 407 de transmisión que genera un flujo de bits que comprende tanto los datos de codificación como los datos paramétricos. Específicamente, el flujo de bits puede ser una señal estéreo codificada compatible con MPEG estando comprendidos los datos paramétricos en partes de datos auxiliares del flujo de bits. Luego se transmite el flujo de bits resultante a la unidad de comunicación apropiada.

La figura 5 ilustra el decodificador 215 en más detalle. El decodificador 215 comprende la unidad 303 de decodificación orientada a objetos que genera réplicas aproximadas de los objetos de voz. Específicamente, la unidad 303 de decodificación puede generar losas de tiempo-frecuencia de los objetos de voz individuales modificando las losas de tiempo-frecuencia correspondientes de la señal de mezcla descendente recibida tal como se indica por la diferencia de nivel relativa correspondiente para ese objeto tal como se da en los datos paramétricos.

Si la señal de voz individual para el objeto  $n$  viene dada por  $x^n(t)$ , con parámetros de energía asociados  $\sigma_{b,t}^n$ , y una señal de mezcla descendente  $m(t)$ , la estimación del lado del decodificador de la señal de voz  $x^n(t)$  para la losa de tiempo/frecuencia  $(b,t)$  puede darse por:

$$\hat{x}_{b,t}^n = m_{b,t} \sqrt{\frac{\sigma_{b,t}^n}{\sum_i \sigma_{b,t}^i}}$$

Los objetos de voz se alimentan a la unidad 305 de presentación que puede proceder para generar una señal de salida para el usuario. Además, en el ejemplo, puede ser que el usuario pueda ajustar diversos parámetros y características de presentación incluyendo por ejemplo el cambio de una posición de uno o más de los objetos de voz en la imagen de sonido generada.

Además, el decodificador 215 comprende un procesador 501 de control que puede generar datos de modificación de codificación en respuesta a una entrada de usuario. Los datos de modificación de codificación se alimentan a una unidad 503 de transmisión que transmite los datos de modificación de codificación al codificador 209.

El codificador 209 comprende un receptor 409 de datos de control que recibe los datos de modificación de codificación. El receptor 409 de datos de control se acopla a la unidad 403 de codificación y la unidad 405 de parámetro que ese disponen para modificar la codificación y generación de datos de parámetro dependiendo de los datos de modificación de codificación recibidos. Por tanto, además del control de la presentación de los objetos de voz en el decodificador, el usuario del mismo también puede controlar la operación de codificación de la codificación orientada a objetos realizada en el lado del codificador.

Como un ejemplo específico, puede controlarse la imagen espacial y las ubicaciones espaciales del objeto en la señal de salida generada del decodificador modificando la operación de presentación del decodificador

mientras que pueden realizarse ajustes de (gran) volumen controlando la mezcla descendente en el codificador.

Por tanto, el usuario del decodificador puede pedir que el volumen de un objeto de voz específico se aumente sustancialmente. Si esto se realiza amplificando el objeto de voz correspondiente en el decodificador, la amplificación también amplificará las componentes de interferencia cruzada de otros objetos de voz que no sólo pueden dar como resultado un mayor volumen de los mismos sino también la distorsión de estos objetos y posiblemente un desplazamiento en la posición de estos objetos.

Sin embargo, según el ejemplo, el decodificador 215 no cambia el ajuste a escala de las réplicas del objeto de voz generadas sino más bien genera datos de modificación de codificación que harán que el codificador modifique los pesos de mezcla descendente para los objetos de voz deseados.

Por tanto, en el ejemplo las desventajas asociadas con niveles de objeto de audio individuales cambiantes en el lado del decodificador se mitigan o eliminan controlando los niveles relativos en el lado del codificador. Específicamente, las modificaciones de nivel deseadas del usuario en el lado del decodificador se transmiten al codificador y se aplican como los pesos de mezcla descendente.

En el ejemplo de teleconferencia, el extremo receptor también transmite la voz producida localmente de vuelta al centro de teleconferencia. Por consiguiente, esta señal de voz puede incluir los pesos de mezcla descendente para todos los objetos que se reciben por el receptor (o por datos que dan como resultado que el receptor cambie los pesos de mezcla descendente, por ejemplo la aplicación de una atenuación o amplificación relativa a un objeto de voz específico). Por ejemplo si el extremo receptor produce una señal "voz 0" y recibe señales "voz 1", "voz 2" y "voz 3" desde otras unidades de comunicación, puede generar y transmitir pesos de mezcla descendente para los objetos "voz 1", "voz 2" y "voz 3". Entonces se usan estos pesos de mezcla descendente por el centro de teleconferencia para generar la señal de mezcla descendente para este extremo receptor.

Una ventaja de este esquema es que el usuario tiene un grado muy alto de libertad para modificar por ejemplo el volumen o distancia de cada señal de voz individual. Además, es probable que los pesos de mezcla descendente (y otros parámetros) sean bastante constantes a través del tiempo y por tanto la tasa de transmisión de datos requerida para los datos de modificación de codificación es normalmente muy baja.

En algunas realizaciones, el codificador 209 puede disponerse para modificar al menos uno de los objetos de audio antes de realizarse la mezcla descendente. Por ejemplo, la unidad 403 de codificación puede ajustar a escala los objetos de audio recibidos antes de realizar la multiplicación de matriz de mezcla descendente. Por tanto, si se reciben los datos de modificación de codificación lo que indica que un objeto de voz específico debe ser menor, las muestras de señales recibidas para este objeto pueden multiplicarse por un factor mayor que uno. La señal resultante puede usarse entonces en la multiplicación de matriz de mezcla descendente para generar la señal de mezcla descendente. Este enfoque puede permitir usar una matriz de mezcla descendente fija y puede permitir específicamente multiplicar coeficientes que van a usarse de manera sencilla (por ejemplo la matriz de mezcla descendente podría contener sólo coeficientes unitarios reduciendo así de manera eficaz la multiplicación de mezcla descendente a un número de adiciones simples).

En el ejemplo, la determinación de los parámetros de objeto puede determinarse basándose en las señales modificadas. Por tanto, los objetos de voz ajustados a escala también pueden alimentarse a la unidad 405 de parámetro que puede determinar los niveles relativos de las losas de frecuencia-tiempo para las señales modificadas. Este enfoque dará como resultado el proceso de mezcla ascendente por el decodificador que genera un objeto de voz que tiene el nivel de volumen deseado. Por tanto, en este enfoque, la modificación de los datos paramétricos dependiendo de los datos de modificación de codificación es indirecta en el sentido de que los datos de modificación de codificación se usan primero para modificar los objetos de voz y entonces se generan los datos de parámetro basándose en los objetos de voz modificados.

En otras realizaciones, los datos paramétricos pueden modificarse más directamente. Por ejemplo, los objetos de voz pueden alimentarse directamente a la unidad 405 de parámetro antes de realizarse cualquier modificación. La unidad 405 de parámetro puede determinar entonces los niveles de intensidad relativos para las diferentes losas de frecuencia-tiempo y posteriormente ajustar los niveles medidos en respuesta a los datos de modificación de codificación. Esta modificación puede realizarse para adaptarse a la modificación del objeto de voz antes de la mezcla descendente garantizando así una generación correcta del objeto de voz con volumen compensado en el decodificador.

En algunas realizaciones, sólo se cambian los datos paramétricos en respuesta a los datos de modificación de codificación y los objetos de voz y se mantiene invariable la mezcla descendente. En este ejemplo, pueden cambiarse los parámetros de objeto de manera que el decodificador generará los objetos de voz requeridos aplicando los parámetros de objeto modificados. En este caso, para modificar un objeto de voz dado, puede ser necesario no sólo cambiar el parámetro de objeto para ese objeto de voz sino también para otros objetos de voz.

En algunas realizaciones, pueden cambiarse los pesos de mezcla descendente (por ejemplo los coeficientes de matriz de mezcla descendente) en respuesta a los datos de modificación de codificación recibidos. Por ejemplo, puede aumentarse el volumen de un objeto de voz específico aumentando el/los coeficiente(s) de matriz de mezcla descendente para ese objeto de voz. En este caso, normalmente una señal de objeto de voz modificada no está disponible y por consiguiente los parámetros de objeto pueden cambiarse directamente en

respuesta a los datos de codificación de manera que reflejen los pesos de mezcla descendente cambiados.

También se apreciará que en algunas de tales realizaciones, la modificación de un objeto de voz también puede afectar a otros objetos de voz. Por ejemplo, cuando se cambia el peso de mezcla descendente de un objeto de voz, los otros pesos de mezcla descendente pueden ajustarse de manera que la energía total de la señal de mezcla descendente permanezca invariable. Alternativa o adicionalmente, pueden modificarse los parámetros de energía relativos para las losas de frecuencia-tiempo de otros objetos de voz para reflejar una energía cambiada de la señal de mezcla descendente generada.

En algunas realizaciones, los datos de modificación de codificación pueden ser específicos de la frecuencia de manera que los diferentes datos de modificación se proporcionen para diferentes frecuencias. Por ejemplo, en vez de sólo indicar un peso de mezcla descendente modificado para un objeto de voz dado, este peso de mezcla descendente puede proporcionarse como una función de la frecuencia. Por tanto, el usuario remoto puede no sólo ajustar la ganancia de un objeto de voz como un todo sino que puede modificar la característica de frecuencia del objeto. Esto puede permitir al usuario remoto controlar de manera eficaz una operación de ecualización para el objeto de voz individual. Por tanto, en el ejemplo, al menos algunos de los datos de modificación de codificación se proporcionan como una función de frecuencia y la unidad 405 de parámetro procede por consiguiente a modificar los datos paramétricos dependiendo de la frecuencia.

Se apreciará que el transmisor 201 puede disponerse para generar señales individuales para diferentes decodificadores. Por ejemplo en la aplicación a modo de ejemplo de un centro de teleconferencia, el transmisor 201 puede recibir diferentes datos de modificación de codificación desde diferentes participantes en la teleconferencia y puede generar datos paramétricos separados y mezcla descendente para los participantes individuales.

En algunas realizaciones, el codificador 209 comprende además funcionalidad para generar la(s) señal(es) de salida como una mezcla descendente espacial. Por tanto, en el ejemplo, el codificador 209 está dispuesto para presentar los objetos de voz como una señal de salida espacial en la que cada objeto de voz se presenta en una ubicación específica con un nivel de volumen y característica de frecuencia, etc. específicos. Específicamente, la salida del codificador 209 puede ser una señal estéreo, una señal multicanal de sonido envolvente y/o una señal envolvente espacial binaural generada, por ejemplo, usando funciones de transferencia relacionadas con la cabeza.

En tales realizaciones, los datos de modificación de codificación recibidos desde el decodificador 215 pueden comprender parámetros de presentación espaciales que afectan a la presentación de los objetos de voz en la señal espacial.

Los parámetros de presentación espaciales pueden indicar por ejemplo que la posición de uno o más de los objetos de audio debe cambiarse en la mezcla de salida espacial. Como otro ejemplo, pueden proporcionarse datos de ecualización que pueden aplicarse a un objeto de audio individual. Como otro ejemplo, la distancia percibida de cada objeto de audio puede controlarse de manera remota desde el extremo decodificador. Por ejemplo, si se reciben datos de modificación de codificación lo que indica que un objeto de audio debe alejarse más en una mezcla descendente espacial, la presentación de este objeto de audio puede cambiarse de manera que el nivel de volumen se reduzca y se aumente la correlación entre los canales frontal y trasero. Se sabe que estas modificaciones afectan a la percepción de la distancia dando como resultado que el usuario que experimenta la fuente de sonido del objeto de audio se aleje más del oyente.

Como otro ejemplo, el usuario remoto puede controlar el modo de presentación espacial del codificador. Por ejemplo, para una señal de salida de dos canales, el usuario puede seleccionar si la presentación debe optimizarse para altavoces o auriculares. Específicamente, el usuario remoto puede seleccionar si la salida debe generarse como una señal estéreo tradicional o como una señal envolvente espacial binaural para su uso con auriculares.

Tal enfoque puede proporcionar una serie de ventajas. Por ejemplo, la tasa de transmisión de bits requerida para transmitir los parámetros de presentación espaciales es normalmente relativamente baja debido a que los parámetros de presentación se definen sólo por fuente de sonido (es decir, normalmente no son dependientes de la frecuencia). Además, es probable que estos parámetros sean bastante constantes a través del tiempo. Los parámetros requeridos para el enfoque de presentación del lado del decodificador, por otro lado, tienen que transmitirse para cada fuente de sonido y para cada losa de tiempo/frecuencia, dando como resultado la transmisión de cantidades significativas de datos. Por tanto, moviendo toda o parte de la presentación al lado del codificador, puede lograrse un sistema de audio eficaz.

También puede lograrse una compatibilidad mejorada con decodificadores legados. El codificador central puede generar un flujo de bits optimizado para cada decodificador de manera independiente (es decir, puede atenderse a todos los decodificadores mono, estéreo o envolventes y puede optimizarse la señal generada para el decodificador de destino específico).

El enfoque puede permitir proporcionar servicios adicionales o mejorados. Por ejemplo, cada cliente puede pagar una tarifa adicional para determinadas posibilidades de presentación (es decir, los ajustes de nivel son un primer nivel de servicio, y la presentación espacial puede ser un segundo nivel de servicio más caro).

Además, como puede disminuirse el requisito de presentación para el decodificador, es posible una complejidad reducida del decodificador de destino en muchas aplicaciones.

La figura 6 ilustra un ejemplo de un procedimiento de codificación de señales de audio según algunas realizaciones de la invención.

El procedimiento se inicia en la etapa 601 en la que se recibe una pluralidad de objetos de audio.

5 La etapa 601 va seguida por la etapa 603 en la que los datos de modificación de codificación se reciben desde una unidad remota.

La etapa 603 va seguida por la etapa 605 en la que la pluralidad de objetos de audio se codifica en un número de señales de audio y datos paramétricos que representan la pluralidad de objetos de audio en relación con el número de señales de audio. Los datos paramétricos comprenden un conjunto de parámetros de objeto para cada uno de los diferentes objetos de audio y se determinan en respuesta a los datos de modificación.

10 La figura 7 ilustra un ejemplo de un procedimiento de decodificación de objetos de audio según algunas realizaciones de la invención.

El procedimiento se inicia en la etapa 701 en la que un número de señales de audio y datos paramétricos que representan los objetos de audio en relación con el número de señales de audio se reciben desde un codificador. Las señales de audio son una mezcla descendente de los objetos de audio y los datos paramétricos comprenden un conjunto de parámetros de objeto para cada uno de los diferentes objetos de audio.

15 La etapa 701 va seguida por la etapa 703 en la que los objetos de audio se decodifican a partir del número de señales de audio en respuesta a los datos paramétricos.

La etapa 703 va seguida por la etapa 705 en la que se genera una señal de salida multicanal espacial a partir de los objetos de audio.

20 La etapa 705 va seguida por la etapa 707 en la que se generan datos de modificación de codificación para el codificador de objetos.

La etapa 707 va seguida por la etapa 709 en la que los datos de modificación de codificación se transmiten al codificador de objetos.

25 Se apreciará que por motivos de claridad la descripción anterior ha descrito realizaciones de la invención con referencia a diferentes unidades y procesadores funcionales. Sin embargo, será evidente que puede usarse cualquier distribución adecuada de funcionalidad entre diferentes unidades o procesadores funcionales sin apartarse de la invención. Por ejemplo, la funcionalidad ilustrada para su realización por procesadores o controladores separados puede realizarse por el mismo procesador o los mismos controladores. Por tanto, las referencias a unidades funcionales específicas sólo deben considerarse referencias a medios adecuados para proporcionar la funcionalidad descrita en vez de indicativas de una estructura u organización física o lógica estricta.

30 La invención puede implementarse de cualquier forma adecuada que incluya hardware, software, *firmware* o cualquier combinación de éstos. La invención puede implementarse opcionalmente al menos en parte como software informático que se ejecute en uno o más procesadores de datos y/o procesadores de señal digital. Los elementos y componentes de una realización de la invención pueden implementarse de manera física, funcional y lógica de cualquier manera adecuada. En efecto la funcionalidad puede implementarse en una sola unidad, en una pluralidad de unidades o como parte de otras unidades funcionales. Como tal, la invención puede implementarse como una sola unidad o puede distribuirse física y funcionalmente entre diferentes unidades y procesadores.

35 Aunque la presente invención se ha descrito en relación con algunas realizaciones, no se pretende que se limite a la forma específica establecida en el presente documento. Más bien, el alcance de la presente invención se limita sólo por las reivindicaciones adjuntas. Adicionalmente, aunque pueda parecer que se describe una característica en relación con realizaciones particulares, un experto en la técnica reconocerá que pueden combinarse diversas características de las realizaciones descritas según la invención. En las reivindicaciones, el término que comprende/comprendiendo no excluye la presencia de otros elementos o etapas.

40 Además, aunque se enumeren individualmente, puede implementarse una pluralidad de medios, elementos o etapas de procedimiento por ejemplo por una sola unidad o procesador. Adicionalmente, aunque pueden incluirse características individuales en diferentes reivindicaciones, éstas pueden combinarse posiblemente de manera ventajosa, y la inclusión en diferentes reivindicaciones no implica que no sea factible y/o ventajosa una combinación de características. Tampoco la inclusión de una característica en una categoría de reivindicaciones implica una limitación a esta categoría sino que más bien indica que la característica puede aplicarse igualmente a otras categorías de reivindicaciones según sea apropiado. Además, el orden de las características en las reivindicaciones no implica ningún orden específico en el que deban aplicarse las características y en particular el orden de las etapas individuales en una reivindicación del procedimiento no implica que las etapas deban realizarse en este orden. Más bien, las etapas pueden realizarse en cualquier orden adecuado. Además, las referencias al singular no excluyen una pluralidad. Por tanto las referencias a "un", "una", "primer/a", "segundo/a", etc. no descartan una pluralidad. Los signos de referencia en las reivindicaciones se proporcionan meramente como ejemplo aclaratorio y no deben interpretarse como que limitan el alcance de las reivindicaciones de ningún modo.

**REIVINDICACIONES**

1. Codificador para codificar objetos de audio, comprendiendo el codificador:
  - medios (401) para recibir una pluralidad de objetos de audio;
  - medios (403) de codificación para codificar la pluralidad de objetos de audio en un número de señales de audio y datos paramétricos que representan la pluralidad de objetos de audio en relación con el número de señales de audio, comprendiendo los datos paramétricos un conjunto de parámetros de objeto para al menos uno de los diferentes objetos de audio;
  - medios (409) para recibir datos de modificación de codificación desde una unidad remota; y
  - medios (405) de parámetro para determinar los datos paramétricos en respuesta a los datos de modificación de codificación.
2. Codificador según la reivindicación 1, en el que los medios de codificación están dispuestos para generar el número de señales de audio mediante una mezcla descendente de los objetos de audio y los medios (405) de parámetro están dispuestos para modificar un peso de mezcla descendente de al menos uno de los objetos de audio en respuesta a los datos de modificación de codificación.
3. Codificador según la reivindicación 1, en el que los medios (405) de parámetro están dispuestos para ajustar a escala al menos un primer objeto de audio en respuesta a los datos de modificación de codificación y para modificar parámetros de objeto para el primer objeto de audio en respuesta al ajuste a escala.
4. Codificador según la reivindicación 1, en el que al menos algunos de los datos de modificación de codificación son específicos de la frecuencia y los medios (405) de parámetro están dispuestos para determinar al menos un parámetro de objeto en respuesta a una característica de frecuencia del parámetro de objeto.
5. Codificador según la reivindicación 1, en el que los medios (403) de codificación están dispuestos para modificar al menos un objeto de audio en respuesta a los datos de modificación de codificación antes de una mezcla descendente de los objetos de audio al número de señales de audio.
6. Codificador según la reivindicación 1, en el que los medios (403) de codificación están dispuestos para generar el número de señales de audio como una mezcla descendente espacial.
7. Codificador según la reivindicación 6, en el que los medios (403) de codificación están dispuestos para modificar en respuesta a los datos de modificación de codificación al menos una característica seleccionada del grupo que consiste en:
  - una ubicación espacial de al menos uno de los objetos de audio;
  - una característica de distancia de al menos uno de los objetos de audio;
  - un modo de presentación espacial del codificador, y
  - una característica de frecuencia de al menos uno de los objetos de audio.
8. Codificador según la reivindicación 1, en el que cada objeto de audio está asociado con un conjunto de fuentes de audio que son independientes de las fuentes de audio de otros objetos de audio.
9. Codificador según la reivindicación 1, en el que el codificador está dispuesto para recibir un primer objeto de audio desde la unidad remota y los medios (409) para recibir los datos de modificación de codificación están dispuestos para extraer los datos de modificación de codificación de datos de codificación recibidos para el primer objeto de audio.
10. Codificador según la reivindicación 1, en el que el codificador está dispuesto para recibir datos de modificación de codificación desde una pluralidad de unidades remotas y para generar diferentes datos paramétricos para las diferentes unidades remotas en respuesta a recibir diferentes datos de modificación de codificación desde las diferentes unidades remotas.
11. Decodificador para decodificar objetos de audio, comprendiendo el decodificador:
  - un receptor (303) para recibir desde un codificador de objetos de audio remoto un número de señales de audio que son una mezcla descendente de una pluralidad de objetos de audio y datos paramétricos que representan la pluralidad de objetos de audio en relación con el número de señales de audio, comprendiendo los datos paramétricos un conjunto de parámetros de objeto para al menos uno de los diferentes objetos de audio;
  - medios (303) de decodificación para decodificar los objetos de audio del número de señales de audio en

respuesta a los datos paramétricos;

- medios (305) de presentación para generar una señal de salida multicanal espacial a partir de los objetos de audio;

5 - medios (501) para generar datos de modificación de codificación para el codificador de objetos de audio remoto; y

- medios (503) para transmitir los datos de modificación de codificación al codificador de objetos de audio remoto.

12. Centro de teleconferencia para soportar una teleconferencia entre una pluralidad de unidades de comunicación remotas, comprendiendo el centro de teleconferencia:

10 - medios (401) para recibir una primera pluralidad de señales de voz desde la pluralidad de unidades de comunicación remotas;

15 - medios (403) de codificación para codificar para una primera unidad de comunicación remota la primera pluralidad de señales de voz en un número de señales de audio y datos paramétricos que representan la pluralidad de señales de voz en relación con el número de señales de audio, comprendiendo los datos paramétricos un conjunto de parámetros de objeto para al menos una de las diferentes señales de voz;

- medios (409) para recibir datos de modificación de codificación desde la primera unidad de comunicación remota; y

- medios (405) de parámetro para determinar los datos paramétricos en respuesta a los datos de modificación de codificación; y

20 - medios (407) para transmitir el número de señales de audio y datos paramétricos a la primera unidad de comunicación remota.

13. Transmisor para transmitir señales de audio, comprendiendo el transmisor un codificador según la reivindicación 1.

25 14. Receptor para recibir señales de audio, comprendiendo el receptor un decodificador según la reivindicación 11.

15. Sistema de comunicación para comunicar señales de audio, comprendiendo el sistema de comunicación:

- un transmisor (201) según la reivindicación 13; y

- un receptor (203) según la reivindicación 14

30 en el que el receptor está configurado para proporcionar los datos de modificación de codificación al transmisor (201).

16. Procedimiento de codificación de objetos de audio, comprendiendo el procedimiento:

- recibir (601) una pluralidad de objetos de audio;

35 - codificar (603) la pluralidad de objetos de audio en un número de señales de audio y datos paramétricos que representan la pluralidad de objetos de audio en relación con el número de señales de audio, comprendiendo los datos paramétricos un conjunto de parámetros de objeto para al menos uno de los diferentes objetos de audio;

- recibir (605) datos de modificación de codificación desde una unidad remota; y

- determinar (603) los datos paramétricos en respuesta a los datos de modificación.

17. Procedimiento de decodificación de objetos de audio, comprendiendo el procedimiento:

40 - recibir (701) desde un codificador de objetos de audio remoto un número de señales de audio que son una mezcla descendente de una pluralidad de objetos de audio y datos paramétricos que representan la pluralidad de objetos de audio en relación con el número de señales de audio, comprendiendo los datos paramétricos un conjunto de parámetros de objeto para al menos uno de los diferentes objetos de audio;

45 - decodificar (703) los objetos de audio a partir del número de señales de audio en respuesta a los datos paramétricos;

- generar (705) una señal de salida multicanal espacial a partir de los objetos de audio;

- generar (707) datos de modificación de codificación para el codificador de objetos de audio remoto; y

- transmitir (709) los datos de modificación de codificación al codificador de objetos de audio remoto.

18. Producto de programa informático para ejecutar el procedimiento según cualquiera de las reivindicaciones 16 a 17.

5 19. Dispositivo (203) de reproducción de audio que comprende un decodificador (215) según la reivindicación 11.

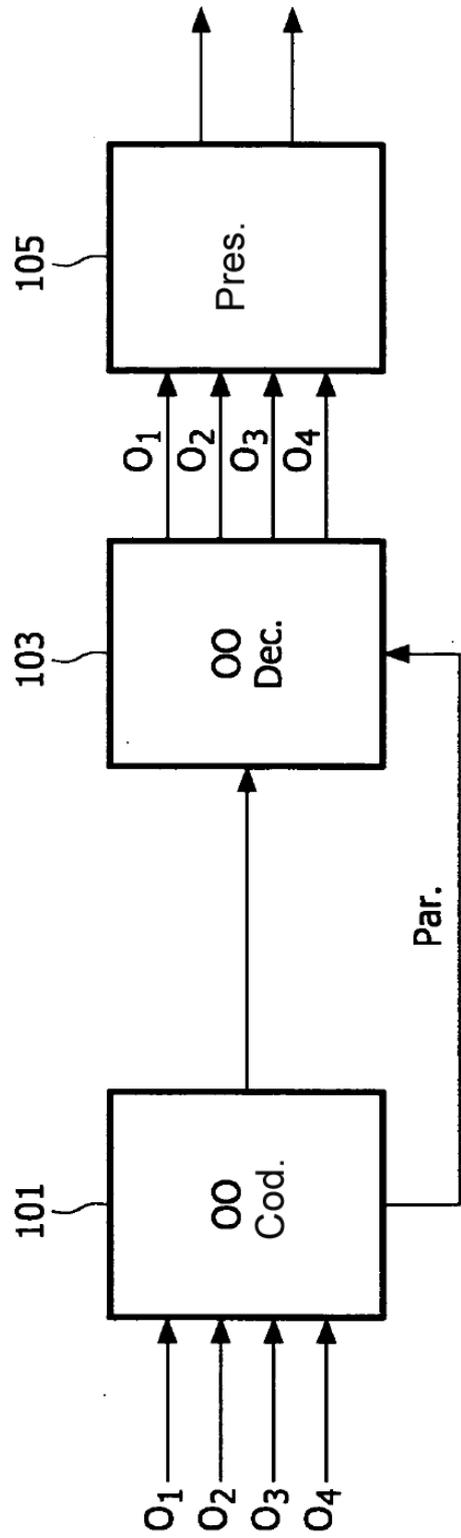


FIG. 1 Técnica anterior

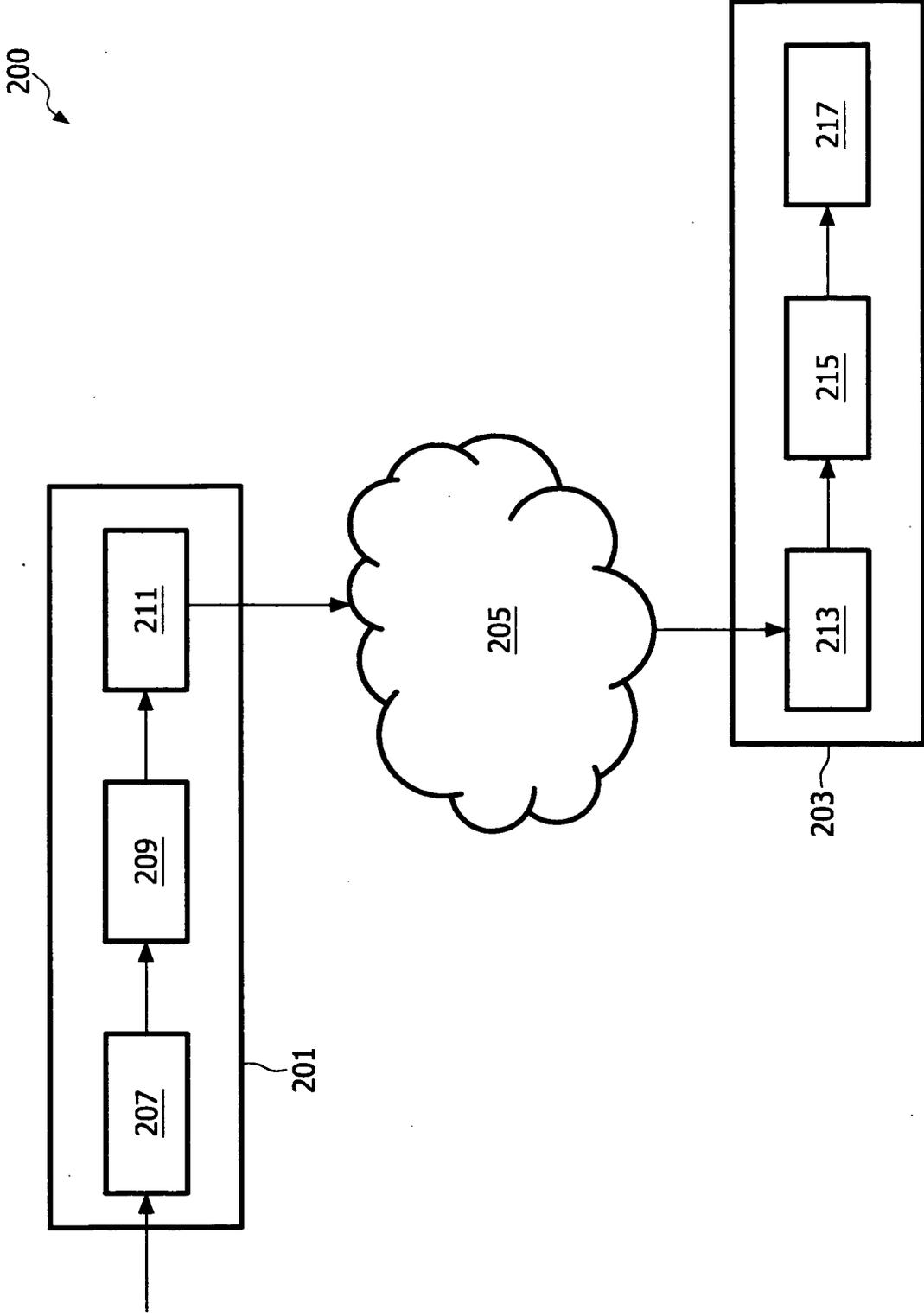


FIG. 2

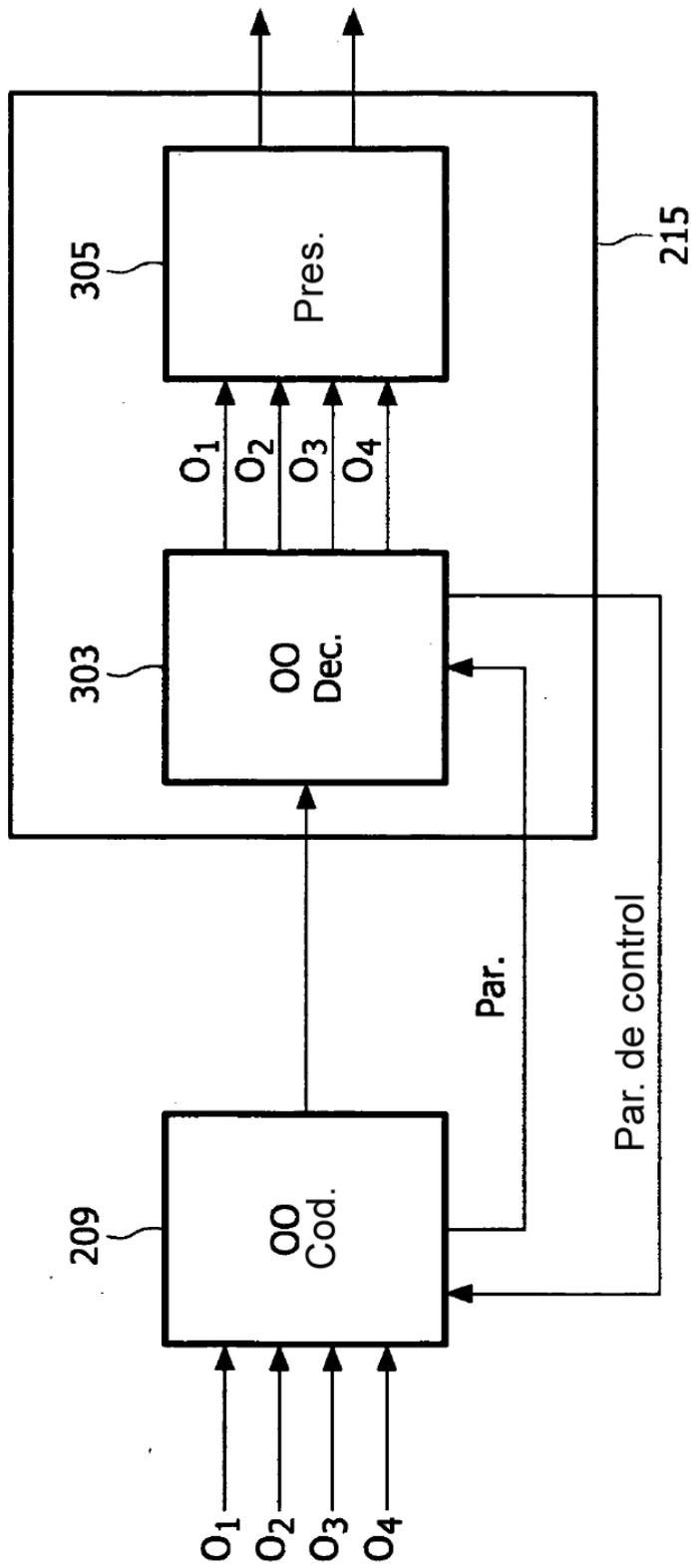


FIG. 3

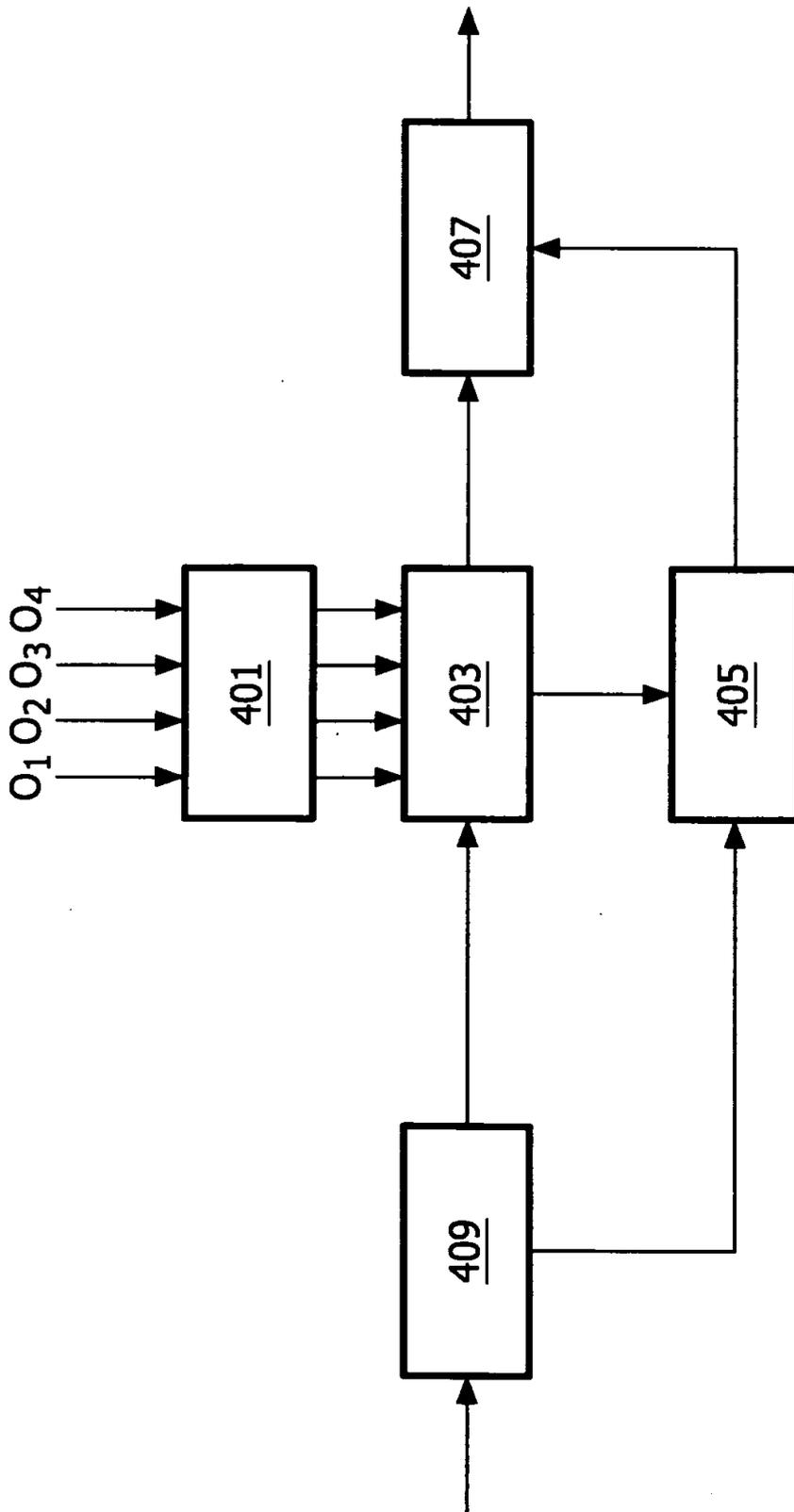
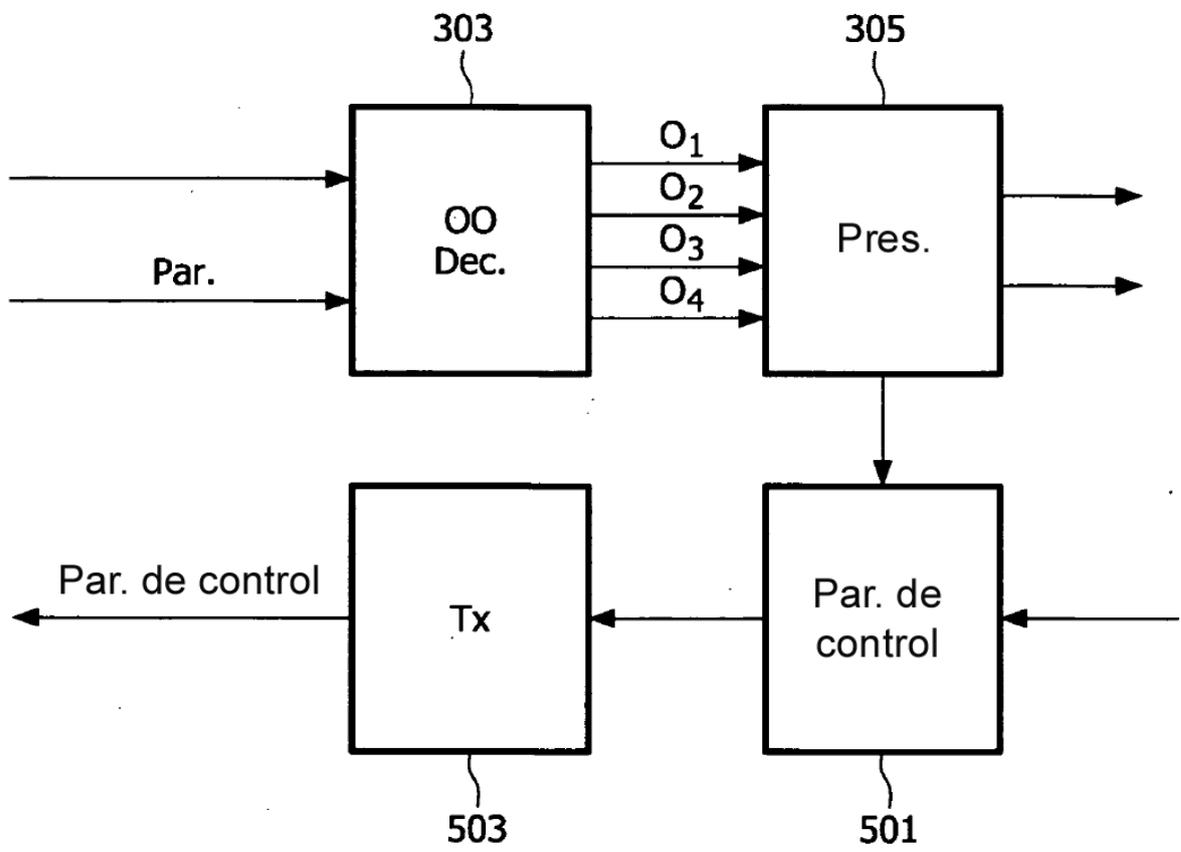
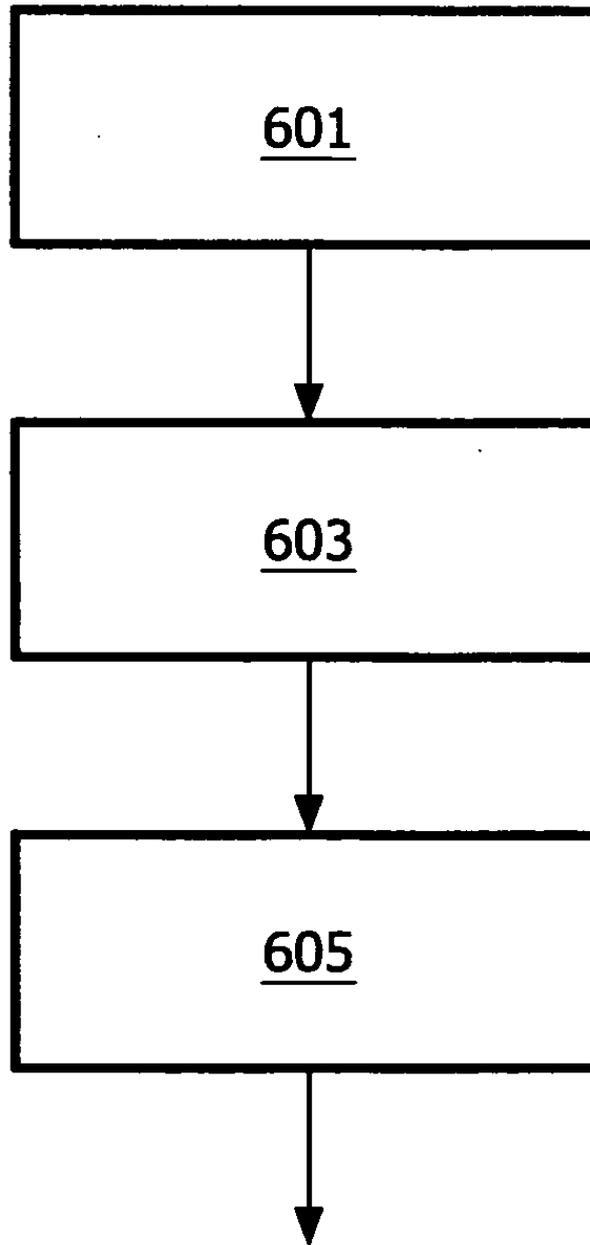


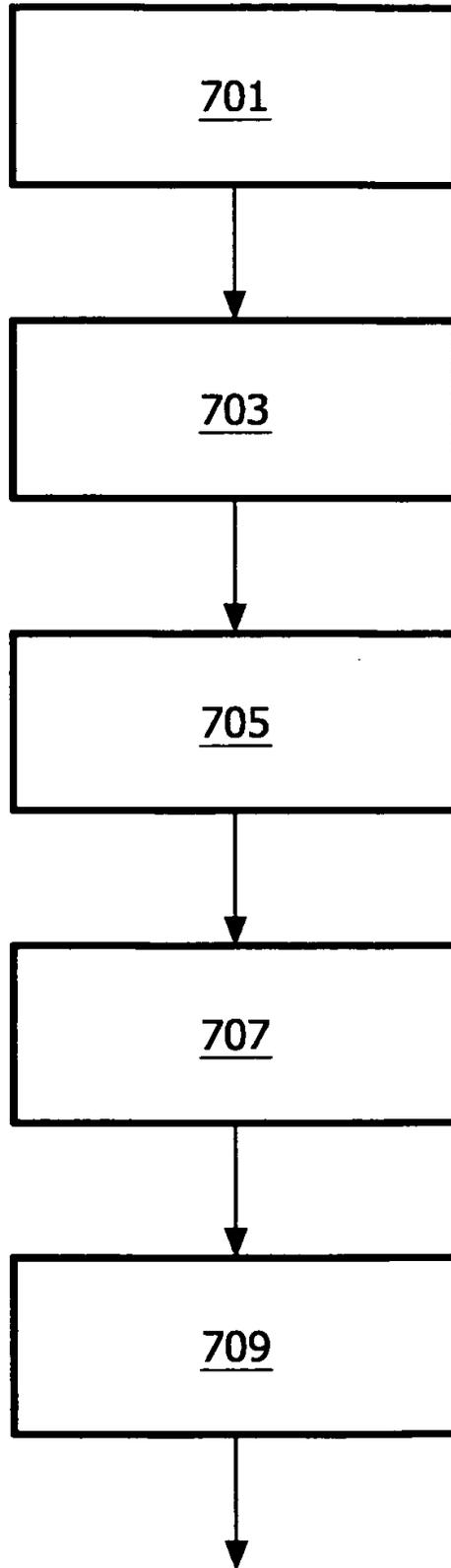
FIG. 4



**FIG. 5**



**FIG. 6**



**FIG. 7**