

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 659 179**

51 Int. Cl.:

**H04S 5/00** (2006.01)

**H04S 7/00** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **11.11.2013 PCT/EP2013/073482**

87 Fecha y número de publicación internacional: **22.05.2014 WO14076030**

96 Fecha de presentación y número de la solicitud europea: **11.11.2013 E 13791783 (7)**

97 Fecha y número de publicación de la concesión europea: **20.12.2017 EP 2920982**

54 Título: **Ajuste por segmentos de señal de audio espacial a diferente configuración de altavoces de reproducción**

30 Prioridad:

**15.11.2012 US 201261726878 P**  
**15.03.2013 EP 13159424**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**14.03.2018**

73 Titular/es:

**FRAUNHOFER-GESELLSCHAFT ZUR FÖRDERUNG DER ANGEWANDTEN FORSCHUNG E.V. (50.0%)**  
**Hansastraße 27c**  
**80686 München, DE y**  
**TECHNISCHE UNIVERSITÄT ILMENAU (50.0%)**

72 Inventor/es:

**ADAMI, ALEXANDER;**  
**HERRE, JÜRGEN;**  
**KUNTZ, ACHIM;**  
**DEL GALDO, GIOVANNI y**  
**KÜCH, FABIAN**

74 Agente/Representante:

**ARIZTI ACHA, Monica**

ES 2 659 179 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).

Ajuste por segmentos de señal de audio espacial a diferente configuración de altavoces de reproducción

## DESCRIPCIÓN

### 5 Campo técnico

La presente invención se refiere en general al procesamiento de señales de audio espaciales, y en particular a un aparato y un método para adaptar una señal de audio espacial destinada para una configuración de altavoces original a una configuración de altavoces de reproducción, que difiere de la configuración de altavoces original. Las realizaciones adicionales de la presente invención se refieren a la conversión en escena de sonido de canales múltiples (multi-canal), de alta calidad, flexible.

### Antecedentes de la invención

15 Los requisitos de un sistema moderno de reproducción de audio han cambiado durante los años. A partir del canal sencillo (mono) al canal doble (estéreo) hasta los sistemas multi-canal, como envolvente 5.1 y 7.1 o incluso la síntesis de campo de onda uniforme, el número de canales de altavoz utilizados se ha incrementado. Incluso los sistemas con altavoces elevados se han de observar en las salas de cine modernas. Esto tiene como objetivo proporcionar al oyente una experiencia de audio de una escena de audio grabada o artificialmente creada, con respecto a la sensación de realidad, inmersión y involucramiento que se asemeja tanto como sea posible a la escena de audio real, o como alternativa refleja mejor las intenciones del ingeniero de sonido (véase por ejemplo, M. Morimoto, "The Role of Rear Loudspeakers in Spatial Impression", en 103rd Convention of the AES, 1997; D. Griesinger, "Spaciousnes and Envelopment in Musical Acoustics", en 101st Convention of the AES, 1996; K. Hamasaki, K. Hiyama y R. Okumura, "The 22.2 Multichanel Sound System and Its Application", en 118th Convention of the AES, 2005). Sin embargo, existen al menos dos inconvenientes: debido a la pluralidad de sistemas de sonido disponibles, con respecto al número de altavoces utilizas y su colocación recomendada, no existe compatibilidad general entre todos estos sistemas. Además, cualquier desviación de la colocación recomendada de los altavoces dará como resultado una escena de audio comprometida y, por lo tanto, disminuye la experiencia de audio espacial del oyente, y por lo tanto, la calidad espacial.

30 En una aplicación en el mundo real, los sistemas de reproducción multi-canal a menudo no se configuran correctamente con respecto a la colocación de los altavoces. Para no distorsionar la imagen espacial original de una escena de audio que resultaría de una colocación errónea, un sistema flexible de alta calidad es necesario, el cual sea capaz de compensar estos desajustes de la configuración. Los enfoques del estado de la técnica frecuentemente carecen de la capacidad para describir una escena de sonido compleja y quizás artificialmente generada donde, por ejemplo, aparece más de una fuente directa por banda de frecuencia e instante de tiempo.

35 Por lo tanto, un objetivo de la presente invención es proporcionar un concepto mejorado para adaptar una señal de audio espacial, de modo que la imagen espacial de una escena de audio se mantenga sustancialmente igual si la configuración de altavoces de reproducción se desvía de la configuración de altavoces original, es decir, la configuración de altavoces para la cual un contenido de audio de la señal de audio espacial se produjo originalmente.

40 La publicación "Multichannel surround format conversion and generalized upmix", Michael M. Goodwin et al., AES 30th International Conference, 15-17 de marzo de 2007, desvela técnicas de separación de mezclado ascendente y de origen multicanal en el contexto del problema de reproducción de una grabación multicanal, destinada para reproducción a través de una distribución de altavoces predefinida, a través de una distribución diferente que comprende potencialmente un número diferente de canales. Particularmente, se describe un método de dominio de frecuencia basado en codificación de audio espacial que usa un mezclado descendente mono intermedia.

50 El documento GB 2 457 508 A desvela un método de adaptación de audio que comprende las etapas de estimación de una posición, donde las señales acústicas de una pluralidad de altavoces son coincidentes (un 'punto dulce') que estima la posición de un usuario, estimando la posición del usuario con relación a la posición estimada de señales coincidentes y ajustando la temporización de salida de señales de datos de sonido para reproducción por respectivos altavoces de acuerdo con la diferencia en las posiciones estimadas para mover la posición efectiva del punto dulce a la posición estimada del usuario.

60 El documento WO 2010/080451 A1 desvela una tecnología de traslación espacial de canal de audio donde M canales de entrada de audio se traducen a N canales de salida de audio, cada uno asociado con una dirección espacial. Particularmente, uno de los canales de entrada de audio está asociado con una dirección espacial distinta de una dirección espacial con la que está asociado uno de los canales de salida de audio. Particularmente, al menos uno de los canales de entrada de audio está mapeado a un conjunto respectivo de tres, cuatro o cinco de los canales de salida.

65

**Sumario de la invención**

Este objetivo se consigue por un aparato de acuerdo con la reivindicación 1, un método de acuerdo con la reivindicación 15, o un programa informático de acuerdo con la reivindicación 16.

5 De acuerdo con una realización de la presente invención, se proporciona un aparato para adaptar una señal de audio espacial para una configuración de altavoces original a una configuración de altavoces de reproducción que difiere de la configuración de altavoces original. La señal de audio espacial comprende una pluralidad de señales de canales. El aparato comprende un agrupador configurado para agrupar al menos dos señales de canal en un segmento. El aparato comprende también un descomponedor de ambiente directo, configurado para descomponer al menos dos señales de canal en el segmento en al menos un componente de sonido directo y al menos un componente ambiental. El descomponedor de ambiente directo puede configurarse además para determinar una dirección de llegada de al menos un componente de sonido directo. El aparato también comprende un representador de sonido directo configurado para recibir una información de la configuración de altavoces de reproducción para al menos un segmento de reproducción asociado con el segmento, y para ajustar al menos un componente de sonido directo utilizando la información de la configuración de altavoces de reproducción para el segmento, de modo que una dirección percibida de llegada de al menos un componente de sonido directo en la configuración de altavoces de reproducción, es idéntica a la dirección de llegada del segmento o más cercana a la dirección de llegada de al menos un componente de sonido directo en comparación con una situación en la cual no ha tenido lugar ningún ajuste. Además, el aparato comprende un combinador configurado para combinar los componentes de sonido directo ajustados, y los componentes ambientales o los componentes ambientales modificados, para obtener las señales del altavoz para al menos dos altavoces de la configuración de altavoces de reproducción.

La idea básica subyacente de la presente invención es agrupar los canales de altavoces vecinos en segmentos (por ejemplo, sectores circulares, sectores cilíndricos o sectores esféricos) y descomponer cada señal de segmento en las partes de señales directa y ambiental correspondientes. Las señales directas conducen a una posición de fuente fantasma (o varias posiciones de fuente fantasma) dentro de cada segmento, mientras que las señales ambientales corresponden a sonido difuso y son responsables del envolvimiento del oyente. Durante el proceso de representación, los componentes directos se vuelven a mapear, ponderar y ajustar por medio de las posiciones de fuente fantasma para ajustar la configuración de altavoces de reproducción real y preservar la localización original de las fuentes. Los componentes ambientales se vuelven a mapear y ponderar para producir la misma cantidad de envolvimiento en la configuración de escucha modificada. Al menos algo del procesamiento puede llevarse a cabo en una base de componentes tiempo-frecuencia. Con esta metodología, incluso puede manejarse un número incrementado o disminuido de altavoces en la configuración de salida.

Un segmento de la configuración de altavoces original puede llamarse también un "segmento original", para referencia más fácil en la siguiente descripción. De igual modo, un segmento en la configuración de altavoces de reproducción puede llamarse también un "segmento de reproducción". Un segmento se extiende o delimita típicamente por dos o más altavoces en una posición de un oyente, es decir, un segmento corresponde típicamente al espacio que se delimita por dos o más altavoces y un oyente. Un altavoz dado puede asignarse a dos o más segmentos. En una configuración de altavoces bidimensional, un altavoz particular se asigna típicamente a un segmento "izquierdo" y un segmento "derecho", es decir, el altavoz emite sonido principalmente hacia los segmentos izquierdo y derecho. El agrupador (o el elemento de agrupamiento) está configurado para reunir aquellas señales de canal que están asociadas con un segmento dado. Ya que cada señal de canal puede asignarse a dos o más canales, puede distribuirse a estos dos o más segmentos por el agrupador o por varios agrupadores.

El descomponedor de ambiente directo puede configurarse para determinar los componentes de sonido directo y los componentes ambientales para cada canal. Como alternativa, el descomponedor de ambiente directo puede configurarse para determinar un componente de sonido directo único y un componente de ambiente único por segmento. La o las direcciones de llegada pueden determinarse mediante el análisis (por ejemplo, la correlación cruzada) de al menos dos señales de canal. Como una alternativa, la o las direcciones de llegada pueden determinarse basándose en la información proporcionada al descomponedor de ambiente directo a partir de un componente adicional del aparato o a partir de una entidad externa.

El representador de sonido directo puede considerar típicamente cómo una diferencia entre la configuración de altavoces original y la configuración de altavoces de reproducción afecta a un segmento actualmente contemplado de la configuración de altavoces original, y qué medidas tienen que tomarse para mantener la percepción de los componentes de sonido directo dentro de dicho segmento. Estas medidas pueden comprender (lista no exhaustiva):

- 60 - modificar una ponderación de amplitud del componente de sonido directo entre los altavoces de dicho segmento;
- modificar una relación de fase y/o una relación de retardo entre los componentes de sonido directo específicos del altavoz para los altavoces de dicho segmento;
- eliminar el componente de sonido directo para el segmento proveniente de un altavoz particular, debido a la disponibilidad de un altavoz mejor adecuado en la configuración de altavoces de reproducción;

- aplicar el componente de sonido directo para un segmento vecino en la configuración de altavoces original a un altavoz en el segmento actualmente contemplado debido a que el altavoz es mejor adecuado para reproducir dicho componente de sonido directo (por ejemplo, debido a un límite de segmento que ha cruzado la dirección de llegada para una fuente fantasma cuando se pasa de la configuración de altavoces original a la configuración de altavoces de reproducción);
- aplicar el componente de sonido directo a un altavoz agregado (altavoz adicional) que está disponible en la configuración de altavoces de reproducción, pero no en la configuración de altavoces original;
- posibles medidas adicionales como se describen más adelante.

10 El representador de sonido directo puede comprender una pluralidad de representadores de segmentos, cada representador de segmento realiza el procesamiento de las señales de canal de un segmento.

15 El combinador puede combinar los componentes de sonido directo ajustados, los componentes ambientales, y/o los componentes ambientales modificados, que se han generados por el representador de sonido directo (o un representador de sonido directo adicional) para uno o más segmentos vecinos con relación a un segmento actualmente contemplado. De acuerdo con algunas realizaciones los componentes ambientales pueden ser sustancialmente idénticos a al menos un componente ambiental determinado por el descomponedor de ambiente directo. De acuerdo con las realizaciones alternativas, los componentes ambientales modificados pueden determinarse basándose en los componentes ambientales determinados por el descomponedor de ambiente directo, teniendo en cuenta una diferencia entre el segmento original y el segmento de reproducción.

20 De acuerdo con una realización adicional, la configuración de altavoces de reproducción puede comprender un altavoz adicional dentro del segmento. Por lo tanto, el segmento de la configuración de altavoces original corresponde a dos o más segmentos del segmento de altavoz de reproducción, es decir, el segmento original en la configuración de altavoces original se ha dividido en dos o más segmentos de reproducción en la configuración de altavoces de reproducción. El representador de sonido directo puede configurarse para generar los componentes de sonido directo ajustados, para al menos dos altavoces y el altavoz adicional de la configuración de altavoces de reproducción.

25 El caso opuesto es también posible: de acuerdo con una realización adicional, la configuración de altavoces de reproducción puede carecer de un altavoz en comparación con la configuración de altavoces original, de modo que el segmento y un segmento vecino de la configuración de altavoces original se unen a un segmento unido de la configuración de altavoces de reproducción. El representador de sonido directo puede configurarse entonces para distribuir componentes de sonido directo, ajustados, de una señal de canal que corresponde al altavoz que carece en la configuración de altavoces de reproducción a al menos dos altavoces restantes del segmento unido de la configuración de altavoces de reproducción. El altavoz que está presente en la configuración de altavoces original, pero no en la configuración de altavoces de reproducción, puede denominarse también como "altavoz carente".

30 De acuerdo con realizaciones adicionales, el representador de sonido directo puede configurarse para reasignar un componente de sonido directo que tiene una dirección determinada de llegada desde el segmento en la configuración de altavoces original hacia un segmento vecino en la configuración de altavoces de reproducción si un límite entre el segmento y el segmento vecino traspasa o cruza la dirección determinada de llegada cuando se pasa de la configuración de altavoces original a la configuración de altavoces de reproducción.

35 De acuerdo con realizaciones adicionales, el representador de sonido directo puede configurarse además para reasignar el componente de sonido directo que tiene la dirección determinada de llegada desde al menos un primer altavoz hacia al menos un segundo altavoz, asignándose el al menos un primer altavoz al segmento en la configuración de altavoces original, pero no al segmento vecino en la configuración de altavoces de reproducción, y asignándose el al menos un segundo altavoz al segmento vecino en la configuración de altavoces de reproducción.

40 De acuerdo con realizaciones adicionales, el representador de sonido directo puede configurarse para generar componentes de sonido directo específicos del segmento de altavoz, para al menos dos pares de segmento de altavoz válido de la configuración de altavoces de reproducción, refiriéndose los al menos dos pares de segmento de altavoz válido al mismo altavoz y dos segmentos vecinos en la configuración de altavoces de reproducción. El combinador puede configurarse para combinar los componentes de sonido directo específicos del segmento de altavoz, para al menos dos pares de segmentos de altavoz válidos con referencia al mismo altavoz, para obtener una de las señales de altavoz para al menos dos altavoces de la configuración de altavoces de reproducción. Un par de segmento de altavoz válido se refiere a un altavoz y a uno de los segmentos al que está asignado este altavoz. El altavoz puede ser parte de pares de segmentos de altavoz válidos, adicionales, si el altavoz se asigna a segmentos adicionales (como es típicamente el caso). De igual modo, el segmento puede ser (y típicamente es) parte de los pares de segmentos de altavoz válidos, adicionales. El representador de sonido directo puede configurarse para considerar esta ambivalencia de cada altavoz y proporcionar componentes de sonido directo específicos del segmento, para el altavoz. El combinador puede configurarse para reunir los diferentes componentes de sonido directo, específicos de segmento (y posiblemente, como pueda ser el caso, también los componentes ambientales

específicos de segmento) destinados para un altavoz particular de la configuración de altavoces de reproducción, a partir de los diversos segmentos a los que se asigna este altavoz particular. Obsérvese que la adición o la eliminación de un altavoz en la configuración de altavoces de reproducción puede tener impacto sobre los pares de segmentos de altavoz, válidos: la adición de un altavoz divide típicamente un segmento original en al menos dos segmentos de reproducción, de modo que los altavoces afectados se asignan a nuevos segmentos en la configuración de altavoces de reproducción. La eliminación de un altavoz puede dar como resultado dos o más segmentos originales que se unen a un segmento de reproducción y una influencia correspondiente sobre los pares de segmento de altavoz válidos.

10 Las realizaciones adicionales de la presente invención proporcionan un método para adaptar una señal de audio espacial destinada para una configuración de altavoces original hacia una configuración de altavoces de reproducción, que difiere de la configuración de altavoces original. La señal de audio espacial comprende una pluralidad de canales. El método comprende el agrupamiento de al menos dos señales de canal en un segmento, y la descomposición de al menos dos señales de canal en el segmento en al menos un componente de sonido directo y al menos un componente ambiental. El método comprende además la determinación de una dirección de llegada de al menos un componente de sonido directo. El método comprende también el ajuste de al menos un componente de sonido directo utilizando una información de la configuración de altavoces de reproducción para el segmento, de modo que una dirección percibida de llegada del componente de sonido directo en la configuración de altavoces de reproducción es sustancialmente idéntica a la dirección de llegada del segmento. Al menos, la dirección percibida de llegada de al menos de un componente de sonido directo es más cercana a la dirección de llegada del segmento en comparación con una situación en la que no ha tenido lugar ningún ajuste. El método comprende además la combinación de los componentes de sonido directos ajustados, y los componentes ambientales o los componentes ambientales modificados para obtener señales de altavoz para al menos dos altavoces de la configuración de altavoces de reproducción.

#### 25 **Breve descripción de las figuras**

A continuación, las realizaciones de la presente invención se explicarán con referencia a los dibujos adjuntos, en las cuales:

- 30 La Figura 1 muestra un diagrama esquemático de bloques de un posible escenario de aplicación;
- La Figura 2 muestra un diagrama esquemático de bloques de una visión general de un sistema de un aparato y un método para ajustar una señal de audio espacial;
- 35 La Figura 3 muestra una ilustración esquemática de un ejemplo para una configuración de altavoces modificada con un altavoz que se ha movido/desplazado;
- La Figura 4 muestra una ilustración esquemática de un ejemplo para otra configuración de altavoces modificada, con un número incrementado de altavoces;
- 40 La Figura 5 muestra una ilustración esquemática de un ejemplo para otra configuración de altavoces modificada, con un número disminuido de altavoces;
- 45 Las Figuras 6A y 6B muestran ilustraciones esquemáticas de los ejemplos para las configuraciones de altavoz modificadas, adicionales, con altavoces desplazados;
- La Figura 7 muestra un diagrama esquemático de bloques de un aparato para ajustar una señal de audio espacial; y
- 50 La Figura 8 muestra un diagrama de flujo esquemático de un método para ajustar una señal de audio espacial.

#### **Descripción detallada de la invención**

55 Antes de analizar la presente invención con detalle adicional utilizando los dibujos, se señala que en las figuras los elementos idénticos, elementos que tienen la misma función o el mismo efecto se proporcionan con los mismos o con similares números de referencia, de modo que la descripción de estos elementos y la funcionalidad de los mismos, ilustrada en las diferentes realizaciones, es mutuamente intercambiable o puede aplicarse entre sí en las diferentes realizaciones.

60 Algunos métodos para ajustar una señal de audio espacial no son lo suficientemente flexibles para manejar una escena de sonido complejo, especialmente aquellas que están basadas en suposiciones físicas globales (véase por ejemplo, V. Pulkki, "Spatial Sound Reproduction with Directional Audio Coding", J. Audio Eng. Soc, vol. 55, n.º 6, pág. 503-516, 2007 y V. Pulkki y J. Herre, "Method and Apparatus for Conversion Between Multi-Channel Audio

- Formats”, *Solicitud de Patente de Estados Unidos Publicación N.º 2008/023616 A1*) o están restringidas a un componente localizable (directo) por banda de frecuencia en la escena de audio completa (véase por ejemplo, M. Goodwin y J.-M. Jot, “Spatial Audio Scene Coding”, en 125th Convention of the AES, 2008 y J. Thompson, B. Smith, A. Warner, y J.-M. Jot, “Direct-Diffuse Decomposition of Multichannel Signals Using a System of Pairwise Correlations”, en 133rd Convention of the AES, 2012, octubre de 2012). La onda de un plano o la suposición de componente directo puede ser suficiente en algunos escenarios especiales pero, en general, no es capaz de capturar una escena de audio compleja con varias fuentes activas a la vez. Esto da como resultado la distorsión espacial y fuentes inestables o incluso que saltan durante la reproducción.
- 10 Existen sistemas que modelan los altavoces de la configuración de entrada que no coinciden con la configuración de salida como altavoces virtuales (la señal de altavoz completa se hace panorámica por los altavoces vecinos a la posición pretendida del altavoz) (A. Ando, “Conversion of Multichannel Sound Signal Maintaining Physical Properties of Sound in Reproduced Sound Field”, *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, n.º 6, pág. 1467-1475, 2011). Esto también puede dar como resultado la distorsión espacial de las fuentes fantasma a las cuales contribuyen esos canales de altavoz. El enfoque mencionado por A. Laborie, R. Bruno y S. Montoya en “Reproducing Multichannel Sound on any Speaker Layout”, 118th Convention of the AES, 2005, necesita que el usuario calibre en primer lugar sus altavoces y después de esto represente las señales para esa configuración a partir de una transformada de señal computacionalmente intensiva.
- 15 Además, un sistema de alta calidad debe ser preservador de la forma de onda. Cuando los canales de entrada se distribuyen a una configuración de altavoces que equivale a la configuración de entrada, la forma de onda no debe cambiar de manera significativa, de otro modo se pierde la información que puede dar como resultado artefactos audibles y disminución de la calidad espacial y de audio. Los métodos basados en objeto pueden sufrir aquí de diafonía adicional que se introduce durante la extracción del objeto (F. Melchior, “Vorrichtung zum Verändern einer Audio-Szene und Vorrichtung zum Erzeugen einer Richtungsfunktion”, *Solicitud de Patente Alemana N.º DE 10 2010 030 534 A1*, 2011). Las suposiciones físicas globales también dan como resultado diferentes formas de onda (véase por ejemplo, M. Goodwin y J.-M. Jot, “Spatial Audio Scene Coding”, en 125th Convention of the AES, 2008; V. Pulkki, “Spatial Sound Reproduction with Directional Audio Coding”, *J. Audio Eng. Soc.* vol. 55, n.º 6, pág. 503-516, 2007; y V. Pulkki y J. Herre, “Method and Apparatus for Conversion Between Multi-Channel Audio Formats”, *Solicitud de Patente de Estados Unidos Publicación N.º 2008/0232616 A1*).
- 20
- 25
- 30

Un elemento de realización de panorámica multi-canal puede utilizarse para colocar una fuente fantasma en algún sitio en la escena de audio. Los algoritmos mencionados por Eppolito, Pulkki y Blauert están basados en suposiciones relativamente sencillas que pueden provocar varias inexactitudes en la localización espacial donde una fuente se hizo panorámica y en dónde se percibe la fuente (A. Eppolito, “Multi-Channel Sound Panner”, *Solicitud de Patente de Estados Unidos Publicación N.º US 2012/0170758 A1*; V. Pulkki, “Virtual Sound Source Positioning Using Vector Base Amplitude Panning”, *J. Audio Eng. Soc.*, vol. 45, n.º 6, pág. 456-466, 1997; y J. Blauert, “Spatial hearing: The psychophysics of human sound localization”, 3rd. Ed. Cambridge and Mass: MIT Press, 2001, sección 2.2.2).

35

- 40 Los métodos de mezclado ascendente de extracción ambiental están diseñados para extraer las partes de señal ambiental y distribuirlas entre los altavoces adicionales para generar una cierta cantidad de envolvimiento (J. S. Usher y Benesty, “Enhancement of Spatial Sound Quality: A New Reverberation-Extraction Audio Upmixer”, *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, n.º 7, pág. 2141-2150, 2007; C. Faller, “Multiple-Loudspeaker Playback of Stereo Signals”, *J. Audio Eng. Soc.*, vol. 54, n.º 11, pág. 1051-1064, 2006; C. Avendano y J.-M. Jot, “Ambience extraction and synthesis from stereo signals for multi-channel audio up-mix”, en *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*, vol. 2, 2002, pág. II-1957 - II-1960; y R. Irwan y R.M. Aarts, “Two-to-Five Channel Sound Processing”, *J. Audio Eng. Soc.* vol. 50, n.º 11, pág. 914-926, 2002). La extracción está basada únicamente en uno o dos canales, lo cual es por lo que la escena de audio resultante ya no es una representación exacta de la escena original, y por lo que éstos no son enfoques útiles para nuestros fines. Esto también se cumple para los enfoques de formación de matrices como se describe por Dressler en “Dolby Surround Pro Logic II Decoder Principles of Operation” (disponible en línea, la dirección se indica más adelante). El enfoque de mezclado ascendente dos a tres mencionado por Vickers en la *Solicitud de Patente de Estados Unidos Publicación N.º US 2010/0296672 A1* “Mezclado ascendente de dos a tres canales para la derivatización de canal central” utiliza algún conocimiento previo respecto a la posición del tercer altavoz y la distribución resultante de la señal entre los otros dos altavoces, y por lo tanto carece de la capacidad para generar señales precisas para una posición arbitraria del altavoz insertado.
- 45
- 50
- 55

Las realizaciones de la presente invención están dirigidas a proporcionar un sistema que es capaz de preservar la escena de audio original en un entorno de reproducción, donde la configuración de altavoces se desvía del original, por el agrupamiento de los altavoces adecuados a segmentos, y aplicando un procesamiento de ajuste de mezclado ascendente, mezclado descendente y/o desplazamiento. Una etapa de post-procesamiento a un códec de audio regular, podría ser un posible escenario de aplicación. Tal caso se describe en la Figura 1, donde  $N$ ,  $\rho_s$ ,  $\vartheta_s$ ,  $\varphi_s$  y  $M$ ,  $\rho_s$ ,  $\vartheta_s$ ,  $\varphi_s$  son el número de altavoces y sus posiciones correspondientes en las coordenadas polares en la configuración de altavoces original y modificada/desplazada, respectivamente. En general, no obstante, el método

60

propuesto es aplicable a cualquier cadena de señales de audio como una herramienta post-procesamiento. En las realizaciones, los segmentos de la configuración de altavoces (configuración de altavoces original y/o de reproducción) representan cada uno un subgrupo de direcciones dentro de un plano bidimensional (2D) o dentro de un espacio tridimensional (3D). De acuerdo con las realizaciones, para una configuración de altavoces bidimensional (2D) planar, el intervalo de ángulo azimutal completo de interés, puede dividirse en múltiples segmentos (sectores) que cubren un intervalo reducido de ángulos azimutales. Análogamente, en el caso 3D el intervalo de ángulo sólido completo (azimutal y elevación) puede dividirse en segmentos que cubren un intervalo de ángulo más pequeño.

Cada segmento puede caracterizarse por una medición de la dirección asociada, que puede utilizarse para especificar o hacer referencia al segmento correspondiente. La medida direccional puede, por ejemplo, ser un vector que apunta hacia el centro del segmento, o un ángulo azimutal en el caso 2D, o un conjunto de un azimut y un ángulo de elevación en el caso 3D. El segmento puede hacerse referencia respecto a un subconjunto de direcciones dentro de un plano 2D o dentro de un espacio 3D. Para simplicidad de presentación, los siguientes ejemplos se describen ejemplarmente para el caso 2D; sin embargo, la expansión a las configuraciones 3D es sencilla.

La Figura 1 muestra un diagrama esquemático de bloques del posible escenario de aplicación anteriormente mencionado para un aparato y/o un método para ajustar una señal de audio espacial. Una señal de audio espacial 1 en el lado del codificador, se codifica por un codificador 10. La señal de audio espacial en el lado del codificador, tiene N canales y se ha producido para una configuración de altavoces original, por ejemplo, una configuración de altavoces 5.0 o una configuración de altavoces 5.1 con posiciones de altavoz a 0 grados, +/- 30 grados, y +/- 110 grados con respecto a una orientación de un oyente. El codificador 10 produce una señal de audio codificada que puede transmitirse o almacenarse. Típicamente, la señal de audio codificada se ha comprimido en comparación con la señal de audio espacial 1 en el lado del codificador, para relajar los requisitos para el almacenamiento y/o la transmisión. Un descodificador 20 se proporciona para descodificar y en particular descomprimir la señal de audio espacial codificada. El descodificador 20 produce una señal de audio espacial descodificada 2 que es altamente similar o incluso idéntica a la señal de audio espacial 1 en el lado del codificador. En este punto en el procesamiento de la señal de audio espacial, puede emplearse un método o un aparato 100 para ajustar una señal de audio espacial. El propósito del método o el aparato 100 es ajustar la señal de audio espacial 2 a una configuración de altavoces de reproductor que difiere de la configuración de altavoces original. El método o el aparato proporcionan una señal de audio espacial ajustada 3 o 4, que está diseñada a la medida para la configuración de altavoces de reproducción disponible.

Una vista general del sistema del método propuesto se representa en la Figura 2. Una representación de dominio de frecuencia de tiempo corto de los canales de entrada, se agrupan en K segmentos por un agrupador 110 (elemento de agrupamiento) y se alimentan en una etapa de Descomposición Directa/Ambiental 130 y de Estimación de DOA 140, donde A es el ambiente y D son las señales directas por altavoz y segmento y  $\theta$ ,  $\varphi$  son las DOA estimadas por segmento. Estas señales se alimentan dentro de un representador ambiental 170 o un representador de sonido directo 150, respectivamente, dando como resultado las señales  $\hat{A}$  y  $\hat{D}$  directa y ambiental recién representadas, por altavoz y segmento para la configuración de salida. Las señales de segmento se combinan por un combinador 180 en señales de salida angularmente corregidas. Para compensar los desplazamientos en la configuración de salida con respecto a la distancia, los canales se escalan y retardan en una etapa de ajuste de distancia 190 para dar como resultado finalmente los canales de altavoz de configuración de reproducción. Dicho método también puede extenderse para manejar configuraciones de reproducción con un número incrementado así como disminuido de altavoces, y se describe más adelante.

En una primera etapa, el método del aparato agrupa las señales de altavoz vecinas, adecuadas a K segmentos, mientras que cada señal de altavoz puede contribuir a varios segmentos y cada segmento consiste en al menos dos señales de altavoz. En una configuración de altavoces como la descrita en la Figura 3, los segmentos de configuración de entrada, por ejemplo, podrían formarse por los pares de altavoz  $Seg_{in} = \{L_1, L_2\}, \{L_2, L_3\}, \{L_3, L_4\}, \{L_4, L_5\}, \{L_5, L_1\}$  y los segmentos de salida podrían ser  $Seg_{out} = \{L_1, L'_2\}, \{L'_2, L_3\}, \{L_3, L_4\}, \{L_4, L_5\}, \{L_5, L_1\}$ . El altavoz  $L_2$  en la configuración de altavoces original (altavoz dibujado en línea discontinua) se modificó a un altavoz movido o desplazado  $L'_2$  en la configuración de altavoces de reproducción.

Durante el análisis, se lleva a cabo una Descomposición Directa/Ambiental basada en correlación cruzada, normalizada, por segmento, dando como resultado los componentes D de señal directa y los componentes A de señal ambiental para cada altavoz (para cada canal) con respecto a cada segmento considerado. Esto significa que el método/aparato propuesto es capaz de estimar las señales directa y ambiental para una fuente diferente dentro de cada segmento. La Descomposición Directa/Ambiental no está restringida al enfoque basado en la correlación cruzada, normalizada, mencionado, sino que puede llevarse a cabo con cualquier algoritmo de descomposición adecuado. El número de señales directa y ambiental generadas por segmento, va desde al menos una hasta el número de altavoces contribuyentes al segmento considerado. Por ejemplo, para la configuración de entrada dada en la Figura 3, existe al menos una señal directa y una señal ambiental o como máximo dos señales directas y dos señales ambientales por segmento.

Además, ya que una señal de altavoz particular está contribuyendo a varios segmentos durante la Descomposición

Directa/Ambiental, las señales pueden disminuirse en escala o dividirse antes de entrar en la Descomposición Directa/Ambiental. La manera más fácil de hacer eso sería una disminución de escala de cada señal de altavoz dentro de cada segmento, por el número de segmentos a los cuales contribuye ese altavoz particular. Por ejemplo, para el caso en la Figura 3, cada canal de altavoz contribuye a dos segmentos, de modo que el factor de  
 5 disminución de escala podría ser 1/2 para cada uno de los canales de altavoz. Pero en general, es también posible una división más sofisticada y no balanceada.

Una etapa de estimación de dirección de llegada (etapa de estimación DOA) 140 puede anexarse a la Descomposición Directa/Ambiental 130. Las DOA que consisten en un ángulo azimutal <sup>9</sup> y posiblemente un ángulo  
 10 de elevación  $\varphi$ , se estiman por segmento y banda de frecuencia y de acuerdo con el método de Descomposición Directa/Ambiental elegido. Por ejemplo, si se utiliza el método de descomposición de correlación cruzada, normalizada, la estimación DOA utiliza las consideraciones de energía de las señales de sonido directas de entrada y extraída para la estimación. En general, no obstante, puede elegirse entre las diversas Descomposiciones Directa/Ambiental y los algoritmos de detección de posición.

En la etapa de representación 170, 150 (Representador de Sonido Ambiental y Directo) tiene lugar la conversión real entre la configuración de altavoces de entrada y salida, con las señales directa y ambiental que se tratan de manera separada y diferente. Cualquier modificación de la configuración de entrada puede describirse como una combinación de tres casos básicos: inserción, eliminación y desplazamiento de los altavoces. Por razones de  
 20 simplicidad, estos casos se describen individualmente pero en un escenario del mundo real éstos podrían ocurrir simultáneamente y, por lo tanto, se tratan también de manera simultánea. Esto se lleva a cabo mediante la superposición de los casos básicos. La inserción y eliminación de altavoces afecta únicamente los segmentos considerados y tiene que observarse como una técnica de mezclado ascendente y descendente, basada en segmento. Durante la representación, las señales directas pueden alimentarse en una función de re-expansión panorámica, que asegura una localización correcta de las fuentes fantasma en la configuración de salida. Para  
 25 hacerlo así, las señales pueden "expandirse panorámicamente de manera inversa" con respecto a la configuración de entrada y expandirse panorámicamente de nuevo con respecto a la configuración de salida. Esto puede lograrse mediante la aplicación de coeficientes de re-expansión panorámica a las señales directas dentro de un segmento. Una implementación posible, por ejemplo, para el caso de desplazamiento, del coeficiente de re-expansión  
 30 panorámica  $c_{D,k}^s$  podría ser como sigue:

$$c_{D,k}^s = \frac{h_k^s + \epsilon}{g_k^s + \epsilon}, \quad (1)$$

donde  $g_k^s$  son las ganancias de expansión panorámica en la configuración de entrada (derivadas de las DOA  
 35 estimadas) y  $h_k^s$  son las ganancias de expansión panorámica para la configuración de salida,  $k = 1 \dots K$  indica el segmento considerado y  $s = 1 \dots S$  el altavoz considerado dentro del segmento.  $\epsilon$  es una constante de regularización pequeña. Esto produce las señales directas re-expandidas panorámicamente:

$$\hat{D}_k^s = c_{D,k}^s \cdot D_k^s. \quad (2)$$

40 En cualquier segmento en el cual los altavoces contribuyentes coincidan en la configuración de entrada y salida, esto da como resultado una multiplicación por 1, y deja sin cambio los componentes directos extraídos.

Se aplica también un coeficiente de corrección a las señales ambientales, que depende en general de cuánto hayan  
 45 cambiado los tamaños del segmento. El coeficiente de corrección podría implementarse como sigue:

$$c_{A,k} = \sqrt{\frac{\angle \text{Seg}_{out}[k]}{\angle \text{Seg}_{in}[k]}}. \quad (3)$$

50 donde  $\angle \text{Seg}_{in}[k]$  y  $\angle \text{Seg}_{out}[k]$  indican el ángulo entre las posiciones del altavoz dentro del segmento  $k$  en la configuración de entrada (configuración de altavoces original) o la configuración de salida (configuración de altavoces de reproducción) respectivamente. Esto produce las señales ambientales corregidas:

$$\hat{A}_k^s = c_{A,k} \cdot A_k^s \quad (4)$$

Como las señales directas, en cualquier segmento en el cual los altavoces contribuyentes coincidan en la configuración de entrada y salida, las señales ambientales se multiplican por uno y se dejan sin cambio. Este comportamiento de suministro directo y ambiental garantiza un procesamiento de preservación de la forma de onda de un canal de altavoz particular, si ninguno de los segmentos a los cuales contribuye el canal de altavoz, sufre de cambios. Además, el procesamiento converge suavemente a la solución de preservación de la forma de onda si las posiciones del altavoz de los segmentos se mueven progresivamente hacia las posiciones de la configuración de entrada.

La Figura 4 visualiza un escenario donde un altavoz ( $L_6$ ) se agregó a una configuración de altavoz convencional 5.1, es decir, un número incrementado de altavoces. La adición de un altavoz puede dar como resultado uno o más de los siguientes efectos: la estabilidad fuera del punto dulce de la escena de audio puede mejorarse, es decir, una estabilidad aumentada de la escena de audio espacial percibida si un oyente se mueve fuera del punto de audición ideal (denominado punto dulce). El involucramiento del oyente puede mejorarse y/o la localización espacial puede mejorarse, por ejemplo si una fuente fantasma se reemplaza por un altavoz real. En la Figura 4, S indica una posición de fuente fantasma estimada en el segmento formado por los altavoces  $L_2$  y  $L_3$ . La posición de la fuente fantasma estimada puede determinarse basándose en la descomposición directa/ambiental llevada a cabo por el descomponedor directo/ambiental 130 y la estimación de dirección de llegada para una o más fuentes fantasma dentro del segmento. Para el altavoz agregado tiene que crearse una señal directa y ambiental y las señales directa y ambiental de los altavoces vecinos tienen que han de ajustarse. Esto da como resultado de manera eficaz un mezclado ascendente para el segmento actual con un manejo de señal como sigue:

Señales directas: en la configuración de altavoces de reproducción (configuración de salida) con el altavoz adicional  $L_6$ , la fuente fantasma S se asigna al segmento  $\{L_2, L_6\}$  en la configuración de altavoces de reproducción. Por lo tanto, las partes de la señal directa que corresponden a S en el altavoz original o el canal  $L_3$ , tienen que reasignarse y recolocarse al altavoz adicional  $L_6$ , y procesarse por una función de re-expansión panorámica, lo cual asegura que la posición percibida de S permanezca igual en la configuración de altavoces de reproducción. La reasignación incluye la eliminación de las señales reasignadas provenientes de  $L_3$ . Las partes directas de S en  $L_2$  tienen que procesarse también por la re-expansión panorámica.

Señales ambientales: la señal ambiental para  $L_6$  se genera a partir de las partes de la señal ambiental en  $L_2$ , y  $L_3$  y se pasan a un descorrelador para asegurar una percepción ambiental de las señales generas. Las energías de las señales ambientales en  $L_2$ ,  $L_6$  y  $L_3$  (cada altavoz de los segmentos de configuración de salida recién formados  $\{L_2, L_6\}$  y  $\{L_6, L_3\}$ ) se ajustan de acuerdo con un Esquema de Remapeo de Energía Ambiental, seleccionable, que a continuación se denomina como AERS. Partes de estos esquemas es un esquema de Energía Ambiental Constante (CAE), donde la energía ambiental general se mantiene constante, y un esquema de Densidad Ambiental Constante (CAD), donde la densidad de energía ambiental dentro de un segmento se mantiene constante (por ejemplo, la densidad de energía ambiental dentro de los nuevos segmentos  $\{L_2, L_6\}$  y  $\{L_6, L_3\}$ ) debe ser la misma que en el segmento original  $\{L_2, L_3\}$ ). Estos esquemas se abrevian a continuación CAE y CAD, respectivamente.

Si S se coloca en el segmento de reproducción  $\{L_6, L_3\}$ , el procesamiento de las señales ambientales sigue las mismas reglas y se lleva a cabo de manera análoga.

Como se ilustra en la Figura 4, la configuración de altavoces de reproducción comprende un altavoz adicional  $L_6$  dentro del segmento original  $\{L_2, L_3\}$  de modo que el segmento original de la configuración de altavoces original corresponde a dos segmentos  $\{L_2, L_6\}$  y  $\{L_6, L_3\}$  de la configuración de altavoces de reproducción. En general, el segmento original puede corresponder a dos o más segmentos de los segmentos de reproducción, es decir, el altavoz adicional subdivide el segmento original en dos o más segmentos. El representador de sonido directo está, en este escenario, configurado para generar los componentes de sonido directo ajustados para los al menos dos altavoces  $L_2$ ,  $L_3$  y para el altavoz adicional  $L_6$  de la configuración de altavoces de reproducción.

La Figura 5 ilustra esquemáticamente una situación de un número disminuido de altavoces en la configuración de altavoces de reproducción en comparación con la configuración de altavoces original. En la Figura 5, se representa un escenario donde un altavoz ( $L_2$ ) se eliminó de una configuración de altavoces convencional 5.1.  $S_1$  y  $S_2$  representan las posiciones de la fuente fantasma, estimadas, por banda de frecuencia en los segmentos de la configuración de entrada  $\{L_1, L_2\}$  y  $\{L_2, L_3\}$ , respectivamente. El manejo de señales, descrito más adelante, da como resultado efectivamente un mezclado descendente de los dos segmentos  $\{L_1, L_2\}$  y  $\{L_2, L_3\}$  a un nuevo segmento  $\{L_1, L_3\}$ .

Señales directas: las partes de señales directas de  $L_2$  tienen que reasignarse a  $L_1$  y  $L_3$ , y unirse, de manera que

las posiciones de la fuente fantasma percibidas  $S_1$  y  $S_2$  no cambian. Esto se realiza por la reasignación de las partes directas de  $S_1$  en  $L_2$  a  $L_3$  y las partes directas de  $S_2$  en  $L_2$  a  $L_1$ . Las señales correspondientes de  $S_1$  y  $S_2$  en  $L_1$  y  $L_3$  se procesan por una función de re-expansión panorámica, que asegura la percepción correcta de las posiciones de la fuente fantasma en la configuración de altavoces de reproducción. La unión se lleva a cabo por una superposición de las señales correspondientes.

Señales ambientales: Las señales ambientales que corresponden a los segmentos  $\{L_1, L_2\}$  y  $\{L_2, L_3\}$  ambas localizadas en  $L_2$  se reasignan a  $L_1$  y  $L_3$ , respectivamente. Nuevamente, las señales reasignadas se escalan de acuerdo con uno de los Esquemas de Remapeo de Energía Ambiental (AER) introducidos, y se unen con las señales ambientales originales en  $L_1$  y  $L_3$ .

Como se ilustra en la Figura 5, la configuración de altavoces de reproducción carece del altavoz  $L_2$  en comparación con la configuración de altavoces original, de modo que el segmento  $\{L_1, L_2\}$  y un segmento vecino  $\{L_2, L_3\}$  se unen a un segmento unido de la configuración de altavoces de reproducción. En general y en particular en una configuración de altavoces tridimensional, la eliminación de un altavoz puede dar como resultado que varios segmentos originales se unan a un segmento de reproducción.

Las Figuras 6A y 6B ilustran esquemáticamente dos situaciones de los altavoces desplazados. En particular, el altavoz  $L_2$  en la configuración de altavoces original se movió a una nueva posición y se denomina como el altavoz  $L'_2$  en la configuración de altavoces de reproducción. Un procesamiento propuesto para el caso de un altavoz desplazado es como sigue.

Dos ejemplos para posibles escenarios de desplazamiento de altavoz se describen en las Figuras 6A y 6B, donde en la Figura 6A solo tiene lugar un reajuste de tamaño del segmento, y no se vuelve necesaria la reasignación de una fuente fantasma, mientras que en la Figura 6B el altavoz desplazado  $L'_2$  se mueve más allá de la posición estimada (dirección) de la fuente fantasma  $S_2$  y, por lo tanto, la fuente necesita reasignarse y unirse al segmento de salida ( $L_1, L'_2$ ). El altavoz original  $L_2$  y su dirección desde la perspectiva del oyente se dibujan en líneas discontinuas en las Figuras 6A y 6B.

En el caso esquemáticamente ilustrado en la Figura 6A, las señales directas se procesan como sigue. Como se estableció anteriormente, no es necesaria una reasignación. De este modo, el procesamiento está confinado a pasar el componente de señal directa de  $S_1$  y  $S_2$  en los altavoces  $L_1, L_2$  y  $L_3$ , respectivamente, a la función de re-expansión panorámica, que ajusta las señales de manera que las fuentes fantasma se perciben en su posición original con el altavoz desplazado  $L'_2$ .

Las señales ambientales en el caso mostrado en la Figura 6A se procesan como sigue. Ya que no existe tampoco necesidad para reasignaciones de señales, las señales ambientales en los segmentos correspondientes y los altavoces simplemente se ajustan de acuerdo con uno de los AER.

Con respecto a la Figura 6B el procesamiento de las señales directas se describe ahora. Si un altavoz se mueve más allá de una posición de fuente fantasma, se vuelve necesario reasignar esta fuente a un segmento de salida diferente. Aquí, la señal fuente correspondiente de  $S_2$  tiene que reasignarse al segmento de salida  $\{L_1, L'_2\}$  y procesarse por la función de re-expansión panorámica para asegurar una percepción de posición de fuente igual. Adicionalmente, las señales fuente correspondientes de  $S_2$   $\{L_1, L_2\}$  tienen que re-expandirse panorámicamente para coincidir con el nuevo segmento de salida  $\{L_1, L'_2\}$  y ambas nuevas partes de señal fuente en cada altavoz  $L_1$  y  $L'_2$ , han de unirse.

Por lo tanto, el representador de sonido directo está configurado para reasignar un componente de sonido directo que tiene una dirección determinada de llegada  $S_2$  desde el segmento  $\{L_2, L_3\}$  en la configuración de altavoces original a un segmento vecino  $\{L_1, L'_2\}$  en la configuración de altavoz de reproducción si un límite entre el segmento y el segmento vecino traspasa la dirección determinada de llegada  $S_2$  cuando se pasa de la configuración de altavoces original a la configuración de altavoces de reproducción. Además, el representador de sonido directo puede configurarse para reasignar el componente de sonido directo que tiene la dirección determinada de llegada desde al menos un altavoz del segmento original  $\{L_2, L_3\}$  a al menos un altavoz en el segmento vecino en la configuración de salida  $\{L_1, L'_2\}$ . En particular, el representador directo puede configurarse para reasignar el componente directo de  $S_2$  en  $L_3$  asignado al segmento  $\{L_2, L_3\}$  en la configuración de entrada al altavoz desplazado  $L'_2$  asignado al segmento  $\{L_1, L'_2\}$  en la configuración de reproducción, y para reasignar el componente directo de  $S_2$  en  $L_2$  asignado al segmento  $\{L_2, L_3\}$  en la configuración de entrada a  $L_1$  asignado al segmento  $\{L_1, L'_2\}$  en la configuración de reproducción. Obsérvese que la acción de reasignación puede también involucrar un ajuste del componente de sonido directo, por ejemplo, por realizando una re-expansión panorámica con respecto a una amplitud relativa y/o un retardo relativo de las señales del altavoz.

Para las señales ambientales en la Figura 6B, puede realizarse un procesamiento similar: las señales ambientales en el segmento  $\{L_2, L_3\}$  se ajustan mediante el uso de los AER. Para desplazamientos grandes, adicionalmente, una

parte de estas señales ambientales puede agregarse al segmento  $\{L_1, L'_2\}$  y ajustado por un AERS.

Dentro de la etapa de combinación 180 (Figura 2), se forman las señales de altavoz reales para la configuración de altavoces de reproducción (configuración de salida). Esto se realiza mediante la adición de las señales directa y ambiental remapeada y representada, correspondientes, del segmento izquierdo y derecho respectivo con respecto al altavoz en medio. (Los términos altavoz “izquierdo” y “derecho” se mantienen para el caso bidimensional, es decir, todos los altavoces están en el mismo plano, típicamente un plano horizontal). En la salida de la etapa de combinación 180, se emiten las señales para la escena de audio original, pero ahora representadas para una nueva configuración de altavoces (la configuración de altavoces de reproducción) con los M altavoces en las posiciones  $\hat{g}_s$  y  $\varphi_s$ .

En este punto, es decir en la salida del combinador o de la etapa de combinación 180, el nuevo sistema proporciona señales de altavoz donde todas las modificaciones con respecto al azimut y al ángulo de elevación de los altavoces en la configuración de salida se han corregido. Si un altavoz en la configuración de salida se movió de manera que su distancia al punto de audición ha cambiado a una nueva distancia  $\hat{r}_s$ , la etapa de ajuste de distancia opcional 190 puede aplicar un factor de corrección y un retardo a ese canal para compensar el cambio de la distancia. La salida 4 de esta etapa da como resultado los canales de altavoz de la configuración de reproducción real.

Otra realización más puede utilizar la invención para implementar un punto dulce de movimiento de la configuración de altavoces de reproducción. Para esto, en un primer paso, el algoritmo o el aparato tiene que determinar la posición del oyente. Esto puede realizarse fácilmente por el uso de una técnica/dispositivo de rastreo para determinar la posición actual del oyente. Posteriormente, el aparato recalcula las posiciones de los altavoces con respecto a la posición del oyente, que significa un nuevo sistema de coordenadas con el oyente en el origen. Esto es el equivalente a tener un oyente fijo y altavoces móviles. El algoritmo calcula entonces las señales óptimamente para esta nueva configuración.

La Figura 7 muestra un diagrama esquemático de bloques de un aparato 100 para ajustar una señal de audio espacial 2 a una configuración de altavoces de reproducción de acuerdo con al menos una realización. El aparato 100 comprende un agrupador 110 configurado para agrupar al menos dos señales de canal 702 dentro de un segmento. El aparato 100 comprende además un descomponedor directo-ambiental 130 configurado para descomponer al menos dos señales de canal 702 en el segmento a al menos un componente de sonido directo 732 y a al menos un componente ambiental 734. El descomponedor directo-ambiental 130 puede comprender opcionalmente un estimador de dirección de llegada 140, configurado para estimar las DOA de al menos un componente de sonido directo 732. Como una alternativa, las DOA pueden proporcionarse a partir de una estimación de DOA externa o como meta información/información secundaria que acompaña la señal de audio espacial 2.

Un representador de sonido directo 150 está configurado para recibir una información de la configuración de altavoces de reproducción para al menos un segmento de reproducción asociado con el segmento y para ajustar al menos un componente de sonido directo 732 utilizando la información de la configuración de altavoces de reproducción para el segmento, de modo que una dirección percibida de llegada de al menos un componente de sonido directo en la configuración de altavoces de reproducción es sustancialmente idéntica a la dirección de llegada del segmento. Al menos la representación realizada por el representador de sonido directo 150 da como resultado que la dirección percibida de llegada sea más cercana a la dirección de llegada de al menos un componente de sonido directo en comparación con una situación en la cual no ha tenido lugar el ajuste. En un inserto en la Figura 7, se ilustra esquemáticamente un segmento original de la configuración de altavoces original y un segmento de reproducción correspondiente de la configuración de altavoces de reproducción. Típicamente, la configuración de altavoces original es conocida o normalizada de modo que la información acerca de la configuración de altavoces original no necesariamente tiene que proporcionarse al representador de sonido directo 150, sino que el representador de sonido directo tiene esta información ya disponible. No obstante, el representador de sonido directo puede configurarse para recibir la información de la configuración de altavoces original. De esta manera, el representador de sonido directo 150 puede configurarse para soportar señales de audio espaciales como entrada que se han registrado o creado para diferentes configuraciones de altavoz originales, tales como las configuraciones 5.1, 7.1, 10.2 o incluso 22.2.

El aparato 100 comprende además un combinador 180 configurado para combinar los componentes de sonido directos 752, ajustados, y los componentes ambientales 734 o los componentes ambientales modificados para obtener las señales de altavoz para al menos dos altavoces de la configuración de altavoces de reproducción. Las señales de altavoz para al menos dos altavoces de la configuración de altavoces de reproducción son parte de la señal de audio espacial ajustada 3 que puede enviarse por el aparato 100. Como se mencionó anteriormente, puede realizarse un ajuste de distancia sobre la señal de audio espacial ajustada por DOA, para obtener la señal de audio espacial 4 ajustada por DOA y por distancia (véase la Figura 2). El combinador 180 puede configurarse también para combinar el componente de sonido directo ajustado 752 y el componente ambiental 734 con los componentes de

sonido directo y/o ambiental provenientes de uno o más segmentos vecinos que comparten el altavoz con el segmento contemplado.

5 La Figura 8 muestra un diagrama de flujo esquemático de un método para ajustar una señal de audio espacial a una configuración de altavoces de reproducción que difiere de una configuración de altavoces original destinada para presentar el contenido de audio transferido por la señal de audio espacial. El método comprende una etapa 802 de agrupar al menos dos señales de canal dentro de un segmento. El segmento es típicamente uno de los segmentos de la configuración de altavoces original. Las al menos dos señales de canal en el segmento se descomponen en componentes de sonido directo y componentes ambientales durante una etapa 804. El método comprende además 10 una etapa 806 para determinar una dirección de llegada de los componentes de sonido directo. Los componentes de sonido directo se ajustan en una etapa 808 utilizando una información de configuración de los altavoces de reproducción para el segmento, de modo que una dirección percibida de llegada de los componentes de sonido directo en la configuración de altavoces de reproducción es idéntica a la dirección de llegada del segmento, o más cercana a la dirección de llegada del segmento en comparación con una situación en la cual no ha tenido lugar el 15 ajuste. El método también comprende una etapa 809 para combinar los componentes de sonido directo ajustados y los componentes ambientales, o los componentes ambientales modificados para obtener las señales de altavoz para al menos dos altavoces de la configuración de altavoces de reproducción.

20 El ajuste propuesto de una señal de audio espacial a una configuración de altavoces de reproducción encontrada, puede estar relacionado con uno o más de los siguientes aspectos:

- Agrupar los canales de altavoz vecinos de la configuración original en segmentos
- La Descomposición Directa/Ambiental basada en segmentos
- Varias Descomposiciones Directas/Ambientales diferentes y algoritmos de extracción de posición, seleccionables
- 25 - El remapeo de los componentes directos de manera que la dirección percibida permanece sustancialmente igual
- El remapeo de los componentes ambientales de manera que el involucramiento percibido permanece sustancialmente igual
- Corrección de distancia del altavoz mediante la aplicación de un factor de escala y/o un retardo
- Varios algoritmos de expansión panorámica seleccionables
- 30 - Remapeo independiente de los componentes directo y ambiental
- Procesamiento selectivo por tiempo y frecuencia
- Procesamiento de preservación de forma de onda general para todos los canales de altavoz si la configuración de salida coincide con la configuración de entrada
- Preservación de la forma de onda en forma de canal para cada altavoz, donde los segmentos a los cuales 35 contribuye el altavoz son se modifican con respecto a la configuración de entrada y salida
- Casos Especiales:
  - "Expansión panorámica inversa" y expansión panorámica de una escena de entrada dada con un algoritmo 40 de expansión panorámica diferente
  - Por segmento, al menos una señal directa y ambiental

45 En segmentos que consisten en dos altavoces: dos señales directas y dos señales ambientales máximo. El número de señales directas y ambientales utilizadas es independiente uno del otro, pero depende de la calidad del objetivo espacial pretendida de las señales directa y ambiental representadas.

- Mezclado Descendente/Ascendente basado en segmentos
- El remapeo ambiental se realiza de acuerdo con los Esquemas de Remapeo de Energía Ambiental (AERS) que 50 comprenden:
  - ° Energía ambiental constante
  - ° Densidad ambiental constante (angular)

55 Al menos algunas realizaciones de la presente invención están configuradas para realizar una conversión de escena de sonido flexible, basada en el canal, que comprende una descomposición de los canales de altavoz originales en partes de señales directas y ambientales de una fuente (fantasma) dentro de y de acuerdo con cada segmento previamente construido. Las direcciones de llegada (DOA) de cada fuente directa se estiman y alimentan, junto con las señales directa y ambiental, dentro de un representador y ajustador de distancia, donde - de acuerdo con la configuración de altavoces de reproducción y a las DOA - las señales del altavoz original se modifican para 60 preservar la escena de audio real. El método y el aparato propuestos funcionan para preservar la forma de onda y son incluso capaces de manejar configuraciones de salida con un número incrementado o disminuido de canales de altavoz que los que están disponibles en la configuración de entrada.

Aunque la presente invención se ha descrito en el contexto de diagramas de bloques donde los bloques representan

los componentes de hardware reales o lógicos, la presente invención puede implementarse también por un método implementado por ordenador. En el último caso, los bloques representan las etapas del método correspondientes donde estas etapas significan las funcionalidades realizadas por los bloques de hardware lógicos o físicos correspondientes.

5 Las realizaciones descritas son meramente ilustrativas para los principios de la presente invención. Se debe entender que modificaciones y variaciones de las disposiciones y los detalles descritos en el presente documento serán evidentes para los expertos en la materia. Por lo tanto, es la intención estar limitados únicamente por el alcance de las reivindicaciones de patente anexas y no por los detalles específicos presentados a manera de descripción y explicación de las realizaciones en el presente documento.

15 Aunque algunos aspectos se han descrito en el contexto de un aparato, es claro que esos aspectos también representan una descripción del método correspondiente, donde un bloque o dispositivo corresponde a una etapa del método o una característica de una etapa del método. Análogamente los aspectos descritos en el contexto de una etapa del método también representan una descripción de un bloque o elemento o característica correspondiente de un aparato correspondiente. Algunas o todas las etapas del método pueden ejecutarse por (o utilizando) un aparato de hardware como, por ejemplo, un microprocesador, un ordenador programable o un circuito electrónico. En algunas realizaciones, una o más de las etapas del método más importantes pueden ejecutarse por tal aparato.

20 Dependiendo de ciertos requisitos de la implementación, las realizaciones de la invención pueden implementarse en hardware o en software. La implementación puede realizarse utilizando un medio de almacenamiento digital, por ejemplo, un disco flexible, un DVD, un Blu-Ray, un CD, una ROM, una EPROM, una EEPROM o una memoria FLASH, que tienen la señal de control electrónicamente legible almacenada en las mismas, que coopera (o pueden cooperar) con un sistema informático programable de manera que el método respectivo se realice. Por lo tanto, el medio de almacenamiento digital puede ser legible por ordenador.

25 Algunas realizaciones de acuerdo con la invención comprenden un soporte de datos que tiene señales de control electrónicamente legibles, que pueden cooperar con un sistema informático programable, de manera que uno de los métodos descritos en el presente documento se lleve a cabo.

30 En general, las realizaciones de la presente invención pueden implementarse como un producto de programa informático con un código de programa, siendo operativo el código de programa para realizar uno de los métodos cuando el producto de programa informático se ejecuta en un ordenador. El código de programa puede almacenarse, por ejemplo, en un portador legible por máquina.

35 Otras realizaciones comprenden el programa informático para realizar uno de los métodos descritos en el presente documento, almacenados en un portador legible en máquina.

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

45 Una realización adicional del método de la invención es por lo tanto un soporte de datos (o un medio de almacenamiento digital, o un medio legible por ordenador) que comprende, grabado en el mismo, el programa informático para llevar a cabo uno de los métodos descritos en el presente documento. El soporte de datos, el medio de almacenamiento digital o el medio grabado son típicamente tangibles y/o no transitorios.

50 Una realización adicional del método de la invención es por lo tanto un flujo de datos o una secuencia de señales que representan el programa informático para llevar a cabo uno de los métodos descritos en el presente documento. El flujo de datos o la secuencia de señales puede, por ejemplo, configurarse para transferirse mediante una conexión de comunicación de datos, por ejemplo mediante Internet.

55 Una realización adicional comprende un medio de procesamiento, por ejemplo un ordenador o un dispositivo lógico programable, configurado para o adaptado para llevar a cabo uno de los métodos descritos en el presente documento.

60 Una realización adicional comprende un ordenador que tiene instalado en el mismo el programa informático para llevar a cabo uno de los métodos descritos en el presente documento.

Una realización adicional de acuerdo con la invención comprende un aparato o un sistema configurado para transferir (por ejemplo electrónicamente u ópticamente) un programa informático para llevar a cabo uno de los métodos descritos en el presente documento hacia un receptor. El receptor puede, por ejemplo, ser un ordenador, un dispositivo móvil, un dispositivo de memoria o similar. El aparato o el sistema, pueden, por ejemplo, comprender un

servidor de archivos para transferir el programa informático al receptor.

En algunas realizaciones, un dispositivo lógico programable (por ejemplo, un campo de matrices de puertas programables) puede utilizarse para realizar algunas o todas las funcionalidades de los métodos descritos en el presente documento. En algunas realizaciones, un campo de matrices de puertas programables puede operar con un microprocesador para llevar a cabo uno de los métodos descritos en el presente documento. En general, los métodos se realizan preferentemente por cualquier aparato de hardware.

Las realizaciones de la presente invención pueden estar basadas en técnicas para la Descomposición Directa-Ambiental. La descomposición directa-ambiental puede llevarse a cabo ya sea basándose en un modelo de señales o en un modelo físico.

La idea detrás de una descomposición directa-ambiental basada en un modelo de señales es una suposición de que un sonido percibido y localizable directo, consiste ya sea de una sola o más señales coherentes o correlacionadas. Mientras que el sonido del ambiente, de este modo no localizable corresponde a las partes de señales no correlacionadas. La transición entre directo y ambiental es sin unión y depende de la correlación entre las señales. Información adicional acerca de la descomposición directa/ambiental puede encontrarse en: C. Faller, "Multiple-Loudspeaker Playback of Stereo Signals", J. Audio Eng. Soc. vol. 54, n.º 11, pág. 1051-1064, 2006, en J. S. Usher y J. Benesty, "Enhancement of Spatial Sound Quality: A New Reverberation-Extraction Audio Upmixer", IEEE Transactions on Audio, Speech, and Language Processing, vol. 15, n.º 7, pág. 2141-2150, 2007; y en M. Goodwin y J.-M. Jot, "Primary-Ambient Signal Decomposition and Vector-Based Localization for Spatial Audio Coding and Enhancement", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), vol. 1, 2007, pág. 1-9-1-12.

La Codificación de Audio Direccional (DirAC) es un método posible para descomponer las señales en energías de señales directas y difusas basándose en un modelo físico. En este punto, las propiedades del campo de sonido para la presión del sonido y la velocidad del sonido (partículas) en el punto de audición se capturan ya sea por una grabación de formato B real o virtual. Después de esto, con la suposición de que el campo de sonido consiste únicamente de una onda plana simple y el resto es energía difusa, la señal puede descomponerse en partes de señales directas y difusas. A partir de las partes directas, puede calcularse las denominadas Direcciones de Llegada (DOA). Con el conocimiento de las posiciones de altavoz reales, las partes de señales directas pueden re-expandirse panorámicamente usando las leyes de expansión panorámica especializadas (véase por ejemplo, V. Pulkki, "Virtual Sound Source Positioning Using Vector Base Amplitude Panning". J. Audio Eng. Soc, vol. 45, n.º 6, pág. 456-466, 1997) para preservar su posición global en la etapa de representación. Finalmente, las partes de señales ambientales descorrelacionadas y directas expandidas panorámicamente se combinan nuevamente, dando como resultado las señales del altavoz (como se describe en, por ejemplo, Pulkki, "Spatial Sound Reproduction with Directional Audio Coding", J. Audio Eng. Soc. vol. 55 n.º 6, pág. 503-516, 2008; o V. Pulkki y J. Herre, "Method and Apparatus for Conversion Between Multi-Channel Audio Formats", *Solicitud de Patente de Estados Unidos Publicación N.º US 2008/0232616 A1*, 2008).

Otro enfoque más se describe por J. Thompson, B. Smith, A. Warner y J.-M. Jot en "Direct-Diffuse Decomposition of multichannel Signals Using a System of Pairwise Correlations" (presentada en la 133ra Convención del AES 2012, octubre de 2012), donde las energías directa y difusa de una señal multicanal se estiman por un sistema de correlaciones por pares. El modelo de señales utilizado en este punto permite detectar una señal directa y difusa dentro de cada canal, incluyendo el desplazamiento de fase de señal directa a través de los canales. Una suposición de este enfoque es que las señales directas a través de todos los canales están correlacionados, es decir, todas representan la misma señal fuente. El procesamiento se lleva a cabo en el dominio de frecuencia y para cada banda de frecuencia.

Una posible implementación de la descomposición directa-difusa (o la descomposición directa-ambiental) se describe ahora en relación con las señales estereofónicas como un ejemplo. Otras técnicas para la descomposición directa-difusa son también posibles, y también las señales diferentes de las señales estereofónicas pueden someterse a descomposición directa-difusa. Típicamente, las señales estereofónicas se graban o mezclan de manera que cada fuente la señal va coherentemente hacia el canal de señales izquierdo y derecho con pistas direccionales específicas (diferencia de nivel, diferencia de tiempo) y las señales independientes reflejadas/reverberadas hacia los canales que determinan la anchura objetivo de auditoría y las pistas de envolvimiento del oyente. Las señales estereofónicas de una sola fuente pueden modelarse por una señal  $s$  que imita el sonido directo proveniente de una dirección determinada por un factor  $a$ , y por señales independientes  $n_1$  y  $n_2$  que corresponden a las reflexiones laterales. El par de señales estereofónicas  $x_1, x_2$  está relacionado a estas señales  $s, n_1, y n_2$  por las siguientes ecuaciones:

$$x_1(k) = s(k) + n_1(k)$$

$$x_2(k) = a s(k) + n_2(k),$$

en las que  $k$  es un índice de tiempo. En consecuencia, la señal de sonido directa  $s$  aparece en ambas señales estereofónicas  $x_1$  y  $x_2$ , sin embargo típicamente con amplitud diferente. La descomposición descrita puede llevarse a cabo en un número de bandas de frecuencia y adaptativamente en el tiempo para obtener una descomposición que no es solamente válida en un escenario de objeto de auditoría, sino también para las escenas sonoras no estacionarias con múltiples fuentes concurrentemente activas. En consecuencia, las ecuaciones anteriores pueden escribirse para un índice de tiempo particular  $k$  y una sub-banda de frecuencia particular  $m$  como:

$$x_{1,m}(k) = s_m(k) + n_{1,m}(k)$$

$$x_{2,m}(k) = A_b s_m(k) + n_{2,m}(k),$$

donde  $m$  es el índice de sub-banda,  $k$  es el índice de tiempo,  $A_b$  el factor de amplitud para la señal  $s_m$  para una cierta banda de parámetro  $b$  que puede comprender una o más sub-bandas de las señales de sub-banda. En cada pieza de tiempo-frecuencia con los índices  $m$  y  $k$  las señales  $s_m$ ,  $n_{1,m}$ ,  $n_{2,m}$  y el factor  $A_b$  se estiman independientemente. Una descomposición de sub-banda actualmente motivada puede utilizarse. Esta descomposición puede estar basada en la transformada rápida de Fourier, el banco de filtros de espejo en cuadratura u otro banco de filtros. Para cada banda de parámetros  $b$ , las señales  $s_m$ ,  $n_{1,m}$ ,  $n_{2,m}$  y  $A_b$  se estiman basándose en los segmentos con una cierta longitud temporal (por ejemplo, aproximadamente 20 ms). Dado el par de señales de sub-banda estereofónicas  $x_{1,m}$  y  $x_{2,m}$ , el objetivo es estimar  $s_m$ ,  $n_{1,m}$ ,  $n_{2,m}$  y  $A_b$  en cada banda de parámetros. Un análisis de las potencias y de la correlación cruzada del par de señales estereofónicas puede llevarse a cabo para este fin. La variable  $\rho_{x_1,b}$  indica una estimación de tiempo corto de la potencia de  $x_{1,m}$  en la banda de parámetros  $b$ . Las potencias de  $n_{1,m}$  y  $n_{2,m}$  puede suponerse que son las mismas, es decir se supone que la cantidad de sonido independiente secundario es la misma para las señales izquierda y derecha:  $p_{n1,b} = p_{1,m1b} = p_{n,b}$ .

La potencia ( $p_{x1,b}$ ,  $p_{x2,b}$ ) y la correlación cruzada normalizada  $\rho_{x1x2,b}$  para la banda de parámetros  $b$  pueden calcularse utilizando la representación de sub-banda de la señal estereofónica. Las variables  $A_b$ ,  $p_{s,b}$  y  $p_{n,b}$  se estiman posteriormente como una función de  $p_{x1,b}$ ,  $p_{x2,b}$  y  $\rho_{x1x2,b}$ , estimados. Tres ecuaciones relacionadas a las variables conocidas y desconocidas son:

$$p_{x_1,b} = p_{s,b} + p_{n,b}$$

$$p_{x_2,b} = A_b^2 p_{s,b} + p_{n,b}$$

$$\rho_{x_1x_2,b} = \frac{A_b p_{s,b}}{\sqrt{p_{x_1,b} p_{x_2,b}}}$$

Estas ecuaciones resueltas para  $A_b$ ,  $p_{s,b}$  y  $p_{n,b}$  producen:

$$A_b = \frac{B_b}{2C_b}$$

$$p_{s,b} = \frac{B_b}{2C_b^2}$$

$$p_{n,b} = \frac{B_b}{2C_b^2}$$

con

$$B_b = p_{x_2,b} - p_{x_1,b} + \sqrt{(p_{x_1,b} - p_{x_2,b})^2 + 4p_{x_1,b}p_{x_2,b}\rho_{x_1x_2,b}^2}$$

$$C_b = \rho_{x_1x_2,b}\sqrt{p_{x_1,b}p_{x_2,b}}$$

A continuación, se calculan las estimaciones de mínimos cuadrados de  $s_m$ ,  $n_{1,m}$ ,  $n_{2,m}$  como una función de  $A_b$ ,  $p_{s,b}$  y  $p_{n,b}$ . Para cada banda de parámetros  $b$  y cada trama de señal independiente, la señal  $s_m$  se estima como

$$\begin{aligned}\hat{s}_m(k) &= w_{1,b}x_{1,m}(k) + w_{2,b}x_{2,m}(k) \\ &= w_{1,b}(s_m(k) + n_{1,m}(k)) + w_{2,b}(A_b s_m(k) + n_{2,m}(k))\end{aligned}$$

5 donde  $w_{1,b}$  y  $w_{2,b}$  son ponderaciones de con valor real. Las ponderaciones  $w_{1,b}$  y  $w_{2,b}$  son óptimas en un sentido de mínimos cuadrados cuando una señal de error  $E$  es ortogonal a  $x_{1,m}$  y  $x_{2,m}$  en la banda de parámetros  $b$ . Las señales  $n_{1,m}$  y  $n_{2,m}$  pueden estimarse de una manera similar. Por ejemplo,  $n_{1,m}$  puede estimarse como

$$\begin{aligned}\hat{n}_{1,m}(k) &= w_{3,b}x_{1,m}(k) + w_{4,b}x_{2,m}(k) \\ &= w_{3,b}(s_m(k) + n_{1,m}(k)) + w_{4,b}(A_b s_m(k) + n_{2,m}(k))\end{aligned}$$

10 El escalamiento posterior puede realizarse entonces sobre las estimaciones de mínimos cuadrados iniciales  $\hat{s}_m$ ,  $\hat{n}_{1,m}$  y  $\hat{n}_{2,m}$  para coincidir la potencia de las estimaciones en cada banda de parámetros a  $p_{s,b}$  y  $p_{n,b}$ . Una descripción más detallada del método de mínimos cuadrados medios puede encontrarse en el capítulo 10.3 del libro de texto "Spatial Audio Processing" por J. Breebart y C. Faller. Uno o más de estos aspectos puede emplearse en conexión con o en el contexto del ajuste propuesto de una señal de audio espacial.

Las realizaciones de la presente invención pueden hacer referencia a o emplear uno o más Expansores Panorámicos Multi-Canal. Los Expansores Panorámicos Multi-Canal son herramientas que hacen posible que el ingeniero de sonido coloque una fuente virtual o fantasma dentro de una escena de audio artificial. Esto puede conseguirse de varias maneras. Después de una función de ganancia especializada o la ley de expansión panorámica, una fuente fantasma puede colocarse dentro de una escena de audio mediante la aplicación de una ponderación de amplitud o retardo o ambas a la señal fuente. Información adicional acerca de los Expansores Multi-Canal puede encontrarse en la *Solitud de Patente de Estados Unidos Publicación N.º* documento US 2012/0170758 A1 "Expansor de Sonido Multi-Canal" por A. Eppolito en V. Pulkki, "Virtual Sound Source Positioning Using Vector Base Amplitude Panning", J Audio Eng. Soc., vol. 45, n.º 6, pág. 456-466, 1997; y en J. Blauert, "Spatial hearing: The psychophysics of human sound localization", sección 2.2.2, 3ª edición. Cambridge and Mass: MIT Press, 2001. Por ejemplo, puede emplearse un expansor panorámico, que puede soportar un número arbitrario de canales de entrada y cambios a las configuraciones al espacio de sonido de salida. Por ejemplo, el expansor panorámico puede manejar los cambios sin interrupciones en el número de canales de entrada. También, el expansor panorámico puede soportar cambios al número y a las posiciones de los altavoces en el espacio de salida. El expansor panorámico puede permitir el control continuo de la atenuación y el colapso. El expansor panorámico puede mantener los canales fuente sobre la periferia del espacio de sonido cuando se colapsan los canales. El expansor panorámico puede permitir el control sobre la trayectoria por la cual colapsan las fuentes. Estos aspectos pueden conseguirse por un método que comprende la recepción de la entrada que requiere el re-balanceo de una pluralidad de canales del audio fuente en un espacio de sonido que tiene una pluralidad de altavoces, en el que la pluralidad de canales del audio fuente se describen inicialmente por una posición inicial en el espacio de sonido y una amplitud inicial, y en el que las posiciones y las amplitudes de los canales definen un balance de los canales en el espacio de sonido. Basándose en la entrada, se determina una nueva posición del espacio de sonido para al menos uno de los canales fuente. Basándose en la entrada, se determina una modificación en la amplitud de al menos uno de los canales fuente, en el que la nueva posición y la modificación a la amplitud consigue el re-balance. En respuesta a la determinación de que la entrada indica que un altavoz particular de la pluralidad de altavoces va a deshabilitarse, el sonido que se iba a originar del altavoz particular puede transferirse automáticamente a otros altavoces adyacentes al altavoz particular. El método se realiza por uno o más dispositivos informáticos. Uno o más de estos aspectos puede emplearse en conjunto con o en el contexto del ajuste propuesto de una señal de audio espacial.

Algunas realizaciones de la presente invención pueden relacionarse a o emplear conceptos para cambiar las escenas de audio existentes. Un sistema para componer o incluso cambiar una escena de audio existente se presentó por IOSONO (como se describe en la Solicitud de Patente Alemana N.º DE 10 2010 030 534 A1, "Vorrichtung zum Verändern einer Audio-Szene und Vorrichtung zum Erzeugen einer Richtungsfunktion"). Utiliza una representación de la fuente basada en el objeto más metadatos adicionales, combinados con una función direccional para colocar la fuente dentro de la escena de audio. Si una escena de audio ya existente, sin el objeto de audio y los meta datos, se alimenta dentro de este sistema, los objetos de audio, las direcciones y las funciones direccionales tienen que determinarse en primer lugar a partir de esa escena de audio. Uno o más de estos aspectos puede emplearse en conexión con o en el contexto del ajuste propuesto de una señal de audio espacial.

Algunas realizaciones de la presente invención pueden relacionarse con o emplear una Conversión de Canal y

Corrección de Posicionamiento. La mayoría de los sistemas que están dirigidos a corregir un posicionamiento de altavoz fallido o una desviación en los canales de reproducción intentan preservar las propiedades físicas del campo de sonido. Para un escenario de mezclado descendente, un posible enfoque podría ser modelar los altavoces omitidos como altavoces virtuales mediante la expansión panorámica y por este medio preservar la presión sonora y la velocidad de partículas en el punto de audición, (como se describe en A. Ando, "Conversion of Multi-channel Sound Signal Maintaining Physical Properties of Sound in Reproduced Sound Field", IEEE Transactions on Audio, Speech, and Language Processing, vol. 19, n.º 6, pág. 1467-1475, 2011). Otro método más podría ser calcular las señales del altavoz en la configuración objetivo para restaurar el campo sonoro original. Esto se realiza mediante la transición de las señales del altavoz original dentro de una representación de campo sonoro, y representando las nuevas señales de altavoz a partir de esta representación (como se describe en A. Laborie, R. Buno y S. Montoya, "Reproducing Multichannel Sound on any Speaker Layout", en 118th Convention of the AES, 2005).

De acuerdo con Ando, una conversión de una señal de sonido multi-canal es posible mediante la conversión de la señal del sistema de sonido multi-canal original en la de un sistema alternativo o con un número diferente de canales, mientras se mantienen las propiedades físicas del sonido en el punto de audición en el campo del sonido reproducido. Un problema de conversión de este tipo puede describirse por la ecuación lineal infradeterminada. Para obtener una solución analítica a la ecuación, el método divide el campo de sonido del sistema alternativo basándose en las posiciones de tres altavoces y resuelve la "solución local" en cada subcampo. Como resultado, el sistema alternativo localiza cada señal de canal del sistema de sonido original en la posición del altavoz correspondiente como una fuente fantasma. La composición de las soluciones locales introduce la "solución global" es decir, la solución analítica al problema de conversión. Los experimentos se realizaron con señales de 22 canales de un sistema de sonido multi-canal 22.2 sin los dos canales de efecto de baja frecuencia convertidos en señales de 10, de 8, y de 6 canales mediante el método. Las evaluaciones subjetivas mostraron que el método propuesto pudo reproducir la impresión espacial del sonido original de 22 canales con ocho altavoces. Uno o más de estos aspectos pueden emplearse en conexión con o en el contexto del ajuste propuesto de una señal de audio espacial.

La Codificación de la Escena de Audio Espacial (SASC) es un ejemplo para un sistema motivado no físico (M. Goodwin y J.-M. Jot, "Spatial Audio Scene Coding" en 125th Convention of the AES, 2008). Realiza un Análisis de Componente Principal (PCA) para descomponer las señales de entrada multi-canal en sus componentes primario y ambiental bajo ciertas restricciones de correlación inter-canal (M. Goodwin y J.-M. Jot, "Primary-Ambient Signal Decomposition and Vector-Based Localization for Spatial Audio Coding and Enhancement", en IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), vol. 1, 2007, pág. 1-9-1-12). El componente primario se identifica aquí como el vector propio de la matriz de correlación de canal de entrada con el valor propio más grande. Después de esto, se realiza un análisis de localización primaria y ambiental, donde se determina el vector de localización directo y ambiental. La representación de las señales de salida se realiza por la generación de una matriz de formato que contiene los vectores unitarios que apuntan hacia la dirección espacial de los canales de salida. Basándose en esa matriz de formato, se deriva un conjunto de ponderaciones nulas, de modo que el vector de ponderación está en el espacio nulo de la matriz de formato. Los componentes direccionales se generan mediante expansión panorámica por pares entre estos vectores y los componentes no direccionales se generan mediante el uso del conjunto completo de vectores en la matriz de formato. Las señales de salida finales se generan mediante la interpolación entre las partes de señales expandidas panorámicamente direccionales y no direccionales. En esta estructura de codificación de escena de audio espacial (SASC), la idea central es representar una escena de audio de entrada de una manera que sea independiente de cualquier formato de reproducción asumido o pretendido. Esta parametrización agnóstica en formato, hace posible la reproducción óptima sobre cualquier sistema de reproducción dado así como la modificación de escena flexible. Se describe el análisis de señales y las herramientas de síntesis necesarias para SASC, incluyendo una presentación de los nuevos enfoques para la descomposición primaria-ambiental multi-canal. Las aplicaciones de SASC a la codificación de audio espacial, al mezclado ascendente, a la descodificación de matriz de amplitud de fase, la conversión de formato multi-canal, y la reproducción binaural pueden emplearse en conexión con o en el contexto del ajuste propuesto de una señal de audio espacial. Uno o más de estos aspectos pueden emplearse en conexión con o en el contexto del ajuste propuesto de una señal de audio espacial.

Algunas realizaciones de la presente invención pueden relacionarse con o emplear técnicas de mezclado ascendente. En general, las técnicas de mezclado ascendente podrían clasificarse en dos categorías principales: el tipo de métodos que alimentan los canales envolventes con el ambiente sintetizado o extraído a partir de los canales de entrada existentes (véase por ejemplo, J. S. Usher y Benesty, "Enhancement of Spatial Sound Quality: A New Reverberation-Extraction Audio Upmixer", IEEE Transactions on Audio, Speech, and Language Processing, vol. 15, n.º 7, pág. 2141-2150, 2007, C. Faller, "Multiple-Loudspeaker Playback of Stereo Signals", J. Audio Eng. Soc., vol. 54, n.º 11, pág. 1051-1064, 2006. C. Avendano y M.-M. Jot, "Ambience extraction and synthesis from stereo signals for multi-channel audio up-mix", en Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on, vol. 2, 2002, pág. II-1957-II-1960, y R. Irwan y R.M. Aarts, "Two-to-Five Channel Sound Processing", J. Audio Eng Soc. vol. 50, n.º 11, pág. 914-926, 2002) y aquellos que crean las señales de excitación para los canales adicionales por formación de matrices de los existentes (véase por ejemplo, R. Dressler (05.08.2004) Dolby Surround Pro Logic II Decoder Principles of Operation [En línea]. *Disponible:*

[http://www.dolby.com/uploadedFiles/Assets/US/Doc/Professional/209\\_Dolby\\_Surround\\_Pro\\_Logic\\_II\\_Decoder\\_Principles\\_of\\_Operation.pdf](http://www.dolby.com/uploadedFiles/Assets/US/Doc/Professional/209_Dolby_Surround_Pro_Logic_II_Decoder_Principles_of_Operation.pdf)). Un caso especial es el método propuesto en la Solicitud de Patente de Estados Unidos Publicación N.º US2010/0296672 A1 "Mezclado Ascendente de Canal Dos a Tres Para Derivación de Canal Central" por E. Vickers, donde en vez de una extracción ambiental se lleva a cabo una descomposición espacial. Entre otros, los métodos de generación ambiental pueden comprender la aplicación de reverberación artificial, calculando la diferencia de las señales izquierda y derecha, aplicando pequeños retardos para los canales de envolvente y los análisis de señales basados en la correlación. Los ejemplos para las técnicas de formación de matrices son los convertidores de matriz lineal y los métodos de dirección de matriz. Una breve vista general de estos métodos se proporciona por C. Avendano y J.-M. Jot en "Frequency Domain Techniques for Stereo to Multichannel Upmix", en 22nd International Conference of the AES on Virtual, Synthetic and Entertainment Audio, 2002, y por los mismos autores en "Ambience extraction and synthesis from stereo signals for multi-channel audio up-mix" en Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on, vol. 2, 2002, pág. II-1957-II-1960. Uno o más de estos aspectos pueden emplearse en conexión con o en el contexto del ajuste propuesto de una señal de audio espacial.

La extracción ambiental y la síntesis a partir de las señales estereofónicas para la mezcla ascendente de audio multi-canal puede conseguirse por una técnica de dominio de frecuencia para identificar y extraer la información ambiental en las señales de audio estereofónicas. El método está basado en el cálculo de un índice de coherencia inter-canal y una función de mapeo no lineal que nos permite determinar las regiones de tiempo-frecuencia que consisten principalmente de componentes ambientales en la señal de dos canales. Las señales ambientales se sintetizan y utilizan a continuación para alimentar los canales de envolvente de un sistema de reproducción multi-canal. Los resultados de la simulación demuestran la efectividad de la técnica en la extracción de la información ambiental y las pruebas de mezcla ascendente sobre el audio real revelan las diversas ventajas y desventajas del sistema, en comparación con las estrategias previas de mezcla ascendente. Uno o más de estos aspectos pueden emplearse en conexión con o en el contexto del ajuste propuesto de una señal de audio espacial.

Las técnicas del dominio de frecuencia para el mezclado ascendente estereofónico a multi-canal pueden emplearse también en conexión con o en el contexto del ajuste de una señal de audio espacial a una configuración de altavoces de reproducción. Están disponibles diversas técnicas de mezclado ascendente para generar el audio multi-canal a partir de las grabaciones estereofónicas. Las técnicas utilizan una estructura de análisis común basada en la comparación entre las Transformadas de Fourier de Tiempo Corto de las señales estereofónicas izquierda y derecha. Una medición de coherencia inter-canal se utiliza para identificar las regiones de tiempo-frecuencia que consisten principalmente de componentes ambientales, que pueden ponderarse a continuación mediante una función de mapeo no lineal, y extraerse para sintetizar las señales ambientales. Una medición de similitud se utiliza para identificar los coeficientes de expansión panorámica de las diversas fuentes en la mezcla en el plano de tiempo-frecuencia, y se aplican diferentes funciones de mapeo para desmezclar (extraer) una o más fuentes, y/o para re-expandir panorámicamente las señales en un número arbitrario de canales. Una aplicación posible de las diversas técnicas se refiere al diseño de un sistema de mezclado ascendente de dos a cinco canales. Uno o más de estos aspectos puede emplearse en conexión con o en el contexto del ajuste propuesto de una señal espacial.

Un descodificador de envolvente puede ser idóneo para llevar a cabo las pistas espaciales ocultas en las grabaciones de música convencionales de una manera convincente, natural. El oyente es atraído hacia un espacio tridimensional en vez de escuchar una presentación bidimensional plana. Esto no solamente ayuda a desarrollar un campo sonoro más envolvente, sino también resuelve el problema de "punto dulce" estrecho de la reproducción estereofónica convencional. En algunos descodificadores lógicos el circuito de control está considerando el nivel relativo y la fase entre las señales de entrada. Esta información se envía a la etapa de matriz de salida variable para ajustar los VCA que controlan el nivel de las señales en contrafase. Las señales en contrafase cancelan las señales de diafonía no deseadas, dando como resultado separación mejorada de los canales. Esto se denomina un diseño con corrección anticipante. Este concepto puede extenderse al mirar las mismas señales de entrada y realizar el control de circuito cerrado, de modo que éstas coincidan con sus niveles. Estas señales de audio coincidentes se envían directamente a las etapas de matriz para derivar los diversos canales de salida. Debido a que las mismas señales de audio que alimentan la matriz de salida se utilizan por sí mismas para controlar el servo bucle, se denomina un diseño lógico de realimentación. El concepto de control de realimentación puede mejorar la exactitud y optimizar las características dinámicas. La incorporación de realimentación global alrededor del proceso de dirección lógica trae beneficios similares en la exactitud de la dirección y el comportamiento dinámico. Uno o más de estos aspectos puede emplearse en conexión con o en el contexto del ajuste propuesto de una señal de audio espacial.

En conexión con la reproducción de altavoces múltiples, puede utilizarse una descomposición espacial perceptualmente motivada para las señales de audio estereofónicas de dos canales, que capturan la información respecto a la etapa de sonido virtual. La descomposición espacial permite la re-síntesis de las señales de audio para la reproducción sobre sistemas de sonido diferentes del estéreo de dos canales. Con el uso de más altavoces frontales, la anchura de la etapa de sonido virtual puede incrementarse más allá de  $\pm 30^\circ$  y se extiende a la región de punto dulce. Opcionalmente, los componentes laterales de sonido, independientes pueden reproducirse de manera separada sobre los altavoces sobre los lados de un oyente, para incrementar el envolvimiento del oyente. La

descomposición espacial puede utilizarse con el sonido de suspensión periférica y los sistemas de audio basados en síntesis de campo de onda. Uno o más de estos aspectos puede emplearse en conexión con o en el contexto del ajuste propuesto de una señal de audio espacial.

- 5 La descomposición de la señal de ambiente primaria y la localización basada en vector para la codificación de audio espacial y el mejoramiento, están dirigidas a la necesidad comercial creciente de almacenar y distribuir audio multi-canal y representar contenido óptimamente sobre sistemas de reproducción arbitrarios. Un esquema de síntesis de análisis espacial puede aplicar el análisis de componente principal a una representación de dominio de STFT (dominio de transformación de frecuencia de tiempo corto) del audio original para separarlo en los componentes
- 10 primario y ambiental, los cuales se analizan después respectivamente para las pistas que describen la percepción espacial de la escena de audio en una base por pieza; estas pistas pueden utilizarse por las síntesis para representar el audio apropiadamente sobre el sistema de reproducción disponible. Esta estructura puede adaptarse a medida para la codificación de audio espacial robusta, o puede aplicarse directamente a los escenarios de mejoramiento en los que no existen restricciones de velocidad sobre los datos espaciales intermediarios y la
- 15 representación de audio.

- Con respecto a la amplitud y al envolvimiento en las acústicas musicales, la sabiduría convencional sostiene que la amplitud y el envolvimiento se provocan por la energía sonora lateral en las habitaciones, y es principalmente la energía lateral que llega tempranamente, la que es más responsable. Sin embargo, por definición las salas o
- 20 habitaciones pequeñas no son espaciales, incluso éstas pueden estar cargadas con reflexiones laterales tempranas. Por lo tanto, los mecanismos perceptuales para la amplitud y el envolvimiento pueden tener una influencia sobre el ajuste de una señal de audio espacial. Se encuentra que las percepciones están relacionadas más comúnmente a la energía lateral (difusa) en las salas en los extremos de las notas (la reverberación de fondo) y menos frecuentemente, pero si importantes, a las propiedades del campo sonoro conforme se mantienen las. Se
- 25 sugiere una medida para la amplitud, denominada tiempo de decaimiento temprano lateral (LEDT). Uno o más de estos aspectos pueden emplearse en conexión con o en el contexto del ajuste propuesto de una señal de audio espacial.

## REIVINDICACIONES

1. Aparato (100) para adaptar una señal de audio espacial (2) para una configuración de altavoces original hacia una configuración de altavoces de reproducción que difiere de la configuración de altavoces original, en el que la señal de audio espacial (2) comprende una pluralidad de señales de canal, siendo cada señal de canal un canal de altavoz que corresponde a un altavoz de la configuración de altavoces original, comprendiendo el aparato:
- 5 un agrupador (110) configurado para agrupar la pluralidad de señales de canal en una pluralidad de segmentos originales, en el que al menos dos señales de canal vecinas se agrupan en un segmento original, y en el que un altavoz se asigna a un primer segmento original y a un segundo segmento original;
- 10 un descomponedor de ambiente directo (130) configurado para descomponer las al menos dos señales de canal en el primer segmento original en al menos un componente de sonido directo (D; 732) y al menos un componente ambiental (A; 734), y para determinar una dirección de llegada de al menos un componente de sonido directo (S, S<sub>1</sub>, S<sub>2</sub>) para el primer segmento original, y para descomponer las al menos dos señales de canal en el segundo segmento original en al menos un componente de sonido directo y al menos un componente ambiental para el segundo segmento original; y para determinar una dirección de llegada del al menos un componente de sonido directo para el segundo segmento original;
- 15 un representador de sonido directo (150) configurado para recibir información de una configuración de altavoces de reproducción para un primer segmento de reproducción asociado con el primer segmento original y para ajustar el al menos un componente de sonido directo (D; 732) del primer segmento original utilizando la información de la configuración de altavoces de reproducción para el primer segmento de reproducción para obtener al menos un componente de sonido directo ajustado de modo que una dirección percibida de llegada del al menos un componente de sonido directo (S, S<sub>1</sub>, S<sub>2</sub>) en la configuración de altavoces de reproducción es idéntica a la dirección de llegada del primer segmento original, o más cercana a la dirección de llegada de al menos un componente de sonido directo del primer segmento original en comparación con una situación en la cual no ha tenido lugar ningún ajuste del al menos un componente de sonido directo; y configurado para recibir una información de configuración de altavoces de reproducción para un segundo segmento de reproducción asociado con el segundo segmento original y para ajustar el al menos un componente de sonido directo del segundo segmento original usando la información de configuración de altavoces de reproducción para el segundo segmento de reproducción para obtener al menos un componente de sonido directo ajustado adicional de modo que una dirección percibida de llegada del al menos un componente de sonido directo en la configuración de altavoces de reproducción es idéntica a la dirección de llegada del segundo segmento original o más cercana a la dirección de llegada del al menos un componente de sonido directo del segundo segmento original en comparación con una situación en la cual no ha tenido lugar ningún ajuste del al menos un componente de sonido directo; y
- 20 un combinador (180) configurado para combinar el al menos un componente de sonido directo ajustado (752) y los componentes ambientales (734), o los componentes ambientales modificados de un primer segmento de reproducción y el al menos un componente de sonido directo ajustado adicional y los componentes ambientales o componentes ambientales modificados de un segundo segmento de reproducción.
2. Aparato (100) de acuerdo con la reivindicación 1, en el que la configuración de altavoces de reproducción comprende un altavoz adicional (L<sub>6</sub>) dentro del primer o segundo segmento original, de modo que el primer o segundo segmento original corresponde a dos o más segmentos del segmento de altavoz de reproducción; en el que el representador de sonido directo (150) está configurado para generar los componentes de sonido directo ajustados (752), para los al menos dos altavoces y el altavoz adicional de la configuración de altavoces de reproducción.
3. Aparato (100) de la reivindicación 1 o 2, en el que la configuración de altavoces de reproducción carece de un altavoz en comparación con la configuración de altavoces original, de modo que el primer o segundo segmento original y una configuración de primer o segundo segmento original vecino se unen a un segmento unido de la configuración de altavoces de reproducción; en el que el representador de sonido directo (150) está configurado para distribuir los componentes de sonido directo ajustados (752) de un canal correspondiente al altavoz que carece en la configuración de altavoces de reproducción a al menos dos altavoces restantes (L<sub>1</sub>, L<sub>3</sub>) del segmento unido de la configuración de altavoces de reproducción.
4. Aparato (100) de acuerdo con una cualquiera de las reivindicaciones 1 a 3, en el que el representador de sonido directo (150) está configurado para reasignar un componente de sonido directo (S<sub>2</sub>) que tiene una dirección determinada de llegada proveniente del primer o segundo segmento original ({L<sub>2</sub>, L<sub>3</sub>}) a un segmento original vecino ({L<sub>1</sub>, L'<sub>2</sub>}) si un límite entre el primer o segundo segmento original ({L<sub>2</sub>, L<sub>3</sub>}) y el segmento vecino ({L<sub>1</sub>, L'<sub>2</sub>}) traspasa la dirección determinada de llegada cuando se pasa de la configuración de altavoces original a la configuración de altavoces de reproducción.
5. Aparato (100) de acuerdo con la reivindicación 4, en el que el representador de sonido directo (150) está además configurado para reasignar el componente de sonido directo (S<sub>1</sub>) que tiene la dirección determinada de llegada

desde al menos un primer altavoz ( $L_3$ ) hacia al menos un segundo altavoz ( $L'_2$ ), estando asignado el al menos un primer altavoz ( $L_3$ ) al primer o segundo segmento original ( $\{L_2, L_3\}$ ) pero no hacia el segmento vecino ( $\{L_1, L'_2\}$ ) en la configuración de altavoces de reproducción, y estando asignado el al menos un segundo altavoz ( $L'_2$ ) al segmento vecino ( $\{L_1, L'_2\}$ ) en la configuración de altavoces de reproducción.

5 6. Aparato (100) de acuerdo con una cualquiera de las reivindicaciones 1 a 5, en el que el representador de sonido directo (150) está configurado para realizar una re-expansión panorámica de al menos un componente de sonido directo ( $S, S_1, S_2$ ) utilizando la información de la configuración de altavoces de reproducción y la dirección de llegada del al menos un componente de sonido directo.

10 7. Aparato (100) de acuerdo con la reivindicación 6, en el que el representador de sonido directo (150) está además configurado para realizar la re-expansión panorámica de al menos un componente de sonido directo ( $S_1$ ) que tiene la dirección determinada de llegada ajustando las señales de altavoz para los altavoces  $\{L_1, L_2\}$  en el primer o segundo segmento original ( $\{L_1, L_2\}$ ) para obtener las señales de altavoz ajustadas para los altavoces ( $L_1, L'_2$ ) en un segmento modificado correspondiente ( $\{L_1, L'_2\}$ ) de la configuración de altavoces de reproducción, si al menos uno de los altavoces ( $L_1, L_2$ ) en el primer o segundo segmento original ( $\{L_1, L_2\}$ ) se desplaza en el segmento modificado correspondiente  $\{L_1, L'_2\}$  de la configuración de altavoces de reproducción sin traspasar la dirección determinada de llegada.

20 8. Aparato (100) de acuerdo con una cualquiera de las reivindicaciones 1 a 7, en el que el representador de sonido directo (150) está configurado para generar los componentes de sonido directo específicos del segmento de altavoz para al menos dos pares válidos de segmento de altavoz de la configuración de altavoces de reproducción, refiriéndose los al menos dos pares de segmentos de altavoz válidos a un mismo altavoz y dos segmentos vecinos en la configuración de altavoces de reproducción; y en el que el combinador (180) está configurado para combinar los componentes de sonido directo específicos del segmento de altavoz para los al menos dos pares de segmentos de altavoz válidos con referencia al mismo altavoz para obtener una de las señales de altavoz para los al menos dos altavoces de la configuración de altavoces de reproducción.

30 9. Aparato (100) de acuerdo con una cualquiera de las reivindicaciones 1 a 8, en el que el representador de sonido directo (150) está además configurado para procesar el al menos un componente de sonido directo ( $D; 732$ ) para un segmento dado de la configuración de altavoces de reproducción y para generar de esta manera los componentes de sonido directo ajustados para cada altavoz asignado al segmento dado.

35 10. Aparato (100) de acuerdo con cualquiera de las reivindicaciones 1 a 9, que comprende además un representador ambiental (170) configurado para recibir la información de la configuración de altavoces de reproducción para el primer o segundo segmento de reproducción y para ajustar el al menos un componente ambiental utilizando la información de la configuración de altavoces de reproducción para el primer o segundo segmento de reproducción, de modo que un involucrimiento percibido del al menos un componente ambiental en la configuración de altavoces de reproducción es idéntico a un involucrimiento del al menos un componente ambiental del primer o segundo segmento original en comparación con una situación en la cual no ha tenido lugar ningún ajuste del al menos un componente ambiental.

45 11. Aparato (100) de acuerdo con una cualquiera de las reivindicaciones 1 a 10, en el que el agrupador (110) está además configurado para escalar al menos dos canales como una función de a cuántos segmentos originales está asignado un canal de los al menos dos canales.

50 12. Aparato (100) de acuerdo con una cualquiera de las reivindicaciones 1 a 11, que comprende además un ajustador de distancia (190) configurado para ajustar al menos una de una amplitud y un retardo de al menos una de las señales de altavoz para los al menos dos altavoces de la configuración de altavoces de reproducción utilizando una información de distancia relacionada con una distancia entre un oyente y un altavoz en la configuración de altavoces de reproducción.

55 13. Aparato (100) de acuerdo con una cualquiera de las reivindicaciones 1 a 12, que comprende además un rastreador de oyente configurado para determinar una posición actual de un oyente con respecto a la configuración de altavoces de reproducción, y para determinar la información de la configuración de altavoces de reproducción utilizando la posición actual del oyente.

60 14. Aparato (100) de acuerdo con una cualquiera de las reivindicaciones 1 a 13, que comprende además un transformador de tiempo-frecuencia configurado para transformar la señal de audio espacial de una representación de dominio de tiempo a una representación de dominio de frecuencia o a una representación de dominio de tiempo-frecuencia, en el que el descomponedor de ambiente directo y el representador de sonido directo están configurados para procesar la representación del dominio de frecuencia o la representación del dominio de tiempo-frecuencia.

15. Método para adaptar una señal de audio espacial (2) para una configuración de altavoces original a una

configuración de altavoces de reproducción que difiere de la configuración de altavoces original, en el que la señal de audio espacial (2) comprende una pluralidad de canales, siendo cada señal de canal un canal de altavoz que corresponde a un altavoz de la configuración de altavoces original, comprendiendo el método:

- 5 agrupar (802) la pluralidad de señales de canal en una pluralidad de segmentos originales, en el que al menos dos señales de canal vecinas se agrupan en un segmento original, y en el que un altavoz se asigna a un primer segmento original y a un segundo segmento original;
- 10 descomponer (804) las al menos dos señales de canal en el primer segmento original, en al menos un componente de sonido directo (D; 732) y al menos un componente ambiental (A; 734), y determinar una dirección de llegada de al menos un componente de sonido directo (S, S<sub>1</sub>, S<sub>2</sub>) para el primer segmento original, y descomponer las al menos dos señales de canal en el segundo segmento original en al menos un componente de sonido directo y al menos un componente ambiental para el segundo segmento original; y determinar una dirección de llegada del al menos un componente de sonido directo para el segundo segmento original;
- 15 ajustar (808) el al menos un componente de sonido directo (D; 732) del primer segmento original utilizando la información de la configuración de altavoces de reproducción para el primer segmento de reproducción para obtener al menos un componente de sonido directo ajustado de modo que una dirección percibida de llegada del al menos un componente de sonido directo (S, S<sub>1</sub>, S<sub>2</sub>) en la configuración de altavoces de reproducción es idéntica a la dirección de llegada del primer segmento original, o más cercana a la dirección de llegada de al menos un componente de sonido directo del primer segmento original en comparación con una situación en la cual no ha tenido lugar ningún ajuste del al menos un componente de sonido directo; y ajustar el al menos un componente de sonido directo del segundo segmento original usando la información de configuración de altavoces de reproducción para el segundo segmento de reproducción para obtener al menos un componente de sonido directo ajustado adicional de modo que una dirección percibida de llegada del al menos un componente de sonido directo en la configuración de altavoces de reproducción es idéntica a la dirección de llegada del segundo segmento original o más cercana a la dirección de llegada del al menos un componente de sonido directo del segundo segmento original en comparación con una situación en la cual no ha tenido lugar ningún ajuste del al menos un componente de sonido directo; y
- 20 combinar (809) el al menos un componente de sonido directo ajustado (752) y los componentes ambientales (734), o los componentes ambientales modificados de un primer segmento de reproducción y el al menos un componente de sonido directo ajustado adicional y los componentes ambientales o componentes ambientales modificados de un segundo segmento de reproducción.

16. Un programa informático que tiene un código de programa para realizar el método de acuerdo con la reivindicación 15 cuando el programa informático se ejecuta en un ordenador.

35

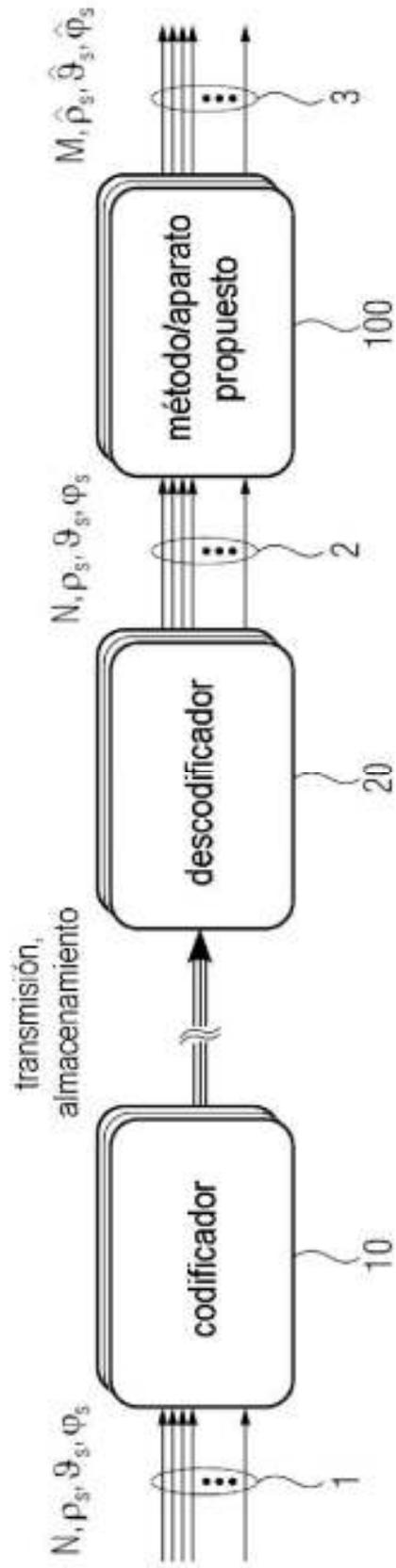


FIG 1

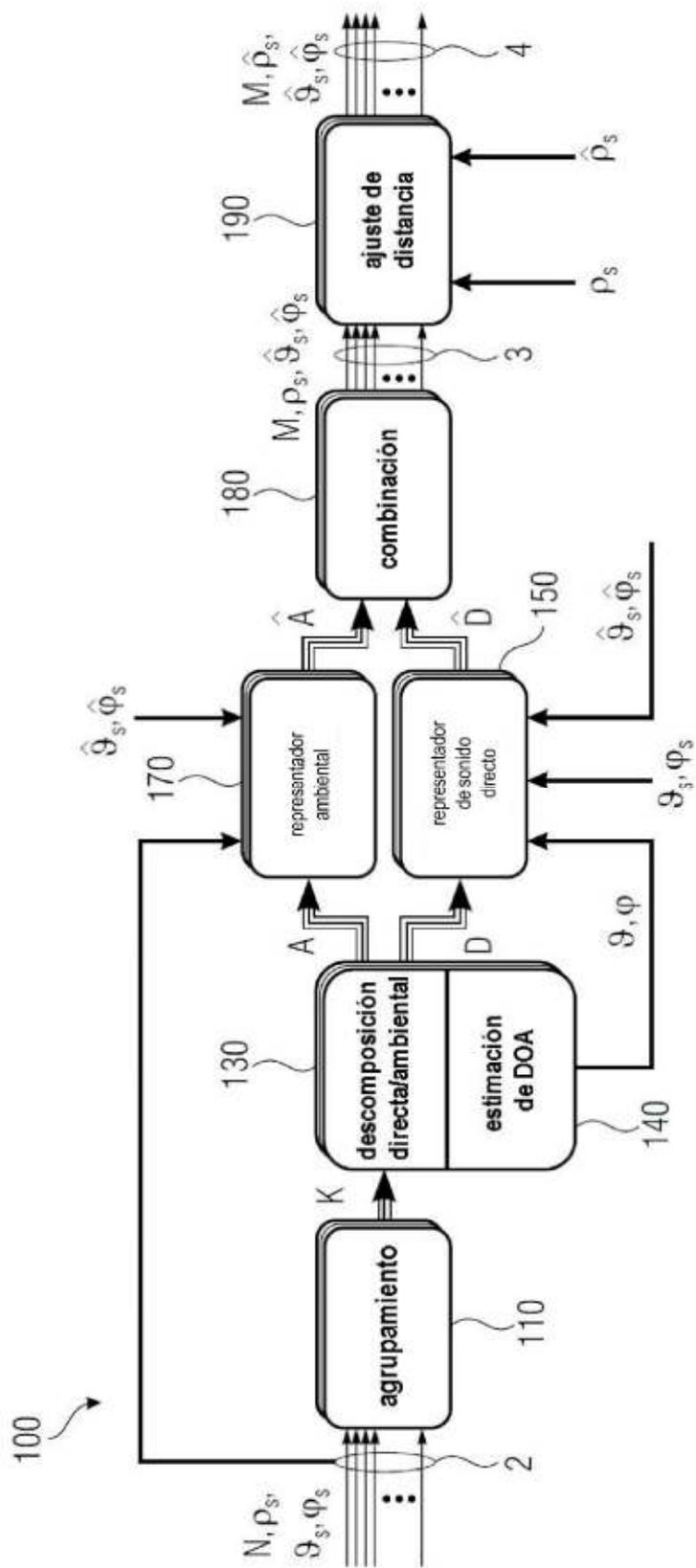


FIG 2

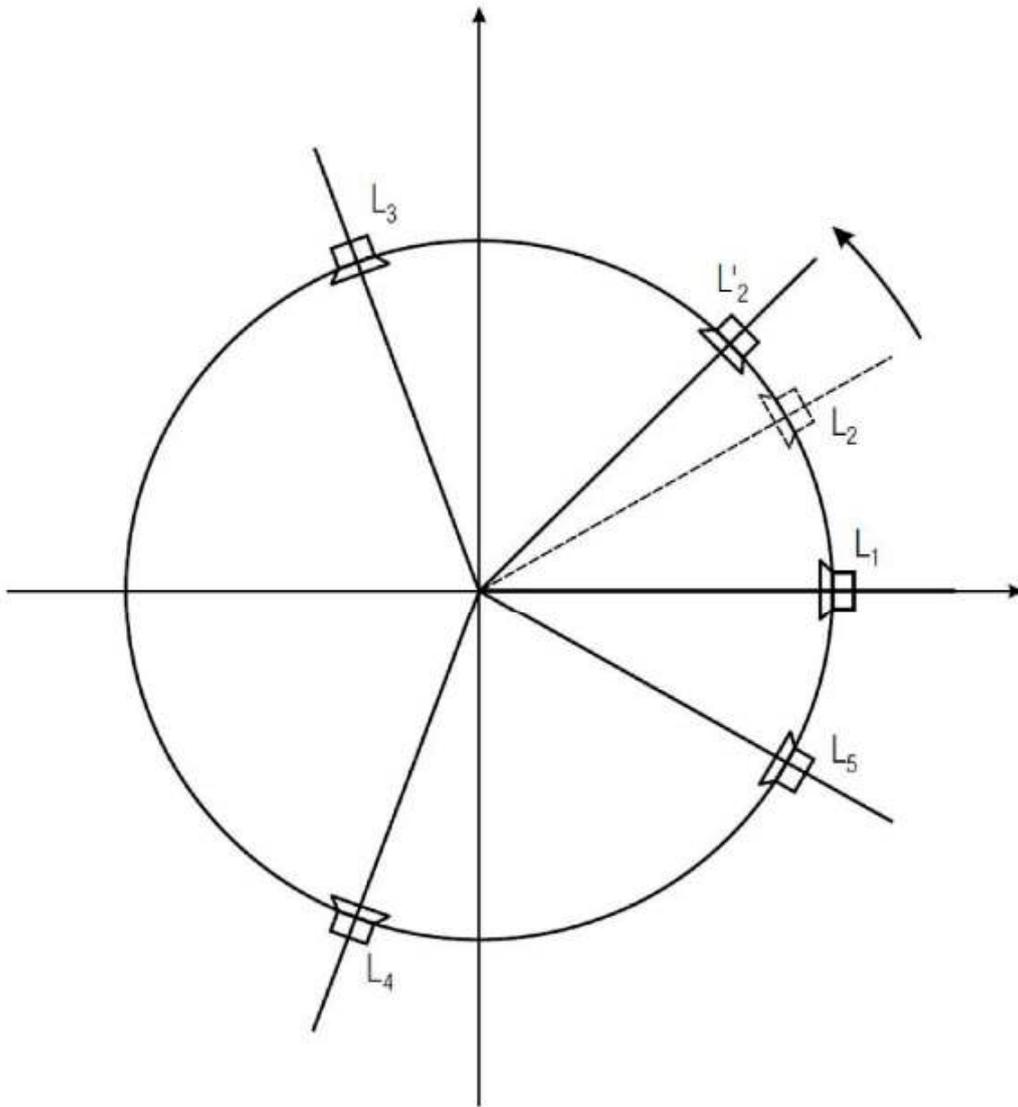


FIG 3

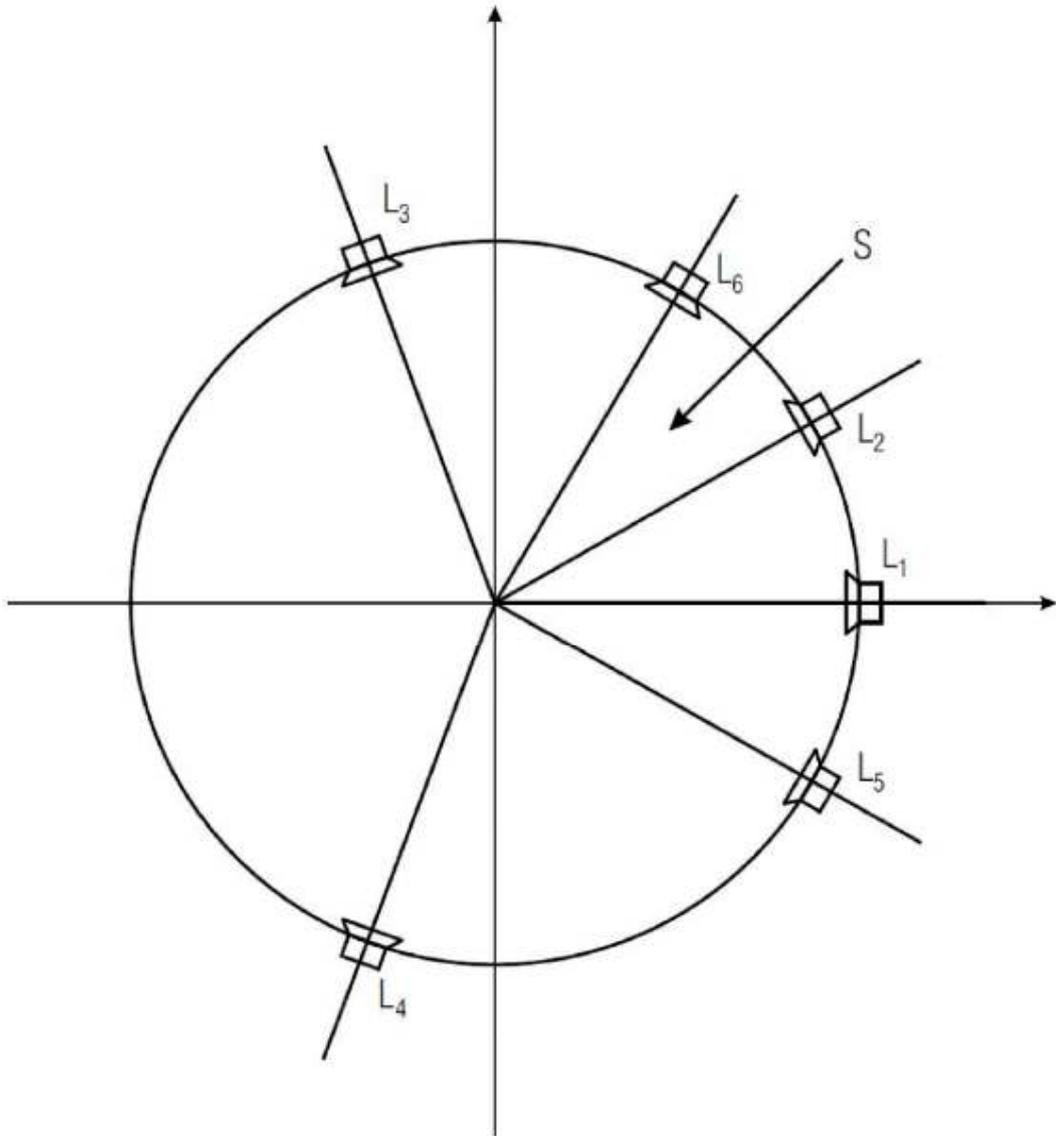


FIG 4

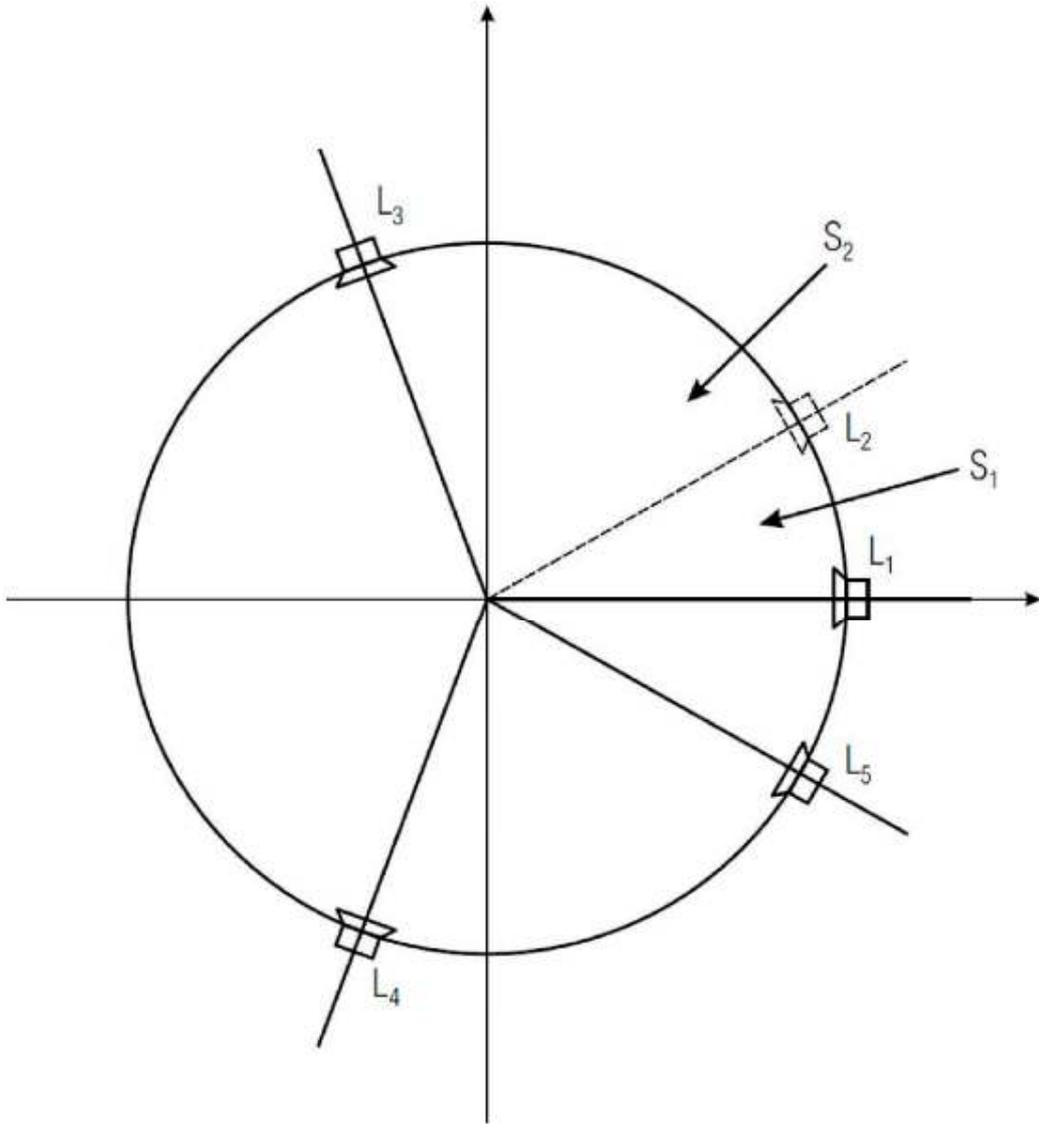


FIG 5

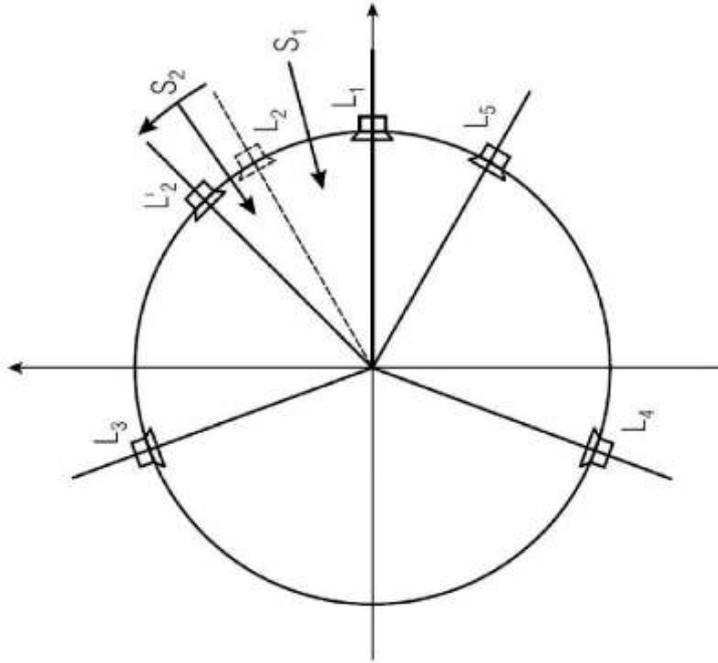


FIG 6A

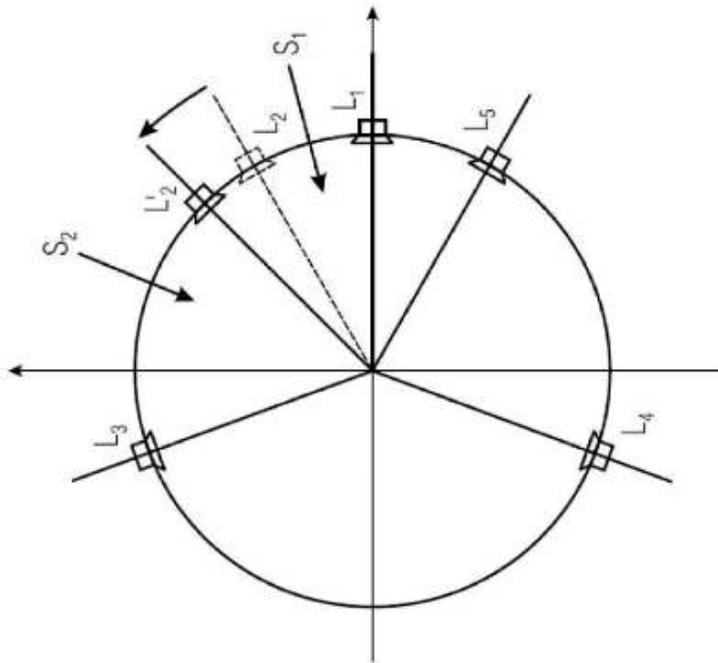


FIG 6B

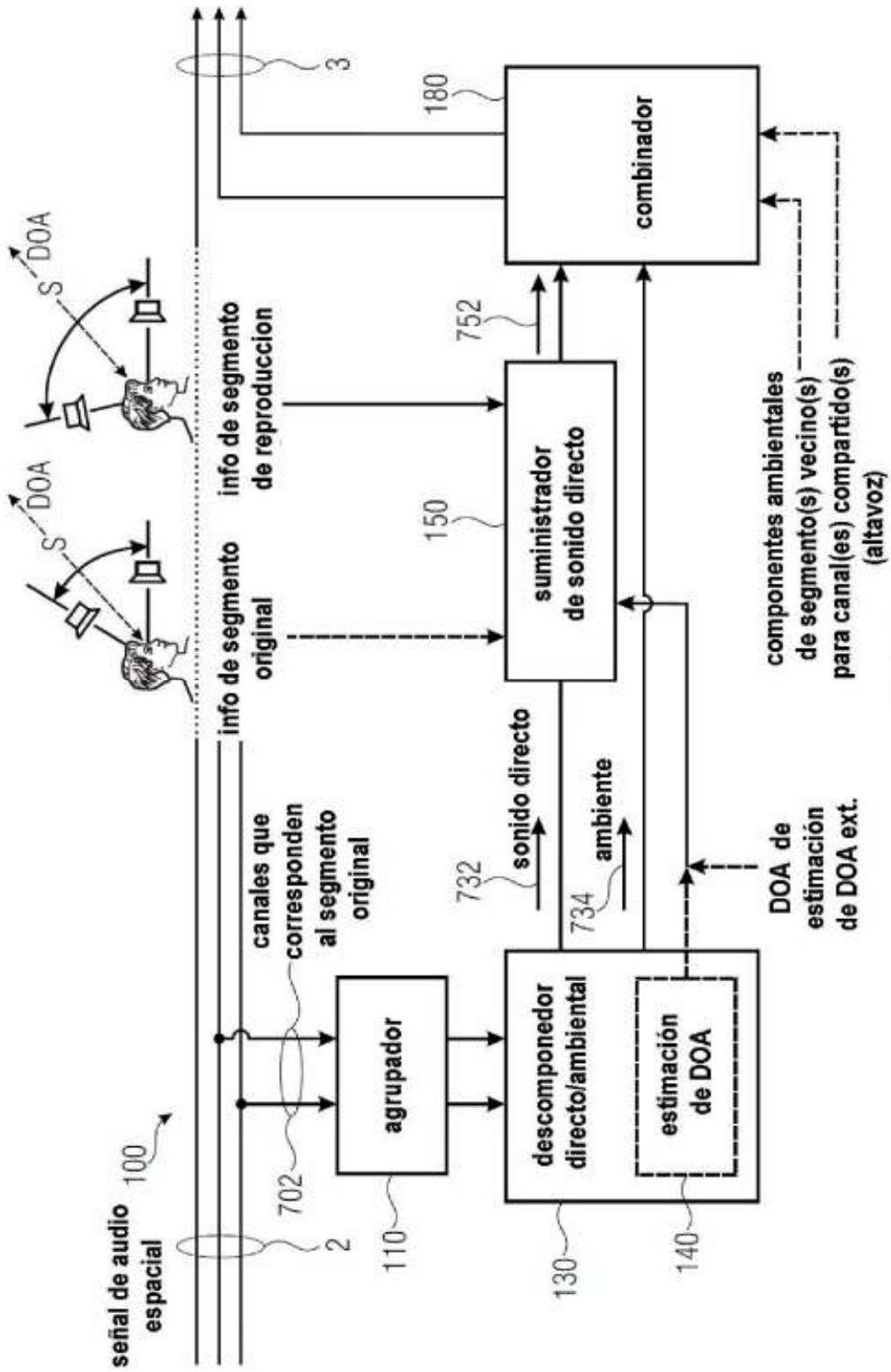


FIG 7

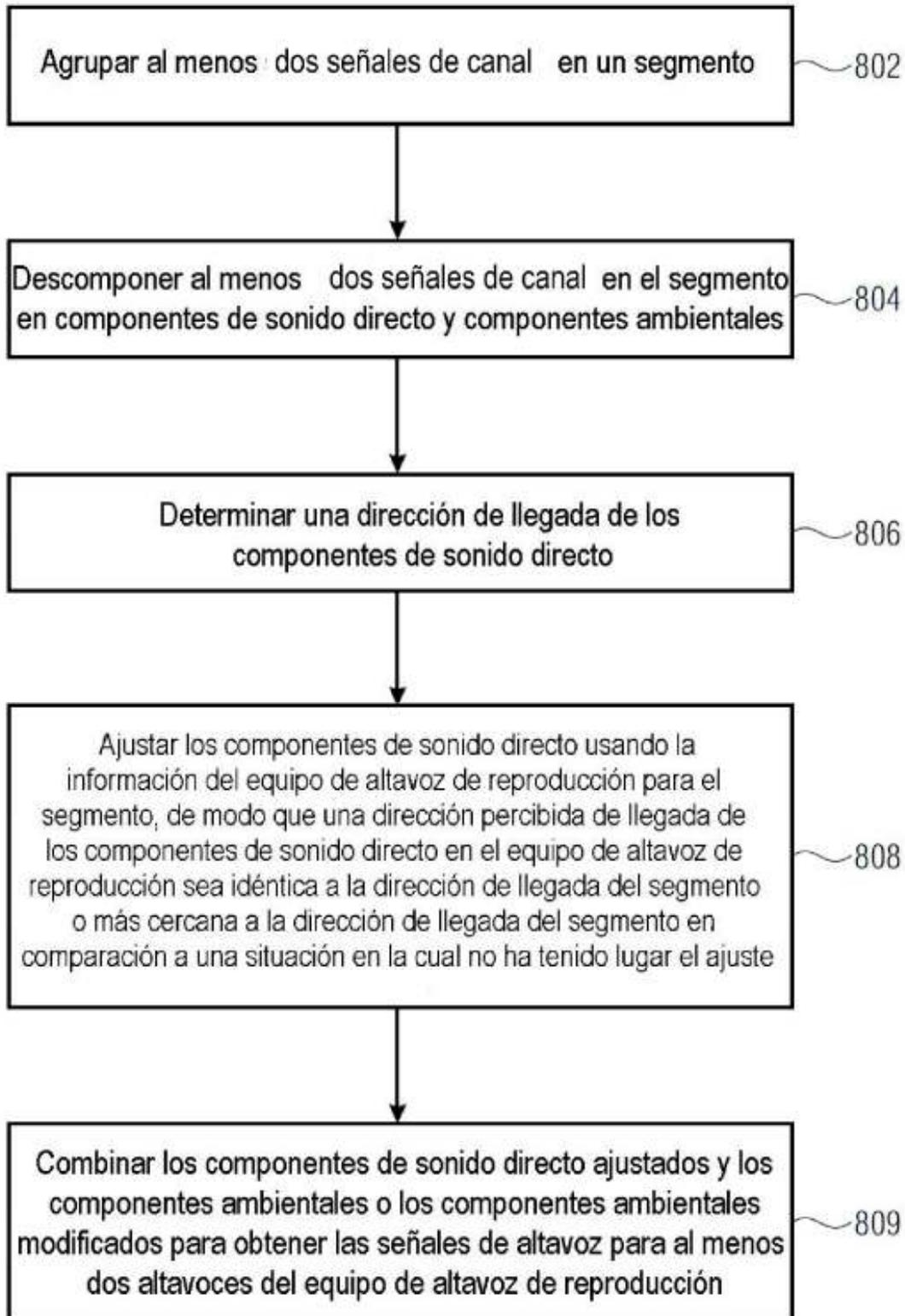


FIG 8