

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 713 685**

51 Int. Cl.:

**G10L 21/0308** (2013.01)

**G10L 25/48** (2013.01)

**H04R 3/00** (2006.01)

**H04R 1/40** (2006.01)

**H04S 3/00** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **26.04.2016** **E 16166989 (0)**

97 Fecha y número de publicación de la concesión europea: **12.12.2018** **EP 3239981**

54 Título: **Métodos, aparatos y programas informáticos relativos a la modificación de una característica asociada a una señal de audio separada**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**23.05.2019**

73 Titular/es:  
**NOKIA TECHNOLOGIES OY (100.0%)**  
**Karaportti 3**  
**02610 Espoo, FI**

72 Inventor/es:  
**ERONEN, ANTTI;**  
**LEHTINIEMI, ARTO;**  
**LEPPÄNEN, JUSSI y**  
**CRICRI, FRANCESCO**

74 Agente/Representante:  
**VALLEJO LÓPEZ, Juan Pedro**

ES 2 713 685 T3

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

**DESCRIPCIÓN**

Métodos, aparatos y programas informáticos relativos a la modificación de una característica asociada a una señal de audio separada

5 **Campo**

La presente especificación se refiere a una modificación de una característica asociada a una señal de audio separada.

10 **Antecedentes**

Las técnicas de procesamiento de señales de audio permiten la identificación y la separación de fuentes sonoras individuales a partir de señales de audio que incluyen componentes desde una pluralidad de diferentes fuentes de sonoras. Una vez se ha separado del resto de la señal una señal de audio que representa una señal de audio identificada, pueden modificarse las características de la señal separada para proporcionar diferentes efectos audibles a un oyente.

20 **Sumario**

En un primer aspecto, la presente especificación describe un método que comprende determinar, basándose en una determinada medida del éxito de una separación de una señal de audio que representa una fuente sonora a partir de una señal de audio compuesta que comprende componentes derivados de al menos dos fuentes sonoras, un valor de un parámetro de modificación de la señal separada, indicando el valor del parámetro de modificación de la señal separada un intervalo de modificación de una característica asociada a la señal de audio separada.

El parámetro de modificación de la señal separada puede ser un parámetro de reposicionamiento espacial que indique un intervalo de reposicionamiento espacial para el reposicionamiento espacial de la señal de audio separada. Otros ejemplos de la característica asociada a la señal de audio separada pueden incluir, pero sin limitación, la amplitud, ecualización, reverberación, distorsión y compresión.

El método puede comprender determinar la medida del éxito de la separación de la señal de audio a partir de la señal de audio compuesta.

35 El método puede comprender limitar una cantidad permitida de modificación de la característica asociada a la señal de audio separada basándose en el valor del parámetro de modificación de la señal separada.

El método puede comprender hacer que se proporcione a un usuario una indicación del valor determinado del parámetro de modificación de la señal separada.

40 El método puede comprender, cuando la medida del éxito indica que el éxito de la separación está por encima de un grado de umbral de éxito, determinar un valor del parámetro de modificación de la señal separada que indique un intervalo completo de modificación de la característica.

45 Cuando la medida del éxito indica que el éxito de la separación está por debajo de un grado de umbral de éxito, el valor determinado del parámetro de modificación de la señal separada puede indicar un intervalo de modificación que tiene una relación directa con el grado de éxito.

50 La medida del éxito puede comprender una correlación entre un resto de la señal de audio compuesta y al menos una señal de audio de referencia. La al menos una señal de referencia puede comprender una o ambas de entre la señal de audio separada y la señal derivada de uno de los dispositivos de grabación adicionales que se asocian a la fuente de audio a la que se refiere la señal de audio separada. El método puede comprender además, si la correlación está por debajo de la correlación de umbral predeterminada, determinar un valor del parámetro de modificación de la señal separada que indica un intervalo completo de modificación y, si la correlación está por encima de la correlación de umbral predeterminada, determinar un valor del parámetro de modificación de la señal separada que indica un intervalo de modificación que tiene una relación inversa con la correlación.

60 En otros ejemplos, la medida del éxito de la separación puede comprender adicional o alternativamente una correlación entre un espectro de frecuencia asociado al resto de la señal de audio compuesta y un espectro de frecuencia asociado a la señal de audio de referencia. En otros ejemplos más, la medida del éxito de la separación puede comprender adicional o alternativamente una correlación entre un resto de la señal de audio compuesta y un componente de una señal de vídeo correspondiente a la señal de audio compuesta.

65 La correlación entre el resto de la señal de audio compuesta y la señal de referencia o entre el resto de la señal de audio compuesta y el componente de la señal de vídeo correspondiente a la señal de audio compuesta puede tener una relación inversa con un grado de éxito de la separación.

El método puede comprender responder a una determinación de que la medida del éxito de la separación indica que, para un marco temporal posterior de la señal de audio compuesta, un grado de éxito de la separación es más bajo que el grado de éxito de la separación para un marco temporal actual de la señal de audio compuesta mediante el reposicionamiento espacialmente de la señal de audio separada a una posición que esté más próxima a una posición espacial original de la señal de audio separada. El reposicionamiento espacial de la señal de audio separada a la posición que está más cerca de la posición espacial original puede realizarse previamente a la representación del marco temporal posterior de la señal de audio compuesta.

El método puede comprender provocar la realización de la separación de audio digital que representa la fuente sonora de la señal de audio compuesta.

El método puede comprender reposicionar la señal de audio separada a una nueva posición espacial basándose en el valor determinado del parámetro de reposicionamiento espacial.

En un segundo aspecto, la presente especificación describe un aparato configurado para realizar un método tal como se ha descrito con referencia al primer aspecto.

En un tercer aspecto, la presente especificación describe instrucciones legibles por ordenador que, cuando se ejecutan por un aparato informático, hacen que el aparato informático provoque la realización del método tal como se ha descrito con referencia al primer aspecto.

### Breve descripción de las figuras

Para una mejor comprensión de la presente solicitud, se hará referencia a modo de ejemplo a los dibujos adjuntos en los que:

la figura 1 es un ejemplo de un sistema de captura de audio que puede usarse para capturar señales de audio para procesamiento de acuerdo con diversos ejemplos descritos en el presente documento;

las figuras 2A a 2C son diagramas de flujo que ilustran diversas operaciones que pueden realizarse por el aparato de procesamiento de audio representado en la figura 1;

la figura 3A es un ejemplo de una interfaz de usuario gráfica que puede proporcionarse de ese modo para indicar a un usuario un valor de un parámetro de modificación de la señal separada;

la figura 3B es otro ejemplo de una interfaz de usuario gráfica que puede proporcionarse de ese modo para indicar a un usuario un valor de un parámetro de modificación de la señal separada;

la figura 3C es otro ejemplo de una interfaz de usuario gráfica que puede proporcionarse de ese modo para indicar a un usuario un valor de un parámetro de modificación de la señal separada;

las figuras 4A a 4C ilustran diversos conceptos descritos en el presente documento con relación al reposicionamiento espacial de señales de audio separadas; y

la figura 5 es una ilustración esquemática de una configuración de ejemplo del aparato de procesamiento de audio representado en la figura 1.

### Descripción detallada de realizaciones

En la descripción y dibujos, números de referencia iguales se refieren a elementos iguales a todo lo largo de ellos.

La figura 1 es un ejemplo de un sistema de captura de audio 1 que puede usarse para capturar señales de audio para procesamiento de acuerdo con diversos ejemplos descritos en el presente documento. En este ejemplo, el sistema 1 comprende un aparato de captura de audio espacial 10 configurado para capturar una señal espacial de audio, y uno o más dispositivos de captura de audio adicionales 12A, 12B, 12C.

El aparato de captura de audio espacial 10 comprende una pluralidad de dispositivos de captura de audio 101A, B (por ejemplo, micrófonos direccionales o no direccionales) que se disponen para capturar señales de audio que posteriormente pueden reproducirse espacialmente en un flujo de audio de tal manera que el sonido reproducido se percibe por un oyente como originado desde al menos una posición espacial virtual. Normalmente, el sonido capturado por el aparato de captura de audio espacial 10 se deriva de una pluralidad de fuentes sonoras diferentes que pueden estar en una o más localizaciones diferentes con relación al aparato de captura de audio espacial 10. Dado que la señal espacial de audio capturado incluye componentes derivados de una pluralidad de diferentes fuentes sonoras, puede denominarse como una señal de audio compuesta. Aunque solo son visibles los dispositivos de captura de audio 102A, B en la figura 1, el aparato de captura de audio espacial 10 puede comprender más de dos dispositivos 102A, B. Por ejemplo, en algunos ejemplos específicos, el aparato de captura de audio 10 puede comprender ocho dispositivos de captura de audio.

En el ejemplo de la figura 1, el aparato de captura de audio espacial 10 se configura también para capturar contenido visual (por ejemplo vídeo) por medio de una pluralidad de dispositivos de captura de contenido visual 102A-G (por ejemplo cámaras). La pluralidad de dispositivos de captura de contenido visual 102A-G del aparato de captura de audio espacial 10 puede configurarse para capturar contenido visual desde diversas direcciones diferentes alrededor del aparato, para de ese modo proporcionar inmersión (o contenidos de realidad virtual) para consumo por los usuarios. En el ejemplo de la figura 1, el aparato de captura de audio espacial 10 es un dispositivo de captura de presencia, tal como la cámara OZO de Nokia. Sin embargo, como se apreciará, el aparato de captura de audio espacial 10 puede ser otro tipo de dispositivo y/o puede componerse de una pluralidad de dispositivos físicamente separados. Como también se apreciará, aunque el contenido capturado puede ser adecuado para

proporcionarse como contenido de inmersión, puede proporcionarse también en un formato regular, no de realidad virtual, por ejemplo a través de un teléfono inteligente o un ordenador tableta.

5 Como se ha mencionado anteriormente, en el ejemplo de la figura 1, el sistema de captura espacial de audio 1 comprende además uno o más dispositivos de captura de audio adicionales 12A-C. Cada uno de los dispositivos de  
 10 captura de audio adicionales 12A-C puede comprender al menos un micrófono y, en el ejemplo de la figura 1, los dispositivos de captura de audio adicionales 12A-C son micrófonos de lavalier configurados para capturar señales de audio derivadas de un usuario asociado 13A-C. Por ejemplo, en la figura 1, cada uno de los dispositivos de captura de audio adicionales 12A-C se asocia a un usuario diferente mediante ser fijado al usuario en alguna forma. Sin embargo, se apreciará que, en otros ejemplos, los dispositivos de captura de audio adicionales 12A-C pueden tener una forma diferente y/o pueden localizarse en localizaciones fijas, predeterminadas dentro de un entorno de captura de audio.

15 Las localizaciones de los dispositivos de captura de audio adicionales 12A-C y/o el aparato de captura de audio espacial 10 dentro del entorno de captura de audio pueden ser conocidas por, o pueden ser determinables por, el sistema de captura de audio 1 (por ejemplo, el aparato de procesamiento de audio 14). Por ejemplo, en el caso de dispositivos/aparatos de captura de audio móvil los dispositivos/aparatos pueden incluir un componente de determinación de la localización para permitir que se determine la localización de los dispositivos/aparatos. En algunos ejemplos específicos, puede emplearse un sistema de determinación de la localización por radiofrecuencia tal como el High Accuracy Indoor Positioning de Nokia, mediante lo que los dispositivos de captura de audio  
 20 adicionales 12A-C (y en algunos ejemplos el aparato de captura de audio espacial 10) transmiten mensajes para permitir al servidor de localización determinar la localización de los dispositivos de captura de audio adicionales dentro del entorno de captura de audio. En otros ejemplos, por ejemplo cuando los dispositivos de captura de audio adicionales 12A-C son estáticos, las localizaciones pueden pre-almacenarse por una entidad que forma parte del sistema de captura de audio 1 (por ejemplo, el aparato de procesamiento de audio 14).  
 25

En el ejemplo de la figura 1, el sistema de captura de audio 1 comprende además el aparato de procesamiento de audio 14. El aparato de procesamiento de audio 14 se configura para recibir y almacenar señales capturadas por el aparato de captura de audio espacial 10 y los uno o más dispositivos de captura de audio adicionales 12A-C. Las  
 30 señales pueden recibirse en el aparato de procesamiento de audio 14 en tiempo real durante la captura de las señales de audio o pueden recibirse posteriormente por ejemplo a través de un dispositivo de almacenamiento intermedio. En dichos ejemplos, el aparato de procesamiento de audio 14 puede ser local respecto al entorno de captura de audio o puede estar geográficamente remoto respecto al entorno de captura de audio en el que se proporcionan el aparato de captura de audio 10 y los dispositivos 12A-C. En algunos ejemplos, el aparato de  
 35 procesamiento de audio 14 puede incluso formar parte del aparato de captura de audio espacial 10.

Las señales de audio recibidas por el aparato de procesamiento de la señal de audio 14 pueden comprender una entrada de audio multicanal en un formato de altavoz. Dichos formatos pueden incluir, pero sin limitación, un formato de señal estéreo, un formato de señal 4.0, un formato de señal 5.1 y un formato de señal 7.1. En dichos ejemplos,  
 40 las señales capturadas por el sistema de la figura 1 pueden haber sido pre-procesadas respecto a su formato en bruto original hacia un formato de altavoz. Alternativamente, en otros ejemplos, las señales de audio recibidas por el aparato de procesamiento de audio 14 pueden estar en un formato de señal multi-micrófono, tal como una señal de entrada de ocho canales en bruto. Las señales multi-micrófono en bruto pueden, en algunos ejemplos, ser pre-procesadas por el aparato de procesamiento de audio 14 usando técnicas de procesamiento espacial de audio para convertir de ese modo las señales recibidas a un formato de altavoz o un formato binaural.  
 45

En algunos ejemplos, el aparato de procesamiento de audio 14 puede configurarse para mezclar las señales derivadas desde los uno o más dispositivos de captura de audio adicionales 12A-C con las señales derivadas del aparato de captura de audio espacial 10. Por ejemplo, pueden utilizarse las localizaciones de los dispositivos de  
 50 captura de audio adicionales 12A-C para mezclar las señales derivadas de los dispositivos de captura de audio adicionales 12A-C a las posiciones espaciales correctas dentro del audio espacial derivado del aparato de captura de audio espacial 10. La mezcla de las señales por el aparato de procesamiento de audio 14 puede estar parcial o totalmente automatizada.

55 El aparato de procesamiento de audio 14 puede configurarse adicionalmente para realizar (o permitir la realización de) el reposicionamiento espacial dentro del audio espacial capturado mediante el aparato de captura de audio espacial 10 de las fuentes sonoras capturadas por los dispositivos de captura de audio adicionales 12A-C.

El reposicionamiento espacial de fuentes sonoras puede realizarse para permitir la reproducción futura en el espacio tridimensional con un punto de vista libre de audio en el que un usuario puede elegir libremente una nueva posición de escucha. También, el reposicionamiento espacial puede usarse para separar fuentes sonoras para de ese modo hacerlas más individualmente distintas. De modo similar, el reposicionamiento espacial puede usarse para  
 60 acentuar/desacentuar ciertas fuentes en una mezcla de audio mediante la modificación de su posición espacial. Otros usos del reposicionamiento espacial pueden incluir, pero ciertamente sin limitación, la colocación de ciertas fuentes sonoras en una localización espacial deseada, para de ese modo obtener la atención de los oyentes (esto puede ser denominado como indicios de audio), limitar el movimiento de las fuentes sonoras para adaptarse a un  
 65

cierto umbral, y ampliar la señal de audio mezclada mediante la ampliación de las localizaciones espaciales de diversas fuentes sonoras. Diversas técnicas para la realización de reposicionamiento espacial son conocidas en la técnica y por ello no se detallarán en el presente documento. Un ejemplo de una técnica que puede usarse implica el cálculo de las ganancias deseadas para una fuente sonora usando Panorámica Vectorial Basada en Amplitud (VBAP, del inglés "Vector Base Amplitude Panning") cuando se mezclan señales de audio en el dominio de señales del altavoz.

Un problema a ser acometido cuando se realiza reposicionamiento espacial es el hecho de que la señal de audio espacial capturada por el aparato de captura de audio espacial 10 incluirá normalmente componentes derivados de la fuente sonora que está siendo reposicionada. Como tal, no será suficiente mover simplemente la señal capturada mediante un dispositivo de captura de audio adicional 12A-C. En su lugar, los componentes de la fuente sonora resultante deberían separarse también de la señal de audio espacial (compuesta) capturada por el aparato de audio espacial 10 y deberían reposicionarse junto con la señal capturada por el dispositivo de captura de audio adicional 12A-C. Si esto no se realiza, el oyente oírá componentes derivados de la misma fuente sonora como procedentes de diferentes localizaciones, lo que es claramente indeseable.

Son conocidas en la técnica diversas técnicas para identificación y separación de fuentes sonoras individuales (tanto estáticas como en movimiento) de una señal compuesta y por ello no serán explicadas con mucho detalle en la presente especificación. Brevemente, el proceso de separación implica normalmente identificar/estimar la fuente a ser separada, y a continuación restar o eliminar en otra forma esa fuente identificada de la señal compuesta. La eliminación de la fuente sonora identificada podría realizarse en el dominio del tiempo mediante la resta de una señal en el dominio del tiempo de la fuente estimada, o en el dominio de la frecuencia. Un ejemplo de método de separación que puede utilizarse por el aparato de procesamiento de audio 14 es el que se describe en la solicitud de patente pendiente PCT/EP2016/051709 que se refiere a la identificación y separación de una fuente sonora móvil de una señal compuesta. Otro método que puede utilizarse puede ser el que se describe en el documento WO 2014/147442 que describe la identificación y separación de una fuente sonora estática.

Otro ejemplo puede hallarse en el documento US 2012/0114130 A1. Independientemente de cómo se identifiquen las fuentes sonoras, una vez que se han identificado, puede restarse o filtrarse a la inversa a partir de la señal de audio espacial compuesta para proporcionar una señal de audio separada y un resto de la señal de audio compuesta. A continuación del reposicionamiento espacial (u otra modificación) de la señal de audio separada, la señal separada modificada puede volverse a mezclar de nuevo dentro del resto de la señal de audio compuesta para formar una señal de audio compuesta modificada.

La separación de una fuente sonora individual de una señal de audio compuesta puede no ser particularmente directa y, como tal, puede no ser posible en todos los casos separar totalmente una fuente sonora individual de la señal de audio compuesta. En dichos casos, algunos componentes derivados de la fuente sonora que se pretende separar pueden permanecer en la señal compuesta restante a continuación de la operación de separación.

Cuando la separación no es totalmente exitosa, y la señal separada se mezcla de nuevo dentro del resto de la señal de audio compuesta en una localización reposicionada, puede degradarse la calidad que se experimenta por el usuario de la representación de audio resultante. Por ejemplo, en algunos casos, el usuario puede oír la fuente sonora en una posición intermedia entre la localización original de la fuente sonora y la localización reposicionada pretendida. En otros ejemplos, el usuario puede oír dos fuentes sonoras distintas, una en la localización original y otra en la localización reposicionada. El efecto experimentado por el usuario puede depender de la forma en la que no tuvo éxito la separación. Por ejemplo, si una parte residual de todos o la mayor parte de los componentes de frecuencia de la fuente sonora permanecen en la señal compuesta a continuación de la separación, el usuario puede oír la fuente sonora en la localización intermedia. Pueden oírse dos fuentes sonoras distintas cuando solo permanecen ciertos componentes de frecuencia (parte del espectro de frecuencia) de la fuente sonora en la señal compuesta, siendo otros componentes de frecuencia exitosamente separados. Como se apreciará, cualquiera de estos efectos puede ser indeseable y, como tal, en ocasiones en las que la separación de la señal de audio no es totalmente exitosa, puede ser beneficioso limitar el intervalo de reposicionamiento espacial que está disponible.

A la vista de este hecho, el aparato de procesamiento de audio 14 se configura para determinar un valor de un parámetro de modificación de la señal separada basándose en una medida determinada del éxito de una separación de una señal de audio que representa una fuente sonora de una señal de audio compuesta, comprendiendo la señal de audio compuesta componentes derivados desde al menos dos fuentes sonoras. El valor del parámetro de modificación de la señal separada (que puede denominarse simplemente como el parámetro de modificación) indica el intervalo para la modificación de una característica de la señal de audio separada que representa la fuente sonora. El intervalo puede corresponder a una cantidad de modificación de la característica de la señal separada más allá de la que la calidad de la señal de audio compuesta modificada (dentro de la que se ha mezclado la señal separada modificada) cae por debajo de un nivel aceptable.

En algunos ejemplos, el parámetro de modificación puede comprender un parámetro de reposicionamiento espacial que indica un intervalo de reposicionamiento espacial para el reposicionamiento espacial de la señal de audio separada. Puesto en otra forma, la característica de la señal separada que ha de modificarse puede ser la posición espacial en el espacio de audio. En otros ejemplos, el parámetro de modificación puede comprender un parámetro de modificación de amplitud que puede indicar un intervalo de modificación de la amplitud de la señal de audio separada. Puesto en otra forma, la característica a ser modificada puede ser la amplitud de la señal de audio

separada. Otros ejemplos de la característica de la señal espacial que pueden modificarse de acuerdo con el éxito de la separación pueden incluir la ecualización, reverberación, distorsión y compresión. Pueden utilizarse los niveles de reverberación aplicados a una señal separada y el volumen de la señal para indicar una distancia de una fuente sonora desde el usuario. Por ejemplo, incrementando la reverberación y disminuyendo el volumen puede dar la impresión de que la fuente sonora está alejada del oyente. A la inversa, disminuyendo la reverberación e incrementando el volumen puede indicar que la fuente sonora está más próxima al oyente. En otros ejemplos más, la característica asociada a la señal separada puede comprender un intervalo de reposicionamiento permitido de la posición del oyente durante la reproducción de audio de punto de vista libre. Como tal, un intervalo permitido de reposicionamiento de la posición del oyente puede depender del éxito de la separación.

Para permitir que se determine el valor del parámetro de modificación, el aparato de procesamiento de audio 14 puede configurarse para determinar la medida del éxito de la separación de la señal de audio que representa la fuente sonora. Sin embargo, en otros ejemplos, la medida del éxito de la separación puede determinarse por otra entidad dentro del sistema y puede proporcionarse al aparato de procesamiento de audio 14, por ejemplo junto con las señales de audio.

El aparato de procesamiento de audio 14 puede configurarse además para limitar una cantidad permitida de modificación de las características de la señal de audio separada basándose en el valor de un parámetro de modificación. De esta forma, puede impedirse la modificación de la señal separada fuera del intervalo indicado por el parámetro de modificación. Esto puede impedir un grado inaceptable de degradación de la señal de audio compuesta modificada.

El aparato de procesamiento de audio 14 puede configurarse además para producir una indicación del valor determinado del parámetro de modificación a ser proporcionado a un usuario, por ejemplo a través de una interfaz de usuario gráfica. La interfaz de usuario gráfica puede configurarse para indicar visualmente al usuario en alguna forma, el valor del parámetro de modificación. Se explican a continuación diversos ejemplos de interfaces de usuario gráficas adecuadas con referencia a las figuras 3A, 3B y 3C.

El aparato de procesamiento de audio 14 puede configurarse de modo que, cuando la medida del éxito indica que el éxito de la separación está por encima de un umbral de grado de éxito, el valor determinado del parámetro de modificación indica que puede realizarse un intervalo completo de modificación de una característica particular de la señal separada. En ejemplos en los que la modificación se refiere al reposicionamiento espacial, el intervalo completo de reposicionamiento espacial puede depender de la configuración del aparato de captura de audio espacial 10. Por ejemplo, si el aparato de captura de audio espacial 10 se configura para capturar audio espacial en los 360 grados que rodean el dispositivo, el intervalo completo de reposicionamiento puede ser 360 grados. Sin embargo, si el aparato de captura de audio espacial 10 se configura para capturar audio espacial desde menos de 360 grados (por ejemplo 180 grados) alrededor del aparato 10, el intervalo completo de reposicionamiento puede estar limitado a esa cantidad.

A la inversa, cuando la medida del éxito indica que el éxito de la separación está por debajo de un umbral de grado de éxito, el aparato de procesamiento de audio 10 puede configurarse de modo que el valor determinado del parámetro de modificación tenga una relación directa con el grado de éxito. Puesto en otra forma, el intervalo de modificación indicado por el valor del parámetro puede incrementarse y disminuirse según se incrementa y disminuye el grado de éxito.

La medida del éxito, en ciertos ejemplos, puede comprender una correlación determinada entre un resto de la señal de audio compuesta y al menos una señal de audio de referencia. La señal de audio de referencia puede ser, en algunos ejemplos, la señal de audio separada. En dichos ejemplos, el aparato de procesamiento de audio 10 puede configurarse así para determinar una correlación entre una parte del resto del audio compuesto correspondiente a la localización original de la señal separada y la señal de audio separada. Una elevada correlación puede indicar que la separación no ha sido particularmente exitosa (un bajo grado de éxito) mientras que una correlación baja (o ninguna) puede indicar que la separación ha tenido éxito (un alto grado de éxito). Se apreciará así que, en dichos ejemplos, la correlación (que es un ejemplo de la media determinada de éxito de la separación) puede tener una relación inversa con el grado de éxito de la separación.

En otros ejemplos, la señal de referencia puede comprender una señal capturada por uno de los dispositivos de grabación adicionales 12A, por ejemplo los dispositivos de grabación adicionales que se asocian a la fuente de audio a la que se asocia la señal separada. Este planteamiento puede ser útil para determinar el éxito de la separación cuando la separación ha dado como resultado un espectro de audio asociado a la fuente sonora que se divide entre el resto de la señal compuesta y la señal separada. Una vez más, la correlación puede tener una relación inversa con el grado de éxito de la separación.

En algunos ejemplos, tanto la correlación entre la señal de audio compuesta y la señal separada como la correlación entre la señal de audio compuesta y la señal derivada del dispositivo de grabación adicional pueden determinarse y utilizarse para determinar el éxito de la separación. Si cualquiera de las correlaciones está por encima de un umbral, puede determinarse que la separación no ha tenido éxito.

La correlación puede determinarse usando la siguiente expresión:

$$\text{Correlación}(\tau, n) = \sum_{k=0}^n R(k)S(k - \tau)$$

5 en la que  $R(k)$  y  $S(k)$  son las  $k$ -ésimas muestras del resto de la señal compuesta y de la señal de referencia respectivamente,  $\tau$  es el tiempo de demora y  $n$  el número total de muestras.

10 El aparato de procesamiento de audio 14 puede configurarse para comparar la correlación determinada con un umbral de correlación predeterminado y, si la correlación está por debajo de la correlación de umbral predeterminada, determinar que la separación ha sido totalmente (o suficientemente) exitosa. A la inversa, si la correlación está por encima de la correlación de umbral predeterminada, el aparato de procesamiento de audio 14 puede configurarse para determinar que la separación no ha sido totalmente (o suficientemente) exitosa o, puesto en otra forma, ha sido solo parcialmente exitosa.

15 Como una alternativa a la expresión mostrada anteriormente, la medida del éxito de la separación, en algunos ejemplos, puede comprender una correlación entre un espectro de frecuencia asociado al resto de la señal de audio compuesta y un espectro de frecuencia asociado a al menos una señal de audio de referencia. Si los componentes de frecuencia de la señal de audio de referencia están también presentes en el resto de la señal de audio compuesta, puede deducirse que la separación no ha sido totalmente exitosa. Por el contrario, si no hay correlación  
20 entre los componentes de frecuencia de la señal de audio separada y el resto de la señal de audio compuesta puede determinarse que la separación ha sido totalmente exitosa. Como se ha descrito anteriormente, la al menos una señal de audio de referencia puede comprender una o ambas de entre la señal de audio separada y una señal derivada de uno de los dispositivos de grabación adicionales.

25 En otros ejemplos, sin embargo, la medida del éxito de la separación puede comprender una correlación entre un resto de la señal de audio compuesta y un componente de una señal de vídeo correspondiente a la señal de audio compuesta. Por ejemplo, en casos en los que la fuente sonora se deriva de una persona hablando, el aparato de procesamiento de audio 14 puede determinar si el resto de la señal de audio compuesta incluye componentes que tienen tiempos que corresponden a movimientos de la boca de la persona desde la que se deriva la fuente de  
30 sonora. Si dichos componentes de audio existen, puede determinarse que la separación no ha sido totalmente exitosa, mientras que si dichos componentes de audio no existen puede determinarse que la separación ha sido totalmente exitosa.

35 Como se apreciará, en todos los ejemplos descritos anteriormente, la correlación determinada tiene una relación inversa con el grado de éxito de la separación.

40 En algunos ejemplos, el aparato de procesamiento de audio 14 puede configurarse para modificar una característica de la señal de audio separada basándose en el valor determinado del parámetro de modificación. Por ejemplo, el aparato de procesamiento de audio 14 puede configurarse para responder a una determinación de que la medida del éxito de la separación indica que, para un marco temporal posterior, un grado de éxito de separación es más bajo que el grado de éxito de separación de un marco temporal actual mediante la modificación de la característica de la señal de audio separada a un valor que esté más próximo a un valor original de la característica de la señal de audio separada. En dichos ejemplos, la modificación de la característica de la señal de audio separada al valor que está más próximo al valor original se realiza previamente al comienzo de la reproducción del marco temporal posterior de  
45 la señal de audio compuesta modificada. La modificación de la característica al valor más próximo al valor original puede realizarse gradualmente de modo que el usuario no experimente un cambio brusco significativo en el valor de la característica en el comienzo de la reproducción del marco temporal posterior de la señal de audio compuesta modificada.

50 Como se entenderá, un marco temporal puede ser un segmento de señal de audio digitalizada  $y(n)$ , por ejemplo,  $y(n)...y(n+M)$ , en la que  $M$  es la longitud de la ventana. Por ejemplo,  $M$  puede ser igual a 2048 muestras o cualquier otro valor adecuado. El tamaño del marco temporal puede predefinirse y puede ser dependiente en algunos ejemplos del tipo o naturaleza de la señal compuesta. Por ejemplo, una señal compuesta que tenga un primer tipo (por ejemplo compuesta de personas hablando) puede analizarse con una primera longitud de marco temporal y una  
55 señal compuesta que tenga un segundo tipo (por ejemplo música) puede analizarse con una segunda longitud de marco temporal. En dichos ejemplos, la primera y segunda longitudes de marco temporal pueden haberse decidido basándose en ensayos sobre qué longitud de marco conduce al mejor éxito de la separación, en promedio, para un tipo de señal particular.

60 La longitud de marco usada durante la separación y la longitud de marco usada durante la reproducción pueden no ser iguales entre sí. Por ejemplo, la separación podría realizarse usando marcos de 2048 muestras de longitud, mientras que la reproducción podría realizarse usando marcos de 512 muestras de longitud.

La figura 2A es un diagrama de flujo que ilustra diversas operaciones que pueden realizarse por el aparato de procesamiento de audio 14 tal como el representado en la figura 1.

- 5 En la operación S201, el aparato de procesamiento de audio 14 recibe una representación de la señal de audio compuesta. Como se ha explicado previamente, la representación puede recibirse en cualquiera de diversos formatos diferentes. Aunque no se representa en la figura 1, dependiendo del formato en el que se recibe la representación, el aparato de procesamiento de audio 14 puede realizar en algunos ejemplos preprocesamiento para reformatear la señal de audio compuesta a otro formato.
- 10 En la operación S202, el aparato de procesamiento de audio 14 realiza la separación de una parte de la señal de audio compuesta que representa una fuente sonora de la señal de audio compuesta. La separación puede realizarse de cualquier manera adecuada. Por ejemplo como se describe en cualquiera de los documentos PCT/EP2016/051709 y WO2014/147442.
- 15 Después de realizar la separación, el aparato de procesamiento de audio 14, en la operación S203, calcula una medida del éxito de la separación de la señal de audio separada de la señal de audio compuesta. Como se ha explicado anteriormente, la medida del éxito puede estar en la forma de una correlación calculada entre el resto de la señal de audio compuesta y cualquiera de al menos una señal de referencia o una parte de un componente de vídeo correspondiente a la señal de audio compuesta. Como se ha explicado anteriormente, la al menos una señal de audio de referencia puede comprender una o ambas de entre la señal de audio separada y una señal derivada desde uno de los dispositivos de grabación adicionales que se asocia a la fuente de audio con la que se relaciona la señal separada.
- 20 Como por supuesto se apreciará, las propiedades de la señal de audio compuesta pueden cambiar a lo largo del tiempo (por ejemplo, pero no exclusivamente, debido al movimiento de las fuentes sonoras dentro del entorno de captura de audio). De esa forma, el éxito con el que puede separarse una fuente sonora de la señal de audio compuesta puede variar a lo largo del tiempo. En consecuencia, la operación S203, así como las operaciones S204 a S207, pueden realizarse para segmentos individuales (o marcos temporales) de la señal de audio compuesta.
- 25 En ejemplos en los que el aparato de procesamiento de audio 14 se configura para calcular la correlación entre el resto de la señal de audio compuesta y la señal de audio de referencia, la correlación puede ser la correlación tanto en el dominio del tiempo como en el dominio de la frecuencia. Cuando la correlación se calcula en el dominio de la frecuencia, el espectro de frecuencia de la señal de audio de referencia puede compararse con un espectro de frecuencia del resto de la señal de audio compuesta.
- 30 En ejemplos en los que el aparato de procesamiento de audio 14 se configura para calcular la correlación entre el resto de la señal de audio compuesta y una parte de un componente de vídeo correspondiente a la señal de audio compuesta esto puede determinarse primero mediante la identificación de una parte del componente de vídeo que corresponde a la localización espacial original de la señal de audio separada. A continuación, se examina el componente de vídeo para determinar si hay cualesquiera características presentes en la parte del componente de vídeo que están sincronizadas en el tiempo con componentes del resto de la señal de audio compuesta. Por ejemplo, el aparato de procesamiento de audio 14 puede determinar si el movimiento de una boca de la persona está sincronizado con componentes de audio del resto de la señal de audio compuesta.
- 35 Independientemente de qué correlación se determine por el aparato de procesamiento de audio 14, un alto grado de correlación puede indicar un bajo grado de éxito de la separación, mientras que un bajo grado de correlación puede indicar un alto grado de éxito de la separación. Puesto en otra forma, puede existir una relación inversa entre la correlación calculada y el grado de éxito de la separación.
- 40 Después de calcular la medida del éxito de la separación, el aparato de procesamiento de audio 14 puede proseguir a la operación S204 en la que determina el valor del parámetro de modificación de la señal separada, que indica un intervalo para la modificación de una característica de la señal de audio separada. Por ejemplo, en algunos ejemplos, el valor del parámetro de modificación puede comprender un valor máximo con el que puede modificarse una característica sin degradar una calidad de la señal de audio compuesta modificada más allá de un nivel aceptable. En otros ejemplos, sin embargo, el valor del parámetro de modificación puede comprender un intervalo permitido de modificación que puede realizarse sin degradar una calidad de la señal de audio compuesta modificada más allá del nivel aceptable. Como se ha explicado anteriormente, el grado de modificación indicado por el valor del parámetro de modificación puede tener una relación directa con el grado de éxito de la separación y una relación inversa con la correlación calculada.
- 45 Ejemplos de diversas sub-operaciones que pueden constituir la operación S204 se ilustran en, y se explican con referencia a, el diagrama de flujo de la figura 2B.
- 50 En la operación S204-1, el aparato de procesamiento de audio 14 puede determinar si la medida del éxito de la separación (tal como se determina en la operación S203) indica que el grado de éxito está por encima de un umbral de éxito. En algunos ejemplos, esta operación puede comprender comparar la correlación calculada con un umbral
- 55
- 60
- 65

de correlación. En dichos ejemplos, si la correlación calculada está por encima de un umbral de correlación, puede determinarse que el grado de éxito está por debajo del umbral de éxito. A la inversa, si se determina que la correlación calculada está por debajo del umbral de correlación, puede determinarse que el grado de éxito de la separación está por encima de un umbral de éxito.

5 Si, en la operación S204-1, se determina que el éxito de la separación está por encima del umbral de éxito, el aparato de procesamiento de audio 14 puede proseguir a la operación S204-2 en la que se determina que la separación fue suficientemente exitosa y como tal que el valor del parámetro de modificación es para indicar que puede realizarse un intervalo completo de modificación. El grado de la modificación que corresponde al "intervalo completo" puede programarse previamente en el aparato de procesamiento de audio 14.

15 A la inversa, si, en la operación S204-1, se determina que el éxito de la separación está por debajo del umbral de éxito, el aparato de procesamiento de audio 14 puede proseguir a la operación S204-3 en la que se determina que la separación no fue suficientemente exitosa y de ese modo puede determinar el valor del parámetro de modificación dependiendo del grado de éxito. Por ejemplo, cuando el grado de éxito está por debajo del umbral, el valor del parámetro de modificación puede indicar un intervalo mayor de modificación para un grado más alto de éxito y puede indicar un intervalo más pequeño de modificación para un grado menor de éxito.

20 Volviendo ahora a la figura 2A, en la operación S205, el aparato de procesamiento de audio 14 puede hacer que el valor del parámetro de modificación sea indicado a través de una interfaz de usuario gráfica a un usuario. Esto puede permitir al usuario determinar el intervalo de modificación que puede realizarse sin degradar la calidad de la señal compuesta modificada más allá de un nivel aceptable.

25 En la operación S206, el aparato de procesamiento de audio 14 puede imponer un límite sobre la cantidad de modificación que puede realizarse con respecto a la señal de audio separada. Como tal, el aparato de procesamiento de audio 14 puede configurarse para impedir la modificación de la característica más allá del intervalo indicado por el valor del parámetro de modificación. De esta forma, un usuario puede ser capaz de modificar solamente la característica, por ejemplo a través de la interfaz de usuario gráfica, dentro de un intervalo permitido.

30 En la operación S207, el aparato de procesamiento de audio 14 puede configurarse para realizar una modificación de la característica de la señal de audio separada. La modificación puede realizarse con respecto al marco temporal con el que se relaciona el grado de éxito espacial. La modificación puede realizarse en respuesta a una entrada por el usuario indicando un grado deseado de modificación. A la vista del límite impuesto sobre el grado de la modificación permitida, la modificación puede limitarse basándose en el valor del parámetro de modificación. De ese modo, en algunos ejemplos, si el usuario indica una modificación deseada que está fuera del intervalo permitido, el aparato de procesamiento de audio 14 puede responder modificando la característica en un grado máximo indicado por el valor del parámetro de modificación incluso aunque este sea menor que la modificación deseada.

40 La figura 2C es un diagrama de flujo que ilustra diversas otras operaciones que pueden realizarse por el aparato de procesamiento de audio 14 tal como el representado en la figura 1. Las operaciones ilustradas en la figura 2C pueden realizarse posteriormente a la realización de la operación S207 y pueden realizarse con respecto a un marco temporal de la señal de audio compuesta que es posterior en el tiempo al marco temporal con respecto al que se realizaron las operaciones S203 a S207 de la figura 2A.

45 En la operación S208, puede determinarse la medida del éxito de la separación de la señal de audio respecto al marco temporal posterior de la señal de audio compuesta. Esto puede realizarse en cualquiera de las formas descritas con referencia a la operación S203.

50 A continuación, en la operación S209, el aparato de procesamiento de audio 14 determina un valor del parámetro de modificación para el marco temporal posterior de la señal de audio compuesta. Esto puede realizarse tal como se ha descrito con relación a la operación S204 en las figuras 2A y 2B.

55 En la operación S210, el valor del parámetro de modificación para la parte posterior puede indicarse al usuario a través de una interfaz de usuario gráfica (ejemplos de las cuales pueden explicarse con más detalle con referencia las figuras 3A, 3B y 3C).

60 En la operación S211, el aparato de procesamiento de audio 14 determina si un grado de modificación de la característica para el marco temporal precedente supera el umbral indicado por el valor del parámetro de modificación para el marco temporal posterior (que se determinó en la operación S209).

65 Si se alcanza una determinación positiva en la operación S211, el aparato de procesamiento de audio 14 prosigue a la operación S212. En la operación S212, el aparato de procesamiento de audio 14, durante la reproducción del marco temporal precedente de la señal de audio compuesta modificada, hace que el grado de modificación de la característica de la señal separada sea reducido a un nivel que esté dentro del intervalo indicado por el valor del parámetro de modificación para el marco temporal posterior. Puesto en otra forma, la realización de la operación

5 S212 puede ser previa al inicio de la reproducción del marco temporal posterior de la señal de audio separada. La modificación al nivel reducido puede realizarse gradualmente tal y como se va reproduciendo la parte anterior. En esta forma, el usuario no experimenta un salto brusco significativo en el valor de la característica modificada. Después de la realización de la operación S211, el aparato de procesamiento de audio 14 puede proseguir a la operación S212.

10 Si se determina en la operación S211 que el grado de modificación de la característica para el marco temporal precedente no excede el umbral indicado por el valor del parámetro de modificación para el marco temporal posterior, el aparato de procesamiento de audio 14 prosigue a la operación S212.

15 En la operación S213, durante la reproducción del marco temporal posterior de la señal de audio compuesta modificada, el aparato de procesamiento de audio 14 impone un límite sobre la modificación permitida. Esto puede ser tal como se ha descrito con referencia a la operación S206.

20 En la operación S214, si, por ejemplo, se recibe una entrada del usuario indicando otra modificación de la característica, el aparato de procesamiento de audio 14 puede responder mediante la modificación de la característica en consecuencia. Esto puede realizarse tal como se ha descrito con referencia a la operación S207. Como se apreciará, si no se recibe una entrada requiriendo la modificación de la característica, puede saltarse la operación S214.

Posteriormente, el aparato de procesamiento de audio 14 vuelve a la operación S208 en la que se determina la medida del éxito de la separación para un marco temporal posterior de la señal de audio compuesta recibida.

25 Como naturalmente se apreciará, las operaciones representadas en las figuras 2A a 2C son solamente ejemplos. Como tales, las operaciones pueden realizarse en un orden diferente, pueden omitirse ciertas operaciones y/o pueden realizarse operaciones adicionales. Por ejemplo, aunque se han descrito diversas determinaciones como realizadas de una forma marco por marco, en otros ejemplos, una medida del éxito de la separación puede determinarse sobre un período extendido, determinándose los marcos temporales utilizados para las finalidades de las operaciones S211 a S214 basándose en la medida del éxito de la separación. En dichos ejemplos, cada marco temporal puede seleccionarse de modo que dentro del marco temporal la medida del éxito de la separación sea relativamente uniforme, correspondiendo los límites entre marcos temporales a momentos en los que hay un cambio significativo (por ejemplo un cambio que es mayor que un umbral) en la medida del éxito de la separación.

30 La figura 3A es un ejemplo de una interfaz de usuario gráfica (GUI) 30 a través de la que puede indicarse al usuario el parámetro de modificación para uno o más marcos temporales de la señal de audio compuesta.

35 La GUI 30, en el ejemplo de la figura 3A, incluye uno o más indicadores 301A-F, correspondiendo cada uno a un marco temporal diferente de la señal de audio compuesta. Los indicadores 301 se configuran para indicar el valor del parámetro de modificación que se determina para cada marco de la señal, para de ese modo indicar un grado de modificación permitido.

40 En algunos ejemplos, tales como el de la figura 3A, los indicadores 301 pueden indicar adicionalmente una duración del marco temporal. En el ejemplo de la figura 3A, una primera dimensión L (por ejemplo longitud) de los indicadores 301A-F indica la duración de cada marco temporal. Más específicamente, una primera dimensión más larga indica un marco temporal con una duración más larga. En el ejemplo de la figura 3A, los indicadores se proporcionan sobre una línea de tiempos, de modo que los marcos temporales correspondientes a las últimas posiciones de la señal compuesta entrante se proporcionan más adelante a lo largo de la línea de tiempos de lo que están los marcos temporales correspondientes a partes anteriores de la señal compuesta entrante.

45 Una segunda dimensión H (por ejemplo altura) de los indicadores puede indicar el valor del parámetro de modificación, de modo que una altura mayor indica un grado mayor de modificación permitida para el marco temporal. Por ejemplo, en la figura 3A las alturas de los indicadores disminuyen sucesivamente de la correspondiente al primer marco temporal hasta la correspondiente al cuarto marco temporal. Esto puede indicar que el valor del parámetro de modificación disminuye sucesivamente desde el primer al cuarto marcos temporales y consecuentemente que el grado permitido de modificación también disminuye desde el primer al cuarto marcos temporales.

50 En algunos casos, tal como el de la figura 3A, los indicadores 301A-F pueden indicar valores de dos parámetros de modificación diferentes. En dichos ejemplos, una tercera dimensión D (por ejemplo profundidad) de los indicadores 301A-F puede indicar un valor del segundo parámetro de modificación. Por ejemplo, en el caso de la figura 3A, el (los) parámetro(s) de modificación son parámetros de reposicionamiento espacial, correspondiendo un primer parámetro al reposicionamiento espacial azimutal y correspondiendo un segundo parámetro al reposicionamiento espacial de elevación. En el ejemplo de la figura 3A, el valor del parámetro de reposicionamiento espacial azimutal se indica por la profundidad y el indicador del valor del parámetro de reposicionamiento espacial de elevación se indica por la altura de los indicadores.

Las figuras 3B y 3C ilustran ejemplos de otros aspectos de la GUI 32, 34 través de los que puede indicarse al usuario un valor del parámetro de modificación para uno o más de los marcos de la señal de audio compuesta.

5 En estos ejemplos, las GUI 32, 34 incluyen un elemento móvil 322, 342, cuya localización indica el grado actual de modificación de la característica (por ejemplo posición espacial) que se aplica.

10 Cada GUI 32, 34 puede incluir además al menos una primera zona 324, 344 delineada que indica un intervalo de modificación que está "permitido" (indicando de ese modo el valor del parámetro de modificación). La GUI 32, 34 puede incluir también una segunda zona 326, 346 que indique grados de modificación fuera del intervalo "permitido". Las dos zonas pueden ser visualmente distintas entre sí (por ejemplo, usando diferentes colores, por ejemplo, verde y rojo). Las GUI 32, 34 pueden incluir adicionalmente demarcaciones 328, 348 que indican el grado de modificación en términos cuantitativos.

15 La GUI 32 de la figura 3B se configura para indicar la modificación en justamente una dimensión (por ejemplo, en donde la modificación se refiere al posicionamiento espacial, solamente el azimut). La GUI 34 de la figura 3C, por otro lado, se configura para indicar la modificación en dos dimensiones (por ejemplo azimut y elevación) en donde la localización del elemento móvil 342 en cualquiera de las direcciones x e y corresponde a la modificación en una dimensión diferente. Como naturalmente se apreciará, pueden proporcionarse dos (o tres) GUI tal como la de la figura 3B en tándem para indicar de ese modo la modificación en dos (o tres) dimensiones.

20 En algunos ejemplos, las GUI 32, 34 pueden visualizarse sobre una interfaz táctil, mediante la que el usuario proporciona entradas táctiles para mover el elemento móvil 322, 342 y de ese modo modificar la característica de la señal separada. En otros ejemplos, sin embargo, las GUI puede ser utilizables con dispositivos de entrada mecánica tales como deslizadores mecánicos o conmutadores/palancas de fuego mecánicas 32, 34, en las que puede hacerse que se mueva el elemento móvil a través de un deslizador, conmutador etc. En dichos ejemplos, los actuadores pueden utilizarse para proporcionar realimentación inercial a los dispositivos mecánicos, para de ese modo impedir o desalentar la modificación de la característica más allá del intervalo "permitido" indicado. En otros ejemplos, la realimentación física puede utilizarse con dispositivos de control mecánico (por ejemplo, deslizadores, conmutadores, palancas de juego, etc.) para indicar el valor del parámetro de modificación (particularmente cuando el usuario está tratando de superar el intervalo de modificación indicado por el parámetro de modificación) en ausencia de las GUI 32, 34.

35 Aunque no se muestra en los ejemplos de las figuras 3A a 3C, se apreciará que puede visualizarse otra información para el usuario a través de la GUI 30, 32, 34. Por ejemplo, puede indicarse un nivel de modificación actual (o pretendido) para uno o más de los marcos temporales con relación a los indicadores correspondientes a esos marcos temporales. Los indicadores 301A-F pueden indicar también o alternativamente diferentes intervalos de modificación para cada marco temporal basándose en la degradación de la calidad de la señal compuesta modificada que se asocia a diferentes intervalos. Por ejemplo, los indicadores pueden indicar un primer intervalo en el que la degradación en la calidad sería baja, un segundo intervalo en el que la degradación de la calidad sería más alta pero aún aceptable y un tercer intervalo en el que la degradación de la calidad sería inaceptable. Los diferentes intervalos pueden indicarse por ejemplo usando diferentes colores (por ejemplo verde, amarillo y rojo).

45 Aunque no se muestra en el ejemplo de las figuras 3A a 3C, las GUI 30, 32, 34 pueden incluir una función para permitir al usuario previsualizar la señal de audio compuesta modificada, por ejemplo en combinación con una versión correspondientemente modificada de una señal derivada de uno de los dispositivos de captura de audio adicionales que corresponde a la fuente sonora separada. En esta forma, el usuario puede ser capaz de verificar la calidad de la señal compuesta modificada antes de confirmar las modificaciones a través de la GUI.

50 Como se apreciará, el reposicionamiento de fuentes sonoras puede realizarse en una, dos o tres dimensiones. El reposicionamiento puede realizarse en un sistema de coordenadas cartesianas con ejes **x**, **y** y **z**, o en un sistema de coordenadas polares con azimut, elevación y distancia. Las GUI pueden así configurarse dependiendo del número de dimensiones (y del sistema de coordenadas) en el que ha de realizarse el posicionamiento.

55 Con referencia ahora a las figuras 4A a 4C, estas figuras sirven para ilustrar la forma en la que puede determinarse el valor de un parámetro de reposicionamiento espacial sobre la base del éxito de una separación de una señal de audio compuesta.

60 La figura 4A ilustra dos fuentes sonoras (en este ejemplo, dos personas 13A, 13B hablando) en diferentes posiciones espaciales con relación a la localización del dispositivo de captura de audio espacial 10 (que puede ser también la localización del oyente cuando se está reproduciendo el audio).

65 Un primer orador 13A se localiza en un ángulo azimutal de -45 grados que está a la izquierda del dispositivo de captura/oyente y un segundo orador 13B se localiza en un ángulo azimutal de +45 grados que está a la derecha del dispositivo de captura/oyente.

Los espectros de frecuencia 40A, 40B de las señales de voz (fuentes sonoras) de cada orador se han representado

en sus posiciones espaciales relativas. El espectro de frecuencia describe la distribución de frecuencia de la señal de voz/fuente sonora. Como se ha explicado anteriormente, sin embargo, se debería apreciar que el espectro de frecuencia varía a lo largo del tiempo y, como tal, la figura 4A representa una situación instantánea en un marco temporal de tiempo corto, por ejemplo una duración de 20 milisegundos.

5 La figura 4B ilustra una separación totalmente exitosa del espectro de frecuencia de la señal de audio compuesta. En este ejemplo, esto se indica por el hecho de que ninguno de los componentes de la señal derivada de la fuente sonora permanece en la localización original.

10 En dicha situación, el aparato de procesamiento de audio 14 puede determinar que el grado de éxito está por encima del umbral de éxito y de ese modo puede establecer el valor del parámetro de reposicionamiento espacial para indicar que puede realizarse el intervalo completo de reposicionamiento espacial. En este ejemplo, el intervalo completo de reposicionamiento es 360 grados y así se indica esto por el parámetro de reposicionamiento espacial.

15 Como puede verse, en este ejemplo, la fuente sonora correspondiente al primer orador 13A (indicado por el espectro de frecuencia 40A) se ha reposicionado dentro del intervalo permitido en de menos 135 grados a menos 180 grados que está por detrás del aparato de captura/oyente.

20 A diferencia de la figura 4B, la figura 4C ilustra una situación en la que la separación no ha sido totalmente exitosa. Esto se indica en la figura 4C por varios componentes 40A-1 del espectro de frecuencia 40A del primer orador 13A que están a la izquierda en su localización original mientras otros componentes 40A-2 se han separado.

25 En un ejemplo tal como el ilustrado en la figura 4C, el aparato de procesamiento de audio 14 determina que la separación no ha sido exitosa. Como tal, el aparato de procesamiento de audio 14 determina un valor del parámetro de reposicionamiento espacial basado en el grado de éxito de la separación. La determinación del valor del parámetro de reposicionamiento espacial puede ser tal que un grado más alto de éxito dé como resultado un parámetro de reposicionamiento espacial que tenga un valor que indique un intervalo más alto de reposicionamiento espacial y un grado más bajo de éxito dé como resultado un reposicionamiento espacial del parámetro que tenga un valor que indique un intervalo más bajo de reposicionamiento espacial.

30 En el ejemplo de la figura 4C, el valor del parámetro de reposicionamiento espacial indica que la fuente sonora separada puede reposicionarse en  $\pm 90$  grados respecto a su localización original. A la vista de esto, la señal separada 40A-2 se ha reposicionado dentro del intervalo indicado por el parámetro de reposicionamiento espacial en -80 grados. De ese modo, la calidad de la señal de audio compuesta modificada resultante no se degrada más allá  
35 de un nivel aceptable.

40 En los ejemplos anteriores descritos con referencia a las figuras 1 a 4C, la señal compuesta a partir de la que se han separado las fuentes sonoras identificadas se genera mediante aparatos de captura de audio espacial 10. Sin embargo, se apreciará naturalmente que los métodos y operaciones descritas en el presente documento pueden realizarse con respecto a cualquier señal de audio que incluya componentes derivados de una pluralidad de fuentes de audio como por ejemplo una señal derivada de los dispositivos de captura de audio adicionales que casualmente incluyan componentes de dos oradores (por ejemplo debido a que ambos oradores están en una proximidad suficientemente próxima para el dispositivo de captura).

45 Aunque los ejemplos anteriores se han explicado principalmente con referencia a la modificación de características de una señal de audio separada, debería apreciarse que diversas operaciones descritas en el presente documento pueden aplicarse a señales que comprendan tanto componentes de audio como visuales (AV). Por ejemplo, el reposicionamiento espacial puede aplicarse a las partes del componente visual de la señal AV. Por ejemplo, el  
50 aparato de procesamiento de audio 14 puede configurarse para identificar y reposicionar un objeto visual en componentes visuales correspondientes a la fuente sonora separada. Más específicamente, el aparato de procesamiento de audio 14 puede configurarse para segmentar (o separar) el objeto visual correspondiente a la fuente sonora separada del resto del componentes de vídeo y sustituir el fondo. El aparato de procesamiento de audio 14 puede configurarse posteriormente para permitir el reposicionamiento del objeto visual separado basándose en el parámetro de reposicionamiento espacial determinado para la señal de audio separada.

55 La figura 5 es un diagrama de bloques esquemático que ilustra un ejemplo de configuración del aparato de procesamiento de audio 14 descrito con referencia las figuras 1 a 4C.

60 El aparato de procesamiento de audio 14 comprende un aparato de control 50 que se configura para realizar diversas operaciones tales como se ha descrito anteriormente con referencia al aparato de procesamiento de audio 14. El aparato de control 50 puede configurarse además para controlar los otros componentes del aparato de procesamiento de audio 14.

65 El aparato de procesamiento de audio 14 puede comprender además una interfaz de entrada de datos 51, a través de la que pueden recibirse señales representativas de la señal de audio compuesta. Las señales derivadas desde los uno o más dispositivos de captura de audio adicionales 12A-C pueden recibirse también a través de la interfaz de

entrada de datos 51. La interfaz de entrada de datos 51 puede ser cualquier tipo adecuado de interfaz por cable o inalámbrica. Los datos representativos de los componentes visuales capturados por el aparato de captura de audio espacial 10 pueden recibirse también a través de la interfaz de entrada de datos 51.

5 El aparato de procesamiento de audio 14 puede comprender además una interfaz de salida visual 52, que puede acoplarse a una pantalla 53. El aparato de control 50 puede hacer que la información indicativa del valor del parámetro de modificación de la señal separada se proporcione al usuario a través de la interfaz de salida visual 52 y la pantalla 53. El aparato de control 50 puede hacer adicionalmente que se visualicen para el usuario una GUI 30, 32, 34 como la descrita con referencia las figuras 3A, 3B y 3C. Puede hacerse también que los componentes de  
10 vídeo que corresponden a las señales de audio se visualicen a través de la interfaz de salida visual 52 y la pantalla 53.

El aparato de procesamiento de audio 14 puede comprender además una interfaz de entrada de usuario 54 a través de la que pueden proporcionarse entradas del usuario al aparato de procesamiento de audio 14 por parte del usuario  
15 del aparato.

El aparato de procesamiento de audio 14 puede comprender adicionalmente una interfaz de salida de audio 55 a través de la que puede proporcionarse audio al usuario, por ejemplo a través de una disposición de altavoz o unos auriculares binaurales 56. Por ejemplo, las señales de audio compuestas modificadas pueden proporcionarse al  
20 usuario a través de la interfaz de salida de audio 55.

Se describirán ahora algunos detalles adicionales de componentes y características del aparato de procesamiento de audio 14 anteriormente descrito y alternativas para ellos, principalmente con referencia a la figura 5.

25 El aparato de control 51 puede comprender circuitos de procesamiento 510 acoplados comunicativamente con la memoria 511. La memoria 511 tiene instrucciones legibles por ordenador 511A almacenadas en ella, que cuando se ejecutan por el circuito de procesamiento 510 hacen que el circuito de procesamiento 510 provoque la realización de varias de las operaciones anteriormente descritas con referencia a las figuras 1 a 5. Puede hacerse referencia al aparato de control 51 en algunos casos, en términos generales, como el "aparato".  
30

Los circuitos de procesamiento de audio 510 de cualquiera de los aparatos de procesamiento de audio 14 descritos con referencia a las figuras 1 a 5 pueden ser de cualquier composición adecuada y pueden incluir uno o más procesadores 510A de cualquier tipo adecuado o combinación adecuada de tipos. Por ejemplo, el circuito de procesamiento 510 puede ser un procesador programable que interprete instrucciones de programa de ordenador 511A y procese datos. El circuito de procesamiento 510 puede incluir una pluralidad de procesadores programables. Alternativamente, el circuito de procesamiento 510 puede ser, por ejemplo, hardware programable con firmware embebido. El circuito de procesamiento 510 puede denominarse medios de procesamiento. El circuito de procesamiento 510 puede incluir alternativa o adicionalmente uno o más circuitos integrados de aplicación específica (ASIC). En algunos casos, el circuito de procesamiento 510 puede denominarse como un aparato de cálculo.  
35

El circuito de procesamiento 510 se acopla a la memoria respectiva (o uno o más dispositivos de almacenamiento) 511 y es operativo para leer/escribir datos a/desde la memoria 511. La memoria 511 puede comprender una única unidad de memoria o una pluralidad de unidades de memoria, sobre la que se almacenan instrucciones (o códigos) legibles por ordenador 511A. Por ejemplo, la memoria 511 puede comprender tanto memoria volátil 511-2 como memoria no volátil 511-1. Por ejemplo, las instrucciones legibles por ordenador 511A pueden almacenarse en la memoria no volátil 511-1 y pueden ejecutarse por el circuito de procesamiento 510 usando la memoria volátil 511-2 para almacenamiento temporal de datos o datos e instrucciones. Ejemplos de memoria volátil incluyen RAM, DDRAM y SRAM, etc. Ejemplos de memoria no volátil incluyen ROM, PROM, EEPROM, memoria flash, almacenamiento óptico, almacenamiento magnético, etc. Puede hacerse referencia a las memorias en general como  
40 medios de memoria legibles por ordenador no transitorios.

El término "memoria", además de cubrir la memoria que comprende tanto memoria no volátil como memoria volátil, puede cubrir también una o más memorias volátiles solamente, una o más memorias no volátiles solamente, o una o más memorias volátiles y una o más memorias no volátiles.  
45

Las instrucciones legibles por ordenador 511A pueden preprogramarse en el aparato de procesamiento de audio 14. Alternativamente, las instrucciones legibles por ordenador 511A pueden llegar al aparato 14 a través de una señal electromagnética portadora o pueden copiarse desde una entidad física 57 (véase la figura 5) tal como un producto de programa informático, un dispositivo de memoria o un medio de registro tal como un CD-ROM o DVD. Las instrucciones legibles por ordenador 511A pueden proporcionar la lógica y rutinas que permitan que el aparato de procesamiento de audio 14 realice la funcionalidad descrita anteriormente. La combinación de instrucciones legibles por ordenador almacenadas en memoria (de cualquiera de los tipos descritos anteriormente) puede denominarse como un producto de programa informático.  
50

65 Donde es aplicable, la capacidad de comunicación inalámbrica del aparato 10, 12, 14 puede proporcionarse mediante un único circuito integrado. Puede proporcionarse alternativamente mediante un conjunto de circuitos

integrados (es decir un conjunto de chips). La capacidad de comunicación inalámbrica puede ser alternativamente un circuito integrado de aplicación específica (ASIC) cableado.

5 Como se apreciará, los aparatos 10, 12, 14 descritos en el presente documento pueden incluir diversos componentes de hardware que pueden no haberse mostrado en las figuras. Por ejemplo, el aparato de procesamiento de audio 14 puede comprender en algunas implementaciones un dispositivo informático portátil tal como un teléfono móvil o un ordenador tableta y de ese modo puede contener componentes comúnmente incluidos en un dispositivo del tipo específico. De modo similar, el aparato de procesamiento de audio 14 puede comprender componentes de software opcionales adicionales que no se han descrito en la presente especificación dado que pueden no ser relevantes para los principios y conceptos principales descritos en el presente documento.

15 Los ejemplos descritos en el presente documento pueden implementarse en software, hardware, lógica de aplicación o una combinación de software, hardware y lógica de aplicación. El software, lógica de aplicación y/o hardware puede residir en una memoria, o cualquier medio informático. En una realización de ejemplo, la lógica de aplicación, software o conjunto de instrucciones se mantiene en uno cualquiera de diversos medios legibles por ordenador convencionales. En el contexto del presente documento, una "memoria" o un "medio legible por ordenador" puede ser cualquier medio o medios que puedan contener, almacenar, comunicar, propagar o transportar las instrucciones para su uso por, o en combinación con, un sistema, aparato o dispositivo de ejecución de instrucciones, tal como un ordenador.

20 La referencia, donde sea relevante, a un "medio de almacenamiento legible por ordenador", "producto de programa informático", "programa informático tangiblemente realizado", etc., o a un "procesador" o "circuito de procesamiento" etc. debería entenderse que engloba no solamente ordenadores que tengan diferentes arquitecturas tales como arquitecturas de procesador simple/multiprocesador y arquitecturas de secuencia/paralelas, sino también circuitos especializados tales como matrices de puertas programables en campo FPGA, circuitos de aplicación específica ASIC, dispositivos de procesamiento de señal y otros dispositivos. Las referencias a programas, instrucciones, código informático, etc. debería entenderse que expresa software para un firmware de procesador programable tal como el contenido programable de un dispositivo de hardware tales como instrucciones para un procesador configuradas o ajustes de configuración para un dispositivo de función fija, matrices de puertas, dispositivo lógico programable, etc.

35 Tal como se usa en la presente solicitud, el término "circuito" se refiere a todos los siguientes: (a) implementaciones de circuitos solo en hardware (tales como implementaciones solo en circuitos analógicos y/o digitales) y (b) a combinaciones de circuitos y software (y/o firmware), tal como (según sea aplicable): (i) a una combinación de procesador(es) o (ii) a partes de procesador(es)/software (incluyendo procesador(es) de señal digital), software, y memorias que trabajan juntos para hacer que un aparato, tal como un teléfono móvil o servidor, realice diversas funciones) y (c) a circuitos, tales como un(os) microprocesador(es) o una parte de un(os) microprocesador(es), que requieren software o firmware para su operación, incluso si el software o firmware no está físicamente presente.

40 Esta definición de "circuitos" se aplica a todos los usos de este término en la presente solicitud, incluyendo en cualquiera de las reivindicaciones. Como un ejemplo adicional, tal como se usa en la presente solicitud, el término "circuito" debería cubrir también una implementación de meramente un procesador (o múltiples procesadores) o una parte de un procesador y su(s) software y/o firmware adjunto. El término "circuitos" debería cubrir también, por ejemplo y si es aplicable a un elemento de reivindicación particular, un circuito integrado en banda base o aplicaciones de circuitos integrados en un procesador para un teléfono móvil o un circuito integrado similar en un servidor, dispositivos de red celular, u otro dispositivo de red.

50 Si se desea, las diferentes funciones explicadas en el presente documento pueden realizarse en un orden diferente y/o simultáneamente entre sí. Adicionalmente, si se desea, una o más de las funciones anteriormente descritas pueden ser opcionales o pueden combinarse. De modo similar, se apreciará que los diagramas de flujo de las figuras 2A a 2C son ejemplos solamente y que pueden omitirse, reordenarse y/o combinarse diversas operaciones representadas en ellos,

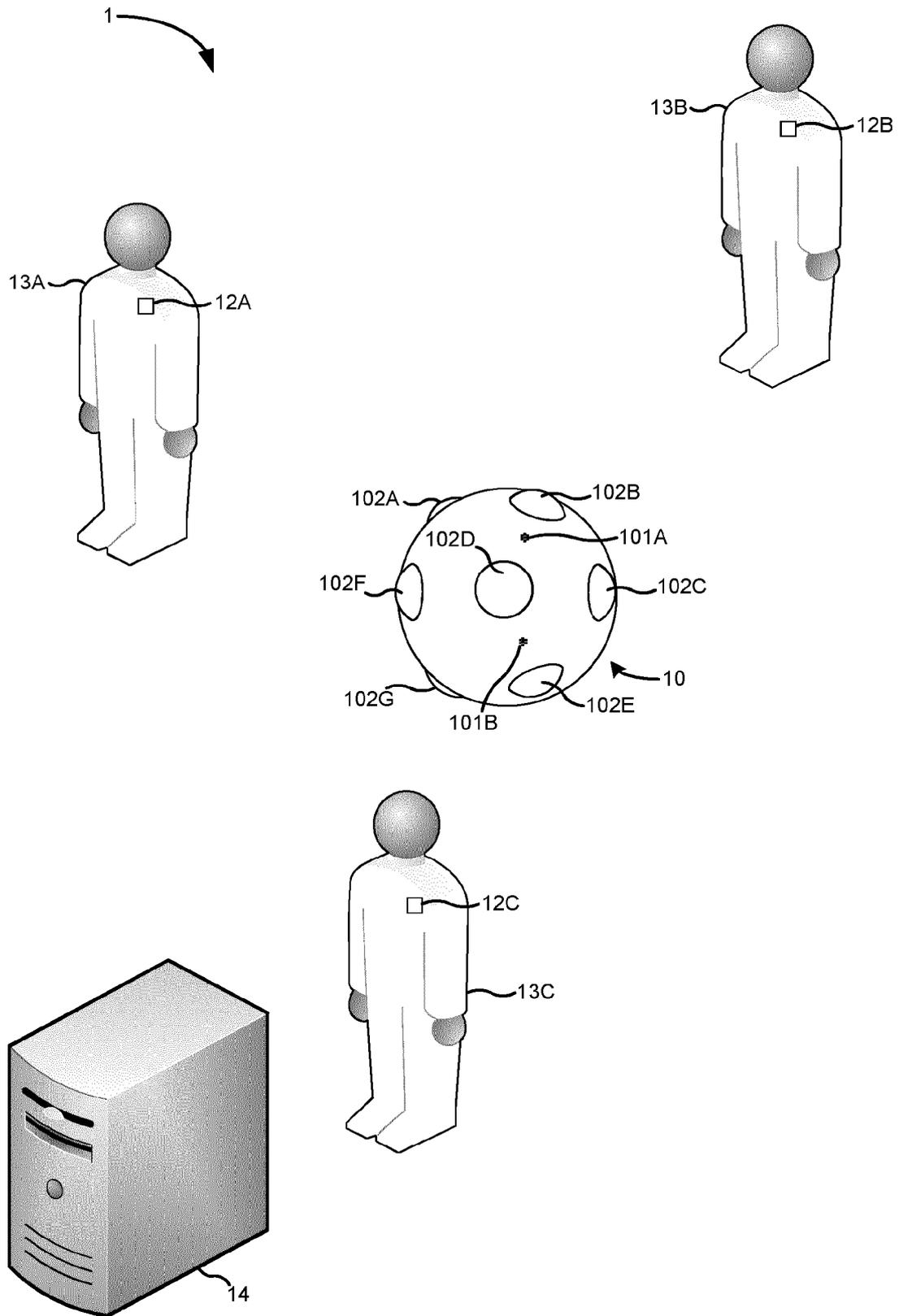
55 Aunque diversos aspectos se exponen en las reivindicaciones independientes, otros aspectos comprenden otras combinaciones de características a partir de las realizaciones descritas y/o las reivindicaciones dependientes con las características de las reivindicaciones independientes, y no solamente las combinaciones explícitamente expuestas en las reivindicaciones. Se señala también en el presente documento que mientras que lo anterior describe diversos ejemplos, estas descripciones no deberían verse en un sentido limitativo. Por el contrario, hay varias variaciones y modificaciones que pueden realizarse sin apartarse del alcance de la presente invención tal como se define en las reivindicaciones adjuntas.

60

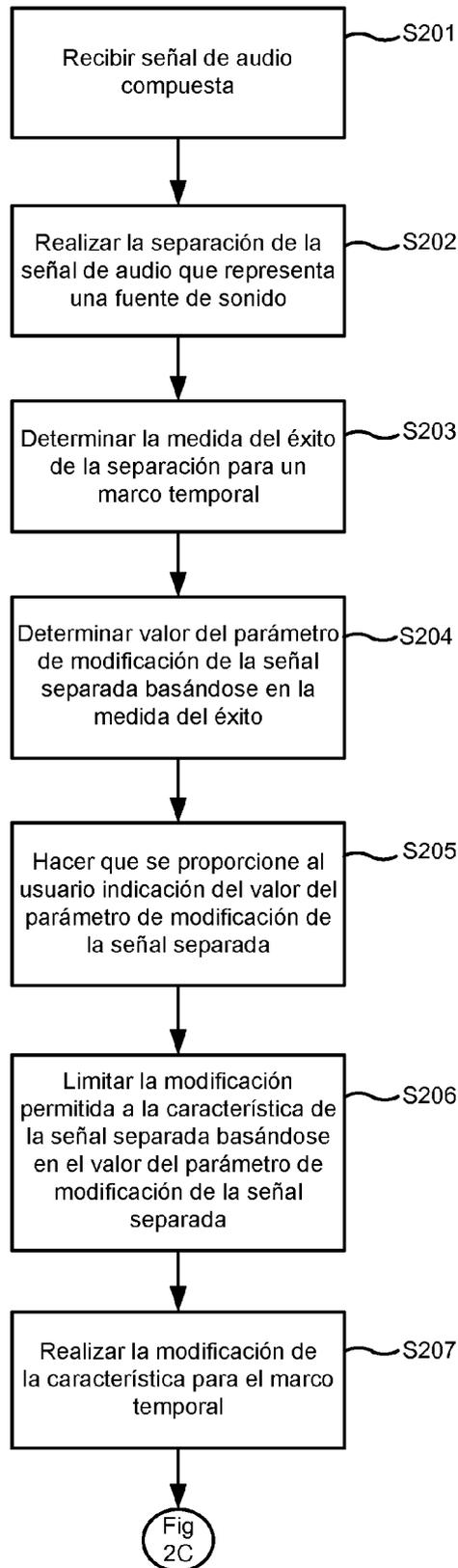
**REIVINDICACIONES**

1. Un método que comprende:  
 5 determinar, basándose en una medida determinada del éxito de una separación de una señal de audio que representa una fuente sonora de una señal de audio compuesta que comprende componentes derivados de al menos dos fuentes sonoras, un valor de un parámetro de modificación de la señal separada, indicando el valor del parámetro de modificación de la señal separada un intervalo de modificación de una característica asociada a la señal de audio separada.
- 10 2. El método de acuerdo con la reivindicación 1, en el que el parámetro de modificación de la señal separada es un parámetro de reposicionamiento espacial que indica un intervalo de reposicionamiento espacial para el reposicionamiento espacial de la señal de audio separada.
- 15 3. El método de acuerdo con la reivindicación 1 o la reivindicación 2, que comprende determinar la medida del éxito de la separación de la señal de audio de la señal de audio compuesta.
- 20 4. El método de acuerdo con cualquiera de las reivindicaciones 1 a 3, que comprende: limitar la cantidad permitida de modificación de la característica asociada a la señal de audio separada basándose en el valor del parámetro de modificación de la señal separada.
- 25 5. El método de acuerdo con cualquier reivindicación anterior, que comprende: hacer que se proporcione a un usuario una indicación del valor determinado del parámetro de modificación de la señal separada.
- 30 6. El método de acuerdo con cualquier reivindicación anterior, que comprende: cuando la medida del éxito indica que el éxito de la separación está por encima de un grado de umbral de éxito, determinar un valor del parámetro de modificación de la señal separada que indique un intervalo completo de modificación de la característica.
- 35 7. El método de acuerdo con cualquier reivindicación anterior en el que, cuando la medida del éxito indica que el éxito de la separación está por debajo de un grado de umbral de éxito, el valor determinado del parámetro de modificación de la señal separada indica un intervalo de modificación que tiene una relación directa con el grado de éxito.
- 40 8. El método de acuerdo con cualquier reivindicación anterior, en el que la medida del éxito comprende una correlación entre un resto de la señal de audio compuesta y al menos una señal de audio de referencia.
- 45 9. El método de acuerdo con la reivindicación 8, que comprende:  
 si la correlación está por debajo de una correlación de umbral predeterminada, determinar un valor del parámetro de modificación de la señal separada que indique un intervalo completo de modificación;  
 si la correlación está por encima de la correlación de umbral predeterminada, determinar un valor del parámetro de modificación de la señal separada que indique un intervalo de modificación que tenga una relación inversa con la correlación.
- 50 10. El método de acuerdo con la reivindicación 8 en el que la medida del éxito de la separación comprende una correlación entre un espectro de frecuencia asociado al resto de la señal de audio compuesta y un espectro de frecuencia asociado a al menos una señal de audio de referencia.
- 55 11. El método de acuerdo con cualquiera de las reivindicaciones 1 a 7, en el que la medida del éxito de la separación comprende una correlación entre un resto de la señal de audio compuesta y un componente de una señal de vídeo correspondiente a la señal de audio compuesta.
- 60 12. El método de acuerdo con cualquier reivindicación anterior, que comprende:  
 responder a una determinación de que la medida del éxito de la separación indica que, para un marco temporal posterior de la señal de audio compuesta, un grado de éxito de la separación es más bajo que el grado de éxito de la separación para un marco temporal actual de la señal de audio compuesta mediante el reposicionamiento espacial de la señal de audio separada a una posición que esté más próxima a una posición espacial original de la señal de audio separada.
- 65 13. El método de la reivindicación 12, en el que el reposicionamiento espacial de la señal de audio separada a la posición que está más próxima a la posición espacial original se realiza previamente a la reproducción del marco temporal posterior de la señal de audio compuesta.
14. Aparato configurado para realizar un método de acuerdo con cualquiera de las reivindicaciones 1 a 13.

15. Instrucciones legibles por ordenador que, cuando las ejecuta un aparato informático, hacen que el aparato informático provoque la realización del método de acuerdo con cualquiera de las reivindicaciones 1 a 13.



**FIGURA 1**



**FIGURA 2A**

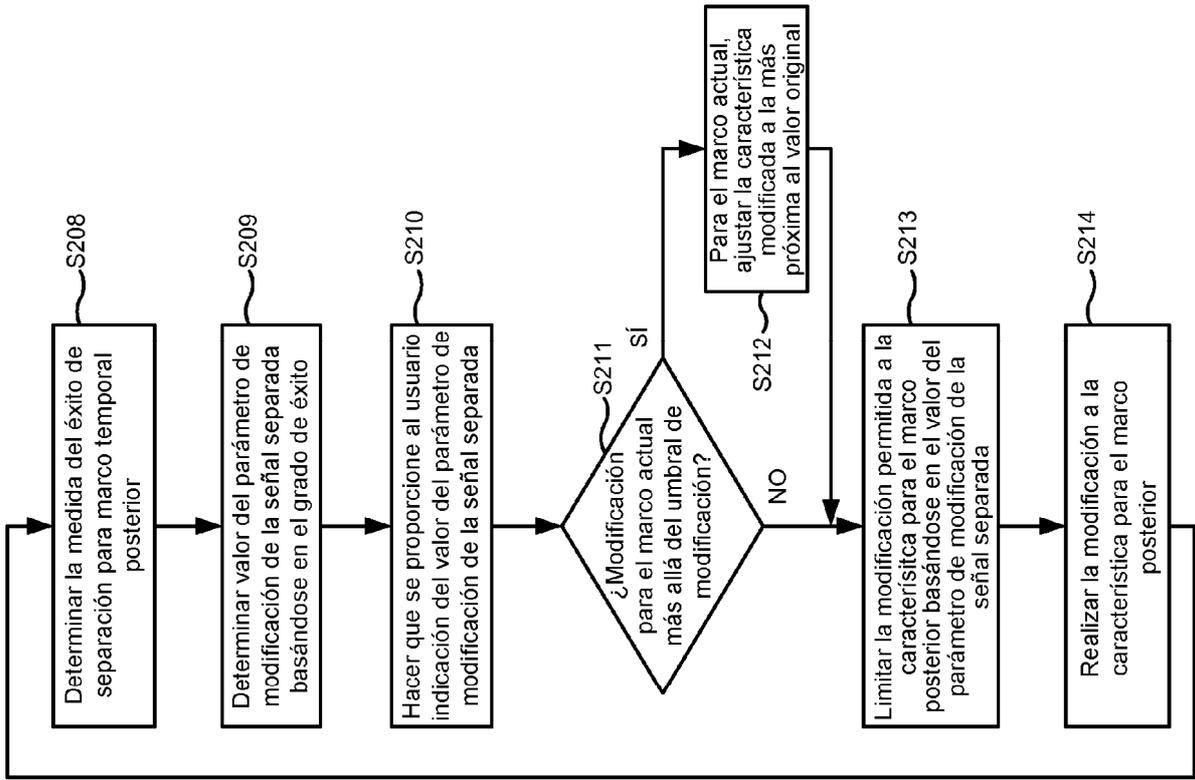


FIGURA 2C

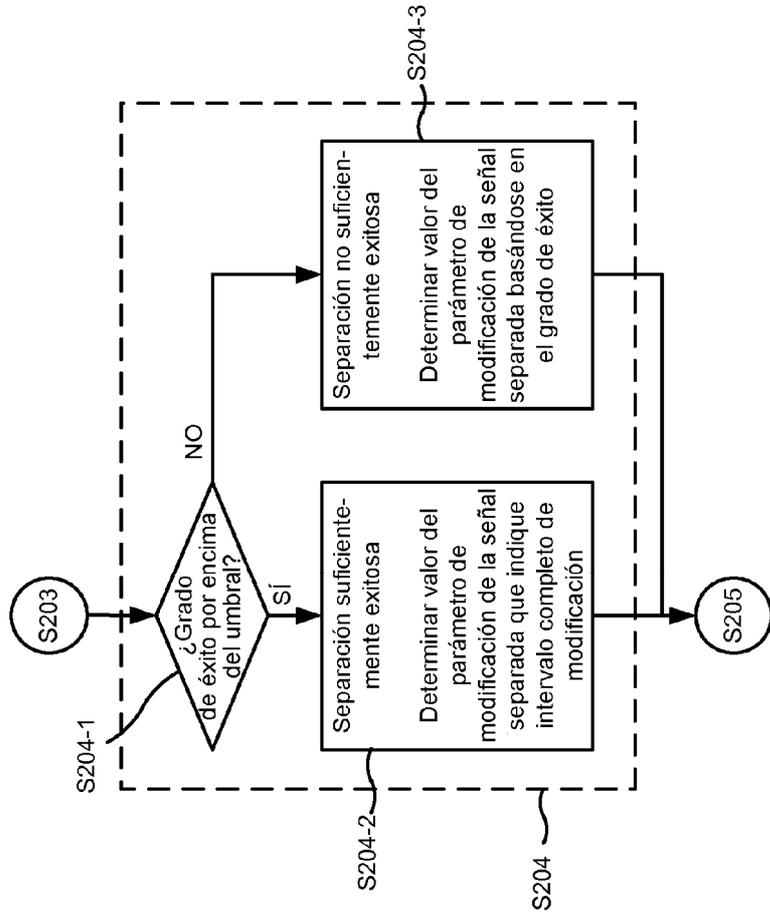


FIGURA 2B

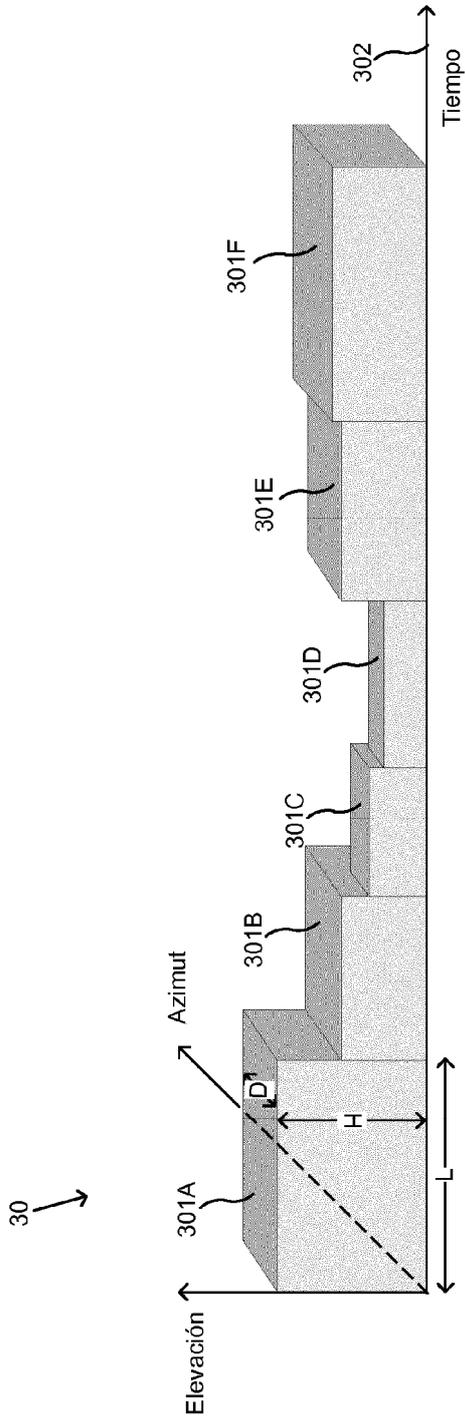


FIGURA 3A

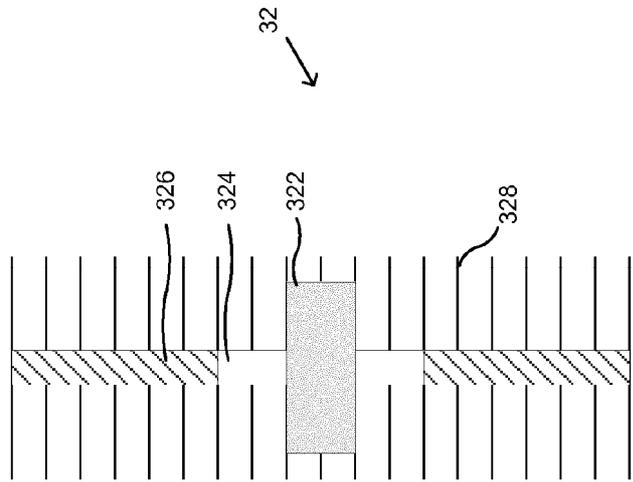


FIGURA 3B

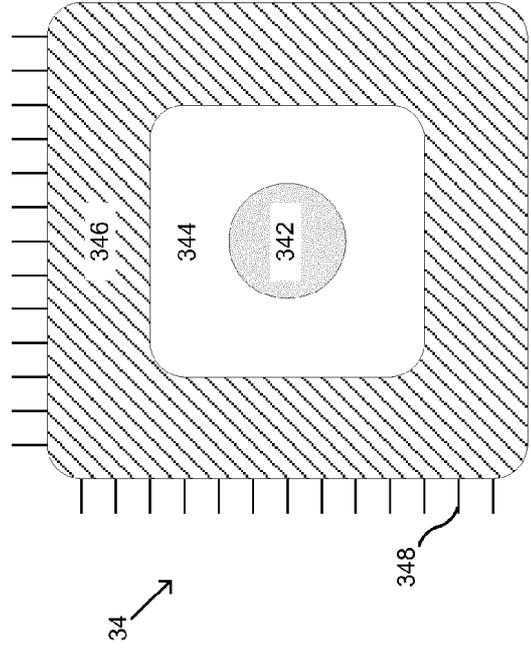
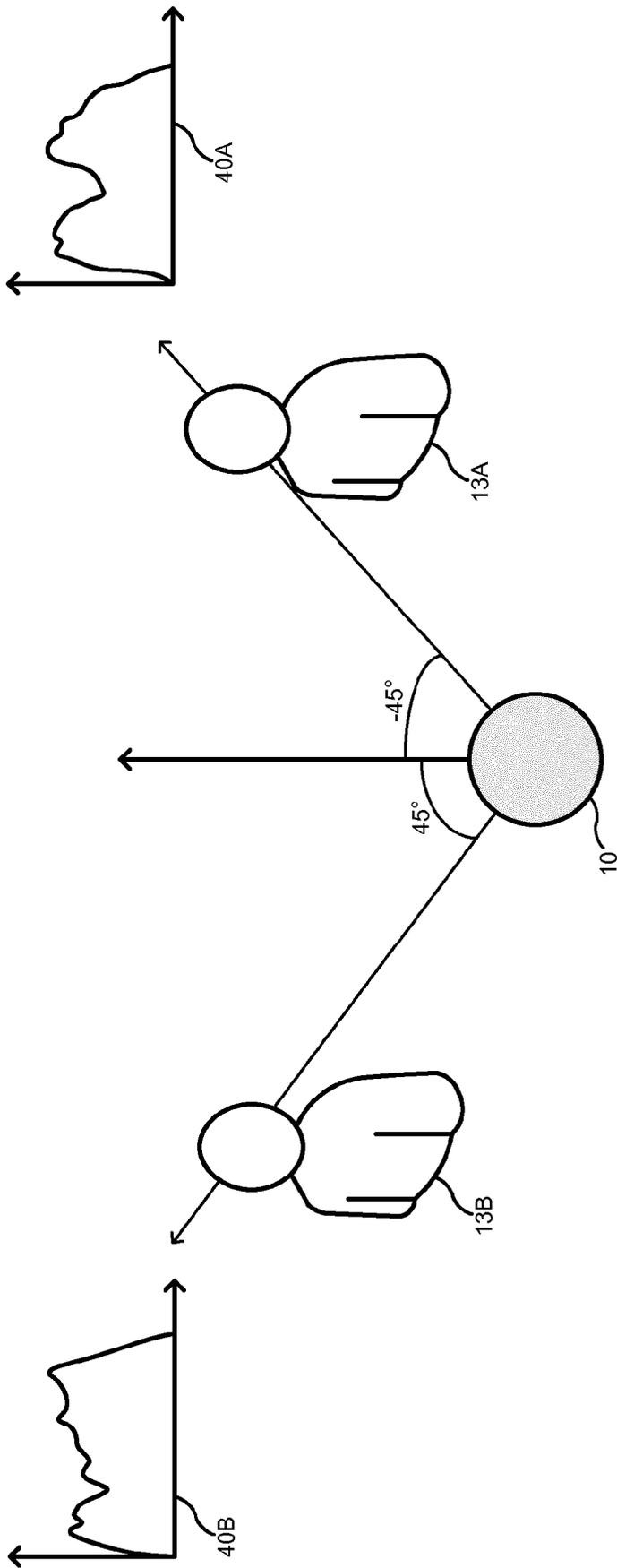


FIGURA 3C



**FIGURA 4A**

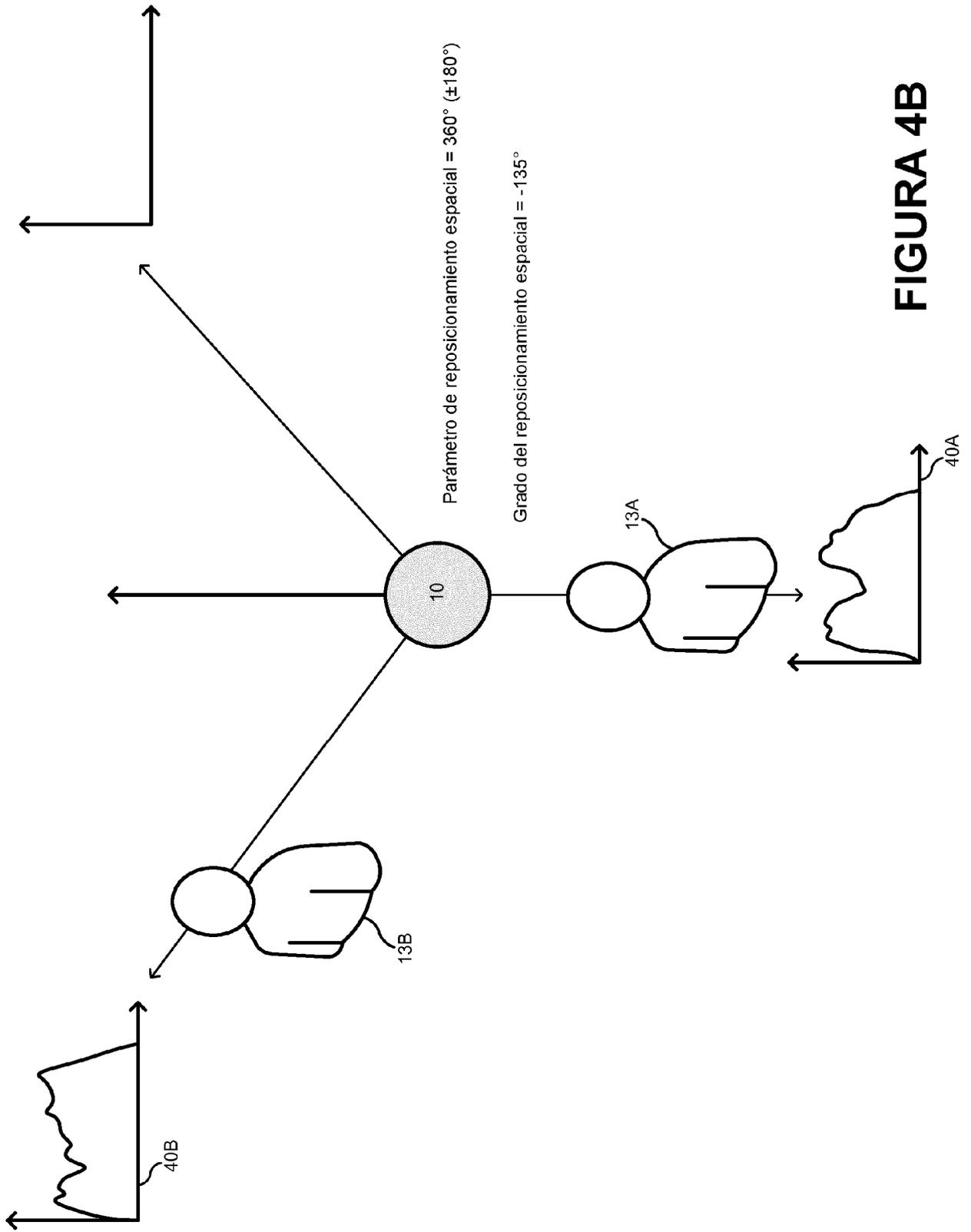


FIGURA 4B

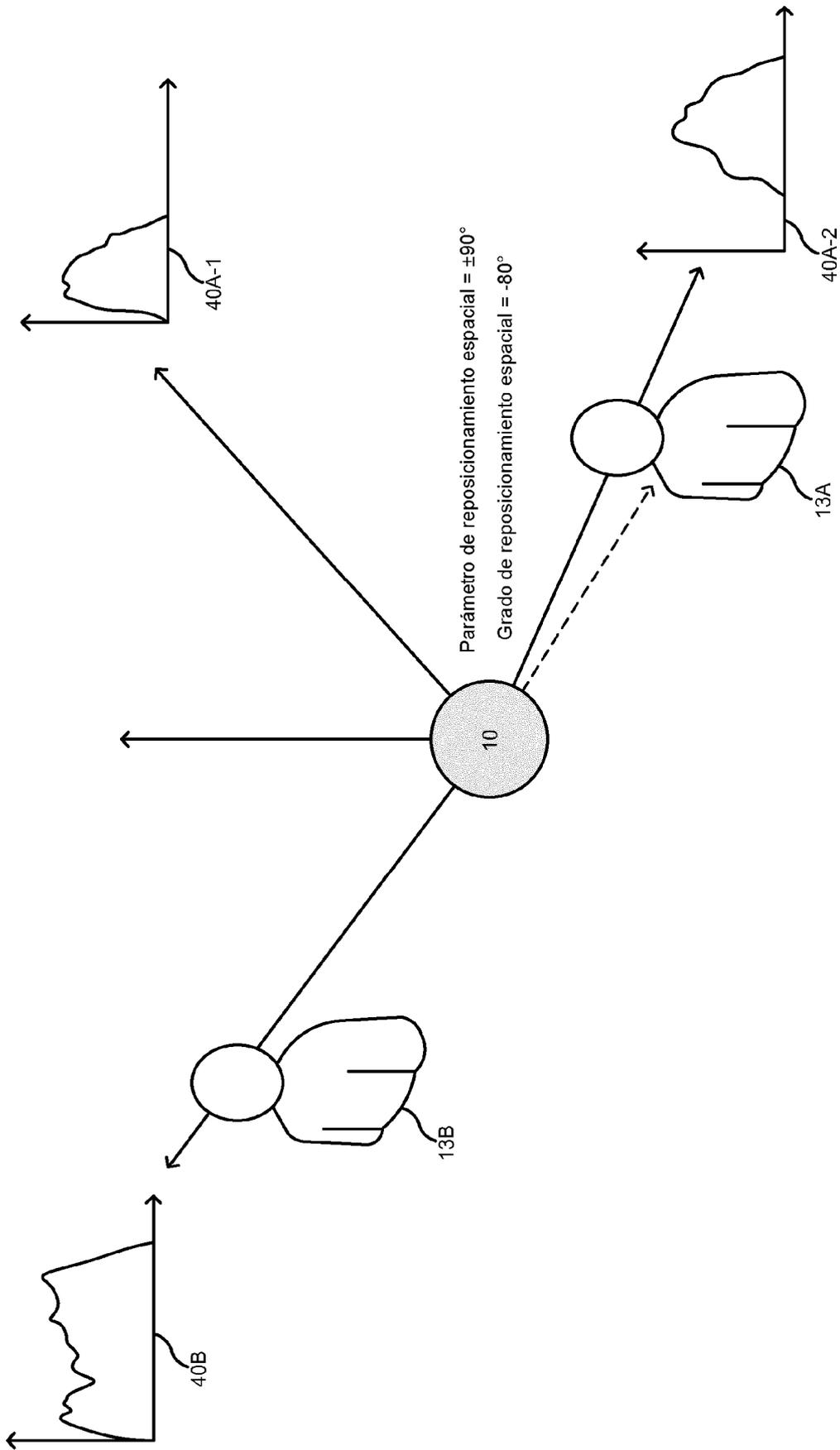
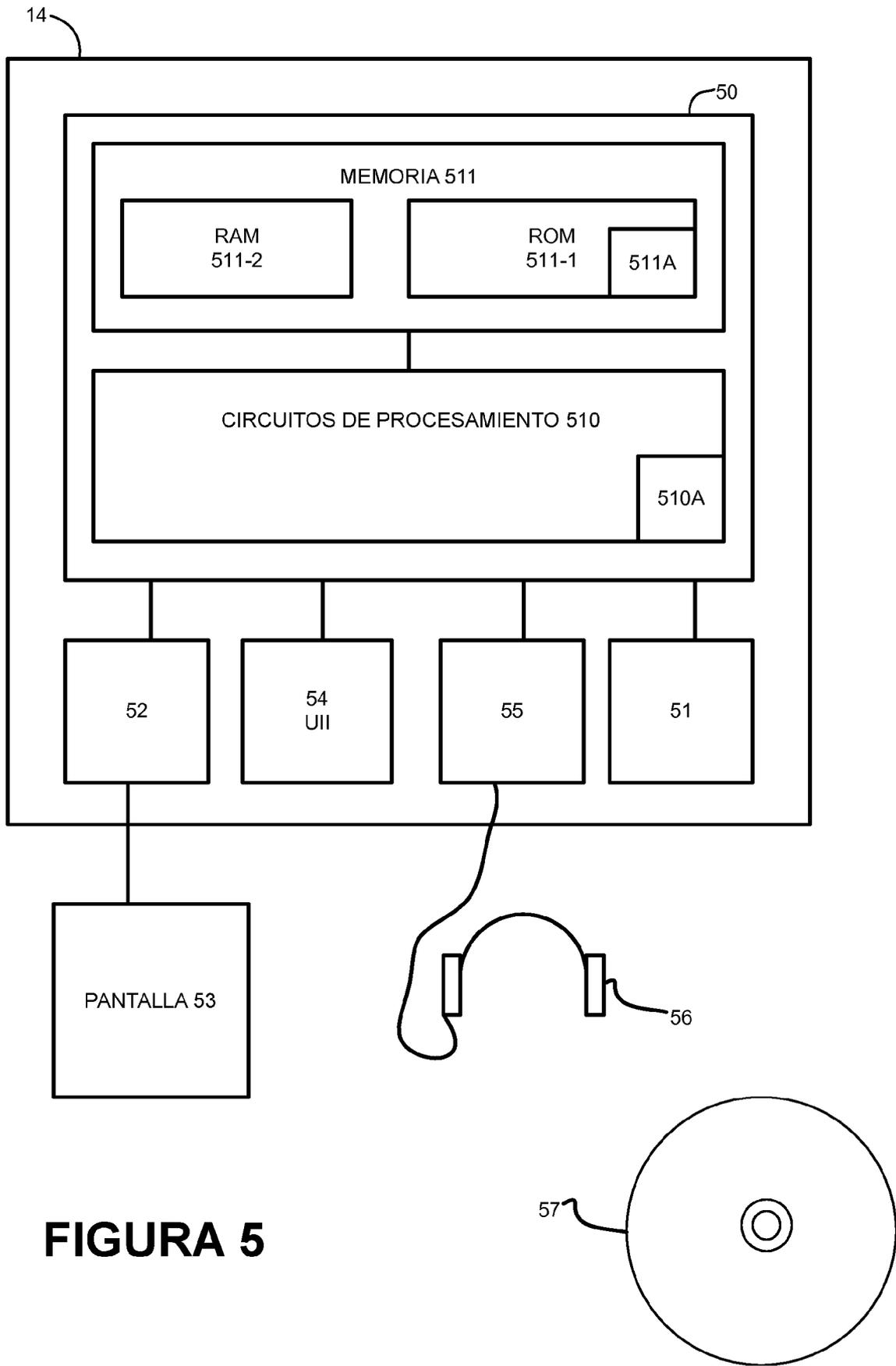


FIGURA 4C



**FIGURA 5**