

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 822 600**

51 Int. Cl.:

H04S 7/00

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **01.09.2016 PCT/EP2016/070673**

87 Fecha y número de publicación internacional: **08.03.2018 WO18041359**

96 Fecha de presentación y número de la solicitud europea: **01.09.2016 E 16762999 (7)**

97 Fecha y número de publicación de la concesión europea: **08.07.2020 EP 3507996**

54 Título: **Método de determinación de una función de transferencia relacionada con la cabeza y una función de diferencia de tiempo interaural personalizadas, y producto de programa informático para realizar el mismo**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
04.05.2021

73 Titular/es:

**UNIVERSITEIT ANTWERPEN (100.0%)
Prinsstraat 13
2000 Antwerpen, BE**

72 Inventor/es:

**REIJNIERS, JONAS;
PEREMANS, HERBERT y
PARTOENS, BART, WILFRIED, M**

74 Agente/Representante:

ELZABURU, S.L.P

ES 2 822 600 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

Método de determinación de una función de transferencia relacionada con la cabeza y una función de diferencia de tiempo interaural personalizadas, y producto de programa informático para realizar el mismo

Campo de la invención

5 La presente invención está relacionada con el campo de la tecnología de sonido 3D. Más concretamente, la presente invención está relacionada con un método implementado por ordenador de estimación de una función de transferencia relacionada con la cabeza individualizada (HTRF) y una función de diferencia de tiempo interaural individualizada (ITDF) de una persona particular. La presente invención también está relacionada con un producto de programa informático y un portador de datos que comprende dicho producto de programa informático, y con un kit de partes que
10 comprende dicho portador de datos.

Antecedentes de la invención

A lo largo de las pasadas décadas ha habido un gran avance en el campo de la tecnología de la realidad virtual, en particular con respecto a la realidad virtual visual. Las pantallas de televisión 3D han terminado llegando al público general, y especialmente los dispositivos de cine en casa y los videojuegos se aprovechan de ello. Pero la tecnología
15 de sonido 3D todavía va por detrás. Sin embargo, es – al menos en teoría – bastante fácil crear un entorno acústico 3D virtual, llamado Espacio de Auditorio Virtual (VAS). Cuando los humanos localizan el sonido en el espacio 3D, utilizan dos señales de audio recogidas por el oído izquierdo y el oído derecho. Por ello un indicio importante es la denominada “diferencia de tiempo interaural” (ITD): dependiendo de la dirección del sonido (con respecto a la cabeza de las personas), el sonido alcanzará primero el oído izquierdo o el oído derecho, y esta diferencia de tiempo contiene información acerca del ángulo lateral θ (véase la Figura 2). La función de diferencia de tiempo interaural (ITDF) describe cómo varía la ITD con la dirección de la fuente del sonido (por ejemplo, un altavoz), véase la Figura 3 para un ejemplo.
20

Otros indicios están contenidos en el contenido espectral del sonido cuando éste es registrado por el oído interno. Al fin y al cabo, antes de que las ondas sonoras procedentes de una cierta dirección alcancen la membrana del tímpano, interfieren con el cuerpo, la cabeza y los pabellones auditivos. Y mediante esta interferencia algunas frecuencias se transmiten con mayor facilidad que otras; por consiguiente, se produce un filtrado espectral que depende de la dirección desde la cual está llegando el sonido. Este filtrado se describe mediante la denominada “Función de Transferencia Relacionada con la Cabeza” (HRTF), (véase ejemplo en la Figura 4), la cual, para cada dirección de la fuente del sonido, describe la proporción de cada frecuencia que es transmitida o eliminada por filtrado. El contenido espectral de las señales recibidas en ambos oídos contiene por tanto información adicional (llamada: indicios espectrales) acerca de la ubicación de la fuente del sonido, y especialmente acerca de la elevación ϕ (véase la Figura 2), acerca de la altura a la cual está ubicada la fuente del sonido con respecto a la cabeza, pero también acerca de si la fuente del sonido está situada delante o detrás de la persona.
25

Para crear una realidad virtual acústica 3D realista (por ejemplo mediante un sistema de reproducción de audio), es por lo tanto fundamental conocer la ITDF y la HRTF de una persona particular. Cuando estas son conocidas, se pueden añadir artificialmente retrasos de tiempo y filtrado espectral apropiados para cualquier dirección específica y, de esta manera, se proporcionan al oyente los indicios necesarios (indicios temporales e indicios espectrales) para reconstruir el mundo 3D.
30

Actualmente, ya existen muchas aplicaciones en el mercado que utilizan la HRTF para crear una impresión 3D virtual, pero hasta el momento no se utilizan de manera generalizada. Al fin y al cabo, utilizan un conjunto de ITDF y HRTF único, generalizado, que se supone que funciona para una amplia audiencia. Exactamente igual que con los sistemas de visión 3D, en los que se supone que la distancia entre los ojos es la misma para todo el mundo, estos sistemas utilizan las ITDFs y HRTFs medias. Aunque esto no plantea problemas significativos para la visión, sí lo hace para el audio 3D. Cuando para un individuo, la distancia entre los ojos es significativamente diferente a la distancia media,
35 puede ocurrir que la percepción de profundidad del usuario no sea óptima, provocando la sensación de que “algo va mal”, pero los problemas relacionados con el audio 3D son mucho más graves. Pequeñas diferencias pueden provocar grandes errores. Equipado con “oídos medios” virtuales, el usuario experimenta efectivamente un efecto espacial – el sonido ya no está dentro de la cabeza -, sino en algún lugar fuera de la cabeza, pero a menudo existe mucha confusión acerca de la dirección de la cual proviene el sonido. La mayoría de los errores se cometen en la percepción de la elevación, pero también, y esto es mucho más molesto: a menudo la parte delantera y la parte trasera se intercambian. Sonido que en realidad debería proceder de la parte delantera, se percibe como procedente de la parte trasera, reduciendo significativamente la utilidad de esta tecnología.
40

Por tanto, a pesar del hecho de que la HRTF y la ITDF de personas diferentes son similares, incluso pequeñas diferencias entre las HRTF e ITDF verdaderas de una persona y las HRTF e ITDF generales provocan errores, los cuales, a diferencia de la visión 3D, son perjudiciales para la experiencia espacial. Esta es probablemente una de las razones por las que el VAS por medio de auriculares estéreo aún no ha alcanzado todo su potencial. Por tanto, para hacer un uso óptimo de la tecnología, es necesario utilizar una HRTF y una ITDF personalizadas. Pero ¿cómo conseguir esto a gran escala, de modo que esta tecnología se pueda poner a disposición del público general?
45

La HRTF y la ITDF de una persona se registran tradicionalmente utilizando infraestructura especializada: en una cámara anecoica, en la cual se posicionan fuentes de sonido alrededor del sujeto, y para cada dirección muestreada se registra la correspondiente señal que llega al oído izquierdo y al oído derecho por medio de micrófonos que están dispuestos en el oído izquierdo y en el oído derecho del sujeto, justo a la entrada del canal auditivo. Aunque en los últimos años se han hecho avances y se han desarrollado nuevos métodos para simplificar este procedimiento, estas mediciones siguen siendo muy engorrosas y caras. Por lo tanto, no es posible medir la HRTF y la ITDF de todos los potenciales usuarios de esta manera. Por lo tanto, existe una necesidad de buscar otras maneras de individualizar la HRTF y la ITDF.

El documento US5729612A describe un método y aparato para medir una función de transferencia relacionada con la cabeza, fuera de una cámara anecoica. En este documento se propone medir la HRTF utilizando una onda de sonido proporcionada como salida por un altavoz montado en un soporte especial. Una señal de audio izquierda y una señal de audio derecha son capturadas por dos micrófonos de oído portados por un sujeto cuyos movimientos de cabeza son rastreados por un sensor de posición y/o que está sentado en una silla que puede ser orientada en direcciones particulares (conocidas). Los datos se procesarán en un ordenador remoto. El documento no dice nada acerca de cómo se calculan exactamente la ITDF y la HRTF a partir de las señales de audio medidas y de las señales de posición. Sin embargo, se utiliza un paso de calibración para determinar una característica de transferencia del altavoz y de los micrófonos, y el método también confía fuertemente en el hecho de que las posiciones relativas de la persona y del altavoz se conocen con exactitud. Técnica anterior adicional se muestra en el documento US 2009/052703.

Todavía existe margen para mejora o para alternativas.

20 **Compendio de la invención**

Es un objeto de las realizaciones de la presente invención proporcionar un buen método y un buen producto de programa informático para determinar o estimar una función de diferencia de tiempo interaural (ITDF) personalizada y una función de transferencia relacionada con la cabeza (HRTF) personalizada.

Es un objeto de las realizaciones de la presente invención proporcionar un método y un producto de programa informático para determinar o estimar una ITDF personalizada y una HRTF personalizada, en base a datos capturados por el propio usuario final, en una disposición de ensayo relativamente simple sin requerir habilidades específicas o equipos profesionales.

Es un objeto de las realizaciones de la presente invención proporcionar un método y un producto de programa informático para realizar ese método en casi cualquier habitación en casa, y básicamente sólo requiere un dispositivo informático apropiado, micrófonos de oído, un altavoz y una unidad de orientación "de gama baja" como la que se encuentra típicamente en los teléfonos inteligentes (año 2016). Con la expresión "de gama baja" se quiere indicar que no es necesario que la información de orientación sea muy precisa (por ejemplo una posición angular de +/- 5° es aceptable), y parte de la información de orientación puede ser incorrecta, y donde la unidad de orientación se puede montar de forma fija en cualquier posición y orientación arbitrarias con respecto a la cabeza, y la persona se puede situar a una distancia arbitraria en el campo lejano del altavoz, y no es necesario que la persona realice movimientos precisos.

Es un objeto de las realizaciones de la presente invención proporcionar un método robusto (por ejemplo, "a prueba de tontos") y un producto de programa informático robusto que sea capaz de determinar o estimar una función de diferencia de tiempo interaural (ITDF) personalizada y una función de transferencia relacionada con la cabeza (HRTF) personalizada utilizando estímulos de audio emitidos por al menos un altavoz, en base a muestras de audio izquierdas y derechas capturadas por micrófonos de oído y en base a información de orientación originada en una unidad de orientación que está montada de forma fija a la cabeza de la persona, pero en donde la posición y/o la distancia y/o la orientación de la cabeza con respecto a los uno o más altavoces no se conoce con precisión en el momento de la captura de dichas muestras de audio.

Es un objeto de realizaciones particulares de la presente invención proporcionar un método y un producto de programa informático que permita estimar dichas ITDF y HRTF personalizadas utilizando una unidad de orientación que mide el campo magnético de la tierra y/o la aceleración y/o la velocidad angular (como la que se puede encontrar por ejemplo en teléfonos inteligentes apropiados del año 2016), y utilizando micrófonos de oído y un altavoz, opcionalmente pero no necesariamente en combinación con otro ordenador (tal como por ejemplo un ordenador portátil o de sobremesa).

Estos y otros objetivos son logrados por realizaciones de la presente invención.

En un primer aspecto, la presente invención está relacionada con un método de estimar una función de transferencia relacionada con la cabeza individualizada y una función de diferencia de tiempo interaural individualizada de una persona particular en un dispositivo informático, comprendiendo el método los pasos de: a) obtener o recuperar una pluralidad de conjuntos de datos, comprendiendo cada conjunto de datos una muestra de audio izquierda que se origina en un micrófono de oído izquierdo y una muestra de audio derecha que se origina en un micrófono de oído derecho e información de orientación que se origina en una unidad de orientación, siendo capturadas la muestra de audio izquierda y la muestra de audio derecha y la información de orientación de cada conjunto de datos de forma substancialmente simultánea en una disposición en la que: el micrófono de oído izquierdo está insertado en un oído

5 izquierdo de la persona, y el micrófono de oído derecho está insertado en un oído derecho de la persona, y la persona está ubicada a una distancia de un altavoz, y la unidad de orientación está montada de forma fija a la cabeza de la persona, y el altavoz está diseñado para reproducir una señal de ensayo acústica que comprende una pluralidad de fragmentos de ensayo de audio, y la persona mueve su cabeza en una pluralidad de orientaciones diferentes durante la reproducción de la señal de ensayo acústica; b) extraer o calcular a partir de los conjuntos de datos una pluralidad de valores de diferencia de tiempo interaural y/o una pluralidad de valores espectrales, y valores de orientación correspondientes de la unidad de orientación; c) estimar una dirección del altavoz con respecto a una posición media del centro de la cabeza de la persona y expresada en el sistema de referencia del mundo, comprendiendo los pasos de: 1) suponer una dirección candidata de la fuente; 2) asignar una dirección a cada miembro de al menos un subconjunto de la pluralidad de valores de diferencia de tiempo interaural y/o a cada miembro de al menos un subconjunto de la pluralidad de valores espectrales, *que corresponda con la dirección supuesta de la fuente expresada en un sistema de referencia de la unidad de orientación*, obteniendo de este modo un conjunto de datos mapeados; 3) calcular un valor de calidad del conjunto de datos mapeados en base a un criterio de calidad predefinido; 4) repetir los pasos 1) a 3) al menos una vez para una segunda dirección candidata y/o más direcciones candidatas de la fuente diferente a las direcciones candidatas de la fuente anteriores; 5) elegir la dirección candidata de la fuente que da como resultado el mayor valor de calidad como la dirección del altavoz con respecto a la posición media del centro de la cabeza de la persona; d) estimar una orientación de la unidad de orientación con respecto a la cabeza; e) estimar la ITDF individualizada y la HRTF individualizada de la persona, en base a la pluralidad de conjuntos de datos y en base a la dirección estimada del altavoz con respecto a la posición media del centro de la cabeza estimada en el paso c) y en base a la orientación estimada de la unidad de orientación con respecto a la cabeza estimada en el paso d); en donde los pasos a) a e) son realizados por al menos un dispositivo informático.

25 Con la última frase "en donde los pasos a) a e) son realizados por al menos un dispositivo informático" se quiere indicar que cada uno de los pasos individuales a) a e) es realizado por exactamente el mismo dispositivo informático o que algunos de los pasos son realizados por un primer dispositivo informático, y algunos otros pasos son realizados por un segundo dispositivo informático o incluso algún dispositivo informático más.

La "asignación de una dirección" del paso c) 2) puede comprender asignar dos coordenadas, por ejemplo dos coordenadas esféricas, u otras coordenadas apropiadas, preferiblemente de tal manera que definan una dirección única. Una ventaja de utilizar coordenadas esféricas es que en ese caso se pueden utilizar funciones esféricas en la determinación del valor de calidad, y que los resultados se pueden visualizar y se pueden interpretar más fácilmente.

30 El mapeado del paso c) 2) puede comprender mapear el conjunto de datos ITD_i, S_i a una esfera.

Es una ventaja de este método que la estimación de la dirección de la fuente en el paso c) se puede basar únicamente en las muestras de audio izquierdas y derechas capturadas y en la información de orientación que se origina en la unidad de orientación, sin tener que utilizar una ITDF o una HRFT general.

35 Es una ventaja de este método que la ITDF y la HRTF se pueden realizar en un ordenador estándar (por ejemplo, un ordenador portátil o de sobremesa) en un tiempo razonable (del orden de aproximadamente 30 minutos).

Es una ventaja del método de la presente invención que el algoritmo es capaz de extraer de manera correcta y precisa la ITDF y la HRTF de los datos capturados, incluso si la posición de la persona con respecto al altavoz no es fija, o no se conoce con precisión cuándo se capturan los datos. O, dicho de otra manera, es una ventaja que no es necesario conocer a priori, y no es necesario calibrar, la posición de la cabeza de la persona con respecto al altavoz.

40 Es una ventaja que la unidad de orientación puede tener una orientación a priori desconocida con respecto a la cabeza, es decir, se puede montar a la cabeza en cualquier orientación arbitraria (por ejemplo orientada o girada hacia la parte delantera de la cabeza, o girada hacia la parte posterior o hacia el lado izquierdo).

45 Es una ventaja de las realizaciones de acuerdo con la presente invención que la estimación de la orientación de la fuente de sonido con respecto a la cabeza se puede basar únicamente en datos de ITD (véase la Figura 27), o se puede basar únicamente en datos espectrales de las muestras de audio izquierdas a una frecuencia particular (por ejemplo a 8100 Hz), o se puede basar únicamente en datos espectrales de las muestras de audio derechas a una frecuencia particular (por ejemplo a 8100 Hz), o se puede basar en datos espectrales de al menos dos frecuencias diferentes (por ejemplo por adición del valor de calidad para cada frecuencia), o se puede basar en datos espectrales de las muestras de audio izquierdas y/o derechas en un rango de frecuencia predefinido (por ejemplo, desde aproximadamente 4 kHz hasta aproximadamente 20 kHz, véanse por ejemplo las Figuras 28 a 30), o cualquier combinación de estos.

55 Es una ventaja de las realizaciones de la presente invención que proporciona una ITDF y una HRTF individualizadas para un individuo, cuyas ITDF y HRTF sólo es necesario estimar una vez, y por consiguiente se pueden utilizar en una variedad de aplicaciones, como por ejemplo en juegos 3D o en aplicaciones de conferencia telefónica para crear una experiencia espacial.

Es una ventaja de las realizaciones de la presente invención que el algoritmo para estimar la ITDF y la HRTF no necesita ser ajustado a un entorno o disposición particular, especialmente en el momento de capturar las muestras de audio y los datos de orientación.

Es una ventaja particular que el método no impone movimientos estrictos cuando se capturan los datos, y puede ser realizado por la mayoría de los individuos en su casa, sin requerir equipos caros. En particular, aparte de un par de micrófonos de oído, otro equipo necesario para realizar la parte de captura está disponible de manera generalizada (por ejemplo: un dispositivo para reproducir audio en un altavoz, un teléfono inteligente, un ordenador).

5 Es una ventaja que no es necesario conocer a priori la característica de filtro espectral del altavoz.

Es una ventaja de las realizaciones de la presente invención que el algoritmo para estimar la ITDF y la HRTF permite estimar la orientación relativa de la cabeza con respecto al altavoz en el momento de la obtención de datos, sin conocimiento de la orientación o posición (exactas) de la unidad de orientación en la cabeza y sin conocimiento preciso de la posición (exacta) del altavoz y/o de la persona en la habitación, y sin requerir una calibración para determinar la posición relativa y/o la orientación de la cabeza con respecto al altavoz.

Es una ventaja de las realizaciones de la presente invención que el algoritmo para estimar la ITDF y la HRTF se puede realizar en el mismo dispositivo, o en otro dispositivo diferente al dispositivo que se utilizó para capturar los datos de audio y de orientación. Por ejemplo, los datos pueden ser capturados por un teléfono inteligente y ser transmitidos a un ordenador remoto o ser almacenados en una tarjeta de memoria en un primer paso, pudiendo a continuación los datos ser obtenidos (por ejemplo recibidos a través de un cable o de forma inalámbrica) o recuperados de la tarjeta de memoria por el ordenador remoto para estimar realmente la ITDF y la HRTF.

Es una ventaja de las realizaciones de la presente invención que el algoritmo para estimar la ITDF y la HRTF no requiere necesariamente información de orientación muy precisa obtenida de la unidad de orientación (por ejemplo un margen de tolerancia de aproximadamente +/- 10° puede ser aceptable), debido a que el algoritmo puede, pero no necesita, confiar únicamente en los datos de orientación para determinar la posición relativa, sino que también puede confiar en los datos de audio.

Aunque la ITDF y la HRTF proporcionadas por la presente invención no serán tan precisas como la ITDF y la HRTF medidas en una cámara anecoica, es una ventaja que se espera que la ITDF y la HRTF personalizadas que pueden ser obtenidas por la presente invención, cuando se utilicen en un sistema 3D-VAS, proporcionen resultados mucho mejores que el uso de ese mismo sistema 3D-VAS con una ITDF y una HRTF "medias" o "generales", especialmente en términos de fallos de percepción delante/detrás.

Es una ventaja de las realizaciones de la presente invención que el algoritmo puede contener una o más iteraciones para derivar la ITDF y la HRTF, mientras que sólo es necesario realizar una vez el paso de captura de datos. Múltiples iteraciones proporcionarán una mejor aproximación de la ITDF y la HRTF verdaderas, a expensas de tiempo de procesamiento.

Es una ventaja de las realizaciones de la presente invención que está basada en el conocimiento de que múltiples incógnitas (tales como por ejemplo la orientación desconocida entre la cabeza de la persona y el altavoz, y/o la característica de transferencia desconocida de los micrófonos y/o la del altavoz, y/o la ITDF y la HRTF desconocidas) se pueden calcular "juntas" utilizando aproximaciones por pasos, por lo cual en cada aproximación se puede utilizar una versión mejorada de las variables desconocidas. El número de iteraciones puede ser seleccionado (y por lo tanto ajustado a un valor predefinido) por la persona con experiencia, en base a la precisión requerida, o se puede determinar dinámicamente durante la medición.

Es una ventaja de las realizaciones de la presente invención que no requiere equipos especiales (por ejemplo una cámara anecoica con una pluralidad de micrófonos dispuestos en una esfera o un arco), sino que puede ser realizada por el propio usuario en su domicilio en una configuración muy simple.

En una realización, el paso b) comprende: localizar una pluralidad de fragmentos de audio izquierdos y de fragmentos de audio derechos en la pluralidad de conjuntos de datos, correspondiendo cada fragmento de audio izquierdo y derecho con un fragmento de ensayo de audio reproducido por el altavoz; calcular un valor de diferencia de tiempo interaural para al menos un subconjunto de las parejas de fragmentos de audio izquierdo y derecho correspondientes; estimar una orientación momentánea de la unidad de orientación para cada pareja de fragmentos de audio izquierdo y derecho correspondientes.

Es una ventaja de esta realización que la estimación de la orientación de la fuente de sonido se puede basar únicamente en datos ITD, si así se desea, como se ilustra en la Figura 27.

En una realización, el paso b) comprende o comprende además: localizar una pluralidad de fragmentos de audio izquierdos y/o fragmentos de audio derechos en la pluralidad de conjuntos de datos, correspondiendo cada fragmento de audio izquierdo y/o derecho con un fragmento de ensayo de audio reproducido por el altavoz; calcular un conjunto de valores espectrales izquierdos para cada fragmento de audio izquierdo y/o calcular un conjunto de valores espectrales derechos para cada fragmento de audio derecho, conteniendo cada conjunto de valores espectrales al menos un valor espectral correspondiente a una frecuencia espectral; estimar una orientación momentánea de la unidad de orientación para al menos un subconjunto de los fragmentos de audio izquierdos y/o de los fragmentos de audio derechos.

Es una ventaja de esta realización que la estimación de la orientación de la fuente de sonido se puede basar en datos espectrales. Esto es especialmente útil si las muestras de ensayo de audio tienen una frecuencia variable, por ejemplo si las muestras de ensayo de audio son "chirps".

En una realización, el criterio de calidad predefinido es un criterio de suavidad espacial de los datos mapeados.

- 5 Sorprendentemente los inventores encontraron que la estimación de la orientación de la fuente de sonido con respecto a la cabeza se puede encontrar buscando la dirección para la cual los datos mapeados son los "más suaves", a diferencia de su previsión original de que una estimación incorrecta de la dirección de la fuente produciría como resultado una simple rotación de los datos mapeados en la esfera. En cambio, los experimentos han mostrado que una estimación incorrecta de la dirección de la fuente produce como resultado una grave distorsión de los datos mapeados y de los datos de ITDF y HRTF resultantes. Hasta donde saben los inventores, esta información no es conocida en la técnica anterior. De hecho, hasta donde saben los inventores, no existe ninguna técnica anterior en la que la fuente de sonido se ubique en una posición/orientación desconocida con respecto al sujeto.

- 10 En una realización, el criterio de calidad predefinido está basado en una desviación o distancia entre los datos mapeados y una superficie de referencia, donde la superficie de referencia se calcula como una variante de paso bajo de dichos datos mapeados.

Es una ventaja de esta realización que la superficie de referencia utilizada para definir "suavidad" se puede derivar de los propios datos mapeados, de esta manera por ejemplo no es necesario extraerlos de una base de datos que contiene funciones IDTF o HRTF utilizando análisis estadístico. Esto simplifica la implementación del algoritmo, aunque es muy flexible y proporciona resultados muy precisos.

- 20 Se observa que muchas superficies "suaves" se pueden utilizar como superficie de referencia, lo cual ofrece oportunidades para mejorar aún más el algoritmo, por ejemplo en términos de complejidad computacional y/o velocidad.

- 25 En una realización, el criterio de calidad predefinido se basa en una desviación o distancia entre los datos mapeados y una superficie de referencia, donde la superficie de referencia se basa en una aproximación de los datos mapeados, definida por la suma ponderada de un número limitado de funciones base.

Es una ventaja de la utilización de un conjunto limitado de funciones base, en particular de un conjunto de funciones base ortogonales que tienen un "orden" menor que un valor predefinido (por ejemplo un valor en el rango de 5 a 15), que son muy apropiadas para aproximar la mayoría de las superficies relativamente suaves, y que se pueden calcular de maneras conocidas, y pueden ser representadas mediante un conjunto de parámetros relativamente pequeño.

- 30 En una realización, las funciones base son funciones armónicas esféricas.

Aunque la invención también funcionará con otras funciones, las funciones armónicas esféricas son funciones base muy convenientes para esta aplicación. Ofrecen las mismas ventajas que las series de Fourier en otras aplicaciones.

En una realización, se utilizan armónicos esféricos reales.

En otra realización, se utilizan armónicos esféricos complejos.

- 35 En una realización, el criterio de calidad predefinido es un criterio que expresa un grado de la antisimetría especular de los datos ITDi mapeados.

Con antisimetría especular se quiere indicar simétrico excepto por el signo.

- 40 Se pueden utilizar varias propiedades generales de la ITDF y/o de la HRTF para definir el criterio de calidad. En la reivindicación 4 se utiliza la suavidad de una ITDF y/o de una HRTF y se contempla aprovechar la simetría especular o simetría cilíndrica de la ITDF. Los ITDi serán mayormente simétricos cilíndricamente alrededor de un eje (de hecho, el eje oído-oído) en caso de que se suponga la dirección real correcta de la fuente. De manera similar, los ITDi mostrarán la máxima simetría especular alrededor de un plano que pasa a través del centro de la esfera en caso de que se suponga la dirección real correcta de la fuente. En el último caso, esto permite determinar la dirección de la fuente excepto por el signo.

- 45 En una realización, el criterio de calidad predefinido es un criterio que expresa un grado de simetría cilíndrica de los datos ITDi mapeados.

- 50 En una realización, el método comprende además: f) estimar los parámetros del modelo de un modelo mecánico relacionado con los movimientos de cabeza que fueron realizados por la persona en el momento de capturar las muestras de audio y la información de orientación del paso a); g) estimar una pluralidad de posiciones de la cabeza utilizando el modelo mecánico y los parámetros del modelo estimados; y en donde el paso c) comprende utilizar las posiciones estimadas de la cabeza del paso g).

Esta es una ventaja de utilizar un modelo mecánico para estimar la posición del centro de la cabeza, en lugar de suponer que la posición de la cabeza es fija. El modelo permite estimar mejor la posición relativa de y/o la distancia entre la cabeza y el altavoz. Esto permite mejorar la precisión de la ITDF y la HRTF.

5 En una realización, el modelo mecánico está adaptado para modelizar al menos la rotación de la cabeza alrededor de un centro de la cabeza, y al menos uno de los siguientes movimientos: rotación de la persona alrededor de un eje vertical estacionario, cuando está sentada en una silla giratoria; movimiento del cuello de la persona con respecto al torso de la persona.

10 Es una ventaja de utilizar un modelo como este, especialmente un modelo que tenga ambos rasgos, que permite estimar mejor la posición relativa de la cabeza con respecto al altavoz, produciendo como resultado una mejora de la precisión de la ITDF y la HRTF.

Es una ventaja que este modelo permite que se capturen los datos en el paso a) de una manera mucho más conveniente para el usuario, que no tiene que intentar mantener el centro de su cabeza en un único punto en el espacio, sin reducir la precisión de la ITDF y la HRTF.

15 En una realización, el paso b) comprende: estimar una trayectoria de los movimientos de la cabeza para una pluralidad de fragmentos de audio; teniendo en cuenta la trayectoria estimada cuando se estime la posición de la cabeza y/o la orientación de la cabeza.

En una realización, se puede utilizar más de un altavoz (por ejemplo, dos altavoces), ubicados en direcciones diferentes con respecto al usuario, en cuyo caso se utilizaría más de una señal de ensayo acústica (por ejemplo dos), y en cuyo caso en el paso c) se estimaría la dirección del altavoz que generó cada estímulo acústico específico.

20 Es una ventaja de utilizar dos altavoces, por ejemplo situados de manera que formen un ángulo de 45° o 90° vistos desde la posición del usuario (por ejemplo en cualquier momento particular en el tiempo durante la captura de datos), el que ello produce como resultado estimaciones mejoradas de las direcciones de los altavoces, porque existen dos puntos de referencia que no cambian de posición. Asimismo, el usuario no tendría que girar su cabeza tanto en comparación con una configuración con sólo un único altavoz, y aún cubriría una parte mayor de la esfera de muestreo.

25 En realizaciones particulares, estímulos de ensayo acústicos individuales pueden ser emitidos por los dos altavoces de forma alternante.

30 En una realización, el paso e) comprende además estimar una característica de filtro combinada del altavoz y los micrófonos, o comprende ajustar la ITDF estimada de tal manera que la energía por cada banda de frecuencia corresponda a la de una ITDF general y comprende ajustar la HTRF estimada de tal manera que la energía por cada banda de frecuencia corresponda a la de una HRTF general.

35 Es una ventaja de las realizaciones de la presente invención que el algoritmo para estimar la ITDF y la HRTF no necesita conocer de antemano la característica de filtro espectral del altavoz y de los micrófonos de oído, sino que puede estimar la característica de filtro espectral combinada del altavoz y del micrófono como parte del algoritmo, o puede compensar de tal manera que la ITDF y la HRTF resultantes tengan aproximadamente la misma densidad de energía o el mismo contenido de energía que la ITDF y la HRTF generales.

40 Esto ofrece la ventaja de que el usuario puede (en principio) utilizar cualquier conjunto de micrófonos de oído (de calidad razonable) y cualquier altavoz (de calidad razonable). Esto ofrece la ventaja de que no es necesario utilizar ningún tipo particular de altavoz y de micrófonos de oído durante la captura de datos, y también de que se puede omitir un paso de calibración específico. Pero por supuesto, también es posible utilizar un altavoz y micrófonos de oído con una característica de filtro espectral conocida, en cuyo caso el algoritmo puede utilizar la característica de filtro espectral conocida, y la estimación de las características de filtro espectral combinadas del altavoz y de los micrófonos de oído se puede omitir.

45 La estimación de una característica de filtro espectral combinada del altavoz y de los micrófonos se puede basar en la suposición o aproximación de que esta característica de filtro espectral combinada es una función espectral en sólo un único parámetro, a saber, la frecuencia, pero es independiente de la orientación. Esta aproximación es válida debido al pequeño tamaño de los micrófonos de oído y a la distancia relativamente grande entre la persona y el altavoz, preferiblemente al menos 1,5 m, más preferiblemente al menos 2,0 m.

50 En una realización, estimar la característica de filtro espectral combinada del altavoz y los micrófonos comprende: utilizar información a priori acerca de una característica de filtro espectral del altavoz, y/o utilizar información a priori acerca de una característica de filtro espectral de los micrófonos.

Las realizaciones de la presente invención pueden utilizar información estadística acerca de micrófonos de oído típicos y acerca de altavoces típicos. Esto puede comprender por ejemplo el uso de una característica de filtro espectral "media" y de una función de "covarianza", que se pueden utilizar en el algoritmo para calcular una medida de "distancia" o medida de desviación o una probabilidad de funciones candidatas.

En una realización, el paso b) estima la orientación de la unidad de orientación teniendo en cuenta también información espacial extraída de las muestras de audio Izquierda y Derecha, utilizando al menos una función de transferencia que relaciona indicios acústicos con información espacial.

5 En esta realización, se utiliza al menos una función de transferencia, tal como por ejemplo una ITDF y/o una HRTF de humanos, por ejemplo una ITDF general y/o una HRTF general de humanos, para permitir la extracción de información espacial (por ejemplo, información de orientación) de las muestras de audio izquierda y derecha.

10 Es una ventaja del algoritmo, que teniendo en cuenta al menos una función de transferencia, permite extraer información espacial de los datos de audio, la cual, en combinación con los datos del sensor de orientación, permite estimar mejor y/o mejorar la precisión de la orientación relativa de la cabeza durante la obtención de datos, sin conocimiento de la posición/orientación (exactas) de la unidad de orientación en la cabeza y sin conocimiento de la posición (exacta) del altavoz. Esto es especialmente útil cuando la precisión de la propia unidad de orientación es bastante baja.

15 Es una ventaja de algunas realizaciones de la presente invención que es capaz de extraer información espacial de datos de audio, necesaria para estimar la ITDF y la HRTF, aunque la ITDF y/o la HRTF exactas aún no se conocen, por ejemplo resolviendo el problema de manera iterativa. En una primera iteración, se puede utilizar una función de transferencia general para extraer información espacial de los datos de audio. Esta información se puede utilizar a continuación para estimar la HRTF y/o la ITDF, las cuales, en una siguiente iteración, se pueden utilizar entonces para actualizar la al menos una función de transferencia, convergiendo finalmente a una estimación mejorada de la ITDF y la HRTF.

20 Se observa que en caso de que se utilice más de un altavoz (por ejemplo dos altavoces) ubicados en diferentes direcciones vistos desde la posición del usuario, es una ventaja que la información espacial se extrae de dos fuentes de sonido diferentes, ubicadas en diferentes direcciones. De manera general, la función de transferencia que relaciona indicios acústicos con información espacial no es espacialmente homogénea, es decir, no todas las direcciones espaciales están igual de bien representadas en términos de indicios acústicos y, por consiguiente, sonidos procedentes de algunas direcciones son más fáciles de localizar en base a su contenido acústico, que aquellos que se originan de otras direcciones. Utilizando más de un altavoz (por ejemplo, dos), se puede hacer frente a estos 'puntos ciegos' en la función de transferencia, porque los dos altavoces muestrean diferentes direcciones de la función de transferencia, y si un altavoz produce un sonido que es difícil de localizar, el sonido que se origina del otro altavoz todavía puede contener la información direccional necesaria para hacer inferencias sobre la orientación de la cabeza.

30 En una realización, la al menos una función de transferencia predefinida que relaciona indicios acústicos con información espacial es una función de diferencia de tiempo interaural (ITDF) predefinida.

Es una ventaja de las realizaciones en las cuales la función de transferencia es una ITDF predefinida que la orientación de la cabeza con respecto al altavoz durante la captura de cada conjunto de datos se calcula únicamente a partir de una ITDF (media o estimada), y no de la HRTF.

35 En una realización, la al menos una función de transferencia que relaciona indicios acústicos con información espacial son dos funciones de transferencia que incluyen una función de diferencia de tiempo interaural predefinida y una función de transferencia relacionada con la cabeza predefinida.

40 Es una ventaja de las realizaciones que la orientación de la cabeza con respecto al altavoz durante la captura de cada conjunto de datos se calcula tanto a partir de una (media o estimación de una) ITDF, como a partir de una (media o estimación de una) HRTF, porque esto permite una estimación mejorada de la orientación de la cabeza con respecto al altavoz durante la obtención de datos, lo cual, a su vez, permite mejorar las estimaciones de la ITDF y de la HRTF.

45 En una realización, el método comprende realizar los pasos b) a e) al menos dos veces, en donde el paso b) de la primera iteración no tiene en cuenta dicha información espacial, y en donde el paso b) de la segunda iteración y de cualquier iteración adicional tiene en cuenta dicha información espacial, utilizando la función de diferencia de tiempo interaural y/o la función de transferencia relacionada con la cabeza estimadas en el paso e) de la primera iteración o de iteraciones adicionales.

Es una ventaja de las realizaciones que la orientación de la cabeza con respecto al altavoz se puede calcular teniendo en cuenta una ITDF y una HRTF, pero no en la primera iteración, sino a partir de la segunda iteración. De esta manera se puede evitar el uso de una ITDF general y/o de una HRTF general, si así se desea.

50 En una realización, el paso e) de estimar la función ITDF comprende utilizar información a priori acerca de la ITDF personalizada basada en análisis estadístico de una base de datos que contiene una pluralidad de ITDFs de diferentes personas.

55 Las realizaciones de la presente invención pueden utilizar información estadística acerca de ITDFs típicas contenidas en una base de datos. Esto puede por ejemplo comprender el uso de una ITDF "media" y de una función "covarianza", las cuales se pueden utilizar en el algoritmo para calcular una medida de "distancia" o medida de desviación o una probabilidad de funciones candidatas.

Es una ventaja de las realizaciones de la presente invención que se tiene en cuenta información de estas bases de datos (algunas de las cuales están disponibles para el público), porque incrementa la precisión de la ITDF individualizada estimada y de la HRTF individualizada estimada.

5 Es una ventaja de las realizaciones particulares de la presente invención que sólo se tiene en cuenta un subconjunto de dichas bases de datos, por ejemplo, en base a la edad o en base al género de la persona particular.

En una realización, el paso e) de estimar la HRTF comprende utilizar información a priori acerca de la HRTF personalizada basada en análisis estadístico de una base de datos que contiene una pluralidad de HRTFs de diferentes personas.

10 Las mismas ventajas que se han mencionado anteriormente cuando se utiliza información a priori acerca de la ITDF, aplican también para la HRTF.

En una realización, la unidad de orientación comprende al menos un sensor de orientación adaptado para proporcionar información de orientación relativa al campo gravitatorio de la tierra y al menos un sensor de orientación adaptado para proporcionar información de orientación relativa al campo magnético de la tierra.

15 Es una ventaja de las realizaciones de la presente invención que se utiliza una unidad de orientación que puede proporcionar información de orientación relativa a un sistema de coordenadas que está fijado a la tierra (también denominado en esta memoria "al mundo"), a diferencia de una unidad de posicionamiento que requiera una unidad de envío y una unidad receptora, porque requiere sólo una única unidad.

En una realización, el método comprende además el paso de: montar de forma fija la unidad de orientación a la cabeza de la persona.

20 El método de la presente invención tiene en cuenta que la orientación relativa de la unidad de orientación y de la cabeza es fija para todas las muestras/fragmentos de audio. No se requiere ninguna orientación específica, cualquier orientación arbitraria es buena, siempre y cuando la orientación relativa entre la cabeza y la unidad de orientación sea constante.

25 En una realización, la unidad de orientación está comprendida en un dispositivo portátil, y en donde el método comprende además el paso de: montar de forma fija a la cabeza de la persona el dispositivo portátil que comprende la unidad de orientación.

En una realización, el método comprende además los pasos de: reproducir una señal de ensayo acústica por medio del altavoz; capturar dichas señales de audio izquierda y derecha que se originan en dichos micrófonos de oído izquierdo y derecho y capturar dicha información de orientación de una unidad de orientación.

30 En una realización, la unidad de orientación está comprendida en un dispositivo portátil, pudiéndose montar el dispositivo portátil a la cabeza de la persona; y el dispositivo portátil comprende además un procesador programable y una memoria, y métodos de interfaz conectados eléctricamente a los micrófonos de oído izquierdo y derecho, y medios para almacenar y/o transmitir dichos conjuntos de datos capturados; y el dispositivo portátil captura la pluralidad de muestras de audio izquierdas y de muestras de audio derechas e información de orientación, y el
35 dispositivo portátil almacena los conjuntos de datos capturados en una memoria intercambiable y/o transmite los conjuntos de datos capturados al dispositivo informático, y el dispositivo informático lee dicha memoria intercambiable o recibe los conjuntos de datos capturados transmitidos, y realiza los pasos c) a e) mientras lee o recibe los conjuntos de datos capturados o después de leerlos o recibirlos.

40 En una realización de este tipo el paso de la captura real de datos es realizado por el dispositivo portátil, por ejemplo por un teléfono inteligente provisto de un dispositivo enchufable con una entrada de audio estéreo o similar, mientras que el procesamiento de los datos capturados puede ser realizado fuera de línea por otro ordenador, por ejemplo en la nube. Dado que la unidad de orientación es parte del propio teléfono inteligente, no se necesita ningún cable extra.

45 Es una ventaja de dicha realización que los cables que van a los micrófonos de oído pueden ser (mucho) más cortos (en comparación con cables enrutados a un ordenador cercano), dando como resultado una mayor libertad de movimiento. Además, las señales de audio izquierda y derecha capturadas pueden tener una mejor SNR debido al menor movimiento de los cables y a los bucles más pequeños formados por los cables, por tanto menos captación de radiación electromagnética no deseada. El dispositivo portátil puede comprender una cantidad de memoria suficiente para almacenar dichas señales de audio, por ejemplo puede comprender 1 GByte de memoria volátil (RAM) o de memoria no volátil (FLASH), y el dispositivo portátil puede comprender por ejemplo un transmisor inalámbrico, por
50 ejemplo un transmisor RF (por ejemplo Bluetooth, wifi, etc.), para transmitir los conjuntos de datos a un dispositivo externo. Experimentos han mostrado que un tamaño de RAM de aproximadamente 100 a 200 MByte puede ser suficiente.

En dicha realización, el ordenador externo realizaría típicamente todos los pasos b) a e), excepto el paso a) de captura de datos, y el dispositivo portátil, por ejemplo un teléfono inteligente, realizaría la captura de datos.

Por supuesto también es posible otra división de la funcionalidad, por ejemplo la primera ejecución del paso c), utilizando una ITDF media y/o una HRTF media también se puede ejecutar en el teléfono inteligente, mientras los otros pasos son realizados por el ordenador.

5 En una realización, el método comprende además los pasos de: insertar el micrófono de oído izquierdo en el oído izquierdo de la persona e insertar el micrófono de oído derecho en el oído derecho de dicha persona; el dispositivo informático se conecta eléctricamente a los micrófonos de oído izquierdo y derecho, y se conecta operativamente a la unidad de orientación; y el dispositivo informático captura la pluralidad de muestras de audio izquierdas y de muestras de audio derechas y recupera o recibe o lee u obtiene de otra manera la información de orientación de dicha unidad de orientación de manera directa o indirecta; y en donde el dispositivo informático almacena dichos datos en una memoria.

10 En una realización como esta, todos los pasos, incluyendo la captura de datos reales, son realizados por el dispositivo informático, el cual puede ser por ejemplo un ordenador de sobremesa o un ordenador portátil equipado con un dispositivo USB con una entrada de audio estéreo o similar. Si se utiliza una unidad de orientación de un teléfono inteligente en esta realización, el ordenador recuperaría la información de orientación del teléfono inteligente, por ejemplo a través de un cable de conexión o a través de una conexión inalámbrica, y la única tarea el teléfono inteligente sería proporcionar los datos de orientación.

En una realización, el dispositivo informático es un dispositivo portátil que también incluye la unidad de orientación.

20 En una realización de este tipo, todos los pasos a) a e), incluyendo la captura de datos reales, son realizados en el dispositivo portátil, por ejemplo por el teléfono inteligente. Se destaca de manera explícita que esto ya es técnicamente posible con muchos teléfonos inteligentes del año 2015, aunque el procesamiento puede llevar un tiempo relativamente largo (por ejemplo, del orden de 30 minutos para código no optimizado), pero se contempla que está velocidad se pueda mejorar aún más en el futuro cercano.

En una realización, el dispositivo portátil es un teléfono inteligente.

25 En una realización, el dispositivo portátil comprende además un altavoz; y en el cual el dispositivo portátil está adaptado además para analizar la información de orientación a fin de verificar si un espacio 3D alrededor de la cabeza se muestrea suficientemente, de acuerdo con un criterio predefinido; y está adaptado además para reproducir respectivamente un primer y un segundo mensajes de audio predefinidos por medio del altavoz del dispositivo portátil dependiendo de la salida del análisis si el espacio 3D se muestrea suficientemente o no.

30 El criterio predefinido para decidir si el espacio 3D se muestrea suficientemente puede estar basado por ejemplo en una densidad predefinida mínima en un subespacio predefinido. El subespacio puede ser por ejemplo un espacio definido por una parte significativa de una esfera completa.

35 Es una ventaja de dicha realización que se proporciona alguna forma de control e interacción durante la captura de datos o brevemente después de la misma, antes de que empiece la estimación real de la ITDF y la HRTF. De esta manera la precisión de las ITDF y HRTF individualizadas estimadas se puede incrementar, y el riesgo de malas percepciones durante la reproducción de datos de audio en un sistema 3D-VAS, debido a interpolación de curvas ITDF y HRTF en un espacio 3D muestreado de forma poco precisa, se puede reducir.

40 Aunque la información de orientación puede tener precisión insuficiente para ser utilizada directamente como información de dirección de la cual procede un sonido cuando se determina la HRTF, la precisión es típicamente suficiente para permitir verificación de si el espacio 3D alrededor de la cabeza de la persona es o no suficientemente muestreado. Por supuesto, puede haber más de dos mensajes predefinidos. Ejemplos de dichos mensajes pueden contener por ejemplo el mensaje de que "el ensayo ha terminado", o de que "es necesario repetir el ensayo", o de que "se requiere muestreo adicional cuando se mira a la derecha y hacia arriba", o cualquier otro mensaje.

45 En una realización, la señal de ensayo de audio comprende una pluralidad de estímulos acústicos, en donde cada uno de los estímulos acústicos tiene una duración en el rango de 25 a 50 ms; y/o en donde un periodo de tiempo entre estímulos acústicos consecutivos es un periodo en el rango de 250 a 500 ms.

En una realización, los estímulos acústicos son estímulos acústicos de banda ancha, en particular chirps.

Se observa que una señal de ensayo acústica con tonos puros probablemente también funcionaría, pero se requeriría mucho más tiempo para obtener la misma calidad de ITDF y de HRFT.

En una realización, los estímulos acústicos tienen una frecuencia instantánea que disminuye linealmente con el tiempo.

50 Es una ventaja la utilización de señales de estímulos acústicos de banda ancha (en vez de señales de tonos puros), porque las señales de gran ancho de banda permiten extracción de la información espectral y por tanto estimación de la HRTF en todo el rango completo de frecuencias de interés para cada orientación de la cabeza, y también porque la precisión de la estimación de la ITD es mayor para señales de gran ancho de banda.

- 5 Es una ventaja la utilización de señales de ensayo con estímulos acústicos que tienen una duración menor de 50 ms, porque para una señal tan corta, se puede suponer de manera razonable que la cabeza está (momentáneamente) quieta, incluso en la práctica puede estar girando (y típicamente lo estará), suponiendo que la persona está girando lentamente su cabeza a una velocidad angular relativamente baja (por ejemplo a menos de 60° por segundo), y no bruscamente.
- Es también una ventaja que dichas señales de corta duración permiten solapamiento entre la recepción a lo largo del camino directo y la recepción de la misma señal a lo largo de un camino en directo que contiene al menos una reflexión adicional sobre una de las fronteras de la habitación, u objetos presentes dentro de la habitación. Por tanto, se pueden evitar técnicas de cancelación de eco complejas.
- 10 En una realización, el método comprende además el paso de: seleccionar, dependiendo de un análisis de los conjuntos de datos capturados, un mensaje de audio predefinido de un grupo de mensajes de audio predefinidos, y reproducir dicho mensaje de audio seleccionado por medio del mismo altavoz que se utilizó para los estímulos de ensayo o por medio de un segundo altavoz diferente al primer altavoz, para proporcionar información o instrucciones a la persona antes y/o durante y/o después de la reproducción de la señal de ensayo de audio.
- 15 En una realización, el segundo altavoz puede ser por ejemplo el altavoz de un dispositivo portátil.
- Dicha realización puede ser útil por ejemplo en un procesamiento en tiempo (casi) real del paso c), por el cual información de posición y/o orientación (precisa o aproximada) se extrae de un subconjunto de las muestras capturadas, o idealmente en el tiempo entre cada muestra de audio sucesiva, y por lo cual el algoritmo verifica además si el espacio tridimensional alrededor de la cabeza es o no muestreado con suficiente densidad, y por el cual se proporciona al usuario realimentación acústica correspondiente, después, o incluso antes de que finalice el fichero de ensayo acústico.
- 20 Pero por supuesto también se podrían proporcionar otros mensajes, por ejemplo una instrucción de texto para que el usuario mantenga su cabeza quieta durante un cierto número de estímulos acústicos (por ejemplo 5 o 10) para permitir el cálculo de la media de las muestras de audio recogidas para esa orientación particular, de modo que se pueda conseguir una relación señal ruido (SNR) mayor.
- 25 Por supuesto, la misma funcionalidad también puede ser proporcionada por una aplicación en tiempo no real, en la cual por ejemplo la señal de ensayo acústico se reproduce una primera vez, y se captura una primera pluralidad de conjuntos de datos, siendo a continuación dicha primera pluralidad de muestras de datos procesada en el paso c), y por lo cual el paso c) comprende además una verificación de si el espacio alrededor de la cabeza es muestreado o no con suficiente densidad, y por la cual se proporciona al usuario un correspondiente mensaje acústico por medio del segundo altavoz, por ejemplo para informarle de que la captura es suficiente, o pidiéndole que repita la medida, opcionalmente de este modo dando instrucciones adicionales para orientar la cabeza en ciertas direcciones.
- 30 De esta manera el paso real de captura de datos se puede hacer bastante interactivo entre el ordenador y la persona, con el efecto técnico de que la HRTF se estima con al menos una densidad predefinida.
- 35 De esta manera el riesgo de muestreo espacial insuficiente, y por tanto el riesgo de tener que interpolar entre dos o más curvas ITDF, respectivamente curvas HRTF para una dirección que no fue muestreada espacialmente de forma suficientemente densa, se puede reducir (aún más).
- En un segundo aspecto, la presente invención está relacionada con un método de reproducir una señal de audio virtual para una persona particular, que comprende: x) estimar una función de transferencia relacionada con la cabeza individualizada y una función de diferencia de tiempo interaural individualizada de dicha persona particular utilizando un método de acuerdo con cualquiera de las reivindicaciones anteriores; y) generar una señal de audio virtual para la persona particular, haciendo uso de la función de transferencia relacionada con la cabeza individualizada y de la función de diferencia de tiempo interaural individualizada estimadas en el paso x); z) reproducir la señal de audio virtual generada en el paso y) utilizando unos auriculares estéreo y/o un conjunto de altavoces de oído.
- 40 En un tercer aspecto, la presente invención está relacionada con un producto de programa informático para estimar una función de transferencia relacionada con la cabeza individualizada y una función de diferencia de tiempo interaural de una persona particular, estando dicho producto de programa informático, cuando es ejecutado en al menos un dispositivo informático que comprende un procesador programable y una memoria, programado para realizar al menos los pasos c) a e) de un método de acuerdo con el primer aspecto o con el segundo aspecto.
- 45 El producto de programa informático puede comprender un módulo software ejecutable en un primer ordenador por ejemplo un ordenador portátil o de sobremesa, estando adaptado el primer módulo para realizar el paso a) relacionado con la captura y almacenamiento de los datos de audio y orientación, incluyendo opcionalmente almacenar los datos en una memoria, y los pasos c) a e) relacionados con la estimación o el cálculo de una IDTF y HRTF personalizadas, cuando el primer ordenador está conectado de manera apropiada a los micrófonos de oído (por ejemplo a través de cables eléctricos) y conectado operativamente (por ejemplo mediante Bluetooth) a la unidad de orientación.
- 50
- 55

5 El producto de programa informático puede comprender dos módulos software, uno ejecutable en un dispositivo portátil que comprende un módulo de orientación, tal como por ejemplo un teléfono inteligente, y un segundo módulo ejecutable en un segundo ordenador, por ejemplo un ordenador portátil o de sobremesa, estando adaptado el primer módulo para realizar al menos el paso a) relacionado con la captura de datos, incluyendo también preferiblemente almacenar los datos en una memoria, estando adaptado el segundo módulo para realizar al menos los pasos c) a e) relacionados con la estimación o el cálculo de una ITDF y una HRTF personalizadas. Durante la captura de datos el dispositivo portátil está conectado de manera apropiada a los micrófonos de oído (por ejemplo por medio de cables eléctricos).

10 El producto de programa informático puede comprender módulos software adicionales para transferir los datos capturados desde el dispositivo portátil al ordenador, por ejemplo por medio de una conexión mediante cables o inalámbrica (por ejemplo, por medio de Bluetooth o wifi). De forma alternativa los datos se pueden transferir del dispositivo portátil al ordenador por medio de una tarjeta de memoria o similar. Por supuesto también es posible una mezcla de mecanismos de transferencia.

15 El cuarto aspecto está relacionado con un portador de datos que comprende el producto de programa informático de acuerdo con el tercer aspecto.

En un ejemplo, el portador de datos comprende además una representación digital de dicha señal de ensayo acústica.

El quinto aspecto está relacionado con la transmisión de un producto de programa informático de acuerdo con el tercer aspecto.

20 La transmisión también puede incluir la transmisión de producto de programa informático en combinación con una representación digital de dicha señal de ensayo acústica.

El sexto aspecto está relacionado con un kit de partes, que comprende: un portador de datos de acuerdo con el cuarto aspecto, y un micrófono de oído izquierdo y un micrófono de oído derecho.

25 Es una ventaja de dicho kit de partes que proporciona todo el hardware que necesita un usuario final típico (además del ordenador y/o del teléfono inteligente y del equipo de audio que ya tiene), para estimar su ITDF individualizada y su HRTF individualizada. Este kit de partes se puede proporcionar como un paquete independiente, o junto con por ejemplo un juego en 3D, u otro paquete software. La señal de ensayo acústica se puede descargar por ejemplo de una página web particular de internet, y se puede grabar en un CD de audio, o se puede escribir en un lápiz de memoria, o se puede obtener de otra manera.

30 En un ejemplo, el kit de partes comprende además: un segundo portador de datos que comprende una representación digital de dicha señal de ensayo acústica.

El segundo portador de datos puede ser por ejemplo un CD de audio que se puede reproducir en un equipo estéreo estándar, o un DVD que se puede reproducir en un reproductor de DVD o en un dispositivo de cine en casa.

Estos y otros aspectos de la invención resultarán evidentes de, y serán explicados con referencia a, las una o más realizaciones descritas en lo que sigue.

35 **Breve descripción de los dibujos**

La Figura 1 ilustra cómo llega el sonido desde una dirección particular en diferentes instantes al oído izquierdo y al oído derecho de una persona, y como ambos oídos imponen un filtrado espectral diferente.

40 La Figura 2 es una representación esquemática de diferentes marcos de referencia que se pueden utilizar en realizaciones de la presente invención: un marco de referencia fijado a la unidad de orientación montada en la cabeza o a la cabeza, un marco de referencia del mundo, que es cualquier marco fijo al mundo (o a la "tierra") utilizado por la unidad de orientación, y un marco de referencia fijado a la cabeza, que se define como el "marco de referencia de la cabeza" utilizado en medidas de HRTF y de ITDF estándar (véanse también la Figura 3 y la Figura 4). La "dirección de la fuente con respecto a la cabeza" (es decir, la dirección de los uno o más altavoces con respecto al marco de referencia de la cabeza fijado en un punto a medio camino entre los dos oídos) está definida por un ángulo lateral θ y una elevación ϕ . El ángulo lateral es el ángulo entre la "dirección de la fuente" y el eje oído-oído, y la elevación es el ángulo entre la "dirección de la fuente" y el plano nariz-oído-oído. La dirección de la fuente es la línea virtual que va desde el altavoz hasta la posición media del centro de la cabeza durante el ensayo.

50 La Figura 3 muestra un ejemplo de una función de diferencia de tiempo interaural (ITDF) de una persona particular, en la cual se utiliza diferente intensidad (escala de grises) para indicar diferentes valores de la diferencia de tiempo interaural (ITD), dependiendo de la dirección de la cual procede el sonido. Con líneas curvas blancas se muestran contornos de iso-ITD.

La Figura 4 muestra un ejemplo de una función de transferencia relacionada con la cabeza (HRTF) monoaural (oído izquierdo) de una persona particular a lo largo del plano mediano, en el cual se utiliza diferente intensidad (escala de grises) para indicar diferentes valores. Con líneas curvas blancas se muestran contornos de iso-respuesta.

5 La Figura 5 muestra una disposición para medir una HRTF fuera de una cámara anecoica, conocida en la técnica anterior.

La Figura 6 muestra un primer ejemplo de una posible configuración hardware para realizar uno o más pasos de un método de acuerdo con la presente invención, por el cual la captura de datos es realizada por un ordenador conectado eléctricamente a micrófonos de oído, y por el cual se obtienen datos de orientación de una unidad sensora presente en un teléfono inteligente montado de manera fija en una posición arbitraria en o a la cabeza de la persona.

10 La Figura 7 muestra un segundo ejemplo de una posible configuración hardware para realizar uno o más pasos de un método de acuerdo con la presente invención, por el cual la captura de datos es realizada por un teléfono inteligente conectado eléctricamente a micrófonos de oído, y por el cual se obtienen datos de orientación de una unidad sensora presente en el teléfono inteligente, y por el cual el procesamiento de datos es realizado también por el teléfono inteligente.

15 La Figura 8 muestra un tercer ejemplo de una posible configuración hardware para realizar uno o más pasos del método de acuerdo con la presente invención, por el cual la captura de datos es realizada por un teléfono inteligente conectado eléctricamente a micrófonos de oído, y por el cual se obtienen datos de orientación de una unidad sensora presente en el teléfono inteligente, y por el cual el procesamiento de datos se descarga a un ordenador o a "la nube".

20 La Figura 9 ilustra las variables que se deben estimar en el método de la presente invención, por tanto ilustra el problema que debe ser resuelto por la parte de procesamiento de datos del algoritmo utilizado en las realizaciones de la presente invención.

La Figura 10 es una representación de diagrama de flujo de una primera realización de un método para determinar una ITDF y una HRTF personalizadas de acuerdo con la presente invención.

25 La Figura 11 es una representación de diagrama de flujo de una segunda realización de un método para determinar una ITDF y una HRTF personalizadas de acuerdo con la presente invención.

La Figura 12 muestra un método para estimar orientaciones del teléfono inteligente con respecto al mundo, como el que se puede utilizar en el bloque 1001 de la Figura 10 y en el bloque 1101 de la Figura 11.

La Figura 13 muestra un método para estimar direcciones de la fuente con respecto al mundo, como el que se puede utilizar en el bloque 1002 de la Figura 10 y en el bloque 1102 de la Figura 11.

30 La Figura 14 muestra un método para estimar orientaciones del teléfono inteligente con respecto a la cabeza, como el que se puede utilizar en el bloque 1003 de la Figura 10 y en el bloque 1103 de la Figura 11.

La Figura 15 muestra un método para estimar la posición del centro de la cabeza con respecto al mundo, como el que se puede utilizar en el bloque 1004 de la Figura 10 y en el bloque 1104 de la Figura 11.

35 La Figura 16 muestra un método para estimar la HRTF y la ITDF, como el que se puede utilizar en el bloque 1005 de la Figura 10 y en el bloque 1105 de la Figura 11.

La Figura 17 muestra un diagrama de flujo de funcionalidad adicional opcional como el que se puede utilizar en las realizaciones de la presente invención.

La Figura 18 ilustra la captura de la información de orientación de una unidad de orientación montada de manera fija a la cabeza.

40 Las Figuras 18(a) a 18(d) muestran un ejemplo de datos de sensor como los que se pueden obtener a partir de una unidad de orientación montada de manera fija a una cabeza.

La Figura 18(e) muestra una plataforma de ensayo robótica como la que se utilizó durante la evaluación.

45 Las Figuras 19(a) a 19(d) son instantáneas de una persona realizando movimientos lentos de cabeza durante la captura de datos de audio y de datos del sensor de orientación para permitir la determinación de la ITDF y la HRTF de acuerdo con la presente invención.

La Figura 20 es un esbozo de una persona sentada en una silla en una habitación típica de una casa, a una distancia típica de un altavoz.

50 La Figura 21 ilustra características de un así llamado "chirp" que tiene una duración de tiempo predefinida y un barrido de frecuencia lineal, que se puede utilizar como estímulos de ensayo de audio en las realizaciones de la presente invención.

Las Figuras 22(a) a 22(c) ilustran posibles pasos para extraer el tiempo de llegada de los chirps y para extraer información espectral de los chirps.

5 La Figura 22(a) muestra el espectrograma de una señal de audio capturada por el micrófono de oído izquierdo, para una señal de ensayo de audio que comprende cuatro chirps consecutivos, cada uno de los cuales tiene una duración de aproximadamente 25 ms con un intervalo entre chirps de 275 ms.

La Figura 22(b) muestra el espectrograma 'rectificado', es decir, cuando se han compensado en él los retrasos de temporización dependientes de la frecuencia conocidos en los chirps.

La Figura 22(c) muestra la intensidad sumada del espectrograma 'rectificado' de una señal de audio capturada por el micrófono de oído izquierdo, en base a la cual se pueden determinar los tiempos de llegada de los chirps.

10 La Figura 23 muestra un ejemplo de los espectros extraídos de la señal de audio izquierda (figura 23a: espectros del oído izquierdo) y de los espectros extraídos de la señal de audio derecha (figura 23b: espectros del oído derecho), y la diferencia de tiempo interaural (figura 23c) para una señal de ensayo de audio ejemplar que comprende cuatro mil chirps.

La Figura 24 muestra parte de los espectros y de los datos de ITD de la Figura 23 con mayor detalle.

15 La Figura 25(a) muestra un mapeado de los datos de ITD de los cuatro mil chirps de la Figura 23 sobre una superficie esférica, utilizando una dirección de la fuente aleatoria (pero incorrecta), dando como resultado una función con un alto grado de irregularidades o con baja suavidad.

20 La Figura 25(b) muestra un mapeado de los datos de ITD de los cuatro mil chirps de la Figura 23 sobre una superficie esférica, utilizando la dirección de la fuente correcta, dando como resultado una función con un alto grado de regularidades o alta suavidad.

Las Figuras 25(a, b) muestran el efecto perjudicial de una dirección de la fuente supuesta de manera incorrecta sobre la suavidad de la superficie proyectada de las medidas ITD.

Las Figuras 25(c, d) muestran el mismo efecto para datos espectrales.

25 La Figura 26(a) muestra un conjunto de funciones base armónicas esféricas reales de bajo orden, que se pueden utilizar para generar o definir funciones que tengan sólo variaciones espaciales que varíen lentamente. Estas funciones se pueden utilizar para definir superficies "suaves".

La Figura 26(b) muestra una técnica para cuantificar suavidad de una función definida sobre la esfera, por ejemplo la ITDF, la cual se puede utilizar como una métrica de suavidad.

30 La Figura 27(a) muestra el valor de suavidad de acuerdo con la métrica de suavidad definida en la Figura 26(b) para dos mil "direcciones de la fuente" candidatas representadas sobre una esfera, cuando se aplican a los valores de ITD, con el orden de los armónicos esféricos configurado a 5. La escala de grises está ajustada en la Figura 27(b).

35 La Figura 28(a) muestra los valores de suavidad, cuando se aplica el criterio de suavidad a espectros binaurales, con el orden de los armónicos esféricos configurado a 5, siendo el valor de suavidad para cada coordenada mostrado en la esfera la suma del valor de suavidad para cada una de las frecuencias en el rango de 4 kHz a 20 kHz, en pasos de 300 Hz. La escala de grises está ajustada en la Figura 28(b).

La Figura 29(a) muestra los valores de suavidad, cuando se aplica el criterio de suavidad a espectros binaurales, con el orden de los armónicos esféricos configurado a 15. La escala de grises está ajustada en la Figura 29(b).

La Figura 30(a) muestra los valores de suavidad, cuando se aplica el criterio de suavidad a espectros monoaurales, con el orden de los armónicos esféricos configurado a 15. La escala de grises está ajustada en la Figura 30(b).

40 La Figura 31 ilustra los parámetros del modelo de un modelo a priori del movimiento del centro de la cabeza. Cuando una persona está sentada en una silla de oficina y se le permite que gire su cabeza libremente en todas direcciones, y que gire libremente junto con la silla con el cuerpo fijado a la silla, entonces el movimiento del centro de la cabeza se puede describir utilizando este modelo mecánico simplificado.

45 La Figura 32 muestra instantáneas de un video que captura a un sujeto cuando se está realizando una medida HRTF en la silla con el giro libre. Utilizando el modelo mecánico de la Figura 31, se extrajo información sobre la posición de la cabeza, (la cual produjo como resultado mejores estimaciones de la dirección de la fuente con respecto a la cabeza), como se puede ver a partir de las visualizaciones de la orientación y la posición estimadas de la cabeza. La línea negra muestra la desviación del centro de la cabeza.

50 La Figura 33 es una representación gráfica de las posiciones estimadas (en coordenadas del mundo X, Y, Z) del centro de la cabeza durante un ensayo de captura de audio ejemplar, utilizando el modelo mecánico de la Figura 31.

La Figura 34 muestra una medida de la distancia entre el centro de la cabeza y la fuente del sonido a lo largo del tiempo, determinada a partir de los retrasos de temporización entre chirps consecutivos. El modelo mecánico de la Figura 31 permite un buen ajuste con estas variaciones de la distancia medida.

5 La Figura 35 muestra una comparación de dos HRTFs de la misma persona: una se midió en una instalación profesional (en Aquisgrán), la otra HRTF se obtuvo utilizando un método de acuerdo con la presente invención, medida en casa. Como se puede ver, existe muy buena correspondencia entre la representación gráfica de la HRTF medida en la instalación profesional y la HRTF medida en casa.

10 Los dibujos son sólo esquemáticos y no son limitativos. En los dibujos, el tamaño de algunos de los elementos puede estar exagerado y no dibujado a escala con fines ilustrativos. Cualquier signo de referencia en las reivindicaciones no se debería interpretar que limita el alcance.

En los diferentes dibujos, los mismos signos de referencia hacen referencia a elementos iguales o análogos.

Descripción detallada de realizaciones ilustrativas

15 La presente invención se describirá con respecto a realizaciones particulares y con referencia a ciertos dibujos, pero la invención no está limitada a ellos sino sólo por las reivindicaciones. Los dibujos descritos son sólo esquemáticos y son no limitativos. En los dibujos, el tamaño de algunos de los elementos puede estar exagerado y no dibujado a escala con fines ilustrativos. Las dimensiones y las dimensiones relativas no se corresponden con puestas en práctica reales de la invención.

20 Además, los términos primero, segundo y similares en la descripción y en las reivindicaciones, se utilizan para distinguir entre elementos similares y no necesariamente para describir una secuencia, ya sea temporalmente, espacialmente, en ranking o en cualquier otra manera. Se debe comprender que los términos así utilizados son intercambiables bajo circunstancias apropiadas y que las realizaciones de la invención descritas en esta memoria son capaces de operar en otras secuencias diferentes a las descritas o ilustradas en esta memoria.

25 Además, los términos encima de, debajo de y similares en la descripción y en las reivindicaciones se utilizan con fines descriptivos y no necesariamente para describir posiciones relativas. Se debe entender que los términos así utilizados son intercambiables bajo circunstancias apropiadas y que las realizaciones de la invención descritas en esta memoria son capaces de operar en otras orientaciones diferentes a las descritas o ilustradas en esta memoria.

30 Se debe observar que el término “que comprende”, utilizado en las reivindicaciones, no se debería interpretar como restringido a los medios enumerados después de él; no excluye a otros elementos o pasos. Por lo tanto, se debe interpretar como que especifica la presencia de los rasgos, números enteros, pasos o componentes indicados como aquellos a los que se hace referencia, pero no impide la presencia o adición de uno o más otros rasgos, números enteros, pasos o componentes, o grupos de los mismos. De esta manera, el alcance de la expresión “un dispositivo que comprende medios A y B” no debería estar limitado a dispositivos que consisten sólo en componentes A y B. Esto significa que, con respecto a la presente invención, los únicos componentes relevantes del dispositivo son A y B.

35 Una referencia a lo largo de toda esta especificación a “una realización” significa que un rasgo, estructura o característica particular descrito en conexión con la realización está incluido en al menos una realización de la presente invención. De esta manera, apariciones de la frase “en una realización” en diferentes lugares a lo largo de toda esta especificación no hacen referencia todas ellas necesariamente a la misma realización, pero pueden hacerlo. Además, los rasgos, estructuras o características particulares se pueden combinar de cualquier manera apropiada, como resultaría evidente para una persona de habilidad ordinaria en la técnica a partir de esta divulgación, en una o más realizaciones.

40 De manera similar se debería apreciar que, en la descripción de realizaciones ejemplares de la invención, a veces se agrupan juntos diferentes rasgos de la invención en una única realización, Figura, o descripción de la misma con el propósito de hacer más eficiente la divulgación y ayudar en la comprensión de uno o más de los diferentes aspectos innovadores. Sin embargo, no se debe interpretar que este método de divulgación refleja una intención de que la invención reivindicada requiere más rasgos de los que son expresamente mencionados en cada reivindicación. Más bien, como reflejan las siguientes reivindicaciones, los aspectos innovadores radican en no todos los rasgos de una única realización descrita anteriormente.

45 En la descripción proporcionada en esta memoria, se describen numerosos detalles específicos. Sin embargo, se entiende que las realizaciones de la invención se pueden poner en práctica sin estos detalles específicos. En otros casos, métodos, estructuras y técnicas bien conocidos no se han mostrado en detalle para no oscurecer una comprensión de esta descripción.

50 En el contexto de la presente invención, con “diferencia de tiempo interaural” o “ITD” se hace referencia a una diferencia de tiempo, que se puede representar mediante un valor (por ejemplo en milisegundos), pero este valor es diferente dependiendo de la dirección de la que procede el sonido (con respecto a la cabeza). En esta memoria se hace referencia a la representación de valores de ITD para diferentes direcciones como la “función de diferencia de tiempo interaural” o “ITDF”, y un ejemplo de dicha función se muestra en la Figura 3.

- 5 En el contexto de la presente invención, con “función de transferencia relacionada con la cabeza” o “HRTF” se hace referencia al conjunto de funciones espectrales binaurales (como se muestra en la Figura 4 sólo para el oído izquierdo, para el plano mediano), representando cada función espectral $S(f)$ (correspondiendo los valores con cada línea horizontal de la Figura 4) las características de filtrado espectral impuestas por el cuerpo, la cabeza, y el oído izquierdo/derecho en el sonido que procede de una dirección particular (con respecto a la cabeza).
- 10 Allí donde en la presente invención se hace referencia a “sistema de referencia del mundo”, lo que se quiere indicar con ello es un sistema de referencia 3D fijo al mundo (o a la “tierra”) en el valor medio del centro de la cabeza del sujeto, que se puede definir eligiendo un eje Z a lo largo del eje gravitatorio que apunte alejándose del centro de la tierra, un eje X que está situado en el plano horizontal y que apunta en la dirección del norte magnético y un eje Y que también está situado el plano horizontal y que forma un sistema de coordenadas 3D ortogonal a derechas con los otros dos ejes.
- 15 Allí donde en la presente invención se hace referencia a “posición de un objeto”, lo que se quiere indicar es una ubicación particular en un espacio 3D, como puede ser indicado por ejemplo por coordenadas X, Y, Z específicas con respecto al marco de referencia del mundo, pero también se pueden utilizar otras coordenadas.
- 20 Allí donde en la presente invención se hace referencia a “orientación de un objeto”, lo que se quiere indicar es la orientación de un marco de referencia 3D fijo al objeto cuya orientación se puede expresar por ejemplo mediante 3 ángulos de Euler con respecto al marco de referencia del mundo, pero también se pueden utilizar otras coordenadas.
- 20 Allí donde en la presente invención se hace referencia a “dirección de la fuente de sonido con respecto a la cabeza”, lo que se quiere indicar es una dirección particular con respecto al marco de referencia de la cabeza como el utilizado en medidas de HRTF e ITDF estándar. Esta dirección se expresa típicamente mediante dos ángulos: un ángulo lateral θ y un ángulo de elevación ϕ como se muestra por ejemplo en la Figura 2, por lo cual el ángulo lateral θ es un valor en el rango de cero a π , y el ángulo de elevación ϕ es un valor en el rango de $-\pi$ a $+\pi$.
- 25 Cuando se hace referencia a “dirección hasta el signo” esto se refiere tanto a la dirección caracterizada por los dos ángulos (θ, ϕ) como a la dirección caracterizada por los dos ángulos $(\pi-\theta, \pi+\phi)$.
- 25 Allí donde en la presente invención se hace referencia a “dirección de la fuente de sonido con respecto al mundo”, lo que se quiere indicar es una dirección particular con respecto al marco de referencia del mundo.
- 30 En la presente invención se hace referencia al “sensor de orientación” o a la “unidad de orientación” en lugar de a un sensor de posición (6D), porque estamos interesados principalmente en la orientación de la cabeza, y la información de posición (X, Y, Z) no es necesaria para estimar la HRTF y la ITDF. Sin embargo, si está disponible, la información de posición (X, Y, Z) también puede ser utilizada por el algoritmo para estimar la posición del centro de la cabeza definida como el punto a medio camino entre las posiciones del oído izquierdo y del oído derecho.
- 35 En este documento, los términos “HRTF media” y “HRTF generalizada” se utilizan como sinónimos, y hacen referencia a un tipo de HRTF promediada o común de un grupo de personas.
- 35 En este documento, los términos “ITDF media” e “ITDF generalizada” se utilizan como sinónimos, y hacen referencia a un tipo de ITDF promediada o común de un grupo de personas.
- En este documento, los términos “HRTF personalizada” y “HRTF individualizada” se utilizan como sinónimos, y hacen referencia a la HRTF de una persona particular.
- En este documento, los términos “ITDF personalizada” e “ITDF individualizada” se utilizan como sinónimos, y hacen referencia a la ITDF de una persona particular.
- 40 Allí donde en la presente invención se utiliza la expresión “la dirección de la fuente con respecto a la cabeza”, lo que se quiere indicar es en realidad la dirección momentánea de la fuente con respecto al “marco de referencia de la cabeza” como se muestra en la Figura 2, en un momento particular en el tiempo, por ejemplo cuando se captura un fragmento de audio izquierdo y un fragmento de audio derecho particulares. Dado que la persona está moviendo su cabeza, la dirección de la fuente cambiará durante el ensayo, aunque la fuente permanezca estacionaria.
- 45 En este documento los términos “información de orientación” y “datos de orientación” se utilizan a veces como sinónimos, o a veces se hace una distinción entre los “datos en bruto” que se pueden obtener de un sensor de orientación, por ejemplo un giróscopo, y los datos convertidos, por ejemplo los ángulos θ y ϕ , en cuyo caso se hace referencia a los datos en bruto como información de orientación, y los datos procesados
- 50 En este documento, la abreviatura “c. r. al mundo” significa “con respecto al mundo”, lo cual es equivalente a “en coordenadas del mundo” también abreviado como “en c. del m.”.
- Allí donde en la presente invención, se utiliza el término “estimar”, “estimado” o “estimada”, este término se debería interpretar de manera general. Dependiendo del contexto, puede significar por ejemplo “medir”, o “medir y corregir” o “medir y calcular” o “calcular” o “aproximar”, etc.

En este documento los términos “datos de audio binaurales” pueden hacer referencia a las “muestras de audio izquierda y derecha” si se refiere a muestras individuales, o a “fragmentos de audio izquierdos y derechos” si se refiere a una secuencia de muestras izquierdas y respectivamente derechas, correspondientes a un chirp.

5 En este documento, el término “fuente” y el término “altavoz” se utilizan como sinónimos, a menos que se indique algo diferente de manera explícita.

A menos que se mencione otra cosa de manera explícita, “modelo mecánico” o “modelo cinemático” se utilizan como sinónimos.

Los inventores se enfrentaron al problema de encontrar una manera de personalizar la HRTF y la ITDF de una manera simple (para el usuario), y a un coste reducido (para el usuario).

10 El método propuesto intenta combinar dos requisitos (contradictorios):

(1) la necesidad de una recogida suficiente de datos informativos de modo que la ITDF y la HRTF se puedan estimar de manera suficientemente precisa (o, dicho de otra manera: de modo que la ITDF y la HRTF verdaderas de cada individuo se puedan aproximar de manera suficientemente precisa), y

15 (2) la limitación de que el procedimiento (o, dicho con mayor precisión: la parte en la que se capturan los datos) se pueda realizar en casa y no sea demasiado difícil para un usuario medio.

Los inventores consiguieron llegar a un método que tiene dos pasos principales:

20 1) un primer paso de captura de datos, que es simple de realizar, y que utiliza hardware que está habitualmente disponible en casa: un dispositivo de reproducción de sonido (por ejemplo cualquier cadena de música mono o estéreo o reproductor MP3 o similar, conectable a un altavoz) y un sensor de orientación (como el que está disponible hoy en día por ejemplo en los teléfonos inteligentes). El usuario sólo necesita comprar un conjunto de micrófonos de oído.

25 2) un segundo paso de procesamiento de datos, que se puede realizar por ejemplo en el mismo teléfono inteligente, o en otro dispositivo informático tal como un ordenador de sobremesa o un ordenador portátil, o incluso en la nube. En el segundo paso se ejecuta un algoritmo que está ajustado a los detalles particulares del paso de captura de datos, y que tiene en cuenta que las características espectrales del altavoz y de los micrófonos pueden no ser conocidas, y que la posición de la persona con respecto al altavoz puede no ser conocida, y que la posición/orientación de la unidad de orientación en la cabeza de la persona puede no ser (exactamente) conocida y opcionalmente también que la precisión de los datos de orientación proporcionados por la unidad de orientación puede no ser muy precisa (por ejemplo tiene una tolerancia de +/- 5°).

30 La ITDF y la HRTF resultantes de este compromiso pueden no ser perfectas, pero son suficientemente precisas para permitir al usuario localizar (aproximadamente) una fuente de sonido en espacio 3D, en particular en términos de discernir la parte de delante de la parte de atrás, creando de esta manera una sensación espacial con un valor añadido para el usuario. Además, el usuario final se enfrenta principalmente a las ventajas del primer paso (de captura de datos), y no se enfrenta a la complejidad del paso de procesamiento de datos.

35 En el resto de este documento, se analizará primero una solución de la técnica anterior con referencia a la Figura 5. A continuación se explicará con mayor detalle el paso de captura de datos de la presente invención con referencia a las Figuras 6 a 8. Por último se explicará con mayor detalle el paso de procesamiento de datos de la presente invención con referencia a las Figuras 9 a 29.

40 Se hace también referencia a una solicitud de patente internacional en tramitación con la presente PCT/EP2016/053020 (WO2016/134982) de los mismos inventores, a la cual se hará referencia en lo que sigue en esta memoria como la “solicitud de patente previa”, la cual aún no se ha publicado, por tanto técnica anterior bajo el Artículo 54(3) en Europa, que tiene algunas características comunes con la presente invención, pero también importantes diferencias, como se explicará más adelante.

I. Solución conocida

45 La Figura 5 es una copia de la Figura 1 del documento US5729612A, e ilustra una realización de una configuración de ensayo conocida, fuera de una cámara anecoica, por la cual una persona 503 está sentada en una silla, a una distancia conocida de un altavoz 502, que está montado en un soporte especial 506 para permitir que se pueda mover el altavoz en la dirección de la altura. Una señal de audio izquierda y una señal de audio derecha son capturadas por dos micrófonos de oído 505 que lleva puestas la persona. Los movimientos de la cabeza de la persona son monitorizados por un sensor de posición 504 montado encima de la cabeza de la persona que está sentada en una silla 507 que se puede orientar en direcciones particulares (como se indica mediante líneas en el suelo). Los micrófonos 505 y el sensor de posición 504 están conectados eléctricamente a un ordenador 501 por medio de cables. El ordenador 501 envía una señal de ensayo acústica al altavoz 502, y controla la posición vertical del altavoz 502 utilizando el soporte especial 506.

Los datos se procesarán en el ordenador 501, pero el documento no dice nada acerca de cómo se calculan exactamente la ITDF y la HRTF a partir de las señales de audio y las señales de posición medidas. El documento menciona un paso de calibración para determinar una característica de transferencia del altavoz 502 y de los micrófonos 505 y el método también confía fuertemente en el hecho de que la posición relativa de la persona 503 y del altavoz 502 se conocen con exactitud.

II. Captura de datos:

Las Figuras 6 a 8 muestran tres ejemplos de posibles disposiciones de ensayo que se pueden utilizar para capturar datos de acuerdo con la presente invención, no estando la presente invención limitada a ellos.

En las configuraciones mostradas, una fuente de sonido 602, 702, 802, por ejemplo un altavoz, está situada a una distancia desconocida de la persona 604, 704, 804, pero aproximadamente a la misma altura que la cabeza de la persona. El altavoz puede estar situado por ejemplo en el borde de una mesa, y no es necesario moverlo. La persona 603, 703, 803 puede estar sentada en una silla o similar. La silla puede ser una silla giratoria, pero esto no es absolutamente necesario, y no es necesario hacer ninguna indicación en el suelo, y no es necesario que el usuario se oriente él mismo en direcciones particulares de acuerdo con las líneas en el suelo.

La persona lleva un micrófono de oído izquierdo en su oído izquierdo, y un micrófono de oído derecho en su oído derecho. Una unidad de orientación 604, 704, 804 está montada de forma fija a la cabeza de la persona, preferiblemente encima de la cabeza de la persona, o en la parte posterior de la cabeza de la persona, por ejemplo por medio de una cinta para la cabeza (no mostrada) o una correa o medios extensibles o medios elásticos. La unidad de orientación 604, 704, 804 puede estar situada en cualquier orientación arbitraria con respecto a la cabeza. La unidad de orientación puede comprender por ejemplo un acelerómetro y/o un giróscopo y/o un magnetómetro, y preferiblemente todos estos, pero también se puede utilizar cualquier otro sensor de orientación apropiado. En las realizaciones preferidas, la unidad de orientación permite determinar la orientación momentánea de la unidad de orientación con respecto al campo gravitatorio de la tierra y al campo magnético de la tierra, y por lo tanto no requiere un transmisor situado por ejemplo en las cercanías del altavoz. La unidad de orientación puede estar comprendida en un dispositivo portátil, tal como por ejemplo un teléfono inteligente. Es una ventaja importante de las realizaciones de la presente invención que no es necesario conocer con exactitud la posición y la orientación de la unidad de orientación con respecto a la cabeza, y que no es necesario que el sensor de orientación sea muy preciso (por ejemplo, una tolerancia de $\pm 10^\circ$ para el individuo puede ser perfectamente aceptable), como se explicará más adelante.

Durante el paso de captura de datos, una señal de ensayo acústica, por ejemplo un fichero de audio grabado previamente presente en un CD de audio, se reproduce en un equipo de reproducción de sonido 608, 708, 808 y se reproducirá a través del (único) altavoz 602, 702, 802. De forma alternativa se pueden utilizar dos o incluso más altavoces. La señal de ensayo acústica comprende una pluralidad de estímulos acústicos por ejemplo chirps que tienen una duración predefinida y un contenido espectral predefinido. En el contexto de esta invención, para mayor facilidad de explicación, los términos "chirp" y "estímulo" se utilizan de forma intercambiable y ambos se refieren a los estímulos acústicos. Preferiblemente se utilizan estímulos acústicos de una duración relativamente corta (por ejemplo en el rango de 25 ms a 50 ms) y con un espectro de banda ancha (por ejemplo en el rango de 1 kHz a 20 kHz), pero la invención no está limitada a ello, y también se pueden utilizar otras señales, por ejemplo tonos puros cortos.

Mientras la señal de ensayo acústica está siendo reproducida a través del altavoz, es necesario que la persona gire su cabeza lentamente en una pluralidad de orientaciones diferentes (véase la Figura 2).

Los estímulos acústicos de interés, por ejemplo los chirps, son capturados o registrados por medio de los micrófonos de oído izquierdo y derecho 605, 705, 805, y para cada estímulo registrado, datos de orientación de la unidad de orientación, también indicativos de la orientación de la cabeza en el momento en que el estímulo llega a los oídos (aunque esta orientación aún no es conocida, porque la unidad de orientación puede estar montada en cualquier posición arbitraria y en cualquier orientación arbitraria con respecto a la cabeza), también son capturados y/o registrados.

En la configuración de la Figura 6, los micrófonos de oído 605 están conectados eléctricamente (por medio de cables relativamente largos) al ordenador 601 que captura los datos de audio izquierdos y derechos, y que también recupera información de orientación de la unidad sensora de orientación 604 (conectada por cables o inalámbrica). El ordenador 601 puede almacenar a continuación la información capturada como conjuntos de datos, comprendiendo cada conjunto de datos una muestra de audio izquierda (L_i) que se origina en el micrófono de oído izquierdo y una muestra de audio derecha (R_i) que se origina en el micrófono de oído derecho e información de orientación (O_i) que se origina en la unidad de orientación. Se observa que el audio se muestrea típicamente a una frecuencia de al menos 40 kHz, por ejemplo a aproximadamente 44,1 kHz o a 48 kHz, pero también se pueden utilizar otras frecuencias. Los conjuntos de datos se pueden almacenar de cualquier manera apropiada, por ejemplo de manera entrelazada en un único fichero, o como ficheros independientes.

Una desventaja de la configuración de la Figura 6 es que los micrófonos de oído y posiblemente también el sensor de orientación, están conectados al ordenador 601 por medio de cables relativamente largos, lo que puede dificultar los movimientos de la persona 603.

La unidad de orientación 604 puede estar comprendida en un dispositivo portátil tal como por ejemplo un teléfono inteligente, o un controlador remoto de una consola de videojuegos, que puede comprender un procesador programable configurado con un programa informático para leer datos de orientación procedentes de los uno o más sensores de orientación, y para transmitir esos datos de orientación al ordenador 601, el cual estaría adaptado con un programa informático para recibir dichos datos de orientación. Los datos de orientación se pueden transmitir por ejemplo a través de un cable o de manera inalámbrica (indicado mediante línea de puntos en la Figura 6). En este último caso se puede omitir un cable entre el ordenador 601 y la unidad sensora 604, lo cual es más conveniente para el usuario 603.

En una variante de este método, los datos de orientación se almacenan en una memoria intercambiable, por ejemplo en una tarjeta flash durante la captura de datos, por ejemplo junto con sellos temporales, pudiendo dicha tarjeta flash ser insertada en el ordenador 601 para su procesamiento.

La configuración de la Figura 7 se puede ver como una variante de la configuración de la Figura 6, por la cual la unidad de orientación 704 es parte de un dispositivo portátil, por ejemplo un teléfono inteligente, que tiene un procesador programable y una memoria, y que está equipado además con medios, por ejemplo un dispositivo añadido que se puede enchufar en una interfaz externa, y que tiene uno o dos conectores de entrada para conexión con los micrófonos de oído izquierdo y derecho 705 para capturar muestras de audio que llegan al oído izquierdo y al oído derecho, llamadas muestras de audio izquierda y derecha. Dado que la unidad sensora de orientación 704 está embebida, el procesador puede leer o recuperar datos de orientación del sensor 704, y almacenar en la memoria las muestras de audio izquierda y derecha capturadas, y la información de orientación correspondiente, por ejemplo capturada simultáneamente, como una pluralidad de conjuntos de datos.

Una ventaja adicional de la realización de la Figura 7, es que los cables entre el dispositivo portátil y los micrófonos de oído 705 pueden ser mucho más cortos, lo cual es mucho más cómodo y conveniente para el usuario 703, y permite más libertad de movimiento. Las señales de audio así capturadas contienen también típicamente menos ruido, por tanto la SNR (relación señal ruido) se puede incrementar de esta manera, produciendo como resultado finalmente una mayor precisión de la ITDF y la HRTF estimadas.

Si el segundo paso, a saber, el procesamiento de datos, también es realizado por el dispositivo portátil, por ejemplo el teléfono inteligente, entonces sólo es necesario que se cargue un único producto de programa software en el teléfono inteligente, y no se requiere ningún ordenador externo.

La Figura 8 es una variante de la última realización descrita con relación a la Figura 7, por la cual el segundo paso, a saber, el procesamiento de datos de los datos capturados, es realizado por un ordenador externo 801, pero el primer paso de captura de datos todavía es realizado por el dispositivo portátil. Los datos capturados se pueden transmitir del dispositivo portátil al ordenador, por ejemplo a través de un cable o de forma inalámbrica, o de cualquier otra manera. Por ejemplo, el dispositivo portátil puede almacenar los datos capturados en una tarjeta de memoria no volátil o similar, y el usuario puede extraer la tarjeta de memoria del dispositivo portátil después de que haya finalizado la captura, e insertarla en una ranura correspondiente del ordenador 801. Los últimos dos ejemplos ofrecen ambos la ventaja de que el usuario 803 tiene mucha libertad para moverse, y no está dificultado por cables. La variante inalámbrica tiene la ventaja adicional de que no es necesario intercambiar ninguna tarjeta de memoria. En todas las realizaciones de la Figura 8, se requiere un primer módulo software para que el dispositivo portátil capture los datos, y almacene o transmita los datos capturados, y se requiere un segundo módulo para que el ordenador 801 obtenga, por ejemplo reciba o recupere o lea los datos capturados, y para procesar los datos capturados a fin de estimar una ITDF personalizada y una HRTF personalizada.

Las siguientes acciones A a G son aplicables a todas las disposiciones hardware para capturar los conjuntos de datos que comprenden información de audio izquierda, información de audio derecha, e información de orientación, en particular, pero no limitadas a las disposiciones mostradas en las Figuras 6 a 8, a menos que se indique otra cosa de manera específica.

En estas secciones, se hará referencia a "chirps" como un ejemplo de los estímulos de audio de interés, para mayor facilidad de explicación, pero la invención no está limitada a ellos, y también se pueden utilizar otras señales, por ejemplo tonos puros cortos, como se ha descrito anteriormente.

En estas secciones, se hará referencia a "teléfono inteligente" como un ejemplo de un dispositivo portátil en el cual está embebida la unidad sensora de orientación, pero la invención no está limitada a ello, y en algunas realizaciones (tales como la mostrada en la Figura 6), también puede funcionar una unidad sensora de orientación autónoma 604, mientras que en otras realizaciones (tales como la mostrada en la Figura 8) el dispositivo portátil necesita tener al menos medios de captura de audio y memoria, mientras que en otras realizaciones adicionales (tales como la mostrada en la Figura 7) el dispositivo portátil necesita además tener medios de procesamiento.

A. Captura simultánea de audio y orientación

Es importante que las muestras de audio izquierda y derecha, es decir, los estímulos registrados, y la información de orientación sean correspondientes. Idealmente, las señales de audio izquierdas y derechas son "muestreadas simultáneamente" (dentro del margen de tolerancia de una señal de reloj), pero existe una cierta tolerancia en cuándo

se miden exactamente los datos de orientación. Lo que es importante para la presente invención es que los datos de orientación obtenidos a partir de la unidad de orientación son representativos para la orientación 3D de la unidad de orientación, e indirectamente también para la orientación 3D de la cabeza (si la orientación relativa de la unidad de orientación y de la cabeza fueran conocidas) en aproximadamente el mismo momento en que se capturan las muestras de audio. Como ejemplo, suponiendo que la cabeza se está girando lentamente durante el paso de captura, (por ejemplo a una velocidad angular de menos de 60° por segundo), y que los estímulos acústicos tienen una duración relativamente corta (por ejemplo aproximadamente 25 ms), no tiene realmente importancia si los datos de orientación se recuperan del sensor al comienzo o al final del estímulo acústico, o durante el estímulo, ya que ello daría como resultado un error de orientación angular de menos de $60^\circ/40^\circ$, que es aproximadamente $1,5^\circ$, lo cual es perfectamente aceptable.

B. La configuración hardware

Durante la captura de datos, una distancia entre el altavoz 602, 702, 802 y las personas 603, 703, 803 es preferiblemente una distancia en el rango de 1,0 a 2,0 m, por ejemplo en el rango de 1,3 m a 1,7 m, por ejemplo de aproximadamente 1,5 m, pero no es necesario conocer la distancia exacta. El altavoz se debería situar aproximadamente a aproximadamente la mitad de la altura de la habitación. La cabeza de la persona se debería situar a aproximadamente la misma altura que el altavoz. El altavoz está dirigido hacia la cabeza. Suponiendo una anchura de la cabeza de aproximadamente 20 cm, una fuente situada a una distancia de 1,5 m, cada oído estaría desalineado un ángulo de $\arctan(0,1/1,5) \text{ rad} = 3,8^\circ$.

Suponiendo que la cabeza de la persona está girada al máximo (alrededor de un punto central de la cabeza) y no está desplazada o sólo está mínimamente desplazada, el lóbulo principal es suficientemente ancho para contener la cabeza totalmente a las frecuencias de interés, para que la diferencia de intensidad sea limitada. Pero los métodos de la presente invención también funcionarán muy bien si el centro de la cabeza no se mantiene exactamente en la misma posición, como se explicará más adelante (véase la Figura 27).

En los ejemplos descritos más adelante, se utiliza un único altavoz, pero por supuesto la invención no está limitada a ello, y también se pueden utilizar múltiples altavoces situados en diferentes puntos en el espacio. Por ejemplo, el sistema de reproducción de sonido puede ser un sistema estéreo, que envía estímulos acústicos de forma alternante al altavoz izquierdo y al altavoz derecho.

C. Posible procedimiento para el usuario final

El procedimiento se ejecuta preferiblemente en una habitación (o espacio) relativamente silenciosa. La persona puede estar provista de un CD de audio que contenga una señal de ensayo acústica así como instrucciones escritas o acústicas. El usuario puede realizar uno o más de los siguientes pasos, en el orden mencionado, o en cualquier otro orden:

1. Colocar el altavoz en un borde de una mesa (pero también se podrían utilizar otros lugares apropiados). Configurar el dispositivo de reproducción de sonido (por ejemplo la cadena estéreo) de modo que sólo uno de los altavoces esté produciendo sonido, (o ambos estén produciendo sonido, pero no al mismo tiempo),

2. Escuchar las instrucciones contenidas en el CD de audio, las cuales pueden comprender por ejemplo instrucciones de con qué frecuencia y/o con qué rapidez y/o cuándo tiene que cambiar el usuario la orientación de su cabeza,

3. Introducir el micrófono de oído izquierdo en el oído izquierdo, y el micrófono de oído derecho en el oído derecho, y conectar los micrófonos al teléfono inteligente (en la Figura 6: al ordenador externo 601),

4. Descargar una aplicación software apropiada (denominada típicamente "app") en el teléfono inteligente, y ejecutar la aplicación, (este paso no es aplicable a la Figura 6)

5. Colocar el teléfono inteligente (o sensor en la Figura 6) encima de la cabeza, y fijar su posición utilizando por ejemplo la cinta para la cabeza especialmente diseñada u otros medios de fijación, para permitir que el teléfono inteligente capture y/o envíe y/o registre cualquier orientación y/o movimiento y/o posición de la cabeza. Se observa que el teléfono inteligente se puede montar en cualquier posición arbitraria y en cualquier orientación arbitraria con respecto a la cabeza,

6. Situarse él mismo (por ejemplo sentarse o ponerse de pie) a una distancia de aproximadamente $1,5 \pm 0,5$ m del altavoz. Asegurarse de que la habitación es suficientemente grande, y de que no están presentes paredes u objetos dentro de un radio de aproximadamente 1,5 metros del altavoz y de la persona (para evitar reflexiones),

7. Cuando se escuchan los estímulos acústicos, por ejemplo sonidos de chirps, girar la cabeza lentamente durante un periodo predefinido (por ejemplo de 5 a 15 minutos, por ejemplo aproximadamente 10 minutos) en todas direcciones, por ejemplo de izquierda a derecha, de arriba a abajo, etc.

En algunas realizaciones (véase la Figura 26), se prefiere que la posición de la cabeza (X, Y, Z) permanezca sin cambios, y sólo se cambia la orientación de la cabeza (por ejemplo 3 ángulos de Euler con respecto al marco de

referencia del mundo), véase la Figura 2, para cambiar el ángulo de incidencia del sonido con respecto a la cabeza. Entre las series de estímulos acústicos (por ejemplo chirps), se pueden proporcionar pautas acerca de cómo moverse. Por ejemplo, se puede dar la instrucción en un cierto momento de girar la cabeza un cuarto de vuelta (90°), o media vuelta (180°) de modo que el hemisferio lateral y el sonido procedente de “detrás” del usuario también se muestreen.

5 En otras realizaciones (véase la Figura 27), se permite al usuario sentarse en una silla giratoria, y no es necesario que mantenga el centro de su cabeza en posición fija, sino que se le permite hacer girar libremente la silla y doblar libremente su cuello. Es evidente que estas realizaciones son mucho más convenientes para el usuario.

8. Después de que se haya completado el ensayo, se le pedirá al usuario que se quite el teléfono inteligente de la cabeza y que detenga la captura o el registro mediante la “app”.

10 Se calculan a continuación una ITDF personalizada y una HRTF personalizada, por ejemplo en el propio teléfono inteligente (véase la Figura 7), en cuyo caso no es necesario que se transfieran los datos capturados a otro ordenador, o se calculan en otro ordenador, por ejemplo en la nube, en cuyo caso es necesario que los datos capturados se transfieran de la “app” al ordenador o a la red.

15 La cantidad de datos a transmitir puede ser por ejemplo de aproximadamente 120 MBytes (para un ensayo acústico de aproximadamente 11 minutos). A una velocidad de transmisión inalámbrica de aproximadamente 8 Mbits/s = 1 MByte por segundo, dicha transferencia sólo requiere aproximadamente 2 minutos.

20 Se calculan a continuación la ITDF y la HRTF utilizando un algoritmo particular (como se explicará más adelante), y a continuación quedan disponibles la ITDF y la HRTF resultantes, y están listas para uso personal, por ejemplo en un entorno de juegos en 3D, o en un entorno de teleconferencia, o en cualquier otra aplicación del Sistema de Audio Virtual 3D.

Son posibles muchas variantes del procedimiento descrito aquí anteriormente, por ejemplo:

- la transmisión de los datos capturados puede comenzar ya antes de que se tomen todas las medidas,
- parte de los cálculos pueden comenzar ya antes de que se reciban todos los datos capturados,
- 25 - en vez de solamente capturar los datos, el teléfono inteligente también puede analizar los datos, por ejemplo los datos de orientación, para verificar si se han medido o no todas las direcciones, y podría reproducir por ejemplo un mensaje apropiado en su propio altavoz con instrucciones correspondientes, por ejemplo girar la cabeza en direcciones particulares, etc.

D. La habitación y la señal de ensayo acústica

30 Se pueden utilizar diferentes estímulos de ensayo para la determinación de la ITDF y la HRTF. En una realización, se propone utilizar estímulos de banda ancha (denominados en esta memoria “chirps”), por lo cual la frecuencia varía al menos de 1 kHz a 20 kHz, no estando limitada la invención a ello. Se podría optar por una banda de frecuencia más estrecha, por ejemplo de 4 kHz a 12 kHz, porque en esta parte del espectro de frecuencias audibles, es donde más varía la HRTF (véanse ejemplos en la Figura 4).

35 Tradicionalmente las medidas de HRTF se realizan utilizando señales bastante largas (por ejemplo aproximadamente de 2 a 5 segundos). Tradicionalmente las medidas de HRTF se realizan en una cámara (semi-) anecoica, en la que las paredes están cubiertas con material absorbente del sonido, de modo que las reflexiones secundarias en las paredes y otros objetos se reducen al mínimo. Dado que el método de la presente invención es para ser realizado en casa, estas reflexiones no se pueden eliminar de esta manera. En vez de esto, se utilizan señales de estímulo, por ejemplo chirps, que tienen una duración suficientemente breve para impedir que el sonido directo y el sonido reflejado (contra paredes y/u objetos en la habitación) se solapen (para una habitación típica), o tengan una duración mayor pero una estructura de barrido de frecuencia que permita diferenciar componentes de la señal que vienen a través del “camino directo” de componentes de la señal que vienen a través de caminos indirectos, por ejemplo reflejados.

45 Supóngase en una disposición ejemplar (véase la Figura 21) que el altavoz está a una altura h_e de 1,40 m, y que la cabeza de las personas está a una altura h_x de 1,40 m, y que la distancia L entre la persona y el altavoz es $d = 1,4$ m, y que la altura de la habitación es de al menos 2,8 m, de modo que la reflexión en el suelo llega antes que la reflexión en el techo, entonces la diferencia en distancia recorrida entre el camino directo y la primera reflexión (en el suelo), es:

$$\Delta x = \sqrt{(h_x + h_e)^2 + d^2} - \sqrt{(h_x - h_e)^2 + d^2} = 1.7m$$

50 y de esta manera la señal reflejada necesita $(1,7 \text{ m}) / (344 \text{ m/s}) =$ aproximadamente 4,94 ms más de tiempo para alcanzar la cabeza.

De esta manera tomando una señal de estímulo con una duración menor de 4,94 ms, por ejemplo como máximo 4,80 ms, o como máximo 4,25 ms, o como máximo 4,0 ms, o como máximo 3,5 ms, o como máximo 3,0 ms, o como máximo 2,5 ms, o como máximo 2,0 ms, o como máximo 1,5 ms, o como máximo 1 ms, la señal directa se puede separar fácilmente de las reflexiones posteriores utilizando una máscara de ventana (que es una técnica conocida por sí misma en la técnica).

5 Otra estrategia sería utilizar un barrido de frecuencia. La duración del estímulo puede ser entonces mucho mayor, más de 10 ms, más de 20 ms, más de 30, más de 40, más de 50 ms, más de 60, más de 100, dado que la señal directa y la reflexión se pueden solapar en el dominio del tiempo, porque pueden estar 'separadas' en el dominio frecuencia-tiempo (espectrograma), véanse la Figura 21 y la Figura 22.

10 En lo que sigue, se supondrá una duración del estímulo de 25 ms, aunque la presente invención no está limitada a ello, y también se pueden utilizar otras duraciones de pulso más cortas o más largas que 25 ms, dependiendo de las características de la habitación. También se contempla que pueda estar presente más de una señal de ensayo acústica en el CD de audio, y que el usuario pueda seleccionar la más apropiada, dependiendo de las características de la habitación.

15 Después de cada estímulo, por ejemplo chirp, es necesario esperar un tiempo suficientemente largo para que todas las reflexiones en el entorno (las reverberaciones) se extingan lo suficiente. Esta duración depende de la cámara y de los objetos dentro de ella. El así llamado tiempo de reverberación, se define como el tiempo necesario para garantizar que la intensidad de la señal de eco ha caído 60 dB en comparación con la señal original. A partir de ensayos en diferentes habitaciones, se determina que un tiempo entre pulsos de aproximadamente 300 ms es suficiente, pero la invención no está limitada a esto, y también se pueden utilizar otros tiempos entre pulsos mayores o menores que 300 ms, por ejemplo un tiempo entre pulsos de aproximadamente 100 ms, por ejemplo de aproximadamente 200 ms, por ejemplo de aproximadamente 400 ms, por ejemplo de aproximadamente 500 ms, por ejemplo de aproximadamente 600 ms, por ejemplo de aproximadamente 800 ms, por ejemplo de aproximadamente 1000 ms. Es ventajoso mantener el tiempo entre chirps lo más pequeño posible, para incrementar el número de chirps durante el tiempo total de ensayo (por ejemplo aproximadamente 15 minutos) o, dicho de otra manera, para reducir el tiempo total de ensayo para un número de chirps dado. Si se proporciona un CD o DVD de audio, también puede ser posible proporcionar múltiples señales de ensayo de audio (por ejemplo pistas de audio), con diferente duración de pulso y/o diferentes tiempos entre pulsos y/o diferente duración total del ensayo, y el procedimiento puede incluir un paso de determinar un fichero de ensayo de audio apropiado, por ejemplo que dependa de la habitación en la que se realiza en el ensayo. Una posible implementación en un CD de audio sería que las instrucciones estén presentes en una primera pista de audio, donde se informa al usuario acerca de las diferentes opciones, y mediante lo cual el usuario puede seleccionar una señal de ensayo apropiada, dependiendo de las características de su habitación y/o de la precisión deseada (cuantas menos muestras se tomen, más rápida puede ser la captura y el procesamiento de datos, pero es de esperar que las ITDF y HRTF resultantes sean menos precisas).

30 No es necesario que los estímulos posteriores sean idénticos, sino que pueden variar en contenido de frecuencia y/o duración. Si se eligieran estímulos posteriores de tal manera que cubran una banda de frecuencia diferente, que sea claramente separable, entonces dicho diseño de señal de ensayo permitiría reducir el tiempo entre estímulos, y por lo tanto acortar el tiempo total de obtención de datos.

40 En la realización en la que se utiliza más de un altavoz, por ejemplo dos en caso de una señal estéreo, cada uno de los altavoces se sitúa en un punto diferente en el espacio, y cada uno de los altavoces reproduce una señal de ensayo acústica diferente (utilizando entrada estéreo), que comprende diferentes estímulos (diferente espectro de frecuencia y/o alternando los estímulos (estímulo/no estímulo) entre altavoces), para poder separar los estímulos tras su recepción y para identificar el altavoz en el cual se originaron. Es una ventaja que la presente invención funciona para un gran número de configuraciones de habitación, sin necesidad de sillas especiales o soporte especial para montar el altavoz, etc., sin requerir que se recoloque el altavoz durante la captura de datos, sin conocer la posición exacta del altavoz, y sin conocer la característica de filtro del altavoz.

E. Midiendo la orientación de la cabeza

50 Para determinar la HRTF y la ITDF, es esencial conocer la dirección de la que procede el sonido con respecto a la cabeza o, más exactamente: con respecto al marco de referencia de la cabeza como se muestra en la Figura 2, en la que el centro de la cabeza está situado en el medio entre los dos oídos, un eje coincide con el eje oído-oído, un eje está orientado hacia "la parte delantera" de la cabeza, y un eje está orientado hacia "arriba".

De acuerdo con la presente invención, la dirección de la fuente (del altavoz) con respecto a la cabeza se puede obtener utilizando una unidad de orientación 201 que comprenda uno o más sensores de orientación, por ejemplo un acelerómetro (que mide principalmente una orientación con respecto al eje gravitatorio), un giróscopo (que mide movimientos de rotación), un magnetómetro (que mide un ángulo con respecto al campo magnético de la tierra), pero también se pueden utilizar otras unidades de orientación u otros sensores de orientación. En opinión de los inventores, esta solución no es trivial, porque la unidad de orientación proporciona información de orientación de la unidad de orientación, no de la cabeza. De acuerdo con los principios de la presente invención, la unidad de orientación 201 está montada de forma fija a la cabeza durante el paso de captura de datos, pero no es necesario conocer de antemano el

posicionamiento y/o la orientación exactos de la unidad de orientación 201 con respecto al marco de referencia de la cabeza, aunque si se tiene algún conocimiento previo acerca de su orientación se puede utilizar éste para determinar la dirección de la fuente con respecto a la cabeza. Es una ventaja de las realizaciones de la presente invención que el método presentado es capaz de determinar la dirección de la fuente sin que el usuario tenga que realizar una medida física, o un ensayo de orientación específico o similar.

Es una ventaja de la presente invención que una potencial inexactitud de la unidad del sensor de orientación se pueda intentar solucionar no sólo confiando en la información de orientación obtenida del sensor de orientación, sino también teniendo en cuenta las señales de audio cuando se determina la orientación de la cabeza, como se explicará con mayor detalle más adelante, cuando se describa el algoritmo.

Es una ventaja que los movimientos de la cabeza sean realizados por la propia persona, de una manera que es mucho más libre y conveniente que en la técnica anterior mostrada en la Figura 5. Además, en algunas realizaciones de la invención, la persona no está obstaculizada por cables que van de los micrófonos de oído al ordenador externo.

Una diferencia importante entre la presente invención y la solicitud de patente en tramitación con la presente PCT/EP2016/05320 de los mismos inventores es que, en la primera solicitud de patente, los inventores eran de la opinión de que la unidad de orientación no era suficientemente precisa para proporcionar datos de orientación fiables. Es cierto que los datos de orientación momentáneos proporcionados por los sensores de orientación previstos son a veces imprecisos en el sentido de que se produce histéresis o "hipos", y de que la detección del campo magnético no es igual de sensible en todas las orientaciones y en todos los entornos. Una idea subyacente de la solicitud de patente previa era que los indicios espaciales extraídos de los datos de audio capturados podrían ayudar a mejorar la precisión de los datos de orientación, pudiendo ser extraídos estos indicios espaciales utilizando una función ITDF y/o HRTF "general", lo cual a su vez era una razón para iterar el algoritmo una vez se encontraba una "primera versión" de la ITDF personalizada y de la HRTF personalizada, debido a que los cálculos se podían repetir a continuación utilizando la ITDF personalizada y/o la HRTF personalizada produciendo resultados más precisos.

La presente invención confía parcialmente en dos conocimientos:

1) que el uso de indicios espaciales para mejorar la precisión de, o para corregir, los datos de orientación en bruto obtenidos de la unidad de orientación no es necesario y, por lo tanto, tampoco es necesario el uso de una ITDF predefinida (por ejemplo una ITDF general) y/o de una HTRF predefinida (por ejemplo una HRTF general) para extraer esos indicios espaciales; y

2) que la estimación conjunta de la dirección de la fuente (c. r. al mundo) y la transformación que mapea el marco de referencia del teléfono inteligente al marco de referencia de la cabeza se puede dividir en dos problemas de estimación más sencillos realizados de manera consecutiva. Esto permite una reformulación del problema de búsqueda pasando de uno realizado en un espacio de búsqueda de 5 dimensiones (2 ángulos para especificar la dirección de la fuente + 3 ángulos para especificar la transformación teléfono inteligente-cabeza) a dos problemas más simples, resolviendo en primer lugar un problema en un espacio de búsqueda de 2 dimensiones (2 ángulos para especificar la dirección de la fuente) y utilizando esos resultados para resolver posteriormente un problema en un espacio de búsqueda de 3 dimensiones (3 ángulos para especificar la transformación teléfono inteligente-cabeza). Esta técnica resulta posible por el hecho de que la ITD y/o la información espectral medidas/calculadas cuando se asignan a una dirección incorrecta de la fuente, dan lugar a una "imagen" completamente distorsionada de la ITDF y de la HRTF cuando se mapean sobre la esfera, con muchos componentes de orden elevado, con poca probabilidad los dibujos relativamente continuos o relativamente suaves mostrados en la Figura 3 y en la Figura 4. La presente invención saca provecho de ese conocimiento, utilizando la "suavidad" de la ITDF y/o de la HRTF mapeadas como criterio de calidad para encontrar en primer lugar la dirección de la fuente con respecto al mundo. Los detalles exactos del algoritmo se describirán más adelante, pero el uso de un criterio de calidad de este tipo es una de las ideas subyacentes de la presente invención. Enunciado en términos simples, esto se reduce a encontrar la dirección de la fuente para la cual la ITDF y/o la HRTF mapeadas sobre una esfera "parecen más suaves" que para todas las otras posibles direcciones de la fuente. Se observa que también se podrían utilizar otros criterios de calidad basados en otras propiedades específicas de la ITDF y/o de la HRTF, por ejemplo simetría (excepto por el signo) de la ITDF con respecto a un plano sagital, simetría cilíndrica de la ITDF alrededor del eje oído-oído. Dada la dirección de la fuente (c. r. al mundo), encontrar la transformación teléfono inteligente-cabeza se reduce entonces a un problema de búsqueda en un espacio de búsqueda tridimensional. Esta búsqueda tridimensional se puede subdividir aún más determinando en primer lugar el eje oído-oído (c. r. al teléfono inteligente) y determinando finalmente el ángulo de rotación alrededor del eje oído-oído.

Una ventaja importante de este conocimiento, a saber, que se puede utilizar la "suavidad de la ITDF mapeada y/o de la HRTF mapeada" como criterio de calidad para encontrar la dirección (más probable) de la fuente, es un conocimiento importante, entre otras cosas porque (1) permite que la ITDF y la HRTF de una persona particular se puedan determinar sin utilizar la ITDF y la HRTF de otra persona (o una ITDF general y/o una HRTF general), y (2) porque ofrece enormes ventajas en términos de complejidad computacional y tiempo de cálculo. Para dar una idea, utilizando un método de acuerdo con la presente invención, los cálculos necesarios para determinar la ITDF y la HRTF en un ordenador portátil estándar con por ejemplo un procesador de 2,6 GHz (año 2016) utilizando código no óptimo, sólo tardan aproximadamente 15 minutos, incluso sin intentos de optimizar el código.

Se contempla que se pueden encontrar varias maneras de cuantificar la “suavidad” de los datos de ITDF y/o HRTF mapeados o dibujados o reproducidos sobre la esfera, dos de las cuales se describirán en esta memoria con referencia a la Figura 31. En una realización, expandimos los datos de HRTF medidos en armónicos esféricos (SH) reales, los cuales son funciones básicas similares a funciones básicas de Fourier, pero definidas sobre la esfera. De manera similar a las funciones básicas de Fourier, las funciones básicas SH reales $Y_{lm}(\theta, \varphi)$ tienen la propiedad de que valores—
 5 l menores corresponden a funciones básicas que varían más lentamente, véase la Figura 26(a). Por tanto, esto significa que si la HRTF se expresa en una base truncada que contiene sólo funciones básicas hasta un orden máximo elegido o predefinido L ($L < \infty$), se aplica de forma efectiva un filtro paso bajo que sólo permite variaciones espaciales lentas. Cuanto mayor es el valor L elegido, más “detalle” espacial incluye la expansión de base. Por tanto, para
 10 cuantificar la “suavidad”, estimamos primero los coeficientes $C_{l,m}^{r,L/R}(f)$ y $C_{l,m}^{r,L}(f)$ de la expansión HRTF (que corresponden respectivamente a la HRTF del oído izquierdo y a la HRTF del oído derecho a la frecuencia f para la dirección r elegida) en la base SH truncada en alguna L elegida. A continuación, calculamos la diferencia al cuadrado entre nuestros puntos de datos medidos y la expansión HRTF obtenida (en la cual se calcula una suma para todas las direcciones medidas y para todas las frecuencias medidas)

$$S_{L/R}^r(f, \mathbf{r}_i) \approx \sum_{l=0}^L \sum_{m=-l}^l C_{l,m}^{r,L/R}(f) Y_{lm}(\mathbf{r}_i)$$

$$\varepsilon_{HRTF}^2(\mathbf{r}) = \sum_f \sum_{\mathbf{r}_i} \left\{ \left[S_L^r(f, \mathbf{r}_i) - \sum_{l=0}^L \sum_{m=-l}^l C_{l,m}^{r,L}(f) Y_{lm}(\mathbf{r}_i) \right]^2 + \left[S_R^r(f, \mathbf{r}_i) - \sum_{l=0}^L \sum_{m=-l}^l C_{l,m}^{r,R}(f) Y_{lm}(\mathbf{r}_i) \right]^2 \right\}$$

Este error cuantifica ahora hasta qué punto la base de funciones base que varían lentamente es adecuada para describir el patrón espacial presente en la HRTF medida en toda la esfera. Cuanto menor sea el error, mejor se aproximarán los datos acústicos utilizando sólo funciones base que varíen lentamente y, por consiguiente, más suave será el patrón HRTF. Por consiguiente, se puede utilizar este error como criterio de calidad.

También se pueden definir otros criterios de calidad. Por ejemplo también se elegiría el siguiente:

$$\varepsilon_{HRTF}^2(\mathbf{r}) = \sum_f \left[\left(C_{L,0}^{r,L}(f) \right)^2 + \left(C_{L,0}^{r,R}(f) \right)^2 \right]$$

o

$$\varepsilon_{HRTF}^2(\mathbf{r}) = \sum_f \sum_{\mathbf{r}_i} \{ [\nabla^2 S_L^r(f, \mathbf{r}_i)]^2 + [\nabla^2 S_R^r(f, \mathbf{r}_i)]^2 \}$$

También se pueden utilizar otras normas además de la norma Euclídea como por ejemplo una norma-p general o una norma de valor absoluto.

F. Hardware

Volviendo a hacer referencia a las Figuras 6 a 8. Aunque no todos los teléfonos inteligentes permiten la captura o el registro de señales de audio estéreo por medio de un conector de entrada estéreo o dos conectores de entrada mono, existen extensiones que permiten el registro estéreo por medio de un puerto USB, por ejemplo “TASCAM iM2 Channel Portable Digital Recorder”, disponible comercialmente. Aunque esta extensión tiene micrófonos que no se pueden insertar en un oído, este ejemplo demuestra que la tecnología está disponible para hacer una extensión dedicada de este tipo, por ejemplo eliminando los micrófonos y proporcionando dos conectores de audio, en los cuales se pueden enchufar los micrófonos de oído. Este es sólo un ejemplo de un posible dispositivo portátil que se puede utilizar en las realizaciones de la Figura 7 y de la Figura 8.

También está disponible tecnología para determinar información de orientación de un dispositivo portátil. Considérese por ejemplo la "Sensor Fusion App". Esta aplicación muestra que está disponible tecnología para recuperar información de orientación de dispositivos portátiles con sensores de orientación embebidos, tales como por ejemplo acelerómetros (para medir principalmente una orientación con respecto al eje gravitatorio), un giróscopo (para medir movimientos de rotación) y/o un magnetómetro (para medir dirección con respecto al campo magnético de la tierra).

G. Proporcionando los datos capturados a los medios informáticos

Después de capturar y/o registrar y/o enviar las señales de audio izquierdas y derechas procedentes de los micrófonos (denominados también los datos de audio binaurales), y las correspondientes orientaciones de la cabeza (obtenidas de la unidad de orientación, aunque la relación exacta entre la unidad de orientación y la cabeza aún no se conoce), el procesamiento de los datos capturados puede ser realizado por un procesador existente en el propio dispositivo portátil (por ejemplo en el teléfono inteligente), o en un ordenador remoto (por ejemplo en la nube, o en un ordenador de sobremesa o un ordenador portátil o una consola de videojuegos) al cual los datos se transmiten o se envían o se proporcionan de cualquier otra manera (por ejemplo por medio de una tarjeta de memoria intercambiable).

III. Procesamiento de datos:

El paso de procesamiento de datos de la presente invención se explicará con mayor detalle con referencia a las Figuras 9 a 16.

La Figura 9 es un diagrama esquemático que ilustra las incógnitas que tienen que ser estimadas. Dicho de otra manera, esta Figura ilustra el problema que debe ser resuelto por la parte de procesamiento de datos del algoritmo utilizado en las realizaciones de la presente invención.

Como se puede ver a partir de la Figura 9, la ITDF personal (o individualizada) y la HRTF personal (o individualizada) no son los únicos conjuntos de variables a determinar. La orientación de la cabeza durante la obtención de datos es desconocida en las configuraciones mostradas en las Figuras 6 a 8, porque, aunque la orientación de la propia unidad de orientación 201 se determina (en base principalmente a los sensores de orientación) la orientación de la unidad de orientación 201 con respecto al marco de referencia de la cabeza no se conoce con precisión, y porque la orientación de la cabeza en el instante de recepción de cada estímulo acústico (por ejemplo en cada chirp) posiblemente no se conoce con precisión, en base a la información del sensor individual recuperada u obtenida durante cada chirp individual particular, y se considera por lo tanto desconocida. Asimismo, la dirección de la fuente de sonido (con respecto al marco de referencia de la cabeza) es desconocida. Además, la característica espectral de la combinación de altavoz y micrófono puede ser desconocida, ya que el usuario puede utilizar cualquier altavoz disponible. La característica de transferencia de los micrófonos de oído puede ser conocida de antemano, especialmente cuando los micrófonos de oído se venden por ejemplo en un pack junto con un CD, pero incluso entonces, los parámetros de los altavoces no son conocidos. En casos en que las características de transferencia del altavoz y de los micrófonos se conocen, el algoritmo puede utilizarlas, pero eso no es absolutamente necesario.

Se encontró que este gran número de incógnitas no se pueden estimar con suficiente precisión a menos que todos los datos se combinen y se estimen en conjunto (en el sentido de: "unos en dependencia de otros"). Este es otro aspecto ventajoso de la presente invención. Por ejemplo, los datos de orientación y de movimiento en bruto individuales que se originan en los uno o más sensores de orientación (por ejemplo, embebidos en un teléfono inteligente) podrían no permitir determinar la orientación del teléfono inteligente individual y por lo tanto la orientación de la cabeza con suficiente precisión, entre otros motivos porque la posición/orientación del teléfono inteligente con respecto a la cabeza no es totalmente conocida, y además, porque puede ser bastante difícil estimar con precisión la orientación de la cabeza, dada la precisión limitada de las medidas individuales del sensor de orientación.

Diferencia principal:

Allí donde los inventores propusieron en "la solicitud de patente previa" extraer opcionalmente información de orientación contenida en los datos de audio izquierdo y derechos, en la presente invención no se confía en este principio, al menos para determinar una primera versión de la IDTF personalizada y de la HRTF personalizada, aunque estos datos todavía se podrían tener en cuenta en una segunda iteración o en iteraciones adicionales de ciertos pasos del algoritmo. En vez de esto, el rasgo clave en el que se basa la presente invención es que la dirección del altavoz (con respecto al mundo) se puede encontrar maximizando un valor de calidad predefinido, relacionado preferiblemente con una "métrica de suavidad".

Y, opcionalmente, si la precisión de la información de orientación obtenida de la unidad de orientación es insuficiente, la precisión y/o la fiabilidad de los datos de orientación se puede mejorar aún más confiando en movimientos lentos de la cabeza. Esto permite por ejemplo generar o corregir información de orientación mediante interpolación entre dos orientaciones correspondientes a chirps que no son "chirps adyacentes", sino por ejemplo que están separados por 2 o 3 duraciones de chirp, por tanto datos de orientación en bruto incorrectos debidos por ejemplo a "hipos" o debidos a histéresis, o debidos a baja sensibilidad de la unidad de orientación en direcciones particulares, se pueden mejorar.

En general, se cree que las ventajas más importantes de la presente invención son las siguientes:

- el método puede ser aplicado en casa por casi cualquier usuario (no se requiere ninguna habitación especial, ni se requiere ninguna habilidad especial);

5 - el usuario no requiere equipos especiales aparte de un par de micrófonos de oído y un fichero de ensayo de audio y una cinta para conectar un teléfono inteligente a la cabeza (se supone que casi todos los usuarios tienen un teléfono inteligente y/o un ordenador portátil);

- el método es muy robusto (no es necesario que se conozcan o que se midan la ubicación relativa del altavoz con respecto a la cabeza, y la orientación relativa del teléfono inteligente con respecto a la cabeza);

- el usuario se puede mover casi libremente, y no tiene que seguir patrones específicos (pero el espacio se debería muestrear suficientemente);

10 - (por último, pero no menos importante) una reducción de la complejidad computacional.

Las incógnitas mostradas en la Figura 9 se pueden optimizar de manera iterativa, de tal manera que la solución obtenida de esta manera corresponde al máximo con los conjuntos de datos capturados. Esto se explicará con mayor detalle cuando se analice la Figura 11.

15 En caso de múltiples altavoces, por ejemplo dos en el caso de una señal estéreo (o dos señales mono no solapadas sincronizadas), los estímulos registrados se pueden identificar como originarios de uno de los altavoces gracias a la elección de la señal de ensayo acústica aplicada, y por tanto se obtienen dos conjuntos de datos independientes, cada uno de ellos correspondiente a uno de los altavoces. Estos conjuntos de datos se pueden utilizar a continuación en conjunto como entrada para el algoritmo para estimar correctamente la dirección del altavoz, y las otras incógnitas del problema mostradas en la Figura 9. El hecho de que uno tiene dos "puntos de referencia" que no cambian de posición, puede mejorar las estimaciones de la orientación de la cabeza, y por consiguiente las estimaciones de la ITDF y de la HRTF.

El algoritmo (alto nivel):

La Figura 10 muestra los dos primeros pasos del algoritmo propuesto por la presente invención.

25 En un primer paso 1011, también denominado en lo que sigue "paso a", se obtiene una pluralidad de conjuntos de datos, comprendiendo cada conjunto de datos una muestra de audio izquierda y una muestra de audio derecha, y correspondientes datos de orientación.

Con "fragmentos de audio izquierdos" y "fragmentos de audio derechos" se hace referencia a una parte de la forma de onda de audio recibida por el micrófono de oído izquierdo y respectivamente por el derecho, que corresponde a un estímulo acústico particular enviado por el altavoz, por ejemplo "un chirp".

30 Se observa que los conjuntos de datos pueden ser "obtenidos" y/o "capturados" y/o "almacenados" en memoria de muchas maneras diferentes, por ejemplo como un único fichero o flujo entrelazado, o como tres ficheros o flujos independientes (por ejemplo uno primero que contiene las muestras de audio izquierdas, uno segundo que contiene las muestras de audio derechas, y uno tercero que contiene los datos de orientación, de modo que cada fichero puede comprender información de sincronización, por ejemplo en forma de sellos temporales), o como paquetes de datos individuales, conteniendo cada paquete de datos una muestra de audio izquierda, y una muestra de audio derecha y datos de orientación con respecto a un sistema de referencia fijado al mundo, pero también pueden ser posibles otras maneras, y la presente invención no está limitada a ninguna de estas maneras.

40 Dependiendo de qué dispositivo hardware realiza la captura de los datos, y de qué dispositivo hardware realiza los cálculos, (por ejemplo, un ordenador independiente, o un ordenador en red, o un teléfono inteligente, o cualquier otro medio informático), "obtener" puede significar: "recibir" datos capturados por otro dispositivo (por ejemplo, por un teléfono inteligente, véase por ejemplo la Figura 8), por ejemplo por medio de una interfaz por cables o inalámbrica, o "recuperar" o "leer" datos de una tarjeta de memoria intercambiable (en la cual los datos fueron almacenados por el dispositivo de captura, y que a continuación es conectada al dispositivo informático), o una transferencia de datos de cualquier otra manera. Pero si el dispositivo que captura los datos es el mismo que el dispositivo que realiza los cálculos, "obtener" puede significar "capturar los conjuntos de datos", ya sea directamente, o indirectamente, y no es necesaria ninguna transmisión de los datos capturados a otro dispositivo. Por lo tanto, es evidente que no es necesario que un método o un producto de programa informático dirigido al procesamiento de los datos, capture también necesariamente los datos.

50 En un segundo paso 1012, también denominado en esta memoria "paso b", los conjuntos de datos se almacenan en una memoria. La memoria puede ser una memoria no volátil o una memoria volátil, por ejemplo RAM o FLAH o una tarjeta de memoria, etc. Típicamente todos los conjuntos de datos se almacenarán en una memoria, por ejemplo en RAM. Se contempla que de 100 MBytes a 150 MBytes, por ejemplo aproximadamente 120 MBytes de memoria, son suficientes para almacenar los datos capturados.

Para mayor facilidad de descripción, se supone que la unidad de orientación está presente en el teléfono inteligente, y que sólo existe un altavoz, pero la invención no está limitada a ello, y también se pueden utilizar otras unidades de orientación y más de un altavoz.

5 La Figura 10 es una representación en forma de diagrama de flujo de una primera realización de un método 1000 de acuerdo con la presente invención. Con fines ilustrativos, para no sobrecargar la Figura 10 y la Figura 11 con una gran cantidad de flechas, se debería interpretar este diagrama de flujo como una secuencia de pasos 1001 a 1005, siendo el paso 1004 opcional, con iteraciones o repeticiones opcionales (flecha hacia arriba situada a la derecha del dibujo), pero, aunque no se muestra de forma explícita, los datos proporcionados a un paso “anterior” también están disponibles para un paso posterior. Por ejemplo los datos del sensor de orientación se muestran como entrada al bloque 1001, pero están también disponibles para el bloque 1002, 1003, etc. De manera similar, la salida del bloque 1001 está disponible no sólo para el bloque 1002, sino también para el bloque 1003, etc.

10 En el paso 1001 se estima para cada fragmento de audio la orientación del teléfono inteligente con respecto al mundo (expresada por ejemplo en 3 ángulos de Euler). Un ejemplo de este paso se muestra con mayor detalle en la Figura 13. Opcionalmente este paso puede tener en cuenta datos de audio binaurales para mejorar la estimación de orientación, pero eso no es absolutamente necesario. Expresado en términos sencillos, el objetivo principal de este paso es determinar la orientación desconocida del teléfono inteligente para cada fragmento de audio.

15 A continuación, en el paso 1002, se determina la “dirección de la fuente” con respecto al mundo, excluyendo el signo (o “sentido” explicado anteriormente). Un ejemplo de este paso se muestra con mayor detalle en la Figura 14. Expresado en términos sencillos, el objetivo principal de este paso es determinar la dirección desconocida del altavoz para cada fragmento de audio (en coordenadas del mundo).

20 A continuación, en el paso 1003, se determina la “orientación del teléfono inteligente con respecto al marco de referencia de la cabeza” (véase la Figura 2) y el signo (o “sentido” explicado anteriormente) de la “dirección de la fuente” con respecto al mundo. Un ejemplo de este paso se muestra con mayor detalle en la Figura 15. Expresado en términos sencillos, el objetivo principal de este paso es determinar la orientación desconocida del teléfono inteligente hacia la cabeza.

A continuación, opcionalmente, en el paso 1004, se puede estimar el centro de la posición de la cabeza con respecto al mundo. Si se supone que el centro de la cabeza no se mueve durante la medida, se puede saltar el paso 1004, de lo cual se describirán dos ejemplos en la Figura 26 y en la Figura 27. En caso del “modelo simple” de la Figura 26,.

25 A continuación, en el paso 1005 se estiman una ITDF personalizada y una HRTF personalizada. Expresado en términos sencillos, el principal objetivo de este paso es proporcionar una función ITDF y una función HRTF capaces de proporcionar un valor para cada dirección de la fuente con respecto a la cabeza, también para direcciones de la fuente medidas de forma no explícita durante el ensayo.

Un ejemplo de esta realización 1000 se describirá en el Apéndice.

30 Los inventores son de la opinión de que tanto la secuencia particular de pasos (para obtener la dirección del sonido con respecto a la cabeza sin imponerla o medirla realmente pero utilizando en cambio un teléfono inteligente que además se puede orientar en cualquier orientación arbitraria), como la solución específica propuesta para el paso 1002 no son triviales.

35 La Figura 11 es una variante de la Figura 10 y muestra una segunda realización de un método 1100 de acuerdo con la presente invención. La principal diferencia entre el método 1100 de la Figura 11 y el método 1000 de la Figura 10 es que el paso 1102 también puede tener en cuenta información a priori de la posición/orientación del teléfono inteligente, si esa es conocida. Esto puede permitir estimar el signo de la fuente ya en el paso 1102.

Todo lo demás que se mencionó en la Figura 10 también es aplicable aquí.

40 La Figura 12 muestra un método 1200 (es decir, una combinación de pasos) que se puede utilizar para estimar orientaciones de teléfonos inteligentes con respecto al mundo, en base a datos de sensor de orientación y datos de audio binaurales, como el que se puede utilizar en el paso 1001 del método de la Figura 10, y/o en el paso 1101 del método de la Figura 11.

En el paso 1201 se obtienen o se leen o se obtienen de otra manera datos de sensor a partir de uno o más sensores de la unidad de orientación, por ejemplo datos procedentes de un magnetómetro y/o datos procedentes de un acelerómetro y/o datos procedentes de un giróscopo, y preferiblemente todos estos.

45 En el paso 1202 se determina una trayectoria de la orientación del teléfono inteligente a lo largo de un intervalo de tiempo dado, por ejemplo maximizando la consistencia interna entre datos del magnetómetro, datos del acelerómetro y datos del giróscopo.

En el paso 1203 se determina el tiempo de llegada de los fragmentos de audio (por ejemplo chirps) a cada uno de los oídos, por ejemplo se extraen de los datos de audio binaurales.

En el paso 1204 se estima la orientación del teléfono inteligente (c. r. al mundo) en un momento igual al tiempo de llegada medio de chirps correspondientes a ambos oídos.

5 La Figura 13 muestra un método ejemplar 1300 para estimar la dirección de la fuente con respecto al mundo, como el que se puede utilizar en el paso 1002 y/o en el paso 1102 de la Figura 10 y de la Figura 11. O, más específicamente, lo que se estima es la dirección de una línea virtual que pasa a través del altavoz y a través de una "posición media" del centro de la cabeza a lo largo de todas las medidas, pero sin un "signo" que apunte a ambos extremos de la línea. Dicho de otra manera, un vector situado sobre esta línea virtual, apuntaría desde la posición del centro medio de la cabeza al altavoz, o en la dirección opuesta.

10 En el paso 1301 se extrae información de ITD de los datos de audio binaurales, por ejemplo calculando una diferencia de tiempo entre el momento de llegada de los fragmentos de audio (correspondiente a los chirps emitidos por el altavoz) en el oído izquierdo y en el oído derecho. Los datos de ITD se pueden representar como una matriz de valores $ITD[i]$, para $i=1$ a m , donde m es el número de chirps. m también es igual al número de fragmentos de audio capturados por cada oído. En el paso 1301 también se extraen datos espectrales binaurales de las muestras de audio izquierda y derecha. Los datos espectrales $S_i(f)$, para $i=1$ a m , se pueden almacenar por ejemplo como una matriz bidimensional de datos, véanse por ejemplo la Figura 23(a) y la Figura 23(b) y la Figura 24(a) y la Figura 24(b), las cuales son representaciones gráficas de estos datos.

15 Los pasos 1302, 1303, 1304, 1305 y 1306 forman un bucle que se ejecuta múltiples veces, cada vez para una "dirección candidata de la fuente" diferente. En cada iteración del bucle, la "dirección candidata de la fuente" se utiliza para mapear los valores de los datos de ITD (para todos los chirps o para un subconjunto de ellos) a una superficie esférica, y/o para mapear los valores espectrales de una o más frecuencias particulares a uno o más otras superficies esféricas. Y para cada uno de estos mapeados, por lo tanto para cada "dirección candidata de la fuente", se calcula un valor de calidad, en base a un criterio de calidad predefinido.

En las realizaciones preferidas, el criterio de calidad está relacionado con, o es indicativo de, una suavidad de los datos mapeados. Este aspecto se describirá con mayor detalle cuando se analice la Figura 26.

25 El bucle se repite varias veces, y la "dirección candidata de la fuente" para la cual se obtuvo el máximo valor de calidad, se selecciona en el paso 1307 como "la dirección de la fuente". Experimentos han mostrado que la dirección de la fuente así encontrada corresponde con la dirección verdadera de la fuente. Por lo que saben los inventores, esta técnica para encontrar la dirección de la fuente no es conocida en la técnica anterior, aunque ofrece varias ventajas importantes, tales como por ejemplo: (1) que no es necesario que la dirección de la fuente se conozca de antemano, (2) que la dirección de la fuente se puede determinar de forma relativamente precisa sobre la base de los datos capturados, y (3) que la dirección de la fuente se puede encontrar relativamente rápido, especialmente si se utiliza una estrategia de búsqueda inteligente.

Se podría utilizar la siguiente estrategia de búsqueda, pero la invención no está limitada a esta estrategia de búsqueda particular, y también se pueden utilizar otras estrategias de búsqueda:

35 a) en una primera serie de iteraciones, se determina el factor de calidad para un conjunto predefinido de por ejemplo 8 a 100, por ejemplo, aproximadamente 32 direcciones candidatas de la fuente, para obtener una idea aproximada de encontrar un buen punto de comienzo cerca de la mejor candidata. Se calcula el factor de calidad para este número predefinido de candidatas, y se elige la ubicación que proporciona el mayor factor de calidad como punto de inicio para una segunda serie de iteraciones.

40 b) en una segunda serie de iteraciones, se ajusta la dirección candidata de la fuente en pequeños pasos, por ejemplo ensayando ocho direcciones cercanas, que tengan un ángulo de elevación ligeramente diferente (por ejemplo, un ángulo de elevación actual de -5° , $+0^\circ$, $+5^\circ$) y/o un ángulo lateral ligeramente diferente (por ejemplo, un ángulo lateral actual de -5° , $+0^\circ$, $+5^\circ$), dando como resultado ocho nuevas candidatas, que son evaluadas, y se elige la mejor candidata.

45 c) repetir el paso b) hasta que el factor de calidad ya no aumenta,

d) repetir el paso b) con un tamaño de paso menor, por ejemplo (-1° , $+0^\circ$, $+1^\circ$) hasta que el factor de calidad ya no aumenta.

Los ensayos han mostrado que la convergencia puede ser relativamente rápida, por ejemplo puede requerir menos de 1 minuto en un ordenador portátil estándar de frecuencia de reloj de aproximadamente 2,6 GHz.

50 La Figura 14 muestra un método 1400 para determinar la orientación del teléfono inteligente con respecto al marco de referencia de la cabeza, como el que puede ser utilizado en el bloque 1003 de la Figura 10 y en el bloque 1103 de la Figura 11, pero la invención no está limitada a ello, y también se pueden utilizar otros métodos.

El paso 1401 es idéntico al paso 1301, pero se muestra con fines ilustrativos. Por supuesto, dado que el paso 1301 ya se ha ejecutado antes, no es necesario que se ejecute de nuevo, pero los datos se pueden reutilizar.

En el paso 1402 se estima la orientación del eje oído-oído con respecto al marco de referencia del teléfono inteligente, en base a la orientación del teléfono inteligente (c. r. al mundo) y la dirección de la fuente hasta el signo (c. r. al mundo) y la ITD y/o información espectral. En la realización descrita en el Apéndice, sólo se utilizaron datos ITD, pero la invención no está limitada a ello.

5 La orientación del eje oído-oído (con respecto al teléfono inteligente) se puede utilizar a continuación en el paso 1403, junto con información espectral monoaural o binaural, complementada con las orientaciones del teléfono inteligente con respecto al mundo, y la dirección de la fuente excepto su signo con respecto al mundo, para estimar la dirección frontal de la cabeza con respecto al marco de referencia del teléfono inteligente, produciendo como resultado la orientación del teléfono inteligente con respecto a la cabeza, y el "signo" de la dirección de la fuente con respecto al mundo.

10 La Figura 15 muestra un método 1500 para determinar la posición del centro de la cabeza con respecto al mundo, como el que se puede utilizar en el bloque opcional 1004 de la Figura 10 y en el bloque 1104 de la Figura 11, pero la invención no está limitada a ello, y también se pueden utilizar otros métodos.

15 En el paso 1501, se extraen los tiempos de llegada de los fragmentos de audio izquierdos y derechos correspondientes.

En el paso 1502 se utilizan estos tiempos de llegada para estimar una variación de distancia entre el centro de la cabeza y la fuente.

20 En el paso 1503 se puede utilizar esta variación de distancia para estimar parámetros del modelo de un movimiento de la cabeza/silla, por ejemplo los parámetros del modelo mostrado en la Figura 31, si se utiliza éste. Como se ha mencionado anteriormente, este modelo es opcional, pero cuando se utiliza, puede proporcionar datos más precisos.

En el paso 1504, se pueden estimar a continuación las posiciones del centro de la cabeza, en base a los parámetros del modelo mecánico, complementados con las orientaciones de la cabeza y la dirección de la fuente con respecto al mundo.

25 La Figura 16 muestra un método 1600 para determinar la HRTF y/o la ITDF, como el que se puede utilizar en el bloque 1005 de la Figura 10 y en el bloque 1105 de la Figura 11, pero la invención no está limitada a ello, y también se pueden utilizar otros métodos.

En el paso 1601, se estiman las direcciones de la fuente con respecto a la cabeza, en base a la dirección de la fuente y a las orientaciones de la cabeza en el mundo, complementadas, si están disponibles, con las posiciones de la cabeza e información a priori sobre la distancia hasta la fuente.

30 El paso 1602 es idéntico al paso 1301, pero se muestra con fines ilustrativos. Por supuesto, dado que el paso 1301 ya se ha ejecutado antes, no es necesario que se ejecute de nuevo, pero los datos se pueden reutilizar.

En el paso 1603 se estiman la ITDF y la HRTF ajustando por mínimos cuadrados los coeficientes armónicos esféricos de una base truncada a respectivamente los datos de ITD y los datos espectrales (en una base por frecuencia) proyectados sobre la esfera, de acuerdo con las direcciones del sonido con respecto a la cabeza.

35 La Figura 18 muestra un diagrama de flujo de funcionalidad adicional opcional como el que se puede utilizar en las realizaciones de la presente invención.

40 En la configuración más simple, un fichero de sonido que contiene la señal de ensayo acústica (una serie de estímulos acústicos, por ejemplo chirps) se reproduce en un altavoz, y los datos son recogidos por el teléfono inteligente. Puede ser beneficioso incluir instrucciones verbales para el sujeto, para guiarle a través del experimento mejorando por tanto la recogida de datos. Estas instrucciones pueden ser fijadas, por ejemplo predeterminadas, como parte del fichero de sonido grabado previamente que se reproducirá a través del altavoz, u otra posibilidad puede ser procesar la recogida de datos hasta cierto punto en tiempo real en el dispositivo informático, por ejemplo el teléfono inteligente, y proporcionar realimentación inmediata o intermedia al usuario, por ejemplo para mejorar la obtención de datos. Esto se podría conseguir mediante el proceso esbozado en la Figura 18, que comprende los siguientes pasos.

45 En un primer paso 1801, el teléfono inteligente captura, almacena y recupera los datos del sensor de orientación y los datos de audio binaurales.

50 En un segundo paso 1802, los datos medidos se procesan (al menos parcialmente) en tiempo real en el teléfono inteligente. Para la pluralidad de conjuntos de datos se puede extraer información de temporización y/o información espectral de las muestras de audio izquierda y derecha. En base a esta información, se puede evaluar la calidad de la señal y la configuración experimental (por ejemplo, relación Señal Ruido de las señales recibidas, solape con ecos, etc.). También se puede extraer información de orientación (precisa o aproximada) para el subconjunto de muestras capturadas, mediante la cual el algoritmo verifica además si el espacio alrededor de la cabeza se muestrea con suficiente densidad. En base a esta información, se pueden identificar problemas, e instrucciones (por ejemplo instrucciones verbales) para mejorar la recogida de datos pueden ser seleccionadas por el algoritmo de un grupo de

mensajes de audio predefinidos, por ejemplo asegúrese de que el techo es suficientemente alto, asegúrese de que no hay objetos reflectantes dentro de un radio de 1,5 m, suba/baje el volumen del altavoz, utilice un altavoz diferente, mueva la cabeza más lentamente, gire un cuarto a la izquierda y mueva la cabeza de izquierda a derecha, etc.

5 En un tercer paso 1803, estas instrucciones se comunican en tiempo real a través de los altavoces del teléfono inteligente.

En un cuarto paso 1804, la persona reacciona a estas instrucciones, cuyas acciones se reflejan en los posteriores registros de los datos de audio binaurales y los datos del sensor del teléfono inteligente, como los obtenidos en el primer paso 1801.

10 En un quinto paso 1805, se utilizan los datos recogidos para estimar la HRTF y la ITDF de acuerdo con los métodos descritos anteriormente.

La Figura 18 ilustra la captura de la información de orientación a partir de una unidad de orientación montada de forma fija a la cabeza. La unidad de orientación puede estar embebida en un teléfono inteligente, pero la presente invención no está limitada a ello.

15 Las Figuras 18(a) a 18(c) muestran un ejemplo de datos de medida en bruto como los que se pueden obtener de una unidad de orientación 1801 que se montó de forma fija a una cabeza robótica 1802.

20 En el ejemplo mostrado, se utilizó como unidad de orientación una Unidad de Medida Inercial (IMU) "PhidgetSpatial Precision 3/3/3 High Resolution" disponible comercialmente de "Phidgets Inc." (Canadá), pero la invención no está limitada a ello, y también se pueden utilizar otras unidades de orientación capaces de proporcionar información de orientación a partir de la cual se puede derivar una orientación única en espacio 3D (por ejemplo en forma de ángulos con respecto al campo magnético de la tierra y al campo gravitatorio de tierra). Esta IMU tiene varios sensores de orientación: un acelerómetro, un magnetómetro, y un giróscopo. Formas de onda de datos ejemplares proporcionadas por cada uno de estos sensores se muestran en las Figuras 18(a) 18(c). Esta información fue leída a través de cables 1904 por un dispositivo informático (no mostrado en la Figura 19). El periodo de muestreo para la medida de la IMU se estableció en 16 ms.

25 En el experimento se utilizaron datos procedentes de los tres sensores, porque eso proporciona los resultados más precisos. La orientación estimada de la IMU se puede representar en la forma de los así llamados cuaterniones, véase la Figura 18(d). La orientación de la IMU se estima cada 100 ms, utilizando un método de procesamiento por lotes que estima la orientación de la IMU sin utilizar sólo datos instantáneos, pero.

30 La Figura 18(e) muestra un dispositivo robótico 1903 que se utilizó durante la evaluación. Una cabeza simulada 1802 con orejas parecidas a las de un ser humano se montó al dispositivo robótico 1803 para simular movimientos de la cabeza. Una unidad de orientación 1801 se montó de forma fija a la cabeza, en el ejemplo encima de la cabeza, pero eso no es absolutamente necesario, y la invención también funcionará cuando la unidad de orientación se monte en cualquier otra posición arbitraria, siempre y cuando la posición sea fija durante el experimento. Asimismo, no es necesario que la orientación de la unidad de orientación esté alineada con la parte delantera de la cabeza, lo que significa por ejemplo que se permite que el lado delantero de la unidad de orientación apunte al oído izquierdo, o al oído derecho, o a la parte delantera de la cabeza, o a la parte trasera, no importa. El lector atento recordará que el método de la Figura 14 puede calcular la orientación de la unidad de orientación 1801 con respecto a la cabeza 1802.

35 En el experimento, el dispositivo robótico se programó para mover la cabeza de acuerdo con un patrón predefinido (conocido). Los resultados de un ensayo mostraron buen acuerdo ($<3^\circ$) entre los movimientos reales de cabeza y la orientación medida. Dado que sensores de orientación similares están embebidos hoy en día en los teléfonos inteligentes (y se están utilizando por ejemplo en aplicaciones de orientación), se contempla que los sensores embebidos en un teléfono inteligente se puedan utilizar para obtener dicha información de orientación. Incluso si la orientación de cada medida individual no fuera perfecta, por ejemplo si se produjeran hipos en uno de los sensores, esto se puede detectar y/o corregir fácilmente utilizando la información del otro sensor, y/o mediante interpolación (suponiendo movimientos lentos de la cabeza), y/o teniendo en cuenta información espacial procedente de los datos de audio capturados. Esta última posibilidad es puramente opcional: algunas realizaciones de la presente invención sólo utilizarán información de orientación obtenida de la unidad de orientación (sin utilizar información espacial procedente de los datos capturados). Otras realizaciones de la presente invención utilizarán tanto información de orientación procedente de la unidad de orientación como información espacial extraída del audio capturado. Los experimentos han mostrado que esto último puede no ser necesario.

40 Las Figuras 20(a) a 20(d) son unas pocas instantáneas de una persona realizando movimientos de cabeza lentos durante el paso de obtención de datos, lo que significa la captura de datos de audio y de datos de orientación.

45 En el ejemplo mostrado, la persona está sentada en una silla giratoria y mueve su cabeza lentamente (es decir, no de forma brusca) en "muchas direcciones diferentes" durante un periodo de tiempo de aproximadamente 10 minutos, mientras una señal acústica está siendo emitida por un altavoz (no mostrado en la Figura 20), comprendiendo la señal acústica una pluralidad de estímulos de ensayo acústicos, por ejemplo pitidos y/o chirps.

En la secuencia de imágenes mostrada en la Figura 20, se muestra una trayectoria de movimientos lentos de la cabeza, que duró aproximadamente 3 segundos.

Es importante que no es necesario que la persona siga trayectorias particulares, sino que puede mover libremente su cabeza, lo cual hace el paso de obtención de datos muy conveniente para el usuario. La intención es que se gire la cabeza substancialmente en “todas las direcciones posibles” sobre la esfera, para permitir determinar la ITDF y la HRTF para sonido procedente de cualquier punto situado en una esfera virtual alrededor de la cabeza de la persona (por ejemplo desde delante, desde detrás, desde la derecha, desde la izquierda, desde arriba, desde abajo, y todas las posiciones entre estas). Por supuesto, algunas áreas de la esfera no se muestrearán, debido a las limitaciones físicas del cuerpo humano.

En los ejemplos mostrados en la Figura 20, la persona está sentada en una silla giratoria, lo cual es muy conveniente para el usuario. Las realizaciones de la presente invención pueden tener esto en cuenta, cuando se determina la posición media de la cabeza, como se describirá con mayor detalle en la Figura 31. Sin embargo, la invención no está limitada a esto, y los datos también se pueden obtener cuando el usuario está sentado en una silla estacionaria, o está sentado sobre sus rodillas o permanece de pie recto. En estos casos, las realizaciones de la presente invención suponen que el centro de la cabeza está situado en una posición fija (aunque desconocida) durante la captura de datos, pero capaz de girar alrededor del centro de la cabeza.

La Figura 20 muestra una disposición típica de la persona sentada en una silla en una habitación típica 2000 de una casa típica durante el paso de captura de datos. La habitación 2000 tiene un techo situado a una altura “hc” por encima del suelo, típicamente en el rango de 2,0 a 2,8 m. Un altavoz 2002 está situado en la habitación a una altura “he”, por ejemplo igual a aproximadamente 80 a 120 cm por encima del suelo. La cabeza 2001 de la persona está situada a una altura “hx” por encima del suelo, por ejemplo a aproximadamente de 120 a 160 cm, y a una distancia “d” del altavoz, típicamente aproximadamente de 1,0 a 2,0 m de distancia.

Es una ventaja de la presente invención que estos valores “he”, “d”, “hx” o cualquier ángulo asociado, en particular la orientación relativa del altavoz con respecto a la cabeza de la persona, no se conozcan, y no sea necesario conocerlos, de antemano, y no sea necesario calibrarlos utilizando algún tipo de medida, pero que el algoritmo pueda no obstante determinar o estimar la “dirección de la fuente” relevante, lo cual es clave para la ITDF y la HRTF, sobre la base de datos de audio binaurales, información de orientación o datos obtenidos de una unidad de orientación fija montada a la cabeza, además en una posición y orientación arbitrarias.

La Figura 21 ilustra características de un así llamado “chirp” como estímulo acústico ejemplar para estimar la ITDF y la HRTF, pero la invención no está limitada a esta forma de onda particular, y también se pueden utilizar otras formas de onda, por ejemplo un chirp con una frecuencia linealmente creciente o un chirp con una frecuencia decreciente de manera no lineal, o un chirp que tenga un perfil de frecuencia en forma de escalera, o incluso un tono puro. Se describirá la invención para el chirp mostrado en la Figura 21.

En el Apéndice al final de la descripción se describen algunos aspectos de cómo se puede diseñar un chirp apropiado teniendo en cuenta algunas características de una habitación típica, y cuál es un intervalo de tiempo apropiado entre dos chirps pero, a fin de entender la presente invención, basta con saber que cada chirp tiene una duración de tiempo predefinida “T”, típicamente un valor en el rango de 25 a 50 ms. El chirp puede comprender un barrido de frecuencia lineal desde una primera frecuencia fH hasta una segunda frecuencia fL, por ejemplo desde 20 kHz hasta 1 kHz. Como se describe en el Apéndice, esto permite medir la ITDF y la HRTF con una resolución de frecuencia δf igual a aproximadamente 300 Hz.

La Figura 22 ilustra los posibles pasos a dar para extraer los tiempos de llegada de los chirps y la información espectral.

La Figura 22(a) muestra el espectrograma de una señal de audio capturada por el micrófono de oído izquierdo, para una señal de ensayo de audio que comprende cuatro chirps consecutivos, cada uno de los cuales tiene una duración de aproximadamente 25 ms con un intervalo entre chirps de 275 ms. Este espectrograma se obtiene aplicando una Transformación Rápida de Fourier después de una aplicación de funciones ventana (“*windowing*”) apropiada a las muestras de audio respectivamente izquierda y derecha, de maneras conocidas por sí misma en la técnica. La Figura 21 también muestra que las señales de eco son una versión amortiguada de la señal emitida después de una o más reflexiones contra partes de la habitación (por ejemplo suelo y techo) o contra objetos presentes en la habitación (reverberaciones). Los métodos de la presente invención preferiblemente sólo trabajan con la “parte directa de la señal”.

La Figura 22(b) muestra el espectrograma “rectificado”, es decir, cuando se han compensado en él los retrasos de temporización dependientes de la frecuencia conocidos en los chirps.

La Figura 22(c) muestra la intensidad sumada de la señal de audio izquierda y derecha, en base a la cual se pueden determinar los tiempos de llegada de los chirps.

La Figura 23 muestra un ejemplo de los espectros extraídos de la señal de audio izquierda (figura 23a: espectros del oído izquierdo) y de los extraídos de la señal de audio derecha (figura 23b: espectros del oído derecho), y la diferencia de tiempo interaural (figura 23c) para una señal de ensayo de audio ejemplar que comprende cuatro mil chirps.

La Figura 24 muestra parte de los espectros y de los datos de ITD de la Figura 23 con mayor detalle.

Las Figuras 25 a 30 se utilizan para ilustrar un principio subyacente importante de la presente invención. Dichas Figuras están relacionadas principalmente con el método 1300 para estimar la dirección de la fuente con respecto al mundo, mostrado en la Figura 13, la cual se puede encontrar de forma iterativa maximizando un valor de calidad predefinido de acuerdo con un criterio de calidad predefinido.

En las realizaciones preferidas, el criterio de calidad está relacionado con una "métrica de suavidad", pero también se pueden utilizar otros criterios de calidad, tales como por ejemplo una función de probabilidad, donde la probabilidad de ciertos rasgos o características como los que se pueden extraer o derivar de los datos de audio binaurales después de ser mapeados sobre una superficie esférica, donde el mapeado se basa en la dirección supuesta de la fuente (altavoz) c. r. al mundo, y donde los datos de audio están asociados con información de orientación también c. r. al mundo.

Haciendo referencia en primer lugar a la Figura 25, la Figura 25(a) es un ejemplo en el que los valores de ITD de los 4000 chirps (véase la Figura 24) están mapeados sobre una superficie esférica, suponiendo una dirección aleatoria (pero incorrecta) de la fuente. Como se puede ver en la Figura 27(a), existen muchos "puntos oscuros" en áreas brillantes y "puntos brillantes" en "áreas oscuras" o, dicho de otra manera, la superficie tiene un alto grado de irregularidad, de discontinuidad, no cambia gradualmente, no es suave. Todas estas expresiones están relacionadas con "suavidad", pero se pueden expresar o calcular de diferentes maneras.

En cambio, si el mapeado se realiza basándose en la dirección correcta de la fuente (c. r. al mundo), como se ilustra en la Figura 27(b), entonces se forma una superficie, la cual cambia de manera mucho más continua, de manera mucho más suave, tiene menos irregularidades, cambia de forma menos brusca, etc. El lector debería ignorar las áreas blancas puras, correspondientes a direcciones para las cuales no está disponible ningún dato real o, dicho de otra manera, que no están mapeadas sobre la superficie.

Como se ha explicado anteriormente, los inventores llegaron a la idea de aprovechar este efecto para "encontrar" la dirección de la fuente, comprobando la calidad, por ejemplo el grado de continuidad, el grado de cambios bruscos, el grado de suavidad, para una pluralidad de direcciones candidatas de la fuente, y eligiendo la dirección candidata de la fuente que proporciona el mayor valor de calidad.

La Figura 25 muestra el efecto perjudicial de una dirección incorrectamente supuesta de la fuente sobre la suavidad de la superficie proyectada de las medidas de ITD.

La Figura 25(a) muestra un mapeado de los datos de ITD de los cuatro mil chirps de la Figura 23 sobre una superficie esférica, utilizando una dirección de la fuente aleatoria (pero incorrecta), dando como resultado una función con un alto grado de irregularidades o con baja suavidad.

La Figura 25(b) muestra un mapeado de los datos de ITD de los cuatro mil chirps de la Figura 23 sobre una superficie esférica, utilizando la dirección correcta de la fuente, dando como resultado una función con un alto grado de regularidades o con alta suavidad.

La Figura 25(c) y la Figura 25(d) muestran el efecto de una dirección incorrectamente supuesta de la fuente sobre la suavidad de los datos espectrales obtenidos a partir de los chirps. En el ejemplo se utilizó información espectral a 8100 Hz, pero también se puede elegir otra frecuencia. Como se puede ver, la superficie de la Figura 27(c) es muy irregular, mientras que la superficie de la Figura 27(d) es mucho "más suave".

Se contempla que se pueden utilizar muchas maneras diferentes para expresar el grado de continuidad o suavidad, denominado en esta memoria "valor de calidad".

En realizaciones preferidas de la presente invención, se determina la suavidad calculando una "distancia total" entre la ITD valores espectrales mapeados y una versión paso bajo filtrada espacialmente de los datos mapeados, que se puede considerar una "superficie de referencia". Se contempla que se puedan utilizar para este fin técnicas de filtrado conocidas. Es importante observar que la "superficie de referencia" así obtenida no está predeterminada, y no se deriva de una base de datos de IDT o de HRTF, sino que se deriva de los propios datos capturados, dicho de otra manera, también la superficie de referencia es personalizada.

La Figura 26 ilustra una manera particular para determinar una "superficie de referencia", basada en aproximar la superficie mediante una serie de un número limitado de funciones base ortogonales, en particular limitando el orden máximo de la serie.

En las realizaciones preferidas, las funciones base ortogonales son "funciones armónicas esféricas".

La Figura 26(a) muestra una representación gráfica de estas funciones base, para dar una idea de qué aspecto tienen las funciones armónicas esféricas. Los lectores familiarizados con las técnicas de procesamiento de imágenes reconocerán similitudes con las series de Fourier, pero ahora las funciones básicas están definidas sobre la esfera.

Se encontraron buenos resultados para órdenes en el rango de 5 a 15, por ejemplo 10. El valor del orden no parece ser crítico.

5 Haciendo referencia a la Figura 27(a), cuando se determina el "factor de calidad" o "valor de suavidad" de la 'dirección candidata de la fuente' dando lugar a esta superficie, se determina primero "una superficie de referencia" para esta superficie, por ejemplo aproximando la superficie con una serie de funciones armónicas esféricas con orden=10.

A continuación, se calcula una "distancia total" entre los datos de medida mapeados y la superficie de referencia (suave), como la suma de las diferencias al cuadrado para todas las medidas (por lo tanto para cada chirp). Se puede utilizar cualquier "criterio de distancia" o "métrica de distancia" apropiados, por ejemplo:

$d1 =$ valor absoluto de la diferencia entre los datos reales y los datos de referencia, o

10 $d2 =$ cuadrado de la diferencia entre los datos reales y los datos de referencia, o

cualquier otro criterio de distancia apropiado. Hacemos referencia al Apéndice para más detalles.

La Figura 26(b) muestra una técnica para cuantificar la suavidad de una función definida sobre la esfera, por ejemplo ITDF, la cual se puede utilizar como una métrica de suavidad.

15 La Figura 27(a) muestra el valor de suavidad (indicado en sombras grises) de acuerdo con la métrica de suavidad definida en la Figura 26(b) para dos mil "direcciones de la fuente" candidatas representadas sobre una esfera, cuando se aplican a los valores de ITD, con el orden de los armónicos esféricos configurado a 5. La escala de grises está ajustada en la Figura 27(b). De esta Figura es evidente que los valores de suavidad sobre la esfera alcanzan un mínimo evidente y, como resultado la dirección de la fuente con respecto al mundo, se puede localizar en esta dirección (o punto sobre la esfera). No es visible en esta Figura, pero la superficie que representa los valores de suavidad exhibe simetría especular y un mínimo local también está situado en el lado opuesto de la esfera. Esto explica por qué sólo
20 se puede estimar la dirección de la fuente en 1002 y 1300, y no el signo. Obsérvese también que, al menos en este ejemplo particular, la superficie que representa los valores de suavidad no tiene otros mínimos locales, simplificando considerablemente la búsqueda.

25 La Figura 28(a) muestra los valores de suavidad indicados cuando se aplica el criterio de suavidad a espectros binaurales, con el orden de los armónicos esféricos configurado a 5, siendo el valor de suavidad para cada coordenada mostrado sobre la esfera la suma del valor de suavidad para cada una de las frecuencias en el rango de 4 kHz a 20 kHz, en pasos de 300 Hz. La escala de grises está ajustada en la Figura 28(b). Se pueden extraer conclusiones similares a las de la Figura 27(a).

30 La Figura 29(a) muestra los valores de suavidad, cuando se aplica el criterio de suavidad a espectros binaurales, con el orden de los armónicos esféricos configurado a 15. La escala de grises está ajustada en la Figura 29(b). Se pueden extraer conclusiones similares a las de la Figura 27(a).

La Figura 30(a) muestra los valores de suavidad, cuando se aplica el criterio de suavidad a espectros monoaurales, con el orden de los armónicos esféricos configurado a 15. La escala de grises está ajustada en la Figura 30(b). Se pueden extraer conclusiones similares a las de la Figura 27(a).

35 Los ejemplos anteriores ilustran que el principio de encontrar la dirección de la fuente c. r. al mundo de la manera descrita anteriormente, basado en minimizar o maximizar un valor de calidad, funciona y es muy preciso. Además, es bastante factible en términos de complejidad computacional, no requiere enormes cantidades de memoria o potencia de procesamiento. Por ejemplo, no se requiere ningún DSP.

40 La Figura 31 ilustra los parámetros del modelo de un modelo a priori del movimiento del centro de la cabeza, que se podrían utilizar en 1004, 1104, 1503. Cuando una persona está sentada en una silla de oficina y se le permite que gire su cabeza libremente en todas las direcciones, y que gire libremente junto con la silla con el cuerpo fijado a la silla, entonces el movimiento del centro de la cabeza se puede describir utilizando este modelo mecánico relativamente simple. El centro de la cabeza (rc) está a una distancia b de la base del cuello (un punto de rotación), la base del cuello está a una distancia a del centro de rotación de la silla.

45 Pero también se contemplan otros modelos mecánicos de movimiento de la cabeza, por ejemplo un modelo como el de la Figura 32, pero sin el movimiento de la silla, suponiendo por lo tanto que la cabeza está montada sobre un cuello (distancia $a=0$).

50 En otra variante de la Figura 32, algo más compleja que el modelo mostrado en la Figura 32, el modelo también tiene en cuenta que la persona se puede inclinar hacia adelante o hacia atrás en la silla, por lo tanto existe un grado de movimiento adicional.

Se contempla que la gran cantidad de datos permite determinar los parámetros (más probables) del modelo, y una vez que se conocen los parámetros del modelo, se pueden utilizar la información de orientación y/o la información acústica para determinar un estado particular del modelo en el momento de capturar cada fragmento de audio.

La Figura 32 muestra imágenes de un video que captura un sujeto cuando realiza una medida HRTF en la silla con el giro libre. Utilizando el modelo mecánico de la Figura 31, se extrajo información sobre la posición de la cabeza, (la cual produjo como resultado mejores estimaciones de la dirección de la fuente con respecto a la cabeza), como se puede ver a partir de las visualizaciones de la orientación y la posición estimadas de la cabeza. La línea negra muestra la desviación del centro de la cabeza con respecto al centro medio de la cabeza. Estas desviaciones tendrán efecto sobre la dirección percibida de la fuente con respecto a la cabeza, especialmente cuando se mueve la cabeza en perpendicular a la fuente. Por tanto, la inclusión de estas traslaciones del centro de la cabeza mejorará la estimación de la HRTF y de la ITDF en 1005 y 1105.

La Figura 33 es una representación gráfica de las posiciones estimadas (en coordenadas del mundo X, Y, Z) del centro de la cabeza durante un ensayo de captura de audio ejemplar, utilizando el modelo mecánico de la Figura 31. Cada punto corresponde a una posición del centro de la cabeza en el momento de llegada de un chirp. Obsérvese que el centro estimado de la cabeza sigue una trayectoria continua (puntos consecutivos están conectados con una línea). Cada instantánea mostrada en la Figura 32 corresponde con un punto particular a lo largo de esta trayectoria.

La Figura 34 muestra una medida de la distancia entre el centro de la cabeza y la fuente del sonido a lo largo del tiempo, determinada a partir de los retrasos de temporización entre chirps consecutivos. En efecto, si el centro de la cabeza no se moviera, entonces el tiempo entre chirps recibidos de manera sucesiva sería constante. Pero si la cabeza se mueve, los chirps se retrasarán cuando la cabeza se aleja de la fuente, o, llegarán más pronto cuando la cabeza se acerca a la fuente. Las diferencias en los tiempos de llegada de los chirps se pueden traducir fácilmente en diferencias de distancia mediante multiplicación por la velocidad del sonido. Estas variaciones de distancia se pueden utilizar a continuación como entrada en 1503, para estimar los parámetros del modelo del modelo mecánico mostrado en la Figura 31. Es evidente de la curva (originariamente) roja que el modelo mecánico de la Figura 31 permite un buen ajuste con estas variaciones de distancia medidas (curva originariamente azul).

La Figura 35 muestra una comparación de dos HRTFs de la misma persona: una se midió en una instalación profesional (en Aquisgrán), la otra HRTF se obtuvo utilizando un método de acuerdo con la presente invención, medida en casa. Como se puede ver, existe muy buena correspondencia entre la representación gráfica de la HRTF medida en la instalación profesional y la HRTF medida en casa.

otras consideraciones:

Un paquete comercial vendido al usuario puede comprender: una pareja de micrófonos de oído, y un CD de audio con la señal de ensayo acústica. Opcionalmente el paquete también puede contener una tira para la cabeza por ejemplo una tira elástica para la cabeza, para fijar el dispositivo portátil o el conjunto de dispositivo portátil a la cabeza de la persona, pero esto último no es esencial. De hecho, tampoco el CD de audio es esencial, ya que el fichero de sonido también se podría descargar de una página web particular, o podría ser proporcionado por otros medios de almacenamiento, tales como por ejemplo un DVD-ROM o una tarjeta de memoria, o similar. Se espera que el usuario final ya posea el otro hardware necesario, en particular un dispositivo que comprenda una unidad sensora de orientación (tal como por ejemplo un teléfono inteligente apropiado), y un sistema de reproducción de sonido con un altavoz (por ejemplo una cadena estéreo, o un ordenador con una tarjeta de sonido, o un reproductor MP3 o similar) y una unidad de captura de audio (por ejemplo dicho teléfono inteligente equipado con un dispositivo añadido, o un ordenador, o similar), pero también podría ser ofrecido como parte del paquete.

El método, el programa informático y el algoritmo de la presente invención no tienen como objetivo proporcionar la HRTF y la ITDF más precisas, sino más bien aproximarlas lo suficientemente de modo que al menos los problemas principales de fallos de percepción de la parte delantera frente a la parte trasera, y/o fallos de percepción de arriba frente a abajo se reduzcan drásticamente, y preferiblemente se eliminen completamente.

La presente invención utiliza tecnologías de uso generalizado hoy en día (teléfonos inteligentes, micrófonos, y altavoces), combinadas con un procedimiento fácil de usar que permite al usuario ejecutar el procedimiento por sí mismo. Aunque los teléfonos inteligentes son de uso generalizado, la utilización de un teléfono inteligente para registrar señales de audio estéreo en combinación con información de orientación no es de conocimiento general, y mucho menos la utilización de las señales de audio para corregir la información de orientación, relacionar la orientación desconocida de la unidad de orientación con el marco de referencia de la cabeza tal como se utiliza en medidas de HRTF e ITDF estándar, y localizar la fuente del sonido. Esto significa que el método propuesto en esta memoria es más flexible (más fácil de usar), y que la complejidad del problema se desplaza del paso/configuración de captura de datos hacia el post-procesamiento, es decir, hacia el algoritmo de estimación.

Lista de referencias:

501, 601, 801: ordenador. 502, 602, 702, 802: altavoz. 503, 603, 703, 803: persona. 504, 604, 704, 804: unidad de orientación. 505, 605, 705, 805: micrófonos de oído. 506: soporte- 507: silla. 608, 708, 808: equipo de reproducción de sonido

Apéndice:

Como una prueba de principio, en lo que sigue se muestran resultados que se obtuvieron utilizando un método de acuerdo con una realización particular de la presente invención.

La configuración de medida

5 Se utilizó un ordenador de una sola placa (SBC) Raspberry PI 2 modelo B para capturar y almacenar datos de audio. Como unidad de orientación se utilizó una unidad de medida inercial (IMU) PhidgetSpatial Precision 3/3/3 High Resolution. Esta IMU mide datos del giróscopo, del magnetómetro y del acelerómetro. El SBC está ampliado con una tarjeta de sonido (Wolfson Audio Card), que permite grabación estéreo a 44,2 kMuestras/sec con una resolución de 16 bits. Las capacidades de detección y almacenamiento de esta configuración son comparables a las de al menos algunos dispositivos de teléfono inteligente actuales (año 2016).

El sonido binaural se capturado mediante micrófonos binaurales comerciales (Soundman OKM II Classic) utilizando la técnica del canal auditivo bloqueado, aunque esto último no es absolutamente necesario.

15 El procesamiento de los datos obtenidos se llevó a cabo en un ordenador portátil (Dell Latitude E5550, Intel Core™ i7 de doble núcleo de 2,6 GHz, con 8 GBytes de RAM, Windows10, 64 bits). Todo el procesamiento de señal se programó en Matlab R2015b. El tiempo total de procesamiento para procesar 15 minutos de sonido estéreo e información de orientación asociada fue de aproximadamente 30 minutos, no estando el código optimizado para lograr una mayor velocidad.

La señal de sonido de estímulo se reprodujo a través de un único altavoz (JBC), utilizando un sistema Hi-Fi ordinario presente en casa.

20 Todas las medidas se llevaron a cabo en casa, en una sala de estudio sin modificar (dimensiones aproximadas 4 m x 3 m x 2,5 m de altura, sólo de madera, paredes enyesadas, cortinas, escritorio, armarios, etc.). El sujeto se sentó en una silla de oficina ordinaria ubicada a aproximadamente 1,5 m del altavoz, el cual apuntaba aproximadamente al eje de rotación de la silla. El sujeto recibió instrucciones de sentarse erguido y doblar su cabeza libremente en todas las direcciones (de arriba a abajo, hacia los lados). Se le dieron instrucciones de hacer girar la silla libremente pero de forma lenta (utilizando sus piernas), sin mover mientras tanto su torso sobre la silla. Aparte de estas instrucciones los movimientos del sujeto no se controlaron de ninguna manera. La IMU se fijó en una posición arbitraria y en una orientación arbitraria a la parte posterior de la cabeza del sujeto. Las dimensiones exactas de la habitación, la altura de la fuente, la posición del sujeto con respecto al altavoz, la posición/orientación de inicio, las configuraciones del altavoz/sistema hi-fi no eran conocidas a priori para el sistema ni estaban controladas a priori por el sistema.

30 Estimación de la orientación de la IMU

La orientación de la IMU se estimó en base a los datos de sensor del giróscopo, del magnetómetro, y del acelerómetro, utilizando el método de Gauss-Newton clásico (de procesamiento por lotes). La orientación de la IMU se representó con cuaterniones. Las Figuras 18(a)-(d) muestran un ejemplo de dichos datos de (a) acelerómetro, (b) magnetómetro y (c) giróscopo registrados y (d) la dinámica (de orientación) del cuaternión estimada a lo largo del tiempo.

35 La señal de estímulo

40 Se diseñó una señal de estímulo acústica que presenta un compromiso razonable entre las diferentes condiciones de contorno (dimensiones medias de la habitación, duración limitada del experimento) permitiendo la extracción de la información acústica relevante (un rango de frecuencia de desde aproximadamente 1 kHz hasta aproximadamente 20 kHz, una resolución de frecuencia de aproximadamente 300 Hz y relación señal-ruido suficiente para una duración total de la medida de entre 10-20 minutos).

45 Para que la medida se pueda llevar a cabo en casa, uno tiene que tratar con las reflexiones de los sonidos que rebotan en el suelo, en las paredes y en el techo. Esto se consigue trabajando con chirps de banda ancha cortos, entrelazados con un periodo de silencio intermitente suficientemente largo (tiempo entre estímulos). Es ventajoso aislar sólo el sonido que viaja a lo largo de la ruta directa, y separarlo de las primeras reflexiones, véase la Figura 20. El tiempo entre la llegada del sonido directo y la primera reflexión en el sujeto es una propiedad de la configuración de medida (posiciones de la cabeza y del altavoz en la habitación). En esta medida, el sujeto estaba sentado separado a una distancia de aproximadamente $d=1,5$ metros del altavoz, tanto la cabeza como el altavoz estaban a una altura de aproximadamente $h_e = h_x = h_c/2 =$ aproximadamente 1,30 m, que es aproximadamente la mitad de la altura de la habitación. (Véase la Figura 20 para las definiciones de h_e , h_x y h_c).

50 La resolución de frecuencia con la cual se puede extraer el contenido espectral del sonido directo, depende del tiempo hasta la primera reflexión (Δt), de la duración (T) y del rango de frecuencia (Δf) del chirp, véase la Figura 21. Cada combinación permite una resolución de frecuencia (δf) particular, la cual se puede obtener utilizando la siguiente desigualdad:

$$\frac{T\delta f}{\Delta f} + \frac{1}{\delta f} < \Delta t$$

En los resultados experimentales mostrados, se utilizó un barrido de chirp linealmente descendente desde $f=20\text{kHz}$ hasta 1Hz durante $T=25\text{ ms}$. Esto permite una resolución de frecuencia δf de aproximadamente 300 Hz , que es similar a la resolución de frecuencia utilizada en las bases de datos de HRTF comunes (véase CIPIC: 223 Hz). Pero también se pueden utilizar estímulos diferentes (barrido exponencial, duración diferente, rango de frecuencia diferente, etc.).

Además, el tiempo entre chirps debería ser suficientemente grande, de tal manera que la grabación de un chirp no se vea influida significativamente por el sonido de los uno o más chirps previos, que todavía reverberan en la habitación. El tiempo de reverberación es una propiedad de la habitación, que depende de las dimensiones y de las propiedades de absorción/reflexión del contenido (por ejemplo paredes, mobiliario, etc.). El tiempo de reverberación se expresa a menudo como el tiempo necesario para que la intensidad del sonido disminuya 60 db . En las habitaciones encontradas durante nuestros ensayos, un tiempo entre chirps de 275 ms fue suficiente para impedir que efectos de reverberación afectaran a la calidad de las medidas. Si se aplica el método, en habitaciones con alta reverberación podría ser necesario incrementar este tiempo entre chirps dando como resultado una duración de la medida más larga.

Extrayendo información de temporización y espectral

A fin de extraer la información de temporización y espectral de las señales de audio capturadas, se utilizó una representación de espectrograma de las señales del micrófono y se dibujó su módulo al cuadrado, proporcionando información espectral en función del tiempo. En la Figura 22(a), se muestra el espectrograma para $1,2\text{ seg}$ de sonido grabado (en un oído). A continuación, el espectrograma se 'rectificó', compensando los retrasos de temporización dependientes de la frecuencia conocidos en los chirps, véase la Figura 22 (b). A continuación se sumó la intensidad a lo largo del eje de frecuencia, como se muestra en la Figura 22(c). El tiempo de llegada estimado de un chirp es ahora el tiempo en el cual el patrón de intensidad sumada correspondiente a este chirp tiene un pico. El contenido espectral se obtiene entonces evaluando el espectro en el tiempo de llegada correspondiente en el espectrograma rectificado mostrado en la Figura 22(b). El contenido espectral correspondiente para los diferentes chirps se muestra en la Figura 23(a, b), respectivamente para el oído izquierdo (a) y para el oído derecho (b) en una escala de dB. Se observa que esta no es la única manera de extraer información de temporización y espectral, existen muchas otras maneras, por ejemplo filtrado inverso.

Estimación de la dirección de la fuente de sonido

Para estimar la "dirección de la fuente de sonido", se utilizan las orientaciones de la IMU (a partir de los datos del sensor de orientación) y la información espectral y/o la información de ITD extraídas (a partir de los datos de audio binaurales). La técnica utilizada está basada parcialmente en el hecho de que la HRTF y la ITDF son funciones espacialmente suaves. El método se puede comprender como se explica a continuación.

Primero se determinan las HRTF/ITDF con respecto a la IMU (no con respecto a la cabeza, lo cual es anti- intuitivo, porque la HRTF se expresa siempre con respecto a la cabeza). Si se conociera la dirección de la fuente exacta r con respecto al marco de referencia del mundo, se podría relacionar con cada medida de orientación de la IMU una única dirección de la fuente muestreada $(\theta_i, \phi_i)=r_i$, lo cual daría como resultado una versión muestreada de forma discreta de la HRTF ($S^r(r_i)$), como se muestra en la Figura 25(d) para $f=8100\text{ Hz}$. Se puede reconocer un patrón relativamente suave en toda la esfera. Sin embargo, si se supone una dirección errónea de la fuente con respecto al marco de referencia del mundo, se llega a un patrón diferente, mucho más caótico y menos suave, como se muestra en la Figura 25(c). Los inventores llegaron a la conclusión de que, a partir de la perspectiva de la IMU, elecciones diferentes para la dirección de la fuente no dan como resultado únicamente una rotación de la HRTF verdadera, sino que, en vez de esto, como se puede entender comparando las Figuras 25(c) y (d), dan lugar a HRTFs que contienen grandes cantidades de variación espuria. Por tanto, se puede utilizar la característica de 'suavidad' de la HRTF y/o de la ITDF para derivar un criterio de calidad para evaluar direcciones candidatas de la fuente. La optimización de este criterio de calidad conduce entonces a la mejor estimación de la dirección de la fuente del sonido.

Se pueden elegir diferentes criterios para cuantificar la 'suavidad'. En esta solicitud de patente, los datos de HRTF medidos se expanden en armónicos esféricos (SH) reales,

los cuales son funciones base similares a funciones base de Fourier, pero definidas sobre una esfera. De forma similar a las funciones base de Fourier, las funciones base SH reales $Y_{lm}(\theta, \phi)$ tienen la propiedad de que valores- l menores corresponden a funciones base que varían más lentamente. Por tanto, esto significa que si la HRTF se expresa en una base truncada que contiene sólo funciones base hasta un orden máximo elegido o predefinido L ($l < L$), se aplica de forma efectiva un filtro paso bajo que sólo permite variaciones espaciales lentas. Cuanto mayor es el valor L elegido, más 'detalle' espacial incluye la expansión de base. Por tanto, para cuantificar 'suavidad', estimamos primero los coeficientes $C_{l,m}^{r,R}(f)$ y $C_{l,m}^{r,L}(f)$ de la expansión HRTF (Correspondiendo $C_{l,m}^{r,R}(f)$ y $C_{l,m}^{r,L}(f)$ respectivamente a la HRTF del oído izquierdo y a la HRTF del oído derecho a frecuencia f para la dirección elegida r) en la base SH truncada a alguna L elegida. A continuación, calculamos la diferencia al cuadrado entre los puntos de datos medidos y la

expansión HRTF obtenida (en la cual se calcula una suma para de todas las direcciones medidas y todas las frecuencias medidas):

$$S_{L/R}^r(f, r_i) \approx \sum_{l=0}^L \sum_{m=-l}^l C_{l,m}^{r,L/R}(f) Y_{lm}(r_i)$$

$$\varepsilon_{HRTF}^2(\mathbf{r}) = \sum_f \sum_{r_i} \left\{ \left[S_L^r(f, r_i) - \sum_{l=0}^L \sum_{m=-l}^l C_{l,m}^{r,L}(f) Y_{lm}(r_i) \right]^2 + \left[S_R^r(f, r_i) - \sum_{l=0}^L \sum_{m=-l}^l C_{l,m}^{r,R}(f) Y_{lm}(r_i) \right]^2 \right\}$$

5 Este error cuantifica hasta qué punto la base de funciones base que varían lentamente es adecuada para describir el patrón espacial presente en la HRTF medida en toda la esfera. Cuanto menor fue el error, mejor se aproximaron los datos acústicos utilizando sólo funciones base que varían lentamente, y por consiguiente, más suave es el patrón HRTF. Por consiguiente, se puede utilizar este error como criterio de calidad. Obsérvese que el mismo procedimiento también se puede aplicar utilizando medidas de HRTF o de ITDF monoaurales.

10 Se utilizó el método de Gauss-Newton para estimar la dirección de la fuente r , por medio de minimización de $\varepsilon_{HRTF}^2(r)$. En la presente implementación, se utiliza $L=10$ para la expansión de la HRTF, pero también pueden servir otros valores mayores de 10, por ejemplo 15, o menores de 10, por ejemplo $L=9$ o $L=8$ o $L=7$ o $L=6$ o $L=5$ o $L=4$. Se observa que se utilizó información de HRTF binaural para un rango de frecuencia de 5 kHz - 10 kHz, pero se podría utilizar también información de ITDF o información espectral monoaural, o también se podría elegir un rango de frecuencias diferente.

15 Se encontró que la dirección óptima de la fuente de sonido era muy parecida a la dirección real. Ejemplos de este error en la esfera se muestran en las Figuras 27, 28, 29 y 30, en base a la información de ITDF y de HRTF monoaural/binaural, para diferentes valores de L .

Los r_i resultantes con sus correspondientes valores $S^r(f, r_i)$ se muestran en la Figura 25(d) para el oído derecho y una frecuencia de 8100 Hz. También se muestra en la Figura 25(b) la ITDF resultante. Se observa que este método sólo permite estimar la dirección de la fuente de sonido excepto por su signo. Así que todavía existe incertidumbre sobre la dirección exacta de la fuente: son posibles dos direcciones opuestas de la fuente. Para resolver esta ambigüedad, se pueden aprovechar otras propiedades de la HRTF.

Se observa que este error también se puede utilizar en un procedimiento iterativo para mejorar aún más la calidad global de la estimación de las HRTF/ITDF; para mejorar la estimación de orientación de la IMU (por ejemplo utilizando los parámetros del modelo del ruido de la IMU); y/o para estimar un retraso de temporización entre datos de orientación y datos de audio (si la captura de datos no fue totalmente sincrona).

También se pueden definir otros criterios de suavidad. Por ejemplo se podría elegir el siguiente:

$$\varepsilon_{HRTF}^2(\mathbf{r}) = \sum_f \left[\left(C_{L,0}^{r,L}(f) \right)^2 + \left(C_{L,0}^{r,R}(f) \right)^2 \right]$$

o

$$\varepsilon_{HRTF}^2(\mathbf{r}) = \sum_f \sum_{r_i} \{ [\nabla^2 S_L^r(f, r_i)]^2 + [\nabla^2 S_R^r(f, r_i)]^2 \}$$

30 También se pueden utilizar otras normas diferentes a la norma Euclidea tales como una norma-p general o una norma de valor absoluto.

Estimación de la orientación del eje oído-oído

Para estimar la orientación del eje oído-oído, se aprovecha la simetría de la ITDF y/o de la HRTF (izquierda vs derecha) con respecto al plano perpendicular al eje oído-oído. En lo que sigue, se utiliza la simetría de la ITDF.

5 En primer lugar se supone un valor particular para la dirección del eje oído-oído a . A continuación se obtiene la imagen especular de todas las direcciones r_i con respecto al plano perpendicular a este eje oído-oído, dando como resultado las direcciones r_i' . A continuación, se supone que los valores de ITD para las direcciones especulares son iguales $ITD_i' = -ITD_i$, y los conjuntos de datos original y especular se fusionan en un único conjunto de datos. Ahora, si se dibuja el conjunto de ITD fusionado, se obtiene como resultado sólo un patrón suave en caso de que la a supuesta sea la dirección verdadera del eje oído-oído. Si se supone un eje oído-oído erróneo, el patrón es de nuevo mucho más caótico.

10 Por tanto, como antes, se utiliza el criterio de 'suavidad' como factor de calidad para estimar la dirección del eje oído-oído, pero ahora proyectando el conjunto de ITD fusionado en una base truncada de armónicos esféricos. De nuevo se utiliza el método de Gauss-Newton para llegar a la mejor estimación de la dirección del eje oído-oído.

Estimación de la dirección frontal del sujeto

15 Se define la dirección frontal de la persona para que coincida con la dirección frontal en medidas HRTF tradicionales (véase la base de datos CIPIC). Expresado en términos sencillos, la dirección hacia delante es parecida a la dirección en la que apunta la nariz de la persona vista desde el centro de la cabeza.

20 Para estimar la dirección frontal del sujeto, se hace girar la HRTF alrededor del eje oído-oído y se compara la HRTF resultante con una HRTF general (por ejemplo la media de una base de datos de HRTFs que ha sido medida bajo circunstancias controladas). Dado que sólo se conoce la dirección de la fuente excepto por su signo, este procedimiento se realiza para las dos direcciones candidatas (=opuestas) de la fuente. A continuación se estiman la dirección frontal y el signo de la dirección de la fuente seleccionando el ángulo de rotación y el signo para el cual la HRTF medida se parece más a la HRTF general.

25 Existen diferentes maneras de comparar dos HRTFs, por ejemplo calculando el producto escalar o calculando la diferencia cuadrática media, etc. En esta implementación, primero se evalúa la HRTF general interpolada en las direcciones muestreadas supuestas, a continuación se normalizan tanto la HRTF general muestreada como la HRTF medida en una base por frecuencia y por último se comparan ambas HRTFs, calculando la diferencia cuadrática media. A continuación se estima la dirección frontal (y el signo de la dirección de la fuente) en base al ángulo (y al signo de la dirección de la fuente) para el cual la diferencia cuadrática media de la HRTF general girada y la HRTF medida es mínima.

30 Estimando la desviación del centro de la cabeza (c. r. al mundo)

Hasta ahora, se supuso que la cabeza está girando alrededor del centro de la cabeza (el cual se define como el punto en el medio entre ambos oídos). Por supuesto, en realidad esto no es así. El centro de la cabeza se moverá hacia atrás y hacia adelante, hacia arriba y hacia abajo, y estas desviaciones con respecto a su posición 'media' tendrán un efecto sobre la dirección que se muestrea realmente, es decir, puede ser diferente a cuando la cabeza permanece fija. 35 Los errores de dirección son mayores cuando la cabeza se mueve alejándose de esta posición 'media', y en particular cuando se mueve en perpendicular a la dirección de la fuente. Incluir estas traslaciones adicionales del centro de la cabeza, mejorará la dirección estimada de la fuente de sonido, y como resultado mejorará también la estimación resultante de la HRTF y de la ITDF.

40 Existen diferentes maneras de 'rastrear' el movimiento del centro de la cabeza. En una implementación, esto se hace sobre la base de un modelo para el movimiento de la cabeza humana, y sobre un análisis de la variación de la temporización entre chirps consecutivos.

45 El modelo describe los movimientos típicos de la cabeza. En esta implementación, se dan instrucciones al sujeto de sentarse erguido en una silla de oficina giratoria, mantener su torso fijo a la silla, y mover sólo su cabeza en todas las direcciones posibles, mientras se realizan rotaciones lentas alrededor de un eje vertical utilizando las capacidades de rotación ofrecidas por la silla. Esto limita los posibles movimientos de la cabeza y se puede modelizar utilizando un modelo mecánico relativamente simple mostrado de manera esquemática en la Figura 31. El centro de la cabeza (r_c) está a una distancia b de la base del cuello (un punto de rotación), la base del cuello está a una distancia a del centro de rotación de la silla. El modelo a priori del centro de la cabeza es entonces:

$$r_c = \begin{matrix} a \cdot \cos(\theta_1) + b \cdot \cos(\theta_1 + \theta_2) \sin(\varphi + \varphi_0) \\ a \cdot \sin(\theta_1) + b \cdot \sin(\theta_1 + \theta_2) \sin(\varphi + \varphi_0) \\ b \cdot \cos(\phi + \varphi_0). \end{matrix} ,$$

50 El ángulo de cabeceo φ del cuello y los ángulos de guiñada θ_1 y θ_2 , indicados en la Figura 31, son incógnitas, pero se pueden estimar en base a las orientaciones de la cabeza. El ángulo de cabeceo φ del cuello es idéntico al ángulo de

cabecero de la cabeza, hasta un desfase φ_0 (el eje del cuello no es necesariamente paralelo al eje z de la cabeza). Además, θ_1 y θ_2 se pueden estimar ambos a partir del ángulo de guiñada de la cabeza θ . En efecto, ya que se dieron instrucciones a la persona del ensayo de realizar muchos movimientos de cabeza en cada posición de la silla, y de sólo girar la silla muy lentamente, se puede suponer que el ángulo de guiñada correspondiente a la silla (θ_1) es la componente que varía lentamente del ángulo de guiñada del total (θ), mientras que el ángulo de guiñada correspondiente con el cuello es la componente que varía rápido (θ_2).

Para estimar el resto de parámetros del modelo (a , b , φ_0), se puede hacer uso del hecho de que la distancia a la fuente varía durante el movimiento de la cabeza/silla. Estos movimientos a lo largo de la dirección de la fuente de sonido se pueden medir mediante inspección de la temporización entre chirps consecutivos. En efecto, si el centro de la cabeza no se moviera, entonces el tiempo entre chirps sucesivos recibidos sería constante. Pero si la cabeza se mueve, los chirps se retrasarán cuando la cabeza se aleja de la fuente, o, llegarán más pronto cuando la cabeza se acerca a la fuente. Las diferencias en tiempos de llegada de los chirps se pueden traducir fácilmente en diferencias de distancia $\Delta r_{meas}(t)$, mediante multiplicación por la velocidad del sonido.

Principalmente un desplazamiento del centro de la cabeza a lo largo de la dirección de la fuente afectará a la distancia a la fuente, y por tanto la variación de distancia de acuerdo con el modelo $\Delta r_{mod}(t)$ se puede escribir como

$$\Delta r_{mod}(t) = a \cdot \cos(\theta_1(t) - \theta_{fuente}) + b \cdot \cos(\theta_1(t) + \theta_2(t) - \theta_{fuente}) \sin(\varphi(t) + \varphi_0)$$

A continuación, estos parámetros del modelo φ_0 , a y b se estiman utilizando el método de estimación de Gauss-Newton por medio de la minimización de

$$\sum_i [\Delta r_{mod}(t_i) - \Delta r_{meas}(t_i)]^2$$

En la Figura 34 se muestra la variación de distancia (con desfase) durante la medida en función del tiempo. Una curva (originariamente la curva azul) es la distancia estimada $\Delta r_{meas}(t)$ basada en el tiempo medido entre chirps, la otra curva (originariamente la curva roja) es la distancia estimada $\Delta r_{mod}(t)$ obtenida del modelo optimizado. Ambas están relativamente de acuerdo la una con la otra.

En la Figura 33 se muestra la trayectoria de las desviaciones del centro de la cabeza (con respecto al centro 'medio') tal como es obtenida por el modelo. Se observa que (0, 0, 0) corresponde a la posición del centro 'medio'. Como se puede ver, la posición del centro verdadero de la cabeza es, en efecto, no constante.

La Figura 32 muestra (filas con números impares) instantáneas de un video que se capturó de un sujeto cuando se realizó una medida de HRTF en la silla con el giro libre, yuxtapuestas (filas con números pares) a visualizaciones que muestran la orientación y la posición estimadas de la cabeza. La línea negra muestra la desviación del centro de la cabeza.

Estimando la característica de transferencia desconocida del altavoz y/o de los micrófonos

Las características de transferencia exactas del altavoz y de los micrófonos no son conocidas, y tampoco lo son las características espectrales del sistema de producción de sonido. Para compensar esta característica de transferencia desconocida, la energía de la información espectral se ajusta en una base por frecuencia, de modo que la energía en cada frecuencia es substancialmente igual a la de una HRTF general (la media de una base de datos de HRTFs que se han medido bajo circunstancias controladas, como la base de datos CIPIC).

Estimando la HRTF y la ITDF sobre toda la esfera

Los pasos precedentes conducen a una versión muestreada de la HRTF y de la ITDF. Pero debido a los movimientos incontrolados, irregulares, de la cabeza, algunas áreas se muestrearán de forma más densa que otras, mientras que otras no se muestrearán en absoluto, debido al rango limitado de movimientos realistas de la cabeza. Obsérvese que, hasta ahora, sólo se utilizó la representación SH para valorar la suavidad de la HRTF y de la ITDF. Por lo tanto, la representación SH sólo se evaluó en los mismos puntos de datos que se utilizaron para 'construir' la representación SH y por tanto la representación SH nunca se evaluó en áreas que no fueron muestreadas.

Sin embargo, para permitir la estimación de la HRTF y la ITD sobre toda la esfera, lo cual es necesario para que un sistema de reproducción de audio cree la ilusión de sonido procedente de cualquier dirección, se aplica una interpolación basada en armónicos esféricos SH reales. Para interpolar la HRTF se considera un orden de truncamiento limitado de la base SH ($l \leq 15$) e ITD ($l \leq 5$), ya que esto captura suficiente detalle espectral. Sin embargo, debido al número limitado de muestras direccionales y al hecho de que algunas partes de la esfera no se han muestreado en absoluto, podrían aparecer problemas de regularización.

5 Para tratar de solucionar estos problemas de regularización cuando se estiman los coeficientes SH, se aplica regularización de Tikhonov cómo se describe en Zotkin et al. De nuevo son posibles diferentes criterios, pero en esta implementación, la norma del vector de coeficientes, que consiste en coeficientes con orden $l > 2$, se minimiza (además de la suma de residuos al cuadrado). De esta manera, se 'fuerza' a la solución a utilizar todo lo posible las funciones base SH de bajo orden que varían lentamente, garantizando que los valores de la HRTF no crecen demasiado en áreas que no han sido muestreadas.

Evaluación de la HRTF

10 La HRTF obtenida utilizando la implementación actual se ha comparado con la HRTF medida en una instalación profesional, de última generación (la cámara anecoica de la Universidad de Aquisgrán). Ambos métodos producen claramente HRTFs similares, véanse la Figura 35, la Figura 35(b) y la Figura 35(d) que se midieron en Aquisgrán, habiéndose determinado la Figura 35(c) y la Figura 35(e) con el método de la presente invención, por supuesto para el mismo sujeto.

Referencias

15 D. Zotkin, R. Duraiswami, N. Gumerov, "Regularized HRTF fitting using spherical harmonics", Applications of signal processing to audio and acoustics, (*WASPAA 2009 IEEE Workshop on*, pp. 257-260, 2009.

REIVINDICACIONES

1. Un método (1000, 1100) de estimar una función de transferencia relacionada con la cabeza (HRTF) individualizada y una función de diferencia de tiempo interaural (ITDF) individualizada de una persona particular en un dispositivo informático, comprendiendo el método los pasos de:
- 5 a) obtener (1011; 1111) o recuperar una pluralidad de conjuntos de datos (L_i, R_i, O_i),
comprendiendo cada conjunto de datos una muestra de audio izquierda (L_i) que se origina en un micrófono de oído izquierdo y una muestra de audio derecha (R_i) que se origina en un micrófono de oído derecho e información de orientación (O_i) que se origina en una unidad de orientación (604; 704; 804),
siendo capturadas la muestra de audio izquierda y la muestra de audio derecha y la información de orientación de cada conjunto de datos de manera substancialmente simultánea en una disposición en la cual:
- 10 el micrófono de oído izquierdo está insertado en un oído izquierdo de la persona, y
el micrófono de oído derecho está insertado en un oído derecho de la persona, y
la persona está ubicada a una distancia de un altavoz (602; 702; 802), y
la unidad de orientación está montada de forma fija a la cabeza de la persona, y
- 15 el altavoz está dispuesto para reproducir una señal de ensayo acústica que comprende una pluralidad de fragmentos de ensayo de audio (C_i), y
la persona mueve su cabeza en una pluralidad de orientaciones diferentes durante la reproducción de la señal de ensayo acústica;
- 20 b) extraer o calcular una pluralidad de valores de diferencia de tiempo interaural (ITDi) y/o una pluralidad de valores espectrales (SLi_f, SRi_f), y correspondientes valores de orientación de la unidad de orientación a partir de los conjuntos de datos (L_i, R_i, O_i);
- c) estimar (1002; 1102) una dirección del altavoz con respecto a una posición media del centro de la cabeza de la persona y expresada en el marco de referencia del mundo, comprendiendo los pasos de:
- 1) suponer una dirección candidata de la fuente (θ, ϕ);
- 25 2) asignar una dirección (θ_i, ϕ_i) a cada miembro de al menos un subconjunto de la pluralidad de valores de diferencia de tiempo interaural (ITDi) y/o a cada miembro de al menos un subconjunto entre la pluralidad de valores espectrales (SLi_f, SRi_f), correspondiente con la dirección de la fuente supuesta expresada en un marco de referencia de la unidad de orientación, obteniendo de este modo un conjunto de datos mapeados (ITDi(θ_i, ϕ_i); Si(f)(θ_i, ϕ_i));
- 30 3) calcular un valor de calidad del conjunto de datos mapeados (ITDi(θ_i, ϕ_i); Si(f)(θ_i, ϕ_i)) en base a un criterio de calidad predefinido;
- 4) repetir los pasos 1) a 3) al menos una vez para una segunda dirección candidata y/o más direcciones candidatas de la fuente diferentes a las direcciones candidatas anteriores de la fuente;
- 5) elegir la dirección candidata de la fuente que de como resultado el mayor valor de calidad como la dirección del altavoz con respecto a la posición media del centro de la cabeza de la persona;
- 35 d) estimar (1003; 1103) una orientación de la unidad de orientación con respecto a la cabeza;
- e) estimar (1005; 1105) la ITDF individualizada y la HRTF individualizada de la persona, en base a la pluralidad de conjuntos de datos (L_i, R_i, O_i) y en base a la dirección estimada del altavoz con respecto a la posición media del centro de la cabeza estimada en el paso c) y en base a la orientación estimada de la unidad de orientación con respecto a la cabeza estimada en el paso d);
- 40 en donde los pasos a) a e) son realizados por al menos un dispositivo informático.
2. El método de la reivindicación 1, en el cual el paso b) comprende:
- localizar una pluralidad de fragmentos de audio izquierdos (LFi) y/o fragmentos de audio derechos (RFi) en la pluralidad de conjuntos de datos (L_i, R_i, O_i), correspondiendo cada fragmento de audio izquierdo y derecho con un fragmento de ensayo de audio (C_i) reproducido por el altavoz;
- 45 - calcular un valor de diferencia de tiempo interaural (ITDi) para al menos un subconjunto de las parejas de fragmentos de audio izquierdo y derecho correspondientes (LFi, RFi);

- estimar una orientación momentánea de la unidad de orientación (604; 704; 804) para cada pareja de fragmentos de audio izquierdo y derecho correspondientes (L_{Fi}, R_{Fi});

3. El método de la reivindicación 1 o 2, en el cual el paso b) comprende:

5 - localizar una pluralidad de fragmentos de audio izquierdos (L_{Fi}) y/o fragmentos de audio derechos (R_{Fi}) en la pluralidad de conjuntos de datos (L_i, R_i, O_i), correspondiendo cada fragmento de audio izquierdo y derecho con un fragmento de ensayo de audio (C_i) reproducido por el altavoz;

10 - calcular un conjunto de valores espectrales izquierdos (S_{Li_f}) para cada fragmento de audio izquierdo (L_{Fi}) y/o calcular un conjunto de valores espectrales derechos (S_{Ri_f}) para cada fragmento de audio derecho (R_{Fi}), conteniendo cada conjunto de valores espectrales (S_{Li_f}; S_{Ri_f}) al menos un valor espectral correspondiente a una frecuencia espectral;

- estimar una orientación momentánea de la unidad de orientación (604; 704; 804) para al menos un subconjunto de los fragmentos de audio izquierdos (L_{Fi}) y/o de los fragmentos de audio derechos (R_{Fi}).

4. El método de acuerdo con cualquiera de las reivindicaciones anteriores, en el cual el criterio de calidad predefinido es un criterio de suavidad espacial de los datos mapeados, o

15 - basado en una desviación o distancia entre los datos mapeados y una superficie de referencia, donde la superficie de referencia se calcula como una variante paso bajo de dichos datos mapeados, o

- basado en una desviación o distancia entre los datos mapeados y una superficie de referencia, donde la superficie de referencia está basada en una aproximación de los datos mapeados, definida por la suma ponderada de un número limitado de funciones base, por ejemplo funciones armónicas esféricas, o

20 - que expresa un grado de la anti-simetría especular de los datos ITD_i mapeados, o

- que expresa un grado de simetría cilíndrica de los datos ITD_i mapeados.

5. El método de acuerdo con cualquiera de las reivindicaciones anteriores, que comprende además:

25 f) estimar (1503) parámetros del modelo de un modelo mecánico relacionados con los movimientos de la cabeza que fueron realizados por la persona en el momento de capturar las muestras de audio y de la información de orientación del paso a);

g) estimar (1504) una pluralidad de posiciones de la cabeza utilizando el modelo mecánico y los parámetros del modelo estimados;

y en donde el paso c) comprende utilizar las posiciones estimadas de la cabeza del paso g).

6. El método de acuerdo con cualquiera de las reivindicaciones anteriores,

30 en el cual el paso b) comprende:

- estimar una trayectoria de los movimientos de la cabeza a lo largo de una pluralidad de fragmentos de audio;

- tener en cuenta la trayectoria estimada cuando se estime la posición de la cabeza y/o la orientación de la cabeza,

35 y/o en el cual el paso e) comprende además estimar una característica de filtro combinada del altavoz y de los micrófonos, o comprende ajustar la ITDF estimada de tal manera que la energía por banda de frecuencia corresponde a la de una ITDF general, y comprende ajustar la HRTF estimada de tal manera que la energía por banda de frecuencia corresponde a la de una HRTF general y en el cual la estimación de la característica de filtro espectral combinada del altavoz y los micrófonos comprende:

utilizar información a priori acerca de una característica de filtro espectral del altavoz, y/o

utilizar información a priori acerca de una característica de filtro espectral de los micrófonos.

40 7. El método de acuerdo con cualquiera de las reivindicaciones anteriores,

en el cual el paso d) estima la orientación de la unidad de orientación teniendo también en cuenta información espacial extraída de las muestras de audio Izquierda y Derecha, utilizando al menos una función de transferencia (ITDF, HRTF) que relaciona indicios acústicos con información espacial,

45 en el cual opcionalmente la al menos una función de transferencia predefinida que relaciona indicios acústicos con información espacial es una función de diferencia de tiempo interaural (ITDF) predefinida, o

- 5 en el cual opcionalmente la al menos una función de transferencia que relaciona indicios acústicos con información espacial son dos funciones de transferencia que incluyen una función de diferencia de tiempo interaural (ITDF) predefinida y una función de transferencia relacionada con la cabeza (HRTF) predefinida; o en el cual el método comprende realizar los pasos b) a e) al menos dos veces, en donde el paso b) de la primera iteración no tiene en cuenta dicha información espacial, y en donde el paso b) de la segunda iteración y de cualquier iteración adicional tiene en cuenta dicha información espacial, utilizando la función de diferencia de tiempo interaural (ITDF) y/o la función de transferencia relacionada con la cabeza (HRTF) estimadas en el paso e) de la primera iteración o de iteraciones adicionales.
- 10 8. El método de acuerdo con cualquiera de las reivindicaciones anteriores, en el cual el paso d) de estimar la función ITDF comprende utilizar información a priori acerca de la ITDF personalizada basada en análisis estadístico de una base de datos que contiene una pluralidad de ITDFs de personas diferentes.
9. El método de acuerdo con cualquiera de las reivindicaciones anteriores, en el cual el paso e) de estimar la HRTF comprende utilizar información a priori acerca de la HRTF personalizada basada en análisis estadístico de una base de datos que contiene una pluralidad de HRTFs de personas diferentes.
- 15 10. El método de acuerdo con cualquiera de las reivindicaciones anteriores, en el cual la unidad de orientación comprende al menos un sensor de orientación adaptado para proporcionar información de orientación relativa al campo gravitatorio de la tierra y al menos un sensor de orientación adaptado para proporcionar Información de orientación relativa al campo magnético de la tierra y/o
- en el cual el método comprende montar de forma fija la unidad de orientación a la cabeza de la persona y/o
- 20 en el cual la unidad de orientación está comprendida en un dispositivo portátil, y en el cual el método comprende además el paso de montar de forma fija el dispositivo portátil que comprende la unidad de orientación a la cabeza de la persona.
11. El método de acuerdo con cualquiera de las reivindicaciones anteriores, que comprende además el paso de:
- reproducir la señal de ensayo acústica por medio del altavoz;
- 25 - capturar dichas señales de audio izquierdas y derechas (Li, Ri) que se originan en dicho micrófono de oído izquierdo y en dicho micrófono de oído derecho y capturar dicha información de orientación (Oi) de una unidad de orientación (604; 704; 804)
- y/o
- 30 en donde la unidad de orientación está comprendida en un dispositivo portátil, pudiéndose montar el dispositivo portátil a la cabeza de la persona; y en donde el dispositivo portátil comprende además un procesador programable y una memoria, y medios de interfaz conectados eléctricamente a los micrófonos de oído izquierdo y derecho, y medios para almacenar y/o transmitir dichos conjuntos de datos capturados; y en donde el dispositivo portátil captura la pluralidad de muestras de audio izquierdas (Li) y de muestras de audio derechas (Ri) e información de orientación (Oi), y en el cual el dispositivo portátil almacena los conjuntos de datos capturados en una memoria intercambiable y/o transmite
- 35 los conjuntos de datos capturados al dispositivo informático y en el cual el dispositivo informático lee dicha memoria intercambiable o recibe los conjuntos de datos capturados transmitidos, y realiza los pasos c) a e) mientras lee o recibe o después de leer o recibir los conjuntos de datos capturados,
- o
- 40 en el cual el método comprende además los pasos de insertar el micrófono en el oído izquierdo en el oído izquierdo de la persona e insertar el micrófono de oído derecho en el oído derecho de dicha persona; en el cual el dispositivo informático se conecta eléctricamente a los micrófonos dentro de los oídos izquierdo y derecho, y se conecta operativamente a la unidad de orientación; y en el cual el dispositivo informático captura la pluralidad de muestras de audio izquierdas y de muestras de audio derechas y recupera o recibe o lee u obtiene de otra manera la información de orientación de dicha unidad de orientación;
- 45 y en el cual el dispositivo informático almacena dichos datos en una memoria.
12. El método de la reivindicación 11,
- en el cual el dispositivo portátil comprende además un altavoz;
- y en el cual el dispositivo portátil está además adaptado para analizar la información de orientación a fin de verificar si un espacio 3D alrededor de la cabeza se muestrea suficientemente, de acuerdo con un criterio predefinido;
- 50 y está adaptado además para reproducir un primer y un segundo mensajes de audio predefinidos respectivamente por medio del altavoz del dispositivo portátil dependiendo del resultado del análisis sobre si el espacio 3D está suficientemente muestreado.

13. El método de acuerdo con cualquiera de las reivindicaciones anteriores,
en el cual la señal de ensayo acústica comprende una pluralidad de estímulos acústicos,
en el cual cada uno de los estímulos acústicos tiene una duración en el rango de 25 a 50 ms;
y/o en el cual un periodo de tiempo entre estímulos acústicos consecutivos es un periodo en el rango de 250 a 500 ms
- 5
- y/o en el cual el método comprende además el paso de:
- seleccionar, dependiendo de un análisis de los conjuntos de datos capturados, un mensaje de audio predefinido de un grupo de mensajes de audio predefinidos, y
 - reproducir dicho mensaje de audio seleccionado por medio del mismo altavoz que se utilizó para los estímulos de ensayo o por medio de un segundo altavoz diferente al primer altavoz, para proporcionar información o instrucciones a la persona antes y/o durante y/o después de la reproducción de la señal de ensayo acústica.
- 10
14. Un método de reproducir una señal de audio virtual para una persona particular, que comprende:
- x) estimar una función de transferencia relacionada con la cabeza (HRTF) individualizada y una función de diferencia de tiempo interaural (ITDF) individualizada de dicha persona particular utilizando un método de acuerdo con cualquiera de las reivindicaciones anteriores;
 - y) generar una señal de audio virtual para la persona particular, utilizando la función de transferencia relacionada con la cabeza (HRTF) individualizada y la función de diferencia de tiempo interaural (ITDF) individualizada estimadas en el paso x);
 - z) reproducir la señal de audio virtual generada en el paso y) utilizando unos auriculares estéreo y/o un conjunto de altavoces de oído.
- 15
15. Un producto de programa informático para estimar una función de transferencia relacionada con la cabeza (HRTF) individualizada y una función de diferencia de tiempo interaural (ITDF) de una persona particular, estando programado dicho producto de programa informático, cuando se ejecuta en al menos un dispositivo informático que comprende un procesador programable y una memoria, para realizar un método de acuerdo con cualquiera de las reivindicaciones anteriores.
- 20
- 25

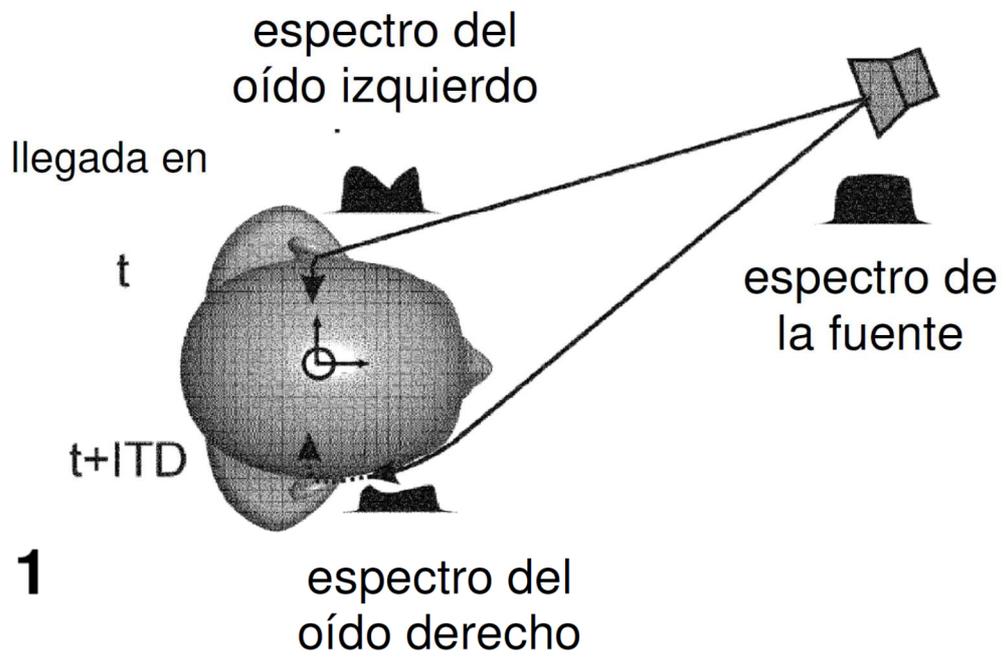


FIG 1

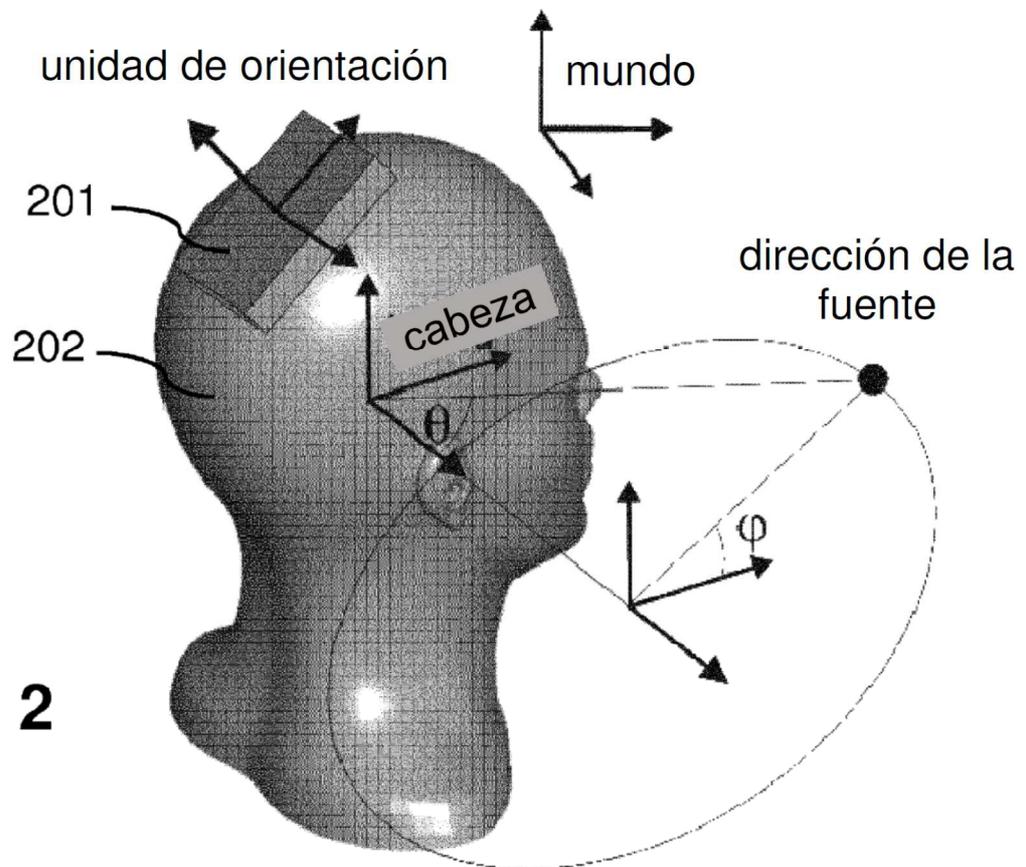


FIG 2

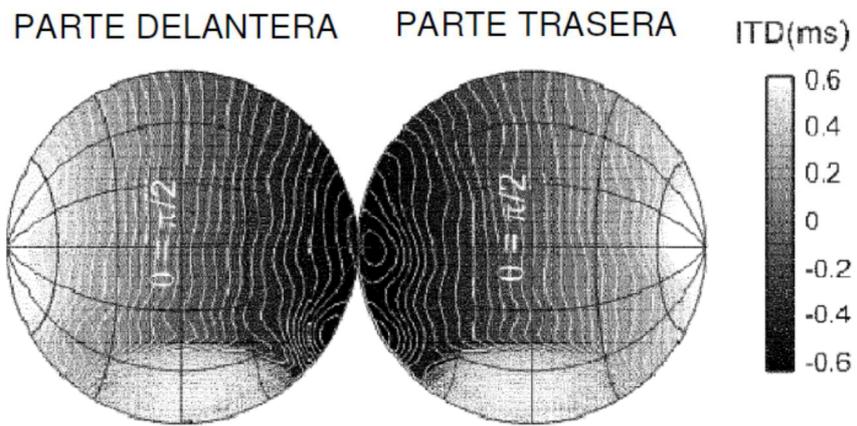


FIG 3

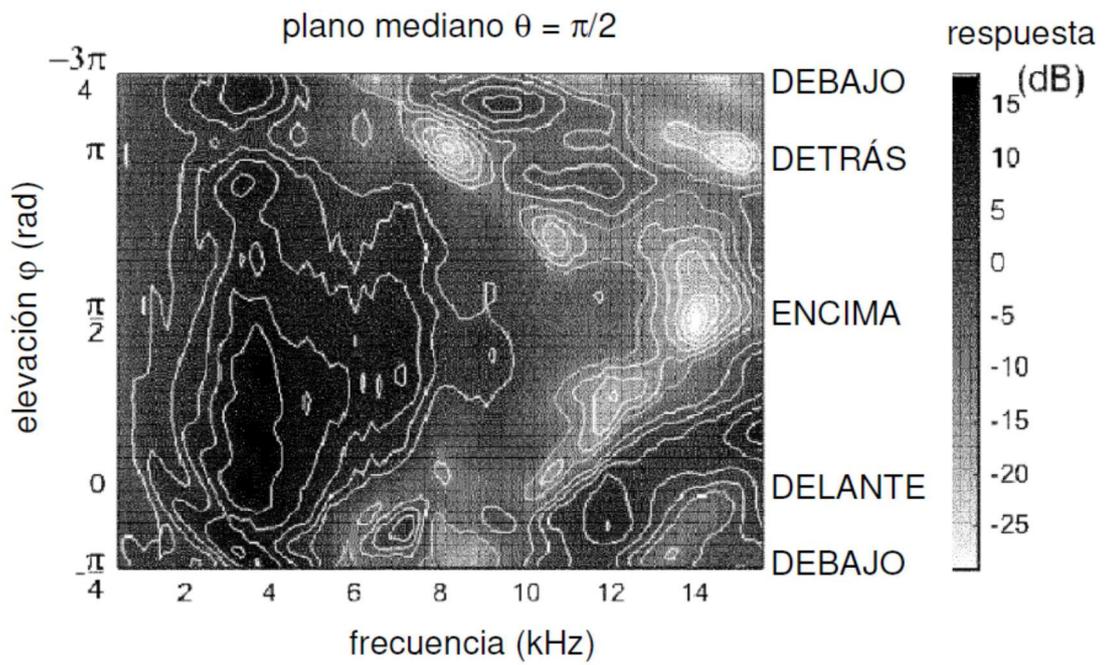


FIG 4

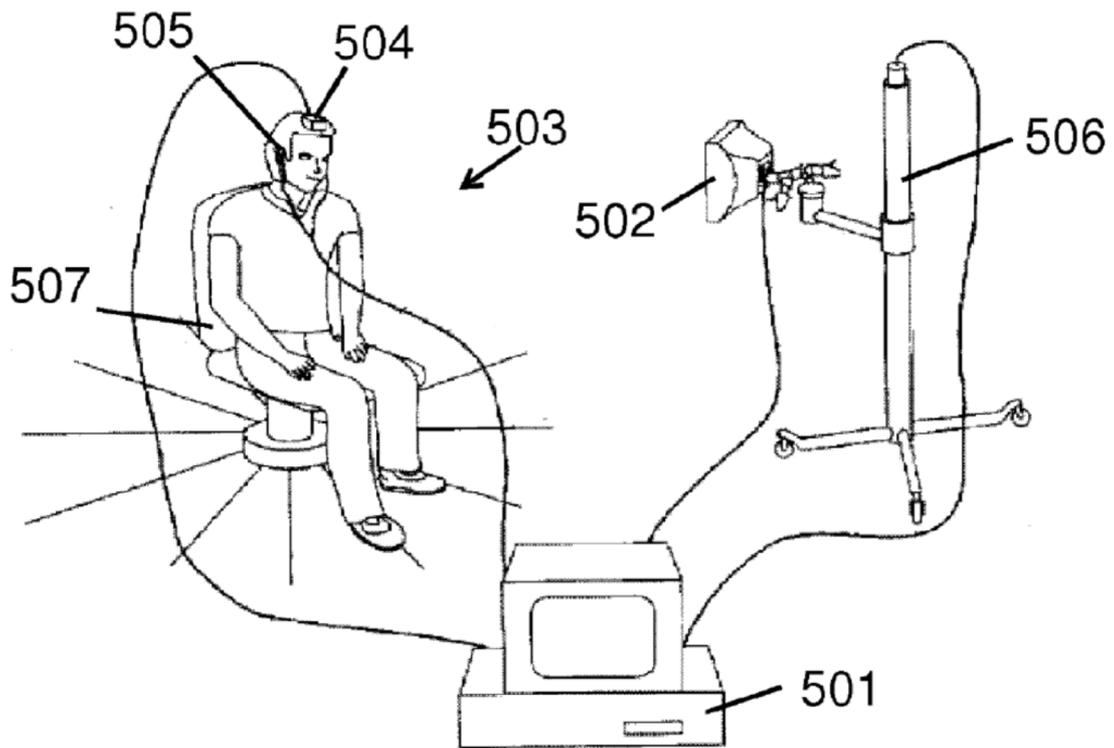
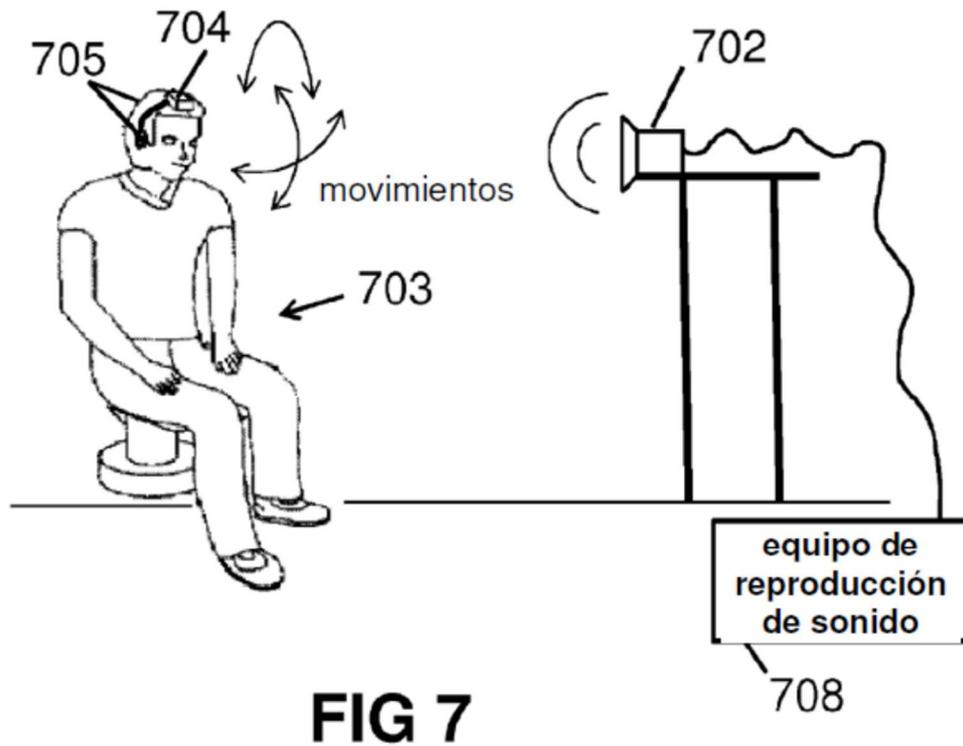
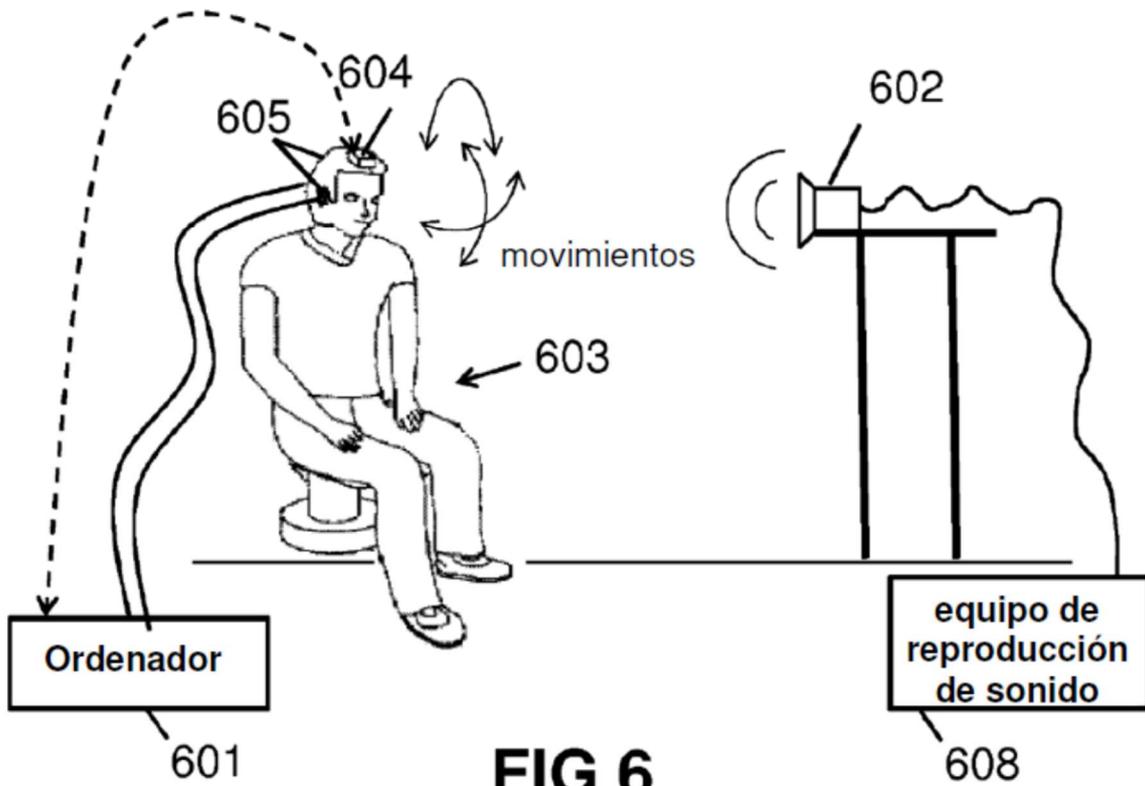
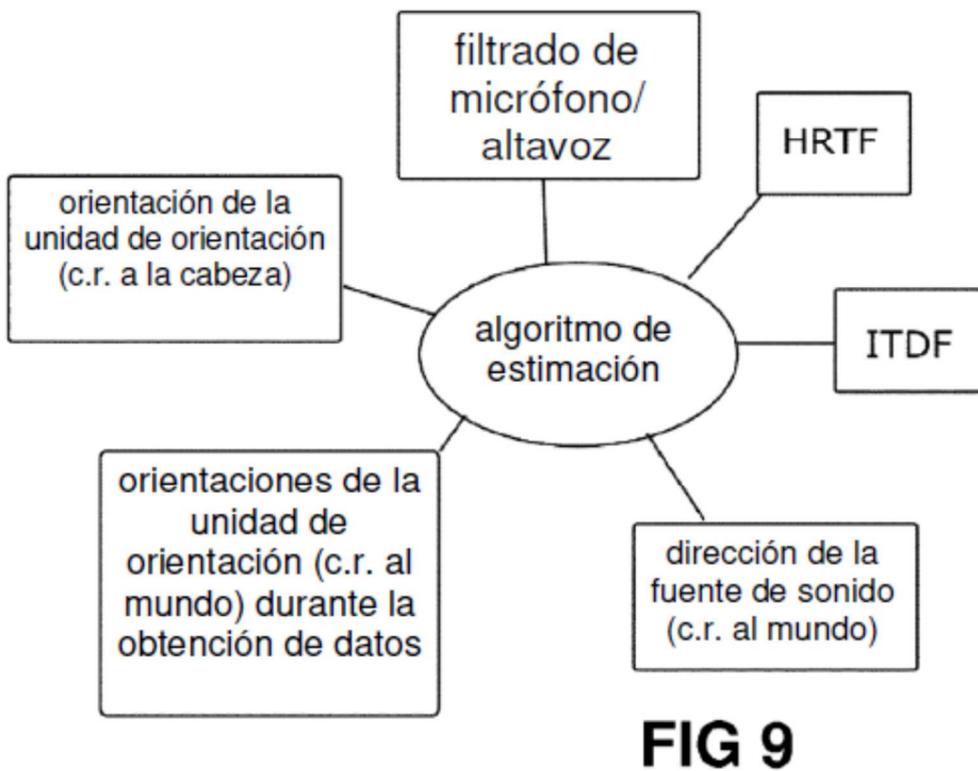
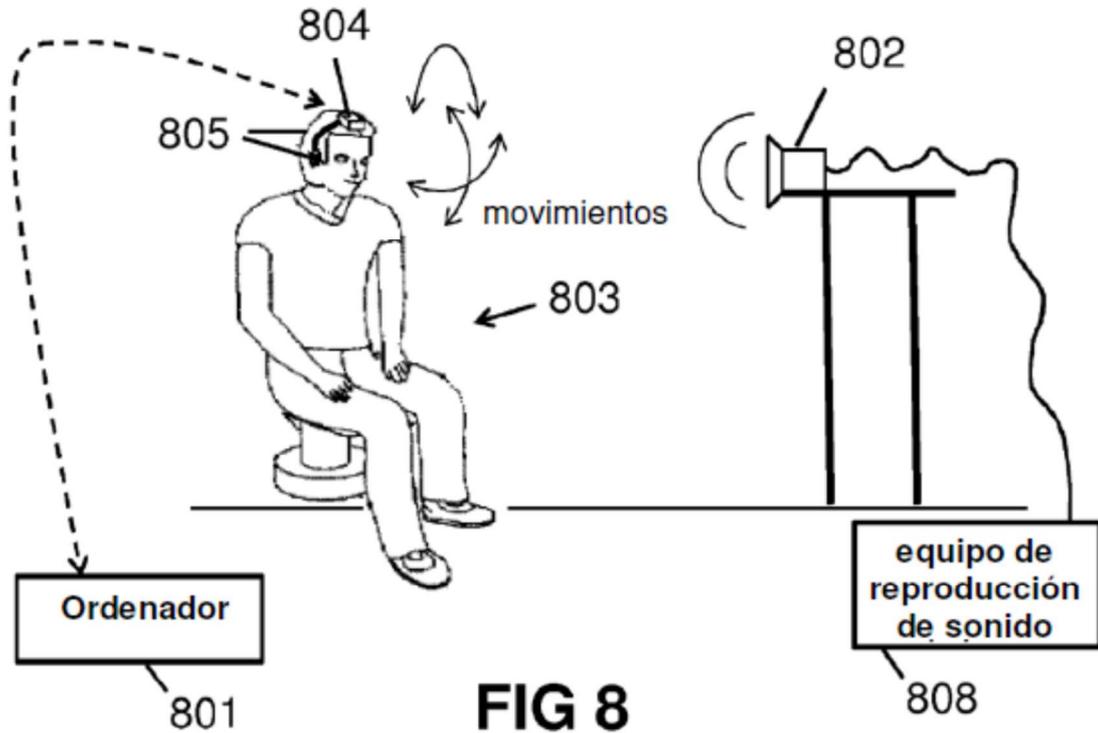
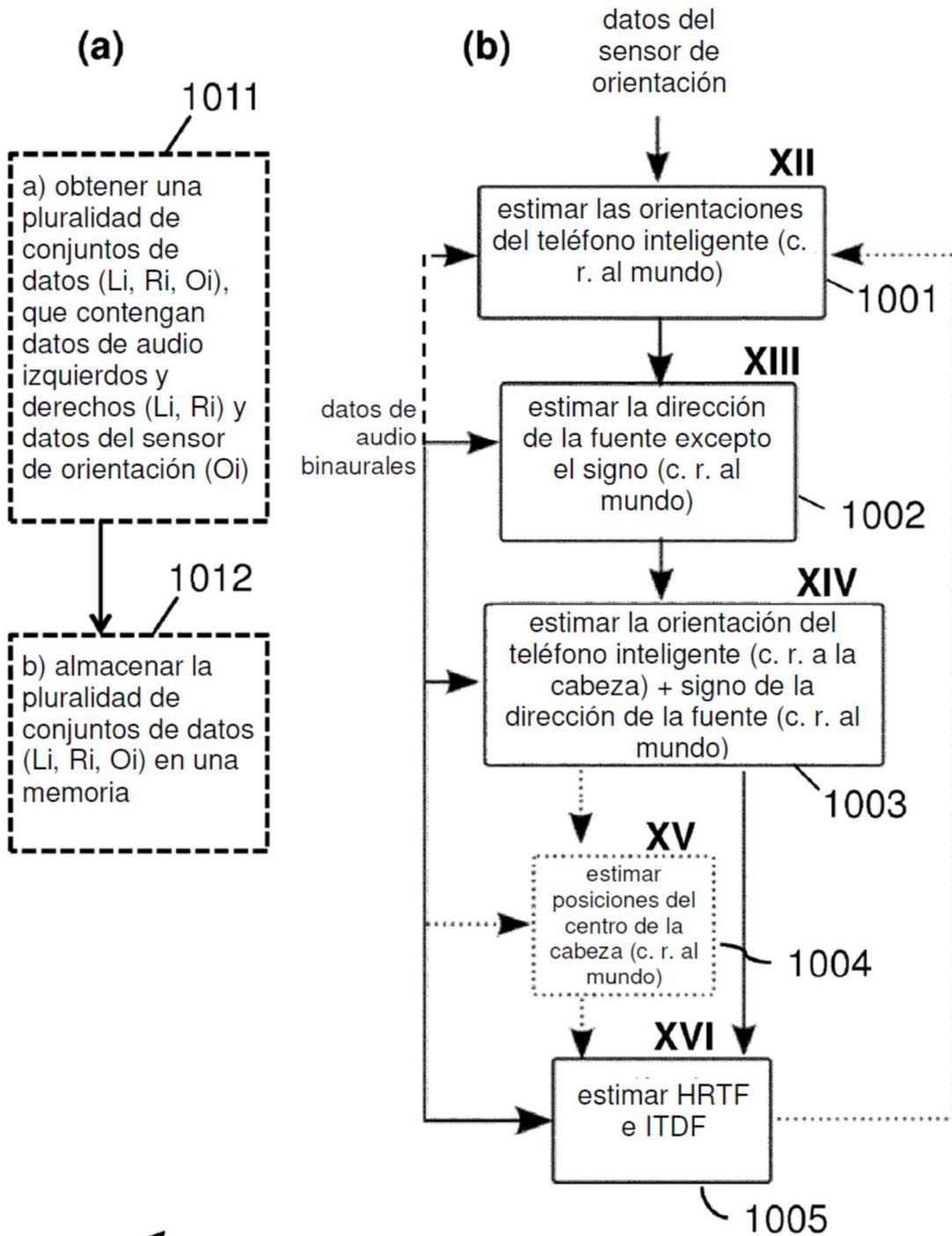


FIG 5 (técnica anterior)







1000 ↗

FIG 10

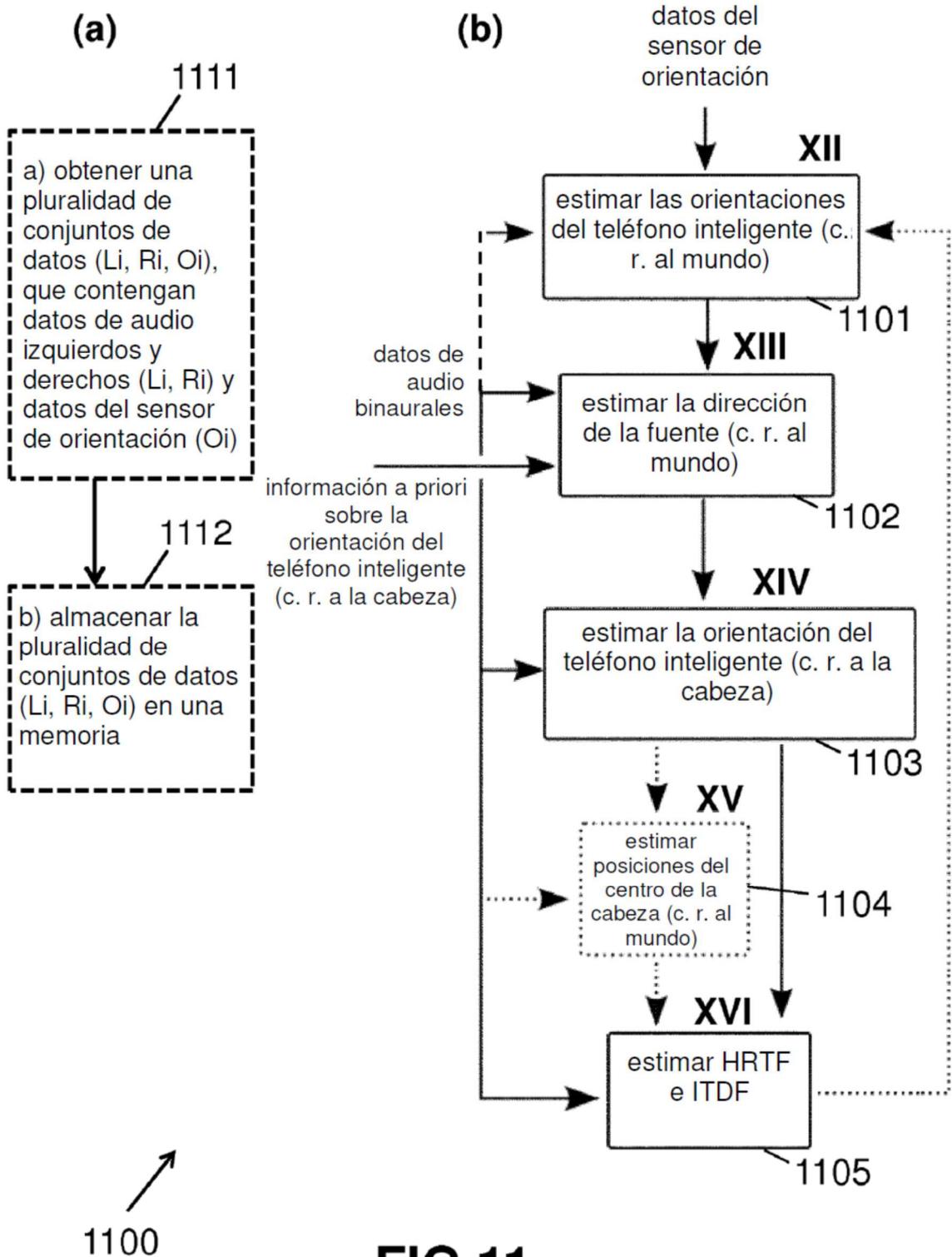
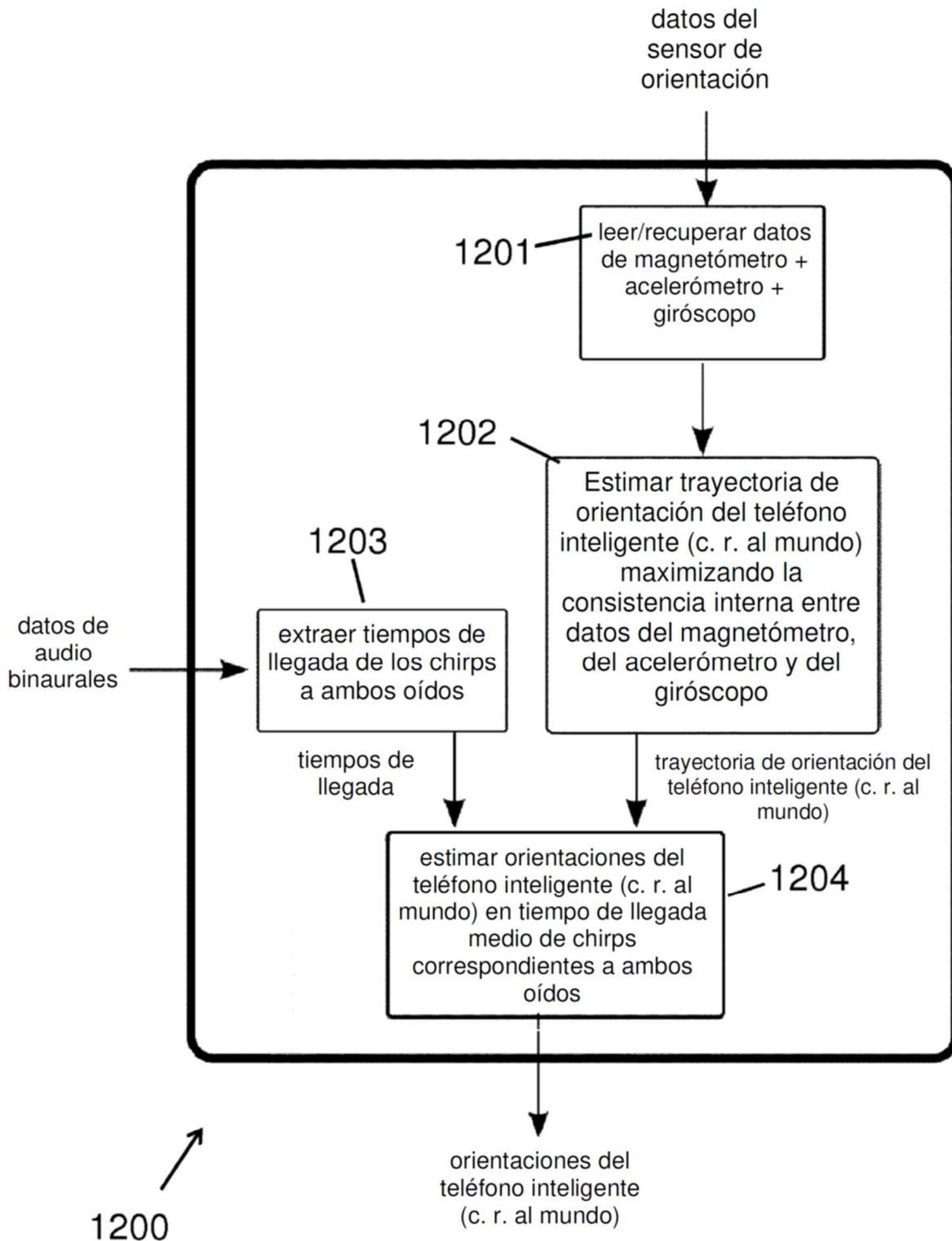
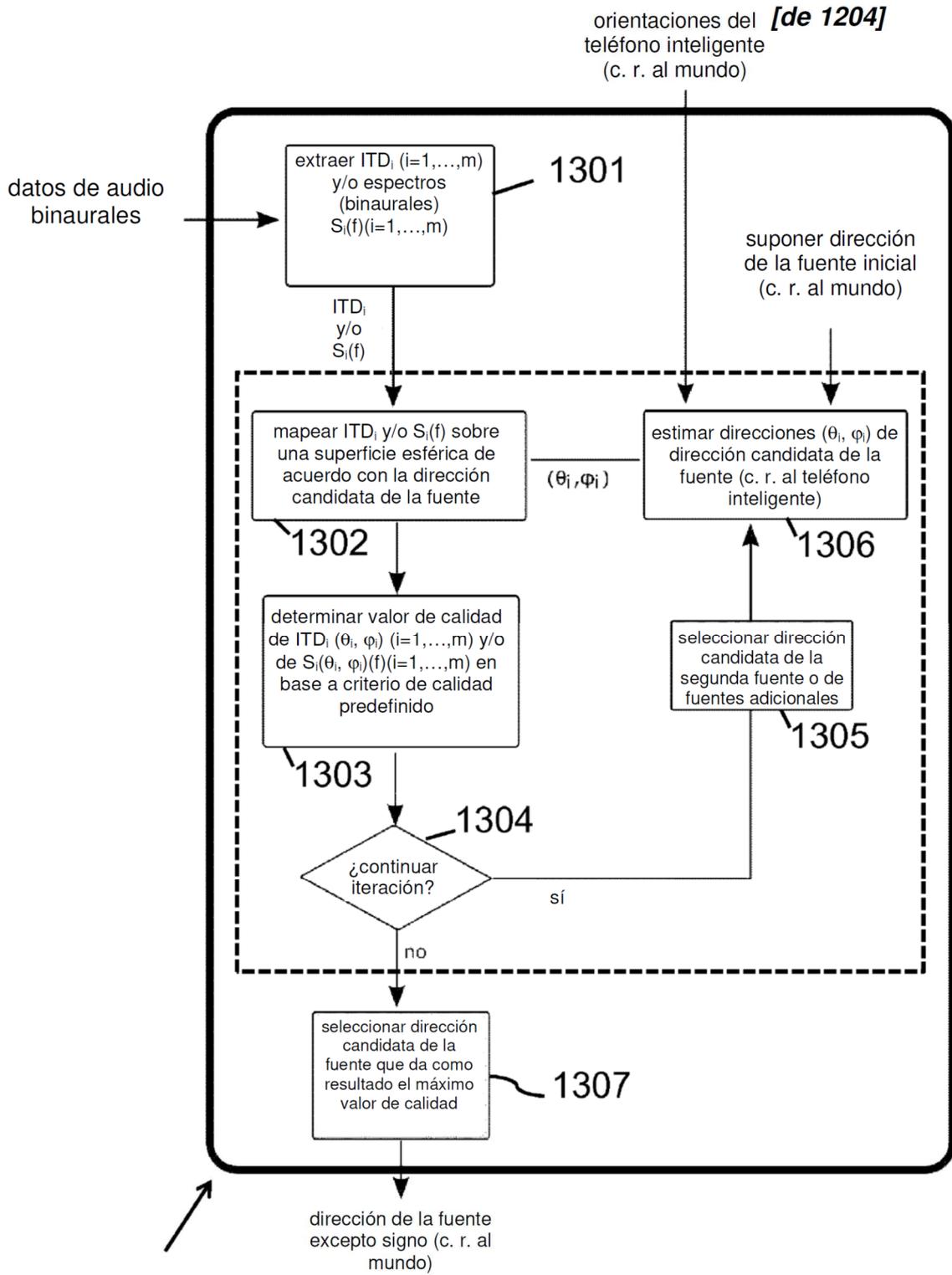


FIG 11



[a 1302, a 1402]

FIG 12



[a 1402]

FIG 13

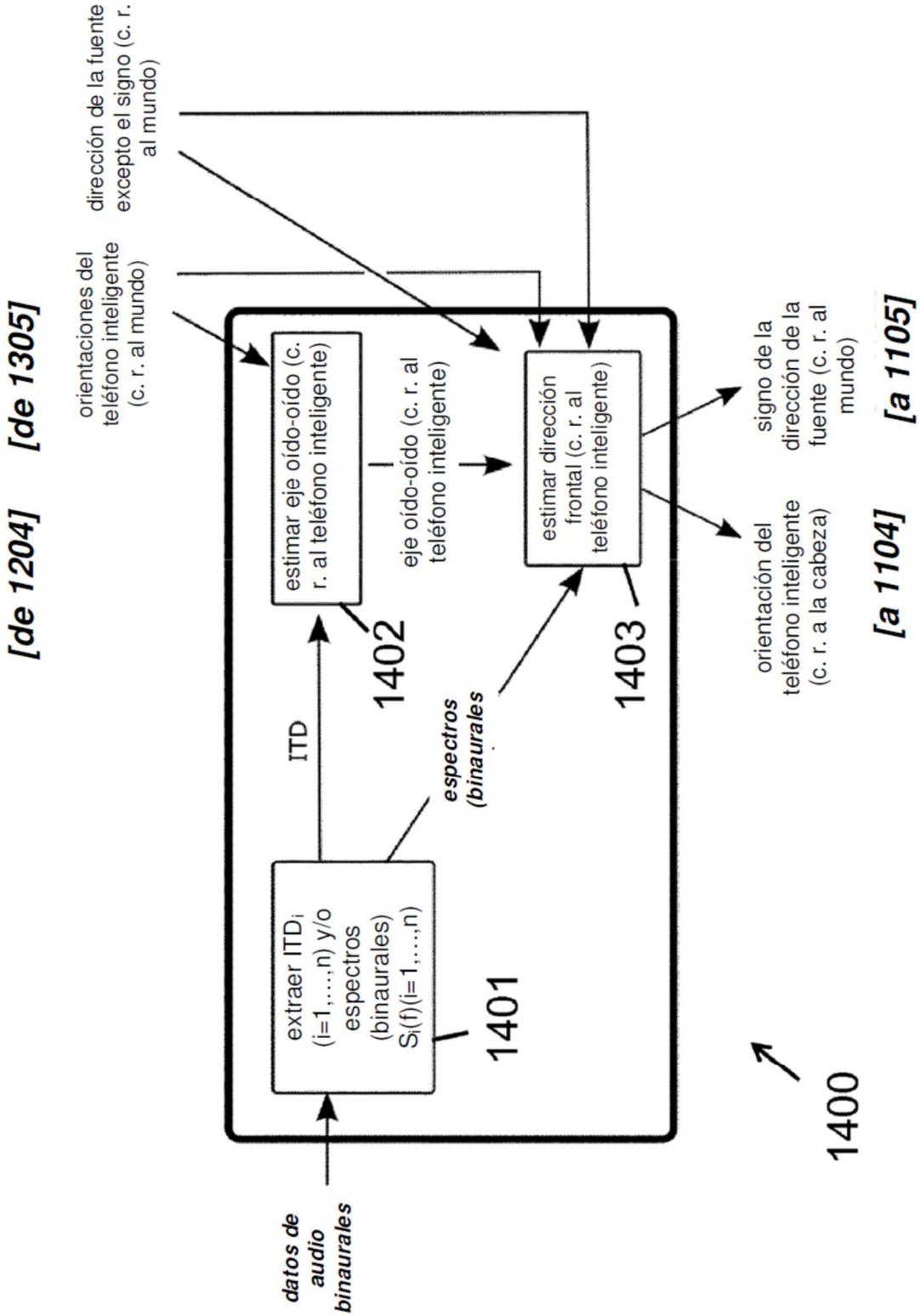


FIG 14

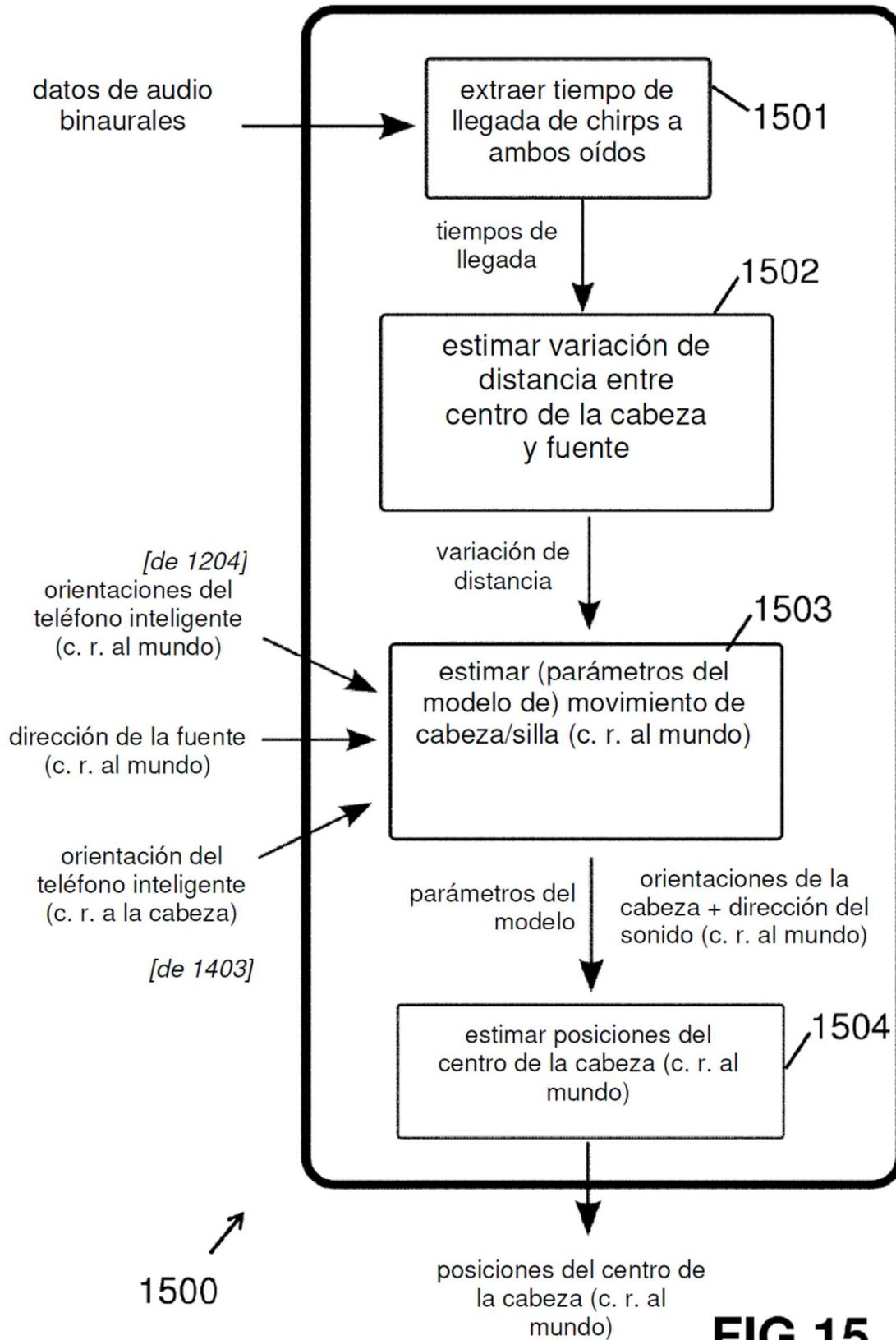


FIG 15

[a 1601]

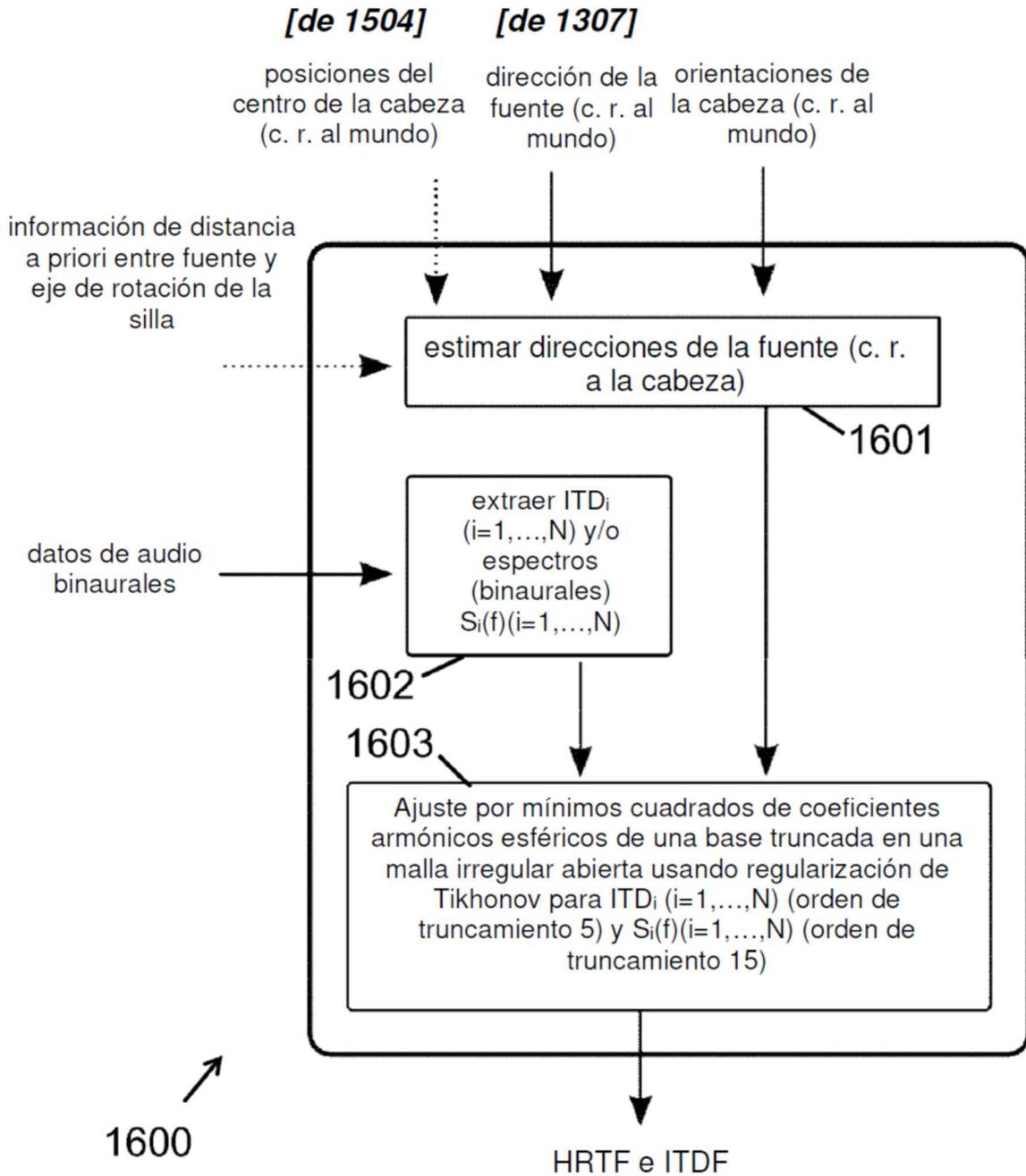


FIG 16

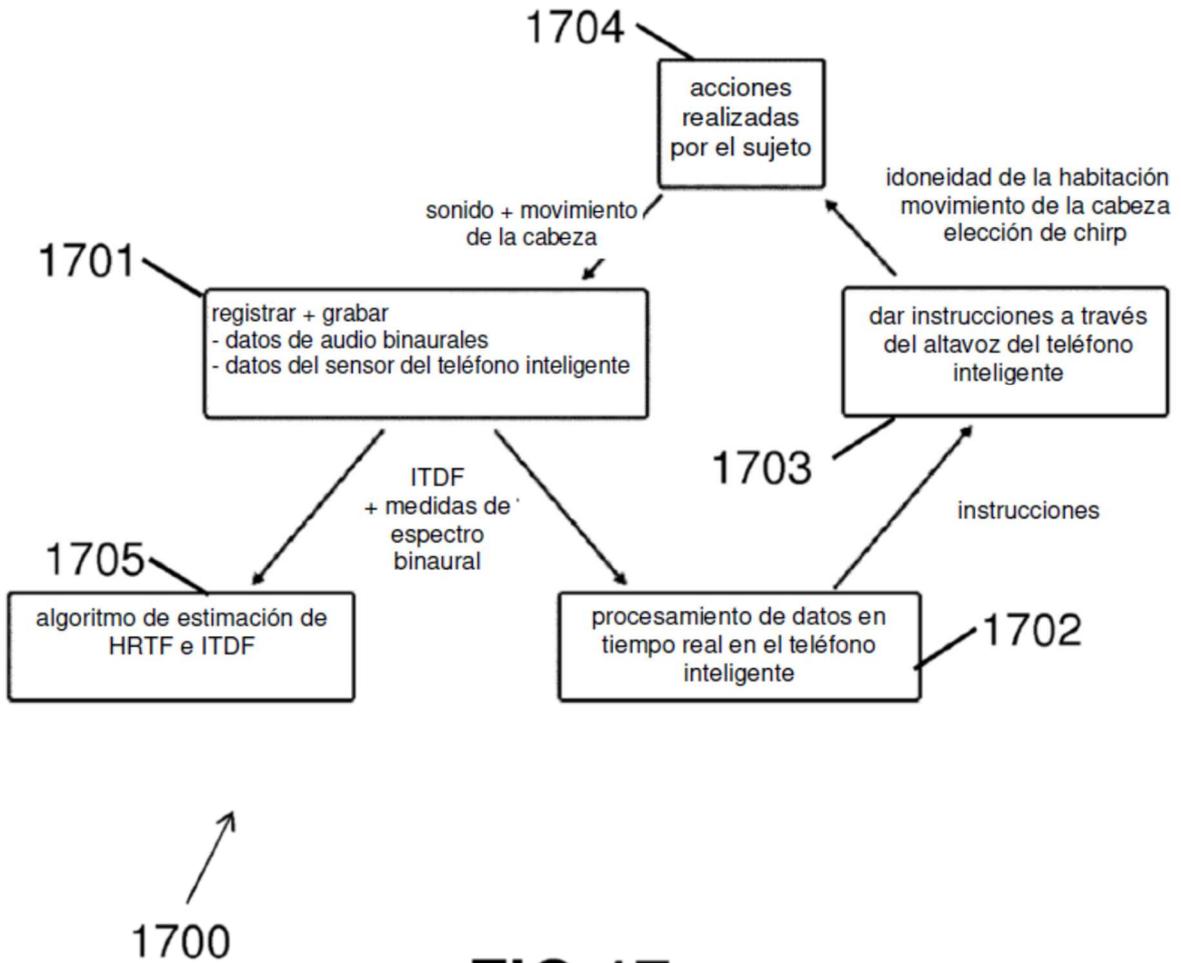


FIG 17

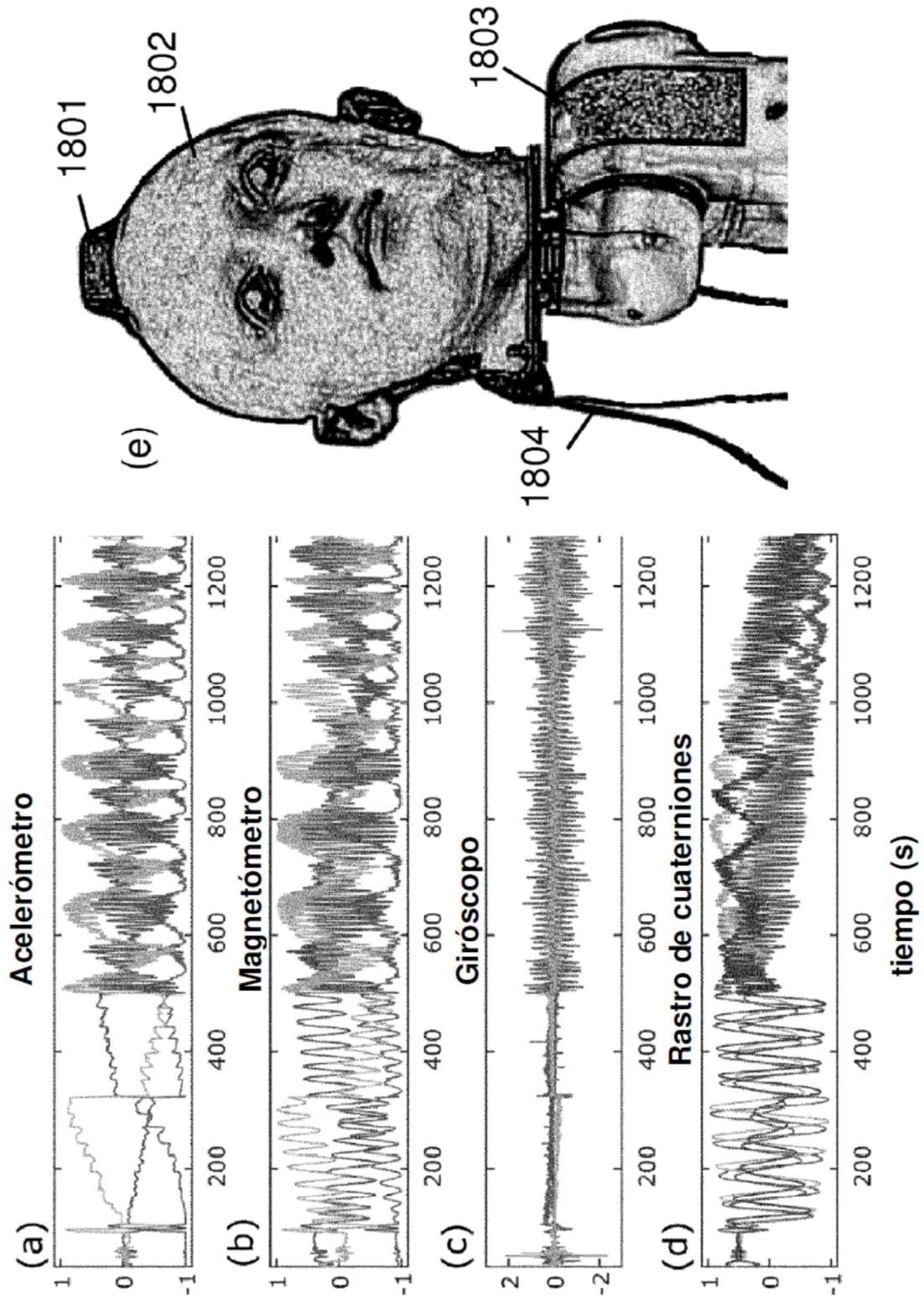
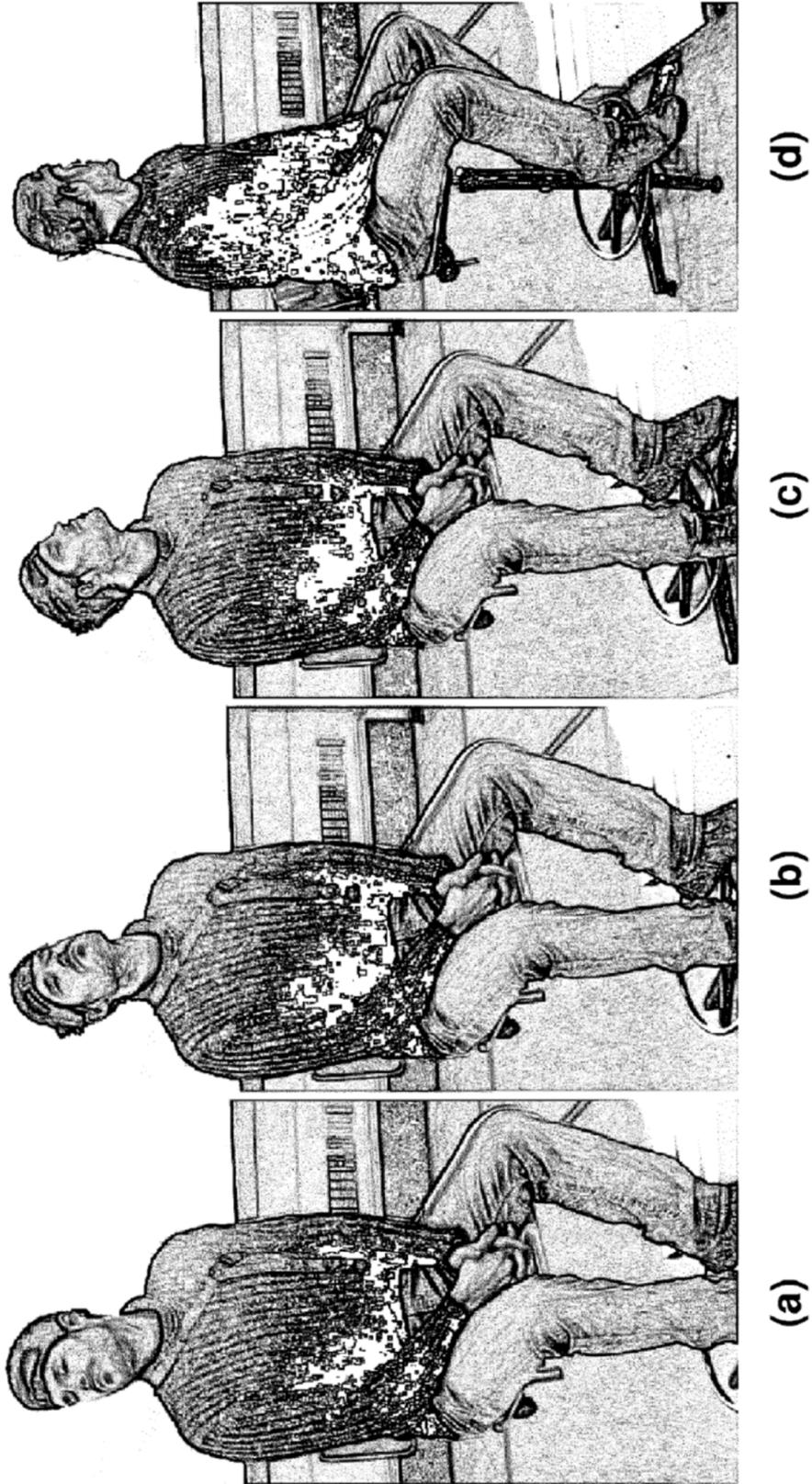


FIG 18



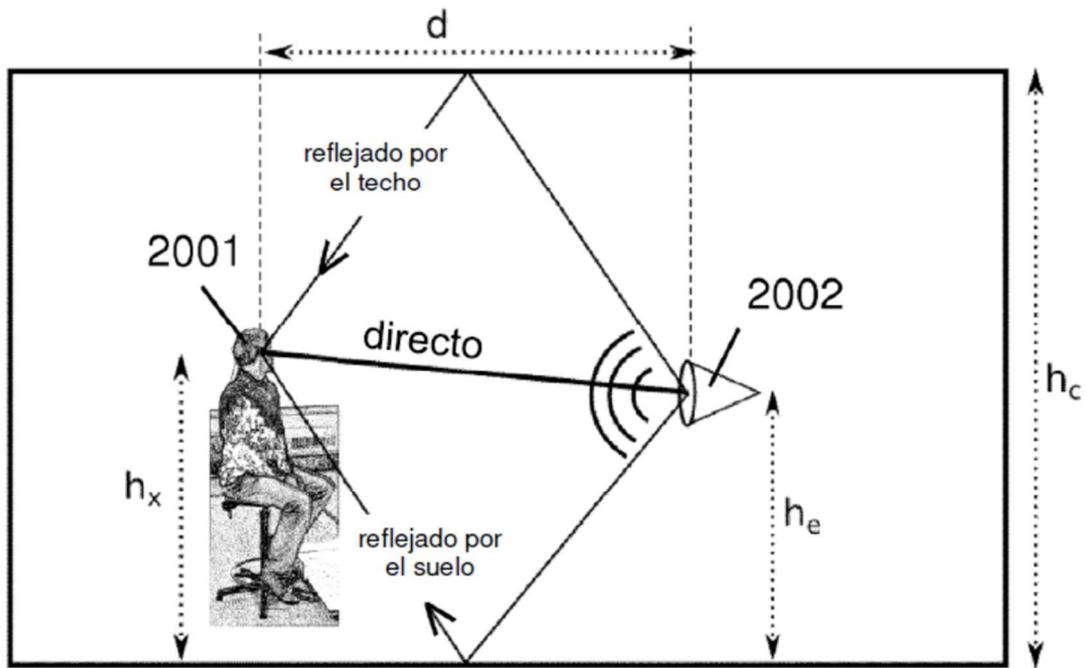


FIG 20

2000

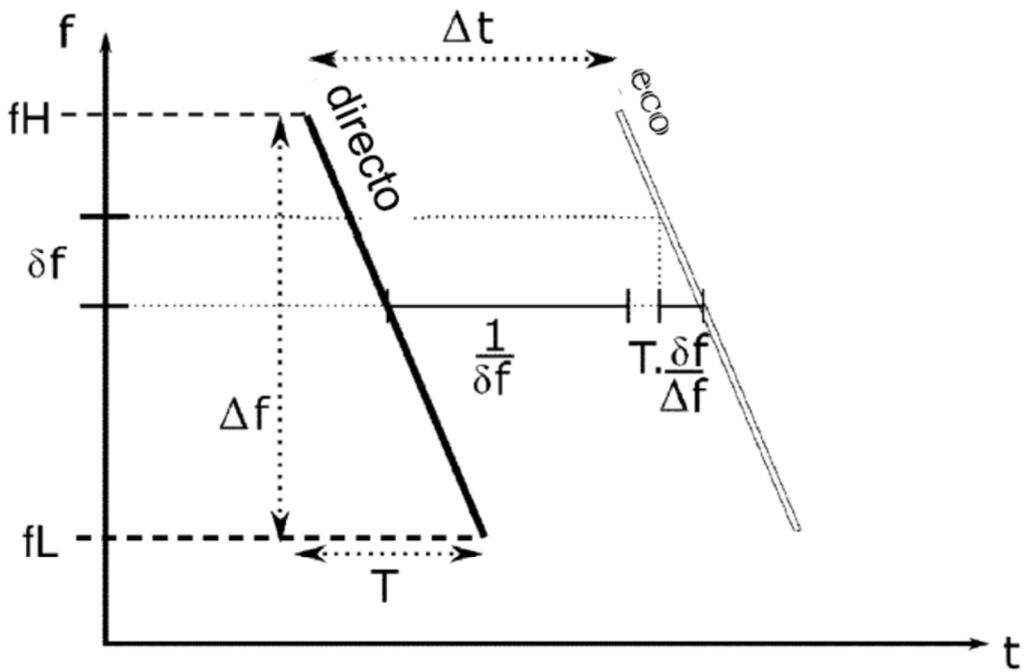


FIG 21

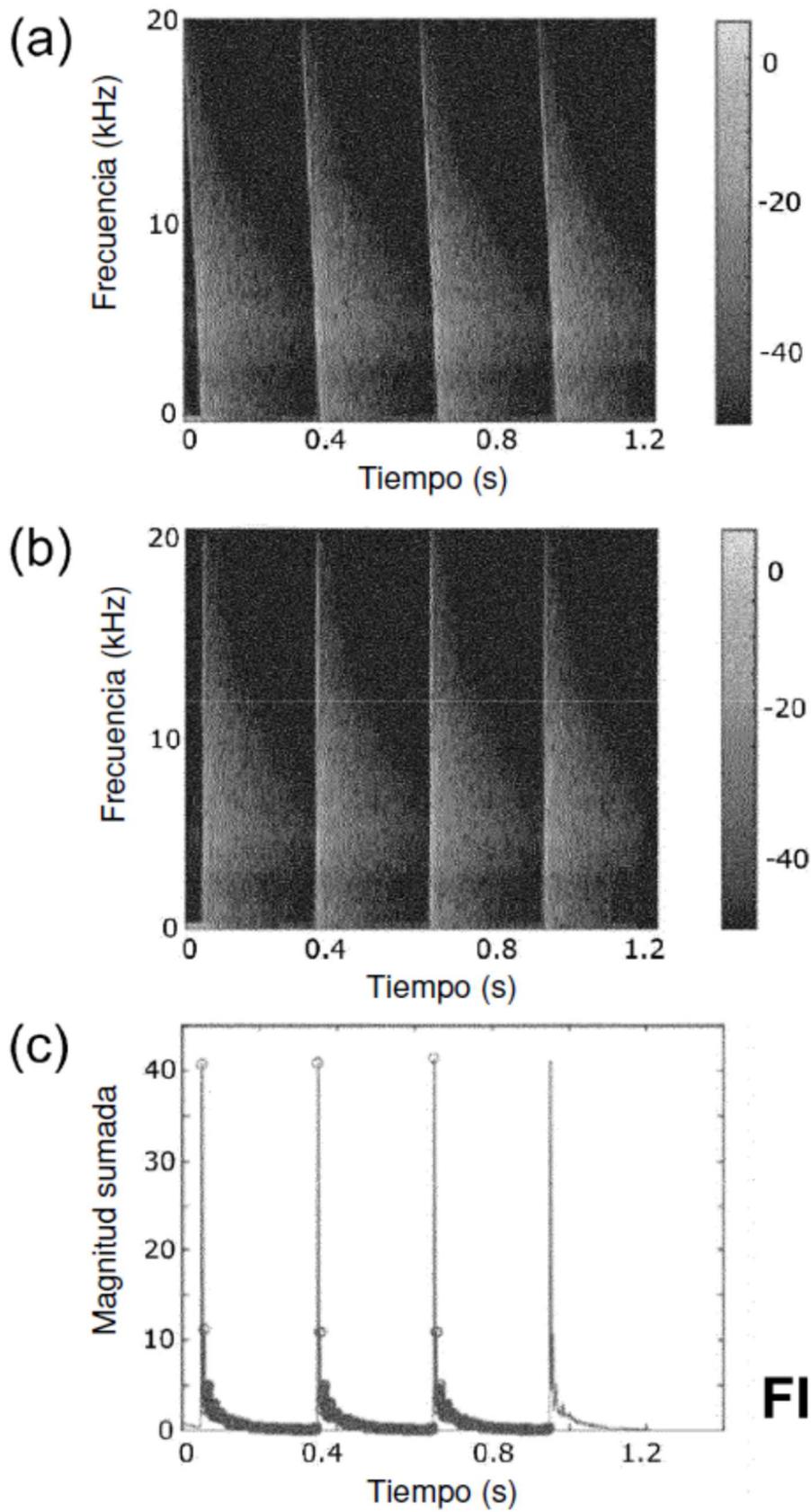


FIG 22

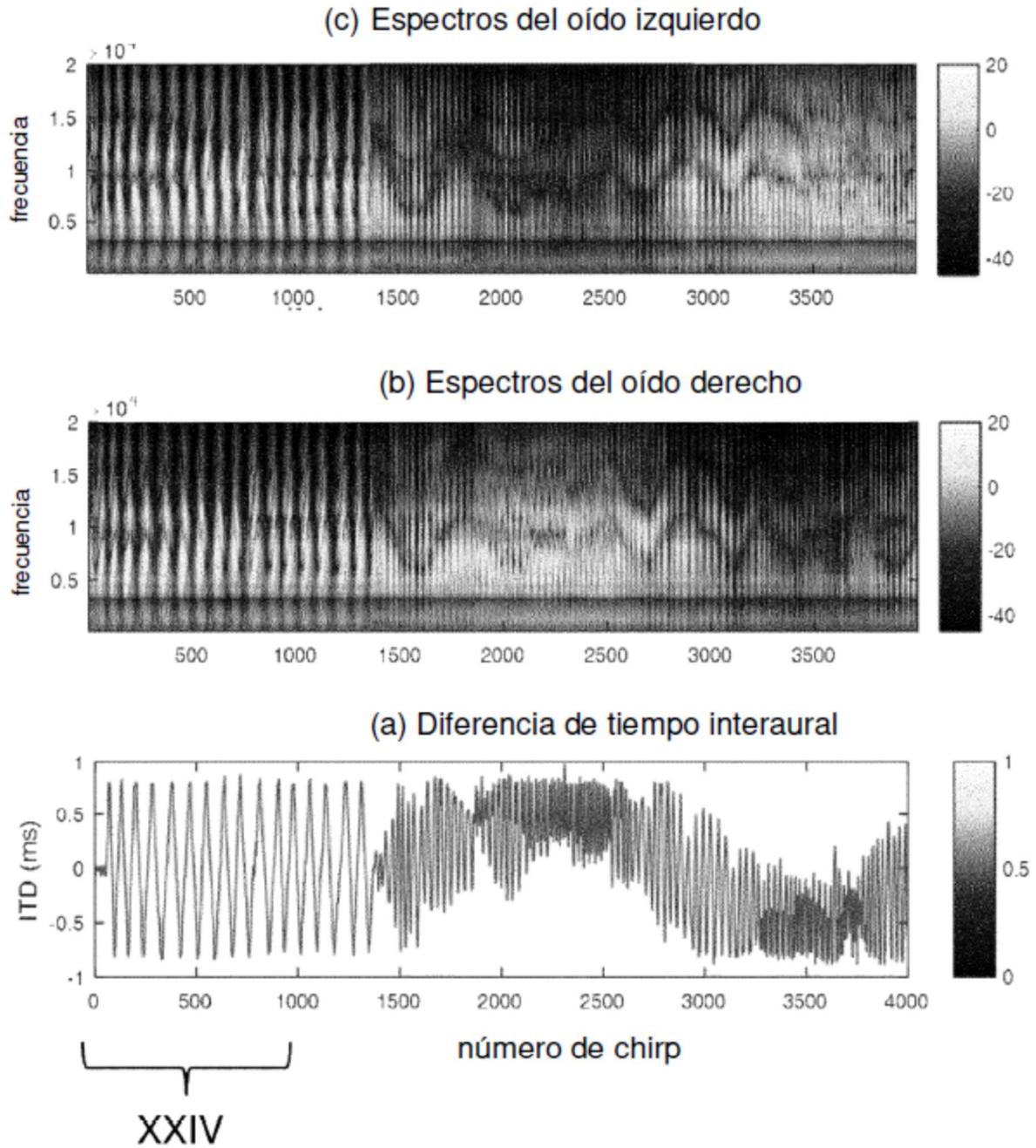


FIG 23

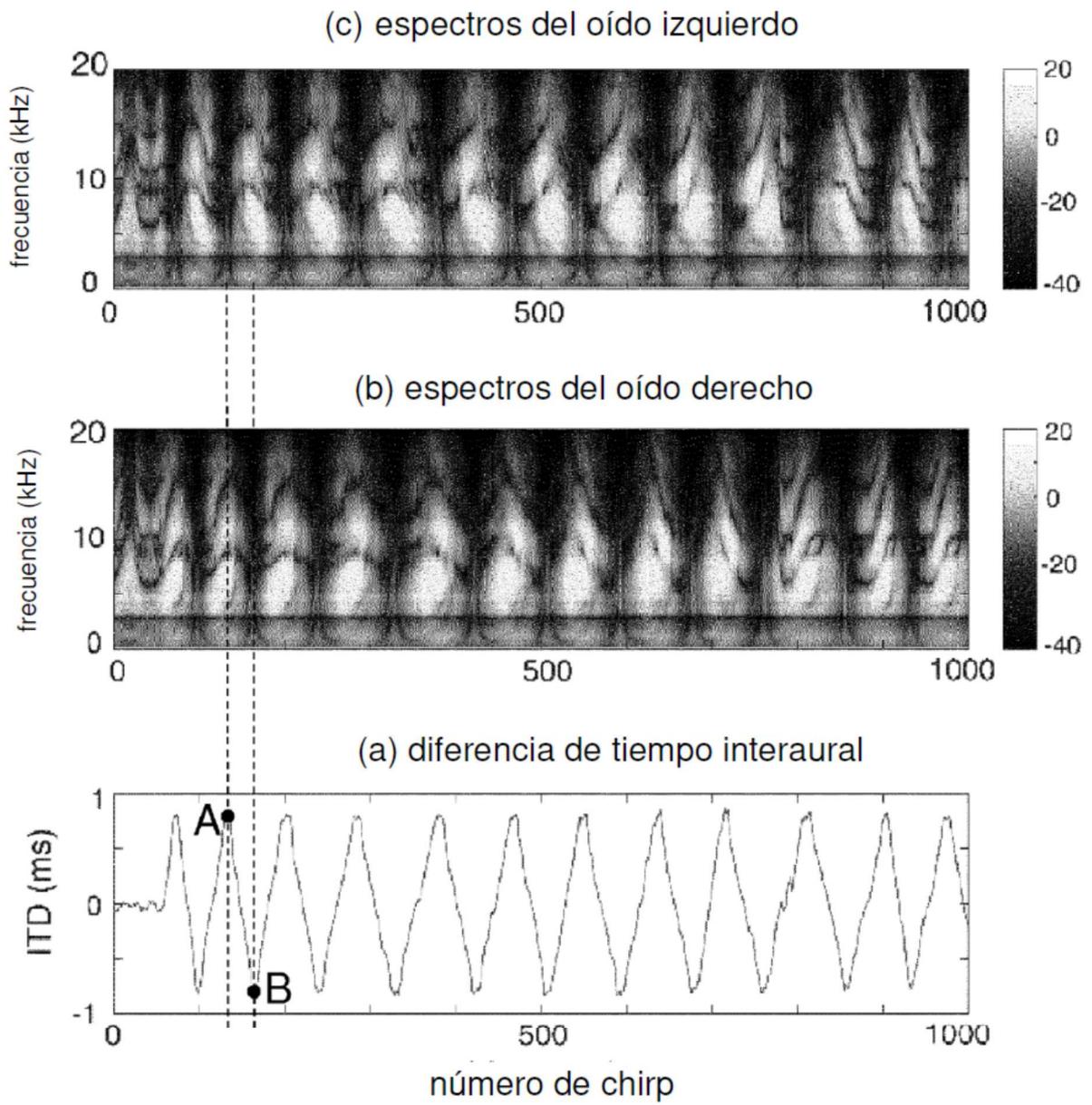


FIG 24

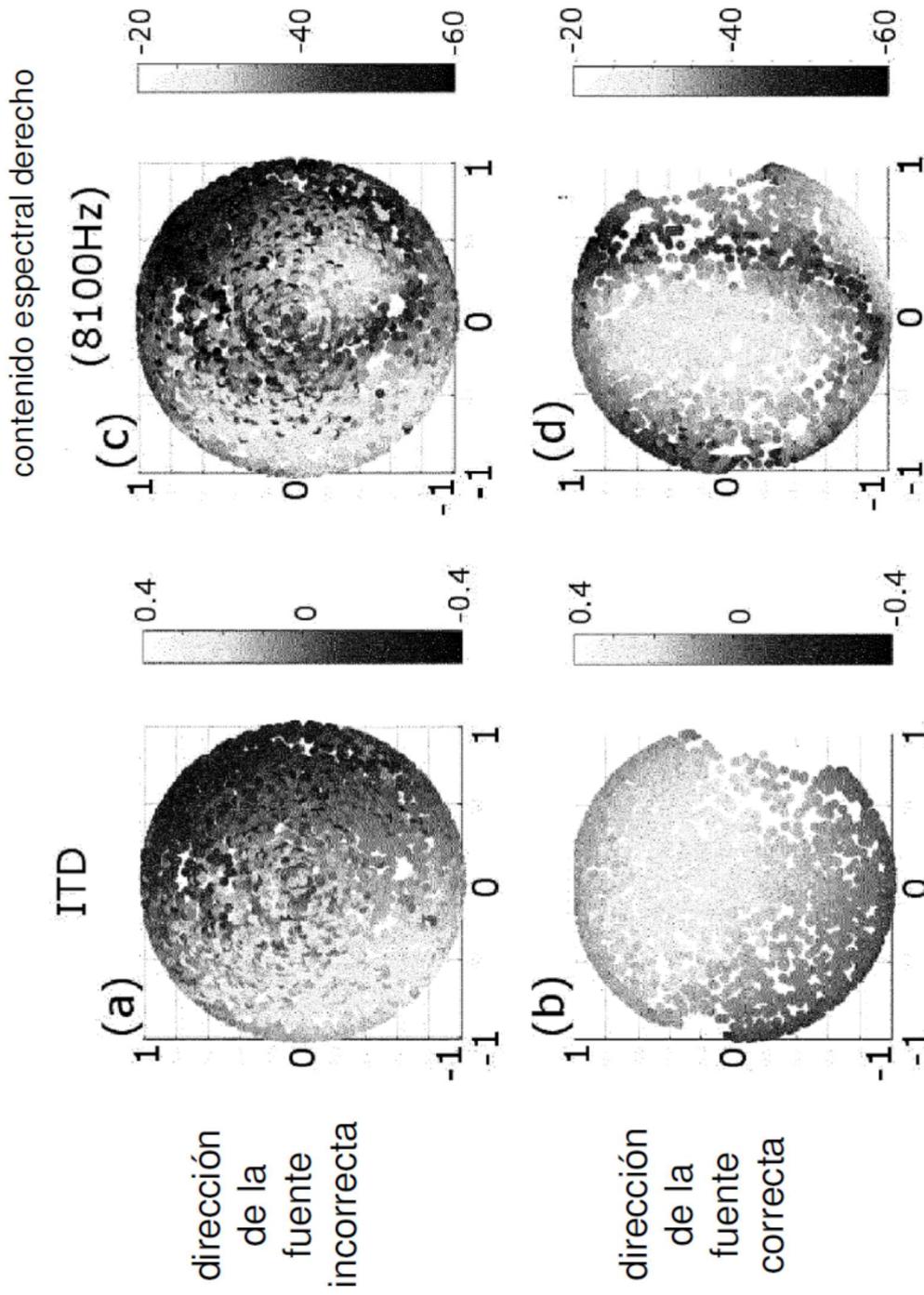


FIG 25

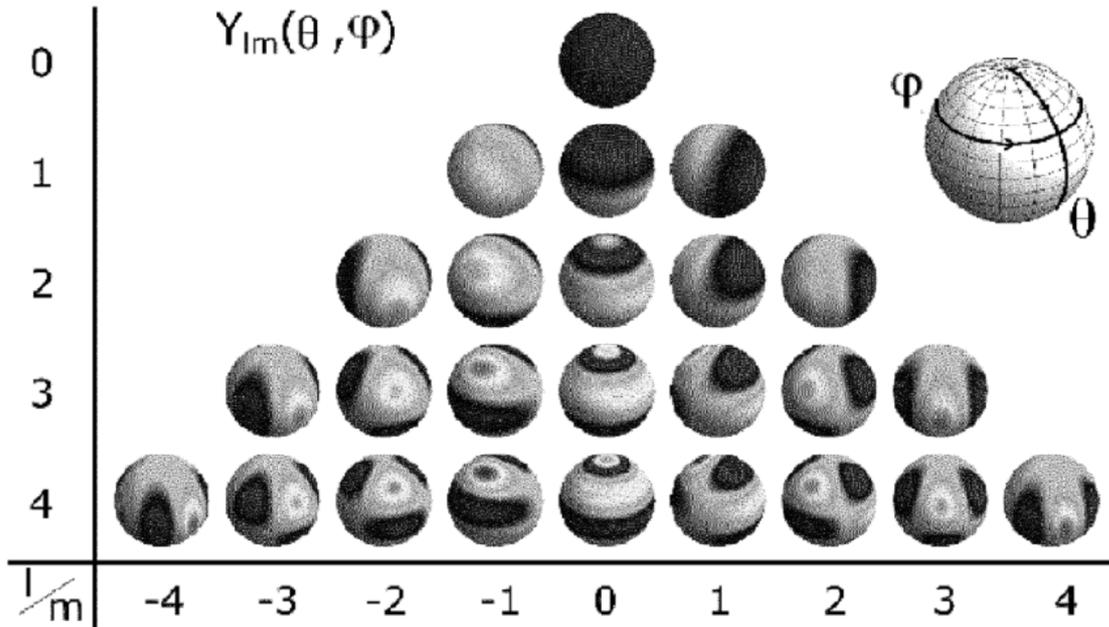


FIG 26(a)

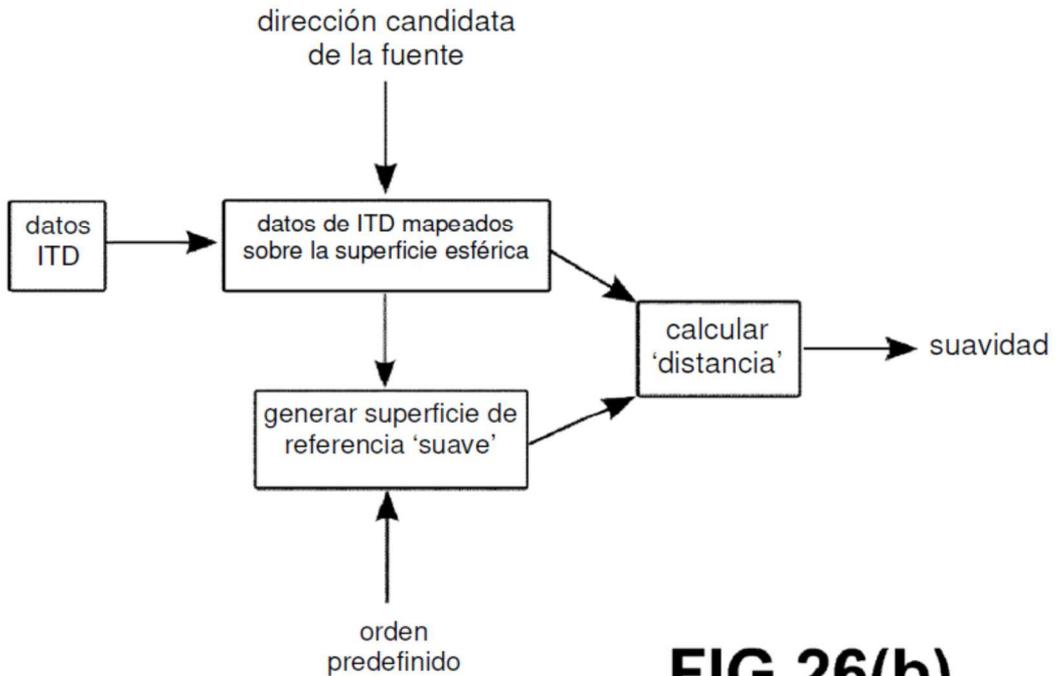


FIG 26(b)

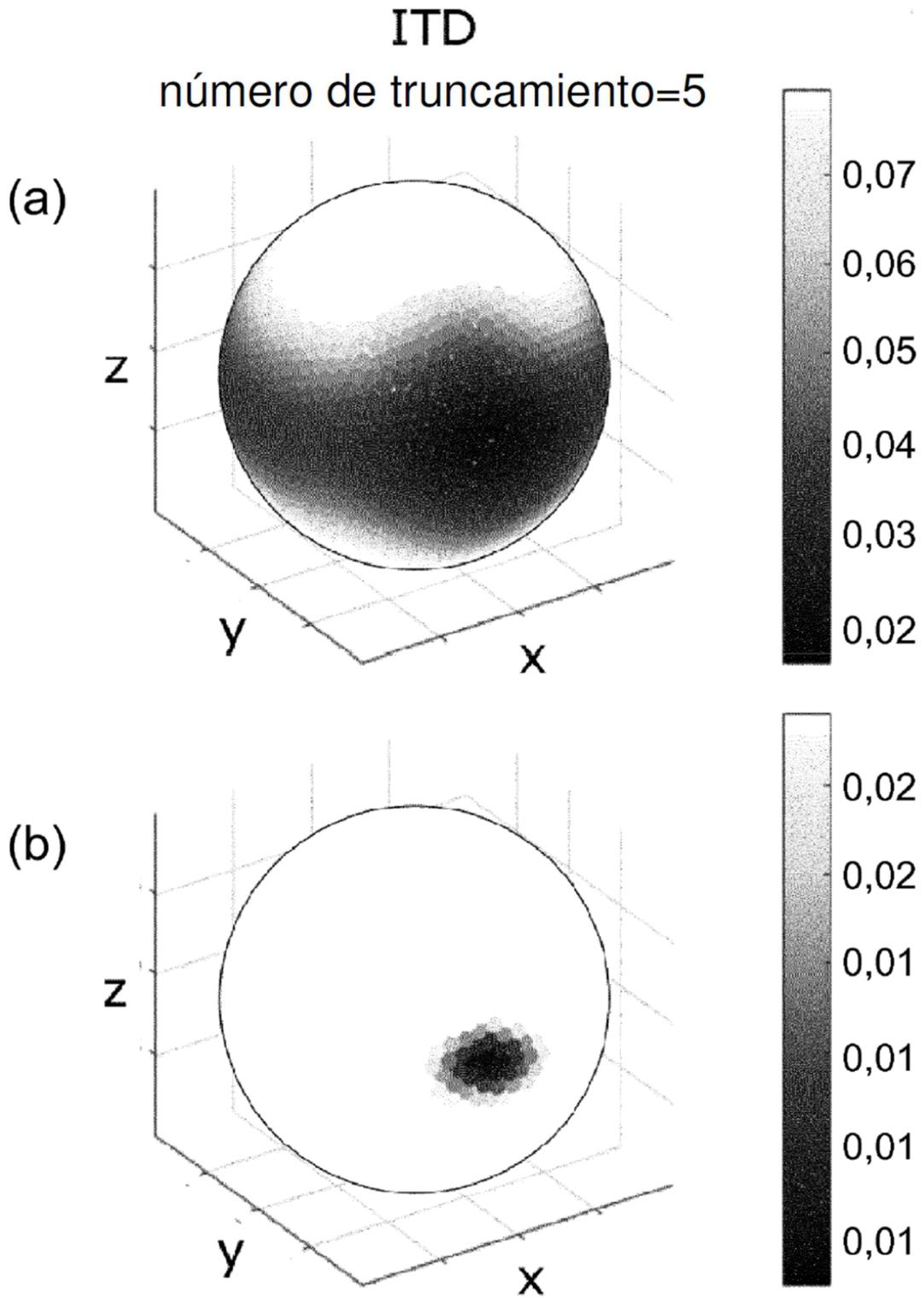
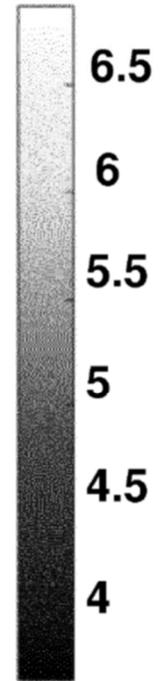
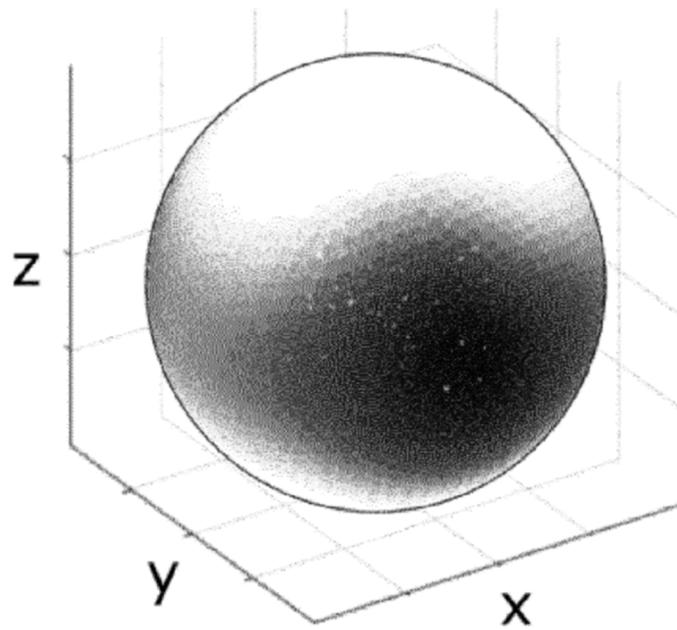


FIG 27

espectros binaurales ($4\text{kHz} < f < 20\text{kHz}$)

número de truncamiento=5

(a)



(b)

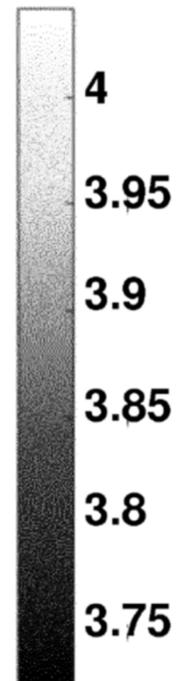
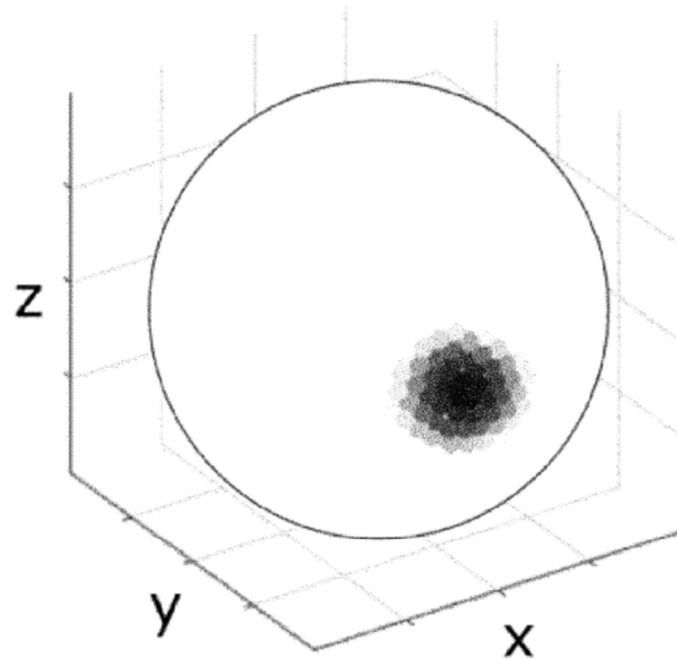


FIG 28

espectros binaurales ($4\text{kHz} < f < 20\text{kHz}$)

número de truncamiento=15

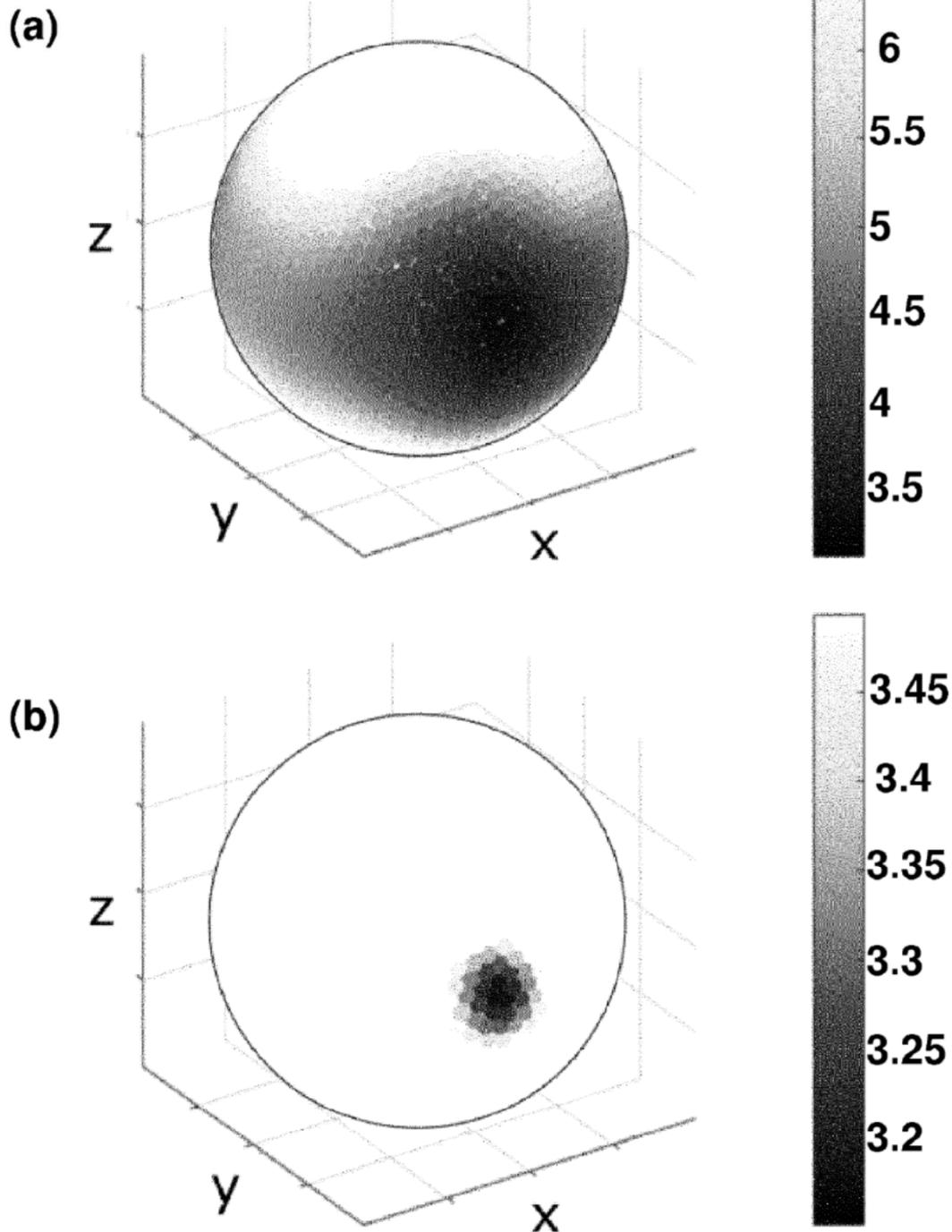


FIG 29

espectros monoaurales ($4\text{kHz} < f < 20\text{kHz}$)

número de truncamiento=15

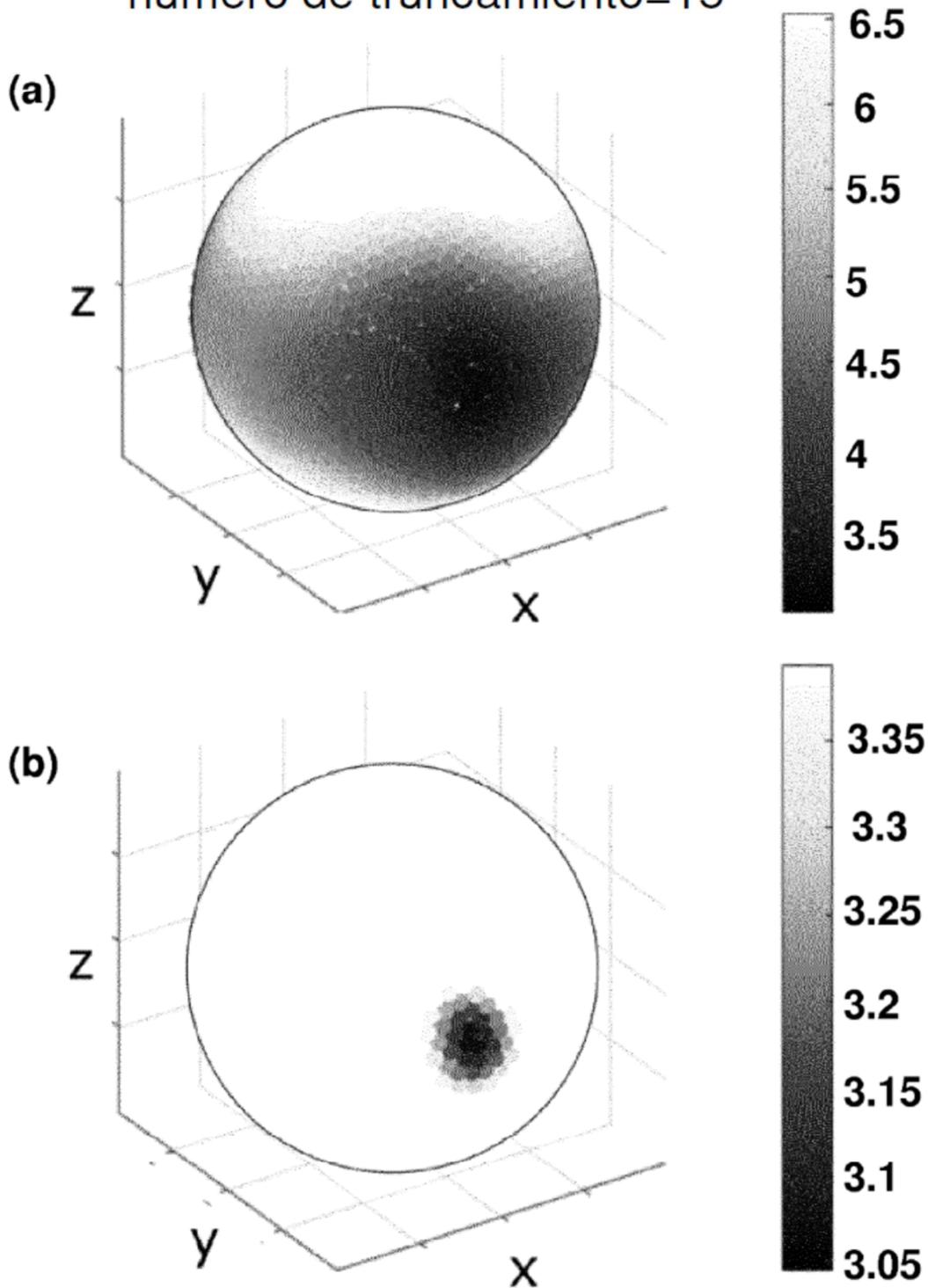


FIG 30

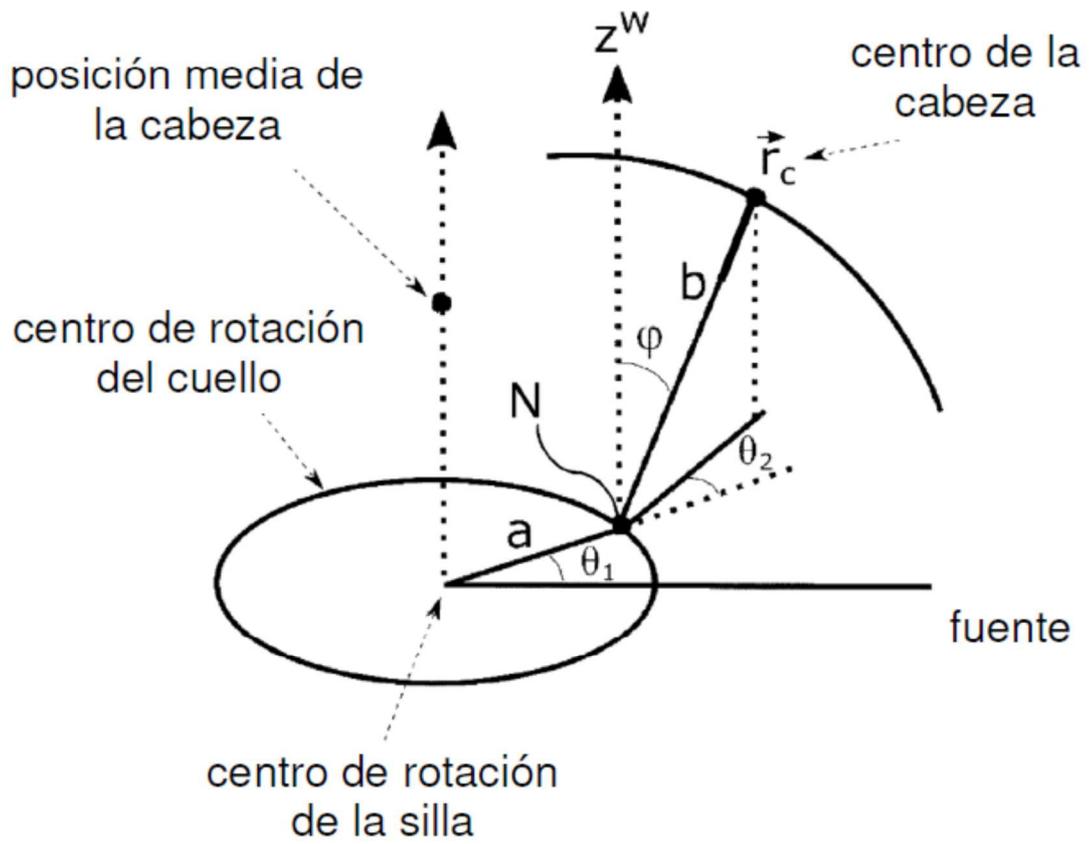


FIG 31

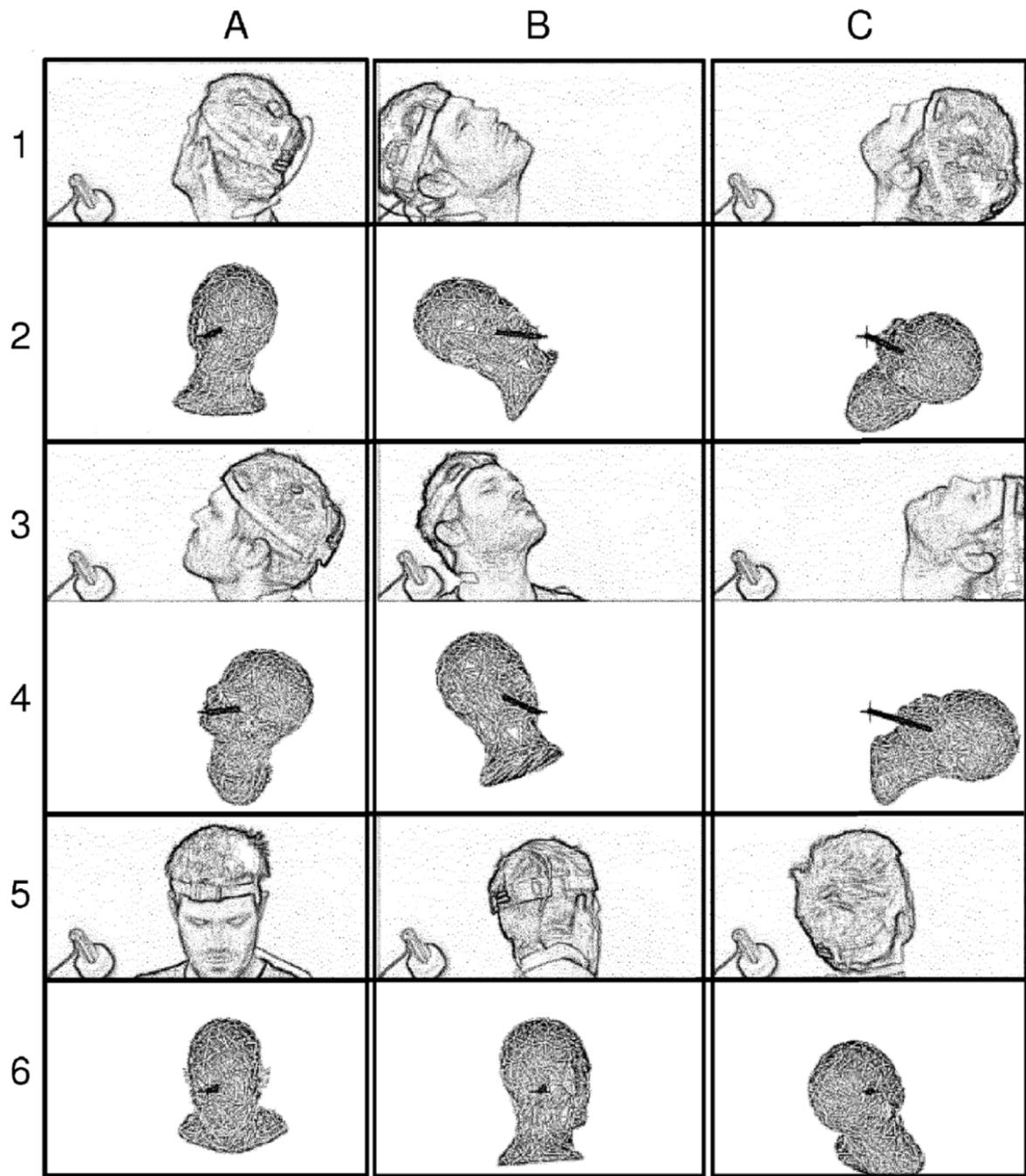


FIG 32

Movimiento estimado del centro de la cabeza (en coord. del mundo)

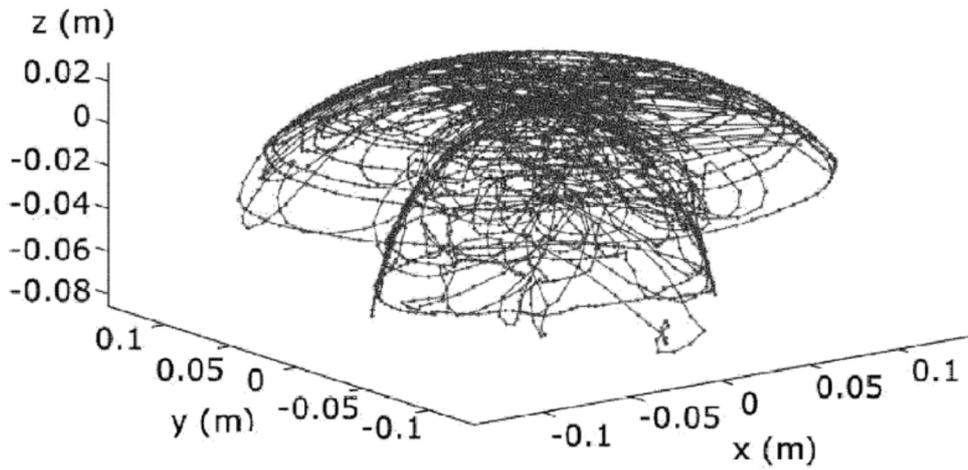


FIG 33

Distancia estimada vs distancia medida

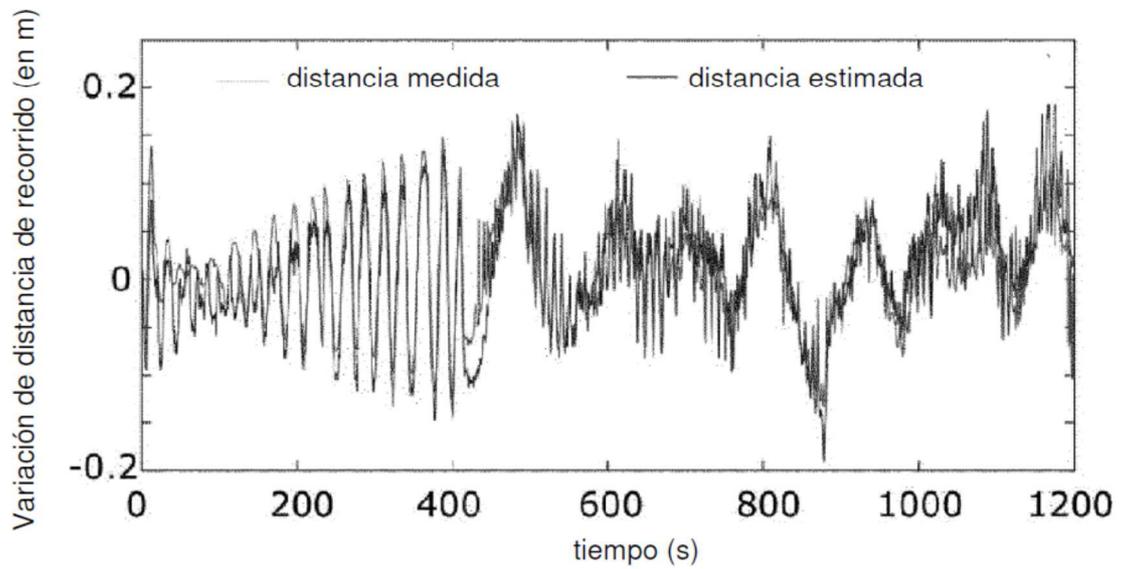


FIG 34

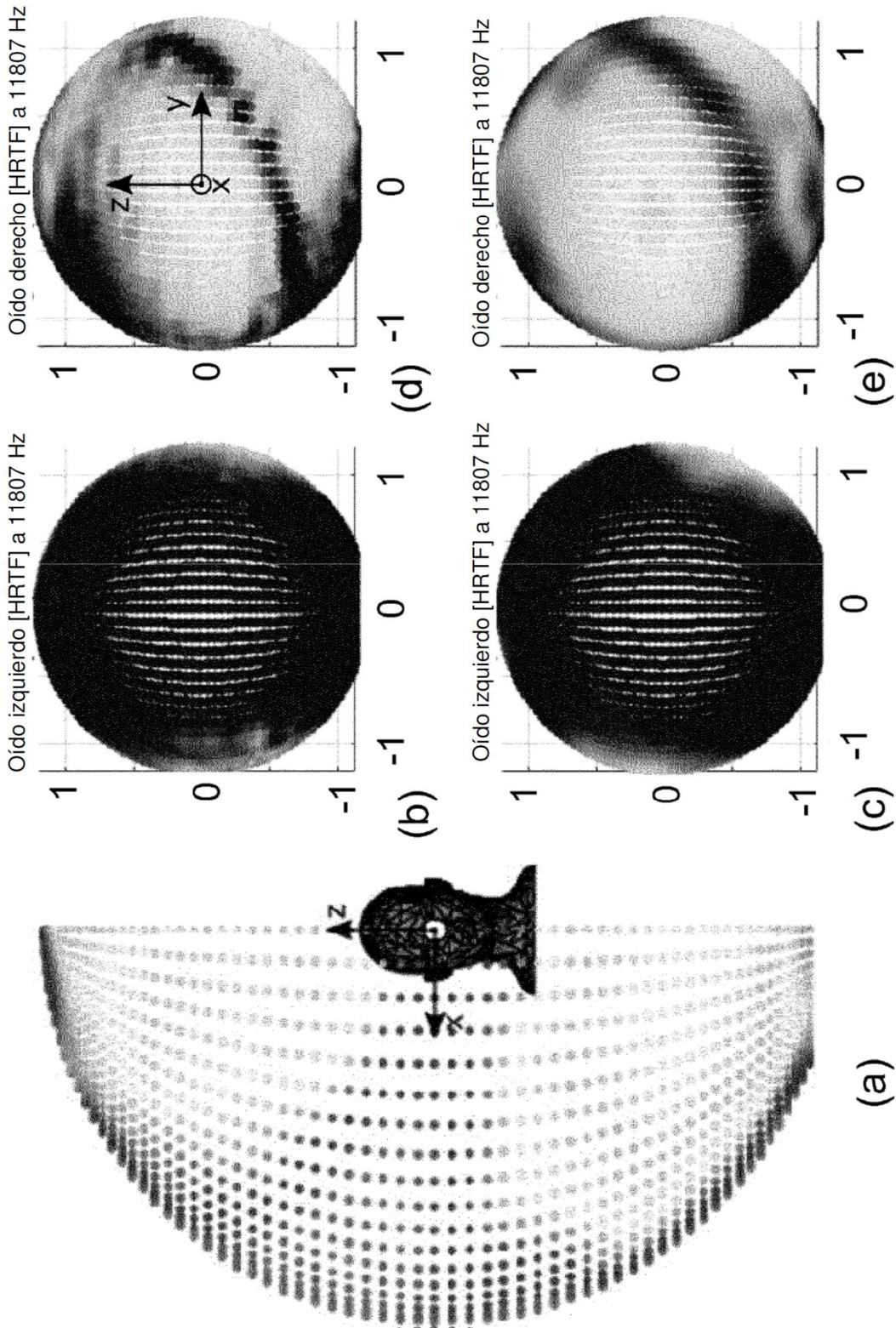


FIG 35