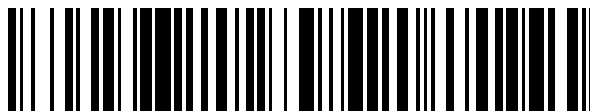


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 639 862**

21 Número de solicitud: 201700463

51 Int. Cl.:

G10L 17/00 (2013.01)

G10L 21/10 (2013.01)

G06T 1/00 (2006.01)

12

SOLICITUD DE PATENTE

A1

22 Fecha de presentación:

31.03.2017

43 Fecha de publicación de la solicitud:

30.10.2017

71 Solicitantes:

UNIVERSIDAD DE MÁLAGA (100.0%)

Avenida Cervantes, Nº 2

29071 Málaga ES

72 Inventor/es:

RAMOS JIMÉNEZ, Gonzalo Pascual

54 Título: **Dispositivo de montaje sobre la cabeza para percepción de realidad aumentada**

57 Resumen:

Dispositivo de montaje sobre la cabeza para percepción de realidad aumentada. La invención refiere un dispositivo que comprende medios de montaje sobre la cabeza del usuario; al menos una pantalla; al menos un micrófono; al menos un altavoz; y un módulo de procesamiento de datos, dicho módulo comprendiendo sub-módulos de tratamiento de sonidos y de imágenes, y de salida de datos, y opcionalmente de posicionamiento y orientación, y responsable de, entre otros, la correlación de al menos uno de los sonidos recibidos con una imagen recibida correspondiente a la entidad física que emite dicho sonido y la eliminación del sonido junto con la imagen correspondiente a la entidad física, así como, alternativamente, la eliminación de una imagen o de un sonido correlacionados si se elimina el sonido o la imagen correspondiente, respectivamente. La invención también refiere un método de generación de un entorno de realidad aumentada mediante el dispositivo referido.

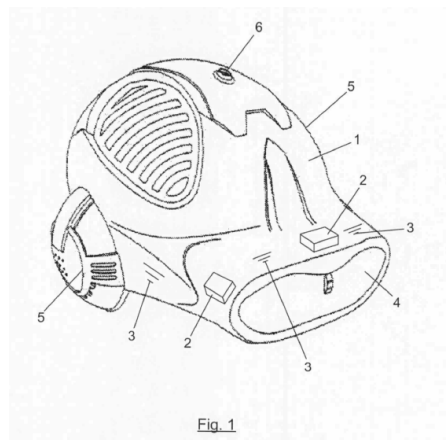


Fig. 1

DESCRIPCIÓN

Dispositivo de montaje sobre la cabeza para percepción de realidad aumentada.

5 Campo de la invención

La presente invención se refiere a un dispositivo de montaje sobre la cabeza de un usuario para la percepción de realidad aumentada. En concreto se pretende que el usuario obtenga, parcialmente, una percepción de la realidad pero con ciertas modificaciones tal y como se explicará en mayor detalle a continuación.

Antecedentes de la invención

Son conocidos diversos dispositivos de montaje sobre la cabeza de un usuario para realidad virtual en los que el usuario tiene una percepción visual de un escenario completamente diferente a la realidad.

Además, son conocidos diversos dispositivos de realidad aumentada que principalmente se refieren a anteojos que superponen ciertas imágenes sobre la realidad para dotar al usuario de una experiencia mixta entre la realidad y la virtualidad. Conforme a lo anterior, y en el contexto de la presente invención, se entenderán comprendidas dentro del concepto de realidad aumentada evoluciones o variantes respecto de la misma, tales como realidad mixta (traducción de la expresión en inglés "mixed reality"), realidad mediada (por computador; traducción de la expresión en inglés "(computer-)mediated reality"), realidad sustitucional (traducción de la expresión en inglés "substitutional reality"), o realidad integrada (traducción de la expresión en inglés "integrated reality"), por ejemplo.

Sin embargo, los dispositivos de la técnica anterior se han enfocado exclusivamente en el aspecto visual dejando de lado, por ejemplo, el sonido por lo que la experiencia de realidad aumentada es incompleta.

Descripción de la invención

La presente invención da a conocer un dispositivo de montaje sobre la cabeza para percepción de realidad aumentada que comprende un módulo de recolección de imágenes, un módulo de recolección de sonidos, un módulo de tratamiento de datos y un módulo de salida que comprende una pantalla y altavoces.

En particular, el dispositivo de la presente invención se caracteriza porque el módulo de tratamiento de datos dispone de diversos módulos para el tratamiento del sonido a fin de dar al usuario una experiencia completa que no solo se basa en la percepción visual sino que le da igual importancia a la percepción auditiva.

En concreto, la presente invención da a conocer un dispositivo de montaje sobre la cabeza de un usuario para la percepción de realidad aumentada que comprende:

- medios de montaje sobre la cabeza del usuario;
- al menos una pantalla;
- al menos un micrófono;
- al menos un altavoz; y

- un módulo de procesamiento de datos, dicho módulo comprendiendo a su vez un sub-módulo de tratamiento de imágenes, un sub-módulo de tratamiento de sonidos y un sub-módulo de salida de datos,

5 comprendiendo el sub-módulo de tratamiento de sonidos medios de descomposición de los sonidos recibidos mediante el al menos un micrófono en una serie de espectrogramas y medios de correlación de dichos espectrogramas con espectrogramas de una biblioteca a fin de identificar qué entidad física produce el sonido.

10 Preferentemente, la correlación de los espectrogramas generados a partir de los sonidos recibidos con espectrogramas de una biblioteca se realiza mediante al menos una de las siguientes técnicas: Template Matching, SVM, Deep Learning, y/o redes neuronales, por ejemplo perceptrón multicapa MLP.

15 Por otra parte, la correlación de los espectrogramas generados a partir de los sonidos recibidos puede comprender, por ejemplo, la generación de imágenes de los espectrogramas recibidos, la biblioteca comprendiendo además imágenes de espectrogramas. De esta manera la correlación y análisis de los espectrogramas no se realiza en función de comparaciones entre seis de datos sino comparación de imágenes.

20 El sub-módulo de salida de datos del dispositivo de la presente invención puede comprender medios de eliminación de algunos sonidos cuyos espectrogramas han sido identificados. De forma análoga, el sub-módulo de salida de datos también puede comprender medios de adición de sonidos, por ejemplo, sonidos almacenados en la base de datos.

25 Preferentemente, el sub-módulo de salida de datos comprende medios de selección de espectrogramas, medios de selección de imágenes y medios de transmisión de las imágenes y/o espectrogramas seleccionados a al menos una pantalla y/o altavoz.

30 El sub-módulo de salida de datos puede estar provisto de una entrada proveniente del sub-módulo de tratamiento de imágenes, una entrada proveniente del sub-módulo de tratamiento de sonidos y dispone de medios de correlación de al menos una de las entradas del sub-módulo de tratamiento de imágenes con al menos una de las entradas del sub-módulo de tratamiento de sonidos.

35 Por otra parte, el sub-módulo de salida de datos puede disponer de medios de eliminación de una imagen correlacionada si se elimina el sonido correspondiente a dicha imagen. De igual manera, el sub-módulo de salida de datos puede disponer de medios de eliminación de un sonido correlacionado si se elimina la imagen correspondiente a dicho sonido.

40 Opcionalmente, el dispositivo objeto de la presente invención dispone de medios de posicionamiento, por ejemplo, mediante un sistema GNSS (siglas de la expresión en inglés "Global Navigation Satellite System", sistema global de navegación por satélite), tal como GPS, GLONASS, Galileo, etc. Del mismo modo, el dispositivo puede disponer de medios para determinar la orientación del dispositivo, tales como acelerómetros,

45 Es otro objetivo de la presente invención dar a conocer un método de generación de un entorno de realidad aumentada mediante un dispositivo que comprende

- medios de montaje sobre la cabeza del usuario;
- al menos una pantalla;

- al menos un micrófono;
 - al menos un altavoz; y
- 5 • un módulo de procesamiento de datos;

que comprende las etapas de:

- 10 I. obtención de imágenes a través de al menos una cámara;
- 10 II. obtención de sonidos a través de al menos un micrófono;
- 15 III. procesamiento de datos que a su vez comprende la correlación de al menos uno de los sonidos recibidos con una imagen recibida correspondiente a la entidad física que emite dicho sonido y la eliminación del sonido junto con la imagen correspondiente a la entidad física; y
- 20 IV. reproducción a través de la al menos una pantalla y/o del al menos un altavoz de las imágenes y/o sonidos que no han sido eliminadas en la etapa III.

Para la correlación de los sonidos recibidos se pueden obtener previamente una serie de espectrogramas correspondientes a los sonidos recibidos y/o realizar alguna comparación de los sonidos recibidos con una base de datos de sonidos.

25 Preferentemente, dicha base de datos de sonidos se encuentra en una memoria del dispositivo aunque, alternativamente, la base de datos de sonidos se puede encontrar en un servidor remoto al dispositivo, por ejemplo, en la nube.

30 Adicionalmente, la correlación de los sonidos recibidos con una imagen recibida puede comprender el procesamiento mediante técnicas de inteligencia artificial como, por ejemplo, Template Matching, SVM, Deep Learning, y/o redes neuronales, por ejemplo perceptrón multicapa MLP.

35 Más preferentemente, la etapa IV comprende la reproducción a través de la al menos una pantalla y/o del al menos un altavoz de al menos una imagen y/o un sonido almacenados en la base de datos y adicionales a los obtenidos en la etapa I.

40 Adicionalmente a la correlación de sonidos e imágenes comprendida en la etapa III de procesamiento de datos, dicha etapa III puede comprender la eliminación de una imagen correlacionada si se elimina el sonido correspondiente a dicha imagen, así como la eliminación de un sonido correlacionado si se elimina la imagen correspondiente a dicho sonido.

Breve descripción de los dibujos

45 En las figuras adjuntas se muestran, de manera ilustrativa y no limitativa, dos ejemplos de realización del sistema según la presente invención, en las que:

50 - La figura 1 es un ejemplo de dispositivo según la presente invención.

- La figura 2 es un diagrama de flujo del funcionamiento de un dispositivo según la presente invención, haciendo énfasis en el sub-módulo de tratamiento de sonidos del módulo de procesamiento de datos y en la interacción de dicho sub-módulo y los sub-módulos de tratamiento de imágenes y de salida de datos.

Descripción detallada de un modo de realización

La figura 1 muestra un dispositivo según la presente invención. En esta figura se observan las partes principales del dispositivo (1) que son al menos una cámara (2) ubicada, por ejemplo, en la parte delantera del dispositivo a fin de obtener las imágenes de un punto determinado al que se encuentre mirando el usuario, al menos un micrófono (3) para captar sonidos del ambiente y, con respecto a la salida hacia el usuario, dispone de al menos una pantalla (4) y al menos un altavoz (5).

Alternativamente, el dispositivo puede comprender cámaras en los laterales (por ejemplo, una en cada lateral) y/o en la parte trasera del dispositivo a fin de permitir la captación de imágenes antes de que el usuario cambie la orientación del dispositivo.

El dispositivo dispone de medios de montaje sobre la cabeza del usuario que pueden ser, por ejemplo, un par de patillas para su montaje sobre la cabeza del usuario o una sección de un casco que cubra al menos parcialmente la cabeza y, preferentemente, las orejas del usuario y disponga de los altavoces en la vecindad de las orejas de dicho usuario, tal y como se muestra en la figura 1.

Dado que el objetivo de la presente invención es presentar al usuario una realidad modificada, se concibe que, en un aspecto especialmente preferente de la presente invención, se dispongan medios para que el usuario vea las imágenes seleccionadas por el dispositivo e, igualmente, escuche los sonidos seleccionados. En consecuencia, se contempla que la al menos una pantalla (4) comprendida en el dispositivo objeto de la presente invención no permita ver al usuario a través de ella (es decir, que no sea transparente), siendo preferentemente de tipo NED (siglas en inglés de la expresión "Near Eye Display" o pantalla cercana al ojo), así como que, adicionalmente, en una realización con dos altavoces (5), éstos tengan forma de auriculares del tipo "sobre la oreja" a fin de que dispongan de medios de reducción de ruido para evitar que se escuchen ruidos exteriores que puedan interferir con la información perceptible por el usuario.

Según la presente invención, se capturan sonidos y/o imágenes del entorno del usuario. Posteriormente, mediante medios de procesamiento de datos, se seleccionan cuáles de las imágenes y/o sonidos capturados serán finalmente enviados al usuario. Adicionalmente, se contempla la posibilidad de añadir sonidos y/o imágenes provenientes de una base de datos y que no se encuentran en el entorno del usuario.

Adicionalmente, el dispositivo de la figura 1 dispone de medios de posicionamiento y orientación (6). De esta manera, se dispone de la posición del dispositivo respecto a un eje de coordenadas X, Y, Z así como la orientación del dispositivo.

En cuanto al tratamiento de datos, el dispositivo dispone de un módulo de tratamiento de datos, preferentemente en tiempo real. Los datos que debe gestionar el dispositivo son principalmente: imágenes, sonidos, posición y orientación del dispositivo y salida de datos. En consecuencia, el módulo de tratamiento de datos dispone de sub-módulos para tratar cada uno de estos tipos de datos.

Sub-módulo de tratamiento de imágenes (40)

El objetivo del sub-módulo de imágenes (40) es, por una parte, obtener una serie de imágenes, en concreto imágenes de lo que vería el usuario si no tuviera el dispositivo así como de su entorno. Para esto el sub-módulo de tratamiento de imágenes dispone de

entrada de imágenes obtenidas por cámaras y medios de tratamiento de dichas imágenes a fin de adaptarlas a pantallas, preferentemente tipo NED.

Preferentemente, el sub-módulo de tratamiento de imágenes puede disponer también de una serie de bibliotecas de objetos a fin de, entre otros, poder identificar los objetos que está observando el usuario y poder clasificarlos en una serie de objetos conocidos. En definitiva, en la base de datos se disponen imágenes de objetos y se dispone información identificativa de cada objeto (por ejemplo, una referencia o su nombre). De esta manera, al captar la imagen de un objeto se procede a comparar con la base de datos si existen objetos similares y, si hay una coincidencia con un objeto de la base de datos, se procede a clasificar el objeto según la información identificativa del mismo.

Además, esta identificación de objetos puede dotar al dispositivo de la capacidad de dar una mayor información al usuario indicándole, por ejemplo mediante texto, información relevante respecto a los objetos que está observando.

Otra posible utilización de la identificación de objetos es la posibilidad de, en el sub-módulo de salida de datos que se describe más adelante, eliminar la imagen de objetos reales de la información perceptible por el usuario; así se puede evitar que el usuario tenga acceso a imágenes determinadas. En línea con esta posible utilización, una realización preferente de la invención comprende no sólo la eliminación de dichas imágenes de objetos reales de dicha información perceptible por el usuario, sino también la inclusión de imágenes de objetos no reales (virtuales), almacenados y disponibles a partir de bibliotecas de objeto como las referidas anteriormente, en dicha información perceptibles por el usuario, así como, en su caso, su posterior eliminación.

En un ejemplo de realización, el usuario observa en su entorno una máquina. En primer lugar, se capta la imagen de la máquina y se procede a consultar la base de datos para localizar imágenes similares. Las imágenes de la base de datos están relacionadas con al menos un campo referente a información identificativa, por ejemplo, que es un objeto realizado por el hombre. En consecuencia, el dispositivo procede a marcar la imagen obtenida con la información de que es un objeto realizado por el hombre.

Tras identificar la imagen, el dispositivo puede estar configurado, por ejemplo, para eliminar todos los objetos realizados por el hombre por lo que, en el sub-módulo de salida, se eliminarla dicha imagen.

Adicionalmente, en el sub-módulo de salida se podría reemplazar dicha imagen añadiendo, por ejemplo, una planta en su lugar a fin de evitar que el usuario pueda tropezar al moverse haciendo uso del dispositivo.

Sub-módulo de tratamiento de sonidos (20)

El sub-módulo de tratamiento de sonidos (20) incluye la captación del sonido entrante que es captado, preferentemente, a través del al menos un micrófono (3) del dispositivo. Para la captación de sonido ambiente, el dispositivo puede comprender dos o más micrófonos (3) dispuestos, por ejemplo, en los laterales del dispositivo. Alternativamente, dichos micrófonos pueden, además, detectar la dirección del sonido recibido, en el caso de micrófonos direccionales. Alternativamente, los medios de procesamiento de datos pueden comprender medios de tratamiento de sonidos para identificar una posición estimada de la proveniencia del sonido.

Una vez captado el sonido, se procede a la realización de un filtrado del ruido (21). El filtrado del ruido puede realizarse mediante cualquiera de las técnicas conocidas y

comprendidas en el estado de la técnica, tales como el uso de un filtro Wiener, alternativamente, la presente invención contempla la utilización también técnicas de Inteligencia Artificial (IA) similares a las que se usarán en las siguientes fases del procesamiento.

Otra de las etapas que contempla la presente invención se refiere a la descomposición e identificación de sonidos (22). El objetivo de esta etapa es discernir entre los diferentes sonidos detectados y, una vez clasificados poder identificar a que corresponden los sonidos detectados.

La presente invención contempla la descomposición por frecuencias del sonido a fin de tener para cada sonido un espectrograma que comprende al menos la frecuencia y la intensidad. Una vez se dispone de la descomposición de cada sonido en frecuencias e intensidades se procede a utilizar algoritmos de inteligencia artificial.

En concreto, la descomposición de e identificación de sonidos (22) se realiza a partir del espectrograma de frecuencias e intensidad, y por medio de distintas técnicas de Aprendizaje Automático, tales como Deep Learning, Template Matching, SVM (siglas de la expresión en inglés "Support Vector Machines"), y otros tipos de redes neuronales.

Una vez se dispone del espectrograma se procede a clasificar los sonidos entrantes y diferenciarlos unos de otros. Dichas técnicas utilizarán una base de datos de sonidos, alternativamente, dicha base de datos de sonido puede ser una base de sonidos accesible mediante internet. Las técnicas mencionadas son complementarias, de manera que combinaciones entre ellas también son posibles para realizar la citada descomposición e identificación.

Con el fin de clasificar los sonidos, mediante Template Matching es posible identificar que tan parecidos son dos espectrogramas de datos. En consecuencia, se pueden comparar los sonidos recibidos con una base de datos de sonidos (26), estando dicha base de datos de sonidos almacenada en el dispositivo o, alternativamente, en un servidor y accediendo a ellos, preferentemente en tiempo real o casi real.

Por otra parte las SVM (siglas de la expresión en inglés "Support Vector Machines") indican qué probabilidad hay de que cada sonido entrante se corresponda con alguno de los almacenados en la base de datos de sonidos (26). Para ello se generan los vectores del sonido, a partir del espectrograma, y se comparan con los vectores disponibles.

Las redes neuronales (p.e. perceptrón multicapa - MLP) también trabajan con vectores y se pueden utilizar como único mecanismo de identificación o en combinación con otras técnicas para complementar la información. Además, si bien su entrenamiento suele ser más lento, su aplicación para clasificar suele ser muy rápida, lo cual nos conviene en vistas a trabajar en tiempo real.

En el caso de las redes neuronales, para su entrenamiento se puede hacer énfasis en diferentes propiedades del sonido, en particular, es especialmente ventajosa la utilización del par frecuencia-intensidad para identificar a que corresponde cada sonido. Alternativamente, ante un par frecuencia-intensidad desconocido se puede preguntar al usuario que sonido es para que el usuario lo identifique y la red neuronal tenga un aprendizaje continuo, una vez el usuario ha identificado un sonido nuevo se puede incorporar a la base de datos de sonidos (26) o mejorar el algoritmo en caso de que ya estuviese almacenado mediante el entrenamiento de la red neuronal.

En una realización especialmente preferentemente, la técnica para identificar a que corresponde cada sonido captado es mediante Deep Learning, utilizando los espectrogramas como imágenes sobre las que aprender.

- 5 Un detalle destacable es que si bien las fases de aprendizaje (y por tanto entrenamiento) de los algoritmos de aprendizaje citados pueden durar un cierto tiempo, lo importante es que la fase de aplicación de lo aprendido (predicción o clasificación) si puede hacerse muy rápido, incluso en tiempo real o casi real.
- 10 Dichos espectrogramas, posteriormente, pasan por una fase de ponderación para obtener la descomposición e identificación de los sonidos. Estos resultados se pasarían a la siguiente fase, la de eliminación de sonidos, pero a su vez también se le pasaría un informe de dichos resultados a los procesos que controlan los requerimientos de la realidad integrada. Estos datos se procesarán en el sub-módulo de salida de datos (30) a
- 15 fin de determinar qué sonidos han de mantenerse, cuáles han de eliminarse (en la fase de eliminación de sonidos (23)) y cuáles nuevos han de incorporarse (en la fase de inclusión de sonidos (24)).
- 20 Volviendo al ejemplo de realización en base al que se explicó el sub-módulo de tratamiento de imágenes, el micrófono del dispositivo capta la sumatoria de una pluralidad de sonidos que se encuentran en el ambiente. Mediante los medios de filtrado se eliminan, por ejemplo, sonidos que no son interesantes para su procesamiento y que pueden considerarse como ruido.
- 25 Mediante la descomposición por frecuencias se separan los sonidos a fin de obtener, por ejemplo, el sonido correspondiente a una máquina tal como un coche o sonidos de aves. Para cada uno de estos sonidos se obtiene un espectrograma en el que se pueden identificar las frecuencias así como las intensidades de sonido en cada una de dichas frecuencias.
- 30 Para al menos uno de dichos espectrogramas se realiza una búsqueda y una comparación con sonidos almacenados en la base de datos a fin de identificar a que objeto corresponde dicho espectrograma.
- 35 Retornando al ejemplo anterior se habrá identificado que un sonido corresponde a un coche y otro sonido corresponde a un ave.

Sub-módulo de posicionamiento y orientación del dispositivo

- 40 El dispositivo objeto de la presente invención dispone de medios de posicionamiento, por ejemplo, mediante un sistema GNSS (siglas de la expresión en inglés "Global Navigation Satellite System", sistema global de navegación por satélite), tal como GPS, GLONASS, Galileo, etc.
- 45 Adicionalmente, el sistema dispone de medios para determinar la orientación del dispositivo, en particular mediante acelerómetros. Alternativamente se podrían utilizar unidades de medición de inercia (IMU) o otros tipos de sensores tridimensionales de orientación más complejos como AHRS (siglas de la expresión en inglés "Attitude and Heading Reference System").
- 50 En una realización preferente de la invención, el dispositivo puede determinar la posición (X, Y, Z) mediante un sistema de geo posicionamiento (por ejemplo, un sistema GNSS). Adicionalmente, el dispositivo puede conocer su orientación (dirección, elevación, ángulo de alabeo) mediante al menos un acelerómetro.

En un ejemplo de realización de la presente invención, el sistema dispone de un mapa virtual geo-referenciado, de manera que el sistema dispone de información posicional del usuario (mediante dichos medios de posicionamiento) y emite a través del módulo de salida al menos parcialmente imágenes de dicho mapa virtual geo-referenciado.

5

Sub-módulo de salida de datos (30)

Los datos de entrada al sub-módulo de salida de datos (30) comprenden, al menos, unas imágenes previamente identificadas en el sub-módulo de tratamiento de imágenes y unos sonidos previamente identificados en el sub-módulo de tratamiento de sonidos.

10

En el sub-módulo de salida de datos se procede a correlacionar las imágenes captadas con los sonidos captados mediante la información identificativa obtenida de las bases de datos, si bien se contempla la posible gestión de imágenes y sonidos que no precisan o para los que no se desea realizar correlación alguna con sonidos o imágenes, respectivamente.

15

Por ejemplo, volviendo a los ejemplos anteriores, mediante las cámaras se ha obtenido una imagen que, tras el procesamiento de dicha imagen en el sub-módulo de tratamiento de imágenes se ha identificado que corresponde a una máquina. Por otra parte, del sub-módulo de tratamiento de sonidos se ha identificado que hay un sonido correspondiente a un coche y un sonido correspondiente a un ave.

20

El sub-módulo de salida analiza las referencias identificativas y correlaciona el espectrograma correspondiente al coche con la imagen captada de la máquina de manera que, si los requerimientos del sistema requieren la eliminación de la máquina, el sub-módulo de salida no solo elimina la imagen de la máquina sino el sonido correspondiente a la misma.

25

Alternativamente, el sub-módulo de salida puede incluir la imagen de un ave (por ejemplo, obteniéndola de la base de datos) ya que ha identificado que hay un sonido que corresponde a la misma.

30

Finalmente, el sub-módulo de salida dispone de medios de comunicación con la al menos una pantalla (4), y el al menos un altavoz (5) a fin de enviar las imágenes y/o sonidos determinados.

35

REIVINDICACIONES

1. Dispositivo de montaje sobre la cabeza de un usuario para la percepción de realidad aumentada que comprende:

- medios de montaje sobre la cabeza del usuario;
- al menos una pantalla;
- al menos un micrófono;
- al menos un altavoz; y
- un módulo de procesamiento de datos, dicho módulo comprendiendo a su vez un sub-módulo de tratamiento de Imágenes, un sub-módulo de tratamiento de sonidos y un sub-módulo de salida de datos,

caracterizado por que el sub-módulo de tratamiento de sonidos comprende medios de descomposición de los sonidos recibidos mediante el al menos un micrófono en una serie de espectrogramas y medios de correlación de dichos espectrogramas con espectrogramas de una biblioteca a fin de identificar qué entidad física produce el sonido.

2. Dispositivo según la reivindicación 1 **caracterizado** por que la correlación de los espectrogramas generados a partir de los sonidos recibidos con espectrogramas de una biblioteca se realiza mediante Template Matching.

3. Dispositivo según la reivindicación 1 **caracterizado** por que la correlación de los espectrogramas generados a partir de los sonidos recibidos con espectrogramas de una biblioteca se realiza mediante redes neuronales.

4. Dispositivo según la reivindicación 1 **caracterizado** por que la correlación de los espectrogramas generados a partir de los sonidos recibidos con espectrogramas de una biblioteca se realiza mediante SVM.

5. Dispositivo según la reivindicación 1 **caracterizado** por que la correlación de los espectrogramas generados a partir de los sonidos recibidos con espectrogramas de una biblioteca se realiza mediante perceptrón multicapa MLP.

6. Dispositivo según la reivindicación 1 **caracterizado** por que la correlación de los espectrogramas generados a partir de los sonidos recibidos con espectrogramas de una biblioteca se realiza mediante Deep Learning.

7. Dispositivo según cualquiera de las reivindicaciones anteriores **caracterizado** por que la correlación de los espectrogramas generados a partir de los sonidos recibidos comprende la generación de imágenes de los espectrogramas recibidos y, además, la biblioteca comprende imágenes de espectrogramas.

8. Dispositivo según cualquiera de las reivindicaciones anteriores **caracterizado** por que el sub-módulo de salida de datos comprende medios de eliminación de algunos sonidos cuyos espectrogramas han sido identificados.

9. Dispositivo según cualquiera de las reivindicaciones anteriores **caracterizado** por que el sub-módulo de salida de datos comprende medios de adición de sonidos.

10. Dispositivo según cualquiera de las reivindicaciones anteriores **caracterizado** porque el sub-módulo de salida de datos comprende medios de selección de espectrogramas, medios de selección de imágenes y medios de transmisión de las imágenes y/o espectrogramas seleccionados a al menos una pantalla y/o altavoz.

11. Dispositivo según cualquiera de las reivindicaciones anteriores **caracterizado** porque el sub-módulo de salida de datos dispone de una entrada proveniente del sub-módulo de tratamiento de imágenes, una entrada proveniente del sub-módulo de tratamiento de sonidos y dispone de medios de correlación de al menos una de las entradas del sub-módulo de tratamiento de imágenes con al menos una de las entradas del sub-módulo de tratamiento de sonidos.

12. Dispositivo según la reivindicación 11 **caracterizado** por que el sub-módulo de salida de datos dispone de medios de eliminación de una imagen correlacionada si se elimina el sonido correspondiente a dicha imagen.

13. Dispositivo según la reivindicación 11 **caracterizado** por que el sub-módulo de salida de datos dispone de medios de eliminación de un sonido correlacionado si se elimina la imagen correspondiente a dicho sonido.

14. Dispositivo según cualquiera de las reivindicaciones 1 a 13 **caracterizado** por que el módulo de procesamiento de datos comprende además un sub-módulo de posicionamiento y orientación del dispositivo.

15. Método de generación de un entorno de realidad aumentada en un dispositivo de acuerdo a cualquiera de las reivindicaciones 1 a 14 **caracterizado** por que comprende las etapas de:

I. obtención de imágenes a través de al menos una cámara;

II. obtención de sonidos a través de al menos un micrófono;

III. procesamiento de datos que a su vez comprende la correlación de al menos uno de los sonidos recibidos con una imagen recibida correspondiente a la entidad física que emite dicho sonido y la eliminación del sonido junto con la imagen correspondiente a la entidad física; y

IV. reproducción a través de la al menos una pantalla y/o del al menos un altavoz de las imágenes y/o sonidos que no han sido eliminadas en la etapa III.

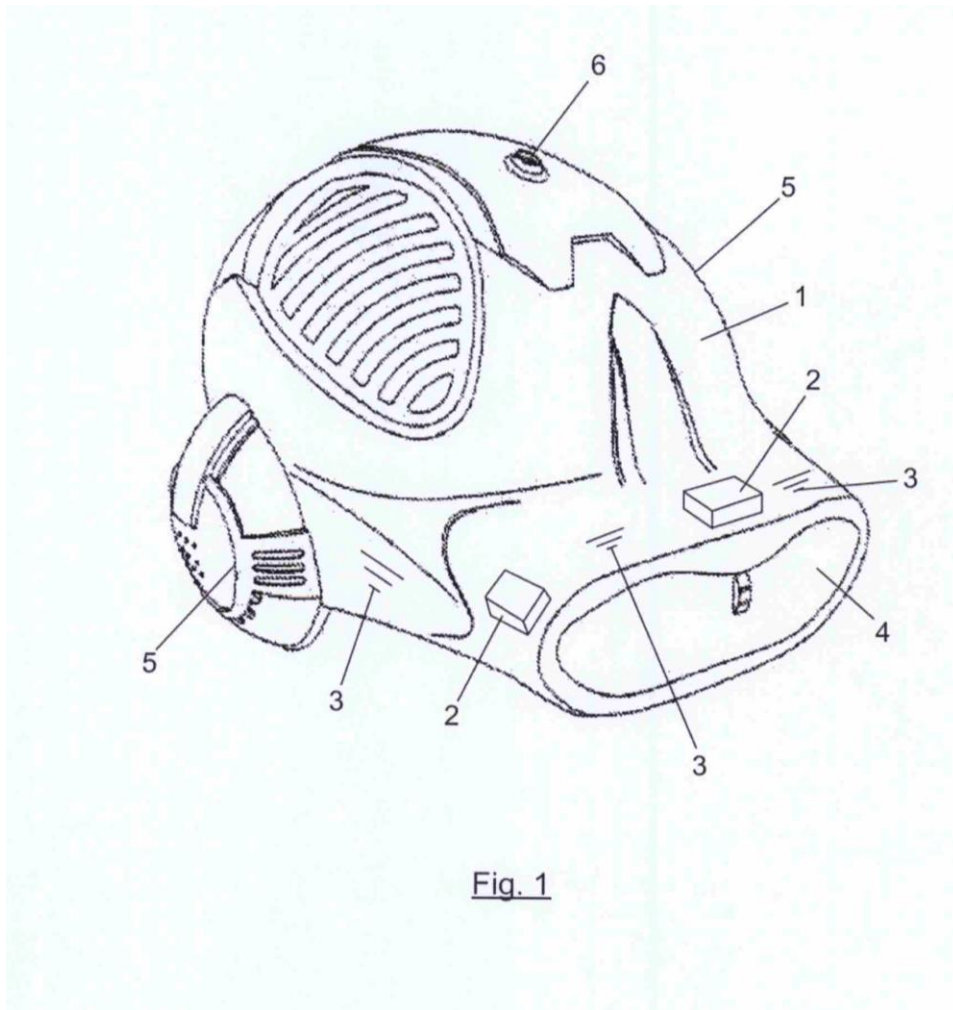
16. Método según la reivindicación 15 **caracterizado** por que para la correlación de los sonidos recibidos se obtienen previamente una serie de espectrogramas correspondientes a los sonidos recibidos.

17. Método según cualquiera de las reivindicaciones 15 ó 16 **caracterizado** por que para la correlación de los sonidos recibidos comprende una comparación con una base de datos de sonidos.

18. Método según la reivindicación 17 **caracterizado** por que la base de datos de sonidos se encuentra en un servidor remoto al dispositivo.

19. Método según cualquiera de las reivindicaciones 15 a 18 **caracterizado** por que la correlación de los sonidos recibidos con una imagen recibida comprenden el procesamiento mediante redes neuronales.

20. Método según cualquiera de las reivindicaciones 15 a 18 **caracterizado** por que la correlación de los sonidos recibidos con una imagen recibida comprenden el procesamiento mediante SVM.
- 5 21. Método según cualquiera de las reivindicaciones 15 a 18 **caracterizado** por que la correlación de los sonidos recibidos con una imagen recibida comprenden el procesamiento mediante perceptrón multicapa.
- 10 22. Método según cualquiera de las reivindicaciones 15 a 18 **caracterizado** por que la correlación de los sonidos recibidos con una imagen recibida comprenden el procesamiento mediante Deep Learning.
- 15 23. Método según cualquiera de las reivindicaciones 15 a 22 **caracterizado** porque la etapa IV comprende la reproducción a través de la al menos una pantalla y/o del al menos un altavoz de al menos una imagen y/o un sonido almacenados en la base de datos adicionales a los obtenidos en la etapa I.
- 20 24. Método según cualquiera de las reivindicaciones 15 a 23 **caracterizado** por que la etapa III comprende la eliminación de una imagen correlacionada con un sonido si se elimina el sonido correspondiente a dicha imagen.
- 25 25. Método según cualquiera de las reivindicaciones 15 a 24 **caracterizado** por que la etapa III comprende la eliminación de un sonido correlacionado con una imagen si se elimina la imagen correspondiente a dicho sonido.
26. Método según cualquiera de las reivindicaciones 15 a 25 **caracterizado** por que comprende la incorporación o asociación de datos de posicionamiento y orientación a los datos procesados en la etapa III.



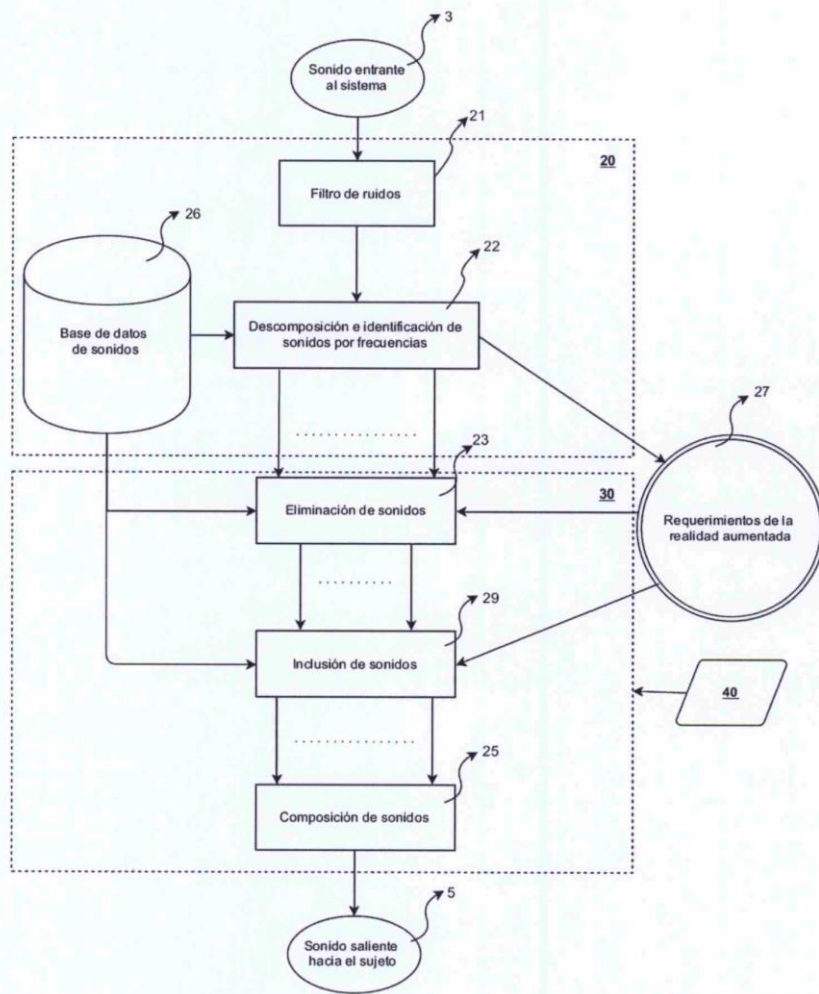


Fig. 2



②① N.º solicitud: 201700463

②② Fecha de presentación de la solicitud: 31.03.2017

③② Fecha de prioridad:

INFORME SOBRE EL ESTADO DE LA TECNICA

⑤① Int. Cl.: Ver Hoja Adicional

DOCUMENTOS RELEVANTES

Categoría	⑤⑥ Documentos citados	Reivindicaciones afectadas
X	US 7676372 B1 (OBA TOSHIHIKO) 09/03/2010, columna 3, línea 12 – columna 6, línea 22; columna 6, línea 61 – columna 7, línea 3; columna 7, líneas 44-53; columna 8, línea 42 – columna 10, línea 2; columna 11, líneas 41-46; columna 12, líneas 25-26; columna 14, línea 59 – columna 16, línea 26; columna 17, líneas 19-26, 34-40; columna 18, líneas 13-51; columna 20, línea 49-56; columna 21, líneas 3-12, 35-65; columna 22, líneas 41-51; columna 24, línea 55 – columna 25, línea 39; columna 25, líneas 50-53, 60-63; columna 26, línea 62 – columna 27, línea 41; columna 28, líneas 24-28; columna 30, líneas 1-51; figuras 1-7	1-11, 14
Y		12-13, 15-26
Y	US 2015279109 A1 (ANDERSON GLEN J et al.) 01/10/2015, párrafos 1, 13, 22-25, 30-40, 45, 48, 53-74, 88-93, 106; figuras 1-6, 8	12-13, 15-26
A	US 4952931 A (SERAGELDIN AHMEDELHADI Y et al.) 28/08/1990, columna 1, líneas 13-21; columna 5, línea 41 – columna 6, línea 39	2
A	Wikipedia. LEARNING VECTOR QUANTIZATION. 26/10/2016 [en línea][Recuperado el 16/10/2017]. Recuperado de Internet <URL: https://en.wikipedia.org/w/index.php?title=Learning_vector_quantization&oldid=746220062 >	3, 19
A	Wikipedia. DEEP LEARNING. 29/03/2017 [en línea][Recuperado el 16/10/2017]. Recuperado de Internet <URL: https://en.wikipedia.org/w/index.php?title=Deep_learning&oldid=772888814 >	4-6, 20-22
A	Sigtia Siddharth et al. AUTOMATIC ENVIRONMENTAL SOUND RECOGNITION: PERFORMANCE VERSUS COMPUTATIONAL COST. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 20161101 IEEE, USA. 01/11/2016, Vol. 24, Nº 11, Páginas 2096 - 2107 [en línea][recuperado el 16/10/2017]. ISSN ISSN 2329-9290, <DOI: doi:10.1109/TASLP.2016.2592698> <URL: http://ieeexplore.ieee.org/abstract/document/7515194/ > Apartado II	4-6, 20-22

Categoría de los documentos citados

X: de particular relevancia

Y: de particular relevancia combinado con otro/s de la misma categoría

A: refleja el estado de la técnica

O: referido a divulgación no escrita

P: publicado entre la fecha de prioridad y la de presentación de la solicitud

E: documento anterior, pero publicado después de la fecha de presentación de la solicitud

El presente informe ha sido realizado

☒ para todas las reivindicaciones

☐ para las reivindicaciones nº:

Fecha de realización del informe
18.10.2017

Examinador
J. M. Vazquez Burgos

Página
1/6

CLASIFICACIÓN OBJETO DE LA SOLICITUD

G10L17/00 (2013.01)**G10L21/10** (2013.01)**G06T1/00** (2006.01)

Documentación mínima buscada (sistema de clasificación seguido de los símbolos de clasificación)

G10L, G06T

Bases de datos electrónicas consultadas durante la búsqueda (nombre de la base de datos y, si es posible, términos de búsqueda utilizados)

INVENES, EPODOC, WPI, INTERNET

Fecha de Realización de la Opinión Escrita: 18.10.2017

Declaración

Novedad (Art. 6.1 LP 11/1986)	Reivindicaciones 2, 4-6, 12-13, 15-26	SI
	Reivindicaciones 1, 3, 7-11, 14	NO
Actividad inventiva (Art. 8.1 LP11/1986)	Reivindicaciones	SI
	Reivindicaciones 1-26	NO

Se considera que la solicitud cumple con el requisito de aplicación industrial. Este requisito fue evaluado durante la fase de examen formal y técnico de la solicitud (Artículo 31.2 Ley 11/1986).

Base de la Opinión.-

La presente opinión se ha realizado sobre la base de la solicitud de patente tal y como se publica.

1. Documentos considerados.-

A continuación se relacionan los documentos pertenecientes al estado de la técnica tomados en consideración para la realización de esta opinión.

Documento	Número Publicación o Identificación	Fecha Publicación
D01	US 7676372 B1 (OBA TOSHIHIKO)	09.03.2010
D02	US 2015279109 A1 (ANDERSON GLEN J et al.)	01.10.2015
D03	US 4952931 A (SERAGELDIN AHMEDELHADI Y et al.)	28.08.1990
D04	Wikipedia. LEARNING VECTOR QUANTIZATION. [en línea][Recuperado el 16/10/2017]. Recuperado de Internet <URL: https://en.wikipedia.org/w/index.php?title=Learning_vector_quantization&oldid=746220062 >	26.10.2016
D05	Wikipedia. DEEP LEARNING. [en línea][Recuperado el 16/10/2017]. Recuperado de Internet <URL: https://en.wikipedia.org/w/index.php?title=Deep_learning&oldid=772888814 >	29.03.2017
D06	Sigita Siddharth et al. AUTOMATIC ENVIRONMENTAL SOUND RECOGNITION: PERFORMANCE VERSUS COMPUTATIONAL COST. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 20161101 IEEE, USA. Vol. 24, Nº 11, Páginas 2096 - 2107 [En línea][recuperado el 16/10/2017]. ISSN 2329-9290, <DOI: doi:10.1109/TASLP.2016.2592698><URL: http://ieeexplore.ieee.org/abstract/document/7515194/ >	01.11.2016

2. Declaración motivada según los artículos 29.6 y 29.7 del Reglamento de ejecución de la Ley 11/1986, de 20 de marzo, de Patentes sobre la novedad y la actividad inventiva; citas y explicaciones en apoyo de esta declaración

El documento del estado de la técnica más próximo a la invención es D01 y divulga un sistema de realidad aumentada para la ayuda a personas discapacitadas, configurado para ser portado en la cabeza, y para reconocer sonidos y locuciones, que son identificados visualmente mediante caracteres o imágenes.

Reivindicación 1

Con el fin de ilustrar de la manera más clara posible la diferencia entre la invención reivindicada en 1 y el documento D01 del estado de la técnica más próximo, se reproduce seguidamente el texto de dicha reivindicación, desprovisto de sus referencias originales si las tuviese, e insertando donde proceda las de D01. Asimismo, aquellas partes del mismo que pudieran no estar comprendidas en D01 se resaltarían entre corchetes y en negrita.

Dispositivo de montaje sobre la cabeza (columna 3, líneas 12-22) de un usuario para la percepción de realidad aumentada (columna 25, líneas 50-62) que comprende:

- medios de montaje sobre la cabeza del usuario (10; columna 4, líneas 9-18);
- al menos una pantalla (7; columna 3, líneas 23-24, 30-35);
- al menos un micrófono (8, 11; columna 3, líneas 36-58);
- al menos un altavoz (9, 25; columna 3, líneas 59-62, columna 9, líneas 12-18); y
- un módulo de procesamiento de datos (3; columna 4, líneas 19-29), dicho módulo comprendiendo a su vez un sub-módulo de tratamiento de imágenes (22; columna 21, líneas 59-65), un sub-módulo de tratamiento de sonidos (22; columna 5, líneas 4-23) y un sub-módulo de salida de datos (23; columna 6, líneas 23-32),

Caracterizado por que el sub-módulo de tratamiento de sonidos comprende medios de descomposición de los sonidos recibidos mediante el al menos un micrófono en una serie de espectrogramas y medios de correlación de dichos espectrogramas con espectrogramas de una biblioteca a fin de identificar qué entidad física produce el sonido (columna 5, líneas 31-33; columna 15, línea 35 – columna 16, línea 26; columna 18, líneas 21-51).

De acuerdo con lo anterior, cabe concluir que, a la luz del documento D01, la invención reivindicada en 1 no es nueva, conforme dicho requisito de novedad se define en el artículo 6 de la Ley de patentes de 1986.

Reivindicaciones 2 a 14

La técnica de clasificación objeto de la reivindicación 2, en cuanto a su empleo para la identificación de sonidos, forma parte del conocimiento general común, como ilustra el documento D03 (columna 1, líneas 13-21; columna 5, línea 41 □ columna 6, línea 39).

La técnica objeto de la reivindicación 3 está comprendida en D01, por cuanto este incluye la técnica de LVQ (columna 18, líneas 21-36), que constituye un caso particular de las redes neuronales, como ilustra el documento D04 (primer párrafo).

Las técnicas objeto de las reivindicaciones 4 a 6, en cuanto a su empleo para la identificación de sonidos, forman parte del conocimiento general común, como ilustran los documentos D05 y D06 (apartado II).

La inclusión de una biblioteca de imágenes asociada a los espectrogramas, que incluya imágenes de estos, reivindicada en 7, está incluida en D01 (columna 6, líneas 1-4, 11-22).

Los medios de eliminación de sonidos reivindicados en 8, y los de adición de estos, objeto de 9, se encuentran incluidos en D01 (columna 14, líneas 59-64; columna 20, líneas 49-56).

Los medios reivindicados en 10 son sinónimos de una configuración de los medios de procesamiento del sub-módulo de salida de datos para seleccionar espectrogramas e imágenes, así como para transmitir dichas imágenes a una pantalla o a un altavoz. Todo ello está incluido en D01 (columna 6, líneas 18-32; 61-67).

Con respecto al objeto de la reivindicación 11, D01 muestra una realización con una misma unidad (22) para el soporte de los procesos de reconocimiento de audio y de vídeo, conectada a la unidad de salida de datos (23), si bien admite la posibilidad de hacerlo mediante diferentes unidades de procesamiento (columna 24, línea 55; columna 25, línea 11). Por otra parte, con relación a la configuración de la unidad de salida de datos para correlar sonido e imágenes, esta se encuentra recogida en D01 (columna 21, líneas 59-65).

El objeto de la reivindicación 12 no está incluido en D01, ya que este divulga una solución donde es posible añadir información de vídeo a la de audio o a la derivada de ella, pero no la supresión de la primera en función de la de la segunda. El efecto técnico que ello tiene es el de no poder definir filtros de información relacionales (vinculados entre diferentes fuentes de información, como audio y vídeo). El problema técnico objetivo a resolver sería el disponer de filtros relacionales de información mostrada en una solución como la divulgada en D01. El documento D02 presenta un sistema de realidad aumentada, donde es posible añadir o suprimir objetos (párrafos 13, 30, 33-38, 45), pudiendo estos incluir información tanto de audio como de vídeo (párrafos 24-25). Esto es, es posible filtrar o modificar la información mostrada, no ya en función de la fuente de la información (vídeo o audio), sino de la etiqueta o concepto asociada a ellos (objeto). En este sentido, el documento D01 muestra una solución en la que es posible identificar objetos mediante sus emisiones de audio, pudiendo asociarse a ellos imágenes. D02 a su vez, muestra una solución enfocada al mismo tipo de sistemas (de realidad aumentada), mediante la cual es posible ligar audio y vídeo a un objeto, e insertar o eliminar este de la información mostrada por el sistema. Un experto en la materia, a la vista de las enseñanzas de ambos documentos, no requeriría de actividad inventiva para, combinando D01 con las partes relevantes de D02, obtener las características reivindicadas en 12 con una expectativa razonable de éxito.

Razonando de manera similar, puede aplicarse al objeto de la reivindicación 13 la misma conclusión que a la 12.

El sub-módulo de posicionamiento y orientación, reivindicado en 14 se encuentra recogido en D01 (columna 26, línea 62; columna 27, línea 32).

En consecuencia, de las consideraciones anteriores, y una vez tenidas en cuenta las correspondientes relaciones de dependencia, cabe concluir que:

A la luz de D01, las invenciones reivindicadas en 3, 7 a 11 y 14 no son nuevas, conforme el requisito de novedad se define en el artículo 6 de la Ley de Patentes de 1986.

A la luz de D01, las invenciones reivindicadas en 2 y 4 a 6 no reúnen el requisito de actividad inventiva, conforme este se define en el artículo 8 de la Ley de Patentes de 1986.

A la luz de la combinación de D01 con D02, las invenciones reivindicadas en 12 y 13 no reúnen el requisito de actividad inventiva, conforme este se define en el artículo 8 de la Ley de Patentes de 1986.

Reivindicación 15

Con el fin de ilustrar de la manera más clara posible la diferencia entre la invención reivindicada en 15 y el documento D01 del estado de la técnica más próximo, se reproduce seguidamente el texto de dicha reivindicación, desprovisto de sus referencias originales si las tuviese, e insertando donde proceda las de D01. Asimismo, aquellas partes del mismo que pudieran no estar comprendidas en D01 se resaltarían entre corchetes y en negrita.

Con el fin de ilustrar de la manera más clara posible la diferencia entre la invención reivindicada en 1 y el documento D01 del estado de la técnica más próximo, se reproduce seguidamente el texto de dicha reivindicación, desprovisto de sus referencias originales si las tuviese, e insertando donde proceda las de D01. Asimismo, aquellas partes del mismo que pudieran no estar comprendidas en D01 se resaltarían entre corchetes y en negrita.

Método de generación de un entorno de realidad aumentada en un dispositivo de acuerdo a cualquiera de las reivindicaciones 1 a 14 caracterizado por que comprende las etapas de:

- i. obtención de imágenes a través de al menos una cámara (columna 7, líneas 44-46);
- ii. obtención de sonidos a través de al menos un micrófono (columna 3, líneas 36-44);
- iii. procesamiento de datos que a su vez comprende la correlación de al menos uno de los sonidos recibidos con una imagen recibida correspondiente a la entidad física que emite dicho sonido y la eliminación del sonido (columna 20, líneas 49-56; columna 21, líneas 49-65) **[junto con la imagen correspondiente a la entidad física]**; y
- iv. reproducción a través de la al menos una pantalla y/o del al menos un altavoz de las imágenes y/o sonidos que no han sido eliminadas en la etapa III (columna 8, línea 65 – columna 9, línea 18).

La diferencia entre el documento D01 del estado de la técnica más próximo y la invención reivindicada en 15, es que el primero divulga una solución donde es posible añadir información de vídeo a la de audio o a la derivada de ella, pero no la supresión de la primera en función de la de la segunda, como ocurre en la invención reivindicada. El efecto técnico que ello tiene es el de no poder definir filtros de información relacionales (vinculados entre diferentes fuentes de información, como audio y vídeo). El problema técnico objetivo a resolver sería el disponer de filtros relacionales de información mostrada en una solución como la divulgada en D01. El documento D02 presenta un sistema de realidad aumentada, donde es posible añadir o suprimir objetos (párrafos 13, 30, 33-38, 45), pudiendo estos incluir información tanto de audio como de vídeo (párrafos 24-25). Esto es, es posible filtrar o modificar la información mostrada, no ya en función de la fuente de la información (vídeo o audio), sino de la etiqueta o concepto asociada a ellos (objeto). En este sentido, el documento D01 muestra una solución en la que es posible identificar objetos mediante sus emisiones de audio, pudiendo asociarse a ellos imágenes. D02 a su vez, muestra una solución enfocada al mismo tipo de sistemas (de realidad aumentada), mediante la cual es posible ligar audio y vídeo a un objeto, e insertar o eliminar este de la información mostrada por el sistema. Un experto en la materia, a la vista de las enseñanzas de ambos documentos, no requeriría de actividad inventiva para, combinando D01 con las partes relevantes de D02, obtener las características reivindicadas en 15 con una expectativa razonable de éxito.

De las consideraciones anteriores cabe en consecuencia concluir que, a la luz de la combinación de D01 con D02, la invención reivindicada en 15 no reúne el requisito de actividad inventiva, conforme este se define en el artículo 8 de la Ley de Patentes de 1986.

Reivindicaciones 16 a 26

Los objetos de las reivindicaciones 16 y 17 están incluidos en D01 (columna 18, líneas 21-51).

El objeto de 18 está también incluido en D01 (columna 30, líneas 26-36), lo mismo que el de 19 (columna 18, líneas 21-36). Con respecto a las técnicas reivindicadas en 20 a 22 pueden hacerse similares consideraciones a las realizadas para las reivindicaciones 4 a 6 respectivamente.

El objeto de la reivindicación 23 también forma parte de D01 (columna 15, línea 53 □ columna 16, línea 26).

Con respecto a los objetos de las reivindicaciones 24 y 25 pueden hacerse las mismas consideraciones que sobre los de las reivindicaciones 12 y 13 respectivamente, con idénticas conclusiones.

La asociación de datos de posicionamiento reivindicada en 26 está incluida en D01 (columna 27, líneas 33-41), así como en D02 (párrafos 53-54, 56).

En consecuencia, de las consideraciones anteriores, y una vez tenidas en cuenta las correspondientes relaciones de dependencia, cabe extraer la conclusión de que, a la luz de la combinación de D01 con D02, las invenciones reivindicadas en 16 a 26 no reúnen el requisito de actividad inventiva, conforme este se define en el artículo 8 de la Ley de Patentes de 1986.