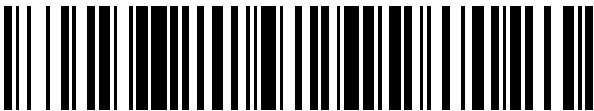


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 995 010**

21 Número de solicitud: 202430620

51 Int. Cl.:

B60W 30/095 (2012.01)
B60W 60/00 (2010.01)
B60W 50/00 (2006.01)
H04R 1/08 (2006.01)
H04R 1/32 (2006.01)

12

SOLICITUD DE PATENTE

A1

22 Fecha de presentación:
22.07.2024

43 Fecha de publicación de la solicitud:
05.02.2025

71 Solicitantes:
UNIVERSITAT POLITÈCNICA DE VALÈNCIA
(100.00%)
Camí de VERA, s/n
46022 VALENCIA (Valencia) ES

72 Inventor/es:
LÓPEZ MONFORT, José Javier y
BANCHERO MARTÍNEZ, Lucas

74 Agente/Representante:
SÁNCHEZ MARGARETO, Carolina

54 Título: **Dispositivo y procedimiento de asistencia a la conducción**

57 Resumen:

Dispositivo de asistencia a la conducción, en un vehículo (2), que incorpora una serie de micrófonos (1), una unidad de control (5) que ejecuta una programación de reconocimiento de sonidos para comprobar la presencia de otro vehículo. El sistema comprende micrófonos (1) dispuestos en las esquinas del vehículo (2) en sendas estructuras aerodinámicas (3).

El procedimiento de asistencia a la conducción, con el dispositivo de las reivindicaciones anteriores comprende las etapas de preprocesado y de análisis mediante redes neuronales.

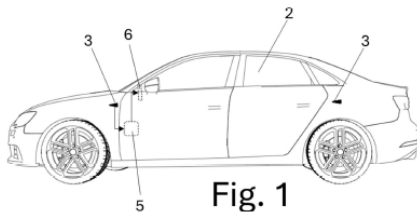


Fig. 1

DESCRIPCIÓN

Dispositivo y procedimiento de asistencia a la conducción

5 SECTOR DE LA TÉCNICA

La presente invención se refiere a un dispositivo que define un sistema de detección y localización de sonidos para asistencia a la conducción. La información así recibida se transmite al conductor para que tome las decisiones correctas.

10

ESTADO DE LA TÉCNICA

15

La conducción segura depende de la percepción visual del conductor, pero también de su capacidad para detectar y responder adecuadamente a señales auditivas, como las sirenas de vehículos de emergencia, claxon de vehículos, etc.

20

En muchas situaciones donde un vehículo de emergencia trata de abrirse paso en el tráfico, los conductores no perciben con claridad el sonido, y casi nunca son capaces de distinguir la dirección por donde les llega. Esto se debe al cada vez mejor aislamiento interior de los vehículos, o también a la reproducción de música o radio en su interior.

25

Estas situaciones pueden generar estrés y reacciones erróneas del conductor. En algunos vehículos de última generación se usan sistemas visuales por cámaras, pero solo sirven para detectar vehículos muy próximos y no son útiles para los casos de uso mencionados. Los sistemas ADAS de información auditiva precisa, podrían reducir el riesgo de accidentes y aumentarla seguridad vial.

30

Se conoce en el estado de la técnica el uso de sistemas de micrófonos para reconocer sonidos y su origen, asociados a la conducción. Se pueden apreciar ejemplos en US2021358300, US2019294169 o US2022284919. Estas invenciones no se han incorporado a sistemas ADAS.

35

La presente invención aborda, por tanto, el desarrollo de una solución integral que complementa el resto de sistemas ADAS de un vehículo, brindando al conductor información auditiva relevante para una conducción más segura y eficiente.

El solicitante no conoce ningún dispositivo que pueda ser considerado similar a la invención.

BREVE EXPLICACIÓN DE LA INVENCION

5

La invención consiste en un dispositivo de asistencia a la conducción según las reivindicaciones independientes y cuyas variantes resuelven los problemas del estado de la técnica.

10 El dispositivo de la invención es un sistema integral de hardware y software diseñado para mejorar las capacidades de Asistencia Avanzada a la Conducción (ADAS) en vehículos. El sistema complementa los sistemas visuales existentes del vehículo al agregar capacidades de detección y análisis de sonido en tiempo real. Consiste en la instalación de micrófonos estratégicamente ubicados alrededor del vehículo, los cuales capturan
15 sonidos del entorno circundante. Estos datos de sonido son procesados por un hardware específico y un software dedicado, que analiza la dirección, distancia, movimiento y proximidad de las fuentes de sonido, particularmente aquellos de interés como sirenas de vehículos de emergencia u otros eventos relevantes para la seguridad en la conducción.

20 En concreto, el dispositivo de asistencia a la conducción, en un vehículo, incorpora una serie de micrófonos, una unidad de control que ejecuta una programación de reconocimiento de sonidos para comprobar la presencia de otro vehículo. El sistema comprende micrófonos dispuestos en las esquinas del vehículo en sendas estructuras aerodinámicas. En la realización más preferida, comprende un único micrófono en cada
25 esquina.

Preferiblemente, las estructuras aerodinámicas comprenden un relleno de material absorbente.

30 Preferiblemente, los orificios de entrada a las estructuras aerodinámicas disponen de filtros antiviento.

El procedimiento de asistencia a la conducción que usa el dispositivo de las reivindicaciones anteriores comprende las etapas de:

35 Una etapa de preprocesado de la señal digitalizada para reducir el ruido de viento y el ruido propio de la rodadura.

Cálculo del espectrograma de MEL de las señales de los micrófonos.

Análisis mediante una primera red neuronal profunda de transformers para la inferencia de la detección del sonido, incluyendo el cálculo del GCC-PHAT de los distintos pares de micrófonos.

- 5 Tratamiento mediante una segunda red neuronal convolucional profunda del resultado del cálculo del GCC-PHAT para la inferencia de la localización del sonido.

Revisión mediante un algoritmo de estimación del movimiento de la fuente sonora.

Transmisión del resultado al usuario.

- 10 Otras variantes se aprecian en el resto de la memoria.

DESCRIPCIÓN DE LAS FIGURAS

- 15 Para una mejor comprensión de la invención, se incluyen las siguientes figuras, que muestran formas de realización ejemplares.

Figura 1: Vista de un vehículo donde se ha instalado un ejemplo de realización de la invención, de forma cónica.

- 20 Figura 2: Ejemplos de estructuras aerodinámicas: (A) tubular, (B) semiesférica.

Figura 3: Corte esquemático de un ejemplo de estructura aerodinámica.

Figura 4: Esquema del procedimiento.

25

MODOS DE REALIZACIÓN DE LA INVENCION

A continuación, se pasa a describir de manera breve un modo de realización de la invención, como ejemplo ilustrativo y no limitativo de esta.

30

- El dispositivo de la realización de la figura 1 parte de una serie de equipos (hardware) y una programación (software). Los equipos parten de cuatro micrófonos (1) dispuestos aproximadamente en las esquinas del vehículo (2). Los micrófonos (1) están dispuestos en estructuras aerodinámicas (3), de forma que se reduce tanto el ruido del viento como el producido por las vibraciones del vehículo en movimiento. Los micrófonos (1) están
35 conectados a convertidores (4) que digitalizan las señales y que las remiten a una unidad

de control (5) que ejecuta la programación en tiempo real. Una pantalla (6) accesible al usuario muestra las conclusiones extraídas por la programación a partir de las señales acústicas. Es decir, está configurada para mostrar la información del vehículo detectado por los micrófonos (1).

5

La programación de tratamiento de las señales parte de una etapa de preprocesado (101) de la señal para reducir aún más el ruido de viento y eliminar el ruido propio de la rodadura. A continuación, se realiza el cálculo (102) del espectrograma de MEL de las señales de los micrófonos (1). Ese espectrograma se trabaja mediante una primera red neuronal (103) profunda de transformers para la inferencia de la detección del sonido, decidiendo si algún sonido puede ser relevante. Esta primera red neuronal (103) permite seleccionar los ruidos que se consideran de interés, descartando el resto. Los ruidos considerados continúan las etapas del procedimiento. En caso afirmativo, se reconoce el tipo de sonido, para mostrarlo en la pantalla (6). En paralelo, se realiza un inicio (108) del estudio del origen del sonido (distancia, dirección desde donde proviene y si la fuente es móvil, su movimiento). El estudio se realiza mediante una red Encoder-Decoder (107) que lleva a cabo un proceso de eliminación del ruido de la señal de interés. Posteriormente se procesa el cálculo del GCC-PHAT (104) entre cada posible par de micrófonos, asegurando que se realiza para todos los posibles pares de micrófonos (seis combinaciones en caso de cuatro micrófonos).

El resultado del cálculo del GCC-PHAT (104) se trata mediante una segunda red neuronal (105) convolucional profunda para la inferencia de la localización del sonido. Esta inferencia se revisa mediante un algoritmo de estimación del movimiento de la fuente sonora (106) y se muestra en la pantalla (6) con la interfaz adecuada.

En un ejemplo preferido de realización, las inteligencias artificiales comprenden

1. Detector: Audio Spectrogram Transformer (AST), adaptando las últimas capas para tener dos capas dense. La primera una dense de 768 neuronas activadas con una función PRelu. La segunda con 3 neuronas, activadas con la función SoftMax. Se realizó un entrenamiento de 100 epochs con un early stop de 10 epoch de paciencia.
2. Limpiador: Se trata de una Red Encoder-Decoder. Las redes Encoder-Decoder son un tipo de arquitectura de red neuronal profunda que se utilizan ampliamente en tareas donde es necesario transformar una entrada en una salida de una forma

estructurada. Son especialmente efectivas para problemas donde se requiere un mapeo complejo entre la entrada y la salida, como en la traducción automática, la generación de texto, y el procesamiento de señales, incluyendo la limpieza de espectrogramas de audio, que es nuestro caso. La arquitectura Encoder-Decoder consta de dos partes principales:

- Encoder (Codificador):
 - a. Toma la entrada (en este caso, un espectrograma de audio ruidoso) y la transforma en una representación latente de dimensión más baja.
 - b. Este proceso implica varias capas (pueden ser capas convolucionales, recurrentes, o de transformers) que comprimen la información relevante en una representación compacta.
- Decoder (Decodificador):
 - a. Toma la representación latente generada por el encoder y la transforma de nuevo en una forma comprensible (en este caso, un espectrograma de audio limpio).
 - b. De igual manera, implica varias capas que expanden la representación latente para reconstruir la salida deseada.

Un ejemplo de red encoder-decoder tipo U-NET se puede apreciar en O. Ronneberger, P. Fischer, and T. Brox, "*U-net: Convolutional networks for biomedical image segmentation*", in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2015. doi: 10.1007/978-3-319-24574-4_28.

La red tiene como entrada y salida matrices de 2D, creando un cuello de botella donde se concentra la información extraída, y posteriormente como se vuelve a ampliar dicha información a raíz del cuello de botella.

3. Localizador: Se trata de una red ResNet18, donde se han adaptado las capas finales. A la salida de la ResNet se le conecta una Dense Layers de 512 neuronas activadas con la función PRelu. Posteriormente dicha capa se conecta a dos capas más en paralelo. Cada una de estas redes paralelas son otras Dense Layers de 512 neuronas con activación PRelu. Por último, a la salida de cada una de las redes paralelas se le conecta una activación de 1 neurona que aplica el arco tangente de los valores de las redes paralelas.

4. Estimación de movimiento: Con los valores de localización a lo largo del tiempo se calcula el movimiento que lleva el sonido.

A continuación, se aumentará el nivel de detalle de la explicación.

5

Los micrófonos (1) se distribuyen estratégicamente, colocando cada uno en, aproximadamente, cada una de las cuatro esquinas del vehículo. Esta disposición permite al sistema utilizar solo cuatro micrófonos (1), lo que conlleva una significativa reducción en la cantidad de datos a procesar y un aumento de la velocidad de respuesta. Esta optimización es fundamental en aplicaciones en tiempo real, mejorando la eficiencia y la velocidad de los diferentes sistemas involucrados.

Los micrófonos (1) deben ser capaces de soportar las condiciones adversas de temperatura, humedad... propias de la intemperie. La energía de funcionamiento se obtendrá del propio vehículo, para lo que reciben un cable de alimentación y envían una señal del sonido, generalmente también por cableado.

Los micrófonos (1) van alojados dentro de sendas estructuras aerodinámicas (3), especialmente diseñadas para minimizar la resistencia al viento y reducir las vibraciones. Esta estructura aerodinámica (3), hecha con cualquier material impermeable, puede adoptar diversas formas geométricas, entre ellas esferas o conos. En la figura 2 se muestran dos ejemplos preferidos. Se aprecia la presencia de una muesca para el paso de cableado.

Las estructuras aerodinámicas (3) mejoran la calidad de captura del sonido. Para ello, las estructuras aerodinámicas (3) se rellenan con un material absorbente (31). Este material absorbente (31) es un aislante termoacústico con coeficiente de absorción acústico mayor a 0,9 y el aislamiento acústico a ruido aéreo superior a 60 dB. La densidad preferida es entre 70 y 90 kg/m³ para sostener el micrófono de forma efectiva, sin necesidad de otra estructura. Por seguridad, es deseable que su reacción al fuego sea B S1 d0 y su resistencia térmica desde 1,00 m²K/W. Un ejemplo de material adecuado es la espuma aglomerada de poliuretano. Este material absorbente (31) ayuda a reducir aún más los efectos de las vibraciones del vehículo y del ruido provocado por el viento.

Los orificios (32) de entrada a las estructuras aerodinámicas (3) disponen de filtros antiviento (33), que deja pasar el sonido, pero no deja pasar el viento. Se compone de

una o más capas de material acústicamente transparente, como por ejemplo tejido de nylon tensado sobre un montaje de soporte, para eliminar casi completamente el efecto del viento en el sistema, asegurando así la precisión en la detección de sonidos de interés en la carretera.

5

La digitalización de las señales, preamplificadas en caso necesario, se digitalizan por convertidores (4) analógico/digital (ADC) a una determinada frecuencia de muestreo (f_s). Se toma como base una $f_s=16000$ Hz, pero la invención soporta cualquier frecuencia de muestreo, superior o inferior a ésta. Los convertidores (4) pueden estar en la propia estructura aerodinámica (3) o asociados a la unidad de control (5).

10

La unidad de control (5) llevará a cabo los siguientes pasos: algoritmos de reducción de ruido, análisis de las muestras de audio por tramas para la extracción de características de los sonidos, ejecución de inferencia de los algoritmos de detección y localización de fuentes sonoras mediante modelos de inteligencia artificial.

15

Es crucial que este sistema funcione de manera óptima en tiempo real, por lo que requerirá una capacidad de procesamiento muy elevada en términos de número de operaciones en coma flotante (FLOPs). Dado que será el núcleo central de todo el sistema, su desempeño será fundamental para el funcionamiento global del vehículo. En general la unidad de control (5) será capaz de realizar procesamiento paralelo profundo, por lo que arquitectura basadas en todo o en parte en GPU son recomendables. Al usar sólo cuatro micrófonos (1), se limita también las necesidades de procesamiento.

20

La unidad de control (5) puede ser un componente separado del ordenador central a bordo del vehículo, o bien puede integrarse dentro del mismo, dependiendo de la arquitectura y las especificaciones del sistema en particular. En vehículos con un sistema computador potente con procesamiento paralelo, basado en todo o parte en GPU, que disponga de tiempo de procesador libre, la unidad de control (5) podría integrarse en el mismo.

30

Una vez procesada la información, el sistema emite los resultados hacia una representación visual o sonora dentro del vehículo para la alerta del conductor. Estos medios de alerta incluirán información gráfica en pantallas (6) del tablero de instrumentos, y opcionalmente alertas auditivas para el conductor utilizando los altavoces del sistema de sonido del vehículo. Adicionalmente se pueden incluir actuaciones automáticas en el

35

comportamiento del vehículo en respuesta a los sonidos detectados. De esta manera, el ordenador central desempeña un papel crucial en mejorar la percepción y la seguridad del vehículo en su entorno auditivo.

REIVINDICACIONES

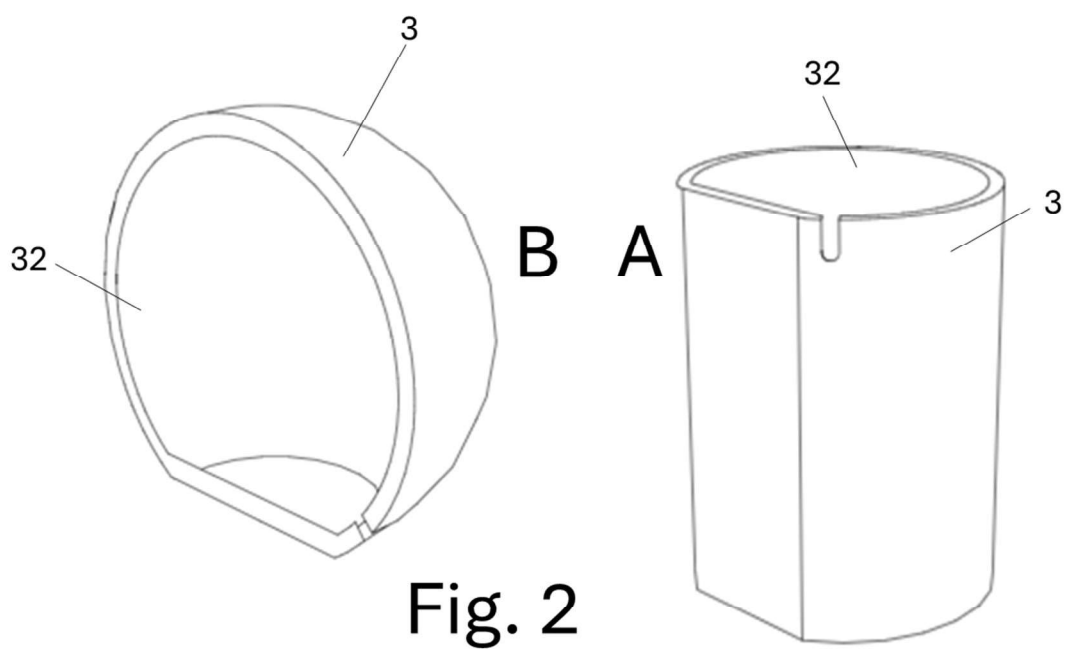
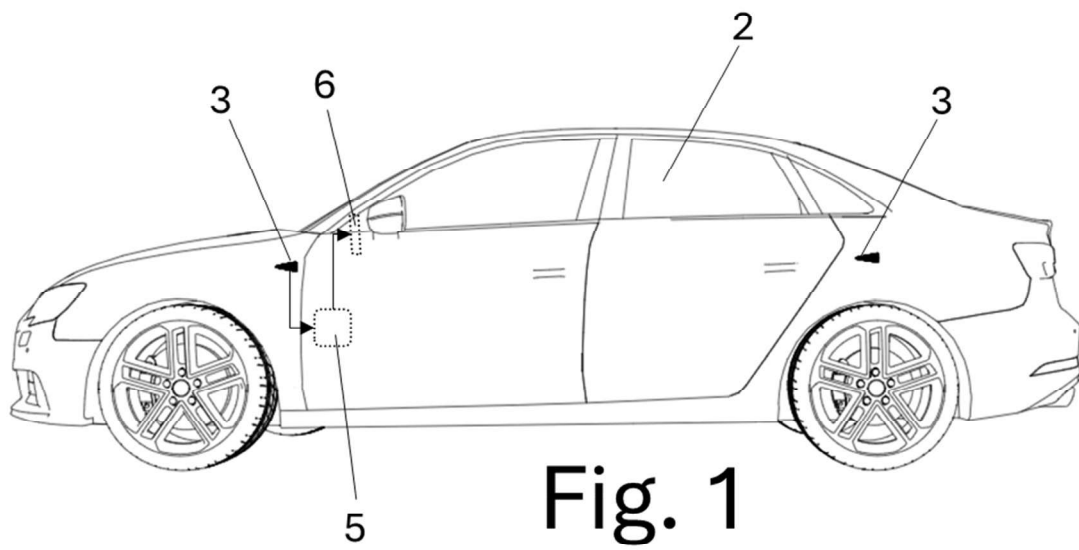
- 1- Dispositivo de asistencia a la conducción, en un vehículo (2), que incorpora una serie
5 de micrófonos (1), una unidad de control (5) que ejecuta una programación de
reconocimiento de sonidos para comprobar la presencia de otro vehículo, caracterizado
por que el sistema comprende micrófonos (1) dispuestos en las esquinas del vehículo (2)
en sendas estructuras aerodinámicas (3).
- 10 2- Dispositivo de asistencia a la conducción, según la reivindicación 1, caracterizado por
que consiste en un único micrófono (1) en cada esquina.
- 3- Dispositivo de asistencia a la conducción, según la reivindicación 1, caracterizado por
que las estructuras aerodinámicas (3) comprenden un relleno de material absorbente
15 (31).
- 4- Dispositivo de asistencia a la conducción, según la reivindicación 1, caracterizado por
que los orificios (32) de entrada a las estructuras aerodinámicas (3) disponen de filtros
antiviento (33).
- 20 5- Dispositivo de asistencia a la conducción, según la reivindicación 1, caracterizado por
que cada micrófono (1) comprende un convertidor (4) analógico/digital en la propia
estructura aerodinámica (3).
- 25 6- Dispositivo de asistencia a la conducción, según la reivindicación 1, caracterizado por
que la unidad de control (5) está integrada en el ordenador central a bordo del vehículo
(2) y conectada a una pantalla (6) accesible al usuario configurada para mostrar la
información del vehículo detectado.
- 30 7- Procedimiento de asistencia a la conducción, con el dispositivo de las reivindicaciones
anteriores, caracterizado por que comprende las etapas de:
una etapa de preprocesado (101) de la señal para reducir el ruido de viento y el ruido
propio de la rodadura;
cálculo (102) del espectrograma de MEL de las señales de los micrófonos (1);

análisis mediante una primera red neuronal (103) profunda de transformers para la inferencia de la detección del sonido, incluyendo el cálculo del GCC-PHAT (104) de los distintos pares de micrófonos;

5 tratamiento mediante una segunda red neuronal (105) convolucional profunda del resultado del cálculo del GCC-PHAT (104) para la inferencia de la localización del sonido;

revisión mediante un algoritmo de estimación del movimiento de la fuente sonora (106);
transmisión del resultado al usuario.

10



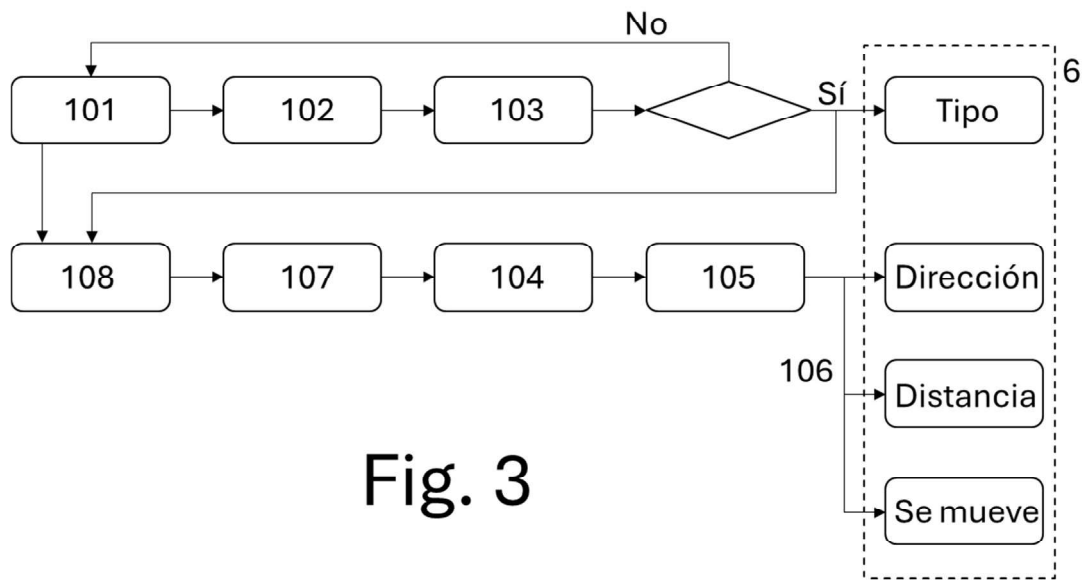


Fig. 3

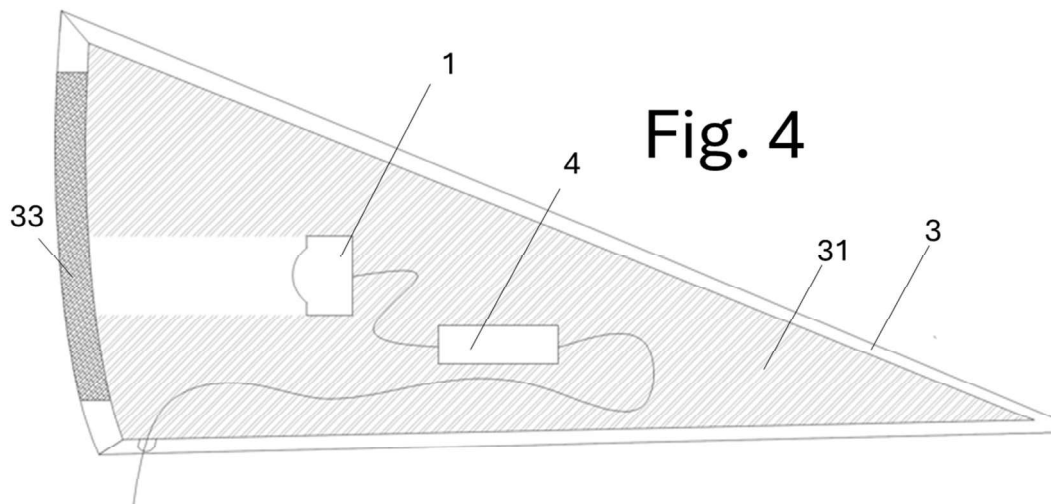


Fig. 4



21 N.º solicitud: 202430620
22 Fecha de presentación de la solicitud: 22.07.2024
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 11242098 B2 (PATNAIK VIJAYSAI et al.) 08/02/2022, todo el documento.	1-7
A	US 2016297437 A1 (HARA YASUHIRO et al.) 13/10/2016, todo el documento.	1, 3
A	US 11889278 B1 (WU CHENG-HAN et al.) 30/01/2024, todo el documento.	1
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 22.01.2025	Examinador G. Barrera Bravo	Página 1/2

CLASIFICACIÓN OBJETO DE LA SOLICITUD

B60W30/095 (2012.01)
B60W60/00 (2020.01)
B60W50/00 (2006.01)
H04R1/08 (2006.01)
H04R1/32 (2006.01)

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

B60W, H04R

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

SEARCH TOOL DATABASES