



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

① Número de publicación: **2 347 517**

② Número de solicitud: 201031215

⑤ Int. Cl.:
H04R 25/00 (2006.01)
G06F 3/01 (2006.01)

⑫ PATENTE DE INVENCION CON EXAMEN PREVIO

B2

⑫ Fecha de presentación: **04.08.2010**

⑬ Fecha de publicación de la solicitud: **29.10.2010**

Fecha de la concesión: **06.05.2011**

⑭ Fecha de anuncio de la concesión: **18.05.2011**

⑮ Fecha de publicación del folleto de la patente:
18.05.2011

⑰ Titular/es: **Universidad Politécnica de Madrid
c/ Ramiro de Maetzu, 7
28040 Madrid, ES**

⑱ Inventor/es: **Pardo Muñoz, José Manuel;
Córdoba Herralde, Ricardo de;
Ferreiros López, Javier;
Montero Martínez, Juan Manuel y
San Segundo Hernández, Rubén**

⑲ Agente: **Carvajal y Urquijo, Isabel**

⑳ Título: **Método y sistema para incorporar información acústica binaural en un sistema visual de realidad aumentada.**

㉑ Resumen:

Método y sistema para incorporar información acústica binaural en un sistema visual de realidad aumentada, que comprende captar señales acústicas binaurales; digitalizar las señales acústicas captadas obteniendo una pluralidad de muestras de señal; agrupar N muestras consecutivas en tramas; extraer, de cada trama, una pluralidad de características que definen dicha trama; seleccionar un subconjunto de las características extraídas; detectar mediante un reconocimiento de patrones el tipo de evento acústico ocurrido; estimar la distancia a la que se encuentra la fuente del evento acústico detectado; calcular la dirección en la que se sitúa la fuente del evento acústico; localizar la fuente del evento acústico en la escena; codificar visualmente la información del evento acústico y generar un vídeo a partir de una secuencia de imágenes; superponer la secuencia de vídeo generada con las imágenes captadas por al menos una cámara de vídeo (13); y mostrar la señal visual combinada.

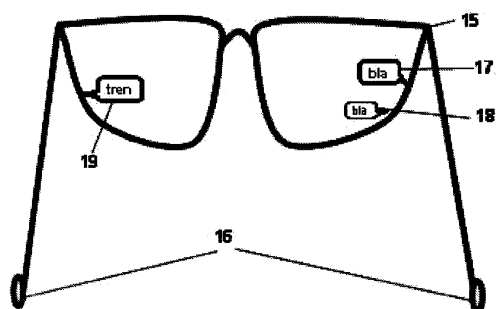


Fig. 2

ES 2 347 517 B2

Aviso: Se puede realizar consulta prevista por el art. 40.2.8 LP.

DESCRIPCIÓN

Método y sistema para incorporar información acústica binaural en un sistema visual de realidad aumentada.

5 **Campo de la invención**

La presente invención pertenece al campo técnico de las tecnologías de la información y las comunicaciones para el desarrollo de sistemas que mejoran la calidad de vida de las personas con discapacidad.

10 **Estado de la técnica**

En la actualidad hay una gran cantidad de trabajos relacionados con la percepción binaural que abordan tanto el problema de localización de la fuente de los eventos acústicos, como los problemas de clasificación de sonidos o incluso reconocimiento de habla. Sin embargo, en estos trabajos no se plantea el problema de clasificar una gran cantidad de tipos de sonidos sino que se centran en la diferenciación entre voz o no voz para saber qué parte de la señal acústica se debe mandar a un reconocedor de voz.

Las solicitudes WO2007063139 A2 y WO2009072040 A1 contemplan la incorporación de información acústica, obtenida de la percepción binaural, para mejorar un sistema de realidad aumentada de audio (no visual) para personas con deficiencias auditivas. En estas solicitudes no se plantea ningún tipo de clasificación del tipo de ruido sino que se proponen algoritmos generales de mejora de las señales acústicas. Su objetivo consiste en mejorar la percepción del habla, rechazando aquellos tipos de eventos acústicos que no sean habla. Por tanto, dichas patentes no detectan el tipo de evento acústico para informar al usuario, únicamente rechazan lo que no sea habla.

Es por tanto deseable un método y un sistema que integre la información que reciba mediante señales binaurales, de forma que partiendo de la detección de ciertos eventos acústicos, pueda formar una realidad aumentada del audio de dicha escena y mostrarla en forma de video, añadiendo los eventos acústicos como iconos o dibujos superpuestos a la señal de video recogida por una o varias cámaras.

30 **Descripción de la invención**

La presente invención resuelve los problemas existentes en el estado de la técnica mediante un método y un sistema capaz de representar de forma visual eventos acústicos en un sistema visual de realidad aumentada.

La invención propuesta consiste en un método y un sistema capaz de captar las señales acústicas recibidas por unos micrófonos situados en la posición de los oídos de una persona (por ejemplo en las patillas de unas gafas), e incorporar, de forma visual, la información acerca de los eventos acústicos detectados (algunos ejemplos de eventos acústicos serían la voz de personas hablando, música, golpes, sirenas, alarmas, pitidos o vehículos en movimiento). El sistema añade esta información de forma complementaria a la visión que se está mostrando a través de un sistema visual de realidad aumentada, por ejemplo, unas gafas de realidad aumentada.

El método y sistema propuesto establecen un marco de integración de tecnologías que aumentan la capacidad de percepción de las personas sordas, mejorando sensiblemente su calidad de vida.

El método para incorporar información acústica binaural en un sistema visual de realidad aumentada, comprende las siguientes etapas:

- a) captar señales acústicas binaurales de una escena;
- 50 b) digitalizar las señales acústicas captadas obteniendo una pluralidad de muestras de señal;
- c) agrupar N muestras consecutivas en tramas, siendo N un número configurable;
- d) extraer, de cada trama, una pluralidad de características que definen dicha trama;
- 55 e) seleccionar un subconjunto de las características extraídas;
- f) detectar mediante un reconocimiento de patrones el tipo de evento acústico ocurrido en la escena analizada a partir del subconjunto de características seleccionadas y de una pluralidad de patrones previamente almacenados en una base de datos del sistema;
- 60 g) estimar la distancia a la que se encuentra la fuente del evento acústico detectado a partir del tipo de evento acústico y de una pluralidad de características de señal previamente almacenadas en una base de datos del sistema de la energía de la señal acústica y de la distribución espectral de dicha señal;
- 65 h) calcular la dirección en la que se sitúa la fuente del evento acústico a partir del retardo existente entre las señales acústicas captadas, de la diferencia de energía entre las señales captadas y de la distribución espectral de la energía de dichas señales;

ES 2 347 517 B2

i) localizar la fuente del evento acústico en la escena a partir de la distancia de la fuente estimada y de la dirección calculada en la que se sitúa dicha fuente;

5 j) codificar visualmente la información del evento acústico, generar un vídeo a partir de una secuencia de imágenes, comprendiendo para cada imagen:

- seleccionar, de una base de datos, un dibujo asociado al tipo de evento acústico;
- determinar la zona de la imagen en la que se representará el dibujo a partir de la localización de la fuente acústica;
- determinar el tamaño del dibujo a partir de la distancia a la que se encuentra la fuente del evento acústico;

10 k) superponer la secuencia de video generada con las imágenes captadas por al menos una cámara de vídeo;

15 l) mostrar la señal visual combinada.

Preferentemente, la extracción de características en cada trama comprende, extraer características en el dominio del tiempo, comprendiendo, la tasa de cruces por cero, la energía de la señal en esa trama y el retardo entre las 20 señales binaurales captadas, y, extraer características en el dominio de la frecuencia, comprendiendo el cálculo de los coeficientes LPC (Linear Prediction Coding), los coeficientes MFC (Mel- Frequency Cepstral), la energía en una selección previa de bandas de frecuencias de la señal y el flujo espectral;

Preferentemente, la selección de un subconjunto de características se realiza mediante LDA (Linear Discriminant 25 Analysis) o mediante árboles de decisión a partir de las características extraídas en la etapa d. El sistema que implementa el método, de manera preferente, comprenderá una base de datos con ejemplos previamente etiquetados y reglas para aprender, de forma automática, las características que mejor ayudan en la detección de los eventos acústicos que serán aquellas que ofrezcan una mayor capacidad de discriminación.

Preferentemente, el reconocimiento de patrones comprende calcular la distancia entre el subconjunto de caracte- 30 rísticas seleccionadas y cada patrón almacenado en la base de datos del sistema, obteniéndose como tipo de evento acústico detectado aquel cuyo patrón ofrezca una menor distancia con el subconjunto de características.

Preferentemente, la estimación de la distancia a la que se encuentra la fuente del evento acústico se obtiene a partir 35 del tipo de evento acústico detectado, de la energía de la señal acústica y de la distribución espectral de dicha señal. Conociendo el tipo de evento acústico y la energía que llega a los micrófonos se puede estimar la distancia a la que está la fuente del evento acústico. Por otro lado, la oreja o pabellón auditivo establece un filtro acústico que conforma y atenúa las señales acústicas percibidas dotándolas de una menor energía y una distribución espectral de su energía 40 (distribución de energía en bandas de frecuencia) característica. Es importante saber si la fuente del evento acústico está situada detrás del oyente para adaptar la estimación de la distancia a esta situación: considerando una atenuación adicional debida al pabellón auditivo y no a una mayor distancia a la fuente acústica.

El sistema para incorporar información acústica binaural en un sistema visual de realidad aumentada, comprende:

- al menos dos micrófonos omnidireccionales configurados para captar señales acústicas binaurales de una esce- 45 na;
- un módulo de extracción de características que comprende:
 - un módulo de extracción de características de las señales captadas por un primer micrófono configurado 50 para digitalizar las señales acústicas captadas obteniendo una pluralidad de muestras de señal, agrupar N muestras consecutivas en tramas, siendo N un número configurable, y extraer, de cada trama, una pluralidad de características que definen dicha trama;
 - un módulo de extracción de características de las señales captadas por un segundo micrófono configurado 55 para digitalizar las señales acústicas captadas obteniendo una pluralidad de muestras de señal, agrupar N muestras consecutivas en tramas, siendo N un número configurable, y extraer, de cada trama, una pluralidad de características que definen dicha trama;
 - un módulo de selección de características configurado para seleccionar un subconjunto de características a partir de las características extraídas;
- un módulo de caracterización de la fuente de sonido que comprende:
 - un módulo de detección del tipo de evento acústico configurado para detectar mediante un reconocimiento 65 de patrones el tipo de evento acústico ocurrido en la escena analizada a partir del subconjunto de caracte- rísticas seleccionadas y de una pluralidad de patrones previamente almacenados en una base de datos del sistema;

ES 2 347 517 B2

○ un módulo de estimación de la distancia de la fuente configurado para estimar la distancia a la que se encuentra la fuente del evento acústico detectado a partir del tipo de evento acústico y de una pluralidad de características de señal previamente almacenadas en una base de datos del sistema;

○ un módulo de localización de la fuente del evento acústico configurado para calcular la dirección en la que se sitúa la fuente del evento acústico a partir del retardo existente entre las señales acústicas captadas, de la diferencia de energía entre las señales captadas y de la distribución espectral de la energía de dichas señales y para localizar la fuente del evento acústico en la escena a partir de la distancia de la fuente estimada y de la dirección calculada en la que se sitúa dicha fuente;

● un módulo de realidad aumentada que comprende:

○ un módulo de codificación configurado para generar un vídeo a partir de una secuencia de imágenes, comprendiendo para cada imagen:

● seleccionar, de una base de datos, un dibujo asociado al tipo de evento acústico;

● determinar la zona de la imagen en la que se representará el dibujo a partir de la localización de la fuente acústica;

● determinar el tamaño del dibujo a partir de la distancia a la que se encuentra la fuente del evento acústico;

○ un módulo de combinación configurado para superponer la secuencia de video generada con las imágenes captadas por al menos una cámara de vídeo de la escena;

○ un módulo de representación visual configurado para mostrar la señal visual combinada.

Preferentemente, el módulo de extracción de características de las señales captadas por un primer micrófono y el módulo de extracción de características de las señales captadas por un segundo micrófono están configurados para extraer características en el dominio del tiempo, comprendiendo, la tasa de cruces por cero, la energía de la señal en esa trama y el retardo entre las señales binaurales captadas, y, para extraer características en el dominio de la frecuencia, comprendiendo el cálculo de los coeficientes LPC (Linear Prediction Coding), los coeficientes MFC (Mel- Frequency Cepstral), la energía en una selección previa de bandas de frecuencias de la señal y el flujo espectral.

Preferentemente, el módulo de extracción de características de las señales captadas por un primer micrófono y el módulo de extracción de características de las señales captadas por un segundo micrófono están configurados para realizar la selección del subconjunto de características mediante LDA (Linear Discriminant Analysis) o mediante árboles de decisión a partir de las características extraídas.

Preferentemente, el sistema comprende una base de datos configurada para aprender automáticamente las características con mayor capacidad de discriminación. La base de datos comprende ejemplos de eventos acústicos previamente etiquetados con sus características y el tipo de evento que es además de una pluralidad de reglas que definen la selección de un dibujo asociado al tipo de evento acústico detectado. El proceso de aprendizaje de la capacidad de discriminación de cada característica se realiza analizando el tipo de evento y las características extraídas para dicho evento.

Preferentemente, el módulo de detección del tipo de evento acústico está configurado para realizar un reconocimiento de patrones, comprendiendo dicho reconocimiento, calcular la distancia entre el subconjunto de características seleccionadas y cada patrón almacenado en la base de datos del sistema, obteniéndose como tipo de evento acústico detectado aquel cuyo patrón ofrezca una menor distancia con el subconjunto de características.

Preferentemente, el módulo de estimación de la distancia de la fuente está configurado para estimar la distancia a la que se encuentra la fuente del evento acústico a partir del tipo de evento acústico detectado, de la energía de la señal acústica y de la distribución espectral de dicha señal.

La presente invención abre un nuevo mundo de posibilidades a las personas con discapacidad auditiva: permitiría desarrollar sistemas de realidad aumentada que suplieran parcialmente esta discapacidad, incluso llegando al desarrollo de videojuegos más accesibles para personas sordas.

Breve descripción de los dibujos

A continuación, para facilitar la comprensión de la invención, a modo ilustrativo pero no limitativo se describirá una realización de la invención que hace referencia a una serie de figuras.

La figura 1 muestra el diagrama de bloques de la realización preferida del sistema propuesto.

La figura 2 muestra una realización del sistema propuesto.

Descripción detallada de un modo de realización

La realización preferente de la invención presenta el diagrama de bloques mostrado en la figura 1. Los módulos que forman esta realización son los siguientes:

- 5
- El primer módulo consiste en un módulo de extracción de características 2 de las señales acústicas binaurales captadas por dos micrófonos omnidireccionales 1 situados en posiciones similares a la localización de los oídos de una persona. Este módulo comprende:
 - 10 ○ un módulo de extracción de características de las señales captadas por un primer micrófono 3. Las señales acústicas se digitalizan a una frecuencia de muestreo de, al menos, 16 Khz. Las muestras se agrupan en tramas (o secuencias de N muestras consecutivas) solapadas en el tiempo con el fin de extraer diferentes características de las señales para cada una de las tramas. Las características a extraer son tanto en el dominio del tiempo (como por ejemplo, tasa de cruces por cero, evolución de la energía o retardo entre señales) como en el dominio de la frecuencia, como por ejemplo, LPC (Linear Prediction Coeficients), MFCC (Mel-frequency cepstral coefficients), energías en subbandas de frecuencias o flujo espectral;
 - 15 ○ un módulo de extracción de características de las señales captadas por un segundo micrófono 4. Este módulo extrae características de forma similar pero independiente al módulo de extracción de características de las señales captadas por el primer micrófono 3;
 - 20 ○ un módulo de selección de características 5. Después de extraer las características de las señales acústicas binaurales captadas por los micrófonos omnidireccionales 1 se selecciona un subconjunto de características a partir de las características extraídas.
 - 25 • El segundo módulo consiste en un módulo de caracterización de la fuente de sonido 6. Este módulo comprende:
 - 30 ○ un módulo de detección del tipo de evento acústico 7. Este módulo consiste en un sistema de reconocimiento de patrones que partiendo de las características extraídas en el módulo de extracción de características 2 decide el tipo de evento acústico ocurrido. Este sistema puede estar basado en mezclas de gaussianas (modelos Bayesianos) o redes neuronales;
 - 35 ○ un módulo de estimación de la distancia de la fuente 8. En función del tipo de evento acústico, de la energía de la señal acústica y de la distribución espectral de la energía de la señal, se puede hacer una estimación de la distancia de los micrófonos a la fuente del evento acústico;
 - 40 ○ un módulo de localización de la fuente del evento acústico 9. Utilizando información sobre la distancia de la fuente y calculando la dirección en la que se sitúa dicha fuente, es posible calcular su localización espacial. La evolución de esta localización a lo largo del tiempo pueda dar información sobre el movimiento de la fuente. Una característica muy útil para el cálculo de la dirección en la que se sitúa la fuente es el retardo con el que las señales acústicas llegan a uno u otro micrófono. Este retardo se puede calcular mediante un análisis de la correlación entre las señales obtenidas de ambos micrófonos omnidireccionales 1.
 - 45 • Una vez estimadas las características del evento acústico (tipo de evento, distancia a la fuente del evento y localización de dicha fuente del evento) se procede a representar dicha información mediante un módulo de realidad aumentada 10. Este módulo comprende:
 - 50 ○ un módulo de codificación 11 que codifica visualmente la información del evento acústico, generando un video (secuencia de imágenes) con la representación visual del evento acústico. Para cada una de las imágenes que forman la secuencia de video, este módulo selecciona, de una base de datos, un dibujo asociado al evento acústico. El dibujo seleccionado y su color, depende del tipo de evento acústico a representar. Por otro lado, dependiendo de la localización de la fuente acústica, el dibujo se mostrará en una zona de la imagen diferente. Y finalmente, el tamaño del dibujo dependerá de la distancia a la que se encuentra la fuente del evento acústico. Las variaciones del tamaño tendrán un límite superior e inferior asociadas a una distancia máxima y mínima, respectivamente. Las imágenes resultantes mostrarán los dibujos de los eventos acústicos, siendo el resto de la imagen transparente;
 - 55 ○ un módulo de combinación 12 que superpone la secuencia de video generada artificialmente por el módulo de codificación 11 (y que incluye información sobre los eventos acústicos detectados) con las imágenes captadas por una o varias cámaras de vídeo 13 de la escena;
 - 60 ○ un módulo de representación visual 14 que muestra la señal visual combinada.

65 La figura 2 muestra un ejemplo de sistema visual de realidad aumentada. El ejemplo consiste en unas gafas 15 con micrófonos omnidireccionales 16 situados en sus patillas. Se puede observar la representación de ciertos eventos acústicos: voz de personas que están cerca y situadas a la derecha 17, voz de personas que están más lejos 18 y ruido de un tren acercándose por la izquierda 19.

ES 2 347 517 B2

En este ejemplo, el módulo de codificación 11 de la información del evento acústico se realiza de la siguiente manera:

- 5 • Según el tipo de evento acústico se representa una imagen o dibujo diferente, con un color diferente. El dibujo será lo menos invasivo posible para evitar que dificulte la visión de objetos reales. En la figura 2 se muestran algunos ejemplos de dibujos posibles: voz de personas que están cerca y situadas a la derecha 17, voz de personas que están más lejos 18 y ruido de un tren acercándose por la izquierda 19.
- 10 • La localización de la imagen o dibujo en pantalla dependerá de la localización de la fuente del evento acústico.
- Y finalmente, el tamaño de la imagen o dibujo ofrece información sobre la distancia a la que se encuentra la fuente del evento acústico.

La invención propuesta se puede utilizar en las siguientes aplicaciones industriales:

- 15 • Diseño y fabricación de un sistema visual de realidad aumentada para la mejora de la interacción de las personas sordas en entornos urbanos. Por ejemplo en una estación de metro, las personas sordas podrían percibir visualmente algunos sonidos (como la llegada de un nuevo tren o alguna alarma acústica). En este tipo de aplicación podrían estar interesadas tanto las administraciones públicas, como las asociaciones de personas sordas.
- 20 • Las empresas de videojuegos podrían incorporar esta información visual para ofrecer juegos más accesibles para este colectivo.
- 25 • Finalmente, una aplicación muy interesante sería la conducción de un coche. Actualmente, las personas sordas pueden conducir un coche aunque no perciban las señales acústicas (como pitidos de otros coches). Con un sistema visual de realidad aumentada que incorpore información sobre los eventos acústicos, se podría hacer que las personas sordas percibieran visualmente eventos acústicos como los pitidos de otros coches.

30 Una vez descrita de forma clara la invención, se hace constar que las realizaciones particulares anteriormente descritas son susceptibles de modificaciones de detalle siempre que no alteren el principio fundamental y la esencia de la invención.

35

40

45

50

55

60

65

REIVINDICACIONES

1. Método para incorporar información acústica binaural en un sistema visual de realidad aumentada, **caracterizado** porque comprende las siguientes etapas:

- a) captar señales acústicas binaurales de una escena;
- b) digitalizar las señales acústicas captadas obteniendo una pluralidad de muestras de señal;
- c) agrupar N muestras consecutivas en tramas, siendo N un número configurable;
- d) extraer, de cada trama, una pluralidad de características que definen dicha trama;
- e) seleccionar un subconjunto de las características extraídas;
- f) detectar mediante un reconocimiento de patrones el tipo de evento acústico ocurrido en la escena analizada a partir del subconjunto de características seleccionadas y de una pluralidad de patrones previamente almacenados en una base de datos del sistema;
- g) estimar la distancia a la que se encuentra la fuente del evento acústico detectado a partir del tipo de evento acústico y de una pluralidad de características de señal previamente almacenadas en una base de datos del sistema de la energía de la señal acústica y de la distribución espectral de dicha señal;
- h) calcular la dirección en la que se sitúa la fuente del evento acústico a partir del retardo existente entre las señales acústicas captadas, de la diferencia de energía entre las señales captadas y de la distribución espectral de la energía de dichas señales;
- i) localizar la fuente del evento acústico en la escena a partir de la distancia de la fuente estimada y de la dirección calculada en la que se sitúa dicha fuente;
- j) codificar visualmente la información del evento acústico, generar un vídeo a partir de una secuencia de imágenes, comprendiendo para cada imagen:
 - seleccionar, de una base de datos, un dibujo asociado al tipo de evento acústico;
 - determinar la zona de la imagen en la que se representará el dibujo a partir de la localización de la fuente acústica;
 - determinar el tamaño del dibujo a partir de la distancia a la que se encuentra la fuente del evento acústico;
- k) superponer la secuencia de video generada con las imágenes captadas por al menos una cámara de vídeo (13);
- l) mostrar la señal visual combinada.

2. Método para incorporar información acústica binaural en un sistema visual de realidad aumentada, según la reivindicación 1, **caracterizado** porque la extracción de características en cada trama comprende, extraer características en el dominio del tiempo, comprendiendo, la tasa de cruces por cero, la energía de la señal en esa trama y el retardo entre las señales binaurales captadas, y, extraer características en el dominio de la frecuencia, comprendiendo el cálculo de los coeficientes LPC (Linear Prediction Coding), los coeficientes MFC (Mel-Frequency Cepstral), la energía en una selección previa de bandas de frecuencias de la señal y el flujo espectral.

3. Método para incorporar información acústica binaural en un sistema visual de realidad aumentada, según cualquiera de las reivindicaciones anteriores, **caracterizado** porque la selección de un subconjunto de características se realiza mediante la selección entre:

- aplicar LDA (Linear Discriminant Analysis) a partir de las características extraídas en la etapa d;
- aplicar árboles de decisión a partir de las características extraídas en la etapa d.

4. Método para incorporar información acústica binaural en un sistema visual de realidad aumentada, según cualquiera de las reivindicaciones anteriores, **caracterizado** porque el reconocimiento de patrones comprende calcular la distancia entre el subconjunto de características seleccionadas y cada patrón almacenado en la base de datos del sistema, obteniéndose como tipo de evento acústico detectado aquel cuyo patrón ofrezca una menor distancia con el subconjunto de características.

ES 2 347 517 B2

5. Método para incorporar información acústica binaural en un sistema visual de realidad aumentada, según cualquiera de las reivindicaciones anteriores, **caracterizado** porque la estimación de la distancia a la que se encuentra la fuente del evento acústico se obtiene a partir del tipo de evento acústico detectado, de la energía de la señal acústica y de la distribución espectral de dicha señal.

6. Sistema para incorporar información acústica binaural en un sistema visual de realidad aumentada, **caracterizado** porque comprende:

- al menos dos micrófonos omnidireccionales (1) configurados para captar señales acústicas binaurales de una escena;
- un módulo de extracción de características (2) que comprende:
 - un módulo de extracción de características de las señales captadas por un primer micrófono (3) configurado para digitalizar las señales acústicas captadas obteniendo una pluralidad de muestras de señal, agrupar N muestras consecutivas en tramas, siendo N un número configurable, y extraer, de cada trama, una pluralidad de características que definen dicha trama;
 - un módulo de extracción de características de las señales captadas por un segundo micrófono (4) configurado para digitalizar las señales acústicas captadas obteniendo una pluralidad de muestras de señal, agrupar N muestras consecutivas en tramas, siendo N un número configurable, y extraer, de cada trama, una pluralidad de características que definen dicha trama;
 - un módulo de selección de características (5) configurado para seleccionar un subconjunto de características a partir de las características extraídas;
- un módulo de caracterización de la fuente de sonido (6) que comprende:
 - un módulo de detección del tipo de evento acústico (7) configurado para detectar mediante un reconocimiento de patrones el tipo de evento acústico ocurrido en la escena analizada a partir del subconjunto de características seleccionadas y de una pluralidad de patrones previamente almacenados en una base de datos del sistema;
 - un módulo de estimación de la distancia de la fuente (8) configurado para estimar la distancia a la que se encuentra la fuente del evento acústico detectado a partir del tipo de evento acústico y de una pluralidad de características de señal previamente almacenadas en una base de datos del sistema;
 - un módulo de localización de la fuente del evento acústico (9) configurado para calcular la dirección en la que se sitúa la fuente del evento acústico a partir del retardo existente entre las señales acústicas captadas, de la diferencia de energía entre las señales captadas y de la distribución espectral de la energía de dichas señales y para localizar la fuente del evento acústico en la escena a partir de la distancia de la fuente estimada y de la dirección calculada en la que se sitúa dicha fuente;
- un módulo de realidad aumentada (10) que comprende:
 - un módulo de codificación (11) configurado para generar un vídeo a partir de una secuencia de imágenes, comprendiendo para cada imagen:
 - seleccionar, de una base de datos, un dibujo asociado al tipo de evento acústico;
 - determinar la zona de la imagen en la que se representará el dibujo a partir de la localización de la fuente acústica;
 - determinar el tamaño del dibujo a partir de la distancia a la que se encuentra la fuente del evento acústico;
 - un módulo de combinación (12) configurado para superponer la secuencia de video generada con las imágenes captadas por al menos una cámara de vídeo (13) de la escena;
 - un módulo de representación visual (14) configurado para mostrar la señal visual combinada.

7. Sistema para incorporar información acústica binaural en un sistema visual de realidad aumentada, según la reivindicación 6, **caracterizado** porque el módulo de extracción de características de las señales captadas por un primer micrófono (3) está configurado para extraer características en el dominio del tiempo, comprendiendo, la tasa de cruces por cero, la energía de la señal en esa trama y el retardo entre las señales binaurales captadas, y, para extraer características en el dominio de la frecuencia, comprendiendo el cálculo de los coeficientes LPC (Linear Prediction Coding), los coeficientes MFC (Mel-Frequency Cepstral), la energía en una selección previa de bandas de frecuencias de la señal y el flujo espectral.

ES 2 347 517 B2

8. Sistema para incorporar información acústica binaural en un sistema visual de realidad aumentada, según cualquiera de las reivindicaciones 6-7, **caracterizado** porque el módulo de extracción de características de las señales captadas por un segundo micrófono (4) está configurado para extraer características en el dominio del tiempo, comprendiendo, la tasa de cruces por cero, la energía de la señal en esa trama y el retardo entre las señales binaurales captadas, y, para extraer características en el dominio de la frecuencia, comprendiendo el cálculo de los coeficientes LPC (Linear Prediction Coding), los coeficientes MFC (Mel-Frequency Cepstral), la energía en una selección previa de bandas de frecuencias de la señal y el flujo espectral.

9. Sistema para incorporar información acústica binaural en un sistema visual de realidad aumentada, según cualquiera de las reivindicaciones 6-8, **caracterizado** porque el módulo de extracción de características de las señales captadas por un primer micrófono (3) está configurado para realizar la selección del subconjunto de características mediante LDA (Linear Discriminant Analysis) a partir de las características extraídas.

10. Sistema para incorporar información acústica binaural en un sistema visual de realidad aumentada, según cualquiera de las reivindicaciones 6-9, **caracterizado** porque el módulo de extracción de características de las señales captadas por un segundo micrófono (4) está configurado para realizar la selección del subconjunto de características mediante LDA (Linear Discriminant Analysis) a partir de las características extraídas.

11. Sistema para incorporar información acústica binaural en un sistema visual de realidad aumentada, según cualquiera de las reivindicaciones 6-10, **caracterizado** porque el módulo de extracción de características de las señales captadas por un primer micrófono (3) está configurado para realizar la selección del subconjunto de características mediante árboles de decisión a partir de las características extraídas.

12. Sistema para incorporar información acústica binaural en un sistema visual de realidad aumentada, según cualquiera de las reivindicaciones 6-11, **caracterizado** porque el módulo de extracción de características de las señales captadas por un segundo micrófono (4) está configurado para realizar la selección del subconjunto de características mediante árboles de decisión a partir de las características extraídas.

13. Sistema para incorporar información acústica binaural en un sistema visual de realidad aumentada, según cualquiera de las reivindicaciones 6-12, **caracterizado** porque comprende una base de datos que comprende eventos acústicos previamente etiquetados y una pluralidad de reglas que definen la selección de un dibujo asociado al tipo de evento acústico detectado.

14. Sistema para incorporar información acústica binaural en un sistema visual de realidad aumentada, según cualquiera de las reivindicaciones 6-13, **caracterizado** porque el módulo de detección del tipo de evento acústico (7) está configurado para realizar un reconocimiento de patrones, comprendiendo dicho reconocimiento, calcular la distancia entre el subconjunto de características seleccionadas y cada patrón almacenado en la base de datos del sistema, obteniéndose como tipo de evento acústico detectado aquel cuyo patrón ofrezca una menor distancia con el subconjunto de características.

15. Sistema para incorporar información acústica binaural en un sistema visual de realidad aumentada, según cualquiera de las reivindicaciones 6-14, **caracterizado** porque el módulo de estimación de la distancia de la fuente (8) está configurado para estimar la distancia a la que se encuentra la fuente del evento acústico a partir del tipo de evento acústico detectado, de la energía de la señal acústica y de la distribución espectral de dicha señal.

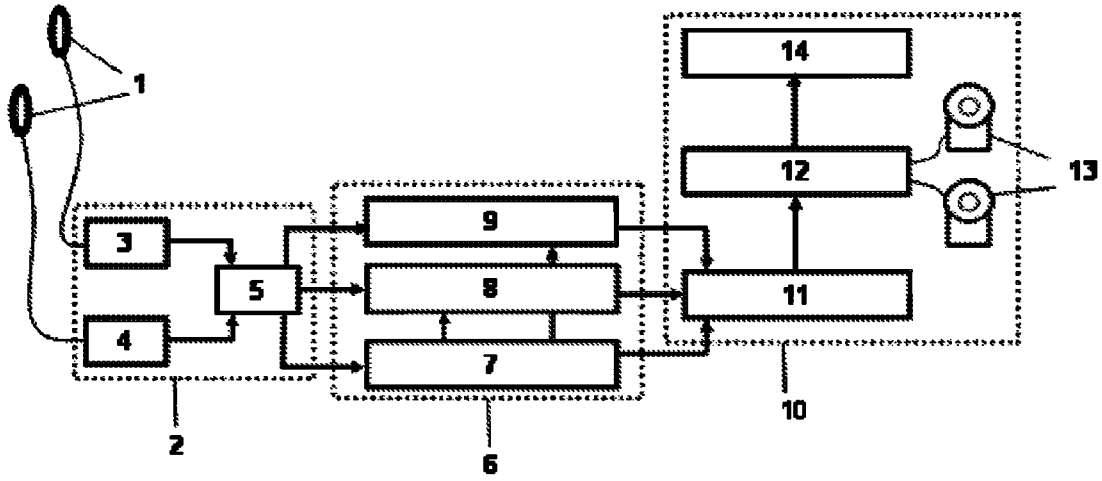


Fig. 1

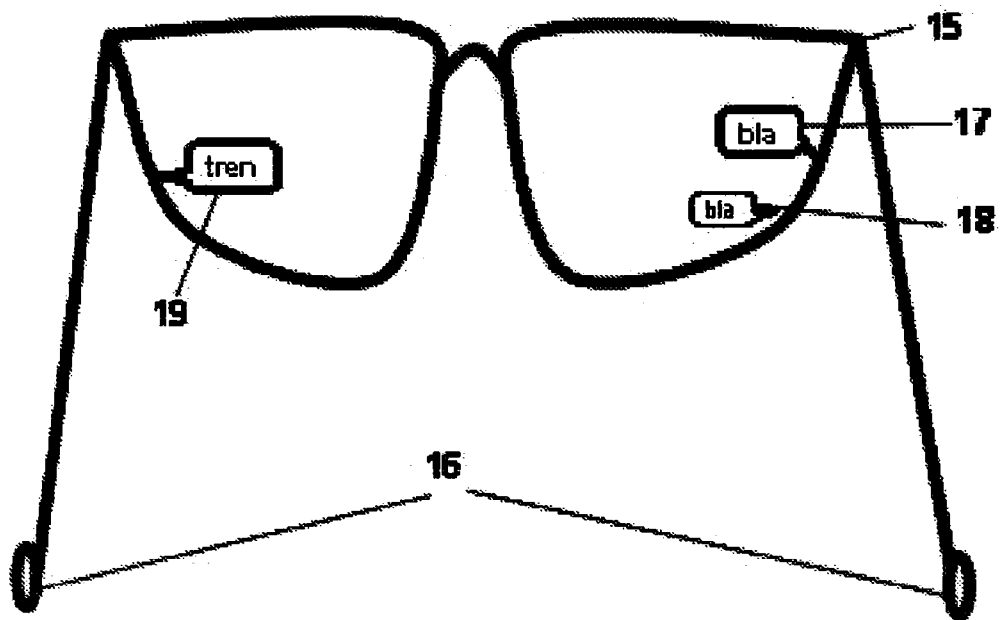


Fig. 2



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

① ES 2 347 517

②1 N° de solicitud: 201031215

②2 Fecha de presentación de la solicitud: **04.08.2010**

③2 Fecha de prioridad:

INFORME SOBRE EL ESTADO DE LA TÉCNICA

⑤1 Int. Cl.: **H04R 25/00** (2006.01)
G06F 3/01 (2006.01)

DOCUMENTOS RELEVANTES

Categoría	⑤6 Documentos citados	Reivindicaciones afectadas
A	US 5029216 A (JHABVALA et al.) 02.07.1991,	1,6
A	US 2004056902 A1 (REKIMOTO et al.) 25.03.2004,	1,6
A	WO 2009072040 A1 (KONINKL PHILIPS ELECTRONICS NV ; ROOVERS DAVID A C M ;) 11.06.2009,	1,6
A	WO 2007063139 A2 (PHONAK AG ; DERLETH RALPH PETER ; LAUNER STEFAN) 07.06.2007,	1,6

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

18.10.2010

Examinador

Mª C. González Vasserot

Página

1/4

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

H04R, G06F

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

Fecha de Realización de la Opinión Escrita: 18.10.2010

Declaración

Novedad (Art. 6.1 LP 11/1986)	Reivindicaciones 1-15	SÍ
	Reivindicaciones	NO
Actividad inventiva (Art. 8.1 LP 11/1986)	Reivindicaciones 1-15	SÍ
	Reivindicaciones	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 ha sido publicada.

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 5029216 A	02-07-1991
D02	US 2004056902 A1	25-03-2004
D03	WO 2009072040 A1	11-06-2009
D04	WO 2007063139 A2	07-06-2007

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

Los documentos citados solo muestran el estado general de la técnica, y no se consideran de particular relevancia. Así, la invención reivindicada se considera que cumple los requisitos de novedad, actividad inventiva y aplicación industrial.

1.- El objeto de la presente solicitud de patente consiste en un método y un sistema que integre la información que reciba mediante señales binaurales, de forma que partiendo de la detección de ciertos eventos acústicos pueda formar una realidad aumentada del audio de dicha escena y mostrarla en forma de video, añadiendo los eventos acústicos como iconos o dibujos superpuestos a la señal de video recogida por una o varias cámaras.

2.- El problema planteado por el solicitante es detectar el tipo de evento acústico localizando la fuente de los eventos acústicos, así como la clasificación del tipo de ruido, para informar al usuario añadiendo esta información de forma complementaria a la visión que se está mostrando a través de un sistema de realidad aumentada.

El documento D01 puede considerarse como el representante del estado de la técnica más cercano ya que en este documento confluyen la mayoría de las características técnicas reivindicadas.

Análisis de las reivindicaciones independientes 1,6

D01 se diferencia del documento de solicitud de patente en que no localiza la fuente del evento acústico de la escena a partir de la distancia de la fuente estimada y de la dirección calculada en la que se sitúa dicha fuente, por tanto no incluye la etapa i) del método propuesto, tampoco superpone la secuencia de video generada con las imágenes captadas, es decir, tampoco incorpora la etapa k) de la reivindicación 1 de la solicitud.

La reivindicación 1 es nueva (Art. 6.1 LP 11/1986) y tiene actividad inventiva (Art. 8.1 LP11/1986).

D01 tampoco tiene el módulo de caracterización de la fuente de sonido de la solicitud conteniendo el módulo de localización de la fuente del evento acústico ni el módulo de realidad aumentada con un módulo de codificación para superponer la secuencia de video generada con las imágenes captadas como tiene la solicitud. Por tanto la reivindicación 6 de la solicitud es nueva (Art. 6.1 LP 11/1986) y tiene actividad inventiva (Art. 8.1 LP11/1986).

Análisis del resto de los documentos

De este modo, ni el documento D01, ni ninguno del resto de los documentos citados en el Informe del Estado de la Técnica, tomados solos o en combinación, revelan la invención en estudio tal y como es definida en las reivindicaciones independientes, de modo que los documentos citados solo muestran el estado general de la técnica, y no se consideran de particular relevancia. Además, en los documentos citados no hay sugerencias que dirijan al experto en la materia a una combinación que pudiera hacer evidente la invención definida por estas reivindicaciones y no se considera obvio para una persona experta en la materia aplicar las características incluidas en los documentos citados y llegar a la invención como se revela en la misma.