(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 923 639**

(51) Int. CI.:

***G11B 27/28*** (2006.01)
***G11B 27/034*** (2006.01)
***G06K 9/62*** (2012.01)
***G06T 7/00*** (2007.01)
***G11B 27/10*** (2006.01)
***G11B 27/32*** (2006.01)
***H04N 5/60*** (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **24.11.2011 PCT/EP2011/070991**

(87) Fecha y número de publicación internacional: **31.05.2012 WO12069614**

(96) Fecha de presentación y número de la solicitud europea: **24.11.2011 E 11788440 (3)**

(97) Fecha y número de publicación de la concesión europea: **22.06.2022 EP 2643791**

(54) Título: **Método y conjunto para mejorar la presentación de sonidos de señal de audio durante una grabación de vídeo**

(30) Prioridad:

**25.11.2010 DE 102010052527**

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
**29.09.2022**

(73) Titular/es:

**INSTITUT FÜR RUNDFUNKTECHNIK GMBH (100.0%)**
**C/O Bayerischer Rundfunk, Rundfunkplatz 1**
**80335 München, DE**

(72) Inventor/es:

**GERSTLBERGER, IRIS;**
**HARTMANN, CHRISTIAN y**
**MEIER, MICHAEL**

(74) Agente/Representante:

**ELZABURU, S.L.P**

## DESCRIPCIÓN

Método y conjunto para mejorar la presentación de sonidos de señal de audio durante una grabación de vídeo

### Campo de la invención

La invención se refiere a un método y un conjunto para mejorar la presentación de audio de sonidos, en particular sonidos específicos de deportes, durante una grabación de vídeo. Tal método y tal conjunto son conocidos a partir del documento DE 10 2008 045 397 A1.

### Descripción de la técnica anterior

Por medio de la introducción de imágenes de televisión de alta definición en formato panorámico acompañadas de sonido de televisión multicanal, en particular durante la transmisión de eventos deportivos, el espectador de televisión está significativamente más implicado en las acciones deportivas en comparación con las tecnologías de televisión convencionales porque son percibibles considerablemente más detalles. Para la grabación de imágenes y audio de eventos deportivos en vivo, con frecuencia no se pueden instalar micrófonos en número suficiente o en la proximidad deseada de fuentes de sonido importantes. Estos son principalmente sonidos que son característicos del deporte específico y enfatizan la franqueza del contenido de las imágenes. Correspondientemente, por ejemplo, en la grabación de televisión de partidos de fútbol, normalmente solo son percibibles unos pocos o ningún sonido específico del partido en el campo debido a que las distancias a los micrófonos direccionales que rodean el campo son demasiado grandes con respecto al ruidoso ambiente del estadio. Para la grabación de televisión de carreras de esquí, una cobertura completa de la pista de esquí de kilómetros de largo con micrófonos sería demasiado costosa. En consecuencia, para grabaciones cercanas de escenas de partidos o carreras, los sonidos característicos típicamente no se capturan por la grabación de audio.

A partir del documento DE 10 2008 045 397 A1, para capturar sonidos específicos de deportes durante la grabación de vídeo de eventos deportivos de pelota, es conocido proporcionar un sistema de micrófono fuertemente direccional con al menos dos micrófonos direccionales, cada uno de ellos alineado hacia la posición actual de la pelota por medio de una entidad de guía móvil en todas las direcciones del espacio. El guiado de los micrófonos durante la producción ocurre automáticamente, sin intervención manual, en dependencia de los datos de posición de la pelota generados por medio de un método de seguimiento de la pelota.

Sin embargo, esta tecnología de grabación de audio conocida no se puede aplicar a todas las grabaciones de vídeo y requiere un esfuerzo técnico comparativamente alto.

El documento US5159140A enseña un aparato de control acústico que se puede aplicar a un instrumento musical electrónico y controla la acústica de un tono musical a ser generado en respuesta a la variación de una imagen. Con el fin de detectar la variación de una imagen, el aparato de control acústico extrae un elemento de imagen predeterminado de la información de imagen a ser dada. Este elemento de imagen se puede identificar como movimiento de imagen, color de imagen o contorno de imagen. El color de imagen se puede detectar detectando el tono y/o el número de colores en la imagen. Además, en respuesta a la periodicidad en la variación de este elemento de imagen, se puede controlar un tempo de ejecución de tono musical.

### Compendio de la invención

El problema a ser resuelto por la invención es crear un método y un conjunto según un diseño descrito al principio que permita(n) una mejor presentación de audio de los sonidos con un esfuerzo técnico reducido durante cualquier grabación de vídeo.

Este problema se resuelve mediante el método expuesto en la reivindicación 1. La reivindicación 2 describe una realización preferida.

### Breve descripción de los dibujos

La invención llegará a estar completamente clara a partir de la siguiente descripción detallada, dada por medio de un mero ejemplo ejemplificativo y no limitativo, para ser leída con referencia a las figuras de los dibujos adjuntos, en donde:

la Fig. 1 muestra un diagrama de bloques esquemático de un conjunto para realizar el método según la invención con los tres componentes centrales: entidad sensora, unidad de control central y base de datos de audio;

la Fig. 2 muestra detalles de la unidad de control central del conjunto según la Fig. 1, y

la Fig. 3 muestra un ejemplo de clasificación de muestras de audio en diferentes categorías (mapeo de muestras) en la base de datos de audio.

**Descripción de la realización preferida**

El conjunto 1 para realizar el método según la invención, que se muestra esquemáticamente en la Fig. 1, comprende una entidad sensora 10 para la detección del contenido de imagen de imágenes de vídeo. El contenido de imagen detectada se suministra por la entidad sensora 10 en forma de datos 11 a una unidad de procesamiento y análisis basada en software 30, que se muestra con más detalle en la Fig. 2 y se va a explicar más adelante.

Por ejemplo, las imágenes en tiempo real de un evento deportivo (denominadas "imagen de transmisión" en la Fig. 1 y la siguiente descripción) suministradas por una cámara de televisión se pueden usar como imágenes de video para la detección del contenido de la imagen. La entidad sensora 10, por ejemplo, realiza un análisis de la imagen de transmisión utilizando algoritmos del campo de la "visión artificial" (visión por ordenador). Estos algoritmos, entre otras cosas, permiten la separación y el seguimiento de objetos en movimiento contra el fondo de una imagen, así como la determinación de sus posiciones en dependencia de la sección de la imagen. Tomando un partido de fútbol como ejemplo, la ubicación de la pelota en el campo así como la posición y el tamaño de todos los jugadores de fútbol mostrados en la sección de la imagen se pueden determinar en consecuencia. Además, es posible asignar a los jugadores a diferentes equipos por medio de sus camisetas, así como calcular la dirección del movimiento y la velocidad de la pelota. La detección (y posterior análisis en la unidad de análisis y procesamiento 30; Fig. 1) de la imagen de transmisión proporciona además la ventaja de ser capaces de deducir la ubicación y la distancia focal de la cámara de televisión actualmente elegida ("cortada") por el director de imagen durante la grabación de un partido usando múltiples cámaras de televisión.

Complementariamente, en la entidad sensorial 10 y en la unidad de análisis y procesamiento 30, también es posible la grabación y análisis automático de señales de audio (denominadas "sonido de transmisión" en la Fig. 1 y la siguiente descripción) que caracterizan acciones específicas dentro de una escena de la imagen de transmisión. Por ejemplo, la información obtenida del sonido de transmisión se utiliza para verificar acústicamente las acciones de imagen detectadas por medio de análisis de video. Además, los sensores que determinan las acciones de los actores que aparecen en la imagen de transmisión de una manera física se pueden proporcionar en la entidad sensora 10, para una definición más precisa y captura de secuencias de movimiento. Esto incluye, por ejemplo, la determinación de la posición actual de los actores por medio de GPS o sistema de marcación por radio. Como los datos 11, dicha información adicional también se suministra a la unidad de análisis y procesamiento basada en software 30.

En la invención, una posibilidad técnicamente menos costosa para la detección del contenido de la imagen es utilizar, para el análisis de vídeo, la señal de una cámara de seguimiento dedicada e instalada estáticamente en lugar de la imagen de transmisión. La cámara de seguimiento se puede calibrar por adelantado para la escena correspondiente y, por lo tanto, simplifica la detección automática de objetos e interacciones en la imagen de video. En este caso, sin embargo, la información sobre la imagen de transmisión real se debe suministrar externamente desde una unidad 20, por ejemplo, sobre (por explicar) los metadatos de la cámara o las señales de GPIO de una consola mezcladora de imágenes no mostrada en los dibujos.

El análisis y procesamiento de los datos 11 suministrados por la entidad sensora 10 se realiza en la unidad 30, que se ilustra con más detalle en la Fig. 2. La unidad 30 dedujo los comandos de control 31 para una base de datos de audio a partir de los datos 11 de la entidad sensora 10, por ejemplo, como parte de un procesamiento basado en PC o DSP. En la unidad 30, en una primera etapa de análisis 32 (que analiza escenas independientemente de la imagen de transmisión), los parámetros determinados en base al sensor para la descripción de la imagen de video se vinculan lógicamente unos con otros según reglas predefinidas y, por medio de la información resultante, los comandos de control 31 se generan para la selección de sonidos individuales archivados, esto es, "muestras de audio", que se almacenan en la base de datos de audio 40. Las reglas predefinidas a su vez son independientes de la aplicación y, para cada propósito, se deben especificar e introducir específicamente en el software de la unidad de análisis y procesamiento 30 por adelantado. La base de datos 40 emite las muestras de audio seleccionadas por el comando de control 31 como la señal de audio 41 que posteriormente se alimenta directamente a la consola mezcladora de producción 50 y, dentro de la misma, se puede mezclar con otros componentes del sonido que acompaña al video, tal como los sonidos del entorno ("sonido original"), así como el "sonido de diálogo" del comentarista del partido cuando sea aplicable. Por este medio, se debe tener cuidado de que no ocurran duplicaciones perturbadoras y desplazadas temporalmente entre las muestras de audio y el sonido que acompaña al vídeo. Durante la selección de las muestras de audio para su adición al sonido que acompaña al vídeo, se hace una distinción entre las siguientes características para obtener una edición de audio de una escena de vídeo que suene realista (en orden de su relevancia):

1. tipo de sonido
2. volumen del sonido (velocidad)
3. adición de reverberación (espacialidad)
4. amplitud panorámica (asignación de dirección al sonido)

Para la aplicación del método según la invención a los partidos de fútbol, tal especificación de reglas en la etapa 32 significa que, por ejemplo, en base a la información extraída del análisis de video con respecto al cambio de vector del

movimiento de la pelota, se puede determinar el origen de un nuevo disparo. La aceleración de la pelota, así como la longitud del vector de movimiento en el campo, proporcionan información sobre: si es un tiro de larga distancia o un pase con características de sonido divergentes; y lo fuerte (valor de "velocidad") que debería ser un sonido correspondiente (señal de audio 41) que se suministra a la consola mezcladora de producción 50 (Fig. 1) desde la base de datos de audio 40 según el comando de control 31 generado por la unidad 30. El volumen del sonido suministrado a la consola mezcladora de producción 50 se puede variar además en dependencia de la posición de la pelota en el campo, por lo que se puede reproducir la distancia del origen del sonido con respecto al espectador.

En la segunda etapa de análisis 33 (Fig. 2), que analiza los parámetros dependiendo de la imagen de transmisión, la información sobre la sección de imagen de la imagen de transmisión se tiene en cuenta para la selección de sonido. Esta información se suministra como los datos 21 por la unidad 20. En caso de que el análisis de video se realice directamente en la imagen de transmisión, la posición de la cámara y la distancia focal se pueden determinar por medio del tamaño de los objetos investigados. En caso de que se emplee una cámara de seguimiento independiente (Fig. 1) u otros sistemas sensores para la detección del contenido de la imagen en la entidad sensora 10, se tiene en cuenta la información externa sobre la naturaleza de la imagen de transmisión. Con este propósito, los metadatos de la cámara extraídos de la unidad de control de una cámara de televisión son aplicables, entre otras cosas. Además, las señales de GPIO de una consola mezcladora son aplicables para señalizar cuál de las múltiples cámaras de televisión se elige actualmente ("cortar") en la imagen de transmisión. En base a estos datos, la segunda etapa de análisis 33 genera un comando de control 34 para la variación del volumen de las señales de audio 410 que se suministran a la consola mezcladora de producción 50. Esta variación ocurre por medio de una etapa 70, que se controla por el comando de control 34, para la edición de sonido en tiempo real de la señal de audio 41 de la base de datos de audio 40. Por medio de la variación adicional del volumen de la señal de audio 410 suministrada a la consola mezcladora 50, en cierta medida, se puede simular la distancia óptica, en la que se sitúa el espectador con respecto al centro de acción de la imagen. Tomando como ejemplo un partido de fútbol, por medio de una diferente nivelación de los sonidos de la pelota, se puede recrear la grabación en primer plano de un placaje o la grabación en gran angular de todo el campo, en donde, en cada caso, el espectador asume una diferente distancia óptica a la acción.

Complementariamente, por medio de la segunda etapa de análisis 33, la etapa 70, que está subordinada a la base de datos de audio 40, para la edición dinámica de sonido en tiempo real se puede controlar de modo que, por medio de ecualización y adición de componentes de reverberación en dependencia de la posición del objeto en la imagen de video, se recrea la influencia de la disipación del aire y la espacialidad.

Durante el suministro en tiempo real descrito de la señal de audio 41 o 410 a la consola mezcladora 50, ocurre un cambio temporal específico entre la señal de audio 41 y la imagen de video como resultado de la detección y el análisis del contenido de la imagen. Sin embargo, este cambio temporal puede estar limitado a un rango de menos de cuatro imágenes completas, por lo que es posible una asociación inequívoca de eventos de audio/video correspondientes.

Un ejemplo para la organización de la base de datos de audio 40 se ilustra para fútbol por medio de un "mapeo de muestras" en la Fig. 3. "Mapeo de muestras" se entiende como la clasificación de las muestras de audio almacenadas en la base de datos 40 en diferentes categorías. La base de datos de audio 40 se puede poner en práctica tanto basada en hardware como basada en software y se basa, por ejemplo, en un muestreador de hardware/software estándar o en un formato de base de datos universal. La transmisión de los comandos de control 31 y 34 a la base de datos 40 puede ocurrir, por ejemplo, a través del protocolo midi. Para clasificar las muestras de audio en la base de datos de audio en dependencia de las características específicas de las muestras de audio, se proporciona un "mapeo de muestras", que varía según surja el propósito de la aplicación.

Tomando como ejemplo un partido de fútbol a ser unido con sonidos realistas, se distinguen muestras de audio para diferentes técnicas de juego en forma de recepciones y entregas de la pelota por el cuerpo, el pie y la cabeza de un jugador. Para las técnicas de juego que involucran al cuerpo, se diversifican aún más las recepciones y entregas de la pelota con el pecho, la rodilla y la cabeza. Las recepciones y entregas con los pies se dividen nuevamente en los grupos "disparos" y "pases".

Para la realización según la Fig. 3, las muestras de audio se eligen con un volumen variable ("velocidad") en dependencia de la potencia de disparo determinada a partir de la imagen de vídeo en la etapa de análisis 33 (Fig. 2) y, mediante el uso de diferentes muestras de audio, se tienen en cuenta las diferencias tonales de diferentes intensidades de juego. Por ejemplo, el sonido tipo pop de un disparo fuerte tiene un volumen más alto y otra composición de frecuencia que el sonido de un disparo menos potente. Por este motivo, no solo se mezcla el sonido de un disparo menos potente a un volumen más bajo, sino que, además, se emplea otra muestra de audio. Con este propósito, las muestras de audio se graban con distancias de grabación variables con respecto a la fuente de sonido (2 metros, 6 metros o 12 metros) cuando se producen con el fin de reproducir un carácter sonoro directo para disparos fuertes así como Un carácter sonoro indirecto para disparos menos potentes.

Durante la reproducción, parámetros, tales como el volumen, el componente de reverberación, la amplitud panorámica y la ecualización, se cambian por la etapa 70 en dependencia del comando de control 34 predominantemente en tiempo real. Esto ofrece la ventaja de que no tiene que ser almacenada una muestra de audio propia para todas y cada una de las parametrizaciones posibles, lo que reduce drásticamente tanto el requisito

de almacenamiento como los gastos durante la producción de tales bases de datos de audio. Con el fin de promover una impresión general auténtica de la escena editada con audio, es necesario emplear diferentes muestras de audio incluso para contenidos de imágenes similares consecutivos. Con este propósito, se puede proporcionar una rotación aleatoria de muestras de audio.

**REIVINDICACIONES**

1. Un método para la edición de audio de una escena, el método que comprende:

grabar (11) un evento por múltiples primeras cámaras (10);

señalizar (21), cuál de las múltiples primeras cámaras (10) suministra actualmente imágenes de vídeo de la escena;

detectar (32) una ubicación y un movimiento de un objeto en la escena por una segunda cámara de seguimiento (10);

seleccionar (32), en base a la información obtenida por el seguimiento, muestras de audio de una base de datos de audio (40) según criterios predefinidos; y

agregar (70) las muestras de audio seleccionadas a un sonido (41) que acompaña a las imágenes de video;

caracterizado por que el método comprende además

adaptar (33) un volumen de las muestras de audio seleccionadas en base a cuál una de las múltiples cámaras proporciona actualmente las imágenes de vídeo de la escena y en base a la información obtenida por el seguimiento.

2. El método de la reivindicación 1, en donde las múltiples primeras cámaras (10) son cámaras de televisión.

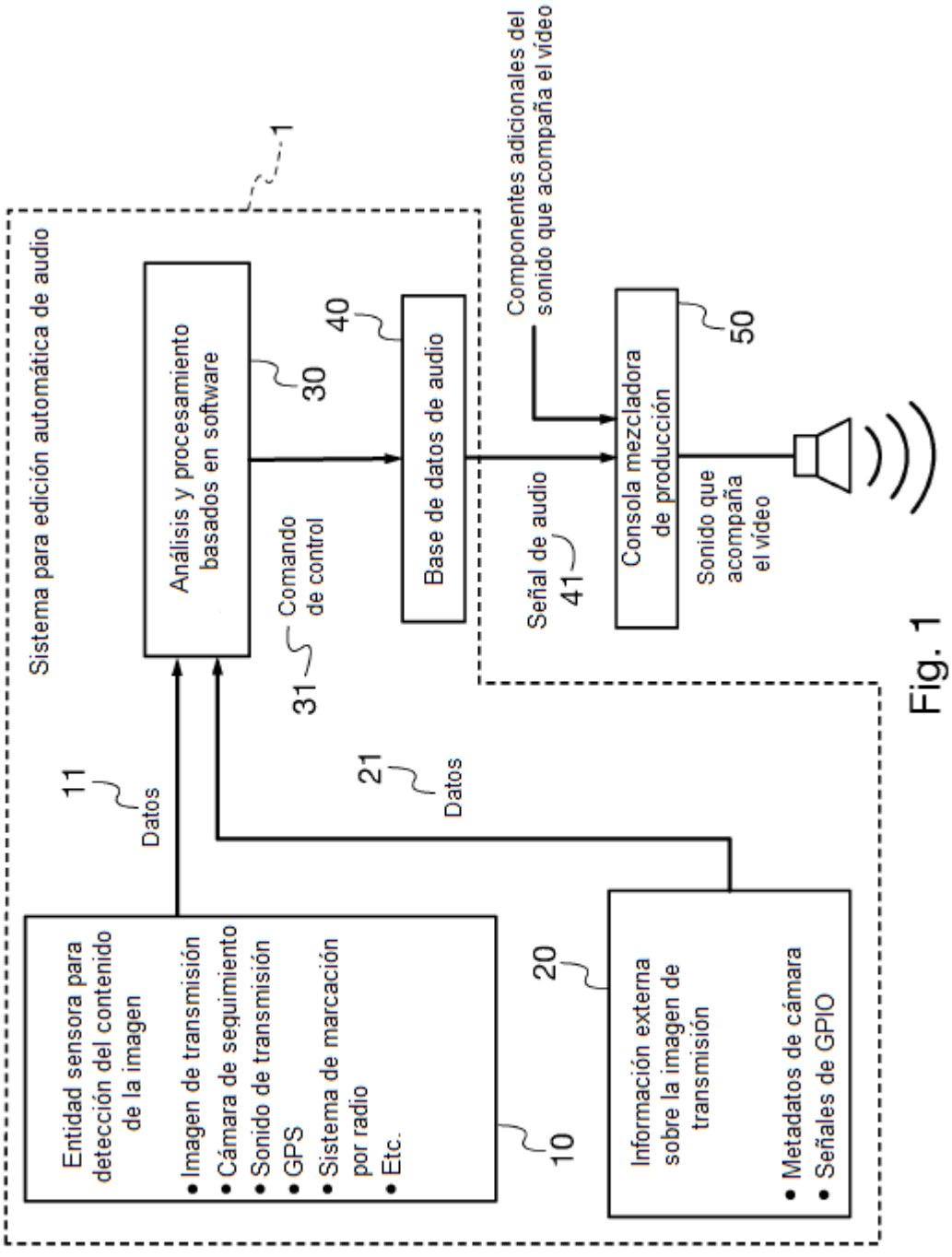
Sistema para edición automática de audio
1
Entidad sensora para detección del contenido de la imagen
• Imagen de transmisión
• Cámara de seguimiento
• Sonido de transmisión
• GPS
• Sistema de marcación por radio
• Etc.
10
11
Datos
Análisis y procesamiento basados en software
30
Comando de control
31
Base de datos de audio
40
21
Datos
Información externa sobre la imagen de transmisión
• Metadatos de cámara
• Señales de GPIO
20
Componentes adicionales del sonido que acompaña el vídeo
Señal de audio
41
Consola mezcladora de producción
50
Sonido que acompaña el vídeo

Fig. 1

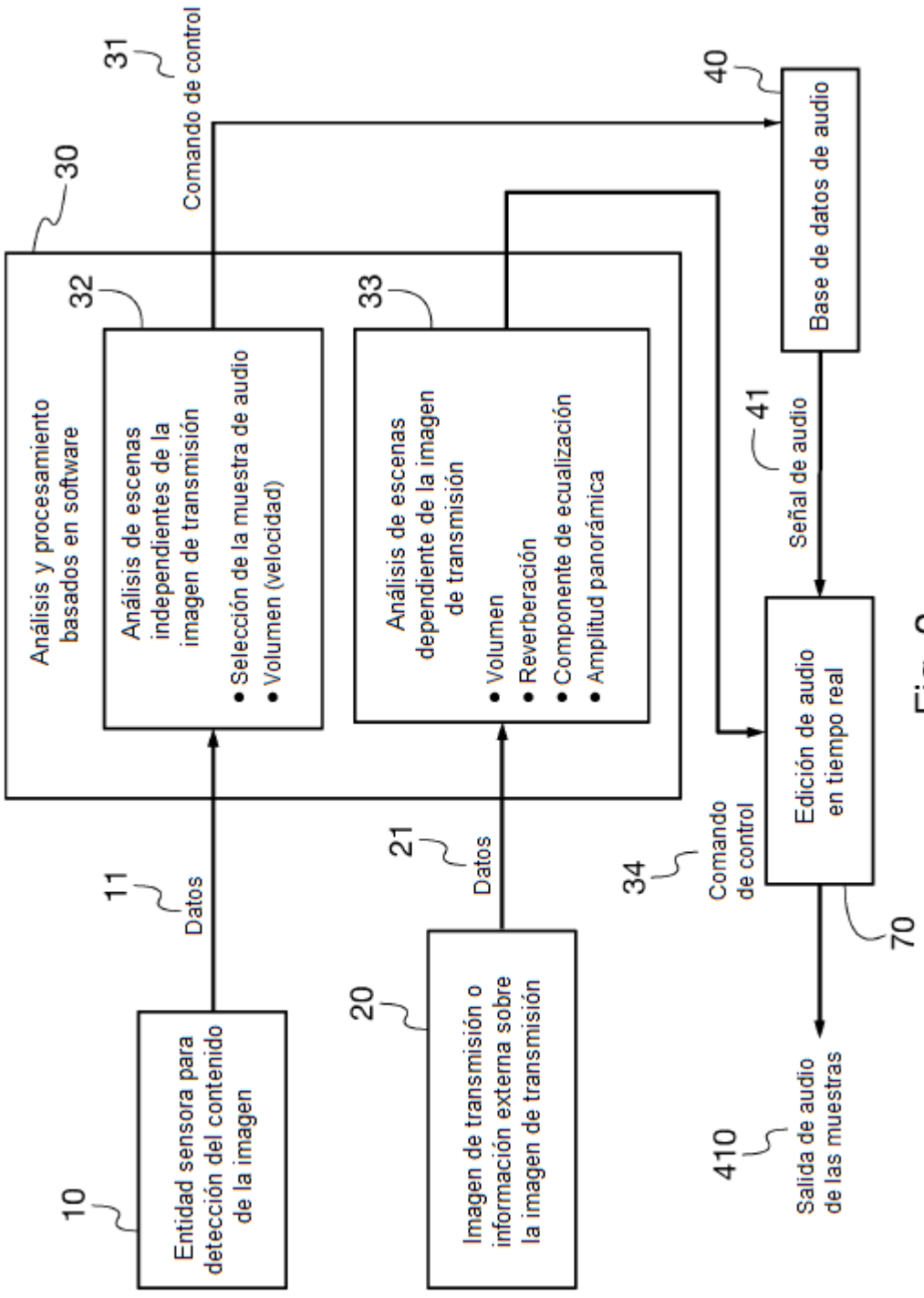
10
Entidad sensora para detección del contenido de la imagen
11
Datos
20
Imagen de transmisión o información externa sobre la imagen de transmisión
21
Datos
30
Análisis y procesamiento basados en software
31
Comando de control
32
Análisis de escenas independientes de la imagen de transmisión
• Selección de la muestra de audio
• Volumen (velocidad)
33
Análisis de escenas dependiente de la imagen de transmisión
• Volumen
• Reverberación
• Componente de ecualización
• Amplitud panorámica
34
Comando de control
40
Base de datos de audio
41
Señal de audio
Edición de audio en tiempo real
70
410
Salida de audio de las muestras

Fig. 2

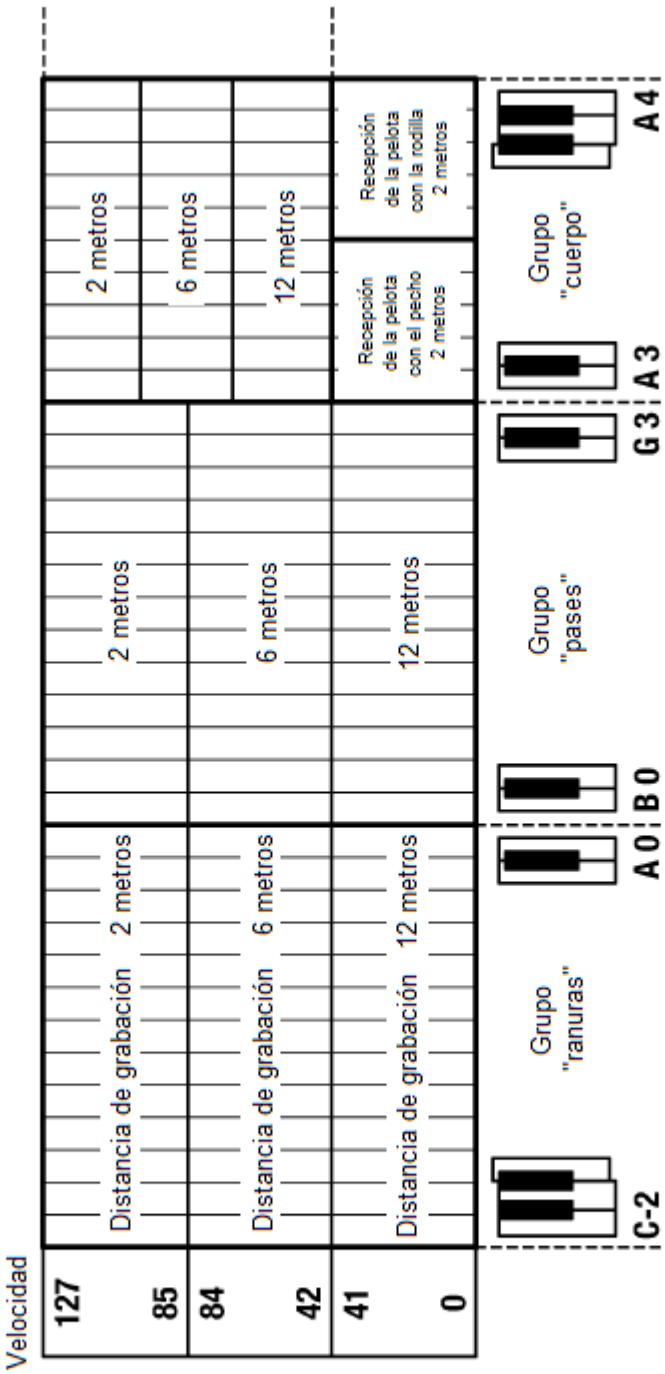
Velocidad
127
85
84
42
41
0
Distancia de grabación
2 metros
6 metros
12 metros
Grupo "ranuras"
Grupo "pases"
Grupo "cuerpo"
Recepción de la pelota con el pecho 2 metros
Recepción de la pelota con la rodilla 2 metros
C-2
A 0
B 0
G 3
A 3
A 4

Fig. 3