



19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 358 099**

51 Int. Cl.:
H04N 7/15 (2006.01)
H04N 7/14 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **02777635 .0**
96 Fecha de presentación : **22.10.2002**
97 Número de publicación de la solicitud: **1449369**
97 Fecha de publicación de la solicitud: **25.08.2004**

54 Título: **Sistema y procedimiento para que personas a distancia conozcan la situación en la sala durante una videoconferencia.**

30 Prioridad: **13.11.2001 US 14200**

45 Fecha de publicación de la mención BOPI:
05.05.2011

45 Fecha de la publicación del folleto de la patente:
05.05.2011

73 Titular/es:
KONINKLIJKE PHILIPS ELECTRONICS N.V.
Groenewoudseweg 1
5621 BA Eindhoven, NL

72 Inventor/es: **Strubbe, Hugo, J.**

74 Agente: **Zuazo Araluze, Alexander**

ES 2 358 099 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

DESCRIPCION

Sistema y procedimiento para que personas a distancia conozcan la situación en la sala durante una videoconferencia.

ANTECEDENTES DE LA INFORMACIÓN

5 1. Campo de la invención

La presente invención se refiere en general a un sistema de comunicación multipunto para establecer una comunicación conectando terminales de comunicación en múltiples ubicaciones, y más particularmente a un procedimiento y sistema para retransmitir la presencia de todos los participantes durante una videoconferencia a un extremo receptor remoto.

10 2. Descripción de la técnica relacionada

Debido al desarrollo avanzado de las técnicas de codificación de compresión de imágenes y las tecnologías de comunicación digital, la demanda de videoconferencias ha crecido exponencialmente. Se han propuesto varios terminales para su uso en videoconferencias. Por ejemplo, de la manera más común se usa una cámara de movimiento horizontal, vertical y zoom (PTZ) montada en la parte superior de un monitor de TV en una configuración de videoconferencia convencional. El operador de la PTZ controla las configuraciones de movimiento horizontal, vertical y zoom de la cámara para capturar una vista de primer plano del hablante, de modo que los participantes a distancia en el extremo receptor puedan ver la cara del hablante. Sin embargo, otros participantes que están presentes en silencio durante la videoconferencia nunca podrían capturarse en los sistemas de videoconferencia convencionales. Por tanto, una persona a distancia que ve la videoconferencia puede no llegar a saber nunca de su presencia. Sin embargo, a menudo se desea percibir la presencia de otros participantes cuando escuchan e interactúan con el hablante.

Por consiguiente, existe la necesidad de proporcionar una vista de otros participantes en videoconferencias, de manera que la presencia de todas las personas en la videoconferencia pueda retransmitirse a los participantes en el extremo receptor remoto.

25 SUMARIO DE LA INVENCION

La presente invención se refiere a un sistema y procedimiento para proporcionar la presencia de todos los participantes en videoconferencias a otros terminales de videoconferencia en extremos receptores remotos.

Según un aspecto de la presente invención, se proporciona un procedimiento según la reivindicación 1. El procedimiento incluye además las etapas de detectar y visualizar una persona nueva que entra a o uno de los participantes que sale de la sala de videoconferencia.

Según otro aspecto de la presente invención, se prevé un aparato según las reivindicaciones 2-11. El aparato es operativo además para reducir el zoom de la cámara en una cantidad predeterminada después de que un periodo detectado de silencio supera el periodo de tiempo umbral.

Las características y ventajas anteriores y otras de la invención resultarán evidentes a partir de la descripción siguiente y más detallada de realizaciones preferidas como se ilustran en los dibujos adjuntos en los que los caracteres de referencia se refieren a las mismas partes a lo largo de las diversas vistas. Los dibujos no son necesariamente a escala, al contrario se hace énfasis en ilustrar los principios de la invención.

BREVE DESCRIPCIÓN DE LA INVENCION

40 Está disponible un entendimiento más completo del procedimiento y del aparato de la presente invención en referencia a la siguiente descripción detallada cuando se toma junto con los dibujos adjuntos en los que:

la figura 1 es un diagrama esquemático que ilustra un sistema de conferencia multipunto según una realización de la presente invención;

la figura 2 es una diagrama simplificado de una sala de videoconferencia a la que puede aplicarse la realización de la presente invención;

45 la figura 3 es un diagrama esquemático que muestra un sistema de videoconferencia según una realización de la presente invención;

la figura 4 es un diagrama de bloques simplificado que muestra los componentes principales del sistema de

videoconferencia según una realización de la presente invención; y,

la figura 5 es un diagrama de flujo que ilustra las etapas de funcionamiento según una realización de la presente invención.

DESCRIPCIÓN DETALLADA DE LAS REALIZACIONES PREFERIDAS

5 En la siguiente descripción, con fines de explicación en vez de limitación, se exponen detalles específicos tales como la arquitectura particular, interfaces, técnicas, etc., con el fin de proporcionar un entendimiento completo de la presente invención. Con fines de simplicidad y claridad, se omiten las descripciones detalladas de dispositivos, circuitos y procedimientos muy conocidos para no confundir la descripción de la presente invención con detalles innecesarios.

10 La figura 1 es un diagrama esquemático que muestra una pluralidad de estaciones terminales conectadas entre sí mediante una línea de comunicación digital configurada para realizar una transmisión bidireccional. Concretamente, el terminal A de videoconferencia está conectado al terminal B de videoconferencia y al terminal D de videoconferencia. El terminal C de videoconferencia está conectado al terminal B de videoconferencia y al terminal D de videoconferencia. Aunque está conectado un número limitado de terminales en forma de lazo con fines
15 ilustrativos, debe entenderse que la presente invención puede soportar comunicaciones concurrentes de videoconferencia entre un número mucho mayor de terminales en diferentes formas. Por tanto, el número de terminales y la disposición de conexión en el dibujo no deben imponer limitaciones al alcance de la invención.

La figura 2 representa una videoconferencia que tiene lugar cuando múltiples participantes interactúan con personas a distancia a través del monitor 2 acoplado a un sistema 10 de videoconferencia que incluye una
20 cámara. La cámara que funciona según la presente invención puede incluir una cámara de movimiento horizontal, vertical y *zoom* (PTZ) convencional. La figura 3 muestra la cámara 8 PTZ que está montada sobre el monitor 2 según la realización de la presente invención. La cámara 8 PTZ es muy conocida por los expertos en la técnica de videoconferencia. Debe observarse que un sistema 10 de videoconferencia dentro del contexto de esta descripción incluye un ordenador portátil, un ordenador de sobremesa convencional, un control remoto de decodificador/TV interactivo, o cualquier dispositivo interactivo dúplex equipado con una cámara para capturar la presencia de
25 participantes en videoconferencias.

En funcionamiento, el sistema 10 con cámara 8 automática incorporada y el software asociado funcionan para retransmitir la presencia de todos los participantes 4-1, 4-2, 4-3,... 4-n en una sala de videoconferencia a otros
30 participantes en el extremo receptor remoto. En particular, el sistema 10 realiza un seguimiento de todos los participantes en el campo de visión de la cámara 8 incluso cuando no están hablando. En un modo normal, la cámara 8 realiza un seguimiento de los participantes basándose en el flujo de las conversaciones y el movimiento de los participantes. En momentos apropiados, es decir cuando haya un silencio largo o cuando una persona habla durante mucho tiempo, la cámara 8 hace *zoom* de manera secuencial en cada participante cada vez. Esto puede alternarse con tomas de la sala completa o tomas que mínimamente enmarquen a todos los participantes. Después,
35 la cámara 9 vuelve a enfocar al hablante o a cualquier otra persona nueva que esté hablando. Mientras tanto, cuando una persona 6 nueva entra o sale de la sala, la cámara 8 automáticamente hace *zoom* de esa persona durante poco tiempo. Como resultado, los participantes a distancia en el extremo receptor sabrán dónde se encuentran todos los participantes en el extremo emisor con respecto al hablante.

La figura 4 muestra los componentes principales del sistema 10 según la realización de la presente
40 invención. El sistema 10 incluye una cámara 8 PTZ, un localizador 12 de audio, un localizador 14 de vídeo, un controlador 16, una interfaz 18 de entrada/salida (E/S), una interfaz 20 de usuario, un procesador 22 de vídeo, un elemento 24 de visualización y una memoria 26.

La cámara 8 PTZ está acoplada al controlador 16 y funciona para realizar un seguimiento de uno de los
45 participantes en la videoconferencia en tiempo real. La cámara 8 PTZ está acoplada al localizador 12 de audio para recibir señales de audio y el localizador 14 de vídeo para señales de vídeo. Debe observarse que pueden usarse otros tipos y disposiciones de conexiones para suministrar señales de vídeo y de audio desde la cámara 8 al controlador 16 u otros sistemas que puedan realizar el seguimiento usando un sistema de seguimiento combinado de audio-vídeo. Además, pueden usarse otros dispositivos de captura de imágenes distintos de una cámara PTZ según las técnicas de la presente invención.

50 El localizador 12 de audio y el localizador 14 de vídeo proporcionan operaciones de seguimiento de audio y vídeo, respectivamente. El localizador 12 de audio según la presente invención puede ser de un tipo descrito en la solicitud de patente estadounidense n.º de serie 09/436.193, presentada el 8 de noviembre de 1999, cuyas enseñanzas están incorporadas en el presente documento como referencia. Brevemente, puede usarse el localizador 12 de audio para discriminar entre los hablantes, es decir, como un subproducto de cancelación de eco.

5 Debe observarse que también pueden usarse otros tipos de localizadores de audio para implementar la presente invención. El localizador de vídeo puede ser cualquier sistema comercialmente disponible muy conocido en la técnica que pueda realizar un seguimiento de personas u otros objetos de interés en una señal de vídeo u otro tipo de señal de imagen. Las señales de salida desde el localizador 12 de audio y el localizador 14 de vídeo se suministran al controlador 16, luego se procesan por el procesador 22 para generar una o más señales de control para controlar las configuraciones de movimiento horizontal, vertical y/o *zoom* de la cámara 8.

10 El dispositivo 18 E/S está acoplado al controlador 16 y adaptado para recibir un flujo de señales de vídeo desde terminales remotos de videoconferencia así como otra variedad de fuentes, incluyendo un proveedor de servicios por cable, una televisión digital de alta definición (HDTV) y/o señales de televisión digital de definición estándar (SDTV), una antena parabólica, una difusión RF convencional y una conexión de Internet. La interfaz 20 de usuario está acoplada al controlador 16 para recibir señales de comando, es decir, señales infrarrojas, del usuario para controlar el sistema 10. El elemento 24 de visualización está acoplado al procesador 22 para recibir señales de audio y de vídeo desde otros terminales de videoconferencia para su visualización.

15 En otras realizaciones de la invención, algunas o todas las funciones realizadas por los elementos del sistema 10 mostrados en la figura 4 pueden combinarse en un solo dispositivo. Por ejemplo, puede implementarse uno o más de los elementos del sistema 10 como circuitos integrados para aplicaciones específicas (ASIC) o una tarjeta de circuito para su incorporación en un ordenador, televisión, decodificador u otro dispositivo de procesamiento.

20 La realización elegida de la presente invención es un software informático que se ejecuta dentro del sistema 10. Los programas informáticos (o lógica de control informática) están almacenados en la memoria 26. Tales programas informáticos, cuando se ejecutan, permiten que el sistema 10 realice la función de la presente invención tal como se comenta en el presente documento.

25 La figura 5 muestra el funcionamiento de la realización de software del sistema 10 de vídeo. Se supone que el sistema 10 de vídeo con el localizador 12 de audio y el localizador 14 de vídeo está activo durante una videoconferencia dada. En la etapa 100, el sistema 10 realiza un seguimiento de todos los participantes usando las configuraciones de movimiento horizontal, vertical y/o *zoom* de la cámara 8. Tanto el localizador 12 de audio y el localizador 14 de vídeo funcionan para generar salidas de seguimiento correspondientes a la presencia de los participantes. Las salidas de seguimiento pueden ser en forma de, por ejemplo, las medidas de direccionalidad en grados que indican una dirección desde un eje central de la cámara 8 para participantes detectados.

30 En la etapa 120, se analizan las salidas de seguimiento procedentes del localizador 12 de audio y del localizador 14 de vídeo para incluir una indicación de la ubicación asociada con la persona que habla en la sala. Con este fin, el sistema 10 se refiere a la salida de seguimiento del localizador de audio que coincide suficientemente con la salida de seguimiento del localizador de vídeo. Por ejemplo, si las medidas de direccionalidad procedentes del localizador 12 de audio y del localizador 14 de vídeo están dentro de un intervalo especificado una respecto a otra, es decir, dentro de 4-5 grados, esto indica que el localizador 12 de audio y el localizador 14 de vídeo coinciden suficientemente en cuanto a la ubicación del hablante actual. Cuando se detecta al hablante, el controlador 16 genera una señal de control que dirige la cámara 8 para hacer *zoom* en la dirección de la persona que habla. En el caso de que hablen más de una persona, el localizador 12 de audio puede ubicar a la persona que hable más fuerte. El localizador 12 de audio descrito en la solicitud de patente estadounidense n.º de serie 09/436.193 anteriormente citada describe el mecanismo de ubicar a la persona que habla más fuerte y no necesita describirse adicionalmente.

35 En la etapa 140, se determina si el hablante actual habla durante más del periodo de tiempo umbral. De no ser así, se determina, en la etapa 160, si la persona actual que habla está en silencio durante más de un periodo de tiempo predeterminado. Si el hablante actual habla o permanece en silencio durante un umbral especificado, el sistema 10 visualiza de manera secuencial una vista de primer plano de los participantes restantes en la etapa 180. Para lograr esto, las configuraciones de la cámara pueden ajustarse basándose en la salida de localizador de vídeo para controlar las configuraciones de movimiento horizontal, vertical y *zoom* de la cámara 8 hacia la dirección de otros participantes. Después, en la etapa 200, el sistema 10 genera una señal de control que dirige la cámara 8 para reducir el *zoom* hasta un grado determinado o para proporcionar una vista del grupo de manera completa. Después, el sistema 10 puede generar una señal de control que dirige la cámara 8 para volver a enfocar al hablante u otra persona que ha comenzado a hablar.

40 Mientras tanto, el localizador 14 de vídeo trata continuamente de hallar a una persona nueva que entra en la sala o uno de los participantes que abandona la sala usando técnicas convencionales muy conocidas basadas en características tales como movimiento y color de cara, de modo que se transmite la presencia o ausencia de una persona que entra o que sale a los participantes a distancia en el extremo receptor.

55 Aunque que se han ilustrado y descrito las realizaciones preferidas de la presente invención, los expertos

5 en la técnica entenderán que pueden realizarse diversos cambios y modificaciones, y que pueden sustituirse equivalentes por sus elementos sin apartarse del verdadero alcance de la presente invención. Además, pueden realizarse muchas modificaciones para adaptarse a una situación particular y a la enseñanza de la presente invención sin apartarse del alcance central. Por tanto, se pretende que la presente invención no esté limitada a la realización particular dada a conocer como el mejor modo contemplado para llevar a cabo la presente invención, sino que la presente invención incluya todas las realizaciones que se encuentran dentro del alcance de las reivindicaciones adjuntas.

REIVINDICACIONES

1. Procedimiento para realizar un seguimiento de una pluralidad de participantes en una videoconferencia, comprendiendo el procedimiento las etapas de:
 - monitorizar dicha videoconferencia para realizar un seguimiento de dichos participantes;
 - 5 generar una salida de localizador de vídeo que representa la ubicación de cada uno de dichos participantes;
 - generar una salida de localizador de audio para determinar una persona que está hablando;
 - ajustar una configuración de una cámara basándose en si dicha salida de localizador de vídeo y dicha salida de localizador de audio se encuentran dentro de un intervalo especificado una respecto a otra; y,
 - 10 si dicho hablante está en silencio durante un periodo de tiempo predeterminado, ajustar dicha configuración de la cámara para hacer *zoom* de manera secuencial en cada participante cada vez.
2. Aparato para realizar un seguimiento de una pluralidad de participantes en una videoconferencia, que comprende:
 - una cámara;
 - un procesador acoplado con dicha cámara y operativo (i) para procesar una salida de localizador de audio y una salida de localizador de vídeo indicativas de la ubicación de cada participante, (ii) ajustar la configuración de una cámara para enfocar a uno de dichos participantes que está hablando; (iii) detectar si dicho participante que habla está en silencio durante un periodo de tiempo umbral; y, (iv) ajustar dicha configuración de la cámara para hacer *zoom* de manera secuencial en cada participante cada vez después de que un periodo detectado de silencio supera dicho periodo de tiempo umbral.
 - 15
- 20 3. Aparato según la reivindicación 2, en el que dicho procesador es operativo además para reducir el *zoom* de dicha cámara en una cantidad predeterminada después de que el periodo detectado de silencio supera dicho periodo de tiempo umbral.
4. Aparato según la reivindicación 2, en el que dicha cámara enfoca a dicho hablante si dichas salidas de localizador de audio y de localizador de vídeo se encuentran dentro de un intervalo especificado una respecto a otra.
- 25 5. Aparato según la reivindicación 2, en el que dicho procesador es operativo además para ajustar dicha configuración de la cámara para visualizar todos los participantes después de que el periodo detectado de silencio supera dicho periodo de tiempo umbral.
6. Aparato según la reivindicación 2, en el que dicho procesador es operativo además para detectar una persona nueva que entra a dicha videoconferencia y en el que dicho procesador es operativo además para utilizar dicha salida de localizador de vídeo para ajustar dicha configuración de la cámara para capturar dicha persona nueva.
- 30 7. Aparato según la reivindicación 2, en el que dicho procesador es operativo además para detectar uno de dichos participantes que abandona dicha videoconferencia y en el que dicho procesador es operativo además para utilizar la salida de localizador de vídeo para ajustar dicha configuración de la cámara para visualizar dicho participante que sale.
- 35 8. Aparato según la reivindicación 2, en el que dicha cámara es una cámara de movimiento horizontal, vertical y *zoom* (PTZ).
9. Aparato según la reivindicación 2, en el que dicho procesador es operativo además para ajustar dicha configuración de la cámara para visualizar de manera secuencial otros participantes que no están hablando si dicho participante que habla continúa hablando durante más de dicho periodo de tiempo umbral.
- 40 10. Aparato según la reivindicación 9, en el que dicho procesador es operativo además para reducir el *zoom* de dicha cámara en una cantidad predeterminada si dicho participante que habla continúa hablando durante más de dicho periodo de tiempo umbral.
- 45 11. Aparato según la reivindicación 9, en el que dicho procesador es operativo además para ajustar dicha configuración de la cámara para visualizar todos los participantes si dicho participante que habla continúa hablando durante más de dicho periodo de tiempo umbral.

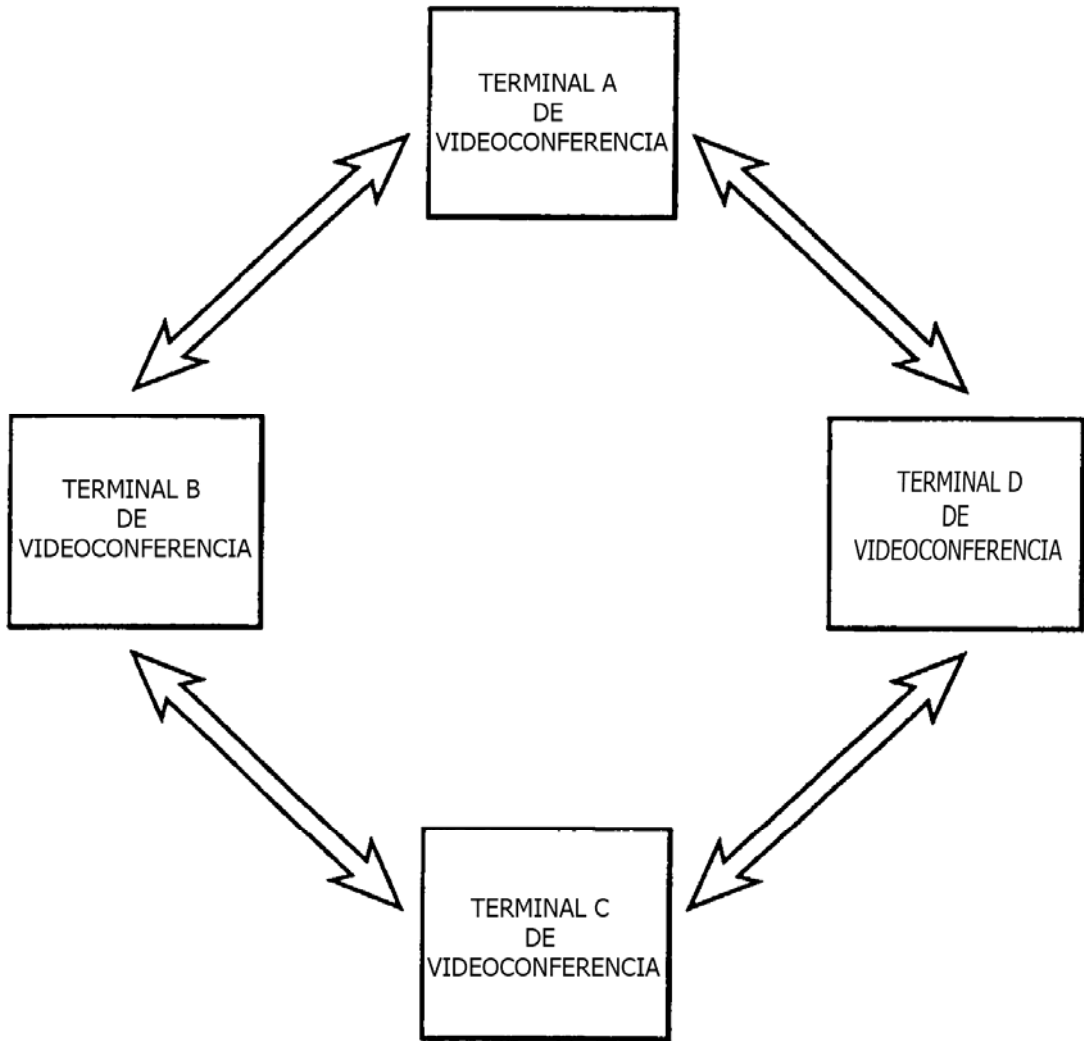
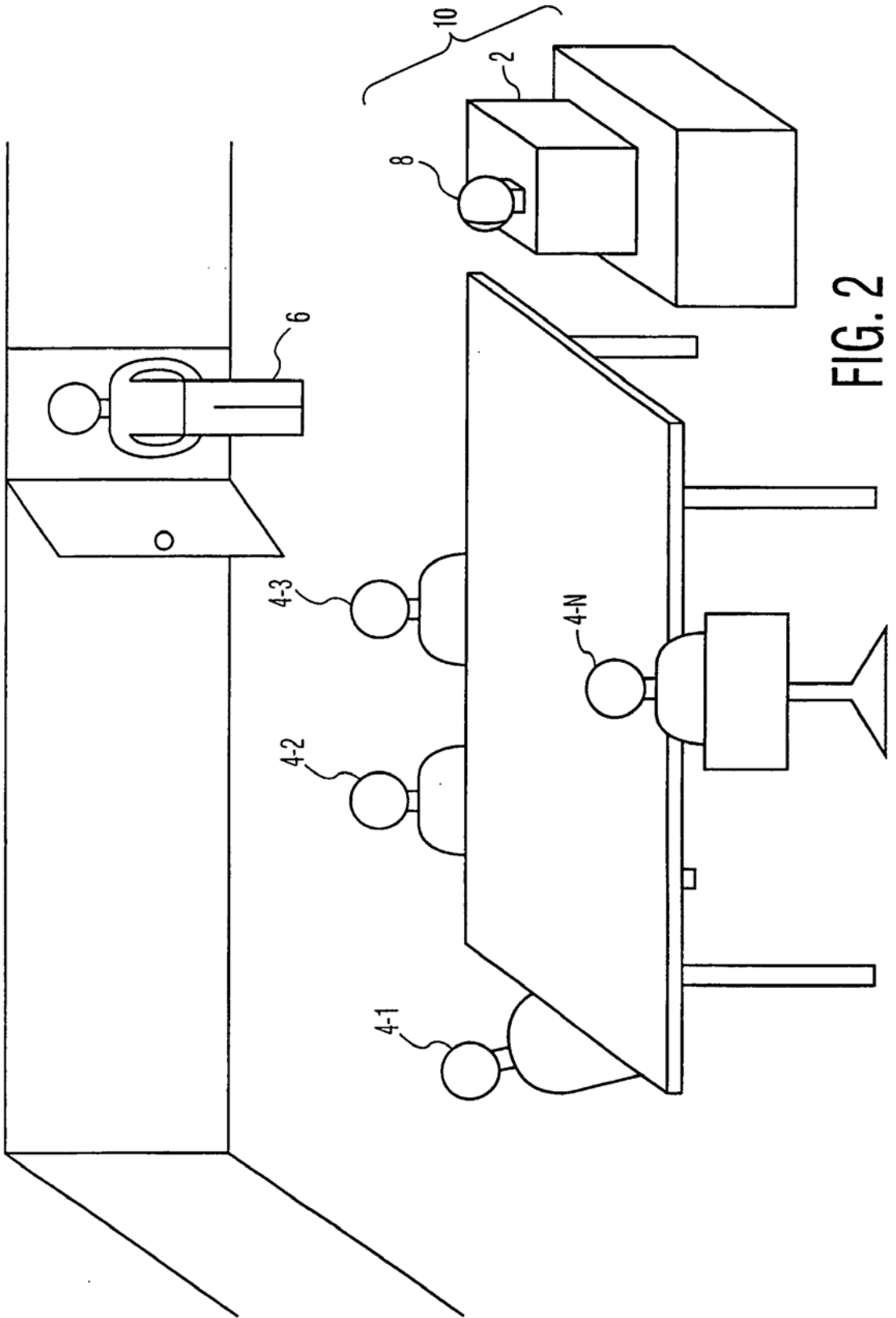


FIG. 1



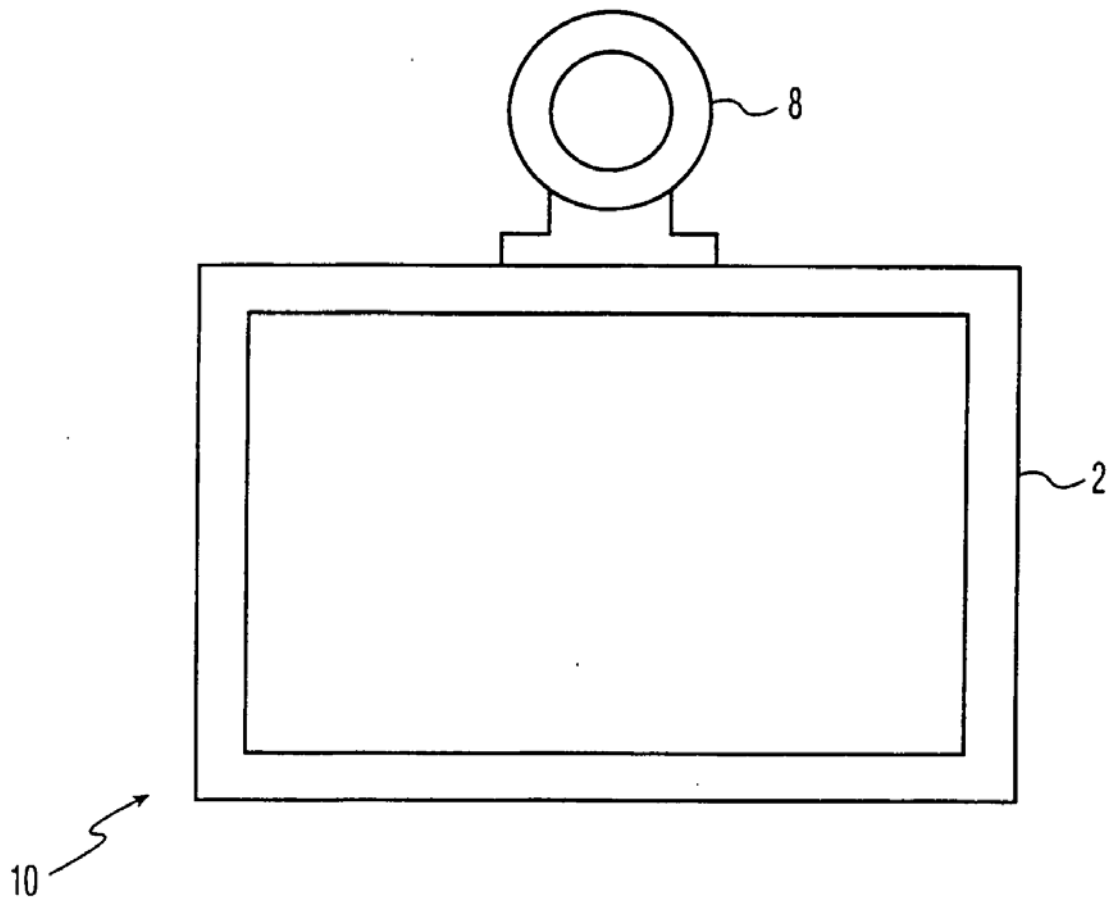


FIG. 3

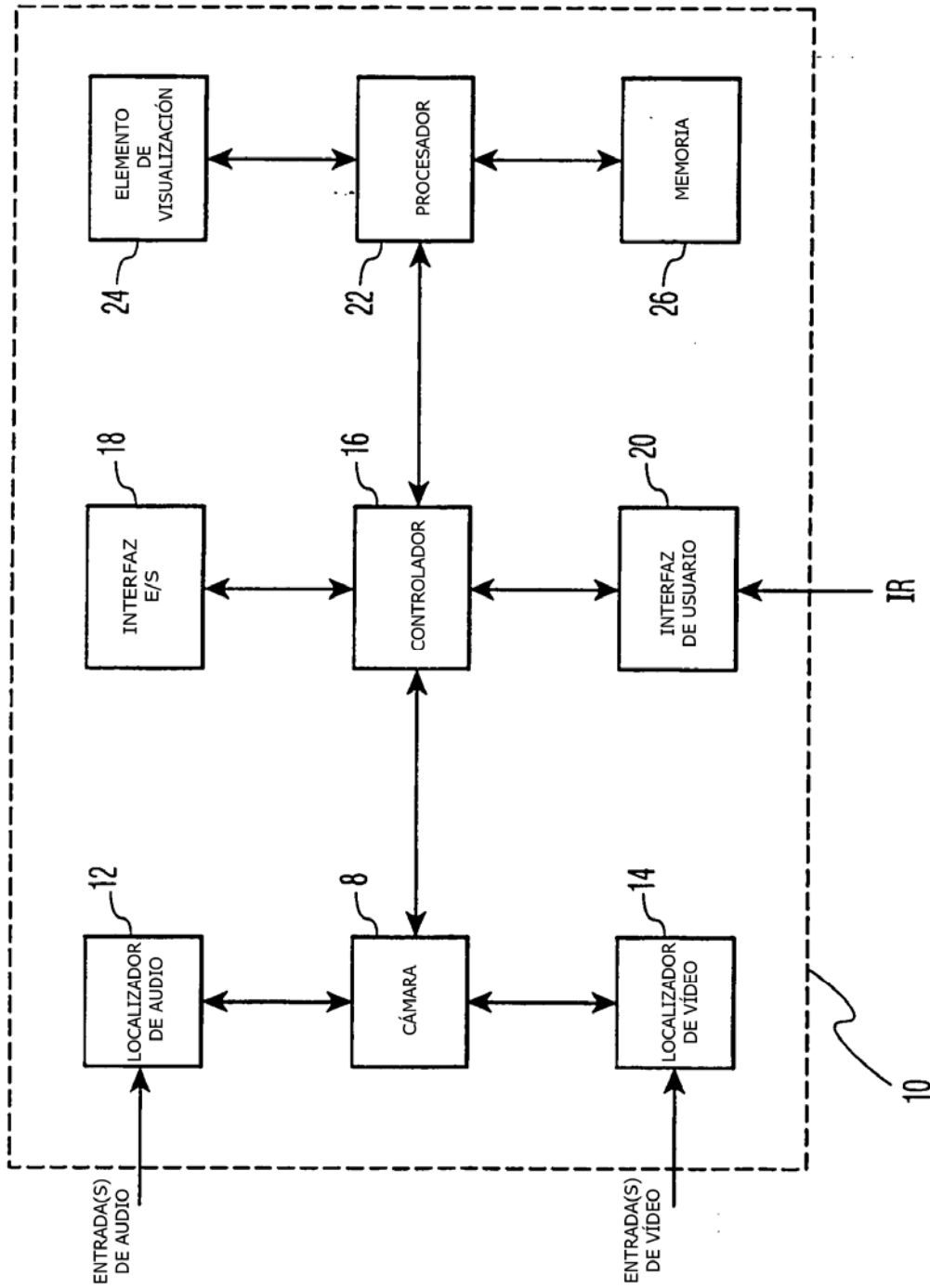


FIG. 4

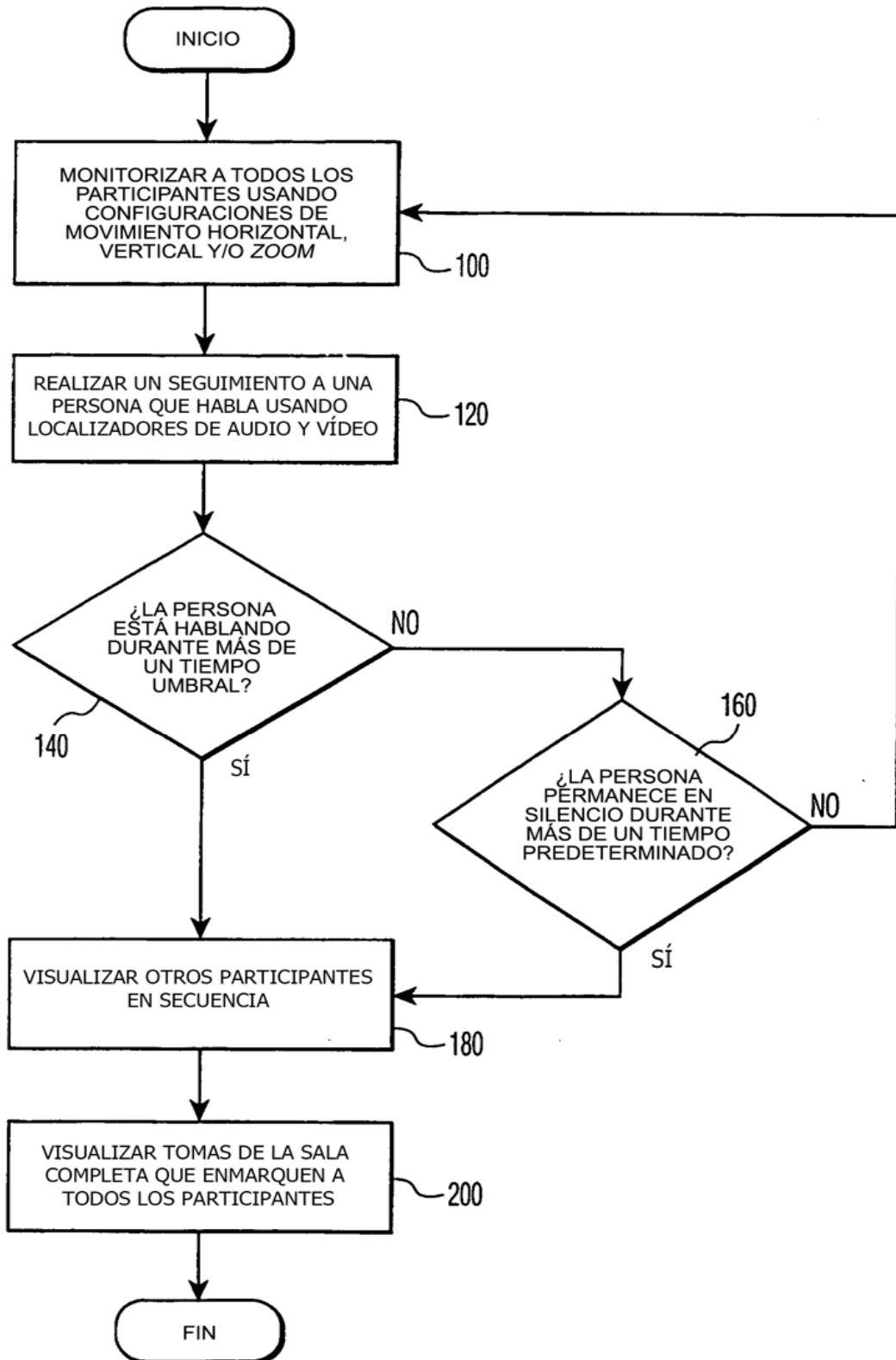


FIG. 5