(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 445 923**

(51) Int. CI.:

H04N 7/15 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Fecha de presentación y número de la solicitud europea: **27.04.2006 E 12164796 (0)**

(97) Fecha y número de publicación de la concesión europea: **13.11.2013 EP 2479986**

(54) Título: **Procesado de audio en una conferencia con múltiples participantes**

(30) Prioridad:

28.04.2005 US 118555

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
06.03.2014

(73) Titular/es:

APPLE INC. (100.0%)
1 Infinite Loop
Cupertino, CA 95014, US

(72) Inventor/es:

JEONG, HYEONKUK y
SALSBURY, RYAN

(74) Agente/Representante:

FÀBREGA SABATÉ, Xavier

DESCRIPCIÓN

Procesado de audio en una conferencia con múltiples participantes

CAMPO DE LA INVENCIÓN

La presente invención se refiere al procesado de audio en una conferencia con múltiples participantes.

ANTECEDENTES DE LA INVENCIÓN

Con la proliferación de los ordenadores de propósito general, ha habido un aumento de la demanda para la realización de conferencias a través de ordenadores personales o de negocios. En tales conferencias, es deseable identificar rápidamente a los participantes que están hablando en un momento dado. Tal identificación, sin embargo, se vuelve difícil a medida que se añaden más participantes, en especial para los participantes que sólo reciben datos de audio. Esto se debe a que las aplicaciones de conferencia anteriores no proporcionan ninguna pista visual o auditiva para ayudar a identificar hablantes activos durante una conferencia. Por lo tanto, existe una necesidad en la técnica de aplicaciones de conferencia que ayuden a un participante a identificar rápidamente los participantes que hablan activamente en la conferencia.

RESUMEN DE LA INVENCIÓN

Algunas realizaciones proporcionan una arquitectura para establecer conferencias de audio con múltiples participantes a través de una red de ordenadores. Esta arquitectura tiene un distribuidor central que recibe señales de audio de uno o más participantes. El distribuidor central mezcla las señales recibidas y las transmite de vuelta a los participantes. En algunas realizaciones, el distribuidor central elimina eco eliminando la señal de audio de cada participante de la señal mezclada que el distribuidor central envía al participante en particular.

En algunas realizaciones, el distribuidor central calcula un indicador de intensidad de señal para la señal de audio de cada participante y pasa los indicios calculados junto con la señal de audio mezclado a cada participante. Algunas realizaciones utilizan entonces los signos de intensidad de señal para mostrar medidores de nivel de audio que indican los niveles de volumen de los diferentes participantes. En algunas realizaciones, los medidores de nivel de audio se muestran junto a la imagen o icono de cada participante. Algunas realizaciones utilizan los indicios intensidad de la señal para activar la panoramización de audio.

En algunas realizaciones, el distribuidor central produce una única señal mezclada que incluye el audio de todos los participantes. Esta corriente (junto con los indicios de intensidad de señal) se envía a todos los participantes. Durante la reproducción de este flujo, un participante silenciará la reproducción de fondo si el participante mismo es el principal contribuyente. Este plan proporciona supresión de eco sin necesidad de flujos separados y distintos para cada participante. Este sistema requiere menos computación del distribuidor central. También, a través de multidifusión IP, el distribuidor central puede reducir sus necesidades de ancho de banda.

Los ordenadores del distribuidor central y de los participantes pueden tomar distintas formas. En otras palabras, estos ordenadores pueden integrarse en cualquier tipo de dispositivo, como un ordenador de mesa independiente, ordenador portátil, y/o ordenador de mano u otro dispositivo de electrónica de consumo o de comunicación, centro multimedia, concentrador, etc.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

Las características novedosas de la invención se exponen en las reivindicaciones adjuntas. Sin embargo con fines de explicación, se exponen varias realizaciones en las siguientes figuras.

La Figura 1 ilustra un ejemplo de la arquitectura de conferencia de audio/video de algunas realizaciones de la invención.

Las Figuras 2 y 3 ilustran cómo algunas realizaciones intercambian contenido de audio en una conferencia de audio/video con múltiples participantes.

La Figura 4 muestra los componentes software de la aplicación de conferencia de audio/video de algunas realizaciones de la invención.

La Figura 5 ilustra el módulo de punto focal de algunas realizaciones de la invención.

La Figura 6 es un diagrama de flujo que muestra la generación de audio mezclado por el punto focal en algunas de las realizaciones.

La Figura 7 ilustra cómo el protocolo RTP es utilizado por el módulo de punto focal en algunas realizaciones para transmitir el contenido de audio.

La Figura 8 ilustra el punto no focal de algunas realizaciones de la invención.

La Figura 9 ilustra cómo el protocolo RTP es utilizado por el módulo de punto no focal en algunas realizaciones para transmitir el contenido de audio.

La Figura 10 ilustra conceptualmente el flujo de la operación de decodificación del punto no focal no en algunas realizaciones.

La Figura 11 ilustra los medidores de nivel de audio que se muestran en algunas realizaciones de la invención.

La Figura 12 muestra una disposición ejemplar de las imágenes de los participantes sobre una de las pantallas de los participantes.

La Figura 13 es un diagrama de flujo que ilustra el proceso mediante el cual algunas realizaciones de la invención realizan la panoramización de audio.

DESCRIPCIÓN DETALLADA DE LA INVENCIÓN

En la siguiente descripción, se exponen numerosos detalles con fines de explicación. Sin embargo, un experto en la técnica se dará cuenta de que la invención puede ser llevada a cabo sin el uso de estos detalles específicos. En otros casos, se muestran estructuras y dispositivos bien conocidos en forma de diagrama de bloques para no oscurecer la descripción de la invención con detalles innecesarios.

Algunas realizaciones proporcionan una arquitectura para establecer conferencias de audio/video con múltiples participantes. Esta arquitectura tiene un distribuidor central que recibe las señales de audio de uno o más participantes. El distribuidor central mezcla las señales recibidas y las transmite de vuelta a los participantes. En algunas realizaciones, el distribuidor central elimina eco eliminando de señal de audio de cada participante de la señal mezclada que el distribuidor central envía al participante en particular.

En algunas realizaciones, el distribuidor central calcula un indicador de intensidad de señal para la señal de audio de cada participante y pasa los indicios calculados junto con la señal de audio mezclado a cada participante. Algunas realizaciones a continuación, utilizan los indicios de intensidad de señal para mostrar medidores de nivel de audio que indican los niveles de volumen de los diferentes participantes. En algunas realizaciones, los medidores de nivel de audio se muestran junto a la imagen o icono de cada participante. Algunas realizaciones utilizan los indicios intensidad de señal para activar la panoramización de audio.

Varias realizaciones detalladas de la invención se describen a continuación. En estas realizaciones, el distribuidor central es el ordenador de uno de los participantes en la conferencia de audio/video. Un experto se dará cuenta de que otras realizaciones se implementan de forma diferente. Por ejemplo, en algunas realizaciones el distribuidor central no es el ordenador de ninguno de los participantes en la conferencia.

I. VISIÓN GENERAL

La Figura 1 ilustra un ejemplo de arquitectura de conferencia 100 de algunas realizaciones de la invención. Esta arquitectura permite que varios participantes participen en una conferencia a través de varios ordenadores que están conectados mediante una red de ordenadores. En el ejemplo ilustrado en Figura 1, cuatro participantes A, B, C y D se participan en la conferencia a través de sus cuatro ordenadores y una red 105-120 (no mostrada) que conecta estos ordenadores. La red que conecta a estos ordenador s pueden ser cualquier red, como una red de área local, una red de área extensa, una red de redes (por ejemplo, Internet), etc.

La conferencia puede ser una conferencia de audio/video, o una conferencia solamente de audio, o una conferencia de audio/video para algunos de los participantes y una conferencia sólo de audio para los demás participantes. Durante la conferencia, el ordenador 105 de uno de los participantes (participante D en este ejemplo) sirve como distribuidor central de contenido de audio y/o vídeo (es decir, contenido de audio/video), como se muestra en Figura 1. Este distribuidor central 125 se denomina más adelante punto focal de la conferencia de múltiples participantes. Los ordenadores de los demás participantes se denominan a continuación máquinas no focales u ordenadores no focales.

Además, la discusión a continuación se centra en las operaciones de audio de los ordenadores focales y no focales. El funcionamiento de vídeo de estos ordenadores se describe más detalladamente en la solicitud de patente estadounidense 11/118,553 titulada "Video Processing in a Multi-Participant Video Conference", presentada con el número expediente de agente APLE.P0091. Además, en la solicitud de patente estadounidense 11/118,931 titulada " Multi-Participant Video Conference Setup", presentada con el número expediente de agente APLE.P0084, describe

cómo establecen algunas realizaciones una conferencia con múltiples participantes a través de una arquitectura de punto focal, como la que se ilustra en la Figura 1. Ambas solicitudes se incorporan aquí por referencia.

Como distribuidor central de contenidos de audio/video, el punto focal 125 recibe las señales de audio de cada participante, mezcla y codifica estas señales, y luego transmite la señal mezclada a cada una de los ordenadores no focales. La Figura 2 muestra un ejemplo de dicho intercambio de señales de audio para el ejemplo con cuatro participantes de Figura 1. Específicamente, la Figura 2 ilustra al punto focal 125 recibiendo las señales de audio comprimidas 205-215 de otros participantes. A partir de las señales de audio recibidas 205-215, el punto focal 125 genera una señal de audio mezclado 220 que incluye cada una de las señales recibidas de audio y la señal de audio del participante utilizando el ordenador de punto focal. El punto focal 125 comprime y transmite la señal de audio mezclado 220 a cada máquina no focal 110, 115, y 120.

En el ejemplo ilustrado en la Figura 2, la señal de audio mezclado 220 que se transmite a cada participante no focal en particular incluye también la señal de audio del participante no focal particular. En algunas realizaciones, sin embargo, el punto focal elimina una señal de audio de participante no focal particular de la señal de audio mezclado que el punto focal transmite al particular no focal particular. En estas realizaciones, el punto focal 125 elimina la señal de audio propia de cada participante de su señal de audio mezclado correspondiente con el fin de eliminar el eco cuando el audio mezclado se reproduce en altavoces del ordenador del participante.

La Figura 3 ilustra un ejemplo de esta eliminación para el ejemplo ilustrado en la Figura 2. Específicamente, la Figura 3 ilustra (1) para un participante, una señal de audio mezclado 305 que no tiene A la señal de audio 205 propia del participante A, (2) para el participante B, una señal de audio mezclado 310 que no tiene la señal de audio propia del participante B 210, y (3) para el participante C, una señal de audio mezclado 315 que no tiene la señal de audio 215 propia del participante C.

Como se muestra en Figura 3, el punto focal 125 en algunas realizaciones calcula indicios de intensidad de señal para las señales de audio de los participantes, y adjunta los indicios de intensidad de señal mezclada a las señales que envía a cada participante. Los ordenadores no focales a continuación, utilizan los indicios de intensidad de señal adjuntos para mostrar medidores de nivel de audio que indican los niveles de volumen de los diferentes participantes. En algunas realizaciones, los medidores de nivel de audio se muestran junto a la imagen o icono de cada participante.

Algunas realizaciones también utilizan los indicios de intensidad de señal transmitida para barrer el audio a lo largo de los altavoces del ordenador de un participante, con el fin de ayudar a identificar oradores durante la conferencia. Esta panoramización crea un efecto tal que el audio asociado con un participante en particular se percibe como originario de una dirección que refleja la posición en la pantalla de la imagen o icono de ese participante. El efecto de panoramización se crea mediante la introducción de pequeños retardos de los canales izquierdo o derecho. El efecto posicional se basa en la percepción cerebral de pequeños retardos y diferencias de fase. Los medidores de nivel de audio y la panoramización de audio se describen a continuación más detalladamente.

Algunas realizaciones son implementadas por una aplicación de conferencia de audio/video que puede llevar a cabo tanto operaciones focales como no focales. La Figura 4 ilustra una arquitectura de software para una aplicación como esta. En concreto, esta Figura muestra una aplicación de conferencia de audio/video 405 que consta de dos módulos, un módulo de punto focal 410 y un módulo de punto no focal 415. Estos dos módulos 410 y 415, y la aplicación de conferencia de audio/video 405, se ejecutan sobre un sistema operativo del ordenador 420 de un participante en la conferencia.

Durante una conferencia con múltiples participantes, la aplicación de conferencia de audio/video 405 utiliza el módulo de punto focal 410, cuando esta aplicación está sirviendo como punto focal de la conferencia, o utiliza el módulo de punto no focal 415 cuando no está actuando como punto focal. El módulo de punto focal 410 lleva a cabo operaciones de procesado de audio de punto focal cuando la aplicación de conferencia de audio/video 405 es el punto focal de una conferencia de audio/video con múltiples participantes. Por otro lado, el módulo de punto focal 415 lleva a cabo las operaciones de procesado de audio de punto no focal cuando la aplicación 405 no es el punto focal de la conferencia. En algunas realizaciones, los módulos de punto focal y no focal 410 y 415 comparten determinados recursos.

El módulo de punto focal 410 se describe en la Sección II de este documento, mientras que el módulo de punto no focal 415 se describe en la Sección III.

II. EL MÓDULO DE PUNTO FOCAL

La Figura 5 ilustra el módulo de punto focal 410 de algunas realizaciones de la invención. El módulo de punto focal 410 se muestra durante una conferencia de audio/video con múltiples participantes. Con el fin de generalizar las operaciones de punto focal, el ejemplo de Figura 5 se ilustra teniendo un número arbitrario de participantes. Este número arbitrario se denota como "n", que representa un número mayor que 2. El módulo de punto focal 410 genera señales de audio mezclados para transmitir a los participantes no focales y lleva a cabo la presentación de audio

para el participante de la conferencia que está utilizando el ordenador de punto focal durante la videoconferencia. Para su funcionamiento de mezcla de audio, el módulo de punto focal 410 utiliza (1) un decodificador 525 y una memoria temporal intermedia 530 para cada señal de audio entrante, (2) una memoria temporal intermedia 532 para la señal de audio de punto focal, (3) un módulo de captura de audio 515, (3) una calculadora de intensidad de señal de audio 580, y (4) un mezclador de audio 535 para cada señal de audio mezclado transmitida, y un codificador 550 para cada señal de audio mezclado transmitida. Para su funcionamiento de presentación de audio el ordenador de punto focal, el módulo de punto focal 410 también utiliza un mezclador de audio 545, un control de panoramización de audio 560 y un control de medidor de nivel 570.

El funcionamiento de mezcla de audio del módulo de punto focal 410 se describirá ahora con referencia al proceso de mezcla 600 que ilustra conceptualmente el flujo de funcionamiento en la Figura 6. El funcionamiento de la presentación de audio del módulo de punto focal se describe a continuación en la Sección III, junto con la presentación de audio del módulo de punto no focal.

Durante el proceso de mezcla de audio 600, dos o más decodificadores 525 reciben (en 605) dos o más señales de audio digitales 510 que contienen muestras de audio de dos o más módulos de puntos no focales. En algunas realizaciones, las señales de audio recibidas están codificadas por los mismos o diferentes códecs de audio en los ordenadores no focales. Ejemplos de estos códecs incluyen Qualcomm PureVoice, GSM, G.711 y códecs de audio ILBC.

El decodificador 525 decodifica y almacena (en 605) las señales de audio decodificadas en dos o más memorias temporales intermedias 530. En algunas realizaciones, el decodificador 525 para el flujo de audio de cada ordenador que no sea el focal, utiliza un algoritmo de decodificación que es adecuado para el códec de audio utilizado por el ordenador no focal. Este decodificador se especifica durante el proceso que establece la conferencia de audio/video.

El módulo de punto focal 410 también captura audio del participante que está utilizando el ordenador de punto focal, a través del micrófono 520 y el módulo de captura de audio 515. En consecuencia, después de 605, el módulo de punto focal (en 610) capta una señal de audio del participante de punto focal y almacena esta señal de audio capturada en su memoria temporal intermedia correspondiente 532.

A continuación, en 615, el calculador de intensidad de señal de audio 580 calcula indicios de intensidad de señal correspondientes a la intensidad de cada señal recibida. El calculador de intensidad de señal de audio 580 asigna un peso a cada señal. En algunas realizaciones, el calculador de intensidad de señal de audio 580 calcula indicios de intensidad de señal como la potencia en valor cuadrático medio (RMS) del flujo de audio proveniente del participante al punto focal. La intensidad RMS se calcula a partir de la siguiente fórmula:

$$\mathrm{RMS} = \sqrt{\frac{\sum_{i=1}^{N} (\,\mathrm{Muestra}_i\,)^2}{N}},$$

donde N es el número de muestras utilizadas para calcular la intensidad RMS y $\mathrm{Muestra}_i$ es la amplitud de i-ésima muestra. El número de muestras, N, que el calculador de intensidad de señal de audio 580 utiliza para calcular el valor RMS depende de la tasa de muestreo de la señal. Por ejemplo, en algunas realizaciones de la invención donde la tasa de muestreo es de 8 KHz, la intensidad RMS puede ser calculada utilizando un trozo de 20 ms de datos de audio que contiene 160 muestras. Otras tasas de muestreo pueden requerir un número diferente de muestras.

A continuación, en 620, el proceso 600 utiliza los mezcladores de audio 535 y 545 para mezclar las señales de audio alacenadas temporalmente. Cada mezclador de audio 535 y 545 genera señales de audio mezclado de uno de los participantes. La señal de audio mezclado para cada participante en particular incluye las señales de audio de todos los participantes, excepto la señal de audio del participante en particular. La eliminación de la señal de audio de un participante en particular de la mezcla que el participante en particular recibe elimina el eco cuando el sonido mezclado se reproduce en los altavoces del ordenador del participante. Los mezcladores 535 y 545 mezclan las señales de audio generando (en 620) una suma ponderada de estas señales. Para obtener un valor de muestra de audio en un instante de muestreo en particular, se añaden todas las muestras en un momento de muestreo en particular en base a los valores de ponderación calculados por el calculador de intensidad de señal de audio 580. En algunas realizaciones, los valores de ponderación se determinan dinámicamente en base a indicios de intensidad de señal calculados en 615 para alcanzar ciertos objetivos. Ejemplos de tales objetivos incluyen (1) la eliminación de las

señales más débiles, que son típicamente atribuibles al ruido, y (2) la prevención de que una señal de audio de un participante abrume las señales de otros participantes, lo que acontece a menudo cuando uno de los participantes siempre habla más fuerte que otro o tiene mejor equipo de audio que el otro.

En algunas realizaciones, los mezcladores 535 y 545 adjuntan (en 625) los indicios de intensidad de señal de todas las señales de audio que se sumaron para generar la señal mezclada. Por ejemplo, la Figura 7 ilustra un paquete RTP (Protocolo de Transporte de Tiempo Real) 700 que algunas realizaciones utilizan para enviar una señal de audio mezclado 705 a un participante en particular. Como se muestra en esta Figura, los indicios de intensidad de señal 710-720 se adjuntan al final del paquete RTP 705.

A continuación, para el audio de los ordenadores no focales, los codificadores (550 a 630) codifican las señales de audio mezclado y las envían (en 635) a sus correspondientes módulos no focales. La señal de audio mezclado para el ordenador de punto focal se envía (en 635) sin codificar al control de panoramización de audio 560. Además, en 635, los indicios de intensidad de señal se envían al medidor de nivel 570 del módulo de punto focal, el cual genera entonces los indicadores de nivel de volumen adecuados para su visualización en el dispositivo de visualización 575 del ordenador de punto focal.

Después de 635, el proceso de mezcla de audio 600 determina (en 640) si el participante de conferencia de audio/video con múltiples participantes ha finalizado. Si es así, el proceso 600 finaliza. De lo contrario, el proceso vuelve a 605 para recibir y decodificar señales de audio entrantes.

Un experto medio se dará cuenta de que otras realizaciones podrían implementar el módulo de punto focal 410 de forma diferente. Por ejemplo, en algunas realizaciones, el punto focal 410 produce una única señal mezclada que incluye audio de cada participante. Este flujo junto con los indicios de intensidad de señal se envía a todos los participantes. Durante la reproducción de este flujo, un participante silenciará la reproducción si el participante mismo es el principal contribuyente. Este sistema ahorra tiempo de computación en el punto focal y proporciona supresión de eco sin necesidad de flujos separados y distintos para cada participante. También, durante multidifusión IP, se puede reducir el ancho de banda del flujo de punto focal. En estas realizaciones, el punto focal 410 tiene un mezclador de audio 535 y un codificador 550.

III. EL MÓDULO DE PUNTO NO FOCAL

La Figura 8 ilustra un módulo de punto no focal 415 de una conferencia de audio/video de algunas realizaciones de la invención. En este ejemplo, el módulo de punto no focal 415 utiliza un decodificador 805, dos memorias temporales 810 y 880, un control de medidor de nivel 820, un control de panoramización de audio 845, un módulo de captura de audio 875 y un codificador 870.

El módulo de punto focal no lleva a cabo operaciones de codificación y decodificación. Durante la operación de codificación, la señal de audio del micrófono del participante punto no foco de 860 es capturado por el módulo de captura de audio 875 y se almacena en su 880 correspondiente tampón intermedia. El codificador 870 codifica entonces el contenido del buffer intermedio 880 y lo envía al módulo de punto focal 410.

En algunas realizaciones que utilizan Protocolo de Transporte en Tiempo Real (RTP) para intercambiar señales de audio, la señal codificada de audio del participante no focal se envía al módulo de punto focal en un paquete 900 que incluye cabeceras RTP 910 junto con audio codificado 920, como se muestra en Figura 9.

El funcionamiento de decodificación del módulo de punto no focal 415 se describirá ahora con referencia al proceso 1000 que ilustra conceptualmente el flujo de operación en la Figura 10. Durante la operación de decodificación, el decodificador 805 recibe (en 1005) paquetes de audio del módulo de punto focal 410. El decodificador 805 decodifica (en 1010) cada paquete de audio recibido para obtener datos de audio mezclado e indicios de intensidad de señal asociados con los datos de audio. El decodificador 805 guarda (en 1010) los resultados en la memoria temporal 810.

Los indicios de intensidad de señal se envían al control de medidor de nivel 820 para mostrar (en 1015) los medidores de nivel de audio en la pantalla del participante no focal 830. En una conferencia de audio/video con múltiples participantes, es deseable identificar altavoces activos. Una característica novedosa de la presente invención es la de representar las intensidades de audio mediante la visualización del nivel de audio correspondiente a la intensidad de voz de cada orador. Los medidores de nivel que aparecen en la pantalla de cada participante expresan el nivel de volumen de los diferentes participantes, mientras que la señal de audio mezclado se está escuchando por los altavoces 855. El nivel de volumen de cada participante puede representarse mediante un medidor de nivel separado, con lo cual se permite al espectador conocer los altavoces activos y el nivel de audio de cada participante en cualquier momento.

Los medidores de nivel son particularmente útiles cuando algunos de los participantes sólo reciben señales de audio durante la conferencia (es decir, algunos de los participantes son " participantes solo de audio"). Estos participantes no tienen las imágenes de vídeo para ayudar a proporcionar una indicación visual de los participantes que están

hablando. La Figura 11 ilustra un ejemplo del uso de medidores de nivel en una conferencia solamente de audio de algunas realizaciones. En esta figura, el nivel de audio de cada participante 1110-1115 se coloca al lado del icono de ese participante 1120-1125. Tal y como se ilustra en Figura 11, algunas realizaciones muestran el nivel de voz del micrófono local 1130 por separado en la parte inferior de la pantalla. Un experto en la técnica debería comprender que la Figura 11 es sólo un ejemplo de la forma de mostrar los medidores de nivel en la pantalla de un participante. Pueden hacerse otras disposiciones de visualización sin alejarse de las enseñanzas de esta invención para calcular y la visualizar la intensidad relativa de las señales de audio en una conferencia.

Tras 1015, la señal de audio mezclado decodificada y los indicios de intensidad de señal almacenados en la memoria temporal intermedia 810 se envían (en 1020) al control de panoramización de audio 845 para controlar los altavoces de los participantes no focales 855. La operación de panoramización de audio se describe más adelante con referencia a las Figuras 12 y 13.

Después de 1020, el proceso de decodificación de audio 1000 determina (en 1025) si la conferencia de audio/video con múltiples participantes ha terminado. Si es así, el proceso 1000 finaliza. De lo contrario, el proceso vuelve a 1005 para recibir y decodificar señales de audio entrantes.

El uso de panoramización de audio para hacer que la ubicación de audio percibida coincida con la ubicación de vídeo es otra característica novedosa de la invención actual. Con el fin de ilustrar cómo se realiza la panoramización de audio, la Figura 12 ilustra un ejemplo de una presentación en pantalla 1200 de videoconferencia en el caso de cuatro participantes en una videoconferencia. Como se muestra en la Figura 12, Las imágenes de los otros tres participantes 1205-1215 se muestran horizontalmente en la pantalla de presentación 1200. La imagen propia del participante local 1220 se muestra opcionalmente con un tamaño más pequeño en relación a las imágenes de los otros participantes 1205-1215 en la parte inferior de la presentación de pantalla 1200.

Algunas realizaciones logran panoramización de audio mediante una combinación de retardo de la señal y ajuste de la amplitud de la señal. Por ejemplo, cuando habla el participante cuya imagen 1205 se coloca en el lado izquierdo de la pantalla, el sonido procedente del altavoz derecho se cambia mediante una combinación de introducir un retardo y ajustar la amplitud para dar la sensación de que la voz viene desde el altavoz izquierdo.

La Figura 13 ilustra un proceso 1300 mediante el que funciona el control de panoramización de audio del módulo no focal 845 en algunas realizaciones de la invención. Los indicios de intensidad de señal de cada señal de audio en la señal de audio mezclado se utiliza (en 1310) para identificar al participante que más contribuye a la señal de audio mezclado decodificada. A continuación, el proceso identifica (en 1315) la ubicación del participante o participantes identificados en 1310. El proceso utiliza entonces (en 1320-1330) una combinación de ajuste de amplitud y retardo de la señal para crear el efecto estéreo. Por ejemplo, si está hablando el participante cuya imagen 1205 se muestra en el lado izquierdo del dispositivo de visualización 1200, se introduce un retardo (en 1325) en el altavoz derecho y la amplitud del altavoz derecho se reduce opcionalmente para hacer que la señal desde el altavoz izquierdo parezca ser más fuerte.

De manera similar, si está hablando el participante cuya imagen 1215 se muestra en el lado derecho del dispositivo de visualización 1200, se introduce un retardo (en 1330) en el altavoz izquierdo y la amplitud del altavoz izquierdo se reduce opcionalmente para hacer que la señal del altavoz derecho parezca ser más fuerte. En contraste, si está hablando el participante cuya imagen 1210 se muestra en el centro del dispositivo de visualización 1200, no se hacen ajustes a las señales enviadas a los altavoces.

La panoramización de audio ayuda a identificar la ubicación de los participantes que están hablando en la pantalla y produce contabilidad estéreo para ubicación. En algunas realizaciones de la invención, se introduce un retardo de aproximadamente 1 milisegundo (1/1000 segundo) y la amplitud se reduce en un 5 a 10 por ciento durante el funcionamiento de la panoramización de audio. Un experto en la técnica, sin embargo, se dará cuenta de que podrían utilizarse otras combinaciones de ajustes de amplitud y retardos para crear un efecto similar.

En algunas realizaciones, algunas acciones de los participantes, tales como unirse a la conferencia, abandonar la conferencia, etc., pueden activar efectos de sonido de la interfaz en los ordenadores de otros participantes. Estos efectos de sonido también pueden ser panoramizados para indicar qué participante realiza la acción asociada.

En las realizaciones en las que el punto focal es también un participante en la conferencia (tales como la realización ilustrada en Figura 1), el módulo de punto focal también utiliza los procedimientos anteriormente descritos para presentar el audio al participante cuyo ordenador sirve como punto focal de la conferencia.

Aunque la invención ha sido descrita con referencia a numerosos detalles específicos, un experto en la técnica reconocerá que la invención puede realizarse de otras formas específicas. En otros lugares, pueden hacerse diversos cambios, y pueden sustituirse equivalentes por elementos descritos sin alejarse del verdadero alcance de la presente invención. Así, un experto en la técnica comprenderá que la invención no está limitada por los detalles ilustrativos anteriores, sino que debe estar definida por las reivindicaciones adjuntas.

REIVINDICACIONES

1. Un procedimiento para crear un efecto de panoramización estéreo en una conferencia multimedia entre una pluralidad de participantes, comprendiendo el procedimiento:

determinar que un segundo participante en la conferencia realiza una acción que provoca un efecto de sonido de interfaz de usuario que se reproducirá en un dispositivo de un primer participante;

identificar una ubicación de una presentación de vídeo del segundo participante en un dispositivo de visualización del primer participante que muestra presentaciones de vídeo de al menos el segundo participante y un tercer participante; y

en base a la ubicación identificada, panoramizar el efecto de sonido para la acción llevada a cabo a través de los altavoces de audio en el dispositivo del primer participante a fin de que el sonido asociado a la acción aparezca como originario del lugar identificado de la presentación del video del segundo participante.

2. El procedimiento según la reivindicación 1, en el que panoramizar el efecto de sonido de interfaz de usuario comprende crear un retardo en al menos uno de los altavoces de audio del primer participante.

3. El procedimiento según la reivindicación 1, en el que la acción que desencadena el efecto de sonido de interfaz de usuario a reproducir comprende unirse a la conferencia.

4. El procedimiento según la reivindicación 1, en el que la acción que desencadena el efecto de sonido de interfaz de usuario a reproducir comprende salirse de la conferencia.

5. El procedimiento según la reivindicación 1, en el que la panoramización comprende reducir una amplitud de audio de al menos uno de los altavoces de audio del primer participante.

6. El procedimiento según la reivindicación 1, en el que el dispositivo del primer participante es un dispositivo distribuidor central para la conferencia multimedia.

7. El procedimiento según la reivindicación 1, en el que el dispositivo del primer participante es un dispositivo distribuidor no central para la conferencia multimedia.

8. Un medio legible por máquina que almacena un programa de ordenador que cuando se ejecuta por al menos una unidad de procesamiento de un dispositivo de un primer participante crea un efecto de panoramización estéreo en una conferencia multimedia entre una pluralidad de participantes, incluyendo al primer participante, comprendiendo el programa de ordenador conjuntos de instrucciones para:

determinar que un segundo participante en la conferencia realiza una acción que provoca un efecto de sonido de interfaz de usuario que se reproducirá en un dispositivo de un primer participante;

identificar una ubicación de una presentación de vídeo del segundo participante en un dispositivo de visualización del primer participante que muestra presentaciones de vídeo de al menos el segundo participante y un tercer participante; y

en base a la ubicación identificada, panoramizar el efecto de sonido para la acción que se realiza a través de los altavoces de audio en el dispositivo del primer participante a fin de que el sonido asociado a la acción aparezca como originario del lugar identificado de la presentación del video del segundo participante.

9. El medio legible por máquina según la reivindicación 8, en el que el conjunto de instrucciones para panoramizar el efecto de sonido comprende conjuntos de instrucciones para:

crear un retardo en al menos uno de los altavoces de audio del primer participante; y

reducir una amplitud de audio del al menos un altavoz de audio del primer participante.

10. El medio legible por máquina según la reivindicación 8, en el que la presentación de vídeo del segundo participante se encuentra en el lado derecho del dispositivo de visualización del primer participante, en el que el al menos un altavoz de audio es un altavoz izquierdo del primer participante.

11. El medio legible por máquina según la reivindicación 8, en el que la acción que desencadena el efecto de sonido de interfaz de usuario a reproducir comprende uno de unirse y salir de la conferencia.

12. El medio legible por máquina según la reivindicación 8, en el que el segundo dispositivo de participante es un dispositivo distribuidor central de la conferencia multimedia, comprendiendo además el programa de ordenador conjuntos de instrucciones para:

recibir una señal de audio mezclada desde el segundo dispositivo participante, la señal de audio mixto comprendiendo señales de audio de los segundos y terceros participantes; y

panoramizar el audio mezclado a través de los altavoces de audio con el fin de crear un efecto de que una localización percibida de una señal de audio de un participante en particular coincide con la ubicación de la representación de vídeo del participante en particular en el dispositivo de visualización.

13. El medio legible por máquina según la reivindicación 8, en el que el primer dispositivo participante es un dispositivo distribuidor central de la conferencia multimedia, comprendiendo además el programa de ordenador conjuntos de instrucciones para:

recibir señales de audio desde los segundo y tercer dispositivos participantes; y

generar señales de audio mezcladas de las señales de audio recibidas y audio capturado localmente en el primer dispositivo participante.

14. El medio legible por máquina según la reivindicación 13, en el que el programa de ordenador comprende un conjunto de instrucciones para transmitir las señales de audio mezcladas a los segundo y tercer dispositivos participantes.

15. El medio legible por máquina según la reivindicación 13, en el que el programa de ordenador comprende un conjunto de instrucciones para entregar una señal de audio mezclada en el primer dispositivo participante.

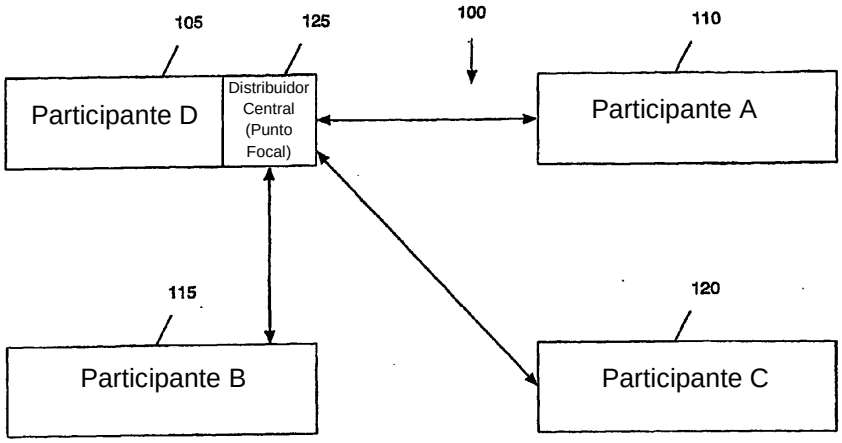

Figura 1

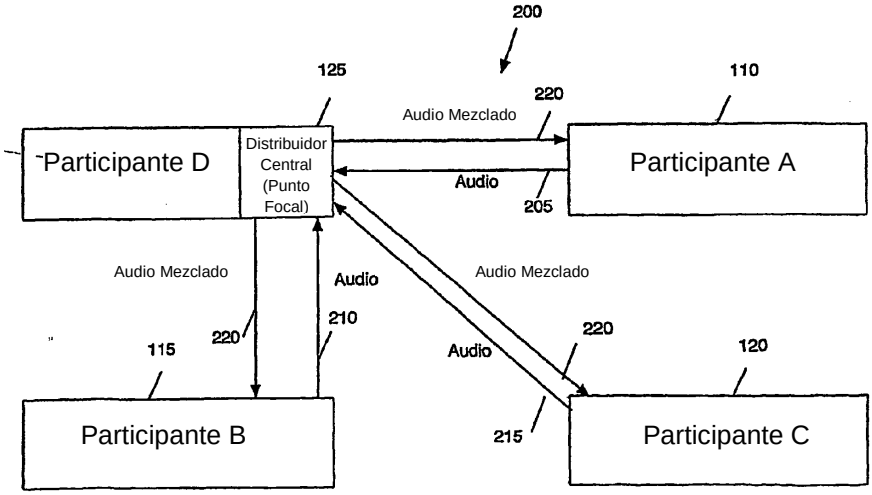

Figura 2

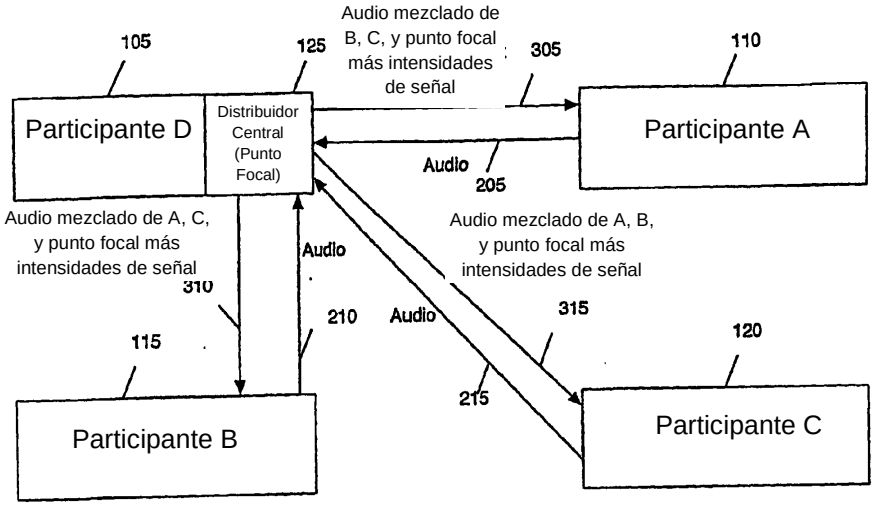

Figura 3

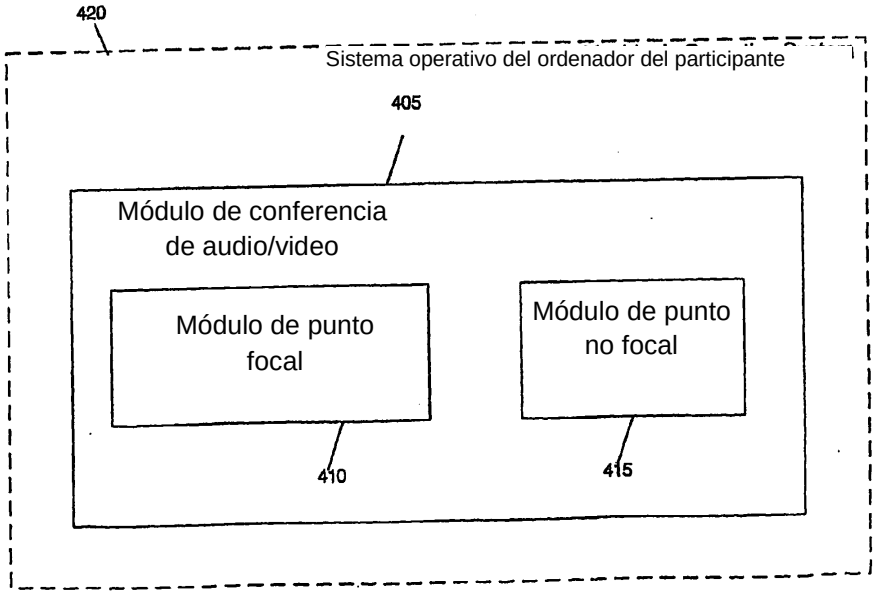

Figura 4

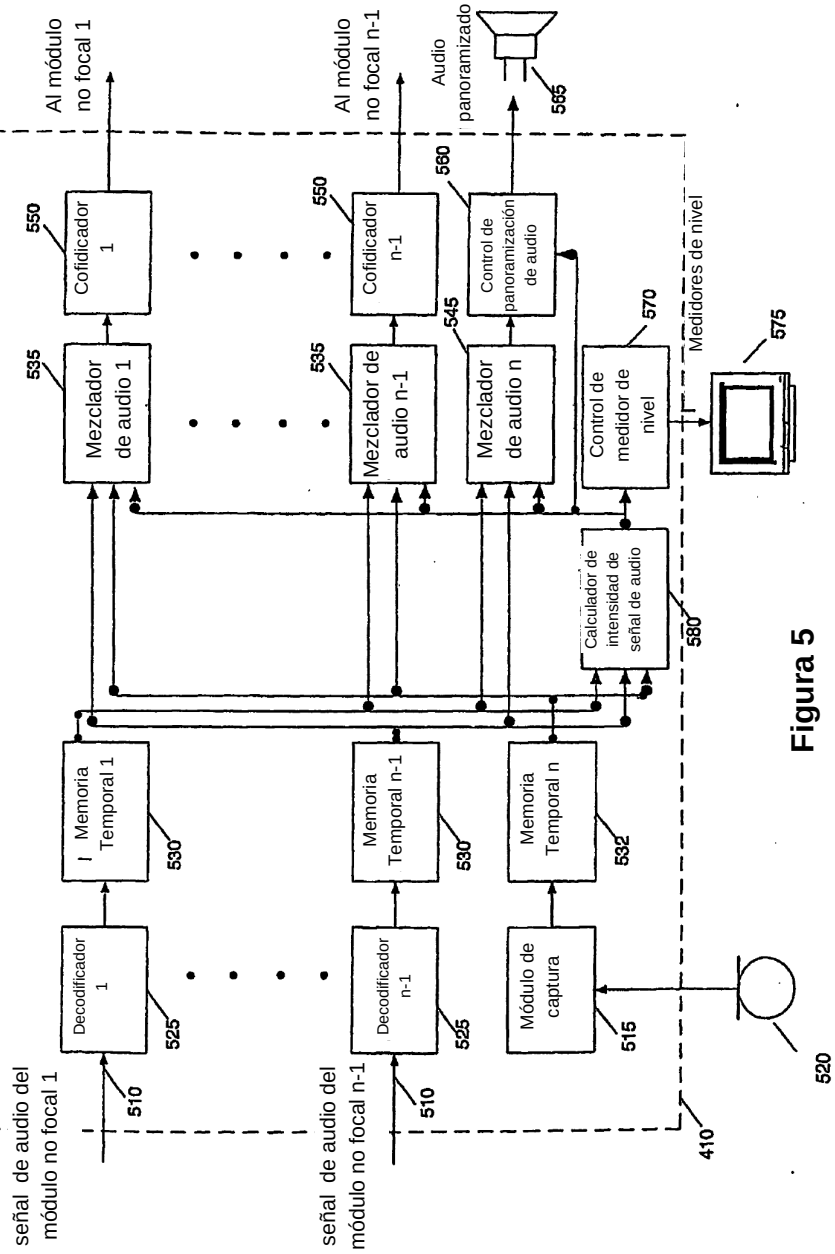
Al módulo no focal 1
Al módulo no focal n-1
Audio panoramizado
565
550
Cofidicador 1
Cofidicador n-1
560
Control de panoramización de audio
535
Mezclador de audio 1
Mezclador de audio n-1
545
Mezclador de audio n
570
Control de medidor de nivel
Medidores de nivel
575
Calculador de intensidad de señal de audio
580
Memoria Temporal 1
Memoria Temporal n-1
530
Memoria Temporal n
532
Decodificador 1
Decodificador n-1
525
Módulo de captura
515
520
510
señal de audio del módulo no focal 1
señal de audio del módulo no focal n-1
410

Figura 5

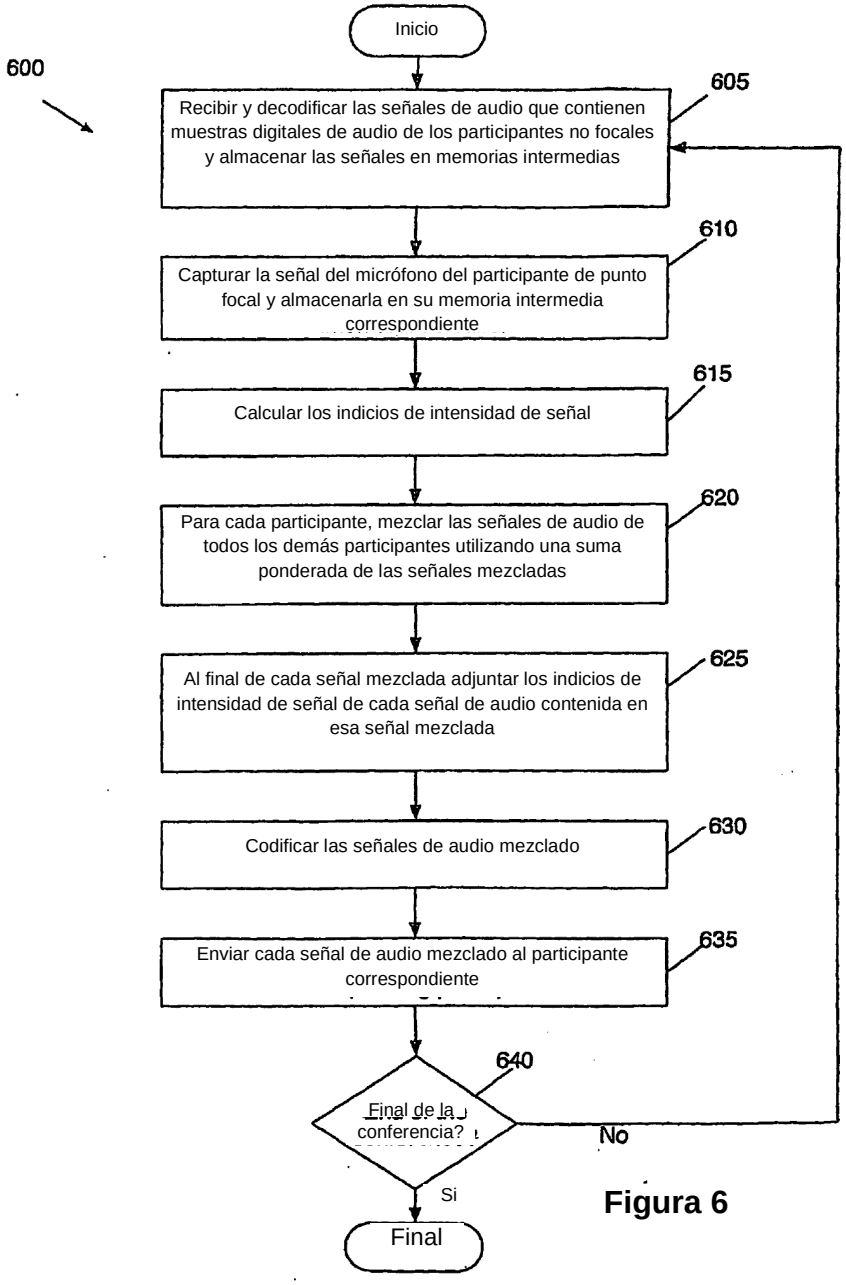
600
Inicio
605
Recibir y decodificar las señales de audio que contienen muestras digitales de audio de los participantes no focales y almacenar las señales en memorias intermedias
610
Capturar la señal del micrófono del participante de punto focal y almacenarla en su memoria intermedia correspondiente
615
Calcular los indicios de intensidad de señal
620
Para cada participante, mezclar las señales de audio de todos los demás participantes utilizando una suma ponderada de las señales mezcladas
625
Al final de cada señal mezclada adjuntar los indicios de intensidad de señal de cada señal de audio contenida en esa señal mezclada
630
Codificar las señales de audio mezclado
635
Enviar cada señal de audio mezclado al participante correspondiente
640
Final de la conferencia?
No
Si
Final

Figura 6

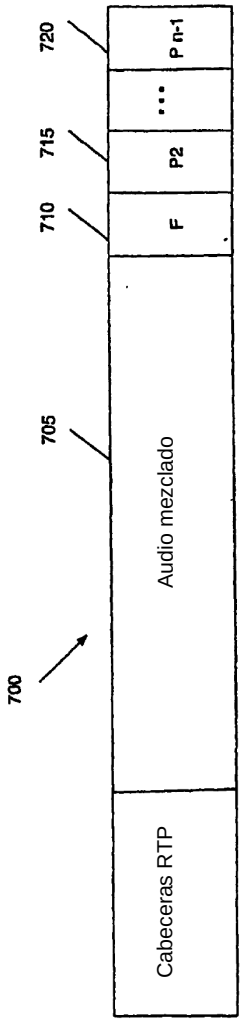

Figura 7

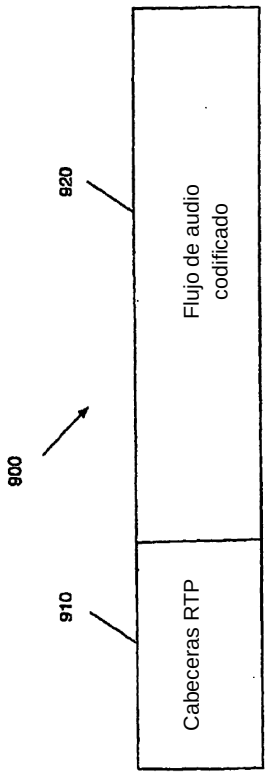

Figura 9

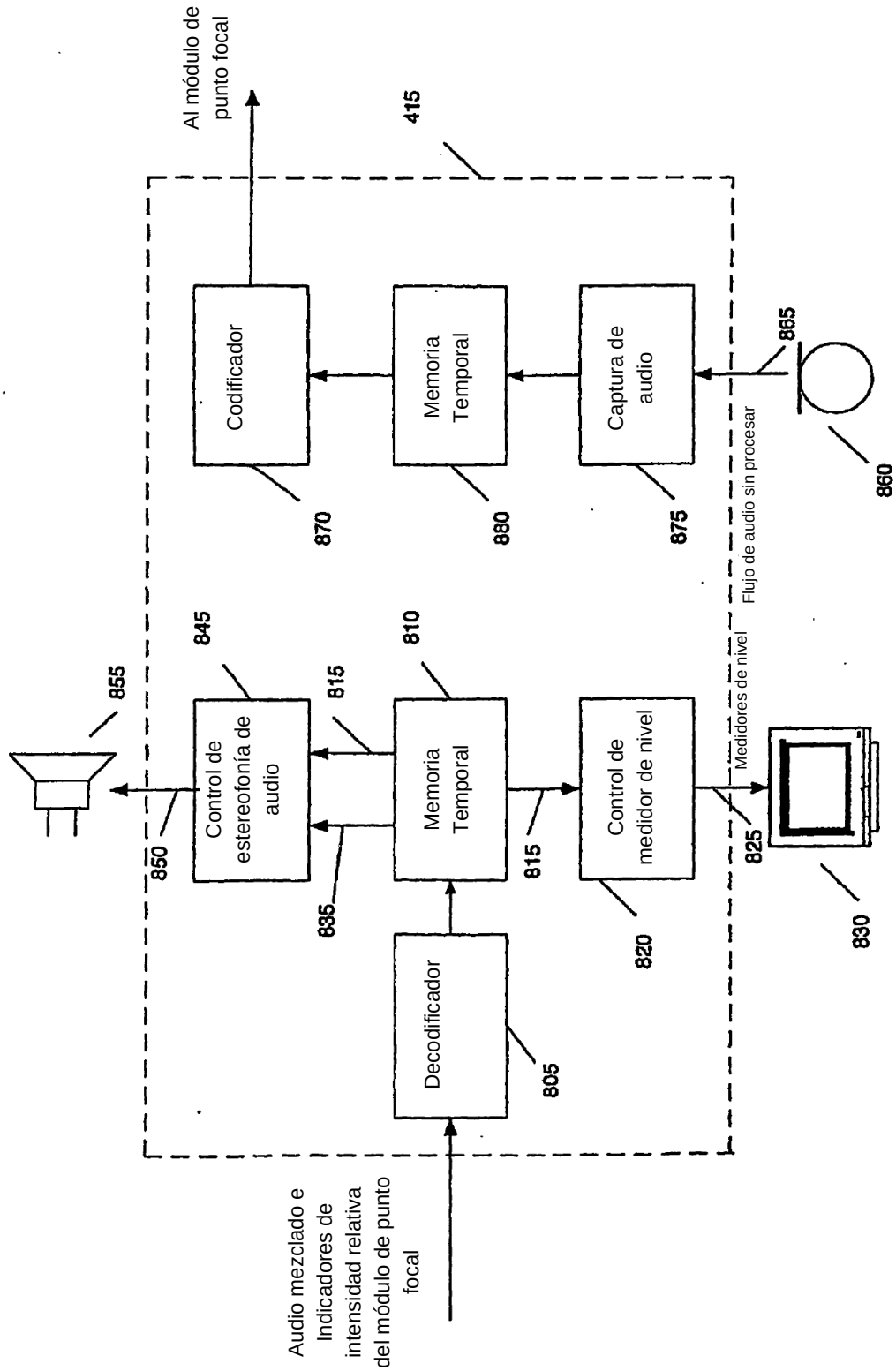
Al módulo de punto focal
415
Codificador
Memoria Temporal
Captura de audio
865
860
870
880
875
Flujo de audio sin procesar
845
855
810
815
Control de estereofonía de audio
Memoria Temporal
Control de medidor de nivel
Medidores de nivel
850
815
825
835
820
830
Decodificador
805
Audio mezclado e Indicadores de intensidad relativa del módulo de punto focal

Figura 8

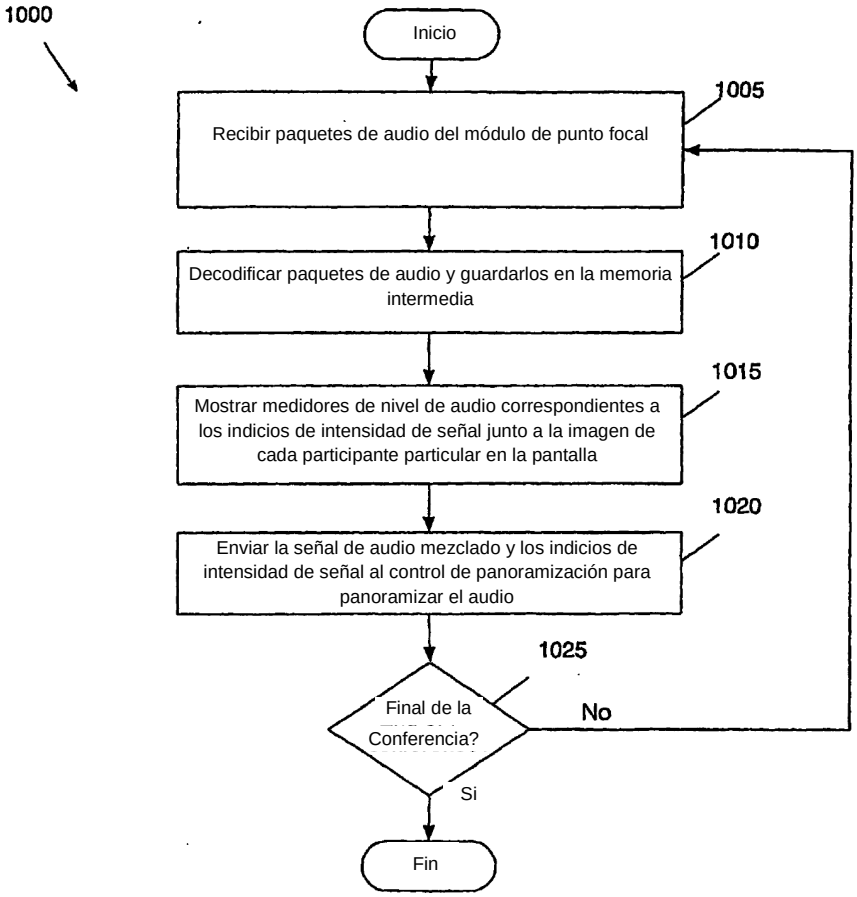
1000
Inicio
1005
Recibir paquetes de audio del módulo de punto focal
1010
Decodificar paquetes de audio y guardarlos en la memoria intermedia
1015
Mostrar medidores de nivel de audio correspondientes a los indicios de intensidad de señal junto a la imagen de cada participante particular en la pantalla
1020
Enviar la señal de audio mezclado y los indicios de intensidad de señal al control de panoramización para panoramizar el audio
1025
Final de la Conferencia?
No
Si
Fin

Figura 10

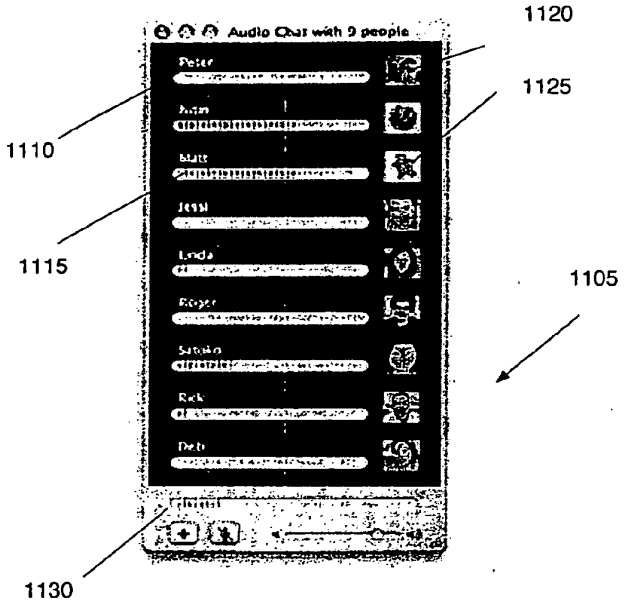

Figura 11

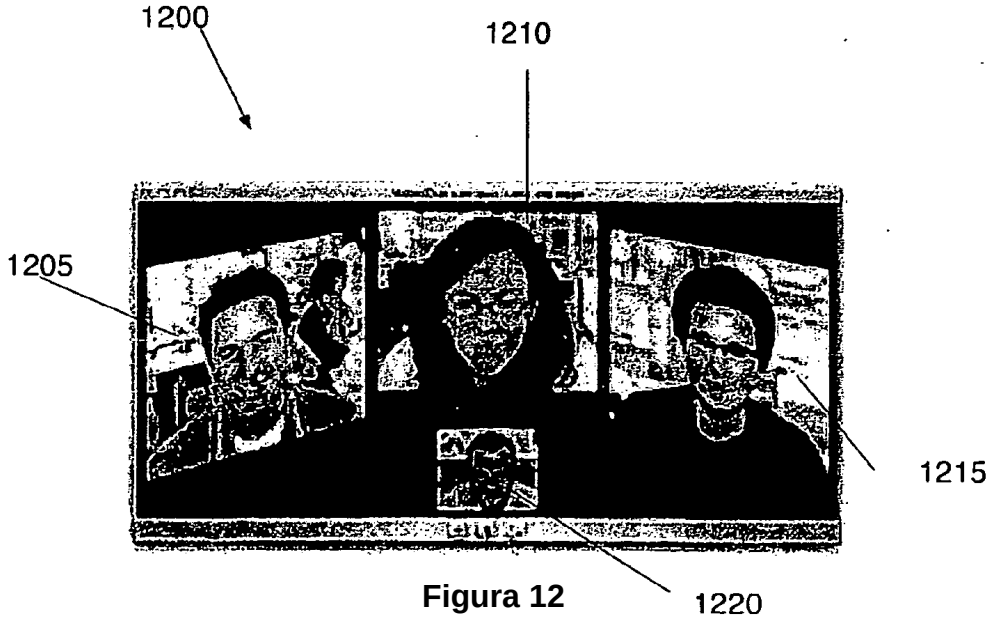

Figura 12

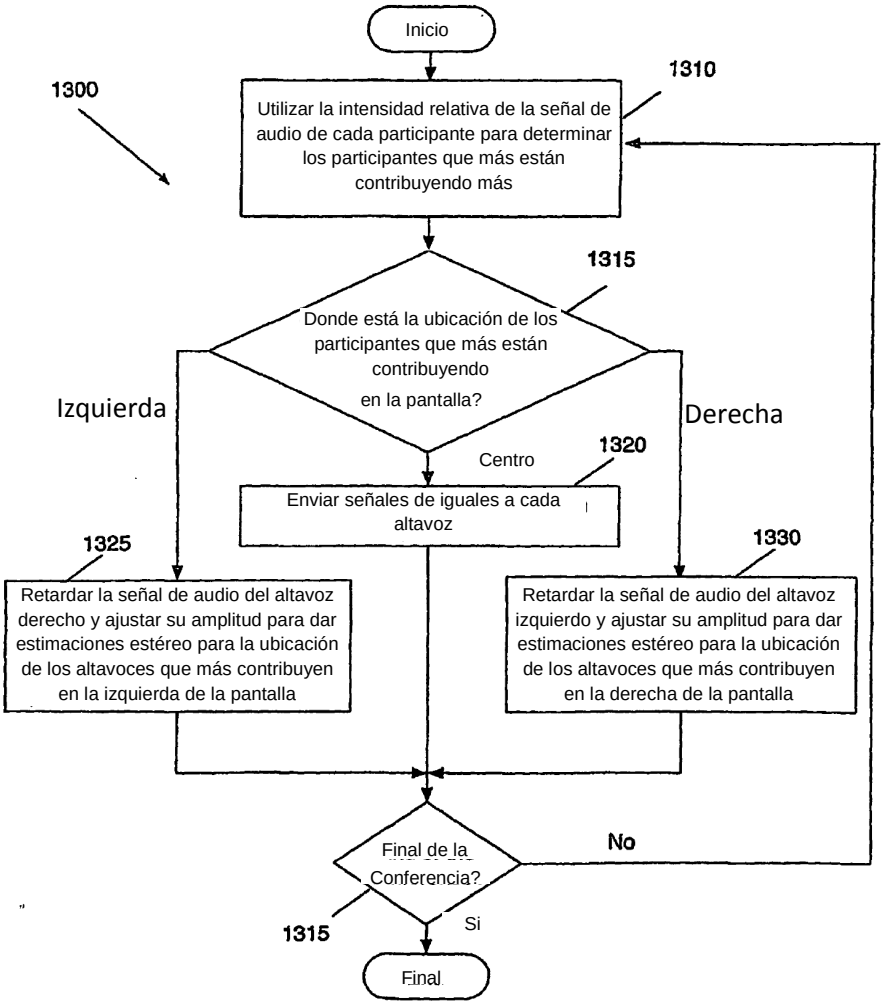
Inicio
1300
1310
Utilizar la intensidad relativa de la señal de audio de cada participante para determinar los participantes que más están contribuyendo más
1315
Donde está la ubicación de los participantes que más están contribuyendo en la pantalla?
Izquierda
Derecha
Centro
1320
Enviar señales de iguales a cada altavoz
1325
Retardar la señal de audio del altavoz derecho y ajustar su amplitud para dar estimaciones estéreo para la ubicación de los altavoces que más contribuyen en la izquierda de la pantalla
1330
Retardar la señal de audio del altavoz izquierdo y ajustar su amplitud para dar estimaciones estéreo para la ubicación de los altavoces que más contribuyen en la derecha de la pantalla
Final de la Conferencia?
No
1315
Si
Final

Figura 13