

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 386 977**

51 Int. Cl.:
G06Q 10/00 (2012.01)
G06Q 30/00 (2012.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 96 Número de solicitud europea: **06838488 .2**
96 Fecha de presentación: **27.11.2006**
97 Número de publicación de la solicitud: **1955458**
97 Fecha de publicación de la solicitud: **13.08.2008**

54 Título: **Aplicaciones sociales e interactivas para medios masivos**

30 Prioridad:
29.11.2005 US 740760 P
29.08.2006 US 823881 P

45 Fecha de publicación de la mención BOPI:
10.09.2012

45 Fecha de la publicación del folleto de la patente:
10.09.2012

73 Titular/es:
GOOGLE INC.
1600 AMPHITHEATRE PARKWAY
MOUNTAIN VIEW, CA 94043, US

72 Inventor/es:
FINK, Michael;
BALUJA, Shumeet y
COVELL, Michele

74 Agente/Representante:
Carvajal y Urquijo, Isabel

ES 2 386 977 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).

DESCRIPCIÓN

Aplicaciones sociales e interactivas para medios masivos.

Campo técnico

Las ejecuciones descritas se relacionan con aplicaciones sociales e interactivas para medios masivos.

5 Antecedentes

10 Los canales de medios masivos (por ejemplo las difusiones de televisión y radio) típicamente suministran contenido limitado a una gran audiencia. Como contraste, la "World Wide Web" suministra vastas cantidades de información que pueden ser solo de interés a pocos individuos. Los intentos de la televisión interactiva convencional crean puentes entre estos dos medios de comunicación al suministrar un medio para que los televidentes interactúen con su televisión y al recibir contenido y/o servicios relacionados con la difusión de televisión

15 La televisión interactiva convencional está típicamente solo disponible a los televidentes a través de redes de cable o satélite para o mediante tarifas de suscripción. Para recibir el servicio de televisión interactiva el televidente tiene que arrendar o comprar un decodificador y que lo instale un técnico. La televisión del observador está conectada al decodificador, lo que le posibilita al observador interactuar con la televisión utilizando el control remoto y otros dispositivos de entrada, y recibir información, entretenimiento y servicios (por ejemplo propaganda, compras en línea, formas y encuestas, juegos y actividades, etc.).

Aunque la televisión interactiva convencional puede mejorar la experiencia de televisión del observador, subsiste la necesidad de aplicaciones sociales e interactivas para medios masivos que no tengan hardware significativo adicional o conexiones físicas entre la televisión o la radio y un decodificador o un ordenador.

20 Una aplicación de televisión social e interactiva que está faltando en los sistemas de televisión convencional e interactiva es la capacidad de suministrar información complementaria al canal de medio masivo sin mayor esfuerzo. Con los sistemas convencionales, un usuario tendría que entrar a un ordenador y consultar tal información lo que disminuiría la experiencia pasiva ofrecida por los medios masivos. Más aún, los sistemas de televisión convencionales no pueden suministrar información complementaria en tiempo real mientras que el usuario está viendo una difusión.

25 Otra aplicación de la televisión social e interactiva que está faltando con los sistemas de televisión interactivos convencionales es la capacidad de enlazar dinámicamente un observador con una comunidad social par nd hoc (por ejemplo un grupo de discusión, un sitio de conversación, etc.) en tiempo real. Imagine que usted está viendo el último episodio de "Friends" en televisión y descubre que el personaje "Monica" está embarazada. Usted quiere conversar, comentar o leer las respuestas de otros observadores a la escena en tiempo real. Una opción sería ingresar a su ordenador, digitar el nombre de "Friends" u otros términos relacionados en un motor de búsqueda, y efectuar una búsqueda para encontrar un grupo de discusión de "Friends." Tal acción requerida por el observador, sin embargo, disminuiría la experiencia pasiva ofrecida por los medios masivos y no le posibilitaría al observador interactuar dinámicamente (por ejemplo comentar, conversar, etc.) con otros observadores que están viendo el programa al mismo tiempo.

35 Otra deficiencia en los sistemas de televisión convencional y los sistemas de televisión interactiva es un método simple para evaluar la popularidad de los eventos de difusión. Las calificaciones de popularidad de los eventos de difusión son de mucho interés para los usuarios, difusores y anunciantes. Estas necesidades se resuelven parcialmente mediante sistemas de medición como las calificaciones de Nielsen®. Estas calificaciones, sin embargo, requieren la instalación de un equipo dedicado y de la cooperación de los observadores participantes.

40 El documento US 2004/0031058 A1 describe sistemas y métodos para navegar hipermedios utilizando múltiples conjuntos de dispositivos de entrada/salida coordinados. Los sistemas y métodos descritos le permiten al usuario y/o a un autor controlar que fuentes se presentan en que dispositivos (sea que estén integrados o no), y suministrar la coordinación de actividades de navegación para posibilitar que tal interfaz se emplee a través de múltiples sistemas independientes.

45 Yan Ke et al., "Computer Vision for Music Identification", Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, páginas 597-604, ISBN: 978/0/7695/2372-9, describe cómo ciertas tareas en el dominio de audio se pueden dirigir efectivamente utilizando aproximaciones de visión con ordenador. Este documento se enfoca en el problema de identificación de la música, donde la meta es identificar de manera confiable una canción dando unos pocos segundos de audio ruidoso. La aproximación trata el espectrograma de cada video musical como una imagen 2D y transforma la identificación de la música en un problema de recuperación de subimagen dañada.

5 El documento US 2004/0199387 A1 describe un método y un sistema que le permite a los usuarios identificar (pregrabar) sonidos tales como música, difusión de radio, comerciales, y otras señales de audio en casi cualquier ambiente. La señal de audio (o sonido) debe ser una grabación representada en una base de datos de grabaciones. El servicio puede identificar rápidamente la señal con solo unos pocos segundos de ejecución, mientras que se tolera alto ruido y distorsión. Una vez que el usuario identifica la señal, el usuario puede efectuar transacciones interactivamente en tiempo real o fuera de línea utilizando la información de identificación.

El documento US 2005/0147256 A1 describe un aparato que incluye un analizador acústico para identificar el audio ambiente recibido y un analizador de contenido. El analizador de contenido es para seleccionar el contenido asociado con el audio identificado para la presentación del contenido a un usuario. Se describen y reivindican otras realizaciones.

10 El documento US 2002/0133499 A1 describe un método para identificar rápida y precisamente un archivo digital, específicamente uno que represente un archivo de audio. La identificación se puede utilizar para hacer un seguimiento a los pagos de regalías a los propietarios de derechos de autor. Una base de datos almacena características de varios archivos de audio y un identificador globalmente único (GUID) para cada archivo. De manera ventajosa, el método permite que sea actualizada una base de datos en el caso de un nuevo archivo de audio al almacenar sus características y generar un nuevo identificador único para el nuevo archivo. El archivo de audio es muestreado para generar una huella que utiliza los residuos del espectro y transforma el tren de ondas de Haar. De manera ventajosa cualquier marca utilizada para el trabajo es actualizada automáticamente si parece haber un error.

Resumen

20 Las deficiencias descritas anteriormente son manejadas por el sistema reivindicado, el método y el producto de programa de ordenador.

Descripción de los dibujos

La FIGURA 1 es un diagrama de bloque de una realización de sistema de personalización masivo.

La FIGURA 2 ilustra una realización de un sistema de identificación de audio ambiente, que incluye la interfaz del lado del cliente mostrada en la FIGURA .1.

25 La FIGURA 3 es un diagrama de flujo de un ejemplo de un proceso para proporcionar aplicaciones de personalización masiva.

La FIGURA 4 es un diagrama de flujo de un ejemplo de un proceso de establecimiento de huella de audio.

La FIGURA 5 es un diagrama de flujo de una realización de una interfaz de usuario para interactuar con aplicaciones de personalización masivas.

30 La FIGURA 6 es un diagrama de bloque de una realización de la arquitectura de hardware para un sistema de cliente para ejecutar la interfaz del lado del cliente mostrada en la FIGURA .1.

La FIGURA 7 es un diagrama de flujo de una realización de un proceso de detección de repetición.

Descripción detallada

Aplicaciones de personalización masivas

35 Las aplicaciones de personalización masivas suministran información personalizada e interactiva relacionada con las difusiones de medios masivos (por ejemplo televisión, radio, cine, difusiones de Internet, etc.). Tales aplicaciones incluyen pero no se limitan a: capas de información personalizada, comunidades sociales pares ad hoc, calificaciones de popularidad en tiempo real y marcadores de video (o audio), etc. Aunque algunos de los ejemplos de medios masivos descritos aquí están en el contexto de difusiones de televisión, las ejecuciones descritas son igualmente aplicables a difusiones de radio y/o música.

40 Las capas de información personalizada suministran información complementaria al canal de medios masivos. Ejemplos de capas de información personalizada incluyen pero no se limitan a: moda, política, negocios, salud, viajes, etc. Por ejemplo, mientras se ve un nuevo segmento de una celebridad, se presenta una capa de moda al observador en una pantalla de televisión o en un dispositivo de pantalla de un ordenador, que suministre información y/o imágenes relacionadas con las prendas y accesorios que está vistiendo la celebridad en el segmento de noticias. Adicionalmente, las capas personalizadas pueden incluir propaganda que promueva productos o servicios con el segmento de noticias, tal como un enlace a un almacén de ropa que está vendiendo prendas que la celebridad está vistiendo.

45 Las comunidades sociales pares ad hoc suministran un sitio para comentarios entre usuarios que estén viendo la misma presentación de televisión o escuchando la misma estación de radio. Por ejemplo, un usuario que esté viendo los últimos titulares de CNN se le puede dar un medio con comentario (por ejemplo un sitio de conversación, un tablero de mensajes, una página wiki, un enlace de video, etc.) que le permita al usuario conversar, comentar o leer las respuestas de otros observadores a la difusión de medios masivos en desarrollo.

Las calificaciones de popularidad en tiempo real le suministra a los proveedores de contenido y a los usuarios una información de calificaciones (similar a las calificaciones de Nielsen). Por ejemplo, a un usuario se le puede suministrar de manera instantánea calificaciones de popularidad en tiempo real de los canales de televisión o de las estaciones de radio que están viendo o escuchando o de la red social del usuario y/o la gente con similar demografía.

5 Las marcas de video o audio les da a los usuarios maneras con poco esfuerzo de crear bibliotecas personalizadas de su contenido de difusión favorito. Por ejemplo, un usuario puede simplemente presionar un botón sobre un ordenador o un dispositivo de control remoto y un fragmento de código de un audio ambiente y/o video de un contenido de difusión que está siendo grabado, procesado y guardado. El fragmento de código se puede utilizar como una marca para referirse al programa, o a porciones del programa, para observación posterior. La marca se puede compartir con amigos o guardarse para futura referencia personal.

Redes de Personalización Masivas

15 La FIGURA 1 es un diagrama de bloque de sistema de personalización masivo 100 para proporcionar aplicaciones de personalización masivas. El sistema 100 incluye una o más interfaces del lado del cliente 102, un servidor de base de datos de audio 104 y un servidor para una aplicación social 106, todas las cuales se comunican sobre una red 108 (por ejemplo, la Internet, una intranet, LAN, red inalámbrica, etc.).

20 Una interfaz de cliente 102 puede ser cualquier dispositivo que le permita a un usuario que le permita a un usuario ingresar y recibir información, y que es capaz de presentar una interfaz de usuario en un dispositivo de pantalla, que incluye pero no se limita a: un ordenador de mesa o portátil; un dispositivo electrónico; un teléfono, un teléfono móvil; un sistema de pantalla, una televisión, un monitor de ordenador; un sistema de navegación; un reproductor/grabador de medios portátil, un ayudante digital personal (PDA); una consola de juegos; un dispositivo electrónico manual; y un dispositivo o aparato electrónico embebido. La interfaz de cliente 102 se describe más completamente con respecto de la FIGURA 2.

25 La interfaz de cliente 102 incluye un detector de audio ambiente (por ejemplo, un micrófono) para supervisar y registrar el audio ambiente de una difusión de medios masivos en un ambiente de difusión (por ejemplo, una sala del usuario). Uno o más segmentos de audio ambiente o "fragmentos de códigos" se convierten en resúmenes estadísticos distintivos y robustos, denominados como "huellas de audio" o "descriptores." En algunos ejemplos, los descriptores son archivos comprimidos que contienen unos componentes de firma audio que se pueden comparar con una base de datos de los descriptores de referencia previamente generados o estadísticas asociadas con la difusión de medios masivos.

30 Una técnica para generar huellas de audio para identificación de música se describe en Ke, Y., Hoiem, D., Sukthankar, R. (2005), Computer Vision for Music Identification, In Proc. Computer Vision and Pattern Recognition. En algunas ejecuciones, el método de identificación de música propuesto por (en lo sucesivo "Ke et al.") se adapta para generar descriptores para datos de audio para televisión y consultas, como se describió con respecto a la FIGURA 4.

35 Una técnica para generar descriptores de audio utilizando trenes de ondas se describe en la Solicitud de Patente Provisional Estadounidense No. 60/823,881, para "Audio Identification Based on Signatures." Esta solicitud describe una técnica que utiliza una combinación de visión por ordenador y unos algoritmos de procesamiento de corriente de datos a gran escala para crear descriptores/huellas compactas de fragmentos de código de audio y se pueda hacer corresponder eficientemente. La técnica utiliza trenes de ondas, que es una herramienta matemática conocida para funciones de jerárquicamente descompuestas.

40 En "Identificación de Audio con Base en Firmas," una implementación de un proceso de recuperación incluye las siguientes etapas: 1) dado el espectro de audio de un fragmento de código de audio, las imágenes espectrales de extracto de, por ejemplo, 11.6*w ms de duración, con un promedio de espaciamiento aleatorio separado d-ms. Para cada imagen espectral: 2) computar los trenes de onda de la imagen espectral; 3) extraer los trenes de onda superiores t; 4) crear una representación binaria de los trenes de onda superiores t; 5) utilizar "min-hash" para crear una subhuella de los trenes de onda superiores t; 6) utilizar LSH con compartimentos b y tablas "hash" para encontrar segmentos de subhuella que se hagan corresponder cercanamente; 7) descargar subhuellas con menos de "v" correspondencias; 8) computar una distancia Hamming desde las subhuellas candidatas restantes a la subhuella de consulta; y 9) utilizar programación dinámica para combinar las correspondencias durante el tiempo.

50 En algunos ejemplos, los descriptores y un identificador de usuario asociado ("usuario id") para identificar la interfaz del lado del cliente 102 se envían al servidor de la base de datos de audio 104 por medio de la red 108. El servidor de la base de datos de audio 104 compara el descriptor con una pluralidad de descriptores de referencia, que se determinaron y almacenaron previamente en una base de datos de audio 110 acoplada al servidor de la base de datos de audio 104. En algunos ejemplos, el servidor de la base de datos de audio 104 actualiza continuamente los descriptores de referencia almacenados en la base de datos de audio 110 desde las difusiones de medios masivos recientes.

55 El servidor de la base de datos de audio 104 determina las mejores correspondencias entre *los descriptores recibidos* y los descriptores de referencia y envía la información de la mejor correspondencia al servidor de aplicación social 106. El proceso de correspondencia se describe más completo con respecto a la FIGURA 4.

5 En algunas ejecuciones, el servidor de aplicación social 106 acepta conexiones de navegador de la red asociados con la interfaz del lado del cliente 102. Utilizando la información de mejor correspondencia, el servidor de aplicación social 106 agrega información personalizada para el usuario y envía la información personalizada a la interfaz del lado del cliente 102. La información personalizada puede incluir pero no se limita a: propagandas, capas de información personalizada, calificaciones de popularidad, e información asociada con un medio en comentario (por ejemplo, comunidades sociales pares ad hoc, foros, grupos de discusión, conferencias de video, etc.).

10 La información personalizada se utiliza para crear una sala de conversación para los observadores sin conocer el espectáculo que los observadores están viendo en tiempo real. Las salas de conversación se crean al comparar directamente los descriptores en las corrientes de datos transmitidas por los sistemas de cliente para determinar las correspondencias. Es decir, las salas de conversación se pueden crear alrededor de observadores que tienen los descriptores que se corresponden. No existe necesidad de comparar los descriptores recibidos de los observadores contra los descriptores de referencia.

15 En algunas ejecuciones, el servidor de aplicación social 106 sirve como una página de red a la interfaz del lado del cliente 102, que se recibe y despliega mediante un navegador de red (por ejemplo, Microsoft Internet Explorer™) que corre en la interfaz del lado del cliente 102. El servidor de aplicación social 106 también recibe la id del usuario desde la interfaz del lado del cliente 102 y/o el servidor de la base de datos de audio 104 para ayudar a agregar contenido personalizado y servir las páginas de red a la interfaz del lado del cliente 102.

20 Debe ser evidente que son posibles otras implementaciones del sistema 100. Por ejemplo, el sistema 100 puede incluir múltiples base de datos de audio 110, servidores de la base de datos de audio 104 y/o servidores de aplicación social 106. Alternativamente, el servidor de la base de datos de audio 104 y el servidor de aplicación social 106 puede ser un servidor o sistema único, o parte de un recurso de red y/o servicio. También, la red 108 puede incluir múltiples redes y enlaces acoplados operativamente en varias topologías y disposiciones que utilizan una variedad de dispositivos de red (por ejemplo, concentradores, enrutadores, etc.) y medios (por ejemplo, cobre, fibra óptica, radiofrecuencias, etc.). Las arquitecturas del servidor del cliente se describen solamente como un ejemplo. Son posibles otras arquitecturas de ordenador.

Sistema de Identificación de Audio Ambiente

30 La FIGURA 2 ilustra un sistema de identificación de audio ambiente 200, que incluye una interfaz del lado del cliente 102 como se muestra en la FIGURA 1. El sistema 200 incluye un sistema de medios masivos 202 (por ejemplo, un aparato de televisión, radio, ordenador, dispositivo electrónico, teléfono móvil, consola de juegos, dispositivos de red, etc.), un detector de audio ambiente 204, una interfaz del lado del cliente 102 (por ejemplo, un ordenador de mesa o portátil, etc.) y un dispositivo de acceso a la red 206. En algunas ejecuciones, la interfaz del lado del cliente 102 incluye un dispositivo de pantalla 210 para presentar una interfaz de usuario (UI) 208 para permitir que un usuario interactúe con una aplicación de personalización masivo, como se describe con respecto a la FIGURA 5.

35 En operación, el sistema de medios masivos 202 genera audio ambiente de una difusión de medios masivos (por ejemplo, audio de televisión), que se detecta mediante el detector de audio ambiente 204. El detector de audio ambiente 204 puede ser cualquier dispositivo que puede detectar el audio ambiente, que incluye un micrófono libre de base y un micrófono que se integra con la interfaz del lado del cliente 102. El audio ambiente detectado se codifica por la interfaz del lado del cliente 102 para suministrar descriptores que identifican el audio ambiente. Los descriptores se transmiten al servidor de la base de datos de audio 104 por medio del dispositivo de acceso a la red 206 y la red 108.

40 En algunas ejecuciones, el software del cliente que corre en la interfaz del lado del cliente 102 monitorea continuamente y registra los archivos de audio de n segundos (por ejemplo, 5 segundos) ("fragmentos de códigos") de audio ambiente. Los fragmentos de códigos luego se convierten en m cuadros (por ejemplo, 415 cuadros) de los descriptores codificados de k-bit (por ejemplo, 32-bits), de acuerdo con un proceso descrito con respecto a la FIGURA 4. En algunas ejecuciones, el monitoreo y el registro están basados en el evento. Por ejemplo, el monitoreo y el registro se puede iniciar automáticamente en una fecha específica y en un momento específico (por ejemplo, Lunes, 8:00 P.M.) y durante un tiempo de duración específico (por ejemplo, entre 8:00-9:00 P.M.). Alternativamente, el monitoreo y el registro se pueden iniciar en respuesta al ingreso del usuario (por ejemplo, una pulsación del ratón, clave de función o combinación de clave) desde un dispositivo de control (por ejemplo, un control remoto, etc.). En algunas ejecuciones, el audio ambiente se codifica utilizando una variación de corriente de características discriminantes de cuadro de 32-bit descrita en Ke et al.

En algunas ejecuciones, el software de cliente corre como una "barra lateral" u otro elemento de interfaz de usuario. Es decir, cuando la interfaz del lado del cliente 102 es arrancada, el muestreo del audio ambiente puede iniciar inmediatamente y correr el "trasfondo" con resultados (opcionalmente) que son desplegados en la barra lateral sin invocar una sección de navegador de la referencia.

55 En algunas ejecuciones, el muestreo del audio ambiente puede iniciar cuando la interfaz del lado del cliente 102 es arrancada o cuando se registra en un servicio o aplicación (por ejemplo, correo electrónico, etc.)

Los descriptores se envían al servidor de la base de datos de audio 104. En algunas ejecuciones, los descriptores son resúmenes estadísticos comprimidos del audio ambiente, descritos en Ke et al. Al enviar los resúmenes estadísticos, la privacidad acústica del usuario se mantiene porque los resúmenes estadísticos no son reversibles, es decir, el audio original no se puede recuperar del descriptor. Así, cualquier conversación del usuario u otros individuos monitoreada y grabada en el ambiente de difusión no se puede reproducir del descriptor. En algunas ejecuciones, los descriptores se pueden encriptar para privacidad y seguridad extra utilizando una o más de las técnicas de encriptación conocida (por ejemplo, encriptación con clave asimétrica o simétrica, encriptación elíptica, etc.).

En algunas ejecuciones, los descriptores se envían al servidor de la base de datos de audio 104 como un envío de consulta (también denominado como un descriptor de consulta) en respuesta a un evento disparador detectado por el proceso de monitoreo en la interfaz del lado del cliente 102. Por ejemplo, un evento disparador puede ser el tema de apertura de un programa de televisión (por ejemplo la sintonización de apertura de "Seinfeld") o los diálogos hechos por los actores. En algunas ejecuciones, los descriptores de consulta se pueden enviar al servidor de la base de datos de audio 104 como parte de un proceso en corriente continua. En algunas ejecuciones, los descriptores de consulta se pueden transmitir al servidor de la base de datos de audio 104 en respuesta al ingreso del usuario (por ejemplo, por medio del control remoto, pulsaciones de ratón, etc.).

Proceso de Personalización Masivo

LA FIGURA 3 es un diagrama de flujo de un proceso de personalización masivo 300. Las etapas del proceso 300 no tienen que ser completadas en un orden particular y al menos algunas etapas se pueden efectuar al mismo tiempo en un ambiente multihilado o de procesamiento paralelo.

El proceso 300 comienza cuando la interfaz del lado del cliente (por ejemplo, interfaz del lado del cliente 102) monitorea y registra los fragmentos de códigos de audio ambiente de una difusión de medios masivos en un ambiente de difusión (302). Los fragmentos de códigos del audio ambiente grabado se codifican en los descriptores (por ejemplo, resúmenes estadísticos comprimidos), que se pueden enviar a un servidor de base de datos de audio (304) como consultas. El servidor de la base de datos de audio compara las consultas contra una base de datos de los descriptores de referencia calculada de las estadísticas de difusión de medios masivos para determinar los descriptores candidatos que mejor se correspondan con la consulta (308). Los descriptores candidatos se envían a un servidor para una aplicación social o a otro recurso de red, que utiliza los descriptores candidatos para agregar información personalizada para el usuario (310). Por ejemplo, si el usuario está viendo el programa de televisión "Seinfeld," los descriptores de consulta generados del audio ambiente del programa se corresponderán con los descriptores de referencia derivados de las difusiones de "Seinfeld" previas. Así, los descriptores candidatos que se correspondan mejor se utilizan para agregar información personalizada que se relaciona con "Seinfeld" (por ejemplo, nuevas historias, grupos de discusión, enlaces a comunidades sociales pares ad hoc o salas de conversación, propagandas, etc.). En algunos ejemplos el procedimiento de correspondencia se efectúa eficientemente utilizando técnicas "hashing" (por ejemplo "hashing" directo o "hashing" sensible al sitio (LSH)) para lograr una lista corta de descriptores candidatos, como se describió con respecto a la FIGURA 4. Los descriptores candidatos luego se procesan en un procedimiento de validación, tal como se describió en Ke et al.

Los descriptores de consulta de los diferentes observadores se corresponden directamente en lugar de hacer corresponder cada consulta con una base de datos de los descriptores de referencia. Esto permite la creación de comunidades sociales pares ad hoc sobre la materia objeto para lo cual no está disponible una base de datos de los descriptores de referencia. Tal realización se puede hacer coincidir con los observadores en tiempo real que estén en la misma forma pública (por ejemplo, estadio, bar, etc.) utilizando dispositivos electrónicos portátiles (por ejemplo, teléfonos móviles, PDA, etc.).

Calificaciones de Popularidad

En algunas ejecuciones, las estadísticas en tiempo real y agregadas se infieren de una lista de observadores que están viendo en el momento la difusión (por ejemplo, programa, publicidad, etc.). Estas estadísticas se pueden recoger en el transcurso mientras que los observadores están utilizando otras aplicaciones. Las estadísticas pueden incluir pero no se limitan a: 1) el número promedio de observadores que ven la difusión; 2) el número promedio de veces que los observadores ven la difusión; 3) otros programas que ven los observadores; 4) el número mínimo y pico de observadores; 5) lo que más a menudo los observadores cambiaron cuando ellos dejaron una difusión; 6) que tanto ven una difusión los observadores; 7) cuantas veces los observadores cambian de canal; 8) que publicidad es vista por los observadores; y 9) lo que los observadores más a menudo cambian cuando ingresan a una difusión, etc. De estas estadísticas, se puede determinar una o más calificaciones de popularidad.

Las estadísticas utilizadas para generar las calificaciones de popularidad se pueden generar utilizando un contador para cada canal de difusión que está siendo monitoreado. En algunas ejecuciones, los contadores se pueden intersectar con datos de un grupo demográfico o con datos de un grupo geográfico. Las calificaciones de popularidad se pueden utilizar por los observadores para "ver lo que es más visto" mientras que la difusión está ocurriendo (por ejemplo, al notar una calificación creciente durante las presentaciones de mitad de tiempo del "Super Bowl" del 2004). Los anunciantes y proveedores de contenidos también pueden utilizar calificaciones de popularidad para ajustar dinámicamente el material mostrado en respuesta a las calificaciones. Esto es especialmente cierto para la publicidad, en razón a que la longitud

unitaria corta y numerosas versiones de la publicidad generada de las campañas publicitarias son fácilmente intercambiadas para ajustar a los niveles de calificación del observador. Otros ejemplos de las estadísticas incluyen pero no se limitan a: popularidad de una difusión de televisión versus la difusión de radio por demografía o tiempo, la popularidad de los momentos del día, es decir los momentos pico de observación/escucha, el número de casas en un área dada, la cantidad de cambio de canales durante un programa particular (género de programas, momentos particulares del día), volumen de la difusión, etc.

La información personalizada se envía a la interfaz del lado del cliente (312). Las calificaciones de popularidad también se pueden almacenar en una base de datos para uso de otros procesos (318), tal como el ajuste dinámico de la publicidad descrita anteriormente. La información personalizada se recibe en la interfaz del lado del cliente (314) en donde se formatea y se presenta en una interfaz de usuario (316). La información personalizada se puede asociar con un medio de comentario (por ejemplo, mensajes de texto en una sala de conversación) que se presenta al usuario en una interfaz de usuario. En algunas ejecuciones, una sala de conversación puede incluir uno o más subgrupos. Por ejemplo, un grupo de discusión para "Seinfeld" podría incluir un subgrupo denominado "Expertos en Seinfeld," o un subgrupo se puede asociar con una demografía particular, tal como las mujeres entre edades de 20-30 que ven "Seinfeld," etc.

En algunas ejecuciones, la información bruta (por ejemplo, valores de contador) utilizada para generar las estadísticas de las calificaciones de popularidad se recolecta y almacena en la interfaz del lado del cliente en lugar del servidor de aplicación social. La información bruta se puede transferir al difusor cuando quiera que el usuario esté en línea y/o invoque una aplicación de personalización masiva.

En algunas ejecuciones, se instala una caja de medición de difusión (BMB) en la interfaz del lado del cliente. El BMB puede ser un dispositivo de hardware simple que es similar a un decodificador pero que no se conecta al dispositivo de difusión. A diferencia del sistema de calificación de Nielsen, que requiere que el hardware sea instalado en la televisión, el BMB se puede instalar cerca al sistema de medios masivos o dentro de un rango de señal de televisión. En algunas ejecuciones, el BMB graba automáticamente los fragmentos de códigos de audio y genera descriptores, que se almacenan en la memoria (por ejemplo, medio flash m). En algunas ejecuciones, el BMB puede incluir opcionalmente uno o más botones de hardware que puede presionar un usuario para indicar que difusión está viendo (similar a las calificaciones Nielsen®). El dispositivo BMB puede ser recogido por el proveedor de calificaciones de vez en cuando para recolectar los descriptores almacenados, o el BMB puede difundir los descriptores almacenados a una o más partes interesadas en una conexión de red (por ejemplo, teléfono, Internet, radio inalámbrico, tal como radio SMS/portadores, etc.) de vez en cuando.

En algunas ejecuciones, la publicidad se puede monitorear para determinar los efectos de los anuncios, que se pueden reportar a los anunciantes. Por ejemplo, que anuncios fueron vistos, no tenidos en cuenta, el nivel de volumen de los anuncios, etc.

En algunas ejecuciones, un dispositivo de captura de imágenes (por ejemplo, una cámara digital, una grabadora de video, etc.) se puede utilizar para medir cuantos observadores están viendo o escuchando una difusión. Por ejemplo, varios algoritmos con correspondencia de patrón conocidos se pueden aplicar a una imagen a una secuencia de imágenes para determinar el número de observadores presentes en un ambiente de difusión durante una difusión particular. Las imágenes y los datos derivados de las imágenes se pueden utilizar en combinación con los descriptores de audio para recoger la información personalizada para un usuario, calcular las calificaciones de popularidad, o para cualquier otro propósito.

Proceso de Estampado de Huella de Audio

La FIGURA 4 es un diagrama de flujo del proceso de estampado de huella de audio 400. Las etapas del proceso 400 no tienen que ser completadas en ningún orden particular y al menos algunas etapas se pueden efectuar al mismo tiempo en un ambiente de procesamiento multihilado o paralelo. El proceso 400 se corresponde con los descriptores de consulta generados por la interfaz del lado del cliente (por ejemplo, interfaz del lado del cliente 102) a los descriptores de referencia almacenados en una o más bases de datos en tiempo real y con baja latencia. El proceso 400 adapta una técnica propuesta por Ke et al. para manejar los datos de audio ambiente (por ejemplo, de una difusión de televisión) y consultas.

El proceso 400 inicia en una interfaz del lado del cliente al descomponer fragmentos de códigos del audio ambiente (por ejemplo, 5-6 segundos de audio) de una difusión de medios masivos capturada por un detector de audio ambiente (por ejemplo, micrófono) en cuadro que se traslapa (402). En algunos ejemplos, los cuadros se separan mediante varios milisegundos (por ejemplo, 12 ms aparte). Cada cuadro se convierte en un descriptor (por ejemplo, un descriptor de 32-bit) que se entrena para superar el ruido de audio y la distorsión (404), como se describe en Ke et al. En algunos ejemplos, cada descriptor representa un resumen estadístico de identificación del fragmento de código de audio.

En algunos ejemplos, los descriptores se pueden enviar como fragmentos de códigos de consulta (también denominado como descriptores de consulta) a un servidor de base de datos de audio donde coinciden con una base de datos de los descriptores de referencia que identifican resúmenes estadísticos de los fragmentos de códigos de audio registrados previamente de la difusión de medios masivos (406). Se puede determinar una lista de los descriptores candidatos que

5 tengan las mejoras correspondencias (408). Los descriptores candidatos se pueden calificar, de tal manera que los descriptores candidatos que son temporalmente consistentes con el descriptor de consulta tienen mayor calificación que los descriptores candidatos que son temporalmente menos consistentes con el descriptor de consulta (410). Los descriptores candidatos con la más alta calificación (por ejemplo, una calificación que excede un valor de un umbral suficientemente alto) se transmiten o de otra manera se suministran a un servidor para una aplicación social (412) donde ellos se pueden utilizar para agregar información personalizada relacionada con la difusión de medios. Utilizando un umbral se asegura que los descriptores tengan una suficiente correspondencia antes de que los descriptores sean transmitidos o suministrados de otra manera al servidor de aplicación social (412).

10 En algunos ejemplos, la base de datos de los descriptores de referencia se puede generar de las difusiones dadas por varias compañías de medios, que se pueden indizar y utilizar para generar los descriptores. En otros ejemplos, los descriptores de referencia también se pueden generar utilizando guías de televisión u otros metadatos y/o información embebida en la señal de difusión.

15 En algunas ejecuciones, la tecnología de reconocimiento de habla se puede utilizar para ayudar a identificar que programa se está viendo. Dicha tecnología podría ayudar a los usuarios a discutir nuevos eventos en lugar de solo programas de televisión. Por ejemplo, un usuario podría ver el lanzamiento de un trasbordador en una canal diferente de otro observador y, por lo tanto, posiblemente conseguir una señal de audio diferente (por ejemplo debido a un presentador diferente). La tecnología de reconocimiento de voz se podría utilizar para reconocer palabras clave (por ejemplo, Trasmordador, lanzamiento, etc.), que se puede utilizar para enlazar el usuario con un medio que lo comente.

Descriptores de "Hashing"

20 Ke et al. utiliza técnicas de visión de ordenador para encontrar estadísticas concretas altamente discriminadas de audio. Su procedimiento entrenado en pares marcados de ejemplos positivos (en donde x y x' son versiones ruidosas del mismo audio) y ejemplos negativos (en donde x y x' son de diferente audio). Durante esta fase de entrenamiento, la técnica de aprendizaje de máquina basada en amplificación utiliza los pares marcados para seleccionar una combinación de 32 filtros y umbrales que crean conjuntamente una estadística altamente discriminada. Los filtros localizan los cambios en la magnitud del estograma, utilizando primeras y segundas diferencias de orden en el tiempo y frecuencia. Un beneficio de utilizar estos filtros de diferencia simple es que se pueden calcular eficientemente utilizando una técnica de imagen integral descrita en Viola, P. and Jones, M. (2002), Robust Real-Time Object Detection, International Journal of Computer Vision.

30 En algunas ejecuciones, las salidas de estos 32 filtros son umbrales, dado un bit único por filtro en cada cuadro de audio. Estos 32 umbrales resultan de solamente descriptores transmitidos de aquel cuadro de audio. Este sobrante de codificación asegura la privacidad del usuario a escucha no autorizada. Además, estos descriptores de 32 bits son robustos a las distorsiones de audio en los datos de entrenamiento, de tal manera que ejemplos positivos (por ejemplo cuadros de correspondencia) tienen pequeñas distancias Hamming (es decir la distancia que mide diferentes números de bits) y ejemplos negativos (por ejemplo cuadros que no se corresponden) tienen grandes distancias Hamming. Cabe notar se pueden utilizar más o menos filtros y más de un bit por filtro se puede utilizar en cada cuadro de audio (por ejemplo más bits que utilizan múltiples pruebas de umbral).

35 En algunas ejecuciones, el descriptor de 32 bits propiamente dicho se utiliza como una clave "hash" para un "hashing" directo. El descriptor es una función "hash" bien balanceada. Las tasas de recuperación se mejoran además al solicitar no solamente el destructor de consulta sino también un pequeño conjunto de destructores similares (hasta una distancia Hamming de 2 desde el descriptor de consulta original).

Consistencia Temporal Dentro de la Consulta

45 Una vez que los descriptores de consulta tienen correspondencia con la base de datos de audio que utiliza el procedimiento de "hashing", las correspondencias con validadas para determinar cuál de las bases de datos que regresa las respuestas son correspondencias precisas. De otra manera, un descriptor candidato podría tener muchos cuadros con correspondencia al descriptor de consulta pero con una estructura temporal equivocada.

50 En algunas ejecuciones, la validación se logra al visualizar cada respuesta de la base de datos como soporte para una correspondencia en un desfase de la fase de datos de consulta específica. Por ejemplo, si el descriptor 8 (q_8) en un fragmento de código de consulta de "Seinfeld" con un largo de 415 cuadros de 5 segundos, q , da el descriptor de 1008 ésimas base de datos (x_{1008}), esto soporta una correspondencia del candidato entre la consulta de 5 segundos y los cuadros 1001 a 1415 en la base de datos de audio. Otras correspondencia entre el q_n y x_{1000+i_n} , ($1 \leq n \leq 415$) soportaría esta misma correspondencia del candidato.

55 Además de la consistencia temporal, necesitamos contar los cuadros cuando las conversaciones apagan temporalmente el audio ambiente. Este se puede modelar como un interruptor exclusivo entre el audio ambiente y los sonidos de interferencia. Para cada cuadro de consulta i , existe una variable oculta, y_i : si $y_i=0$, el cuadro i ésimo de la consulta se modela como interferencia solamente; si $y_i=1$, el cuadro i ésimo se modela desde un audio ambiente limpio. Tomando una visión extrema (ambiente puro o interferencia pura) se justifica por la precisión extremadamente baja con la cada cuadro de audio está representado (32 bits) y suavizado al suministrar probabilidades de bit-flip adicionales

para cada una de las 32 posiciones del vector de cuadro bajo cada una de las dos hipótesis ($y_i=0$ y $y_i=1$). Finalmente, modelamos las transiciones entre cuadros entre los estados de solo ambiente y solo interferencias como un proceso Markov de primer orden oculto, con las probabilidades de transición derivadas de los datos de entrenamiento. Por ejemplo, podemos reutilizar el modelo de probabilidad de 66 parámetros dado Ke et al., CVPR 2005.

- 5 El modelo final de la probabilidad de correspondencia entre el vector de consulta, q , y , un vector de base de datos ambiente en un desfase de N cuadros, x_N , es:

$$P(q | x^N) = \prod_{n=1}^{415} P(\langle q_n, x_{N+n} \rangle | y_n) P(y_n | y_{n-1}) \quad (1)$$

en donde $\langle q_n, x_m \rangle$ denota las diferencias de bit entre los vectores de cuadro de 32 bits q_n y x_m . Este modelo incorpora tanto la restricción de consistencia temporal como el modelo Markov oculto ambiente/interferencia.

10 Filtro de Consistencia Post-Correspondencia

La gente a menudo habla con otros mientras que ve televisión, dando como resultado una interferencia acústica esporádica pero fuerte, especialmente cuando utiliza micrófonos de los portátiles para muestrear el audio ambiente. Dado que la mayoría de las expresiones de conversación son de 2 y 3 segundos de duración, intercambio de comunicación simple entre los observadores podría generar una consulta de 5 segundos no reconocible.

- 15 En algunas ejecuciones, se utiliza filtrado post-correspondencia para manejar estas faltas de correspondencia de baja confianza intermitentes. Por ejemplo, podemos utilizar un modelo Markov oculto de tiempo continuo de un canal que cambia con un tiempo de vida esperado (es decir el tiempo entre cambio de canales) de L segundos. El servidor de aplicación social 106 indica la correspondencia de más alta confianza dentro del pasado reciente junto con su confianza "discontinua" como parte de la información de estado asociada con cada sesión de cliente. Utilizando esta información el servidor 106 selecciona la correspondencia del índice de contenido del pasado reciente o la coincidencia del índice corriente, con base en aquello que tenga la mayor confianza.

Utilizamos M_h y C_h para referirnos a la mejor correspondencia para la etapa de tiempo previa (hace 5 segundos) y su calificación de confianza de la probabilidad de ingreso. Si nosotros simplemente aplicamos el modelo Markov a esta mejor coincidencia previa, sin tomar otra observación, entonces nuestra expectativa es que la mejor correspondencia para el tiempo corriente es aquella misma para la secuencia de programa, solo 5 segundos adicionales, y nuestra confianza a esa expectativa es de $C_h - 1/L$, en donde $1 = 5$ segundos es la etapa del tiempo de consulta. Este descuento de $1/L$ en la probabilidad de entrada corresponde a la probabilidad del modelo Markov, $e = 1/L$, de no cambiar canales durante la etapa con un tiempo de duración 1 .

- 30 Una hipótesis alternativa se genera por la correspondencia de audio para la consulta corriente. Utilizamos M_0 para referirnos a la mejor correspondencia para el fragmento de código de audio corriente: esto es, la correspondencia que se genera por el proceso de estampado de huella de audio 400. C_0 es la calificación de confianza de la probabilidad de entrada dada por el proceso de estampado de huella de audio 400.

Si estas dos correspondencias (la expectativa histórica actualizada y la observación del fragmento de código corriente) dan diferentes correspondencias, seleccionamos la hipótesis con mayor calificación de confianza:

$$\{M_0, C_0\} = \begin{cases} \{M_h, C_h - 1/L\} & \text{si } C_h - 1/L > C_0 \\ \{M_0, C_0\} & \text{de otra forma} \end{cases} \quad (2)$$

- 35 en donde M_0 es la correspondencia que se utiliza por el servidor de aplicación social 106 para seleccionar el contenido relacionado y M_0 y C_0 son llevados adelante a la siguiente etapa de tiempo como M_1 , y C_1 .

Interfaz de Usuario

- 40 La FIGURA 5 es un diagrama de flujo de una realización de una interfaz de usuario 208 para interactuar con aplicaciones de personalización masivas. La interfaz de usuario 208 incluye un área de pantalla de capa personalizada 502, un área de pantalla con medio de comentario 504, un área de pantalla con enlaces patrocinados 506 y un área de pantalla de contenido 508. El área de pantalla de capa personalizada 502 proporciona información complementaria y/o imágenes relacionadas con el contenido de video mostrado en el área de pantalla de contenido 508. Las capas personalizadas se pueden navegar utilizando una barra de navegación 510 y un dispositivo de entrada (por ejemplo, un ratón o un control remoto). Cada capa tiene una marca asociada en la barra de navegación 510. Por ejemplo, si el usuario selecciona la marca "Moda", entonces la capa de moda, que incluye contenido relacionado con la moda asociado con "Seinfeld," se presentará en el área de pantalla 502.

- 5 En algunas ejecuciones, la interfaz del lado del cliente 102 incluye un dispositivo de pantalla 210 capaz de presentar la interfaz de usuario 208. En algunas ejecuciones, la interfaz de usuario 208 es una página de red interactiva servida por el servidor de aplicación social 106 y presentada en una ventana del navegador en el dispositivo de pantalla 210. En algunas ejecuciones, la interfaz de usuario 208 es persistente y estará disponible para interacción después del audio de difusión utilizado en el proceso de correspondencia de contenido que ha cambiado con el tiempo. En algunas ejecuciones, la interfaz de usuario 208 es dinámicamente actualizado durante el tiempo o en respuesta a un evento de disparo (por ejemplo el ingreso de nuevas personas a la sala de conversación, el inicio de comerciales, etc.). Por ejemplo, cada vez que un comercial es difundido, el área de pantalla con enlaces patrocinados 506 se puede actualizar con enlaces frescos 518 relacionados con la materia objeto del comercial.
- 10 En algunas ejecuciones, la información personalizada y los enlaces patrocinados se pueden enviar por correo al observador o ser mostrados en la barra lateral un momento después.
- 15 En algunas ejecuciones, la interfaz del lado del cliente 102 recibe información personalizada desde el servidor de aplicación social 106. La información puede incluir una página de red, correo electrónico, un tablero de mensajes, enlaces, mensajes instantáneos, una sala de conversación, o una invitación a unirse a un grupo de discusión en proceso, "eRoom", video conferencia o encuentros en la red, llamadas de voz (por ejemplo, Skype®), etc. En algunas ejecuciones, la interfaz de usuario 208 proporciona acceso a comentarios y/o enlaces a comentarios de discusiones o películas previamente vistas. Por ejemplo, si el usuario está viendo actualmente un DVD de "Shrek" el puede querer ver lo que la gente dijo acerca de la película del pasado.
- 20 En algunas ejecuciones, el área de pantalla 502 incluye una región de calificación 512, que se utiliza para visualizar calificaciones de popularidad relacionadas con una difusión. Por ejemplo, el área de pantalla 512 puede mostrar como diferentes observadores están en este momento viendo "Seinfeld" comparado con otro programa de televisión que está siendo difundido al mismo tiempo.
- 25 En algunas ejecuciones, el área de pantalla con medio de comentario 504 presenta un ambiente tipo sala de conversación en donde múltiples usuarios pueden comentar acerca de las difusiones. En algunas ejecuciones, el área de pantalla 504 incluye un cuadro texto 514 para ingresar los comentarios que son enviados a la sala de conversación utilizando el mecanismo de entrada 516 (por ejemplo, un botón).
- El área de pantalla con enlaces patrocinados 506 incluye información, imágenes y/o enlaces relacionados con las propagandas que se asocian con la difusión. Por ejemplo, uno de los enlaces 518 puede llevar al usuario a un sitio de red que está vendiendo mercancía de "Seinfeld".
- 30 El área de pantalla de contenido 508 es en donde se visualiza el contenido de difusión. Por ejemplo, una escena de la difusión corriente se puede visualizar con otra información relevante (por ejemplo, el número de episodio, título, inserción de la marca de tiempo etc.). En algunas ejecuciones, el área de pantalla 508 incluye controles 520 (por ejemplo, botones de desplazamiento) para navegar a través del contenido visualizado.

Marcadores de Video

- 35 En algunas ejecuciones, se incluye un botón 522 en el área de pantalla de contenido que se puede utilizar para marcar video. Por ejemplo, al pulsar el botón 522, el episodio de "Seinfeld" mostrado en el área de pantalla 508 se agrega a la biblioteca de video de los favoritos del usuario, que puede ser entonces visualizada por demanda a través de una aplicación en flujo basada en la red u otros métodos de acceso. De acuerdo con la política establecida por el propietario del contenido, este servicio en flujo puede suministrar una reproducción de visualización única libre, recolectar pagos como el agente para los propietarios del contenido, o insertar publicidad que suministraría pago a los propietarios de contenido.

Arquitectura de Hardware de Interfaz del Lado del Cliente

- 45 La FIGURA 6 es un diagrama de bloque de la arquitectura de hardware 600 para la interfaz del lado del cliente 102 mostrada en la FIGURA .1. Aunque la arquitectura de hardware 600 es típica de un dispositivo de computo (por ejemplo, un ordenador personal), las implementaciones descritas se pueden efectuar en cualquier dispositivo capaz de presentar una interfaz de usuario en un dispositivo de pantalla, que incluye pero no se limita a: ordenadores de mesa o portátil; dispositivos electrónicos; teléfonos; teléfono móviles; sistemas de pantalla; televisores; monitores; sistemas de navegación; reproductores/grabadores de medios portátiles; ayudantes digitales personales; sistemas de juegos; dispositivos electrónicos manuales; y dispositivos o aparatos electrónicos embebidos.
- 50 En algunas ejecuciones, el sistema 600 incluye uno o más procesadores 602 (por ejemplo, CPU), opcionalmente uno o más dispositivos de pantalla 604 (por ejemplo, CRT, LCD, etc.), una interfaz de micrófono 606, una o más interfaces de red 608 (por ejemplo, puertos USB, Ethernet, FireWire®, etc.), opcionalmente uno o más dispositivos de entrada 610 (por ejemplo, ratón, teclado, etc.) y uno o más medios legibles por ordenador 612. Cada uno de estos componentes está operativamente acoplado a uno o más buses 614 (por ejemplo, EISA, PCI, USB, FireWire®, NuBus, PDS, etc.).
- 55 En algunas ejecuciones, no existen dispositivos de pantalla o dispositivos de entrada y el sistema 600 solo efectúa muestreo y codificación (por ejemplo, generar descriptores, etc.) en el trasfondo sin el ingreso del usuario.

5 El término "medio legible por ordenador" se refiere a cualquier medio que participa en suministrar instrucciones a un procesador 602 para ejecución, que incluye sin limitación medios no volátiles (por ejemplo, discos ópticos o magnéticos), medios volátiles (por ejemplo, memoria) y medios de transmisión. Los medios de transmisión incluyen, sin limitación, cables coaxiales, alambres de cobre y fibra óptica. Los medios de transmisión también puede tomar la forma de ondas acústicas, de luz o radiofrecuencia.

10 Los medios legibles por ordenador 612 incluyen además un sistema operativo 616 (por ejemplo, Mac OS®, Windows®, Unix, Linux, etc.), un módulo de comunicaciones en red 618, el software del cliente 620 y una o más aplicaciones 622. El sistema operativo 616 puede ser multiusuario, multiproceso, multitarea, multihilado, en tiempo real y similar. El sistema operativo 616 efectúa tareas básicas, que incluyen pero no se limitan a: reconocer el ingreso de los dispositivos de entrada 610; enviar la salida a los dispositivos de pantalla 604; mantener el seguimiento de archivos y directorios en dispositivos de almacenamiento 612; controlar dispositivos periféricos (por ejemplo unidades de disco, impresoras, dispositivos de captura de imágenes, etc.); y manejar tráfico en uno o más buses 614.

15 El módulo de comunicaciones de red 618 incluye varios componentes para establecer y mantener conexiones de red (por ejemplo, software para ejecutar protocolos de comunicación, tales como TCP/IP, HTTP, Ethernet, USB, FireWire®, etc.).

20 El software del cliente 620 proporciona diversos componentes de software para ejecutar el lado del cliente de las aplicaciones de personalización masivas y para efectuar las diversas funciones del lado del cliente descritas con respecto a las FIGURAS 1-5 (por ejemplo, identificación de audio ambiente). En algunas ejecuciones, algunos o todos los procesos efectuados por el software de cliente 620 se pueden integrar en el sistema operativo 616. En algunas ejecuciones, los procesos pueden al menos ser parcialmente ejecutados en un circuito electrónico digital o en un hardware de ordenador, firmware, software, o en cualquier combinación de los mismos.

Otras aplicaciones 624 pueden incluir cualquier otra aplicación de software, que incluye pero no se limita a: procesadores de palabra, navegadores, correo electrónico, mensajería instantánea, reproductores de medios, software de telefonía, etc.

25 Detección de Publicidad y Repetición de Difusiones

30 Cuando se prepara una base de datos para búsqueda, ayuda poder premarcar material repetido utilizando los descriptores previamente descritos. El material repetido puede incluir pero no se limita a programas repetidos, publicidad, sub-segmentos (por ejemplo, información en la parte inferior de las pantallas en los programas de noticias), etc. Utilizando estas marcas, el material repetido se puede presentar de manera que no arrastre el otro material más allá del campo de atención del usuario que conduce una búsqueda (por ejemplo más allá de los primeros 10-20 resultados). El proceso 700 descrito adelante proporciona una forma de detectar aquellos duplicados antes de cualquiera de las consultas de búsqueda en la base de datos.

Remoción de Publicidad del Video

35 Una de las quejas que los difusores han tenido a cerca de permitir que el material sea buscado y reproducido es la nueva difusión de propaganda embebida. Desde el punto de vista de los difusores, la repetición de la difusión es contraproducente: baja el valor de las difusiones que el anunciador paga directamente, en razón a que este le da al anunciante publicidad libre. A menos que sea removida toda la publicidad vieja y se ponga nueva publicidad en su lugar de alguna manera que retorne alguna remisión a los difusores originales, ellos no ganan nada con la repetición del material previamente difundido. El proceso 700 descrito adelante suministra una manera de detectar publicidad embebida al observar las repeticiones, posiblemente en conjunto con otros criterios (por ejemplo duración, volumen, actividad visual, cuadros blancos entre paréntesis, etc.).

Resumen de Video

45 Si un "resumen" (es decir una versión más corta) de un material de programa no repetido se requiere, una manera de conseguir esto es retirando la publicidad (como se detectó por el material repetido) y tomar segmentos del material solo siguiendo la ubicación de publicidad. En la difusión de televisión, estas posiciones en el programa contienen normalmente titulares iniciales (antes de los anuncios y resúmenes (solo después de los anuncios). Si se hace un resumen de un programa de noticias que incluye una mezcla de material de publicidad no repetido y repetido, normalmente el material sin anuncios repetido corresponde a un bit de sonido. Estos segmentos generalmente dan menos información que la narración del presentador de la historia de las noticias y son buenos candidatos para remoción. Si se va a hacer un resumen de un programa narrativo (por ejemplo una película o una serie por capítulos), las pistas de audio repetidas corresponden normalmente a los sonidos del tema, la música de la escena, o silencio. De nuevo estos son típicamente buenos segmentos para retirar de un video de resumen. El proceso 700 descrito adelante suministra una manera de detectar estas pistas de audio repetidas de tal manera que ellas se puedan retirar del video de resumen.

55 Proceso de Detección de Repetición

La FIGURA 7 es un diagrama de flujo de un ejemplo de un proceso de detección de repetición 700 de acuerdo con esto. Las etapas del proceso 700 no se tienen que completar en ningún orden particular y al menos algunas etapas se pueden efectuar al mismo tiempo en un ambiente de procesamiento multihilado o paralelo.

5 El proceso 700 inicia al crear una base de datos de estadísticas de audio de un conjunto de contenido tal como alimentaciones de televisión, cargas de video etc. (702). Por ejemplo, la base de datos puede contener descriptores de 32 bit/cuadro, como se describe en Ke et al. Las consultas son tomadas de las bases de datos y corren contra la base de datos para ver en donde ocurren repeticiones (704). En algunas ejecuciones, un segmento corto de estadísticas de audio se toma como una consulta y correr revisando correspondencias no identificadas (coincidencias que no son idénticas) que utilizan técnicas de "hashing" (por ejemplo "hashing" directo o "hashing" sensible a sitio (LSH)) para lograr una lista corta de posibles correspondencias de auditoría. Estas correspondencias candidatas son entonces procesadas en un procedimiento de validación, por ejemplo, como se describe en Ke, et al. El contenido que corresponde a una correspondencia candidata validada se puede identificar como contenido repetido (706).

15 Las correspondencias no idénticas que son las más fuertes "crecen" hacia adelante y hacia atrás en el tiempo, para encontrar los puntos de inicio y finalización del material repetido (708). En algunos ejemplos, esto se puede hacer utilizando técnicas de programación dinámica conocidas (por ejemplo, decodificación Viterbi). Al extender la coincidencia hacia adelante del tiempo, la última porción de tiempo en la correspondencia de "semilla" fuerte se ajusta como "que corresponde" y la última porción de tiempo de la primera correspondencia fuerte creíble adelante para el mismo desfase de la base de datos entre la consulta y la correspondencia se ajusta como "no corresponde". En algunos ejemplos, las calificaciones del correspondencia para los cuadros individuales entre estos dos puntos fijos se utilizan como observaciones y se utiliza un modelo marco de primer orden que permite dentro de las transiciones de estado más una transición única de los estados de "corresponde" a "no corresponde". La probabilidad de transición desde corresponde a no corresponde a $1/L$ se puede ajustar un poco arbitrariamente, en donde L es el número de cuadros entre estos dos puntos fijos, que corresponde a por lo menos el conocimiento de la ubicación de transición dentro del rango permitido. Otra posibilidad para seleccionar las probabilidades de transición utilizaría los perfiles de refuerzo de correspondencia para empujar este estimado a una transición anterior o posterior. Pero esto incrementaría la complejidad del modelo de programación dinámica y es probable que no mejore los resultados, en razón a que las fortalezas de la correspondencia ya se utilizan como observaciones dentro de este periodo. El mismo proceso se utiliza para crecer el segmento que corresponde hacia atrás en el tiempo (por ejemplo solo cambiar pasado/futuro y correr el mismo algoritmo).

30 En algunos ejemplos las claves de audio se combinan con información no auditiva (por ejemplo claves visuales) para obtener decisiones de mayor correspondencia. Por ejemplo, las correspondencias que se encuentran con la correspondencia de audio se pueden entonces verificar (o revisar una segunda vez) al utilizar métricas de similitud visual simples (710). Estas métricas pueden incluir pero no se limitan a: histogramas de color (por ejemplo, frecuencias de colores similares en dos imágenes), estadísticas del número y distribución de bordes, etc. Esto no requiere ser calculado solamente sobre la imagen completa, sino que es calculado para subregiones de las imágenes también, y comparado con las subregiones correspondientes en la imagen objetivo.

40 Para aquellas aplicaciones que están buscando publicidad (en contraste con todos los tipos de material repetidos), los resultados de la detección del material repetido se pueden combinar con métricas destinadas a distinguir la publicidad de la no publicidad (712). Estas características distinguibles se pueden pasar en convenciones de publicidad, tales como vibraciones (por ejemplo, son comunes franjas de 10/15/30 segundos), en volumen (por ejemplo, la publicidad tiende a ser más dura que el material del programa que la rodea, así si el material repetido es más duro que el material en otro lado, es más probable que sea una publicidad), sobre actividad visual (por ejemplo, la publicidad tiende a tener transiciones más rápidas entre tomas y más dentro del movimiento de la toma, así si el material repetido tiene diferencias de cuadro mayores que el material en otro lado, es más probable que haya publicidad), y los cuadros blancos entre (publicidad localmente insertada que normalmente no llena completamente la franja que se deja para esta para alimentación nacional, que resulta en cuadros blancos y silencio en un espacio que es múltiplo de 30 segundos).

50 Una vez que se identifica la publicidad, el material que rodea la publicidad se puede analizar y se pueden generar estadísticas. Por ejemplo, las estadísticas se pueden generar a cerca de cuantas veces se publicitó un producto particular utilizando un creativo particular (por ejemplo, imágenes, textos), o cuantas veces estuvo en el aire un segmento particular, etc. En algunas ejecuciones, una o más publicidades viejas se pueden retirar o reemplazar con nuevas publicidades. Las técnicas adicionales para supresión de publicidad y reemplazo se describen en Covell, M., Baluja, S., Fink, M., Advertisement Detection and Replacement Using Acoustic and Visual Repetition, IEEE Signal Processing Society, MMSP 2006 International Workshop on Multimedia Signal Processing, October 3-6, 2006, BC Canadá.

55 En algunos ejemplos, la información de los propietarios del contenido a cerca de la estructura detallada del contenido (por ejemplo, en donde se insertó material de publicidad, donde se repitieron programas) se puede utilizar para aumentar el proceso 700 e incrementar las precisiones de correspondencia. En algunos ejemplos, se pueden utilizar estadísticas de video para determinar la repetición en lugar del audio. En otros ejemplos, se puede utilizar la combinación de estadísticas de video y audio.

60 Subastas de Fragmentos de Código de Audio

5 En algunas ejecuciones, los anunciantes pueden participar en subastas relacionadas con la presencia de audio ambiente que se relaciona con el producto o servicio que el anunciante quiere vender. Por ejemplo, múltiples anunciantes podrían presentar propuestas en una subasta por el derecho a asociar sus productos o servicios con el fragmento de código de audio o de escritura asociado con "Seinfeld." El ganador de la subasta pondría alguna información relacionada al frente del observador (por ejemplo, enlaces patrocinados) siempre y cuando esté presente el audio ambiente objeto. En algunas ejecuciones, los anunciantes podrían hacer propuestas sobre fragmentos de códigos de audio ambiente que tengan una descripción a nivel meta. Por ejemplo, los anunciantes podrían presentar propuestas de audio que estén asociadas con una anuncio de televisión (por ejemplo este el audio asociado con el anuncio de TV de Ford Explorer), en leyendas cerradas (por ejemplo, la leyenda dice "Yankees baseball"), un sitio del segmento de programa (por ejemplo, este audio ocurrirá 15 minutos dentro de "Seinfeld" y ocurrirá 3 minutos después del corte de comerciales ocurrido y un minuto antes del siguiente corte de comerciales, o una acústica de bajo nivel o propiedades visuales (por ejemplo, "música de fondo," "voces de conversación," "similar a explosión", etc.).

10 En algunas ejecuciones, una o más aplicaciones de personalización masivas se pueden correr en el trasfondo mientras que el usuario efectúa otras tareas tales como navegar otro sitio de red (por ejemplo, un enlace patrocinado). El material que se relaciona con una difusión de medio (por ejemplo, contenido de televisión) puede participar en las mismas subastas de enlace patrocinados como material que está relacionado con otra fuente de contenido (por ejemplo contenido del sitio de red). Por ejemplo, los anuncios relacionados con televisión se pueden mezclar con anuncios que corresponden al contenido de una página de red corriente.

15 Se pueden hacer varias modificaciones a las ejecuciones descritas y aún están dentro del alcance de las siguientes reivindicaciones

REIVINDICACIONES

1. Un método ejecutado por ordenador que comprende:
 - 5 Recibir un primer descriptor que identifica un primer audio ambiente de una primera difusión de medio masivo, en donde el primer audio ambiente se genera mediante un primer sistema de medios masivos (202) y se detecta mediante un primer detector de audio ambiente (204);
 - Recibir un segundo descriptor que identifica un segundo audio ambiente de una segunda difusión de medio masivo, en donde el segundo audio ambiente se genera mediante un segundo sistema de medios masivos (202) y se detecta mediante un segundo detector de audio ambiente (204);
 - 10 Comparar el primer y segundo descriptores para determinar si la primera y segunda difusiones de medios masivos son iguales; y
 - 15 Agregar información personalizada con base en el resultado de la comparación, en donde la información personalizada incluye información asociada con un medio comentado, en donde se crea una sala de conversación al comparar directamente los primeros y segundos descriptores en el flujo de datos transmitido por los sistemas del cliente (102) para determinar las correspondencias.
2. Un producto de programa de ordenador que comprende un programa de ordenador operable para hacer que el sistema de procesamiento de datos lleve a cabo todas las etapas del método de acuerdo con la reivindicación.
3. Sistema (100) adaptado para llevar a cabo todas las etapas del método de acuerdo con la reivindicación 1.

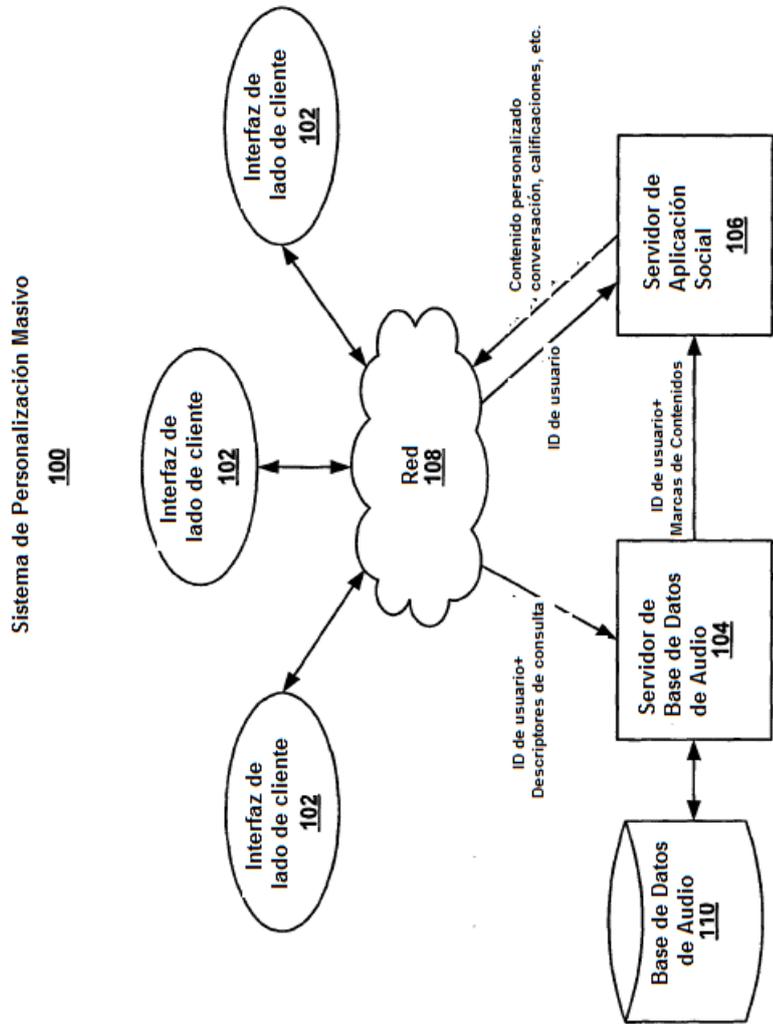


Figura 1

Sistema de Identificación de Audio Ambiente (Lado del Cliente)
200

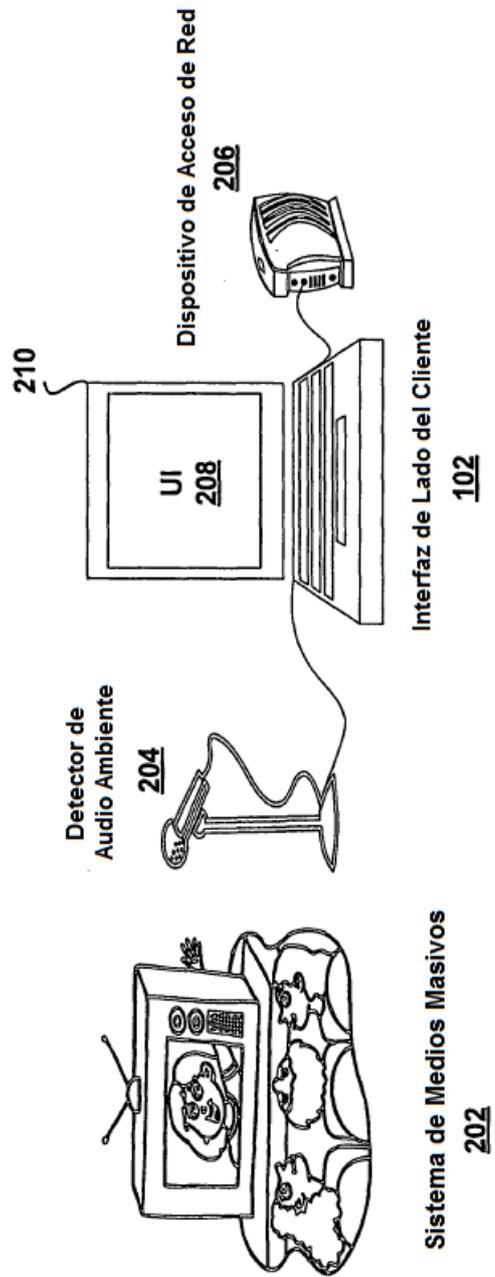


Figura 2

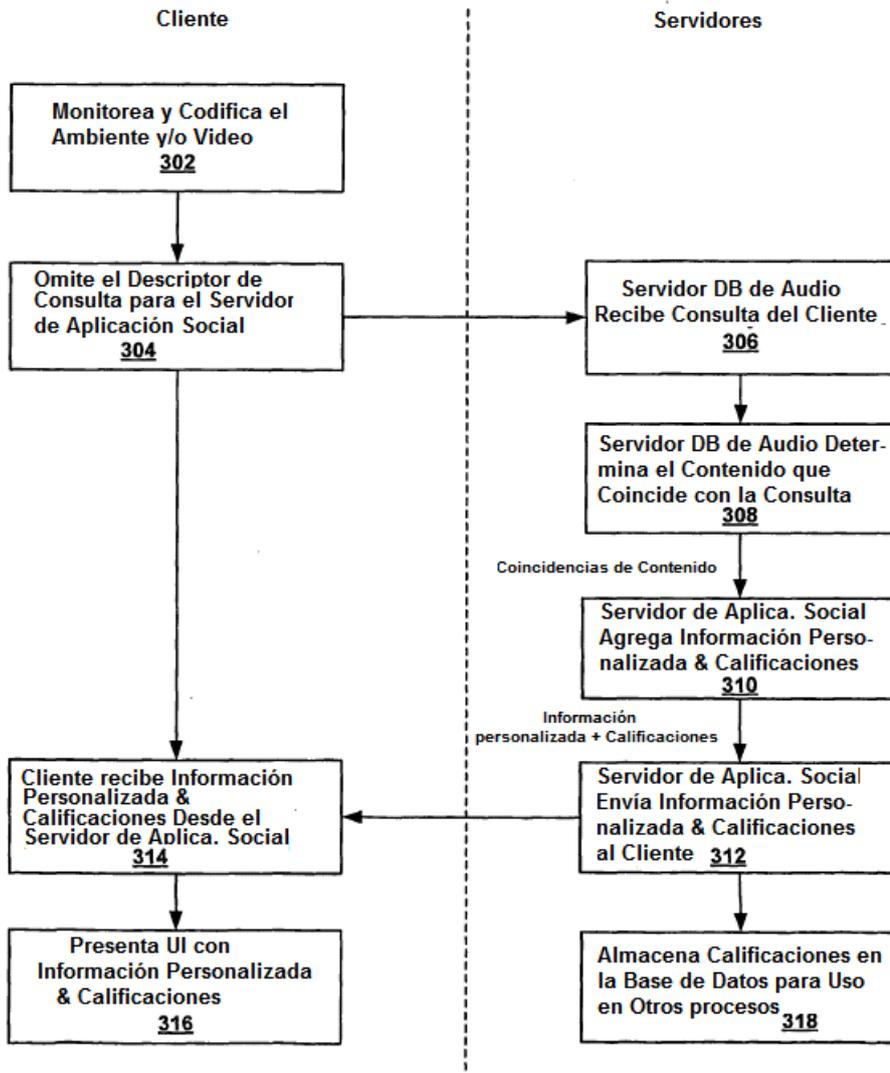
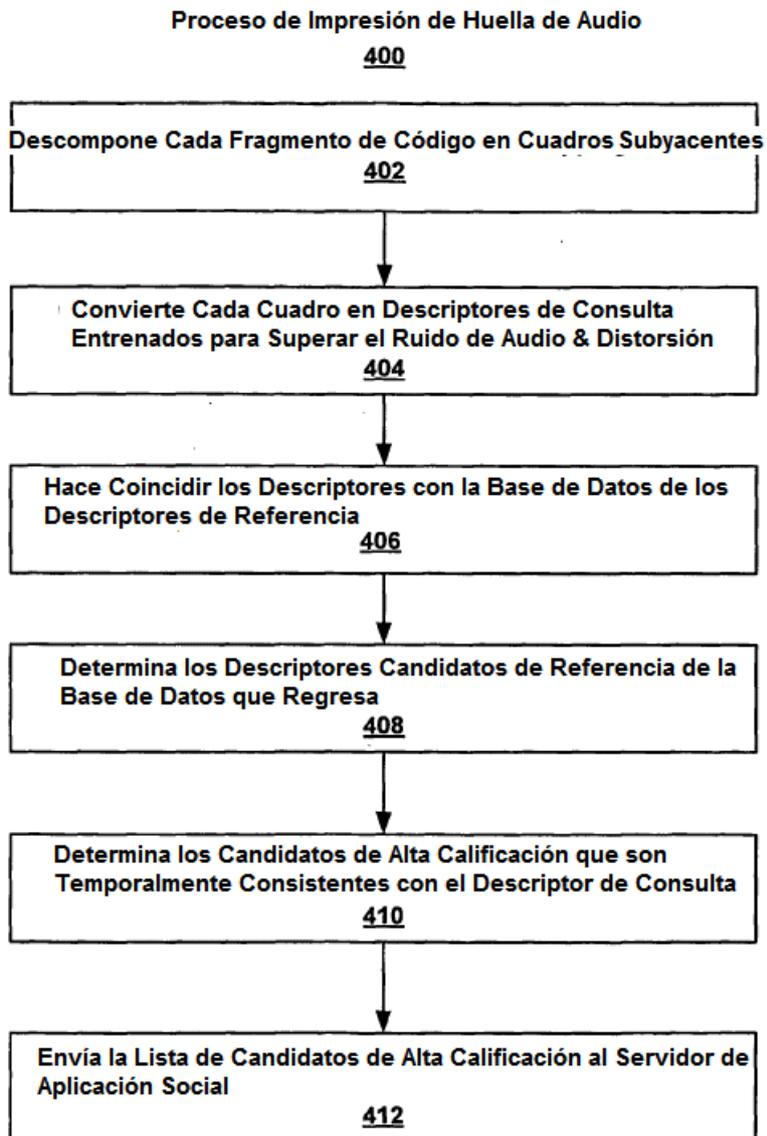


Figura 3



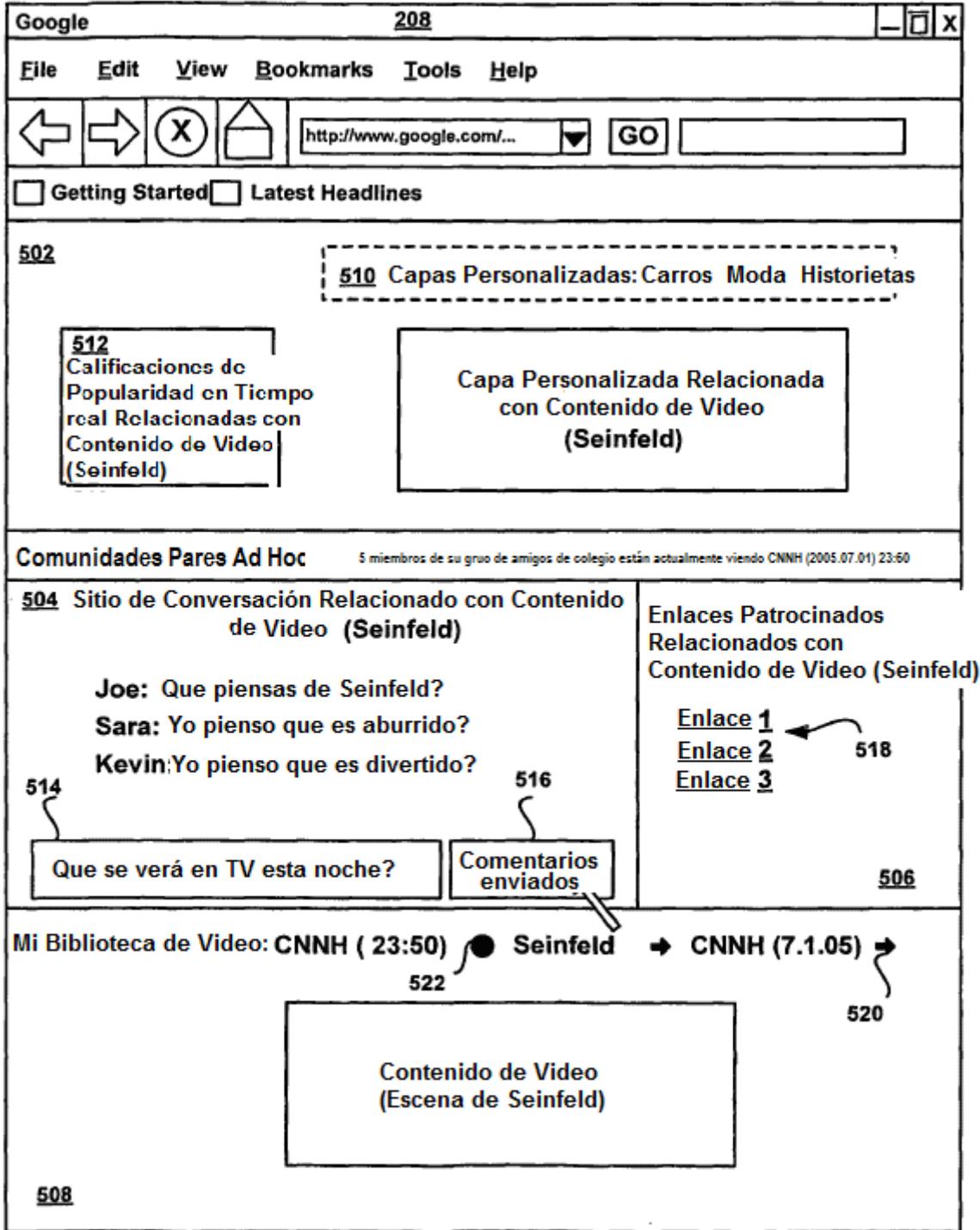


Figura 5

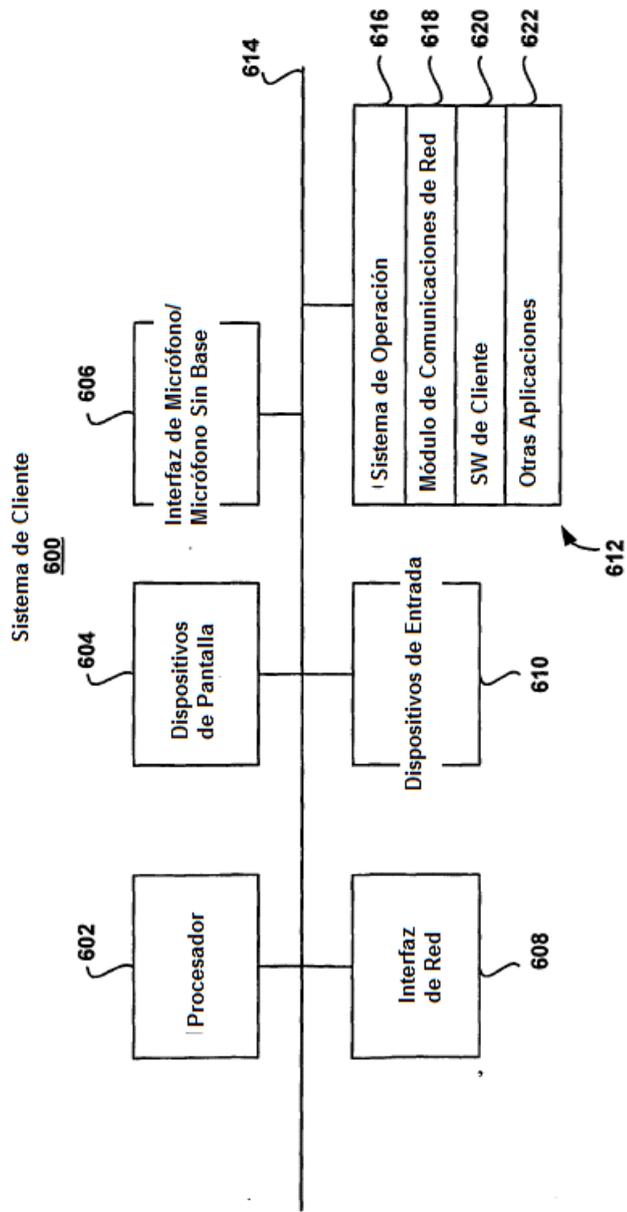


Figura 6

