

COMPLETAMIENTO, SIN CABEZA, DE TAREAS DENTRO DE **ASISTENTES PERSONALES DIGITALES**

ANTECEDENTES

5

A medida que la tecnología de cómputo ha avanzado, se han hecho disponibles dispositivos de cómputo enormemente poderosos. Por ejemplo, los dispositivos de cómputo están agregando enormemente características tales como el reconocimiento de voz. El habla puede ser una manera efectiva para que un usuario pueda comunicarse con un dispositivo de cómputo y se están desarrollando aplicaciones controladas por voz, como asistentes digitales personales controlados por voz.

Un asistente digital personal puede utilizarse para realizar tareas o servicios para un individuo. Por ejemplo, el asistente digital personal puede ser un módulo de *software* que se ejecuta en un dispositivo móvil o una computadora de escritorio. Ejemplos de tareas y servicios que pueden ser realizados por el asistente digital personal pueden incluir la recuperación de las condiciones meteorológicas y previsiones de tráfico, resultados deportivos, instrucciones y condiciones de tráfico, noticias locales y/o nacionales y precios de acciones; el manejo de un programa del usuario mediante la creación de nuevas entradas de programa y recordar al usuario de eventos que vienen; y almacenamiento y recuperación de recordatorios.

No obstante, es probable que el asistente digital personal pueda realizar cada tarea que un usuario puede desear haber realizado. Por lo tanto, existe una amplia oportunidad de mejora en tecnologías relacionadas con asistentes digitales personales controlados por voz.

BREVE DESCRIPCIÓN DE LA INVENCIÓN

Esta Breve Descripción se proporciona para presentar una selección de conceptos en una forma simplificada que se describen a continuación en la Descripción Detallada. Esta Breve Descripción no pretende identificar las principales características o características esenciales de la materia objeto reclamada, ni está destinada a ser utilizada para limitar el alcance de la materia objeto reclamada.

Se describen técnicas y herramientas para completar, sin cabeza (sin periféricos), una tarea de una aplicación en el fondo de un asistente digital personal. Por ejemplo, un método puede ser implementado por un dispositivo de cómputo que comprende un micrófono. El método puede comprender recibir, por un asistente digital personal controlado por voz, una entrada de voz digital generada por un usuario. La entrada de voz digital puede ser recibida a través del micrófono. Puede realizarse procesamiento de lenguaje natural usando la entrada de voz digital para determinar un comando de voz de usuario. El comando de voz de usuario puede incluir una solicitud para realizar una función predefinida de una

aplicación activada por voz de tercera parte. La función predefinida puede ser identificada mediante una estructura de datos que define funciones soportadas por aplicaciones activadas por de tercera parte disponibles utilizando la entrada de voz. Se puede hacer que la

5 aplicación activada por voz de tercera parte ejecute la función predefinida como un proceso de fondo sin una interfaz de usuario de la aplicación activada por voz de tercera parte aparezca en la pantalla del dispositivo de cómputo. Una respuesta puede ser recibida de la aplicación activada por voz de tercera parte indicando

10 un estado asociado con la función de predefinida. Una interfaz de usuario del asistente digital personal controlado por voz puede proporcionar una respuesta al usuario, basándose en el estado recibido asociado con la función de predefinida, de modo que la respuesta proviene de dentro del contexto de la interfaz de usuario

15 del asistente digital personal controlado por voz sin revestimiento de la interfaz de usuario de la aplicación activada por voz de tercera parte.

Como otro ejemplo, los dispositivos de cómputo que comprenden unidades de procesamiento, memoria y uno o más

20 micrófonos pueden ser proporcionados para realizar las operaciones aquí descritas. Por ejemplo, un método realizado por el dispositivo de cómputo puede incluir recibir la entrada de voz generada por un usuario a través de uno o más micrófonos. El reconocimiento de voz puede realizarse utilizando la entrada de voz para determinar un

25 comando hablado. El comando hablado puede comprender una

solicitud para realizar una tarea de una aplicación de tercera parte. La tarea puede ser identificada utilizando una estructura de datos que define tareas de aplicaciones de tercera parte que se invocan por el comando hablado. Se puede determinar si la tarea de la aplicación de tercera parte es capaz de ser ejecutada sin cabeza. Se puede hacer que la aplicación de tercera parte se ejecute como un proceso de fondo para ejecutar sin cabeza la tarea cuando se determine que la tarea de la aplicación de tercera parte es capaz de ser ejecutada sin cabeza. Una respuesta de la aplicación de tercera parte puede ser recibida indicando un estado asociado con la tarea. Una interfaz de usuario del asistente digital personal controlado por voz puede proporcionar una respuesta al usuario, basándose en el estado recibido asociado con la tarea, por lo que la respuesta proviene de dentro del contexto de la interfaz de usuario del asistente digital personal controlado por voz sin revestimiento de la interfaz de usuario de la aplicación de tercera parte.

Como otro ejemplo, se pueden proporcionar dispositivos de cómputo compuestos de unidades de procesamiento y memoria para realizar las operaciones aquí descritas. Por ejemplo, un dispositivo de cómputo puede realizar operaciones para completar una tarea de una aplicación activada por voz dentro del contexto de un asistente digital personal controlado por voz. Las operaciones pueden comprender recibir una entrada de voz digital generada por un usuario en el asistente digital personal controlado por voz. La entrada de voz digital puede ser recibida a través de un micrófono.

Puede realizarse procesamiento de lenguaje natural usando la entrada de voz digital para determinar un comando de voz de usuario. El comando de voz de usuario puede incluir una solicitud para realizar la tarea de la aplicación activada por voz. La tarea
5 puede ser identificada utilizando una estructura de datos extensible que asigna los comandos de voz de usuario a las tareas de aplicaciones activadas por voz. Se puede determinar si la tarea de la aplicación activada por voz es una tarea en primer plano o una tarea de fondo. Cuando se determina que la tarea es una tarea de fondo,
10 se puede hacer que la aplicación activada por voz ejecute la tarea como una tarea de fondo y dentro de un contexto del asistente digital personal controlado por voz sin una interfaz de usuario de la aplicación activada por voz como revestimiento. Una respuesta de la aplicación activada por voz puede ser recibida. La respuesta puede
15 indicar un estado asociado con la tarea. Una respuesta puede ser proporcionada al usuario basándose en el estado recibido asociado con la tarea. La respuesta puede ser proporcionada dentro del contexto del asistente personal digital controlado por voz sin una interfaz de usuario del revestimiento de aplicación de voz cuando se
20 determina que la tarea es una tarea de fondo.

Tal como se describe aquí, una variedad de otras características y ventajas pueden incorporarse a las tecnologías como se desee.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

La Figura 1 es un diagrama que ilustra un ejemplo de un sistema para completar, sin cabeza, una tarea de una aplicación en el fondo de un asistente digital personal.

La Figura 2 es un diagrama que ilustra un ejemplo de una arquitectura de *software* para completar, sin cabeza, una tarea de una aplicación en el fondo de un asistente digital personal.

La Figura 3 es un diagrama de una máquina de estado de ilustrativa para una aplicación que se interconecta con un asistente digital personal.

La Figura 4 es un ejemplo de una definición de comando que puede utilizarse para crear una estructura de datos para habilitar una interfaz entre una aplicación y un asistente digital personal.

La Figura 5 es un ejemplo de un diagrama de secuencia que ilustra la comunicación de varios subprocesos para realizar, sin cabeza, una tarea de una aplicación dentro de un asistente digital personal.

La Figura 6 es un diagrama de flujo de un método ilustrativo para completar, sin cabeza, una tarea de una aplicación en el fondo de un asistente digital personal.

La Figura 7 es un diagrama de flujo de un método ilustrativo para determinar si se debe calentar una aplicación mientras un usuario está hablando con un asistente digital personal.

La Figura 8 es un diagrama de un sistema de cómputo

ilustrativo en el que algunas modalidades pueden ser implementadas.

La Figura 9 es un ejemplo de un dispositivo móvil que puede ser usado en conjunto con las tecnologías aquí descritas.

5 La Figura 10 es un ejemplo de un entorno soportado por nube que puede utilizarse en conjunto con las tecnologías aquí descritas.

DESCRIPCIÓN DETALLADA

10 **Vista general**

A medida que un usuario se siente más cómodo utilizando el asistente digital personal, el usuario puede preferir realizar más acciones dentro del contexto del asistente digital personal. Sin embargo, el proveedor del asistente digital personal no puede
15 predecir o gastar tiempo en desarrollar cada aplicación que un usuario puede querer utilizar. Por lo tanto, puede ser deseable que el asistente digital personal sea capaz de llamar o lanzar aplicaciones de tercera parte que son creadas por entidades distintas a las del proveedor del asistente digital personal.

20 En una solución típica, la interfaz de usuario de la aplicación surge cuando el asistente digital personal inicia la aplicación y el control del programa pasa del asistente digital personal a la aplicación. Una vez que la interfaz de usuario de la aplicación surge, el usuario puede verificar el estado de la solicitud y el usuario puede
25 realizar tareas adicionales desde dentro de la aplicación. Para volver

a la interfaz de usuario del asistente digital personal, el usuario debe salir de la aplicación antes de que el control pueda ser devuelto al asistente digital personal.

Como un ejemplo concreto de utilización de un asistente digital personal de un teléfono móvil, el usuario puede solicitar que una película sea añadida a la cola del usuario utilizando una aplicación de película instalada en el teléfono móvil. Por ejemplo, el usuario puede decir "Aplicación de Película, agregar película-X a mi cola" a la interfaz de usuario del asistente digital personal. Después de que el comando es hablado y reconocido por el asistente, el asistente puede iniciar la aplicación de película que presentará la interfaz de usuario de la aplicación de película. La película puede ser agregada a la cola del usuario y la cola se puede presentar al usuario como una verificación de que la película fue agregada. El usuario puede continuar utilizando la aplicación de película o el usuario puede cerrar la aplicación de película para volver a la interfaz de usuario del asistente digital personal.

Cuando el asistente digital personal pasa el control a la aplicación, la carga de la aplicación y su interfaz de usuario en la memoria puede tomar una cantidad de tiempo perceptible. El retraso puede impactar potencialmente en la productividad del usuario, tal como, al retrasar al usuario llevar a cabo una tarea de seguimiento y/o interrumpiendo el tren de pensamiento del usuario. Por ejemplo, la atención del usuario puede dirigirse a cerrar la aplicación antes de regresar a la interfaz de usuario del asistente digital personal.

Además, al pasar el control a la aplicación, la información contextual disponible al asistente digital personal puede no estar disponible a la aplicación. Por ejemplo, el asistente digital personal puede entender la identidad y la información de contacto del cónyuge del usuario, la
5 ubicación de la casa u oficina del usuario o la ubicación de un proveedor de cuidado infantil del usuario, pero la aplicación no puede tener acceso a la información contextual.

En las técnicas y soluciones aquí descritas, un asistente digital personal puede determinar si una tarea de una aplicación de tercera
10 parte puede realizarse en un segundo plano, de modo que las operaciones para la realización de la tarea se realizan dentro del contexto del asistente digital personal y sin una interfaz de usuario del surgimiento de la aplicación activada por voz. Así, el usuario puede experimentar que un conjunto determinado de tareas se
15 realizan dentro del contexto del asistente digital personal, en oposición al contexto de la aplicación que se está haciendo la tarea de usuario. Además, la potencia del dispositivo puede potencialmente reducirse (y la vida de la batería prolongarse) ya que la interfaz de usuario de la aplicación no está cargada en la memoria
20 cuando la tarea de la aplicación se realiza en el segundo plano.

Las aplicaciones se pueden registrar con el asistente digital personal para ampliar la lista de capacidades nativas que el asistente proporciona. Las aplicaciones se pueden instalar en un dispositivo o llamarse a través de una red (tal como Internet) como
25 un servicio. Una definición de esquema puede habilitar las

solicitudes para registrar un comando de voz con una solicitud para ser lanzada sin cabeza cuando un usuario solicita ese comando/tarea. Por ejemplo, las aplicaciones pueden incluir un archivo de definición de comando de voz (VCD) accesible por el asistente digital personal, en donde el archivo VCD identifica las tareas que pueden ser lanzadas sin cabeza. La definición puede especificar que la tarea de la aplicación siempre será lanzada sin cabeza, o la definición puede especificar que la tarea de la aplicación va a ser lanzada sin cabeza bajo determinadas circunstancias. Por ejemplo, una aplicación puede elegir hacer algo sin cabeza si el usuario está pidiendo que la tarea sea realizada en un dispositivo que no tiene una superficie de visualización (tal como una banda de entrenamiento inalámbrica) o cuando el usuario está operando en modo de manos libres (tal como cuando el usuario está conectado a un auricular Bluetooth).

Las aplicaciones pueden proporcionar una respuesta en el completamiento en progreso, de falla y exitoso de la tarea solicitada y la salida relacionada con los estados puede ser provista por la interfaz de usuario del asistente digital personal. Las aplicaciones pueden ofrecer muchos tipos diferentes de datos de regreso al asistente digital personal incluyendo texto de visualización, texto que se puede leer en voz alta, un profundo enlace para volver a la aplicación, un enlace a una página o sitio web, y el Lenguaje de Marcación de Hipertexto (HTML) basado en el contenido web, por ejemplo. Los datos de la aplicación para que el asistente pueden

presentarse como si vinieran de una función nativa del asistente a través de la interfaz de usuario del asistente.

Si el usuario proporciona una solicitud a la aplicación que puede tener varios significados o resultados, la aplicación puede proporcionar al asistente digital personal con una lista de opciones y la interfaz de usuario del asistente se puede utilizar para desambiguar entre las opciones. Si el usuario proporciona una solicitud para la aplicación que puede ser destructiva o es importante (tal como cuando el usuario solicita que una aplicación bancaria realice una transferencia de saldo), una interfaz de confirmación del asistente se puede utilizar para confirmar la solicitud antes de completar la tarea destructiva o importante.

Las aplicaciones pueden ser cargadas de forma especulativa o calentadas a medida que los comandos están siendo hablados. Por ejemplo, cuando el usuario completa la frase "Aplicación de película" desde el comando, "Aplicación de película, agregar película-X a mi cola", la memoria se puede asignar y diversas subrutinas de una aplicación de película instalada pueden ser recuperadas del almacenamiento y se cargan en la memoria asignada en la preparación para el uso de las subrutinas cuando se completa el comando. Cuando la aplicación es un servicio web, el calentamiento puede incluir el establecimiento de una sesión de comunicación y recuperar información específica de usuario de una base de datos en un servidor remoto, por ejemplo. Al activar la aplicación, el tiempo para responder al usuario puede ser potencialmente disminuido, de

modo que la interacción sea más natural y para que el usuario pueda moverse a la siguiente tarea más rápido, haciendo que los usuarios sean más productivos.

Al utilizar las tecnologías de la presente, el usuario que desea
5 agregar una película a la cola del usuario con una aplicación de película puede tener una experiencia diferente que cuando se utiliza la solución típica de lanzamiento de la aplicación de película y pasar el control a la aplicación. En este ejemplo, el comando de agregar-película-a-cola de la aplicación de película puede definirse como sin
10 cabeza en una estructura de datos de comando, tal como un archivo VCD. Cuando el usuario dice "Aplicación de película" desde el comando, "Aplicación de película, agregar película-X a mi cola", la aplicación de película se puede activar de manera que el tiempo de respuesta para el usuario puede ser reducido. Cuando el comando se
15 completa, la película puede ser agregada a la cola del usuario utilizando la aplicación de película, pero sin emerger la interfaz de usuario de la aplicación de película. La película puede ser agregada a la cola del usuario y el asistente digital personal puede confirmar (utilizando la interfaz de usuario del asistente) que la película fue
20 agregada. El usuario puede experimentar un tiempo de respuesta más rápido y puede realizar menos pasos para completar la tarea (por ejemplo, la aplicación de película no necesita ser cerrada).

Sistema Ilustrativo Incluyendo un Asistente Digital Personal

25 La Figura 1 es un diagrama de sistema que representa un

ejemplo de un sistema 100 para completar, sin cabeza, una tarea 112 de una aplicación activada por voz 110 en el fondo de un asistente digital personal 120. La aplicación activada por voz 110 y el asistente digital personal 120 pueden ser módulos de *software* instalados en un dispositivo de cómputo 130. El dispositivo de cómputo 130 puede ser una computadora de escritorio, un portátil, un teléfono móvil, un teléfono inteligente, un dispositivo portátil (tal como un reloj o banda electrónica inalámbrica), o una computadora de tableta, por ejemplo. El dispositivo de cómputo 130 puede incluir una estructura de datos de comando 140 para identificar aplicaciones y tareas de aplicaciones que pueden ser lanzadas por el asistente digital personal 120. Las aplicaciones pueden ser lanzadas por el asistente digital personal 120 en primer plano (tal como cuando una interfaz de usuario de la aplicación aparece cuando se inicia la aplicación) y/o en segundo plano (tal como cuando la interfaz de usuario de la aplicación no aparece cuando se inicia la aplicación). Por ejemplo, algunas de las tareas de una aplicación pueden ser lanzadas en el primer plano y tareas distintas de la misma aplicación pueden ser lanzadas en segundo plano. La estructura de datos de comando 140 puede definir cómo la aplicación y/o las tareas de la aplicación deben ser lanzadas desde el asistente digital personal 120.

El dispositivo de cómputo 130 puede incluir un micrófono 150 para convertir el sonido a una señal eléctrica. El micrófono 150 puede ser un micrófono dinámico, un condensador o piezoeléctrico

utilizando inducción electromagnética, un cambio en la capacitancia o piezoelectricidad, respectivamente, para producir la señal eléctrica a partir de variaciones de presión de aire. El micrófono 150 puede incluir un amplificador, uno o más filtros analógicos o digitales, y/o un convertidor analógico-digital para producir una entrada de sonido digital. La entrada de sonido digital puede incluir una reproducción de la voz del usuario, tal como cuando el usuario solicita que el asistente digital personal 120 realice una tarea. El dispositivo de cómputo 130 puede incluir una pantalla táctil o teclado (no mostrado) para permitir que el usuario introduzca la entrada de texto.

La entrada de sonido digital y/o la entrada textual pueden ser procesadas por un módulo de procesamiento de lenguaje natural 122 del asistente digital personal 120. Por ejemplo, el módulo de procesamiento de lenguaje natural 122 puede recibir la entrada de sonido digital y traducir las palabras pronunciadas por un usuario en el texto. El texto extraído puede ser semánticamente analizado para determinar un comando de voz de usuario. Mediante el análisis de la entrada de sonido digital y tomar acciones en respuesta a comandos hablados, el asistente digital personal 120 puede ser controlado por voz. Por ejemplo, el asistente digital personal 120 puede comparar el texto extraído con una lista de comandos potenciales de usuario para determinar el comando que más probable coincide con la intención del usuario. La coincidencia puede basarse en métodos probabilísticos o estadísticos, árboles de decisión u otras reglas, otros criterios de coincidencia adecuados o combinaciones de los

mismos. Los comandos potenciales de usuario pueden ser comandos nativos del asistente digital personal 120 y/o comandos definidos en la estructura de datos de comando 140. Así, mediante la definición de los comandos en la estructura de datos 140, la gama de tareas que se pueden realizar en nombre del usuario mediante el asistente digital personal 120 se puede ampliar. Los comandos potenciales pueden incluir realizar la tarea 112 de la aplicación activada por voz 110, que puede ser definida como una tarea sin cabeza o de fondo en la estructura de datos de comando 140.

10 El módulo de procesamiento de lenguaje natural 122 puede generar una secuencia de texto a medida que el habla es procesada para que cadenas de texto intermediarias puedan ser analizadas antes de que una expresión de usuario se complete. Así, si el usuario inicia un comando con el nombre de una aplicación, la aplicación puede ser identificada tempranamente en la expresión y la aplicación se puede activar antes de que el usuario complete el comando. La activación de la aplicación puede incluir recuperar instrucciones de la aplicación desde la memoria no volátil relativamente lenta (tal como un disco duro o memoria Flash) y almacenar las instrucciones en la memoria volátil relativamente más rápida (tal como memoria principal o memoria caché).

Cuando el asistente digital personal 120 determina que un comando está asociado a una tarea de una aplicación, la tarea de la aplicación puede ser ejecutada. Si el asistente digital personal 120 determina que la tarea de la aplicación se va a ejecutar como un

proceso en segundo plano (tal como, por ejemplo, mediante el análisis de la definición de la estructura de datos de comando 140), la aplicación se puede ejecutar en segundo plano. La aplicación, tal como la aplicación de voz 110, puede comunicarse con el asistente digital personal 120. Por ejemplo, la aplicación puede secuenciar a través de un conjunto de estados asociados con el completamiento de la tarea y el estado de la aplicación se puede comunicar al asistente digital personal 120. Por ejemplo, la aplicación puede comenzar en un estado "inicial", la transición a un estado de "progreso" mientras la tarea se está realizando y, a continuación, la transición a un estado "final" cuando la tarea se completa.

El asistente digital personal 120 puede informar sobre el progreso de la tarea a través de una interfaz de usuario 124. La interfaz de usuario 124 puede comunicar la información al usuario de diversas maneras, por ejemplo, al presentar texto, gráficos o hiperenlaces en una pantalla del dispositivo de cómputo 130, al generar salidas de audio desde una bocina del dispositivo de cómputo 130, o al generar otras salidas sensoriales tales como vibraciones de un motor eléctrico conectado a un peso descentrado del dispositivo de cómputo 130. Por ejemplo, la interfaz de usuario 124 puede hacer que una rueda giratoria sea presentada en la pantalla del dispositivo de cómputo 130 cuando la tarea está en el estado de progreso. Como otro ejemplo, la interfaz de usuario 124 puede generar un discurso simulado indicando el completamiento satisfactorio de la tarea cuando la tarea se encuentra en el estado

final y la tarea se completó con éxito. Mediante el uso de la interfaz de usuario 124 del asistente digital personal 120 para informar sobre el estado de la tarea, la respuesta puede venir desde dentro de un contexto de la interfaz de usuario 124 sin emerger una interfaz de usuario de la aplicación.

Se debe observar que la aplicación activada por voz 110 puede ser creada por el productor del asistente digital personal 120 o por una tercera parte que es diferente del productor. La interoperación del asistente digital personal 120 y la aplicación activada por voz 110 puede lograrse cumpliendo con un contrato de *software* de aplicación a aplicación y definiendo la funcionalidad en la estructura de datos de comando 140. La aplicación activada por voz 110 puede ser capaz de funcionar como una aplicación independiente o solo como un componente del asistente digital personal 120. Como una aplicación independiente, la aplicación activada por voz 110 puede ser lanzada fuera del asistente digital personal 120 como un proceso de primer plano como, por ejemplo, tocando o haciendo doble clic sobre un icono asociado con la aplicación activada por voz 110 y presentarse en la pantalla del dispositivo de cómputo 130. La aplicación activada por voz 110 puede presentar una interfaz de usuario cuando es lanzada y el usuario puede interactuar con la interfaz de usuario para realizar tareas. La interacción puede ser sólo con la entrada de voz, o también se pueden utilizar otros modos de entrada, tales como texto de entrada o gestos. Las aplicaciones denominadas por el asistente digital personal 120 pueden ser

instaladas en el dispositivo de cómputo 130 o pueden ser servicios web.

El asistente digital personal 120 puede llamar a servicios web, tales como el servicio web 162 que se ejecuta en la computadora de servidor remoto 160. Los servicios Web son funciones de *software* proporcionados en una dirección de red a través de una red, como una red 170. La red 170 puede incluir una red de área local (LAN), una red de área amplia (WAN) o Internet, una intranet, una red cableada, una red inalámbrica, una red celular, y combinaciones de los mismos, o cualquier red adecuada para proporcionar un canal de comunicación entre el dispositivo de cómputo 130 y la computadora de servidor remoto 160. Cabe destacar que la topología de red que se ilustra en la Figura 1 se ha simplificado y que múltiples redes y dispositivos de red se pueden utilizar para interconectar los diferentes sistemas de cómputo aquí descritos. El servicio web 162 puede ser llamado como parte del núcleo o parte principal del asistente digital personal 120. Por ejemplo, el servicio web 162 puede ser llamado como una subrutina del módulo de procesamiento de lenguaje natural 122. Además, o como alternativa, el servicio web 162 puede ser una aplicación definida en la estructura de datos de comando 140 y puede ser susceptible de ser lanzada, sin cabeza, desde el asistente digital personal 120.

Ejemplo de Arquitectura de Software incluyendo un Asistente Digital Personal

La Figura 2 es un diagrama que ilustra un ejemplo de arquitectura de *software* 200 para completar, sin cabeza, una tarea de una aplicación en el fondo de un asistente digital personal 120. Cuando se realiza una tarea de una aplicación, sin cabeza, la tarea puede ejecutarse en segundo plano y una interfaz de usuario de la aplicación no emerge como resultado de la tarea que se está realizando. Más bien, la interfaz de usuario del asistente digital personal 120 puede utilizarse para proporcionar una salida y/o entrada del usuario de modo que el usuario interactúa dentro del contexto del asistente digital personal 120 y no en el contexto de la aplicación. Por lo tanto, una tarea ejecutada sin cabeza de una aplicación se puede ejecutar en segundo plano para la duración de la ejecución de la tarea y la interfaz de usuario de la aplicación nunca emerge. Un dispositivo de cómputo, como el dispositivo de cómputo 130, puede ejecutar el *software* para un asistente digital personal 120, un núcleo de sistema operativo (OS) 210 y una aplicación 230 organizada según la arquitectura 200.

El núcleo de sistema operativo 210 generalmente proporciona una interfaz entre los componentes de *software* y los componentes de *hardware* del dispositivo de cómputo 130. El núcleo de sistema operativo 210 puede incluir componentes para presentar (por ejemplo, presentar salida visual a una pantalla, generar salida de voz y otros sonidos para un altavoz, y generar una salida vibrante para

un motor eléctrico), componentes para redes, componentes para el manejo de procesos, componentes para el manejo de memoria, componentes para el rastreo de ubicaciones, y componentes para el reconocimiento de voz y otro procesamiento de entrada. El núcleo de sistema operativo 210 puede administrar funciones de entrada de usuario, funciones de salida, funciones de acceso al almacenamiento, funciones de comunicación de red, funciones de manejo de memoria, funciones de manejo de procesos, y otras funciones para el dispositivo de cómputo 130. El núcleo de sistema operativo 210 puede proporcionar acceso a esas funciones al asistente digital personal 120 y la aplicación 230, tal como a través de varias llamadas de sistema.

Un usuario puede generar la entrada del usuario (tal como, por ejemplo, voz, tacto y movimiento) para interactuar con el asistente digital personal 120. El asistente digital personal 120 puede hacerse consciente de la entrada del usuario a través del núcleo de sistema operativo 210, que puede incluir la funcionalidad de creación de mensajes en respuesta a la entrada de usuario. Los mensajes pueden ser utilizados por el asistente digital personal 120 u otro *software*. El usuario puede incluir entrada táctil, tal como entrada de pantalla táctil, pulsaciones de botón o pulsaciones de teclas. El núcleo de sistema operativo 210 puede incluir la funcionalidad de reconocer toques, gestos de dedo, etc. en una pantalla táctil de entrada táctil, entrada de botón o entrada de pulsación de tecla. El núcleo de sistema operativo 210 puede recibir la entrada del

micrófono 150 y puede incluir la funcionalidad de reconocimiento de comandos de voz y/o palabras de la entrada de voz. El núcleo de sistema operativo 210 puede recibir la entrada de un acelerómetro y puede incluir la funcionalidad de reconocimiento de orientación de o
5 movimiento, tal como temblor.

El motor de procesamiento de entrada de interfaz de usuario (UI) 222 del asistente digital personal 120 puede esperar para mensajes de evento de entrada de usuario del núcleo de sistema operativo 210. Los mensajes de evento de interfaz de usuario
10 pueden indicar una palabra reconocida desde la entrada de voz, un gesto de paneo, gesto de parpadeo, gesto de arrastre, u otro gesto en una pantalla táctil del dispositivo, un toque en la pantalla táctil, entrada de pulsaciones de teclas, un gesto de agitación, u otros eventos de interfaz de usuario (por ejemplo, botones direccionales o
15 entrada de seguibola). El motor de procesamiento de entrada de interfaz de usuario 222 puede traducir los mensajes de evento de interfaz de usuario del núcleo de sistema operativo 210 en información enviada a la lógica de control 224 del asistente digital personal 120. Por ejemplo, el motor de procesamiento de entrada de
20 interfaz de usuario 222 puede incluir capacidades de procesamiento de lenguaje natural y puede indicar que un determinado nombre de aplicación se ha hablado o escrito o que un comando de voz ha sido dado por el usuario. Alternativamente, las capacidades de procesamiento de lenguaje natural pueden incluirse en la lógica de
25 control 224.

La lógica de control 224 puede recibir información de varios módulos del asistente digital personal 120, tales como el motor de procesamiento de entrada de interfaz de usuario 222, almacenamiento de información personalizada 226 y la estructura de datos de comando 140 y la lógica de control 224 puede tomar decisiones y realizar operaciones basándose en la información recibida. Por ejemplo, la lógica de control 224 puede determinar si el asistente digital personal 120 debe realizar una tarea en nombre del usuario, como, por ejemplo, mediante el análisis de una secuencia de texto hablado para determinar si un comando de voz ha sido dado.

La lógica de control 224 puede esperar para que todo el comando de usuario sea hablado antes de actuar sobre el comando o la lógica de control 224 puede empezar a actuar sobre el comando a medida que sigue siendo hablado y antes de que haya terminado. Por ejemplo, la lógica de control 224 puede analizar cadenas intermediarias del comando hablado y tratar de hacer coincidir las cadenas con una o más aplicaciones definidas en la estructura de datos de comando 140. Cuando la probabilidad de que una aplicación se llamará excede un umbral, la aplicación puede ser activada para que la aplicación pueda responder al usuario más rápidamente. Múltiples aplicaciones y/o funciones pueden ser activadas de forma especulativa en previsión de que se llamen y las aplicaciones pueden detenerse si se determina que la solicitud no será llamada. Por ejemplo, cuando el usuario inicia el comando hablado con el nombre

de una aplicación en particular, existe una alta probabilidad de que la aplicación particular será llamada, y así la aplicación será activada. Como otro ejemplo, algunas cadenas de comando parciales pueden ser limitadas a un pequeño conjunto de aplicaciones definidas en la estructura de datos de comando 140, y el conjunto de aplicaciones se puede activar en paralelo cuando hay una coincidencia en la cadena de comando parcial. Específicamente, la estructura de datos de comando 140 puede tener sólo dos aplicaciones con comandos que tengan la palabra "tomar", tal como una aplicación de cámara con un comando "tomar una foto", y una aplicación de notas con un comando "tomar nota". La lógica de control 224 puede comenzar la activación tanto de la aplicación de cámara como la aplicación de notas cuando la palabra "tomar" es reconocida y luego la aplicación de notas puede ser detenida cuando el comando completo de "tomar una foto" es reconocido. La activación de la aplicación puede incluir la asignación de memoria, pre-búsqueda de instrucciones, establecer una sesión de comunicación, recuperar información de una base de datos, iniciar un nuevo hilo de ejecución, plantear una interrupción, o de otro tipo adecuado de operaciones específicas de la aplicación. Durante la activación pueden ser llamados servicios del núcleo de sistema operativo 210, tal como el servicio de manejo de procesos, el servicio de manejo de memoria, y el servicio de red, por ejemplo.

El texto hablado puede incluir información contextual y la lógica de control 224 puede resolver la información contextual, de

modo que el comando de voz del usuario es libre de contexto. La información contextual puede incluir una ubicación actual, un tiempo actual, una orientación del dispositivo de cómputo 130 y la información personal almacenada en el almacenamiento de información personalizada 226. La información personal puede 5 incluir: relaciones de usuario tales como un cónyuge del usuario o el nombre del niño; ubicaciones específicas de usuario tales como direcciones del hogar, el trabajo, la escuela, la guardería, o doctor; información de la lista de contactos o calendario del usuario; el color 10 favorito del usuario, restaurante, o método de transporte; cumpleaños importantes, aniversarios u otras fechas; y otra información específica del usuario. El usuario puede dar un comando con información contextual y la lógica de control 224 puede traducir el comando en un comando libre de contexto. Por ejemplo, el usuario 15 puede dar el comando, "Ap de camión, me dice los camiones a casa dentro de la hora siguiente." En este ejemplo, la información contextual en el comando es la fecha y la hora actual, la ubicación actual y la ubicación de la casa del usuario.

La lógica de control 224 puede obtener la hora actual del 20 núcleo de sistema operativo 210 que puede mantener o tener acceso a un reloj de tiempo real. La lógica de control 224 puede obtener los datos de ubicación actual para el dispositivo de cómputo 130 desde el núcleo de sistema operativo 210, que puede obtener los datos de ubicación actual desde un componente local del dispositivo de 25 cómputo 130. Por ejemplo, los datos de ubicación pueden

determinarse a partir de los datos procedentes de un sistema de posicionamiento global (GPS), mediante la triangulación entre torres de una red celular, por referencia a ubicaciones físicas de enrutadores Wi-Fi en las proximidades, o por otro mecanismo. La

5 lógica de control 224 puede obtener la ubicación de la casa del usuario del almacenamiento de información personalizada 226. El almacenamiento de información personalizada 226 puede almacenarse en un almacenamiento auxiliar u otro almacenamiento no volátil del dispositivo de cómputo 130. De esta forma, la lógica de

10 control 224 puede recibir la información personalizada a través del núcleo de sistema operativo 210 que puede acceder a los recursos de almacenamiento (por ejemplo, el almacenamiento de información personalizada 226). Cuando la información contextual puede ser resuelta, el comando puede ser traducido a un comando libre de

15 contexto. Por ejemplo, si es el viernes a las 6:00 p.m., el usuario está en el 444 de Main Street y el domicilio del usuario es 123 Pleasant Drive, entonces el comando libre de contexto puede ser la "ap de camión, dígame los camiones que llegan cerca de 444 Main Street y pasan cerca de 123 Pleasant Drive entre las 6:00 y 7:00

20 p.m. del viernes".

El comando de usuario puede ser realizado por la lógica de control 224 (tal como cuando el comando es un comando nativo del asistente digital personal 120), una aplicación 230 instalada en el dispositivo de cómputo 130 (tal como cuando el comando está

25 asociado con la aplicación 230), o el servicio web 162 (tal como

cuando el comando está asociado con el servicio web 162). La estructura de datos de comando 140 puede especificar qué comandos están asociados con las aplicaciones y si el comando puede realizarse en primer plano o en segundo plano. Por ejemplo, 5 la estructura de datos de comando 140 puede asignar comandos de voz de usuario a las funciones apoyadas por aplicaciones activadas por voz disponibles de tercera parte.

La lógica de control 224 puede causar que una función predefinida 232 de la aplicación 230 sea ejecutada cuando la lógica 10 de control 224 determina que el comando de usuario está asociado con la función predefinida 232 de la aplicación 230. Si la lógica de control 224 determina que la función predefinida 232 de la aplicación 230 se va a ejecutar como un proceso de fondo, la función predefinida 232 se puede ejecutar en segundo plano. Por ejemplo, la 15 lógica de control 224 puede enviar una solicitud 240 a la función predefinida 232 levantando una interrupción, escribiendo a la memoria compartida, escribiendo a una cola de mensajes, transmitiendo un mensaje o iniciando un nuevo hilo de ejecución (tal como a través del componente de manejo de proceso del núcleo de 20 sistema operativo 210). La aplicación 230 puede realizar la función pre-definida 232 y devolver una respuesta 242 a la lógica de control 224 al levantar una interrupción, al escribir a la memoria compartida, al escribir a una cola de mensajes o al transmitir un mensaje. La respuesta puede incluir un estado de la aplicación 230 y/u otra 25 información en respuesta al comando de usuario.

La lógica de control 224 puede causar que el servicio web 162 sea llamado cuando la lógica de control 224 determina que el comando está asociado con el servicio web 162. Por ejemplo, una solicitud 260 puede ser enviada al servicio web 162 a través del componente de red del núcleo de sistema operativo 210. El
5 componente de red puede dar formato y reenviar la solicitud a través de la red 170 (tal como al encapsular la solicitud en un paquete de red según el protocolo de la red 170) al servicio web 162 para realizar el comando de usuario. La solicitud 260 puede incluir varias
10 medidas tales como la abertura de un canal de comunicación (por ejemplo, un conector) entre la lógica de control 224 y el servicio web 162 y el envío de información relacionada con el comando de usuario. El servicio web 162 puede responder a la solicitud 260 con una respuesta que puede ser transmitida a través de la red 170 y
15 transmitida por el componente de red a la lógica de control 224 como una respuesta 262. La respuesta del servicio web 162 puede incluir un estado del servicio web 162 y otra información en respuesta al comando de usuario.

La lógica de control 224 puede generar una salida (con la ayuda de un motor de representación de salida de IU 228 y un
20 componente de representación del núcleo de sistema operativo 210) que será presentada al usuario basándose en las respuestas de las aplicaciones. Por ejemplo, la estructura de datos de comando 140 puede asignar estados recibidos desde las funciones de respuestas
25 provistas al usuario del asistente digital personal controlado por voz

120. En general, la lógica de control 224 puede proporcionar comandos de salida de alto nivel a al motor de representación de salida de UI 228 que pueden producir primitivos de salida de nivel inferior al componente de representación del núcleo de sistema operativo 210 para salida visual en una pantalla, salida de audio y/o voz a través de un altavoz o auriculares y salida de vibración de un motor eléctrico. Por ejemplo, la lógica de control 224 puede enviar un comando de texto-a-habla con una cadena de texto en el motor de representación de salida de UI 228 que puede generar los datos de audio digital simulando una voz hablada.

La lógica de control 224 puede determinar qué información va a proporcionar al usuario basándose en el estado de la aplicación. Los estados pueden corresponder al comienzo, procesamiento, confirmación, desambiguación o finalización de un comando de usuario. La estructura de datos de comando 140 puede asignar los estados de la aplicación a diferentes respuestas que se proporcionarán a los usuarios. Los tipos de información que pueden ser proporcionados incluyen texto en pantalla, voz simulada, un enlace profundo para volver a la aplicación, un enlace a una página web o sitio web y el Lenguaje de marcado de hipertexto (HTML) basado en contenido web, por ejemplo.

Estados de Aplicación Ilustrativos

La Figura 3 es un diagrama de una máquina de estado ilustrativa 300 para una aplicación que interconecta con el asistente

digital personal 120 en modo sin cabeza. La aplicación puede comenzar en un estado de activación 310 o un estado inicial 320. El estado de activación 310 puede ser introducido cuando el asistente digital personal 120 hace que la aplicación se active, tal como cuando se conoce el nombre de la aplicación, pero el comando hablado no está completo. La aplicación permanecerá en el estado de activación 310 hasta las operaciones de activación se completan. Cuando las operaciones de activación se completan, la aplicación puede realizar la transición al estado inicial 320.

10 El estado inicial 320 puede ser introducido después de que el estado de activación 310 se ha completado o después de que el comando de usuario se proporciona por el asistente digital personal 120 a la aplicación. Durante el estado inicial de 320, el comando de usuario es procesado por la aplicación. Si el comando es no ambiguo pero tardará más de un cantidad de tiempo pre-determinada para el completamiento (tal como cinco segundos), el estado puede pasarse a un estado de progreso 330 cuando el comando se está llevando a cabo. Si el comando es no ambiguo y puede resultar en una operación importante o destructiva que se está realizando, el estado puede pasarse a un estado de confirmación 340. Si el comando es algo ambiguo, pero la ambigüedad puede ser aclarada al elegir entre varias opciones, el estado puede pasarse a un estado de desambiguación 350. Si el comando es ambiguo y no se puede desambiguar con pocas opciones, el estado puede pasarse a un estado final 360, tal como un estado de falla o un estado de

15

20

25

redirección. Si el comando no se efectúa, el estado puede pasarse a un estado final 360, tal como el estado de falla. Si el comando se puede completar en menos de una cantidad predeterminada de tiempo para completar y no se desea solicitar la confirmación del usuario, el estado puede pasarse a un estado final de 360, tal como un estado de éxito. Cabe señalar que el estado final 360 puede ser un estado individual con varias condiciones (como, por ejemplo, donde las condiciones son éxito, fracaso, redirección y tiempo agotado) o un grupo de estados finales (tal como en donde los estados son el éxito, falla, redirección y tiempo agotado).

El estado de progreso 330 puede indicar qué operaciones del comando de usuario se están llevando a cabo o se están intentado. La aplicación puede proporcionar información al usuario durante el estado de progreso 330 enviando una cadena texto a voz (TTS) o una cadena de interfaz gráfica de usuario (GUI) al asistente digital personal 120 de manera que la información pueda ser presentada al usuario utilizando la interfaz de usuario del asistente digital personal 120. Además, o alternativamente, información predeterminada (tal como una rueda que gira libremente, un reloj de arena, y/o un botón de cancelar) se puede presentar al usuario durante el estado de progreso 330 utilizando la interfaz de usuario del asistente digital personal 120.

Durante el estado de progreso 330, la aplicación puede supervisar el progreso de las operaciones y determinar si la aplicación puede permanecer en el estado de progreso 330 o pasar

al estado final 360. En una modalidad, la aplicación puede iniciar un temporizador (como por cinco segundos) y si la aplicación no hace un progreso suficiente antes de que expire el temporizador, el estado puede pasarse al estado final 360, como un estado de tiempo agotado. Si la aplicación está haciendo un progreso suficiente, el temporizador puede reiniciarse y el progreso puede ser examinado de nuevo en la siguiente expiración del temporizador. La aplicación puede tener un límite máximo de tiempo para permanecer en el estado de progreso 330 y si el límite máximo de tiempo se supera, el estado puede pasarse al estado final 360, tal como el estado de tiempo agotado. Las operaciones asociadas con el comando de usuario pueden completarse (con éxito o sin éxito) y el estado puede pasarse al correspondiente estado final 360. El usuario puede cerrar la aplicación cuando está en el estado de progreso 330 dando un comando en la interfaz de usuario del asistente digital personal 120. Por ejemplo, el usuario puede pulsar o hacer clic en un botón de "cancelar" o "regresar" en una pantalla o decir "cancelar". La cancelación del comando puede causar que el asistente digital personal 120 detenga la aplicación y muestre una pantalla de inicio del asistente digital personal 120 o salga.

El estado de confirmación 340 puede indicar que la aplicación está esperando la confirmación del usuario antes de completar una tarea. Cuando el asistente digital personal 120 detecta que la aplicación se encuentra en el estado de confirmación 340, un mensaje de respuesta sí/no puede ser presentado al usuario

utilizando la interfaz de usuario del asistente digital personal 120. La aplicación puede proporcionar al asistente digital personal 120 una cadena TTS, que es una cuestión que tiene una respuesta de sí o no. El asistente digital personal 120 puede hablar con la cadena TTS proporcionada a la aplicación y puede escuchar una respuesta de "Sí\No". Si la respuesta del usuario no se resuelve en una respuesta de sí o no, el asistente digital personal 120 puede continuar preguntando al usuario la cuestión hasta un número predefinido de veces (tal como tres veces). Si todos los intentos han sido agotados, el asistente digital personal 120 puede decir una frase predeterminada, como, por ejemplo, "lo siento, no entiendo. Toque abajo para seleccionar una respuesta" y el asistente digital personal 120 puede dejar de escuchar. Si el usuario toca sí o no, el asistente digital personal 120 puede enviar la elección del usuario a la aplicación. Si el usuario toca un icono de micrófono, el asistente digital personal 120 puede volver a intentar reconocer una respuesta hablada (tal como restableciendo un contador que cuenta el número de intentos de respuesta verbal). El asistente digital personal 120 puede hacer un bucle hasta que hay una coincidencia o el usuario cancele o golpee el botón de regreso en la pantalla. Si la aplicación recibe una respuesta afirmativa del asistente digital personal 120, la aplicación puede intentar completar la tarea. Si la tarea se completa correctamente, el estado puede realizar la transición al estado final 360 con una condición de éxito. Si la tarea no se completa correctamente o la solicitud se cancela, el estado puede realizar la

transición al estado final 360 con una condición de falla. Si la tarea va a tomar más de un predeterminado período de tiempo, el estado puede pasarse al estado de progreso 330 mientras la tarea está siendo llevada a cabo.

5 El estado de desambiguación 350 puede indicar que la aplicación está esperando a que el usuario aclare entre un número limitado (tal como diez o menos) de opciones antes de completar una tarea. La aplicación puede proporcionar al asistente digital personal 120 una cadena TTS, una cadena GUI y/o una lista de artículos que
10 el usuario puede elegir. La lista de artículos se proporciona como una plantilla con una o más piezas de información para proporcionar al usuario cada artículo, tal como un título, una descripción y/o un icono. El asistente digital personal 120 puede presentar la lista de artículos al usuario utilizando la información proporcionada por la
15 aplicación. El asistente digital personal 120 puede enviar y escuchar una selección del usuario. El usuario puede seleccionar de la lista utilizando selección flexible o no flexible. La selección no flexible significa que el usuario sólo puede seleccionar de la lista en una forma mientras que la selección flexible significa que el usuario
20 puede seleccionar de la lista en varias formas diferentes. Por ejemplo, el usuario puede seleccionar de la lista basándose en el orden numérico en el que los artículos se enlistan, tal como al decir "primero" o "segundo" para seleccionar el primer artículo o el segundo artículo, respectivamente. Como otro ejemplo, el usuario
25 puede seleccionar de la lista basándose en relaciones espaciales

entre los artículos tales como "el de abajo", "el de arriba", "el de la derecha", o "el segundo desde abajo." Como otro ejemplo, el usuario puede seleccionar de la lista diciendo que el título del artículo.

Como un ejemplo específico de desambiguación, el usuario
5 puede decirle al asistente digital personal 120, "Aplicación de película, agregar película-X a mi cola." Sin embargo, puede haber tres versiones de la película-X, como la original y dos continuaciones: Película-X I, película X-II, y la película X-III. En respuesta al comando hablado, el asistente digital personal 120
10 puede lanzar la Aplicación de película en el fondo con el comando para agregar película-X a la cola. La Aplicación de película puede buscar la película-X y determinar que hay tres versiones. Por lo tanto, la Aplicación de película puede pasar al estado de desambiguación 350 y enviar las tres opciones alternativas al
15 asistente digital personal 120. El asistente digital personal 120, a través de su interfaz de usuario, puede presentar al usuario las tres opciones y el usuario puede seleccionar una de la lista. Cuando se hace una adecuada selección por el usuario, el asistente digital personal 120 puede enviar la respuesta a la Aplicación de película y
20 la película correcta puede ser agregada a la cola.

Si la respuesta del usuario no se puede resolverse a un artículo en la lista, el asistente digital personal 120 puede seguir preguntando al usuario la pregunta hasta un número predefinido de veces. Si todos los intentos se han agotado, el asistente digital
25 personal 120 puede decir una frase predeterminada, como, por

ejemplo, "lo siento, no entiendo. Toque abajo para seleccionar una respuesta" y el asistente digital personal 120 puede dejar de escuchar. Si el usuario toca uno de los artículos de la lista mostrada, el asistente digital personal 120 puede enviar la elección del usuario a la aplicación. Si el usuario toca un icono de micrófono, el asistente digital personal 120 puede volver a intentar reconocer una respuesta hablada (tal como restableciendo un contador que cuenta el número de intentos de respuesta verbal). El asistente digital personal 120 puede hacer un bucle hasta que haya una coincidencia o el usuario cancele o golpee el botón de regresar en la pantalla. Si la aplicación recibe una respuesta válida del asistente digital personal 120, la aplicación puede intentar completar la tarea. Si la tarea requiere la confirmación del usuario antes de tomar acción, el estado puede realizar la transición al estado de confirmación 340. Si la tarea se completa exitosamente, el estado puede realizar la transición al estado final 360 con una condición de éxito. Si la tarea no se completa exitosamente o la aplicación se cancela, el estado puede realizar la transición al estado final 360 con una condición de falla. Si la tarea va a tomar más de un período predeterminado de tiempo para completarse, el estado puede pasarse al estado de progreso 330 mientras la tarea está siendo llevada a cabo.

Debe entenderse que la máquina de estado ilustrativa 300 puede ampliarse con estados adicionales o alternativos para permitir varias conversaciones multi-vueltas entre el usuario y una aplicación. La desambiguación (a través del estado de desambiguación 350) y la

confirmación (a través del estado de confirmación 340) son ejemplos específicos de conversación multi-vueltas. Por lo general, en una conversación multi-vueltas, una aplicación sin cabeza puede solicitar información adicional del usuario sin activación de su interfaz de usuario. Por el contrario, la información puede ser obtenida del usuario por el asistente digital personal 120 en nombre de la aplicación. Así, el asistente digital personal 120 puede actuar como un conducto entre el usuario y la aplicación.

El estado final 360 puede indicar a la aplicación ha completado con éxito la tarea, ha fallado para completar la tarea, se ha agotado el tiempo o está sugiriendo que la aplicación debe ser lanzada en primer plano (redirección). Como se ha descrito anteriormente, el estado final 360 puede ser un estado individual con varias condiciones (por ejemplo, éxito, falla, redirección y tiempo agotado) o un grupo de estados finales (por ejemplo, éxito, falla, redirección y tiempo agotado). La aplicación puede proporcionar al asistente digital personal 120 con una cadena TTS, una cadena GUI, una lista de artículos (provisto a través de una plantilla), y/o un parámetro de lanzamiento. El asistente digital personal 120 puede presentar la información proporcionada por la aplicación al usuario utilizando la interfaz de usuario del asistente digital personal 120. Además, o alternativamente, el asistente digital personal 120 puede presentar respuestas previamente definidas o conservadas asociadas con las diferentes condiciones. Por ejemplo, si se produce un tiempo agotado o la tarea falla, el asistente digital personal 120 puede decir

"¡Lo siento! No pude conseguir eso para usted. ¿Por favor puede que vuelva a intentarlo más tarde?" Como otro ejemplo, si la aplicación está solicitando la redirección, el asistente digital personal 120 puede decir "lo siento. <Nombreapl> no está respondiendo. Inicie
 5 <Nombreapl>" y el asistente digital personal 120 puede intentar iniciar la aplicación en primer plano con el comando de voz original y el parámetro de inicio (si un parámetro de lanzamiento es proporcionado por la aplicación). Como otro ejemplo, si la aplicación completa la tarea exitosamente, el asistente digital personal 120
 10 puede decir "he hecho eso para usted".

Definición de Comando Ilustrativa

La Figura 4 es un ejemplo de una definición de comando 400 que se conforma a un esquema que puede utilizarse para crear una
 15 estructura de datos, tal como la estructura de datos de comando 140, para permitir una interfaz entre una aplicación de tercera parte y el asistente digital personal 120. La definición de comando 400 puede ser escrita en varios lenguajes, tal como el Lenguaje de marcación extensible (XML) o un subconjunto de XML que está definido por un
 20 esquema. Por ejemplo, el esquema puede definir la estructura de la definición de comando, tal como los elementos legales, la jerarquía de elementos, atributos legales y opcionales para cada elemento y otros criterios adecuados. La definición de comando 400 puede ser utilizada por el asistente digital personal 120 para ayudar con el
 25 análisis de una pronunciación de usuario en diferentes componentes

tales como una aplicación, un comando o una tarea, y un artículo de datos o ranura, en donde el artículo de datos es opcional. Por ejemplo, el comando "Servicio de aplicación de película, agregar película X a mi cola" se puede analizar en una aplicación ("Servicio de Aplicación de Película"), un comando ("Agregar"), y un artículo de datos ("PelículaX"). La definición de comando 400 puede incluir elementos para definir un nombre de aplicación, tareas o comandos de la aplicación, fraseo alternativo para el procesamiento de lenguaje natural y respuestas asociadas con los diferentes estados de aplicación.

Una o más aplicaciones pueden definirse en la definición de comando 400. Las aplicaciones pueden ser aplicaciones de tercera parte u otras aplicaciones que son instaladas en el dispositivo de cómputo o servicios web. La información relacionada con la aplicación puede ser delimitada por un elemento que define la aplicación. Por ejemplo, el nombre de la aplicación puede ser definido por un elemento <Nombreapl> y los elementos entre los elementos <Nombreapl> pueden estar asociados con el elemento principal <Nombreapl>. En la definición de comando 400, el nombre de la aplicación es "Servicio de Aplicación de Película" y los elementos que siguen al elemento <Nombreapl> están asociados con la aplicación "Servicio de Aplicación de Película".

Los comandos que siguen al nombre de aplicación son los comandos de la aplicación. Los comandos pueden ser identificados con un elemento <Comando>. Los atributos del elemento de

comando pueden incluir un nombre (por ejemplo, "Nombre") del comando y un tipo de activación (por ejemplo, "TipoActivación") del comando. Por ejemplo, el tipo de activación puede ser "en primer plano" para los comandos que se van a lanzar en el primer plano y el
 5 tipo de activación puede ser "de fondo o segundo plano" para los comandos que se van a lanzar en el segundo plano. El atributo "TipoActivación" puede ser opcional, con un tipo de activación predeterminado estando en primer plano.

El elemento <Escuchar> puede estar anidado dentro del
 10 elemento <Comando> y se puede utilizar para definir una o más formas en las que el comando puede ser hablado. Pueden proporcionarse palabras opcionales o portadoras como indicios al asistente digital personal 120 cuando realiza el procesamiento de lenguaje natural. Las palabras portadoras pueden ser identificadas
 15 dentro de corchetes: []. Los artículos de datos pueden ser identificados dentro de corchetes o llaves: { }. En la definición de comando 400, generalmente existen dos formas alternativas para llamar al comando "Agregar" según lo definido por los dos elementos <Escuchar>. Por ejemplo, diciendo ya sea "agregar películaX a mi
 20 cola" o "agregar películaX a mi cola Servicio de aplicación de película" puede ser usado para que el asistente digital personal 120 lance el comando "Agregar" del Servicio de aplicación de película en el fondo o segundo plano. Cabe señalar que frases predefinidas puede ser identificada con la palabra clave "incorporar:" dentro de un
 25 conjunto de llaves: {incorporar:<identificador de frase>}.

El elemento <realimentación> puede estar anidado dentro del elemento <Comando> y se puede utilizar para definir una frase que se diga al usuario cuando el asistente digital personal 120 ha reconocido exitosamente un comando hablado desde el usuario.

- 5 Además, o alternativamente, el elemento <realimentación> puede definir una cadena de texto que se mostrará al usuario a medida que el comando hablado está siendo analizado por el asistente digital personal 120.

- El elemento <respuesta> puede estar anidado dentro del
 10 elemento <Comando> y se puede utilizar para definir una o más respuestas proporcionadas por el asistente digital personal 120 al usuario. Cada respuesta está asociada con un estado de la aplicación que se define como un atributo de "Estado". Los estados pueden ser para estados finales, tales como éxito y falla, o para
 15 estados intermediarios, tales como progreso. Puede haber varios tipos de respuestas definidas, tal como <PresentarCadena> para mostrar texto en una pantalla, <CadenaTTS> para el texto que se hablará al usuario, <EnlaceProfundoAplicación> para un enlace profundo a un sitio web, y <EnlaceWeb> para un enlace menos
 20 profundo a un sitio web, por ejemplo. Las respuestas definidas por el elemento <respuesta> pueden ser aumentadas con información de respuesta adicional proporcionada por la aplicación.

Diagrama de Secuencia Ilustrativo

- 25 La Figura 5 es un diagrama de secuencia ilustrativo 500 que

ilustra la comunicación de múltiples subprocesos de ejecución (510, 520 y 530) para realizar, sin cabeza, una función de una aplicación de tercera parte dentro del asistente digital personal 120. El subproceso de interfaz de usuario 510 y el subproceso de control 520 puede ser subprocesos paralelos de una modalidad de múltiples subprocesos del asistente digital personal 120. El subproceso de interfaz de usuario 510 puede ser primordialmente responsable de capturar de entrada de y mostrar la salida a la interfaz de usuario del asistente digital personal 120. Por ejemplo, la entrada de voz, entrada táctil y/o entrada de texto pueden ser capturadas por el subproceso de interfaz de usuario 510. En una modalidad, el subproceso de interfaz de usuario 510 puede realizar procesamiento de lenguaje natural en la entrada y puede hacer coincidir los comandos hablados de usuario con comandos en la estructura de datos de comando 140. Cuando se determina que el comando hablado coincide con un comando en la estructura de datos de comando 140, el comando puede ser comunicado al subproceso de control 520 para procesamiento adicional. En una modalidad alternativa, el subproceso de interfaz de usuario 510 puede capturar la voz en texto de entrada y palabras individuales pueden ser comunicadas al subproceso de control 520 que puede realizar el procesamiento de lenguaje natural en la entrada y puede hacer coincidir los comandos hablados del usuario con comandos de la estructura de datos de comando 140.

El subproceso de control 520 puede ser principalmente

responsable de la comunicación y el seguimiento del progreso de la aplicación e interconexión con el subproceso de interfaz de usuario 510. Por ejemplo, el subproceso de control 520 puede ser notificado por el subproceso de interfaz de usuario 510 que el usuario ha
5 hablado con la interfaz de usuario del asistente digital personal 120. Palabras o comandos pueden ser recibidas por el subproceso de control 520 y el subproceso de control 520 puede notificar al subproceso de interfaz de usuario 510 cuando un comando de usuario ha sido reconocido por el subproceso de control 520. El
10 subproceso de interfaz de usuario 510 puede indicar al usuario, a través de la interfaz de usuario del asistente digital personal 120, que el progreso se está haciendo en el comando. El subproceso de interfaz de usuario 510 o el subproceso de control 520 puede determinar que el comando se lanzará sin cabeza, al recuperar los
15 atributos del comando desde la estructura de datos de comando 140. El subproceso de control 520 puede iniciar un nuevo subproceso o comunicarse con un subproceso existente, tal como el subproceso ServicioAplicación 530, cuando el comando se va a lanzar sin cabeza. Para reducir el tiempo de respuesta para el usuario, puede
20 ser deseable que el subproceso ServicioAplicación 530 sea un subproceso existente, en lugar de que el subproceso de control 520 inicie un nuevo subproceso. Por ejemplo, el subproceso ServicioAplicación 530 se puede iniciar cuando se activa la aplicación o durante el arranque del dispositivo de cómputo 130.

25 El subproceso ServicioAplicación 530 se puede ejecutar en el

dispositivo de cómputo 130 o se puede ejecutar en un servidor remoto, tal como la computadora de servidor remoto 160. El subproceso ServicioAplicación 530 puede ser principalmente responsable de completar la función especificada por el comando de usuario. El subproceso ServicioAplicación 530 puede mantener una máquina de estado (tal como la máquina de estado 300) para rastrear el progreso de la ejecución de la función y puede proporcionar actualizaciones sobre el estado del subproceso de control 520. Proporcionando actualizaciones de estado al subproceso de control 520, el subproceso ServicioAplicación 530 puede ser sin cabeza, en donde la salida al usuario se proporciona por el asistente digital personal 120 y no de una interfaz de usuario del subproceso ServicioAplicación 530.

El subproceso de control 520 puede rastrear el progreso de la aplicación (por ejemplo, subproceso ServicioAp 530) al recibir actualizaciones de estado de la aplicación y comprobar si la aplicación está haciendo progresos. Por ejemplo, el subproceso de control 520 puede iniciar un temporizador de una duración predefinida (tal como cinco segundos) cada vez que se comunica con el subproceso ServicioAplicación 530 (ya sea enviando información al subproceso ServicioAplicación 530 o recibiendo información desde el subproceso ServicioAplicación 530). Si el temporizador expira antes de que el subproceso ServicioAplicación 530 responda, el subproceso de control 520 puede indicar al subproceso de interfaz de usuario 510 que la aplicación no responde y el subproceso de

interfaz de usuario 510 puede presentar un mensaje de falla al usuario a través de la interfaz de usuario del asistente digital personal 120. El subproceso ServicioAplicación 530 puede ser terminado o ignorado por el subproceso de control 520 después de
5 que el temporizador expira. Alternativamente, si el subproceso ServicioAplicación 530 responde antes de que el temporizador expire, el temporizador se puede restablecer si se espera otra respuesta de la aplicación (tal como cuando la aplicación responde con el estado de progreso), o el temporizador puede ser cancelado
10 (tal como cuando la aplicación haya completado la función (un estado final) o cuando una respuesta de usuario está siendo solicitada (un estado de confirmación o desambiguación)).

Cuando el subproceso de control 520 recibe un estado de confirmación o desambiguación del subproceso ServicioAplicación
15 530, el subproceso de control 520 puede indicar al subproceso de interfaz de usuario 510 que la confirmación o desambiguación es solicitada por el usuario. El subproceso de interfaz de usuario 510 puede presentar opciones de confirmación o desambiguación al usuario a través de la interfaz de usuario del asistente digital
20 personal 120. Cuando el usuario responde o no responde, el subproceso de interfaz de usuario 510 puede proporcionar la respuesta de usuario, o la falta de ella definitiva, al subproceso de control 520. El subproceso de control 520 puede pasar la respuesta del usuario al subproceso ServicioAplicación 530 de manera que el
25 subproceso ServicioAplicación 530 puede llevar a cabo la función. Si

el usuario no responde, el subproceso de control 520 puede terminar el subproceso ServicioAplicación 530.

El subproceso de interfaz de usuario 510 puede mostrar varios tipos de salida a través de la interfaz de usuario del asistente digital personal 120. Por ejemplo, el subproceso de interfaz de usuario 510
5 puede generar una salida de audio, tal como salida de voz simulada digital del texto. La voz simulada digital puede ser enviada a un chip de procesamiento de audio que puede convertir la voz simulada digital a una señal analógica (tal como con un convertidor de digital
10 a analógico) que puede ser salida como sonido a través de un altavoz o auriculares. Como otro ejemplo, el subproceso de IU 510 puede proporcionar salida visual, tal como imágenes, animación, salida de texto e hipervínculos para ser vistos por el usuario en una pantalla del dispositivo de cómputo 130. Si los hipervínculos se
15 tocan o se hace clic en ellos, el subproceso de interfaz de usuario 510 puede iniciar una aplicación de navegador para ver un sitio web correspondiente al hipervínculo seleccionado. Como otro ejemplo, el subproceso de interfaz de usuario 510 puede generar salida táctil, tal como mediante el envío de una señal de vibración a un motor
20 eléctrico que puede hacer que el dispositivo de cómputo 130 vibre.

Método Ilustrativo para Completar, Sin Cabeza, la Tarea

La Figura 6 es un diagrama de flujo de un método ilustrativo
600 para completar, sin cabeza, una tarea de una aplicación en el
25 fondo del asistente digital personal 120. En 610, una entrada de voz,

generada por un usuario, puede ser recibida por el asistente digital personal 120. La entrada de voz puede ser capturada localmente en el dispositivo de cómputo 130 o de forma remota desde el dispositivo de cómputo 130. Como un ejemplo, la entrada de voz generada por el usuario puede ser localmente capturada por un micrófono 150 del dispositivo de cómputo 130 y digitalizada por un convertidor analógico-a-digital. Como otro ejemplo, la voz de entrada generada por el usuario puede ser remotamente capturada por un micrófono (tal como un dispositivo complementario de blue-tooth) conectado de forma inalámbrica al dispositivo de cómputo 130. El asistente digital personal 120 puede ser controlado por voz y/o texto introducido en la interfaz de usuario del asistente digital personal 120.

En 620, el procesamiento de lenguaje natural de la entrada de voz puede realizarse para determinar un comando de voz de usuario. El comando de voz de usuario puede incluir una solicitud para realizar una función predefinida de una aplicación, como una aplicación activada por voz de tercera parte. La función predefinida puede ser identificada utilizando una estructura de datos que define aplicaciones y funciones de aplicaciones que son soportadas por el asistente digital personal 120. Por ejemplo, las aplicaciones compatibles pueden ser identificadas en un archivo de definición de comando, tal como la definición de comando 400. Al utilizar un archivo de definición de comando extensible para definir funciones de aplicaciones de tercera parte que pueden ser realizadas, sin cabeza, por el asistente digital personal 120, el asistente digital

personal 120 puede permitir al usuario realizar más tareas con la interfaz de usuario del asistente digital personal 120.

En 630, el asistente digital personal 120 puede hacer que la aplicación ejecute, sin cabeza, la función predefinida sin una interfaz de usuario de la aplicación apareciendo en la pantalla del dispositivo de cómputo 130. El asistente digital personal 120 puede determinar ejecutar la aplicación, sin cabeza, porque la aplicación está definida como sin cabeza en la estructura de datos de comando 140 o porque el usuario está utilizando el dispositivo de cómputo en modo manos libres y la ejecución de la aplicación en primer plano podría distraer potencialmente al usuario. Por ejemplo, el asistente digital personal 120 puede llamar a un servicio web para ejecutar la función predefinida de la aplicación. Como otro ejemplo, el asistente digital personal 120 puede comenzar un nuevo subproceso en el dispositivo de cómputo 130 para ejecutar la función predefinida de la aplicación después de que el comando de usuario se determina. Como otro ejemplo, el asistente digital personal 120 puede comunicarse con un subproceso existente, tal como un subproceso iniciado durante una activación de la aplicación, para ejecutar la función predefinida de la aplicación. La función predefinida puede ser ejecutada como un proceso de fondo o segundo plano. La aplicación puede supervisar el progreso de la función predefinida, tal como mediante el seguimiento del estado de la función predefinida.

En 640, se puede recibir una respuesta desde la aplicación que indica un estado asociado con la función de predefinida. Por

ejemplo, los estados pueden incluir estados de activación, iniciales, progreso, confirmación, desambiguación y finales. La respuesta puede incluir información adicional, tal como una lista en plantilla, una cadena de texto, una cadena de texto-a-habla, una imagen, un
5 hiperenlace u otro tipo adecuado de información que se pueda mostrar al usuario a través de la interfaz de usuario del asistente digital personal 120.

En 650, la interfaz de usuario del asistente digital personal 120 puede proporcionar una respuesta para el usuario basándose en el
10 estado recibido asociado con la función predefinida. De esta manera, la respuesta puede venir desde dentro de un contexto de la interfaz de usuario del asistente digital personal 120 sin la activación de la interfaz de usuario de la aplicación. Además, las capacidades de confirmación y desambiguación del asistente digital personal 120
15 pueden utilizarse para confirmar y/o aclarar un comando de usuario para la aplicación.

Método Ilustrativo para Determinar si Activar una Aplicación

La Figura 7 es un diagrama de flujo de un método ilustrativo
20 700 para determinar si se debe activar una aplicación mientras un usuario está hablando con el asistente digital personal 120. En 710, el usuario puede escribir, pronunciar, o hablar con el asistente digital personal 120. El texto o discurso del usuario puede ser analizado utilizando técnicas de procesamiento de lenguaje natural y se
25 pueden reconocer palabras individuales por la voz. Las palabras

individuales pueden analizarse por separado y dentro de la frase intermedia en donde se hablan. Por ejemplo, el usuario puede decir, "Oye, asistente, de MiAp, hace ..." La palabra "oye" puede ser una palabra portadora y cae. La palabra "asistente" puede usarse para
5 permitir que el asistente digital personal 120 sepa que el usuario está solicitando que realice una acción. La palabra "MiAp" puede interpretarse como una aplicación.

En 720, las palabras habladas o escritas pueden ser comparadas con las funciones nativas del asistente digital personal
10 120 y las funciones provistas en las definiciones de comando extensible. Colectivamente, las funciones nativas y las funciones definidas en el archivo de definición de comando pueden ser referidas como los "ServiciosAp conocidos". Las palabras pronunciadas pueden ser analizadas y comparadas con los
15 ServiciosAp conocidos a medida que las palabras son pronunciadas. En otras palabras, el análisis del habla puede ocurrir antes de la frase entera sea hablada o escrita por el usuario. Si ninguno de los Servicios Ap conocidos coincide, entonces en 730, el asistente digital personal 120 puede abrir un navegador web para recuperar
20 una página web de motor de búsqueda con una cadena de búsqueda correspondiente a la frase hablada no reconocida. El control del programa puede ser transferido al navegador web de manera que el usuario pueda refinar la búsqueda en la web y/o ver los resultados. Sin embargo, si un servicioAp conocido coincide, entonces el método
25 700 puede continuar en 740.

En 740, se puede determinar si la aplicación ServicioAp es una tarea de primer plano o de segundo plano. Por ejemplo, la definición de comando puede incluir un atributo que define la aplicación ServicioAp como una aplicación de primer plano o de fondo o
5 segundo plano. Si la aplicación ServicioAp es una tarea de primer plano, en 750, la aplicación ServicioAp puede ejecutarse en primer plano y el control puede ser transferido a la aplicación ServicioAp para completar el comando. Si la aplicación ServicioAp es una tarea de fondo o de segundo plano, entonces el método 700 puede
10 continuar con los pasos paralelos 760 y 770.

En 760, el asistente digital personal 120 puede proporcionar al usuario información sobre el análisis del habla. Específicamente, el asistente digital personal 120 puede generar para una salida para una pantalla en progreso de la interfaz de usuario del asistente
15 digital personal 120. La salida puede ser definida en un elemento <Realimentación>, anidado dentro de un elemento <Comando>, de la definición de comando, por ejemplo. La salida puede ser una cadena de texto y puede ser actualizada continuamente mientras el usuario sigue hablando.

20 En 770, el asistente digital personal 120 puede activar la aplicación ServicioAp sin esperar que la pronunciación de usuario termine. La activación de la aplicación ServicioAp puede incluir asignación de memoria, pre-búsqueda de instrucciones, establecer una sesión de comunicación, recuperar información de una base de
25 datos, iniciar un nuevo hilo de ejecución, plantear una interrupción, u

otras operaciones específicas de aplicación adecuadas. La aplicación puede ser activada basándose en una función especulativa. Por ejemplo, las instrucciones correspondientes a la función especulativa pueden ser buscadas incluso si la función no se
5 conoce con certeza. Al activar la aplicación antes de que el usuario complete el comando hablado, el tiempo para responder al usuario puede ser disminuido potencialmente.

En 780, el asistente digital personal 120 puede continuar analizando el resultado de reconocimiento de voz parcial hasta que
10 la pronunciación se complete. El final de la pronunciación puede ser detectado basándose en el comando que está siendo analizado y/o basándose en una pausa del usuario durante más de una cantidad predeterminada de tiempo. Por ejemplo, el fin del comando, "ServicioApPelícula, agregar películaX a mi cola" puede ser
15 detectado cuando la palabra "cola" es reconocida. Como otro ejemplo, el fin del comando, "ApTexto, diga a mi esposa que voy a llegar tarde a casa para la cena", puede ser más difícil de detectar porque el comando finaliza con un artículo de datos de longitud desconocida. Por lo tanto, se puede usar una pausa para indicar al
20 asistente digital personal 120 que el comando se completa.

En el 790, el final del comando hablado puede ser detectado y el último resultado de reconocimiento de voz puede ser pasado a la aplicación. La aplicación y el asistente digital personal 120 pueden comunicarse entre sí para completar el comando hablado descrito
25 con referencia a las Figuras anteriores.

Sistemas de Cómputo

La Figura 8 muestra un ejemplo generalizado de un sistema de cómputo 800 adecuado en donde las innovaciones descritas pueden ser implementadas. El sistema de cómputo 800 no pretende sugerir
5 ninguna limitación en cuanto al alcance de uso o la funcionalidad, ya que las innovaciones pueden ser implementadas en diversos sistemas de cómputo de propósito general o propósito especial.

Con referencia a la Figura 8, el sistema de cómputo 800 incluye una o más unidades de procesamiento 810, 815 y memoria
10 820, 825. En la Figura 8, esta configuración básica 830 está incluida dentro de una línea discontinua. Las unidades de procesamiento 810, 815 ejecuta instrucciones ejecutables por computadora. Una unidad de procesamiento puede ser una unidad de procesamiento central de propósito general (CPU), procesador en un circuito integrado de
15 aplicación específica (ASIC) o cualquier otro tipo de procesador. En un sistema de procesamiento múltiple, múltiples unidades de procesamiento ejecutan instrucciones ejecutables por computadora para aumentar la potencia de procesamiento. Por ejemplo, la Figura 8 muestra una unidad de procesamiento central 810 así como una
20 unidad de procesamiento de gráficos o unidad de co-procesamiento 815. La memoria tangible 820, 825 puede ser memoria volátil (por ejemplo, registros, memoria caché, RAM), memoria no volátil (por ejemplo, ROM, EEPROM, memoria flash, etc.), o alguna combinación de las dos, accesible a través de la(s) unidad(es) de procesamiento.
25 La memoria 820, 825 almacena *software* 880 que implementa una o

más innovaciones aquí descritas, en la forma de instrucciones ejecutables por computadora adecuadas para la ejecución por la(s) unidad(es) de procesamiento.

Un sistema de cómputo puede tener funciones adicionales. Por ejemplo, el sistema de cómputo 800 incluye el almacenamiento 840, uno o más dispositivos de entrada 850, uno o más dispositivos de salida 860 y una o más conexiones de comunicación 870. Un mecanismo de interconexión (no mostrado) tal como un conductor común, controlador o red que interconecta los componentes del sistema de cómputo 800. Normalmente, el *software* del sistema operativo (no mostrado) proporciona un entorno operativo para otro *software* que ejecuta en el sistema de cómputo 800, y coordina las actividades de los componentes del sistema de cómputo 800.

El almacenamiento tangible 840 puede ser removible o no removible, e incluye discos magnéticos, cintas magnéticas o casetes, CD-ROMs, DVDs, o cualquier otro medio que pueda ser utilizado para almacenar información y que puede ser accedido dentro del sistema de cómputo 800. El almacenamiento 840 almacena las instrucciones para el *software* 880 implementando una o más innovaciones aquí descritas.

El dispositivo(s) de entrada 850 puede ser un dispositivo de entrada táctil, tal como un teclado, un ratón, un lápiz o seguidor, un dispositivo de entrada de voz, un dispositivo de escaneo, u otro dispositivo que proporciona una entrada al sistema de cómputo 800. Para la codificación de vídeo, el dispositivo(s) de entrada 850 puede

ser una cámara, tarjeta de vídeo, tarjeta sintonizadora de TV o un dispositivo similar que acepta la entrada de vídeo en forma analógica o digital, o un CD-ROM o CD-RW que lee muestras de vídeo en el sistema de cómputo 800. El dispositivo(s) de salida 860 puede ser
5 una pantalla, impresora, altavoces, grabadora de CD, u otro dispositivo que proporciona la salida del sistema de cómputo 800.

La conexión(es) de comunicación 870 permite la comunicación a través de un medio de comunicación a otra entidad de cómputo. El medio de comunicación transmite información tal como instrucciones
10 ejecutables por computadora, entrada o de salida de audio o vídeo, o cualquier otro tipo de datos en una señal modulada de datos. Una señal modulada de datos es una señal de que tiene una o más de sus características fijadas o cambiadas de tal manera que codifica la información en la señal. A modo de ejemplo, y sin limitación, los
15 medios de comunicación pueden usar un portador eléctrico, óptico, RF u otro portador.

Las innovaciones pueden ser descritas en el contexto general de instrucciones ejecutables por computadora, tales como aquellas incluidas en los módulos de programa, que se ejecutan en un
20 sistema de cómputo en un objetivo real o procesador virtual. Generalmente, los módulos de programa incluyen rutinas, programas, bibliotecas, objetos, clases, componentes, estructuras de datos, etc. que realizan determinadas tareas o implementan determinados tipos de datos abstractos. La funcionalidad de los módulos de programa se
25 puede combinar o dividir entre los módulos de programa como desee

en varias modalidades. Las instrucciones ejecutables por computadora para los módulos de programa pueden ser ejecutadas dentro de un sistema de cómputo local o distribuido.

Los términos "sistema" y "dispositivo" se utilizan
5 indistintamente en este documento. A menos que el contexto indique claramente lo contrario, ningún término implica ninguna limitación en un tipo de sistema de cómputo o dispositivo de cómputo. En general, un sistema de cómputo o dispositivo de cómputo puede ser local o distribuido y puede incluir cualquier combinación de *hardware* de
10 propósito especial y/o *hardware* de propósito general con software que implementa la funcionalidad aquí descrita.

En aras de la presentación, la descripción detallada utiliza términos como "determinar" y "usar" para describir operaciones de computadora en un sistema de cómputo. Estos términos son
15 abstracciones de alto nivel para operaciones realizadas por una computadora y no deben confundirse con los actos realizados por un ser humano. Las operaciones de computadora reales que corresponden a estos términos varían dependiendo de la implementación.

20

Dispositivo Móvil

La Figura 9 es un diagrama de sistema que muestra un dispositivo móvil ilustrativo 900 que incluye una variedad de componentes de *hardware* y *software* opcionales, mostrado
25 generalmente en 902. Los componentes 902 en el dispositivo móvil

se pueden comunicar con cualquier otro componente, aunque no todas las conexiones se muestran, para facilidad de la ilustración. El dispositivo móvil puede ser cualquiera de una variedad de dispositivos de cómputo (por ejemplo, teléfono móvil, teléfono
5 inteligente, computadora portátil, Asistente digital personal (PDA), etc.) y puede permitir comunicaciones de dos vías inalámbricas con una o más redes de comunicaciones móviles 904, tal como un celular, satélite u otra red.

El dispositivo móvil ilustrado 900 puede incluir un controlador o
10 un procesador 910 (por ejemplo, procesador de señal, microprocesador, ASIC o cualquier otro sistema de circuito lógico de control y procesamiento) para realizar tareas tales como la codificación de señales, procesamiento de datos, procesamiento de entrada/salida, control de la alimentación y/u otras funciones. Un
15 sistema operativo 912 puede controlar la asignación y uso de los componentes 902 y soporte para el asistente digital personal 120 y uno o más programas de aplicación 914. Los programas de aplicación pueden incluir aplicaciones de cómputo móviles comunes (por ejemplo, aplicaciones de correo electrónico, calendarios,
20 administradores de contactos, navegadores web, aplicaciones de mensajería, aplicaciones para películas, aplicaciones de banca), o cualquier otra aplicación de cómputo. Los programas de aplicación 914 pueden incluir aplicaciones que tienen tareas que pueden ser ejecutadas, sin cabeza, por el asistente digital personal 120. Por
25 ejemplo, las tareas pueden ser definidas en la estructura de datos de

comando 140. La funcionalidad 913 para acceder a un almacenamiento de aplicaciones también puede ser utilizada para la adquisición y actualización de programas de aplicación 914.

El dispositivo móvil ilustrado 900 puede incluir la memoria 920.

5 La memoria 920 puede incluir memoria no removible 922 y/o memoria removible 924. La memoria no removible 922 puede incluir RAM, ROM, memoria flash, un disco duro u otras tecnologías de almacenamiento en memoria bien conocidas. La memoria removible 924 puede incluir memoria flash o una tarjeta de Módulo de Identidad de Suscriptor (SIM), que es bien conocida en sistemas de comunicación GSM u otras tecnologías de almacenamiento de memoria bien conocidas, tales como "tarjetas inteligentes". La memoria 920 puede utilizarse para almacenar datos y/o código para que se ejecute el sistema operativo 912 y las aplicaciones 914.

10 Datos ilustrativos pueden incluir páginas web, texto, imágenes, archivos de sonido, datos de vídeo u otros conjuntos de datos que serán enviados y/o recibidos de uno o más servidores de red u otros dispositivos a través de una o más redes de cable o inalámbricas. La memoria de 920 puede ser utilizada para almacenar un identificador de suscriptor, tal como una Identidad Internacional de Suscriptor Móvil (IMSI) y un identificador de equipo, tal como un Identificador Internacional de Equipos Móviles (IMEI). Estos identificadores pueden ser transmitidos a un servidor de red para identificar a los usuarios y equipos.

25 El dispositivo móvil 900 puede soportar uno o más dispositivos

de entrada 930, tales como una pantalla táctil 932, micrófono 934, cámara 936, teclado físico 938 y/o seguidor 940 y uno o más dispositivos de salida 950, como, por ejemplo, un altavoz 952 y una pantalla 954. Otros posibles dispositivos de salida (no mostrados) pueden incluir dispositivos piezoeléctricos u otros dispositivos de salida hápticos. Algunos dispositivos pueden servir a más de una función de entrada/salida. Por ejemplo, la pantalla táctil 932 y la pantalla 954 pueden ser combinadas en un solo dispositivo de entrada/salida.

Los dispositivos de entrada 930 pueden incluir una Interfaz de Usuario Natural (NUI). Una NUI es cualquier tecnología de interfaz que permite a un usuario interactuar con un dispositivo en forma "natural", libre de limitaciones artificiales impuestas por los dispositivos de entrada, tales como ratones, teclados, controles remotos y similares. Ejemplos de métodos NUI incluyen aquellos que se basan en el reconocimiento de voz, reconocimiento de tacto y lápiz, reconocimiento de gestos tanto en la pantalla como adyacente a la pantalla, gestos de aire, seguimiento de cabeza y ojos, voz y habla, visión, tacto, gestos e inteligencia artificial. Otros ejemplos de una NUI incluyen detección de gesto de movimiento utilizando acelerómetros/giroscopios, reconocimiento facial, pantallas 3D, rastreo de cabeza, ojo y mirada, realidad aumentada inmersiva y sistemas de realidad virtual, todos los cuales proporcionan una interfaz más natural, así como tecnologías para la detección de la actividad cerebral mediante electrodos sensores de campo eléctrico

(EEG y métodos relacionados). Así, en un ejemplo específico, el sistema operativo 912 o aplicaciones 914 pueden comprender *software* de reconocimiento de voz como parte de una interfaz de usuario de voz que permite al usuario operar el dispositivo 900 a través de comandos de voz. Además, el dispositivo 900 puede comprender dispositivos de entrada y *software* que permite la interacción de usuario a través de un gesto espacial del usuario, tal como gestos de detección e interpretación para proporcionar entrada a una aplicación de juegos.

Un módem inalámbrico 960 puede ser acoplado a una antena (no mostrada) y puede soportar comunicaciones bidireccionales entre el procesador 910 y los dispositivos externos, como es bien conocido en la técnica. El módem 960 se muestra de forma genérica y puede incluir un módem celular para comunicarse con la red de comunicación móvil 904 y/u otros módems basados en radio (por ejemplo, Bluetooth 964 o Wi-Fi 962). El módem inalámbrico 960 normalmente está configurado para la comunicación con una o más redes celulares, tal como una red GSM para comunicaciones de voz y datos en una sola red celular, entre redes celulares, o entre el dispositivo móvil y una red telefónica pública conmutada (PSTN).

El dispositivo móvil puede incluir además al menos un puerto de entrada/salida 980, un suministro de energía 982, un receptor de sistema de navegación por satélite 984, tal como un receptor de Sistema de Posicionamiento Global (GPS), un acelerómetro 986 y/o un conector físico 990, que puede ser un puerto USB, puerto IEEE

1394 (FireWire), y/o puerto RS-232. Los componentes ilustrados 902 no son necesarios o inclusivos, ya que los componentes pueden ser eliminados y otros componentes pueden ser agregados.

5 Entorno Soportado por Nube

La Figura 10 ilustra un ejemplo generalizado de un entorno soportado por nube adecuado 1000 en donde pueden ser implementadas las modalidades, técnicas y tecnologías descritas. En el entorno ilustrativo 1000, diversos tipos de servicios (por ejemplo, servicios de cómputo) son proporcionados por una nube 1010. Por ejemplo, la nube 1010 puede comprender una colección de dispositivos de cómputo, los cuales pueden ser ubicados de forma central o distribuida, que proporcionan servicios basados en la nube para diferentes tipos de usuarios y dispositivos conectados a través de una red, tal como Internet. El entorno de implementación 1000 puede ser utilizado de diferentes maneras para realizar tareas de cómputo. Por ejemplo, algunas tareas (por ejemplo, el procesamiento de la entrada del usuario y presentar una interfaz de usuario) pueden realizarse en los dispositivos de cómputo locales (por ejemplo, dispositivos conectados 1030, 1040, 1050), mientras que otras tareas (por ejemplo, almacenamiento de datos que serán utilizados en procesamiento subsecuente) pueden realizarse en la nube 1010.

En el entorno ilustrativo 1000, la nube 1010 proporciona servicios para dispositivos conectados 1030, 1040, 1050 con una variedad de capacidades de pantalla. El dispositivo conectado 1030

representa un dispositivo con una pantalla de computadora 1035 (por ejemplo, una pantalla de tamaño mediano). Por ejemplo, el dispositivo conectado 1030 podría ser un computadora personal tal como computadora de escritorio, laptop, portátil, netbook, o
5 similares. El dispositivo conectado 1040 representa un dispositivo con una pantalla de dispositivo móvil 1045 (por ejemplo, una pantalla de pequeño tamaño). Por ejemplo, el dispositivo conectado 1040 podría ser un teléfono móvil, un teléfono inteligente, un asistente digital personal, computadora de tableta y similares. El dispositivo
10 conectado 1050 representa un dispositivo con una pantalla grande 1055. Por ejemplo, el dispositivo conectado 1050 podría ser una pantalla de televisión (por ejemplo, una televisión inteligente) u otro dispositivo conectado a un televisor (por ejemplo, una caja de tv por cable o consola de juegos) o similar. Uno o más de los dispositivos
15 conectados 1030, 1040, 1050 puede incluir capacidades de pantalla táctil. Las pantallas táctiles pueden aceptar la entrada de diferentes maneras. Por ejemplo, las pantallas táctiles capacitivas detectan la entrada táctil cuando un objeto (por ejemplo, un dedo o el lápiz óptico) distorsiona o interrumpe una corriente eléctrica que corre a lo
20 largo de la superficie. Como otro ejemplo, las pantallas táctiles pueden utilizar sensores ópticos para detectar la entrada táctil cuando los haces de los sensores ópticos son interrumpidos. El contacto físico con la superficie de la pantalla no es necesario para que la entrada sea detectada por algunas pantallas táctiles. También
25 pueden ser utilizados dispositivos sin capacidades de pantalla en el

entorno ilustrativo 1000. Por ejemplo, la nube 1010 puede proporcionar servicios para una o más computadoras (por ejemplo, computadoras de servidor) sin pantallas.

Se puede proporcionar servicios por la nube 1010 a través de
5 proveedores de servicio 1020, o a través de otros proveedores de servicios en línea (no representados). Por ejemplo, los servicios en la nube pueden personalizarse al tamaño de la pantalla, la capacidad de visualización y/o la capacidad de pantalla táctil de un determinado dispositivo conectado (por ejemplo, dispositivos conectados 1030,
10 1040, 1050).

En el entorno ilustrativo 1000, la nube 1010 proporciona las tecnologías y las soluciones aquí descritas a los diversos dispositivos conectados 1030, 1040, 1050 utilizando, al menos en parte, los proveedores de servicios 1020. Por ejemplo, los
15 proveedores de servicio 1020 pueden proporcionar una solución centralizada para diversos servicios basados en nube. Los proveedores de servicio 1020 pueden manejar suscripciones de servicio para los usuarios y/o dispositivos (por ejemplo, los dispositivos conectados 1030, 1040, 1050 y/o sus respectivos
20 usuarios).

Implementaciones Ilustrativas

Aunque las operaciones de algunos de los métodos presentados se describen en un determinado orden secuencial, para
25 una presentación conveniente, debe entenderse que esta forma de

descripción abarca la reorganización, a menos que un determinado orden sea requerido por un lenguaje específico que se establece a continuación. Por ejemplo, las operaciones descritas secuencialmente en algunos casos pueden ser reorganizadas o realizadas concurrentemente. Además, en aras de la simplicidad, las figuras adjuntas no pueden mostrar las diversas formas en las que los métodos descritos pueden utilizarse junto con otros métodos.

Cualquiera de los métodos descritos puede ser implementado como instrucciones ejecutables por computadora o un producto de programa de computadora almacenado en uno o más medios de almacenamiento legibles por computadora y se ejecutan en un dispositivo de cómputo (por ejemplo, cualquier dispositivo de cómputo, incluyendo teléfonos inteligentes y otros dispositivos móviles que incluyen *hardware* computacional). Los medios de almacenamiento legibles por computadora son cualquier medio tangible que se puede acceder dentro de un entorno de cómputo (por ejemplo, uno o más discos de medios ópticos tales como CD o DVD, componentes de memoria volátil (tal como SRAM o DRAM), o componentes de memoria no volátil (tal como unidades de disco duro o memoria flash)). A modo de ejemplo y con referencia a la Figura 8, los medios de almacenamiento legibles por computadora incluyen memoria 820 y 825 y almacenamiento 840. A modo de ejemplo y con referencia a la Figura 9, los medios de almacenamiento legibles por computadora incluyen memoria y almacenamiento 920, 922, y 924. El término medios de almacenamiento legibles por computadora no

incluye señales y ondas portadoras. Además, el término medios de almacenamiento legibles por computadora no incluye conexiones de comunicación (por ejemplo, 870, 960, 962 y 964).

Cualquiera de las instrucciones ejecutables por computadora
5 para implementar las técnicas descritas, así como los datos creados y utilizados durante la implementación de las modalidades descritas pueden ser almacenadas en uno o más medios de almacenamiento legibles por computadora. Las instrucciones ejecutables por computadora pueden ser parte de, por ejemplo, una aplicación de
10 *software* dedicada o una aplicación de *software* que se accede o descargar a través de un navegador web u otra aplicación de *software* (tal como una aplicación de cómputo remota). Este *software* puede ser ejecutado, por ejemplo, en una computadora local individual (por ejemplo, cualquier computadora comercialmente
15 disponible adecuada) o en un entorno de red (por ejemplo, a través de Internet, una red de área amplia, una red de área local, una red cliente-servidor (tal como una red de cómputo de nube), o cualquier otra red) utilizando una o más computadoras de red.

Para mayor claridad, sólo determinados aspectos
20 seleccionados de las implementaciones basadas en *software* se describen. Otros detalles que son bien conocidos en la técnica son omitidos. Por ejemplo, debe entenderse que la tecnología descrita no está limitada a ningún lenguaje o programa informático específico. Por ejemplo, la tecnología descrita puede ser implementada por
25 *software* escrito en C++, Java, Perl, JavaScript, Adobe Flash, o

cualquier otro lenguaje de programación adecuado. Asimismo, la tecnología descrita no se limita a cualquier computadora en particular o tipo de *hardware*. Ciertos detalles de computadoras y *hardware* adecuados son bien conocidos y no necesitan ser
5 establecidos en detalle en esta descripción.

Además, ninguna de las modalidades basadas en *software* (que comprenden, por ejemplo, instrucciones ejecutables por computadora para causar que una computadora realice cualquiera de los métodos descritos) se pueden cargar, descargar o acceder de forma remota a
10 través de un medio de comunicación adecuado. Tales medios de comunicación adecuados incluyen, por ejemplo, el Internet, la Gran Red Mundial, una Intranet, aplicaciones de *software*, cable (incluyendo cable de fibra óptica), comunicaciones magnéticas, comunicaciones electromagnéticas (incluyendo comunicaciones RF,
15 de microondas e infrarrojas), comunicaciones electrónicas u otros medios de comunicación.

Los métodos, aparatos y sistemas descritos no deben ser contruidos como una limitación de ninguna manera. En cambio, la presente descripción se dirige hacia todas las características y
20 aspectos novedosos y no obvios de las varias modalidades descritas, solas y en varias combinaciones y subcombinaciones de uno con el otro. Los métodos, aparatos y sistemas descritos no se limitan a cualquier aspecto o característica específica o combinación de los mismos, ni tampoco las modalidades descritas requieren que
25 cualquiera de una o más ventajas específicas estén presentes o los

problemas puedan resolverse.

Las tecnologías de cualquier ejemplo pueden ser combinadas con las tecnologías descritas en uno o más de los otros ejemplos. En vista de las muchas posibles modalidades en donde los principios de la tecnología descrita pueden aplicarse, debe reconocerse que las modalidades ilustradas son ejemplos de la tecnología descrita y no deben ser tomadas como una limitación del alcance de la tecnología descrita.

10

15

20

25