



19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 323 230**

51 Int. Cl.:
G10L 15/22 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **02800417 .4**

96 Fecha de presentación : **01.10.2002**

97 Número de publicación de la solicitud: **1433165**

97 Fecha de publicación de la solicitud: **30.06.2004**

54 Título: **Interfaz de usuario por voz global.**

30 Prioridad: **03.10.2001 US 327207 P**
30.09.2002 US 260906

45 Fecha de publicación de la mención BOPI:
09.07.2009

45 Fecha de la publicación del folleto de la patente:
09.07.2009

73 Titular/es: **Agile TV Corporation**
Building 202, 333 Ravenswood Avenue
Menlo Park, California 94025, US

72 Inventor/es: **Jordan, Adam;**
Maddux, Scott, Lynn;
Plowman, Tim;
Stanbach, Victoria y
Williams, Jody

74 Agente: **Arizti Acha, Mónica**

ES 2 323 230 T3

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

DESCRIPCIÓN

Interfaz de usuario por voz global.

5 **Campo de la invención**

Esta invención se refiere en general a la tecnología de las comunicaciones interactivas, y más en particular a una interfaz de usuario activada por voz usada en un sistema de comunicaciones para televisión por cable u otros servicios.

10 **Antecedentes de la invención**

Los sistemas de reconocimiento de voz llevan desarrollándose más de un cuarto de siglo, lo que ha dado como resultado una variedad de herramientas de hardware y software para ordenadores personales. Los productos y servicios que emplean reconocimiento de voz están desarrollándose a gran velocidad y se aplican de manera continua a nuevos mercados.

Con la sofisticación de las tecnologías de reconocimiento de voz, las tecnologías de conexión en red y las tecnologías de telecomunicación, un sistema de comunicaciones activado por voz multifuncional, que incorpore servicios de programas de TV, servicios de vídeo bajo demanda (VOD), y servicios de Internet etc., se vuelve una realidad. Sin embargo, esta tendencia de integración crea nuevos retos técnicos, uno de los cuales es proporcionar una interfaz de usuario activada por voz para gestionar el acceso a los diferentes servicios. Por ejemplo, una interfaz de usuario activada por voz sencilla y fácil de usar es esencial para implementar un sistema de servicios por cable que sea más cómodo para los usuarios y más interactivo.

En un sistema de vídeo bajo demanda (VOD), los abonados de cable pagan una tarifa por cada programa que quieren ver, y pueden tener acceso al vídeo durante varios días. Mientras tienen tal acceso, pueden iniciar el vídeo en cualquier momento, verlo cuantas veces quieran y usar controles como los de un reproductor de vídeo para ir hacia delante y hacia atrás. Uno de los problemas con los sistemas de vídeo bajo demanda habilitados por teclas es que la navegación es incómoda. Los abonados de cable a menudo deben pulsar las teclas arriba/abajo de la página repetidas veces hasta que encuentran la película que quieren. Esto es poco práctico en sistemas habilitados por voz porque existen límites en cuanto al número de elementos que el sistema de reconocimiento de voz puede gestionar de una vez. Lo que se desea es una interfaz potente que dé a los usuarios más opciones de navegación sin perjudicar la precisión del reconocimiento. Por ejemplo, la interfaz podría permitir a los usuarios, cuando están viendo una lista de películas, decir el nombre de una película en la lista y establecer el enlace a la pantalla de información sobre la película.

La guía de programas interactiva (IPG) es la aplicación que usan los abonados de cable para encontrar lo que están emitiendo en la televisión. Uno de los problemas con las guías de programas habilitadas por teclas es que la navegación es incómoda. Los abonados de cable a menudo deben pulsar las teclas arriba/abajo de la página repetidas veces hasta que encuentran el programa que quieren. Lo que se desea adicionalmente es una interfaz fluida en la que puedan realizarse numerosas funciones comunes con pocas órdenes de voz. Por ejemplo, la interfaz permite el uso de órdenes habladas para controlar todas las funcionalidades IPG.

Otro problema es que el usuario debe cambiar a la guía de programa para encontrar lo que se está emitiendo y entonces volver a cambiar para ver el programa. Existen algunos atajos, pero encontrar programas y entonces cambiar a los mismos todavía requiere numerosas pulsaciones de teclas. Lo que se desea adicionalmente es una aplicación que permita a los abonados de cable obtener acceso en un solo paso a los programas que quiere ver sin tener que salir de la pantalla actual.

Otra cuestión importante en el diseño de una interfaz de usuario activada por voz es la capacidad de respuesta. Para interactuar con el sistema de comunicaciones de manera eficaz, se requiere que el usuario dé órdenes aceptables, y se requiere que el sistema de comunicaciones proporcione realimentación instantánea. Sin embargo, puede que un usuario regular no pueda recordar las órdenes habladas usadas en el sistema de interfaz por voz. Lo que se desea adicionalmente es un mecanismo eficaz para proporcionar mensajes de realimentación visuales inmediatos y coherentes que consistan en órdenes usadas con frecuencia, texto que pueda emitirse de forma hablada y acceso al menú principal, así como ofrecer niveles crecientes de ayuda en caso de reconocimiento de voz no satisfactorio.

El documento US-A-5.774.859 se refiere a una interfaz de reconocimiento de voz para una unidad terminal de abonado en un sistema de información para implementar control hablado de dispositivos electrónicos, que comprende una unidad terminal que tiene un procesador para ejecutar un algoritmo de reconocimiento de voz. La interfaz comprende además un control de visualización en pantalla incluido en la unidad terminal de abonado que está adaptado para seleccionar diferentes modos de funcionamiento al activar una tecla de "reconocimiento". El control de visualización en pantalla genera de manera selectiva caracteres en pantalla y visualizaciones gráficas en lugar de o superpuestos a la señal de vídeo.

65 **Sumario de la invención**

Esta invención proporciona una interfaz de usuario por voz global (GSUI) que soporta el uso de voz como un mecanismo de control de televisión digital y otros contenidos. La funcionalidad y el diseño visual de la GSUI son

consistentes en todas las aplicaciones y servicios activados por voz. El diseño visual puede incluir el uso de un agente como un asistente para introducir conceptos y guiar al usuario a través de la funcionalidad del sistema. El contenido específico en la GSUI puede ser sensible al contexto y estar personalizado a la aplicación o servicio particular. La invención se define mediante las reivindicaciones independientes adjuntas.

La realización preferida actualmente de la GSUI consiste en los siguientes elementos:

- (1) un sistema de entrada, que incluye un micrófono incorporado en un mando a distancia convencional con una tecla de pulsar para hablar, para recibir la orden hablada del usuario (es decir orden de voz); (2) un sistema de reconocimiento de voz para transcribir una orden hablada en una o más órdenes aceptables por el sistema de comunicaciones; (3) un sistema de navegación para navegar entre aplicaciones ejecutadas en dicho sistema de comunicaciones; y (4) un conjunto de superposiciones en la pantalla para ayudar a los usuarios a entender el sistema y proporcionar realimentación de usuario en respuesta a entradas; y (5) una aplicación de centro de usuario que proporciona ayuda adicional, formación y tutoriales, ajustes, preferencias y entrenamiento del emisor de voz.

Las superposiciones se clasifican en cuatro categorías: (1) un conjunto de superposiciones de realimentación de voz inmediatas; (2) una superposición o superposiciones de ayuda que proporcionan una lista sensible al contexto de órdenes activadas por voz usadas con frecuencia para cada pantalla de cada aplicación activada por voz; (3) un conjunto de superposiciones de realimentación que proporcionan información sobre un problema que experimenta dicho sistema de comunicaciones; y (4) una superposición de menú principal que muestra una lista de servicios disponibles para el usuario, pudiendo accederse a cada uno de dichos servicios mediante orden hablada.

Una superposición de realimentación de voz inmediata es una pequeña pestaña, que proporciona realimentación simple, no textual y entendida rápidamente al usuario sobre el funcionamiento básico de la GSUI. Muestra al usuario cuándo el sistema de comunicaciones está escuchando o procesando una emisión de voz, si la aplicación está habilitada por voz o no y si la emisión de voz se ha entendido o no.

Las últimas tres categorías de superposiciones son cuadros de diálogo, cada uno de los cuales puede contener una pestaña que indica un estado específico del sistema de reconocimiento de voz, uno o más cuadros de texto para transmitir información de servicio y una o más teclas virtuales que pueden seleccionarse o bien mediante una orden hablada o bien pulsando las teclas correspondientes reales del dispositivo de mando a distancia.

La superposición de ayuda proporciona una lista sensible al contexto de órdenes habladas para la aplicación activada por voz actual y es accesible en todo momento. También proporciona instrucciones cortas sobre qué texto en pantalla puede emitirse de forma hablada y enlaces a más ayuda en el centro de usuario y el menú principal. En el presente documento, el término “puede emitirse de forma hablada” es sinónimo de “activado por voz” y “habilitado por voz”.

Las superposiciones de realimentación incluyen superposiciones de realimentación de reconocimiento y superposiciones de realimentación de aplicación. Las superposiciones de realimentación de reconocimiento informan al usuario de que ha habido un problema con el reconocimiento. El tipo de realimentación que se da al usuario incluye mensajes genéricos de “no entiendo”, listas de posibles coincidencias de reconocimiento y ayuda más detallada para mejorar el reconocimiento. Las superposiciones de realimentación de aplicación informan al usuario acerca de errores o problemas con la aplicación que no están relacionados con reconocimiento no satisfactorio.

La superposición de menú principal proporciona la lista de servicios por cable digital que están disponibles para el usuario. La superposición de menú principal pretende ser más rápida y menos intrusiva que cambiar a la lista de servicios en pantalla completa del operador del sistema múltiple.

Una implantación de la GSUI es para la guía de programas interactiva (IPG), que es la aplicación que usan los abonados de cable para encontrar lo que se está emitiendo en la televisión. La GSUI proporciona una interfaz fluida en la que numerosas funciones comunes pueden realizarse con mayor facilidad por voz. La GSUI para la IPG permite el uso de órdenes habladas para controlar toda la funcionalidad IPG. Esto incluye: (1) seleccionar “teclas” en pantalla; (2) acceder directamente a cualquier programa o canal en la ranura de tiempo actual; y (3) realizar cualquier función que puede ejecutarse con pulsaciones de teclas del mando a distancia.

Otra implantación de la GSUI es para vídeo bajo demanda (VOD), que funciona como una versión electrónica de un videoclub. La GSUI proporciona una interfaz fluida en la que numerosas funciones comunes pueden realizarse con mayor facilidad por voz. La GSUI para VOD permite el uso de órdenes habladas para controlar toda la funcionalidad VOD. Esto incluye: (1) seleccionar “teclas” en pantalla; (2) acceder directamente a cualquier título de película en una lista particular; y (3) realizar cualquier función que puede ejecutarse con pulsaciones de teclas del mando a distancia.

Otra implantación de la GSUI es para un centro de usuario, que es una aplicación que proporciona: (1) formación y tutoriales sobre cómo usar el sistema; (2) más ayuda con aplicaciones activadas por voz específicas; (3) gestión de cuentas de usuario; y (4) ajustes y preferencias de usuario para el sistema.

Otro aspecto de la invención es la incorporación de una función ID de emisor de voz en la GSUI. La ID de emisor de voz es una tecnología que permite al sistema de reconocimiento de voz identificar a un usuario particular a partir de sus emisiones de voz emitidas de forma hablada. Para que el sistema identifique al usuario, el usuario debe entrenar brevemente al sistema, con unos 45 segundos de voz. Cuando el sistema está totalmente entrenado, puede identificar a ese emisor de voz particular de entre otros numerosos emisores de voz. En la presente realización, la ID de emisor de voz mejora la precisión del reconocimiento. En otras realizaciones, la ID de emisor de voz permite al servicio por cable mostrar una interfaz personalizada y contenido de televisión personalizado para un emisor de voz entrenado particular. La ID de emisor de voz también puede permitir control parental sencillo e inmediato. Por tanto, por ejemplo puede usarse una propia emisión de voz, en lugar de un PIN, para verificar el acceso a contenido bloqueado.

Las ventajas de la GSUI dada a conocer en el presente documento son numerosas, por ejemplo: en primer lugar, proporciona realimentación acerca del funcionamiento de los sistemas de reconocimiento y entrada de voz; en segundo lugar, muestra las órdenes usadas con frecuencia en pantalla y un usuario no tiene que memorizar las órdenes; en tercer lugar, proporciona referencia visual consistente a texto activado por voz; y en cuarto lugar, proporciona información de ayuda de una manera que no obstruye la visión de la pantalla.

Breve descripción de los dibujos

La figura 1 es un diagrama de bloques que ilustra un sistema de comunicaciones a modo de ejemplo que proporciona servicios por cable digital según la invención;

la figura 2A muestra seis pestañas básicas usadas para indicar información de realimentación inmediata;

la figura 2B, 2C, 2D y 2E son diagramas de flujo que ilustran un proceso a modo de ejemplo mediante el cual el sistema de comunicaciones visualiza superposiciones de realimentación inmediatas en la pantalla;

la figura 3A es un diagrama de secuencia que muestra el la línea de tiempo de una orden hablada normal;

la figura 3B es un diagrama de secuencia que muestra la línea de tiempo cuando la orden hablada se interrumpe por una entrada de tecla (caso 1);

la figura 3C es un diagrama de secuencia que muestra la línea de tiempo cuando la orden hablada se interrumpe por una entrada de tecla (caso 2);

la figura 3D es un diagrama de secuencia que muestra la línea de tiempo cuando la orden hablada se interrumpe por una entrada de tecla (caso 3);

la figura 3E es un diagrama de secuencia que muestra la línea de tiempo en un caso en el que la ejecución de una orden hablada se interrumpe por una nueva entrada de voz;

la figura 4 es un diagrama de flujo que ilustra un proceso mediante el cual la superposición de ayuda aparece y desaparece;

la figura 5 es un diagrama de flujo que ilustra un proceso mediante el cual la superposición de menú principal aparece y desaparece;

la figura 6A es un diagrama gráfico que ilustra un cuadro de diálogo de superposición de ayuda a modo de ejemplo usado en la interfaz de usuario de pantalla de TV; y

la figura 6B es una captura de pantalla que muestra la apariencia del cuadro de diálogo de superposición de ayuda ilustrado en la figura 6A.

Descripción detallada

Un sistema de comunicaciones que proporciona servicio por cable digital

En la figura 1 se ilustra un sistema 100 de comunicaciones a modo de ejemplo para facilitar un servicio por cable digital interactivo en el que está integrada una interfaz de usuario por voz global (GSUI). El usuario interactúa con el sistema de comunicaciones dando órdenes habladas a través de un dispositivo 110 de mando a distancia, que combina funcionalidad de mando a distancia universal con un micrófono y una tecla de pulsar para hablar que actúa como un interruptor. El dispositivo de mando a distancia en la realización preferida actualmente de la invención es totalmente compatible con Motorola DCT-2000 (están presentes todas las teclas a distancia de DCT-2000 convencionales). Las órdenes habladas se transmiten desde el dispositivo 110 de mando a distancia al receptor 120 cuando el abonado de cable pulsa la tecla pulsar para hablar y habla hacia el micrófono. El receptor 120 recibe y envía la entrada de voz recibida a un módulo 130 descodificador (STB) de televisión.

El STB 130 reenvía la entrada de voz al terminal 150 de cabecera, que es el centro de control central de un sistema de televisión por cable. El terminal 150 de cabecera incluye un motor 160 de voz, que comprende un dispositivo 170

de reconocimiento de voz y un envolvedor (*wrapper*) 180 de aplicaciones. El dispositivo 170 de reconocimiento de voz intenta transcribir la entrada de voz recibida en información de texto representada por flujos binarios. La salida del dispositivo 170 de reconocimiento de voz se procesa por el envolvedor 180 de aplicaciones, que genera de manera dinámica un conjunto de gramáticas de navegación y un vocabulario, e intenta determinar si una entrada de voz se ha reconocido o no. En el presente documento, una gramática de navegación significa una colección estructurada de palabras y frases unidas mediante reglas que definen el conjunto de todas las emisiones de voz que pueden reconocerse por el motor de voz en un punto dado en el tiempo.

Cuando la entrada de voz se reconoce, el envolvedor 180 de aplicaciones transforma la entrada de voz en órdenes aceptables por el servidor 190 de aplicaciones, que a continuación lleva a cabo la petición del usuario. El servidor 190 de aplicaciones puede residir o no en el motor 160 de voz. Durante el proceso, el sistema 100 de comunicaciones devuelve un conjunto de información de realimentación a la pantalla de TV a través del STB 130. La información de realimentación se organiza en una superposición en la pantalla.

Interfaz de pantalla de televisión - funcionalidad y flujos

Los elementos de la interfaz de la pantalla de televisión de la interfaz de usuario por voz global (GSUI) incluyen (1) superposiciones de realimentación de voz inmediatas; (2) superposiciones de realimentación de voz de instrucción; (3) superposiciones de ayuda; (4) superposiciones de menú principal; e (5) indicadores de texto que puede emitirse de forma hablada.

Realimentación de voz inmediata

La realimentación de voz inmediata proporciona realimentación en tiempo real, sencilla, gráfica y entendida rápidamente, al abonado de cable sobre el funcionamiento básico de la GSUI. Esta sutil realimentación no textual da la información necesaria sin suponer una distracción. La figura 2A ilustra varias pestañas a modo de ejemplo usadas para indicar tal información de realimentación. En la realización preferida, la realimentación de voz inmediata visualiza los siguientes seis estados básicos (los expertos en la técnica apreciarán que la invención también abarca otros estados o representaciones):

(1) La tecla pulsar para hablar pulsada - el sistema ha detectado que la tecla del mando se ha pulsado y escucha al abonado de cable. En la pantalla, se visualiza una pequeña pestaña 211 que incluye, por ejemplo, un logotipo de marca o indicador de identidad resaltado u opaco.

(2) La aplicación o pantalla no está habilitada por voz. Cuando el usuario pulsa la tecla pulsar para hablar, se visualiza una pequeña pestaña 212 que incluye un símbolo de prohibición (⊘) superpuesto sobre un logotipo de marca no resaltado.

(3) El sistema está procesando una emisión de voz, es decir que cubre la duración entre la liberación de la tecla pulsar para hablar y la acción resultante del sistema de comunicaciones. En la pantalla, se visualiza una pequeña pestaña 213 que incluye un logotipo de marca parpadeante en transparencia o semitransparencia (40% de transparencia por ejemplo). La pestaña 213 se alterna con una pestaña vacía para lograr el efecto parpadeante.

(4) La aplicación está bajo alerta. En la pantalla, se visualiza una pequeña pestaña 214 que incluye un signo de exclamación amarillo superpuesto sobre un logotipo de marca no resaltado. Puede tener diferentes variantes. Por ejemplo, puede ir con un mensaje de diálogo corto (variante 214A) o un mensaje de diálogo largo (variante 214B).

(5) Ha tenido lugar un reconocimiento satisfactorio y el sistema está ejecutando una acción. En la pantalla, se visualiza una pequeña pestaña 215 que incluye una marca de comprobación verde superpuesta sobre un logotipo de marca no resaltado.

(6) Ha tenido lugar un reconocimiento no satisfactorio. Tras un primer intento, la superposición de realimentación de reconocimiento también se visualiza. En la pantalla, se visualiza una pequeña pestaña 216 que incluye un signo de interrogación rojo superpuesto sobre un logotipo de marca no resaltado.

Estos estados se muestran en el siguiente conjunto de cuatro diagramas de flujo (figura 2B a figura 2E). Obsérvese que en la realización preferida, las teclas de mando a distancia convencionales están deshabilitadas mientras se pulsa la tecla pulsar para hablar, y que una vez que el sistema ha empezado a procesar una orden hablada, la tecla pulsar para hablar se deshabilita hasta que el abonado de cable recibe notificación de que el reconocimiento ha sido satisfactorio, no satisfactorio o se ha interrumpido.

Las figuras 2B, 2C, 2D y 2E son diagramas de flujo que ilustran un proceso 200 a modo de ejemplo en el que el sistema de comunicaciones visualiza superposiciones de realimentación inmediatas en la pantalla.

La figura 2B ilustra las etapas 200(a)-200(g) del proceso:

200(a): Comprobar si una pantalla actual está habilitada por voz cuando se pulsa la tecla pulsar para hablar.

ES 2 323 230 T3

200(b): Si la pantalla actual está habilitada por voz, visualizar una primera pestaña 211 que señala que un sistema de entrada de voz está activado. Esta primera pestaña 211 incluye un logotipo de marca resaltado u opaco.

5 200(c): Si la pantalla actual no está habilitada por voz, visualizar una segunda pestaña 212 que señala una alerta de no habilitada por voz. Esta segunda pestaña 212 incluye un símbolo de prohibición (⊘) superpuesto sobre un logotipo de marca no resaltado. Permanece en la pantalla durante un intervalo de aproximadamente, por ejemplo, diez segundos.

10 200(d): Si se pulsa la tecla pulsar para hablar antes o después de que desaparezca la segunda pestaña 212, repetir 200(a).

La etapa 200(b) va seguida por las etapas 200(e), 200(f) y 200(g).

15 200(e): Si la tecla pulsar para hablar no se libera dentro de un segundo intervalo (de aproximadamente 10 segundos, por ejemplo), interrumpir el reconocimiento.

20 200(f): Si la tecla pulsar para hablar se libera después de que transcurra un tercer intervalo (de aproximadamente 0,1 segundos, por ejemplo) pero antes de que transcurra el segundo intervalo en la etapa 200(e), visualizar una tercera pestaña 213 que señala que el reconocimiento de voz está en proceso. Esta tercera pestaña incluye un logotipo de marca parpadeante en transparencia o semitransparencia.

200(g): Si la tecla pulsar para hablar si ha liberado antes de que transcurra el tercer intervalo, eliminar cualquier pestaña en la pantalla.

25 Obsérvese que la figura 2B incluye una pulsación doble de la tecla de hablar. La acción que ha de adoptarse puede diseñarse según sea necesario. Tiene lugar una pulsación doble cuando hay 400 ms o menos entre cuando “se eleva la tecla” de una pulsación principal y cuando “se hunde la tecla” de una pulsación secundaria.

30 La figura 2C ilustra las etapas 200(f)-200(k) del proceso. Obsérvese que cuando no hay congestión del sistema, pocas veces tendrá el abonado de cable que pulsar una tecla de mando a distancia mientras está procesándose una orden hablada. sin embargo, cuando hay congestión del sistema el abonado de cable debería poder usar la tecla de mando a distancia para mejorar el tiempo de respuesta. Después de la descripción de este proceso se da un amplio comentario sobre cuándo los abonados de cable pueden emitir una segunda orden mientras la primera está todavía en progreso y lo que sucede cuando lo hacen.

35 La etapa 200(f) va seguida por las etapas 200(h) y 200(i):

40 200(h): Si el módulo 130 descodificador de televisión en la figura 1 dura más que un cuarto intervalo (cinco segundos, por ejemplo) medido a partir del momento en que el abonado de cable libera la tecla pulsar para hablar hasta el momento en que se envía el último dato de voz al terminal 150 de cabecera, se interrumpe el procesamiento de reconocimiento de voz y se visualiza una cuarta pestaña 214V (que es una variante de la pestaña 214), que señala una alerta de aplicación. La cuarta pestaña 214V incluye un signo de exclamación amarillo con un mensaje de diálogo corto tal como un mensaje “procesamiento demasiado largo”. Permanece en la pantalla durante un quinto intervalo (de aproximadamente 10 segundos, por ejemplo).

45 200(i): Si se pulsa una tecla de mando a distancia distinta de la tecla pulsar para hablar mientras está procesándose una orden hablada, interrumpir el procesamiento de reconocimiento de voz y eliminar cualquier pestaña en la pantalla.

La etapa 200(h) puede ir seguida además por las etapas 200(j) y 200(k):

50 200(j): Si la tecla pulsar para hablar vuelve a pulsarse mientras la cuarta pestaña 214V está en la pantalla, eliminar la cuarta pestaña y repetir 200(a). Esta etapa ilustra una situación específica en la que el procesamiento de reconocimiento dura demasiado. Obsérvese que esto no sucede cada vez que la cuarta pestaña está en la pantalla.

55 200(k): Cuando transcurre dicho quinto intervalo o si se pulsa una tecla de mando a distancia distinta de la tecla pulsar para hablar mientras dicha cuarta pestaña 214V está en la pantalla, eliminar dicha cuarta pestaña de la pantalla.

60 La figura 2D ilustra las etapas 200(l)-200(u) tras un reconocimiento completo de 200(f). Obsérvese que el sistema hace un seguimiento del número de reconocimientos no satisfactorios seguidos. Este número se restablece a cero tras un reconocimiento satisfactorio y cuando el abonado de cable pulsa cualquier tecla de mando a distancia. Si este número no se restablece, el abonado de cable continúa viendo el mensaje de realimentación de reconocimiento largo siempre que hay un reconocimiento no satisfactorio. Si los abonados de cable tienen dificultades con el sistema, el mensaje largo está bien, incluso aunque hayan pasado varias horas entre reconocimientos no satisfactorios. La realimentación de reconocimiento sólo permanece en pantalla durante alrededor de un segundo, por lo que no es necesario eliminarla cuando se pulsa cualquiera de las teclas de mando a distancia. Cuando se vuelve a pulsar la tecla 65 pulsar para hablar, la realimentación de reconocimiento debería sustituirse por la pestaña 211 de activación de voz.

200(l): Comprobar si el reconocimiento de voz es satisfactorio.

ES 2 323 230 T3

200(m): Si el reconocimiento de voz es satisfactorio, visualizar una quinta pestaña 215 que señala un reconocimiento de voz positivo. La quinta pestaña incluye una marca de comprobación verde superpuesta sobre un logotipo de marca no resaltado. Permanece en la pantalla durante un intervalo de aproximadamente, por ejemplo, un segundo.

5 200(n): Si la tecla pulsar para hablar vuelve a pulsarse antes de que desaparezca la quinta pestaña 215, repetir 200(a).

200(l) va seguida por las etapas 200(o), 200(q) y 200(r).

10 200(o): Si el reconocimiento de voz no es satisfactorio, comprobar el número de reconocimientos no satisfactorios. El sistema de comunicaciones hace un seguimiento automático del número y se restablece a cero con cada reconocimiento satisfactorio o cuando se pulsa cualquier tecla del dispositivo de mando a distancia.

15 200(p): Si el reconocimiento completo es el primer reconocimiento no satisfactorio, visualizar una sexta pestaña 216 que señala un mal reconocimiento de voz. Esta sexta pestaña 216 incluye un signo de interrogación rojo superpuesto sobre dicho logotipo de marca. Permanece en la pantalla durante aproximadamente, por ejemplo, un segundo.

20 200(q): Si la tecla pulsar para hablar vuelve a pulsarse antes de que desaparezca la sexta pestaña 216, repetir 200(a).

La etapa 200(o) va seguida por las etapas 200(r) y 200(s):

25 200(r): Si el reconocimiento completo es el segundo reconocimiento no satisfactorio, visualizar una primera variante 216A de la sexta pestaña que señala un mal reconocimiento de voz y visualizar un mensaje de texto corto. Esta primera variante 216A de la sexta pestaña comprende un signo de interrogación rojo superpuesto sobre dicho logotipo de marca y un cuadro de diálogo corto que visualiza un mensaje de texto corto. La primera variante 216A permanece en la pantalla durante aproximadamente, por ejemplo, diez segundos.

30 200(s): Si la tecla pulsar para hablar vuelve a pulsarse antes de que desaparezca la primera variante 216A de la sexta pestaña, repetir 200(a).

La etapa 200(o) va seguida por las etapas 200(t) y 200(u):

35 200(t): Si se trata el tercer reconocimiento no satisfactorio, visualizar una segunda variante 216B de la sexta pestaña que señala un mal reconocimiento de voz y visualizar un mensaje de texto largo. La segunda variante de la sexta pestaña permanece en la pantalla durante un intervalo de aproximadamente, por ejemplo, diez segundos.

40 200(u): Si la tecla pulsar para hablar se pulsa antes de que desaparezca la segunda variante 216B de la sexta pestaña, repetir 200(a).

La figura 2E ilustra las etapas 200(v)-200(x) que siguen a la etapa 200(e). Obsérvese que en la realización preferida, hay dos mensajes diferentes cuando la tecla de hablar se pulsa durante un intervalo largo. El primer mensaje trata el caso relativamente normal en el que el abonado de cable tarda más de diez segundos en pronunciar la orden.

45 El segundo trata el caso atípico en el que la tecla pulsar para hablar está atascada. No hay transición entre los dos mensajes. El segundo mensaje permanece en la pantalla hasta que se libera la tecla.

200(e): Si la tecla pulsar para hablar no se libera dentro de un segundo intervalo (de aproximadamente diez segundos, por ejemplo), interrumpir el reconocimiento.

50 200(v): Visualizar una primera variante 214A de la cuarta pestaña. La primera variante 214A incluye un signo de exclamación amarillo y un primer mensaje de texto. Esta pestaña permanece en la pantalla durante un intervalo de aproximadamente, por ejemplo, diez segundos.

55 200(w): Eliminar la primera variante 214A de la cuarta pestaña de la pantalla si la tecla pulsar para hablar se libera después de que haya transcurrido el intervalo.

60 200(x): Visualizar una segunda variante 214B de la cuarta pestaña. La segunda variante 214B incluye un signo de exclamación amarillo y un segundo mensaje de texto. Esta pestaña no se elimina a menos que se libere la tecla pulsar para hablar.

Secuenciación de órdenes

65 A continuación se describen diversos problemas relativos a la secuenciación de órdenes. Estos problemas resultan de la latencia entre una orden y su ejecución. Las órdenes habladas introducen latencias más largas porque la voz requiere más ancho de banda hasta el terminal de cabecera, y puede verse afectada por la congestión de la red. Además, algunas aplicaciones se implementan por un agente. En estos casos, el reconocimiento se realiza en el motor del sistema de comunicaciones y la orden se envía entonces al servidor de aplicaciones del agente. Las aplicaciones en el

ES 2 323 230 T3

motor y aquéllas en el servidor del agente deberían parecer iguales para los abonados de cable. En particular, es muy deseable que la realimentación de reconocimiento para una orden hablada y los resultados de la ejecución aparezcan en la pantalla de televisión al mismo tiempo. Sin embargo, si es probable que haya latencia en la comunicación con un servidor de aplicaciones fuera del motor o en la ejecución de la orden, la realimentación de reconocimiento debe aparecer en cuanto esté disponible.

Cuando hay congestión y las órdenes habladas están tardando mucho en procesarse, el abonado de cable puede intentar usar las teclas del mando a distancia o emitir otra orden hablada. Los diagramas de secuencia siguiente describen lo que sucede cuando el abonado de cable intenta emitir otra orden. Hay condiciones de carrera en el sistema subyacente. Las directrices para tratar estos problemas de secuenciación promueven dos objetivos generales:

En primer lugar, el abonado de cable deberá tener el control. Si una orden tarda demasiado, el abonado de cable debe poder emitir otra orden. En los diagramas de secuencia, cuando un abonado de cable pulsa una tecla de mando a distancia mientras está procesándose una orden hablada, se reasigna la prioridad de la orden hablada, si es posible, para devolver el control al abonado de cable. Una descripción detallada de cuándo es posible la reasignación de la prioridad y qué parte del sistema es responsable de la reasignación de la prioridad acompaña a los diagramas de secuencia.

En segundo lugar, el sistema debería ser lo más consistente posible. Para lograr esto, es necesario minimizar las condiciones de carrera en el sistema subyacente. Esto puede hacerse al menos de dos maneras:

(1) Impedir que el abonado de cable emita una segunda orden de voz hasta que el STB reciba una indicación de si el reconocimiento para la primera orden ha sido satisfactorio o no. Esto hace que sea muy probable que la aplicación haya recibido la primera orden y la esté ejecutando para cuando el abonado ve la realimentación de reconocimiento. Si la orden todavía tarda mucho en ejecutarse, hay dos explicaciones, o bien hay un problema de red entre el motor y el servidor de aplicaciones que ejecuta la orden, o bien la latencia está en la aplicación, y no en el sistema de reconocimiento de voz. Los problemas de red pueden tratarse mediante la secuenciación de órdenes descrita a continuación. Las aplicaciones en las que puede haber latencias largas deberían tener ya mecanismos integrados para abordar múltiples solicitudes en procesamiento al mismo tiempo. Por ejemplo, puede que se tarde mucho en recuperar una página web, y el navegador web estaría preparado para descartar la primera solicitud cuando llega una segunda solicitud.

(2) Hacer que las aplicaciones secuencien la ejecución de órdenes como sigue. Si el abonado de cable emite órdenes en el orden orden hablada (A), seguida por tecla orden (B) y la aplicación las recibe en el orden A, B, ambas órdenes se ejecutan. Si la aplicación las recibe en el orden B, A, la orden B se ejecuta y cuando llega la orden A, se descarta porque está obsoleta.

La figura 3A a la figura 3E son diagramas de secuencia que muestran los instantes en el tiempo en los que puede emitirse una segunda orden y describen lo que deberá pasar cuando se emite la segunda orden.

La figura 3A muestra la línea de tiempo de una orden hablada normal. Los puntos 310 redondos son eventos. Una barra 320 que une eventos indica actividad. Por ejemplo, la barra entre tecla pulsar para hablar (PTT) pulsada y tecla PTT liberada indica que la tecla PTT se hunde y están generándose paquetes de voz. Las etiquetas en el lado izquierdo del diagrama indican los componentes en el sistema. STB/VoiceLink se refiere al sistema de entrada que incluye el módulo 130 descodificador, el mando 110 a distancia y el receptor 120 tal como se ilustra en la figura 1.

El envolver de aplicaciones y el servidor de aplicaciones se indican como componentes separados. Cuando toda la aplicación reside en el motor, el envolver y el servidor son el mismo componente y la secuenciación de órdenes es más sencilla.

Un punto en la misma línea horizontal que el nombre del componente significa que el evento tuvo lugar en este componente. Las etiquetas 330 en la parte inferior del diagrama describen los eventos que han tenido lugar. Los eventos están ordenados por el momento en que tuvieron lugar.

Hay cuatro casos en los que puede emitirse una orden de tecla o hablada mientras otra orden ya está en progreso. Se muestran con la etiqueta 340 “casos de interrupción” en la parte superior derecha del diagrama. El resto de diagramas (figuras 3B-3E) describen lo que sucede en cada uno de estos casos.

La figura 3B muestra la línea de tiempo cuando se interrumpe la orden hablada por una entrada de tecla (caso nº 1). En este caso, el abonado de cable pulsó una tecla de mando a distancia antes de que el STB/VoiceLink enviase todos los paquetes para la orden hablada al sistema de reconocimiento. El diagrama muestra que la orden hablada se cancela y se ejecuta la orden de la tecla de mando a distancia. El STB/VoiceLink y el sistema de reconocimiento deben actuar conjuntamente para cancelar la orden hablada.

La figura 3C muestra la línea de tiempo cuando la orden hablada se interrumpe por una entrada de tecla (caso nº 2). En este caso, el abonado de cable pulsa una tecla de mando a distancia después de que se haya recibido el último paquete por el sistema de reconocimiento y antes de que se haya procesado la lista de los n mejores por el envolver de aplicaciones. En ambas situaciones, la orden hablada se descarta y se ejecuta la orden de la tecla. Este diagrama muestra que el STB/VoiceLink y el sistema de reconocimiento pueden haber actuado conjuntamente para cancelar la

orden hablada en el subcaso A, y la aplicación no tendría que haberse visto implicada. En el subcaso B, la aplicación cancela la orden hablada porque llegó fuera de la secuencia.

La figura 3D muestra la línea de tiempo en la que la orden hablada se interrumpe por una entrada de tecla (caso nº 3). En este caso, el abonado de cable pulsó una tecla de mando a distancia después de que se recibiera el acuse de recibo de reconocimiento positivo y antes de que se ejecutara la orden hablada. Es responsabilidad de la aplicación determinar cuál de las dos órdenes ejecutar. En el subcaso A, la orden hablada se recibe fuera de la secuencia y se ignora. En el subcaso B, la orden hablada se recibe en orden y se ejecutan tanto la orden hablada como la orden de tecla de mando a distancia.

La figura 3E muestra la línea de tiempo en un caso en el que la orden hablada se interrumpe por una entrada de voz. El abonado de cable emite una segunda orden hablada después de que se haya recibido el acuse de recibo de reconocimiento positivo y antes de que se ejecute la primera orden hablada. Es responsabilidad de la aplicación determinar cuál de las dos órdenes ejecutar. En el subcaso A las órdenes habladas se reciben en orden y se ejecutan ambas órdenes. En el subcaso B, las órdenes habladas se reciben desordenadas, se ejecuta la segunda orden y se ignora la primera orden.

Superposición de ayuda

La superposición de ayuda visualiza una lista sensible al contexto corta de órdenes habladas usadas con frecuencia para cada pantalla única de cada aplicación habilitada por voz. La superposición de ayuda pretende conseguir dos objetivos: en primer lugar, proporcionar indicaciones a nuevos usuarios para permitirles controlar la funcionalidad básica de una aplicación habilitada por voz particular; y en segundo lugar, proporcionar un recordatorio de órdenes básicas a usuarios experimentados en caso de que olviden estas órdenes. Además de visualizar órdenes específicas de la aplicación, la superposición de ayuda siempre muestra las órdenes para acceder a la superposición de menú principal y “más ayuda” del centro de usuario. La superposición de ayuda también explica el indicador de texto que puede emitirse de forma hablada, si está activado. Obsérvese que la superposición de ayuda ayuda al abonado de cable en el uso y las órdenes habladas. No describe la funcionalidad de la aplicación.

Las superposiciones de ayuda están organizadas como sigue:

- Órdenes específicas de la aplicación (aproximadamente cinco órdenes básicas)
- Orden de “más ayuda” (enlace al centro de usuario)
- Orden de “menú principal” para visualizar la superposición de menú principal
- “Salir” para que desaparezca la superposición.

La figura 4 es un diagrama de flujo que ilustra un proceso mediante el cual la superposición de ayuda aparece y desaparece. El proceso incluye las siguientes etapas:

400(a): Visualizar una primera superposición de ayuda si el reconocimiento de voz es satisfactorio. La primera superposición 410 de ayuda es un cuadro de diálogo que incluye (1) una pestaña que señala un reconocimiento de voz positivo - por ejemplo puede ser una marca de comprobación verde superpuesta sobre un logotipo de marca no resaltado; (2) un cuadro de texto para información de ayuda textual, que puede incluir además un enlace de “más ayuda” y texto que puede emitirse de forma hablada; y (3) teclas virtuales - una para el menú principal y otra para salir para que desaparezca la superposición. La primera superposición de ayuda podría permanecer en la pantalla durante un primer intervalo, por ejemplo, veinte segundos.

400(b): Eliminar la primera superposición 410 de ayuda de la pantalla si (1) transcurre el primer intervalo; (2) se pulsa accidentalmente cualquier tecla del dispositivo de mando a distancia; o (3) se selecciona la tecla salir.

400(c): Visualizar una segunda superposición 420 de ayuda mientras se está pulsando la tecla pulsar para hablar para dar una nueva entrada de voz. Estructuralmente, la superposición 420 de ayuda es igual que la superposición 410 de ayuda. La única diferencia es que la pestaña de realimentación inmediata en la superposición 420 de ayuda señala la activación de pulsar para hablar en lugar de un reconocimiento positivo como en la superposición 410 de ayuda.

Superposiciones de realimentación

Hay dos tipos de superposiciones de realimentación: superposiciones de realimentación de reconocimiento y superposiciones de realimentación de aplicación. Las superposiciones de realimentación de reconocimiento informan al abonado de cable de que se ha producido un problema con el reconocimiento de voz. Las superposiciones de realimentación de aplicación informan al abonado de cable acerca de errores o problemas relacionados con la interfaz de voz de la aplicación. Las superposiciones de realimentación de reconocimiento existen en tres estados y responden a varias condiciones diferentes. Los tres estados de realimentación de reconocimiento diferentes corresponden a varios reconocimientos no satisfactorios que suceden de manera secuencial. Este comportamiento sucede cuando el abonado

ES 2 323 230 T3

de cable intenta emitir múltiples veces una orden que no se reconoce por el sistema; los tres estados ofrecen progresivamente más realimentación al abonado de cable con cada intento. La respuesta a cada intento incluiría enlaces a niveles crecientes de ayuda.

5 Los tres estados de realimentación de reconocimiento son: (1) el primer reconocimiento no satisfactorio - el indicador de realimentación de voz inmediato cambia a un signo de interrogación que proporciona realimentación de comprensión mínima rápida al abonado de cable; (2) el segundo reconocimiento no satisfactorio - se visualiza la superposición de realimentación con un mensaje y enlace a la superposición de ayuda; y (3) el tercer reconocimiento no satisfactorio - se visualiza la superposición de realimentación con otro mensaje y enlaces a la superposición de ayuda
10 y de más ayuda en el centro de usuario.

Las diferentes condiciones de realimentación de reconocimiento que corresponden a la cantidad de información que el dispositivo de reconocimiento tiene sobre la emisión de voz del abonado de cable y a la latencia en el sistema subyacente incluyen:

- 15 • Puntuación de confianza baja. Se visualiza un conjunto de mensajes genéricos de “no entiendo”.
- Puntuación de confianza media. Puede visualizarse una lista de posibles coincidencias.
- 20 • Nivel de sonido de emisión de voz demasiado bajo. Se visualiza el mensaje “hablar más alto o situar el mando más cerca”.
- Nivel de sonido de emisión de voz demasiado alto. Se visualiza el mensaje “hablar más bajo o situar el mando más lejos”
- 25 • Demasiado tiempo hablando. En la realización preferida, hay un límite de diez segundos para la cantidad de tiempo que puede pulsarse la tecla pulsar para hablar. Si se supera el límite de tiempo, se descarta la emisión de voz y se visualiza el mensaje “demasiado tiempo hablando”.
- 30 • Tecla pulsar para hablar atascada. Si la tecla pulsar para hablar se ha hundido, por ejemplo, durante veinte segundos, se visualiza el mensaje “tecla pulsar para hablar atascada”.
- Procesamiento demasiado largo. Tal como se describió en 200(h) anteriormente, si el mando a distancia y el STB no pueden transferir una emisión de voz al terminal de cabecera en, por ejemplo, cinco segundos después de que se
35 libere la tecla pulsar para hablar, se visualiza el mensaje “procesamiento demasiado largo”.

Las superposiciones de realimentación de aplicación se visualizan cuando ha de comunicarse una información específica de la aplicación al abonado de cable. Un indicador diferente en la parte superior de la superposición (por ejemplo, pestaña 214) diferencia la realimentación de aplicación de la realimentación de reconocimiento. La realimentación de aplicación incluiría mensajes de respuesta o deficiencia relativos a la interfaz de voz de la aplicación.
40

Superposiciones de menú principal

45 En la realización preferida, la superposición de menú principal proporciona una lista de servicios por cable digital habilitados por voz que están disponibles para el abonado de cable. La superposición de menú principal pretende ser más rápida y menos intrusiva que conmutar a una pantalla independiente para obtener la misma funcionalidad. La lista de servicios puede incluir, por ejemplo: (1) “Ver TV” para ver la TV en pantalla completa; (2) “Guía de Programas”; (3) “Vídeo bajo demanda”; (4) “Internet/Sistema cerrado”; y (5) “Centro de usuario”. El servicio actual está resaltado.
50 Órdenes visualizadas adicionales incluyen “Salir” para que desaparezca la superposición.

La figura 5 es un diagrama de flujo que ilustra el proceso mediante el cual aparece y desaparece la superposición de menú. El proceso incluye las siguientes etapas implementadas por ordenador:

55 500(a): Visualizar una primera superposición de menú principal si el reconocimiento de voz es satisfactorio. La primera superposición 510 de menú principal es un cuadro de diálogo que incluye (1) una pestaña que señala un reconocimiento de voz positivo - por ejemplo puede ser una marca de comprobación verde superpuesta sobre un logotipo de marca no resaltado; (2) un cuadro de texto para información textual acerca del menú principal, que puede incluir además texto que puede emitirse de forma hablada; y (3) una o más teclas virtuales tales como la tecla de ayuda
60 y la tecla de salida. La superposición de menú principal permanece en la pantalla durante un primer intervalo, de unos 20 segundos por ejemplo.

500(b): Eliminar la primera superposición 510 de menú principal de la pantalla si (1) transcurre el primer intervalo; (2) se pulsa accidentalmente cualquier tecla del mando a distancia; o (3) se selecciona la tecla de salida.
65

500(c): Visualizar una segunda superposición 520 de menú principal mientras está pulsándose la tecla pulsar para hablar para dar una nueva entrada de voz de navegación. Estructuralmente, la segunda superposición 520 de menú principal es igual que la primera superposición 510 de menú principal. La única diferencia es que la pestaña de

realimentación inmediata en la segunda superposición 520 de menú principal señala activación de pulsar para hablar en lugar de un reconocimiento positivo como en la primera superposición 510 de menú principal.

Indicador de texto que puede emitirse de forma hablada

El indicador de texto que puede emitirse de forma hablada parece estar intercalado a modo de capa sobre aplicaciones habilitadas por voz como parte de la GSUI. Este tratamiento puede aplicarse a texto estático o dinámico. Se usa texto estático en etiquetas para gráficos en pantalla o teclas que pueden seleccionarse moviendo un área resaltada con las teclas de dirección del mando a distancia. Como tal, la mayor parte de las pantallas normalmente tienen varias teclas con etiquetas de texto y por lo tanto requieren un número correspondiente de indicadores de texto que puede emitirse de forma hablada. Se usa texto dinámico en contenido tal como la lista de películas para la aplicación de vídeo bajo demanda (VOD). Cada línea de texto dinámico puede incluir indicadores de texto que puede emitirse de forma hablada para indicar qué palabras pueden emitirse de forma hablada. El indicador de texto que puede emitirse de forma hablada es actualmente un punto verde y puede cambiar a un indicador diferente. Es importante que el indicador sea visible pero que no suponga una distracción. Además, el abonado de cable deberá poder activar y desactivar los indicadores de texto que puede emitirse de forma hablada.

Interfaz de pantalla de televisión - interfaz gráfica de usuario (GUI)

Las superposiciones de la GSUI descritas anteriormente se crean a partir de un conjunto de elementos de herramientas. Los elementos de herramientas incluyen disposición, indicador de marca, pestaña de realimentación, cuadro de diálogo, cuadro de texto, estilo de letra de imprenta, imágenes de fondo, zona resaltada de selección e indicador de texto que puede emitirse de forma hablada.

El operador de sistema múltiple (MSO) tiene cierta flexibilidad para especificar dónde debería aparecer la GSUI. La GSUI está anclada a la pestaña de realimentación de voz inmediata, que debe aparecer a lo largo de uno de los bordes de la pantalla. El punto de anclaje y el tamaño y la forma de los cuadros de diálogo pueden ser diferentes para cada MSO.

La identidad de marca del proveedor de servicios o el diseñador de sistema puede aparecer sola o junto con la identidad de marca del MSO. Siempre que aparezca la identidad de marca, debe ser consistente preferiblemente en cuanto al tratamiento de la ubicación, tamaño y color. La colocación estática del indicador de marca es clave a la hora de reforzar que la realimentación GSUI procede del producto del diseñador. Varios estados de color y animación en el indicador de marca se usan para indicar la funcionalidad del sistema. Las pantallas que contienen el indicador de marca contienen información relativa al reconocimiento de voz. El indicador de marca tiene varios estados de transparencia y color para proporcionar claves visuales del estado o resultado de una solicitud de voz. Por ejemplo: un logotipo indicador con un 40% de transparencia se usa como una indicación de marca, que aparece en todos los aspectos de la GSUI; un logotipo indicador opaco se usa para indicar que la tecla pulsar para hablar del mando se está pulsando actualmente; y un logotipo indicador parpadeante con un 40% de transparencia se usa para indicar que el sistema ha oído lo que el usuario ha dicho y está procesando la información. Un indicador de marca puede situarse en cualquier lugar en la pantalla, aunque preferiblemente se sitúa en la esquina superior izquierda de la pantalla y mantiene el mismo tamaño en toda la GSUI.

La pestaña de realimentación es el elemento gráfico en pantalla usado para implementar realimentación de voz inmediata tal como se describió anteriormente. La pestaña de realimentación usa una variedad de gráficos para indicar el estado y el resultado de una solicitud de voz. Por ejemplo: una marca de comprobación verde superpuesta sobre el indicador de marca podría indicar “realimentación de reconocimiento de voz positivo”; un signo de interrogación rojo superpuesto sobre el indicador de marca podría indicar “realimentación de mal reconocimiento de voz”; un logotipo indicador de marca parpadeante con un 40% de transparencia podría indicar “procesamiento de reconocimiento de voz”; un logotipo indicador de marca opaco podría indicar “activación de la tecla pulsar para hablar”; un signo de exclamación amarillo superpuesto sobre el logotipo indicador de marca podría indicar “alerta de aplicación”; un símbolo de prohibición superpuesto sobre el logotipo indicador de marca podría indicar “alerta de no habilitada por voz”. Las reglas de diseño de pestañas actualmente preferidas incluyen: (1) cualquier color usado debe ser consistente (por ejemplo, R: 54, V: 152, A: 217); (2) siempre debe tener un fondo transparente; (3) siempre debe estar alineado de manera consistente, por ejemplo, en la parte superior de la pantalla de TV; (4) el tamaño debe ser siempre consistente, por ejemplo, 72 de ancho x 67 de alto píxeles; (5) el indicador de marca debe estar siempre presente; (6) las esquinas inferiores deben ser redondeadas; (7) los indicadores de estrella y gráfico deben estar centrados en la pestaña.

El cuadro de diálogo implementa la superposición de realimentación, la superposición de ayuda, la superposición de menú principal y la superposición de lista de órdenes descritas anteriormente. El cuadro de diálogo tiene una forma sencilla limitada. Puede contener un cuadro de texto para comunicar información asociada con el producto del proveedor de servicios. También puede contener teclas virtuales que pueden seleccionarse o bien mediante voz o bien mediante las teclas del mando a distancia. Diferentes cuadros de diálogo pueden usar conjuntos diferentes de teclas virtuales. Cuando dos cuadros de diálogo diferentes usan una tecla virtual, deben aparecer preferiblemente en el mismo orden con respecto al resto de las teclas y tener la misma etiqueta en cada cuadro de diálogo.

En la figura 6A se ilustra un cuadro 600 de diálogo de ayuda a modo de ejemplo. La figura 6B es una captura de pantalla que muestra la apariencia del cuadro de diálogo de ayuda ilustrado en la figura 6A. El cuadro 600 de diálogo

ES 2 323 230 T3

5 incluye un cuadro 610 de fondo usado para visualizar información gráfica y textual, un cuadro 630 de texto usado para visualizar información textual, un logotipo 640 indicador de marca y teclas 650 y 655 virtuales. El cuadro 630 de texto está superpuesto sobre el cuadro 610 de fondo. Las reglas de diseño de cuadros de diálogo actualmente preferidas incluyen: (1) el cuadro de diálogo siempre debe estar alineado con la parte superior de la pantalla de TV; (2) las esquinas inferiores deben ser redondeadas; (3) las imágenes de fondo del proveedor de servicios siempre deben estar presentes; (4) la altura del cuadro puede variar, pero el ancho debe mantenerse consistente; y (5) el cuadro siempre debe aparecer en el lado izquierdo de la pantalla de TV.

10 El cuadro 630 de texto comunica información asociada con el producto de proveedor. Esta información debe resaltar de las imágenes 620 de fondo. Para lograr esto, el cuadro 630 de texto es una forma limitada situada dentro de la forma limitada del cuadro 610 de fondo. En una realización típica, la información textual en el cuadro 630 de texto se presenta siempre en un cuadro opaco de color azul, que se superpone entonces sobre el cuadro 610 de fondo. Puede haber más de un cuadro de texto por cada cuadro de diálogo. Por ejemplo, la superposición de menú principal contiene un cuadro de texto por cada elemento del menú principal. La navegación secundaria, tal como la tecla 655 15 “menú” y la tecla 650 “salir”, puede visualizarse fuera del cuadro de texto en las imágenes de fondo del cuadro de diálogo. Las reglas de diseño de cuadros 630 de texto actualmente preferidas incluyen (1) el color debe ser siempre R: 42, V: 95, A: 170; (2) el cuadro de texto debe disponerse siempre ocho píxeles hacia el interior desde cada lateral del cuadro de diálogo; (3) todas las esquinas deben ser redondeadas; y (4) todo el texto dentro de un cuadro de texto debe estar alineado a la izquierda.

20 El uso de una única familia de fuentes con una combinación de estilos de tipos de letra ayuda a reforzar la identidad de marca. Cuando se usan diferentes tipos de letra, cada uno debería usarse para un fin específico. Esto ayuda al abonado de cable a familiarizarse con la interfaz de usuario. Cualquier tipo de letra usado debería ser legible en la pantalla de TV.

25 Las imágenes 620 de fondo se usan para reforzar el logotipo de marca. El uso consistente de las imágenes de fondo de logotipo ayuda a la marca e indica visualmente que la información que está visualizándose forma parte del producto de reconocimiento de voz.

30 La zona resaltada de selección es un elemento gráfico convencional usado para resaltar un elemento seleccionado en pantalla. En una realización típica, es una regla amarilla de dos píxeles usada para destacar texto o un cuadro de texto indicando que es el elemento actualmente seleccionado.

35 El indicador de texto que puede emitirse de forma hablada es preferiblemente un elemento gráfico consistente. Debería mantener siempre el mismo tratamiento. Debería colocarse cerca de cualquier texto que puede emitirse de forma hablada que aparezca en pantalla. En una realización preferida, el indicador de texto que puede emitirse de forma hablada es un punto verde. El punto verde debería ser consistente en tamaño y color en toda la GSUI y en todas las aplicaciones habilitadas por voz. Tal vez la única excepción a esta regla es que el punto verde es mayor en el texto de ayuda acerca del propio punto verde.

40 La pestaña de realimentación es el elemento gráfico usado para la realimentación de voz inmediata. Este elemento aparece sobre cualquier otra superposición de GSUI en pantalla. Por ejemplo, si la superposición de ayuda está en pantalla, y el abonado de cable pulsa la tecla pulsar para hablar, la pestaña de activación de tecla pulsar para hablar, es decir la imagen de logotipo sólido, aparece sobre la superposición de ayuda.

45 La superposición de ayuda contiene información útil acerca de la interfaz de usuario por voz y las teclas de menú y salir. El diseño visual de la superposición de ayuda es un cuadro de diálogo que usa estos elementos gráficos: indicador de marca, cuadro de texto, imágenes de fondo, tipo de letra y zona resaltada de menú, así como un título de cuadro de diálogo que indica para qué servicio es la ayuda. El contenido en el cuadro de texto cambia en relación con el servicio por cable digital que esté usándose. La superposición de ayuda nunca debería cambiar la disposición de diseño aunque puede aumentar o reducir sus dimensiones según las necesidades del cuadro de texto. 50

55 La superposición de realimentación se visualiza en caso de mal reconocimiento de órdenes de voz. El diseño visual actualmente preferido de la superposición de realimentación es un cuadro de diálogo que usa los siguientes elementos gráficos: indicador de marca, cuadro de texto, imágenes de fondo, tipo de letra y zona resaltada de menú, así como un título de cuadro de diálogo que indica para qué servicio es la realimentación. La superposición de realimentación nunca debería cambiar la disposición de diseño aunque puede aumentar o reducir sus dimensiones según las necesidades del cuadro de texto.

60 La superposición de menú principal es un cuadro de diálogo que contiene un título de cuadro de diálogo, teclas con enlaces a varios servicios por cable digital y una tecla de salir. El menú principal actualmente preferido usa los siguientes elementos gráficos: cuadro de diálogo, imágenes de fondo, tipo de letra, zona resaltada de menú y cuadro de texto. Cada selección en el menú principal es un cuadro de texto.

65 *Navegación*

La GSUI incorpora varias funciones de navegación. Por ejemplo, el usuario navega por la información basada en listas en pantalla a través de control de voz. Puede manipularse y navegarse por la información basada en listas de

ES 2 323 230 T3

varias formas incluyendo órdenes tales como: “ir al letra (nombre de letra)” y “subir/bajar página”. También puede accederse a elementos de las listas de películas y programas de forma aleatoria simplemente diciendo el nombre del elemento. Cuando se ve una lista de películas, el usuario puede simplemente decir el nombre de una película en esa lista y acceder al enlace de la pantalla de información de la película.

Como ejemplo adicional, el usuario puede navegar directamente entre aplicaciones a través de órdenes habladas o del menú principal habilitado por voz. El usuario también puede navegar directamente a páginas de favoritos previamente “marcadas”.

Como ejemplo adicional, el usuario puede iniciar la función de navegación en pantalla completa, lo que permite al usuario realizar lo siguiente:

- (1) navegar, buscar, filtrar y seleccionar programas mediante orden hablada. Esta funcionalidad es similar a numerosas características que se encuentran en las guías de programas interactivas pero es accesible sin la interfaz visual permitiendo así zapear de una que interrumpa menos.
- (2) Iniciar a través de control de voz una búsqueda de tipo “exploración” automática de programas en categorías o géneros. Por ejemplo, el usuario dice “explorar deportes” para iniciar un ciclo automático de programas deportivos. Cada programa permanecerá en pantalla unos segundos antes de pasar al siguiente programa en la categoría. Cuando el usuario encuentra el que quiere ver, puede decir “parar”. Las categorías incluyen, sin ánimo exhaustivo, deportes, programación infantil, películas, noticias, humor, series, drama, favoritos, telerrealidad (*reality shows*), recomendaciones, clásico, etc. Hay disponible una característica como medio para explorar todos los programas sin división por categoría.
- (3) Añadir programas o canales de televisión a las categorías tales como “favoritos”; editar programas o canales de televisión en las categorías; y borrar programas o canales de televisión de las categorías. El usuario también puede establecer “control parental” usando estas funciones de “añadir”, “editar” y “borrar”.
- (4) Buscar, usando órdenes habladas, programas particulares basándose en atributos específicos. Por ejemplo, “encontrar Los Soprano”, “encontrar película por Coppola”, etc.
- (5) Filtrar, usando órdenes habladas, grupos de programas por atributos específicos tales como género, director, actor, clasificación, estreno, popularidad, recomendación, favoritos, etc. Por ejemplo, “encontrar películas de acción” o “mostrar fútbol universitario”, etc.

Control de la guía interactiva de programas

Una implantación de la GSUI es para la guía de programas interactiva (IPG) habilitada por voz, que es la aplicación que el abonado de cable usa para encontrar lo que están emitiendo por televisión. La IPG soporta varias funcionalidades. Permite al usuario hacer lo siguiente a través de órdenes habladas:

- (1) acceder a información de programas de televisión detallada. Por ejemplo, con el programa seleccionado en la guía o visto a pantalla completa, el usuario emite una orden “obtener información” para acceder al enlace de la pantalla de información de programa.
- (2) Clasificar programas por categoría. Por ejemplo, con la IPG activa, el usuario emite la orden “mostrar deportes”. Otras categorías incluyen favoritos, películas, música, noticias, etc.
- (3) Acceder y establecer controles parentales para limitar que los niños puedan ver programación censurable.
- (4) Acceder y establecer recordatorios de programas que van a reproducirse en el futuro. Por ejemplo, con la IPG activa, el usuario emite la orden “ir a viernes 8 p.m.”, y entonces con el programa seleccionado, emite la orden “establecer recordatorio”.
- (5) Buscar programas basándose en criterios específicos. Por ejemplo, con la IPG activa, el usuario emite la orden “encontrar fútbol en viernes noche” o “encontrar Oscars”.
- (6) Realizar una compra de pago por visión.
- (7) Mejorar o acceder a servicios de televisión por cable exclusivos.

Servicio de vídeo bajo demanda

Otra implantación de la GSUI es para vídeo bajo demanda (VOD), que funciona como una versión electrónica de un videoclub. La GSUI proporciona una interfaz sencilla en la que pueden realizarse numerosas funciones comunes con más facilidad mediante órdenes habladas. La aplicación VOD permite al usuario realizar lo siguiente a través de órdenes habladas:

- (1) acceder a información de película detallada.

ES 2 323 230 T3

- (2) Clasificar por género incluyendo, sin ánimo exhaustivo, acción, infantil, humor, romance, aventura, estreno, etc.
- (3) Establecer control parental para evitar que los niños accedan a información de vídeo controlada.
- (4) Buscar por título de película, actor, premios y recomendaciones, etc.
- (5) Obtener recomendación automática basándose en identificación de huella vocal.
- (6) Navegar por Internet.

Otras funciones

La GUSI puede incorporar adicionalmente funcionalidades para permitir al usuario realizar lo siguiente a través de órdenes habladas:

- (1) iniciar comunicación de mensajería instantánea.
- (2) Acceder y jugar a juegos.
- (3) Controlar todos los ajustes de televisión incluyendo, sin ánimo exhaustivo, control de volumen, avanzar/retroceder canal, color, brillo, activación y posición de función PiP (*picture-in-picture*, una imagen sobre otra).
- (4) Control de preferencias personales y opciones de ajuste.
- (5) Enlace a información detallada de producto, tal como especificación de producto, precios y transporte, etc., basándose en anuncios de televisión o anuncios tipo *banner* incluidos en la pantalla de aplicación.
- (6) Aceptar anuncios o *banners* basándose en identificación de huella vocal.
- (7) Aceptar recomendaciones de programación basándose en identificación de huella vocal.
- (8) Aceptar información personalizada basándose en identificación de huella vocal.
- (9) Obtener configuración automática de preferencias basándose en identificación de huella vocal.
- (10) Completar todos los aspectos de una transacción de compra basándose en identificación de huella vocal (también denominada transacción de “una palabra”).
- (11) Iniciar una compra de producto integrada en programación de difusión. Por ejemplo, la orden “comprar ahora” del usuario mientras está viendo el canal QVC inicia el procedimiento de compra.
- (12) Controlar servicios domésticos tales como seguridad doméstica, sistema de entretenimiento doméstico y estéreo, y dispositivos domésticos tales como CD, radio, DVD, VCR y PVR a través de la interfaz de control por voz basada en TV.

Control por voz - órdenes y directrices

Cada orden hablada se procesa en un contexto que incluye órdenes para acceder a cualquier contenido con nombre en la pantalla que está viendo el abonado de cable, órdenes para acceder a características de aplicación, órdenes para acceder a la interfaz de usuario por voz global (GSUI), órdenes para simular pulsaciones de tecla de mando a distancia y órdenes para navegar a otras aplicaciones. Muchas de las directrices descritas en el presente documento se han desarrollado para intentar minimizar el potencial de palabras o frases de una fuente para confundirse con las de otra. Por ejemplo, el contenido en la aplicación de guía de programas interactiva (IPG) contiene los nombres de programas de televisión. Podría haber fácilmente un programa de televisión con nombre “Salir” que entraría en conflicto con el uso de “salir” como el equivalente de voz de pulsar la tecla salir en el mando a distancia. La especificación de una orden describe la forma en que encaja en el entorno.

La especificación actualmente preferida incluye: (1) alcance de la orden, que caracteriza cuándo está disponible la orden; (2) lenguaje, que define las palabras que los abonados de cable usan para invocar la orden; y (3) comportamiento, que especifica lo que sucede cuando se invoca la orden.

Las órdenes globales están disponibles siempre. Las aplicaciones sólo pueden deshabilitarlas para forzar al usuario a realizar una elección de un conjunto de elecciones específicas de la aplicación. Sin embargo, esto debería ocurrir en raras ocasiones. Las interfaces por voz están diseñadas preferiblemente para hacer que el abonado de cable sienta que tiene el control. Es muy deseable que las órdenes navegación estén habilitadas por voz y disponibles globalmente. Esto permite a los abonados de cable desplazarse de una aplicación a otra a través de voz. Cuando todas las aplicaciones

que soporta un MSO están habilitadas por voz, tanto las órdenes de navegación como las órdenes GSUI pasan a ser globales. Las órdenes GSUI están siempre disponibles para aplicaciones habilitadas por voz.

Las órdenes de navegación preferiblemente están siempre disponibles. Las órdenes de navegación incluyen órdenes específicas para permitir al abonado de cables ir a cada aplicación que soporta el MSO y órdenes generales que soportan el modelo de navegación. Por ejemplo, "vídeo bajo demanda" es una orden específica que lleva al abonado de cable a la aplicación VOD, y "último" es una orden general que lleva al abonado de cable a la pantalla apropiada según se haya definido por el modelo de navegación. El lenguaje para las órdenes de navegación puede ser diferente para cada MSO dado que cada MSO soporta un conjunto diferente de aplicaciones. El modelo de navegación determina el comportamiento de las órdenes de navegación. Puede haber un modelo de navegación global y diferentes modelos de navegación para diferentes aplicaciones. En donde ya existen modelos de navegación, la navegación se realiza a través de teclas de mando a distancia. Las órdenes habladas para navegación deberían ser preferiblemente las mismas que al pulsar las teclas de mando a distancia correspondientes. Cuando una pantalla contiene teclas virtuales para navegación y el abonado de cable invoca la orden hablada correspondiente a la tecla virtual, la tecla virtual se resalta y se invoca la orden.

El alcance de las teclas de mando a distancia varía ampliamente. Algunas teclas de mando a distancia se usan en raras ocasiones en cualquier aplicación, por ejemplo, las teclas "a", "b" y "c". Algunas se usan en la mayoría de aplicaciones, por ejemplo, los cursores. Dado que el reconocimiento puede mejorarse limitando las elecciones, se prefiere que cada contexto incluya únicamente órdenes habladas para teclas de mando a distancia aplicables. El comportamiento de las órdenes habladas para teclas de mando a distancia se mantiene igual que al pulsar las teclas de mando a distancia. Sin embargo, cuando una pantalla contiene teclas virtuales que representan teclas del mando a distancia y el abonado de cable invoca la orden hablada correspondiente a una tecla virtual, la tecla virtual se resalta y se invoca la orden.

Los abonados de cable deberían verse forzados en raras ocasiones a decir una de las elecciones en un cuadro de diálogo. Las órdenes globales están preferiblemente siempre disponibles a menos que el abonado de cable se vea forzado a decir una de las elecciones en un cuadro de diálogo. Esto debería ser un suceso infrecuente. Las personas normalmente dicen frases tales como "muestra" o "ir a" antes de emitir una orden. Las órdenes específicas de la aplicación deberían incluir estas frases para hacer las aplicaciones más cómodas de usar y más adecuadas a los lenguajes coloquial o natural.

Aunque la invención se ha descrito en el presente documento con referencia a la realización preferida, un experto en la técnica apreciará fácilmente que otras aplicaciones pueden sustituir a las expuestas en el presente documento sin alejarse del alcance de la invención. Por ejemplo, aunque la invención en el presente documento se ha descrito en conexión con servicios de televisión, los expertos en la técnica apreciarán que la invención también comprende cualquier forma de representación de información con la que interactúa un usuario tal como, por ejemplo, tecnologías habilitadas por navegador e incluirían la World Wide Web y acceso a redes de información.

Por consiguiente, la invención sólo debería estar limitada por las reivindicaciones incluidas a continuación.

REIVINDICACIONES

5 1. Medio de almacenamiento legible por ordenador que contiene un producto de programa informático adaptado para implementar una interfaz de usuario por voz global (GSUI), comprendiendo dicho producto de programa informático:

10 código de programa para transcribir órdenes habladas introducidas en un receptor y transferidas desde un módulo descodificador de televisión a un terminal de cabecera para reconocer y procesar en el terminal de cabecera en órdenes aceptables para un sistema de comunicaciones;

código de programa para navegar entre aplicaciones alojadas en dicho sistema de comunicaciones;

15 código de programa para devolver un conjunto de información de realimentación desde dicho terminal de cabecera a un dispositivo para visualizar un conjunto de superposiciones de realimentación de voz inmediatas, cada una de las cuales proporciona información de realimentación simple, no textual sobre un estado de dicho sistema de comunicación;

20 código de programa para visualizar un conjunto de superposiciones de ayuda en dicho dispositivo, cada una de las cuales proporciona una lista de órdenes activadas por voz usadas con frecuencia sensibles al contexto para cada pantalla de cada aplicación activada por voz;

25 código de programa para visualizar un conjunto de superposiciones de realimentación en dicho dispositivo, cada una de las cuales proporciona información sobre un problema que experimenta dicho sistema de comunicación, en el que dicho conjunto de superposiciones de realimentación comprende el reconocimiento de superposiciones de realimentación existentes en tres estados que corresponden a un número de reconocimientos no satisfactorios que se producen de manera secuencial, en el que los tres estados ofrecen de manera progresiva más información de realimentación a medida que aumenta el número de reconocimientos no satisfactorios, en el que el primer estado se visualiza tras un primer reconocimiento no satisfactorio, en el que el segundo estado se visualiza tras un segundo reconocimiento no satisfactorio y en el que el tercer estado se visualiza tras un tercer reconocimiento no satisfactorio; y

30 código de programa para visualizar una superposición de menú principal en dicho dispositivo que muestra una lista de servicios disponibles para el usuario, pudiendo accederse a cada uno de dichos servicios mediante orden hablada.

35 2. Medio de almacenamiento legible por ordenador según la reivindicación 1, en el que el producto de programa informático comprende además código de programa para proporcionar un centro de usuario que proporciona cualquiera de:

40 formación y tutoriales sobre cómo usar dicho sistema de comunicaciones;

más ayuda con aplicaciones activadas por voz específicas;

45 gestión de cuentas de usuario; y

ajustes y preferencias de usuario para dicho sistema de comunicaciones.

50 3. Medio de almacenamiento legible por ordenador según la reivindicación 1, en el que cada una de dichas superposiciones de realimentación de voz inmediatas proporciona información de realimentación simple, no textual sobre un estado de dicho sistema de comunicaciones, siendo dicho estado cualquiera de:

escuchar la orden hablada del usuario;

55 alerta de no habilitada por voz;

procesamiento de reconocimiento de voz;

60 alerta de aplicación;

reconocimiento de voz positivo; y

reconocimiento de voz no satisfactorio.

65 4. Medio de almacenamiento legible por ordenador según la reivindicación 2, en el que puede accederse a cada una de dichas superposiciones de ayuda en todo momento.

5. Medio de almacenamiento legible por ordenador según la reivindicación 2, en el que dicha lista de órdenes activadas por voz proporcionada por dicha superposición de ayuda comprende cualquiera de:

- un conjunto de órdenes específicas de la aplicación;
- una orden asociada con el centro de usuario para más ayuda;
- una orden asociada con dicha visualización de menú principal; y
- una orden para hacer que desaparezca dicha superposición.

6. Medio de almacenamiento legible por ordenador según la reivindicación 2, en el que dicho conjunto de superposiciones de realimentación comprende cualquiera de:

- un conjunto de superposiciones de realimentación de reconocimiento que informa al usuario de una situación relacionada con el reconocimiento; y
- un conjunto de superposiciones de aplicación que informan al usuario de un error o un problema relacionado con una aplicación usada en dicha GSUI.

7. Medio de almacenamiento legible por ordenador según la reivindicación 6, en el que dicho conjunto de superposiciones de realimentación de reconocimiento, en respuesta a reconocimientos no satisfactorios que se suceden inmediatamente entre sí, se visualiza en tres modos diferentes que comprenden:

- un primer modo en el que dicho indicador de realimentación de voz inmediato cambia a un signo de interrogación en respuesta al primer reconocimiento no satisfactorio;
- un segundo modo en el que un mensaje de texto y un enlace a dicha superposición de ayuda se visualizan en respuesta al segundo reconocimiento no satisfactorio; y
- un tercer modo en el que un mensaje de texto, un enlace a dicha superposición de ayuda y un enlace a dicha superposición de más ayuda se visualizan en respuesta al tercer y posteriores reconocimientos no satisfactorios.

8. Medio de almacenamiento legible por ordenador según la reivindicación 1, en el que el producto de programa informático comprende además código de programa para visualizar un tratamiento de texto en pantalla que puede activarse mediante una orden hablada.

9. Medio de almacenamiento legible por ordenador según la reivindicación 8, en el que dicho texto en pantalla comprende cualquiera de:

- un texto estático usado en etiquetas para gráficos en pantalla o en teclas virtuales que pueden seleccionarse mediante un cursor; y
- un texto dinámico usado en contenido en el que una o más palabras pueden activarse mediante una orden hablada.

10. Medio de almacenamiento legible por ordenador según la reivindicación 1, en el que cualquiera de dichas superposiciones de ayuda, superposiciones de realimentación y superposición de menú principal se implementa en un cuadro de diálogo, comprendiendo dicho cuadro de diálogo cualquiera de:

- uno o más cuadros de texto para información de texto; y
- una o más teclas virtuales.

11. Medio de almacenamiento legible por ordenador según la reivindicación 10, en el que dicho cuadro de diálogo comprende además un indicador de identidad.

12. Medio de almacenamiento legible por ordenador según la reivindicación 1, en el que dicha superposición de menú principal comprende:

- una primera superposición de menú secundario específicamente para el acceso a un sistema de guía de programas interactiva que proporciona servicio de televisión por cable;
- una segunda superposición de menú secundario específicamente para el acceso a un sistema de vídeo bajo demanda que proporciona servicio de vídeo por cable; y

ES 2 323 230 T3

una tercera superposición de menú secundario específicamente para el acceso a un sistema cerrado que proporciona servicios de Internet basados en navegador;

en el que cada uno de dichos menús secundarios proporciona un conjunto de teclas virtuales activadas por voz.

13. Medio de almacenamiento legible por ordenador según la reivindicación 1, que comprende además un mecanismo de personalización e identificación de emisor de voz que permite a un usuario entrenar a dicho sistema de comunicaciones con aproximadamente cuarenta segundos de voz e identifica al usuario por la voz.

14. Método para proporcionar información de ayuda visualizando un conjunto de superposiciones en la pantalla de un usuario en un sistema de comunicaciones habilitado por voz para facilitar un servicio de información digital, incluyendo dicho sistema de comunicaciones un dispositivo de visualización, un módulo descodificador de televisión, un sistema de entrada de voz y un terminal de cabecera, en el que dicho usuario activa dicho sistema de entrada de voz activando un interruptor asociado con el funcionamiento de un dispositivo de entrada de voz, comprendiendo dicho método las etapas implementadas por ordenador de:

recibir entrada de voz usando un receptor;

transferir dicha entrada de voz desde dicho receptor a dicho módulo descodificador de televisión;

transferir dicha entrada de voz desde dicho módulo descodificador de televisión a dicho terminal de cabecera;

reconocer dicha entrada de voz y transformar dicha entrada de voz en órdenes en dicho terminal de cabecera;

generar información de realimentación;

transferir dicha información de realimentación desde dicho terminal de cabecera a dicho dispositivo de visualización;

visualizar una primera superposición de ayuda en dicho dispositivo de visualización si se reconoce satisfactoriamente una orden de ayuda, permaneciendo dicha superposición de ayuda en dicho dispositivo de visualización durante un intervalo específico;

eliminar dicha primera superposición de ayuda de dicho dispositivo de visualización si sucede cualquiera de los siguientes:

transcurre dicho intervalo específico;

se activa accidentalmente cualquier tecla de dicho dispositivo de entrada de voz; y

se selecciona una tecla de salida incorporada en dicha primera superposición de ayuda;

visualizar una segunda superposición de ayuda mientras dicho interruptor está activado para introducir una nueva orden hablada;

en el que la información de realimentación comprende un conjunto de superposiciones de realimentación de reconocimiento existentes en tres estados que corresponden a un número de reconocimientos no satisfactorios que se producen de manera secuencial, en el que los tres estados ofrecen progresivamente más información de realimentación a medida que aumenta el número de reconocimientos no satisfactorios, y

en el que la etapa de generar información de realimentación comprende:

hacer un seguimiento del número de reconocimientos no satisfactorios de entrada de voz seguidos,

visualizar un primer estado de dicha superposición de realimentación de reconocimiento tras un primer reconocimiento no satisfactorio,

visualizar un segundo estado de dicha superposición de realimentación de reconocimiento tras un segundo reconocimiento no satisfactorio, y

visualizar un tercer estado de dicha superposición de realimentación de reconocimiento tras un tercer reconocimiento no satisfactorio.

15. Método según la reivindicación 14, que comprende además las etapas adicionales de:

(a) comprobar si una pantalla actual está habilitada por voz cuando se activa dicho interruptor;

- (b) si la pantalla actual está habilitada por voz, visualizar una primera pestaña que señala que un sistema de entrada de voz está activado;
- (c) si la pantalla actual no está habilitada por voz, visualizar una segunda pestaña que señala una alerta de no habilitada por voz, permaneciendo dicha segunda pestaña en pantalla durante un primer intervalo; y
- (d) si dicho interruptor vuelve a activarse, repetir la etapa (a).

16. Método según la reivindicación 15, en el que dicha etapa (b) comprende además las etapas de:

- (e) si dicho interruptor no se desactiva dentro de un segundo intervalo, interrumpir el reconocimiento;
- (f) si dicho interruptor se desactiva después de que transcurra un tercer intervalo pero antes de que transcurra dicho segundo intervalo en la etapa (e), visualizar una tercera pestaña que señala que el reconocimiento de voz está en procesamiento; y
- (g) si dicho interruptor se desactivó antes de que transcurriera dicho tercer intervalo en la etapa (f), eliminar cualquier pestaña en la pantalla.

17. Método según la reivindicación 16, en el que dicha etapa (f) comprende además las etapas de:

- (h) si dicho módulo descodificador de televisión tarda más de un cuarto intervalo medido desde el momento en que el usuario libera dicho interruptor hasta el momento en que se envía el último dato de voz a dicho terminal de cabecera, interrumpir el procesamiento de reconocimiento de voz y visualizar una cuarta pestaña que señala una alerta de aplicación, permaneciendo dicha cuarta pestaña en la pantalla durante un quinto intervalo; y
- (i) si se pulsa una tecla de mando a distancia diferente de dicho interruptor mientras está procesándose una orden hablada, interrumpir el procesamiento de reconocimiento de voz y eliminar cualquier pestaña en la pantalla.

18. Método según la reivindicación 17, en el que dicha etapa (h) comprende además las etapas de:

- (j) si dicho interruptor vuelve a activarse mientras dicha cuarta pestaña está en la pantalla, eliminar la cuarta pestaña y repetir la etapa (a); y
- (k) cuando transcurre dicho quinto intervalo o si se pulsa una tecla de mando a distancia diferente de dicho interruptor mientras dicha cuarta pestaña está en la pantalla, eliminar dicha cuarta pestaña.

19. Método según la reivindicación 16, en el que dicha etapa (f), una vez completado el reconocimiento, comprende además las etapas de:

- (l) comprobar si el reconocimiento de voz es satisfactorio;
- (m) si el reconocimiento de voz es satisfactorio, visualizar una quinta pestaña que señala un reconocimiento de voz positivo, permaneciendo dicha quinta pestaña en la pantalla durante aproximadamente un segundo; y
- (n) si dicho interruptor vuelve a activarse antes de que desaparezca dicha quinta pestaña, repetir la etapa (a).

20. Método según la reivindicación 14, en el que dicha etapa (l) comprende además las etapas de:

- (o) restablecer dicho número a cero después de cada reconocimiento satisfactorio o cuando se pulse cualquier tecla de dicho dispositivo de mando a distancia;
- (p) si el reconocimiento completo es el primer reconocimiento no satisfactorio, visualizar una sexta pestaña que señala voz mal reconocida, permaneciendo dicha sexta pestaña en la pantalla durante aproximadamente un segundo; y
- (q) si dicho interruptor vuelve a pulsarse antes de que desaparezca dicha sexta pestaña, repetir la etapa (a).

21. Método según la reivindicación 20, en el que dicha etapa (o) comprende además las etapas de:

- (r) si el reconocimiento completo es el segundo reconocimiento no satisfactorio, visualizar una primera variante de dicha sexta pestaña que señala una voz mal reconocida y visualizar un breve mensaje de texto, permaneciendo dicha primera variante de dicha sexta pestaña en la pantalla durante aproximadamente diez segundos; y

ES 2 323 230 T3

- (s) si dicho interruptor vuelve a pulsarse antes de que desaparezca dicha primera variante de dicha sexta pestaña, repetir la etapa (a).

22. Método según la reivindicación 20, en el que dicha etapa (o) comprende además las etapas de:

- (t) si el reconocimiento completo es el tercer reconocimiento no satisfactorio, visualizar una segunda variante de dicha sexta pestaña que señala voz mal reconocida y visualizar un mensaje de texto largo, permaneciendo dicha segunda variante de dicha sexta pestaña en la pantalla durante aproximadamente diez segundos; y
- (u) si dicho interruptor vuelve a activarse antes de que desaparezca dicha segunda variante de dicha sexta pestaña, repetir la etapa (a).

23. Método según la reivindicación 16, en el que dicha etapa (e) comprende además las etapas de:

- (v) visualizar una primera variante de dicha cuarta pestaña, permaneciendo dicha primera variante en la pantalla durante un sexto intervalo;
- (w) eliminar dicha primera variante de dicha cuarta pestaña de la pantalla si dicho interruptor se desactiva después de que transcurra dicho sexto intervalo; y (x) visualizar una segunda variante de dicha cuarta pestaña, permaneciendo dicha segunda variante en la pantalla hasta que se desactive dicho interruptor.

24. Sistema de interfaz de televisión interactiva habilitada por voz, que comprende:

un dispositivo de interconexión que conecta un equipo de televisión con un proveedor de servicios de televisión;

un dispositivo de mando a distancia habilitado por voz que transforma las órdenes habladas de un usuario en señales aceptables para dicho dispositivo de interconexión; y

medios para visualizar un conjunto de superposiciones de ayuda, cada una de las cuales proporciona una lista sensible al contexto de órdenes activadas por voz usadas con frecuencia para cada pantalla de cada aplicación activada por voz; y

medios para visualizar un conjunto de superposiciones de realimentación, cada una de las cuales proporciona información sobre un problema que experimenta dicho sistema de comunicaciones,

en el que dicho conjunto de superposiciones de realimentación comprende superposiciones de realimentación de reconocimiento existentes en tres estados que corresponden a un número de reconocimientos no satisfactorios que se producen de manera secuencial, en el que los tres estados ofrecen progresivamente más información de realimentación a medida que aumenta el número de reconocimientos no satisfactorios, en el que el primer estado se visualiza tras un primer reconocimiento no satisfactorio, en el que el segundo estado se visualiza tras un segundo reconocimiento no satisfactorio y en el que el tercer estado se visualiza tras un tercer reconocimiento no satisfactorio.

25. Sistema según la reivindicación 24, en el que dichos medios para visualizar proporcionan realimentación visual en tiempo real inmediata que indica varios estados de actividades de reconocimiento de voz.

26. Sistema según la reivindicación 24, en el que cada uno de dicho conjunto de superposiciones proporciona información de realimentación simple, no textual sobre un estado de actividades de reconocimiento de voz, siendo dicho estado cualquiera de:

recibir emisión de voz hablada;

procesar la emisión de voz;

reconocimiento satisfactorio;

reconocimiento no satisfactorio; y

orden no autorizada.

27. Sistema según la reivindicación 24, en el que dichos medios para visualizar proporcionan realimentación de ayuda creciente cuando la orden hablada del usuario no se reconoce con un grado de confianza predefinido.

28. Sistema según la reivindicación 27, en el que dicha realimentación de ayuda creciente comprende un conjunto de superposiciones de realimentación para revelar información de ayuda progresiva.

29. Sistema según la reivindicación 28, en el que cada una de dichas superposiciones de realimentación proporciona una lista sensible al contexto de órdenes habilitadas por voz usadas con frecuencia para cada pantalla.

ES 2 323 230 T3

30. Sistema según la reivindicación 29, en el que dicha lista de órdenes habilitadas por voz usadas con frecuencia comprende cualquiera de:

un conjunto de órdenes específicas de la aplicación;

una orden asociada con un centro de usuario para más información de ayuda;

una orden asociada con una visualización de menú principal; y

una orden para hacer que desaparezca dicha superposición de la pantalla.

31. Sistema según la reivindicación 24, en el que cuando la orden hablada del usuario no se reconoce con un grado de confianza predefinido, dichos medios para visualizar presentan una lista de órdenes predichas que piden al usuario que haga una selección de dicha lista.

32. Sistema según la reivindicación 24, que comprende además:

medios para navegar por la información basada en la lista en pantalla a través de órdenes habladas.

33. Sistema según la reivindicación 24, que comprende además:

medios para permitir al usuario navegar directamente entre aplicaciones a través de órdenes habladas o un menú habilitado por voz.

34. Sistema según la reivindicación 24, que comprende además:

medios para permitir al usuario navegar directamente a páginas previamente marcadas a través de órdenes habladas.

FIG. 1

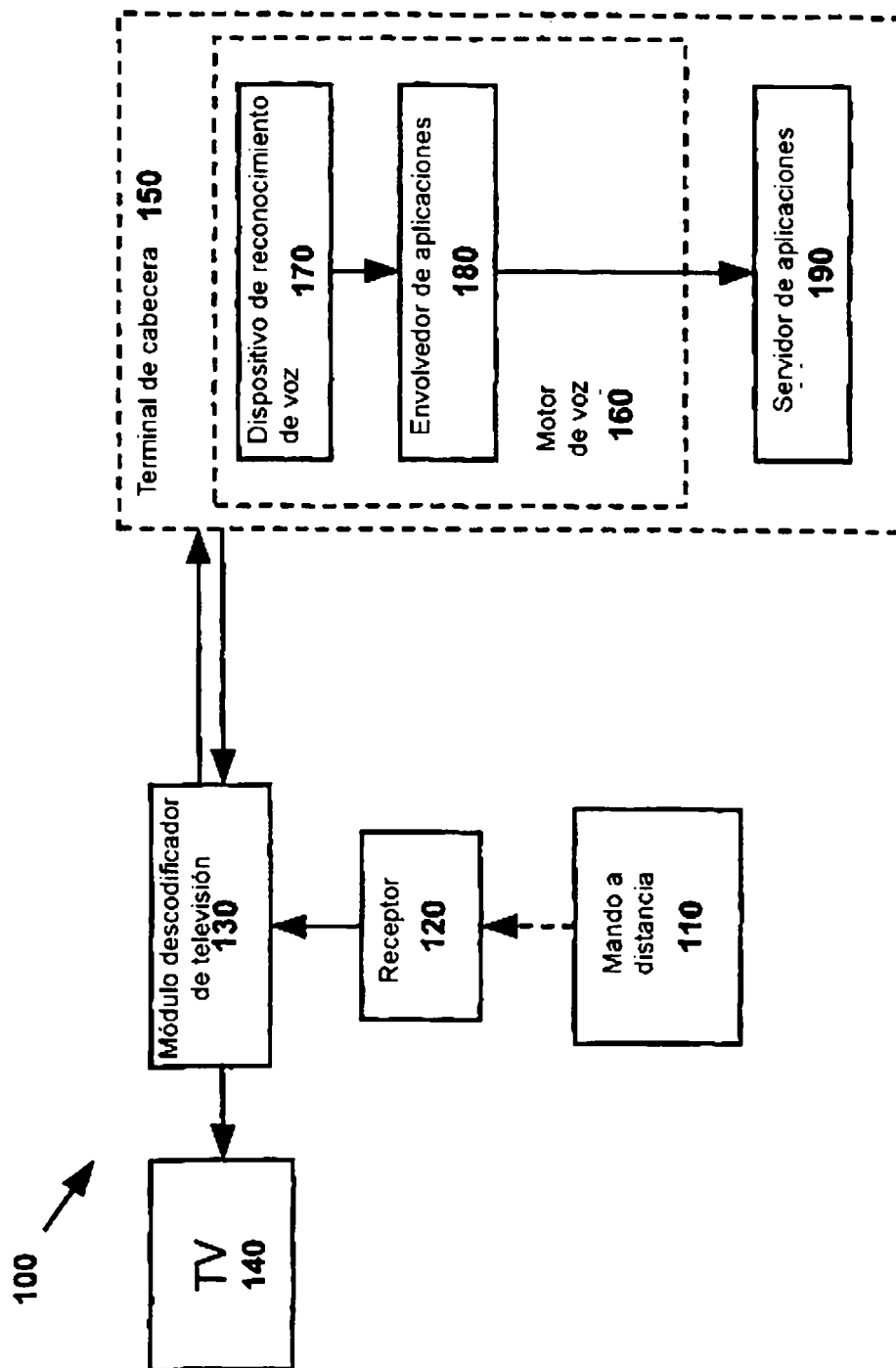


FIG. 2A

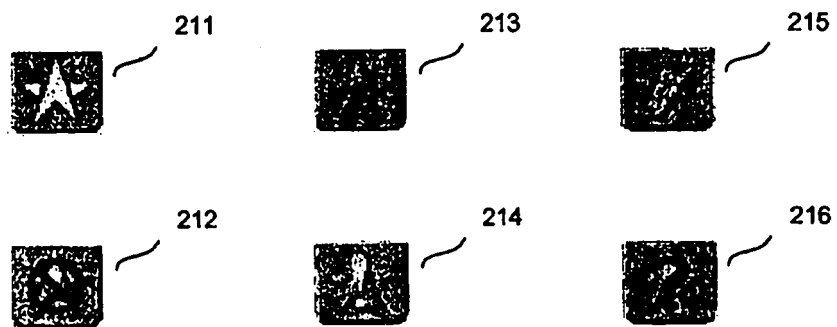


FIG. 2B

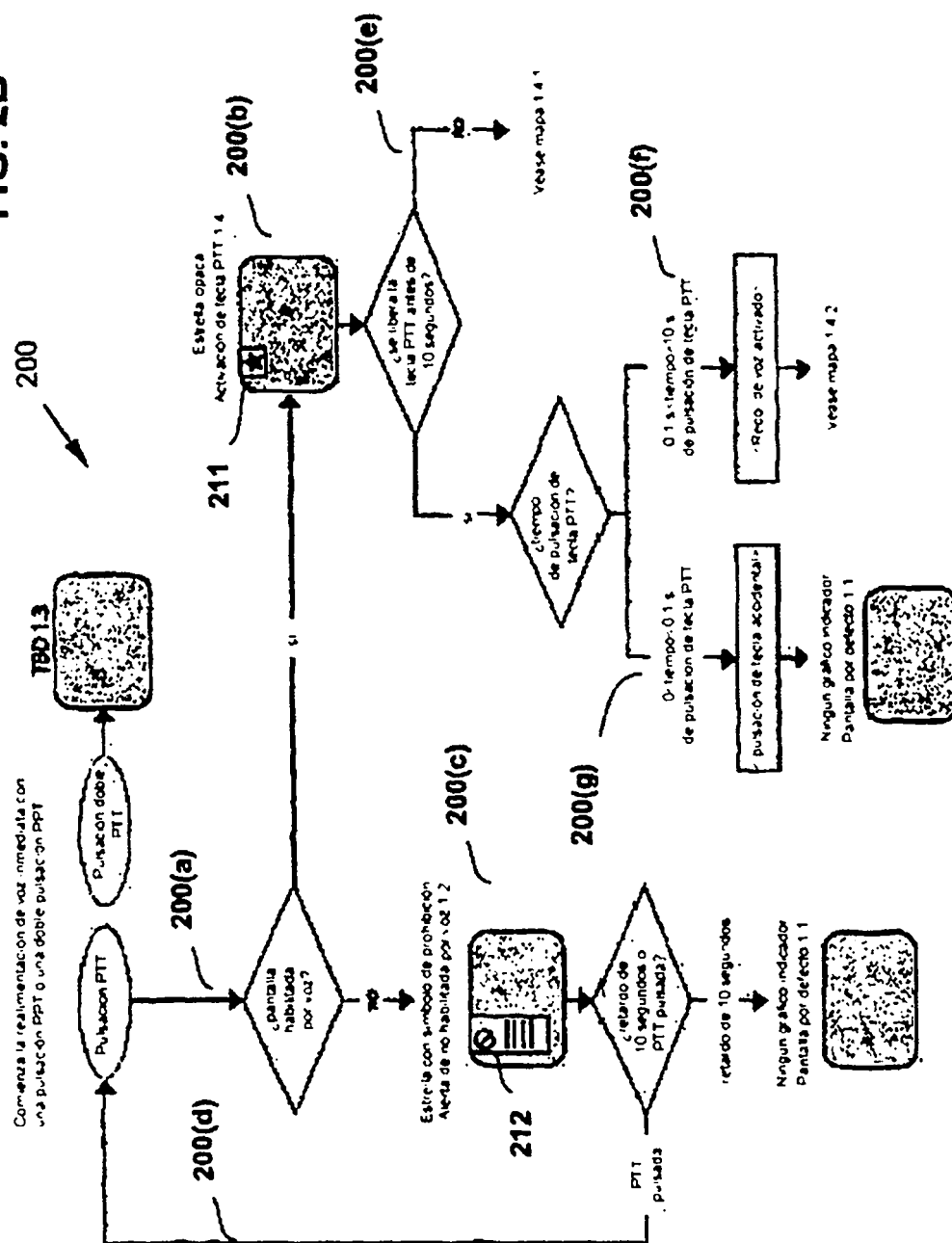
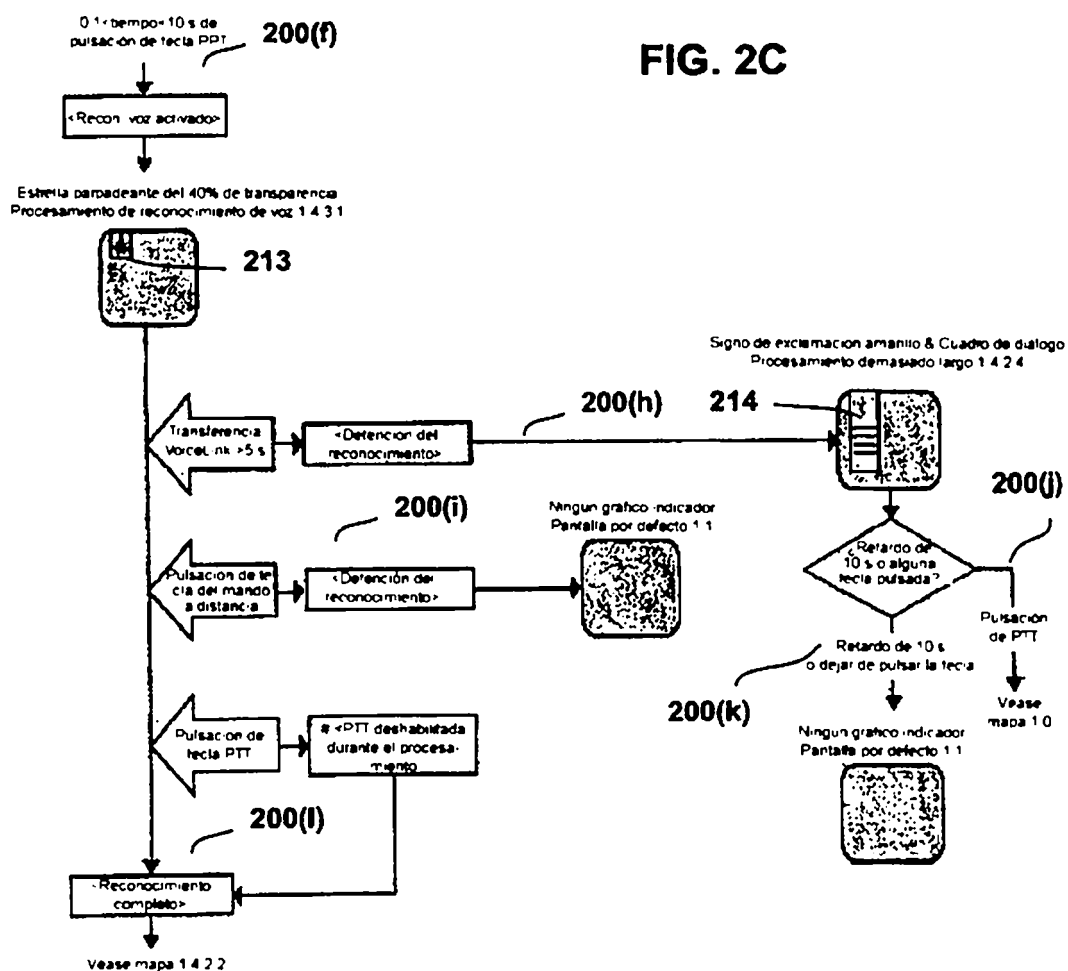


FIG. 2C



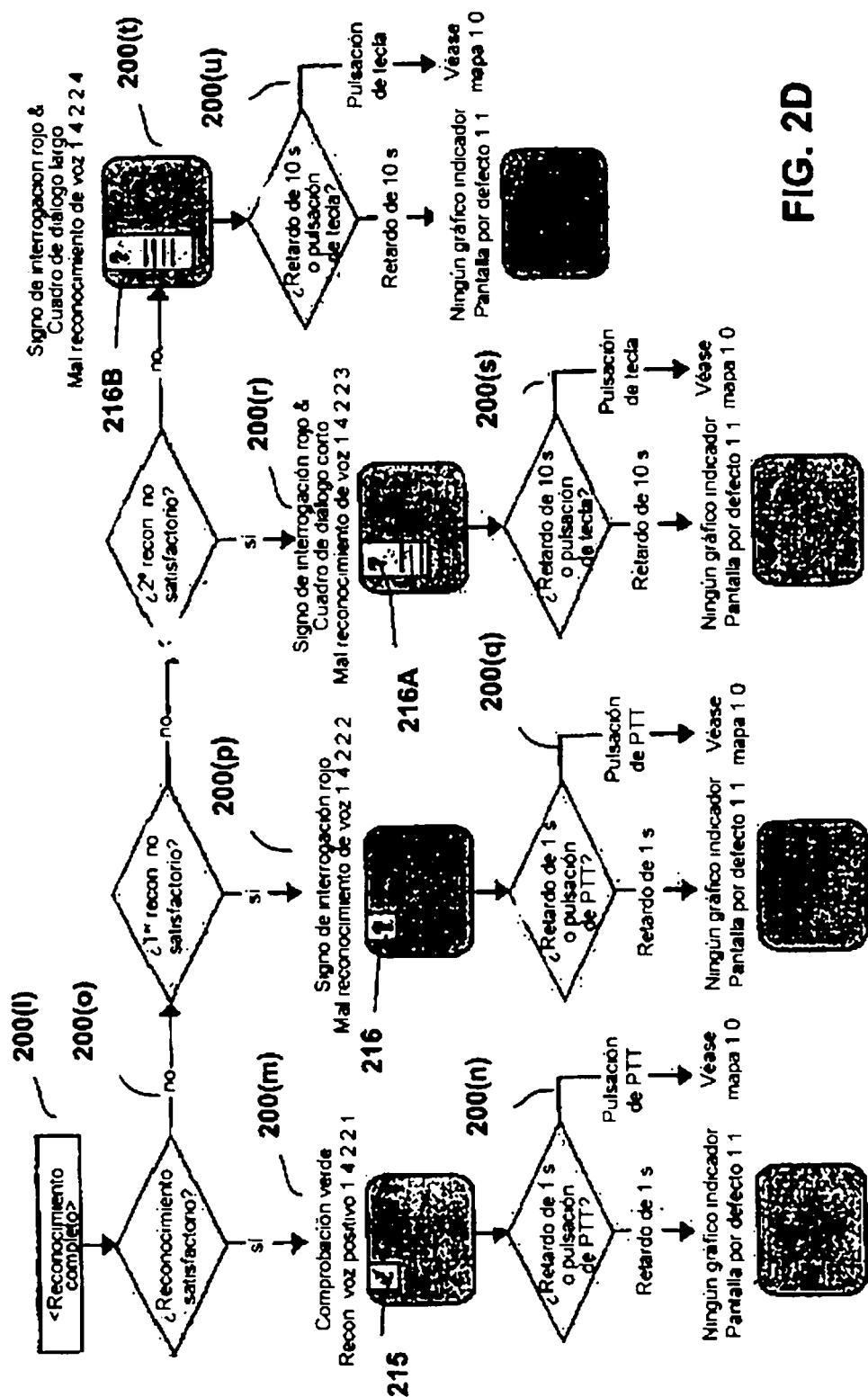


FIG. 2D

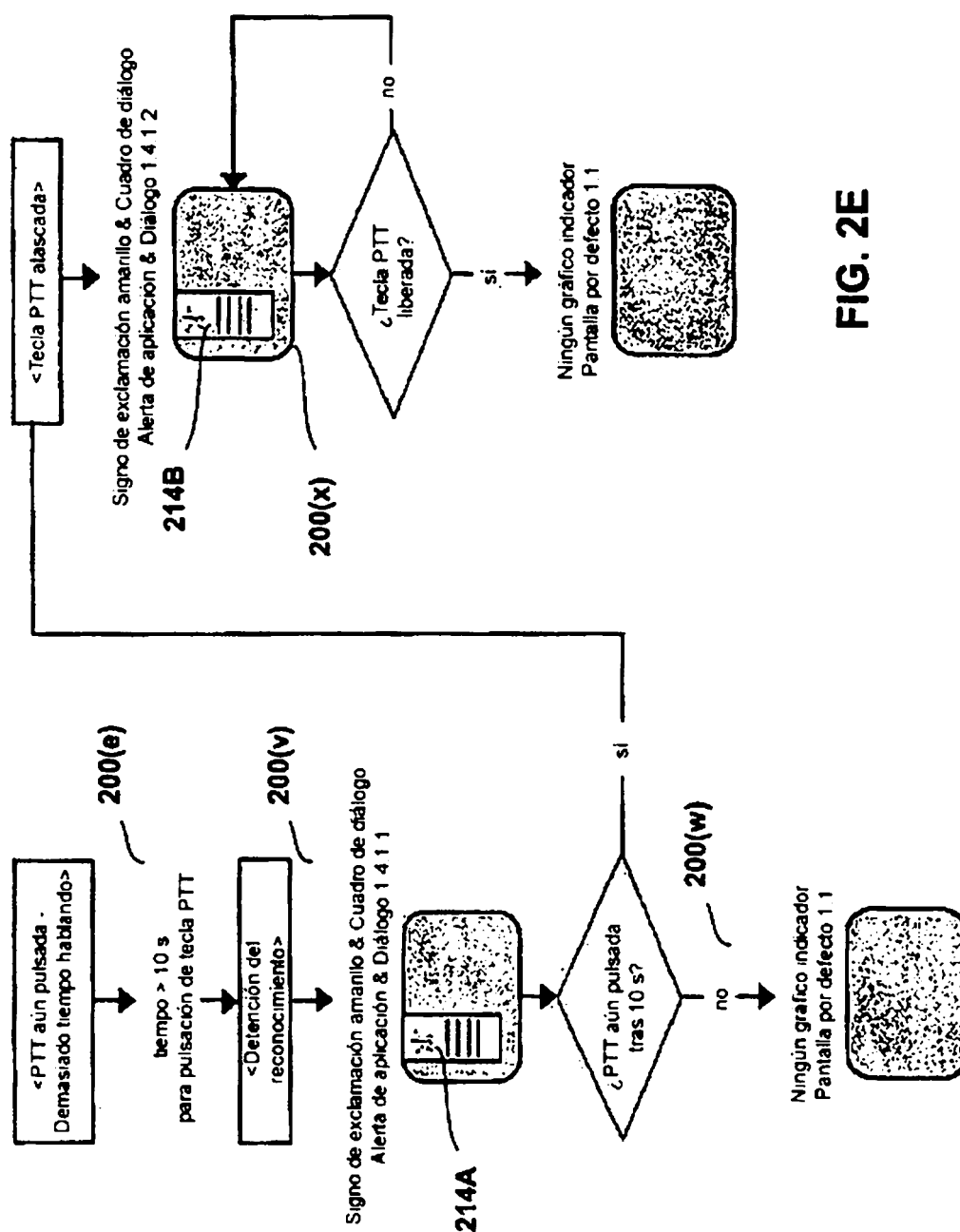


FIG. 2E

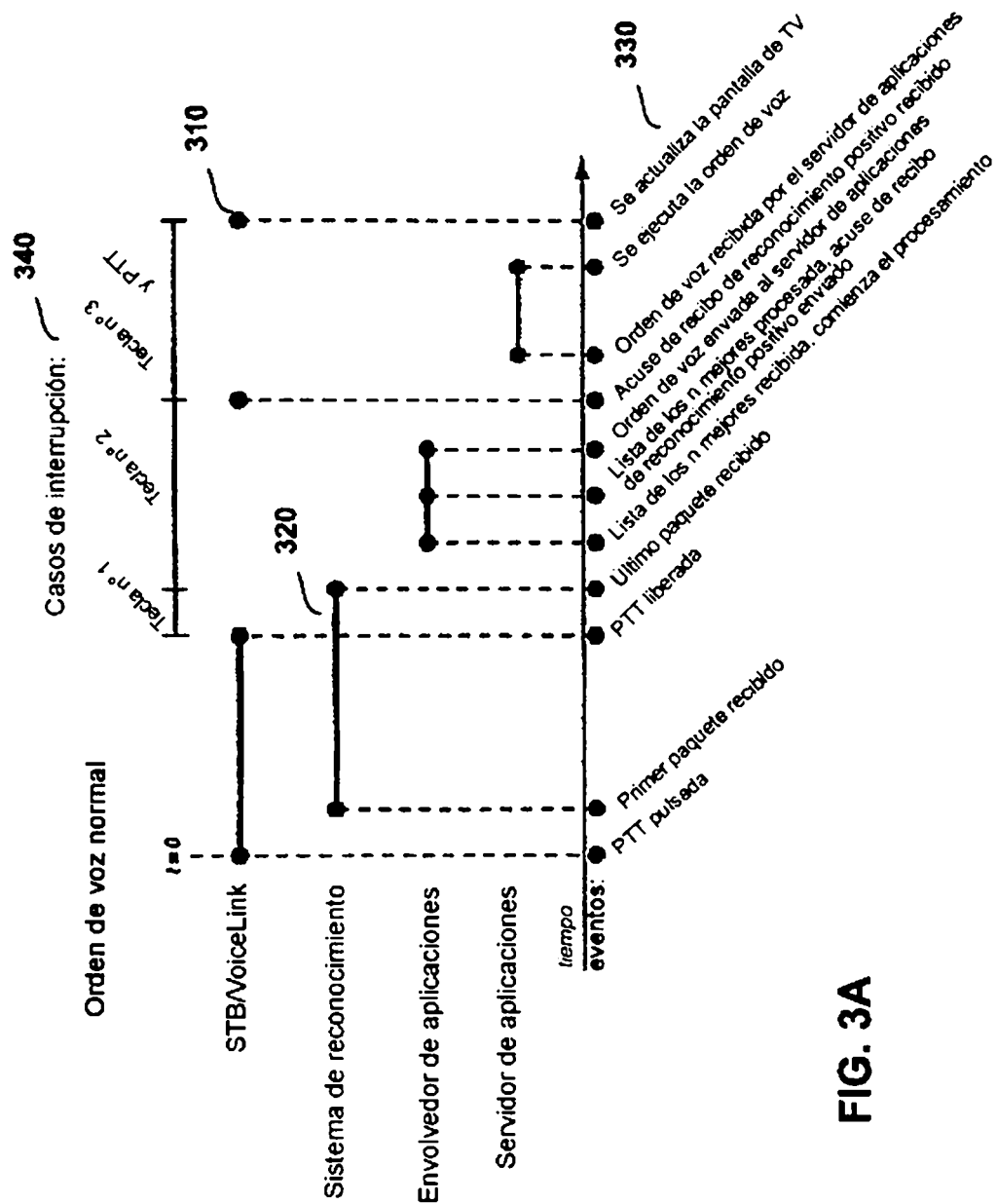


FIG. 3A

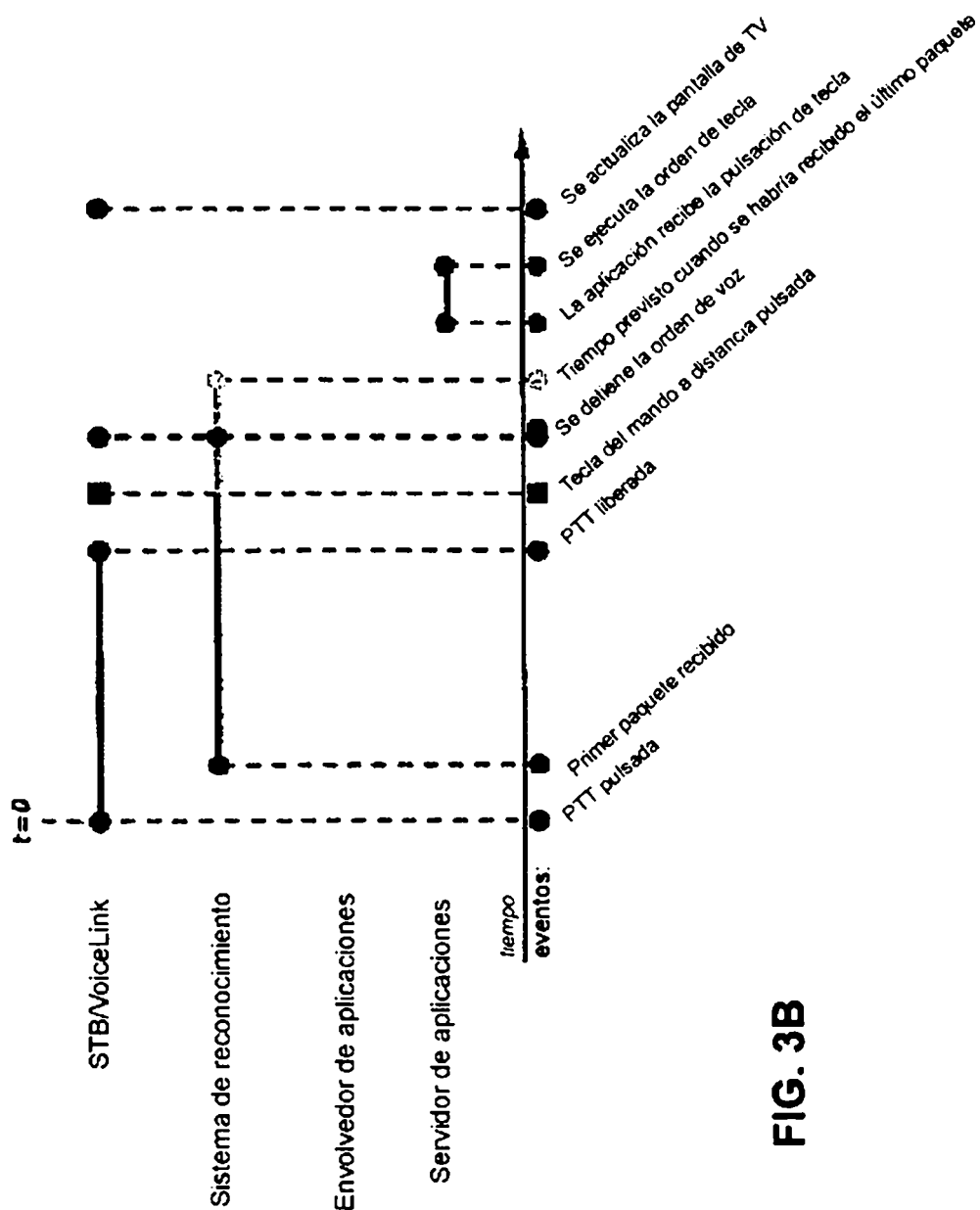


FIG. 3B

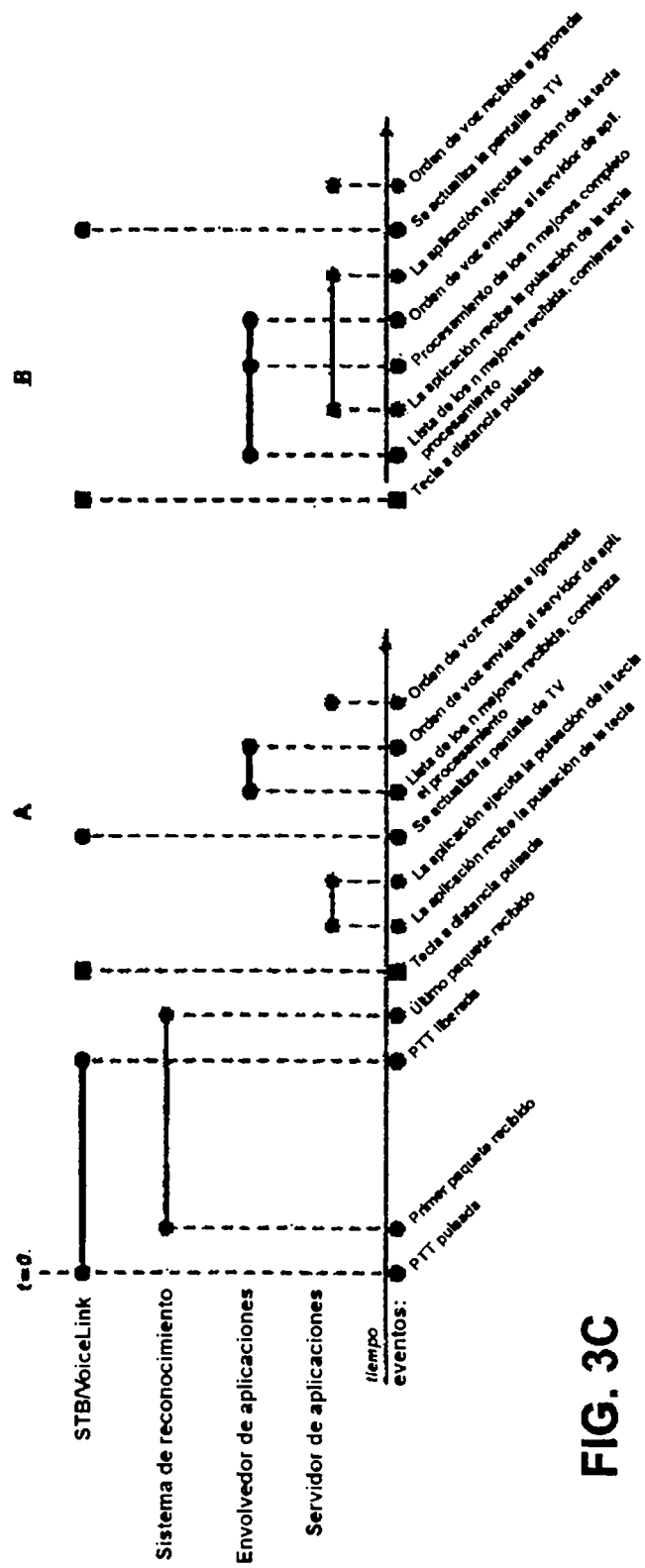
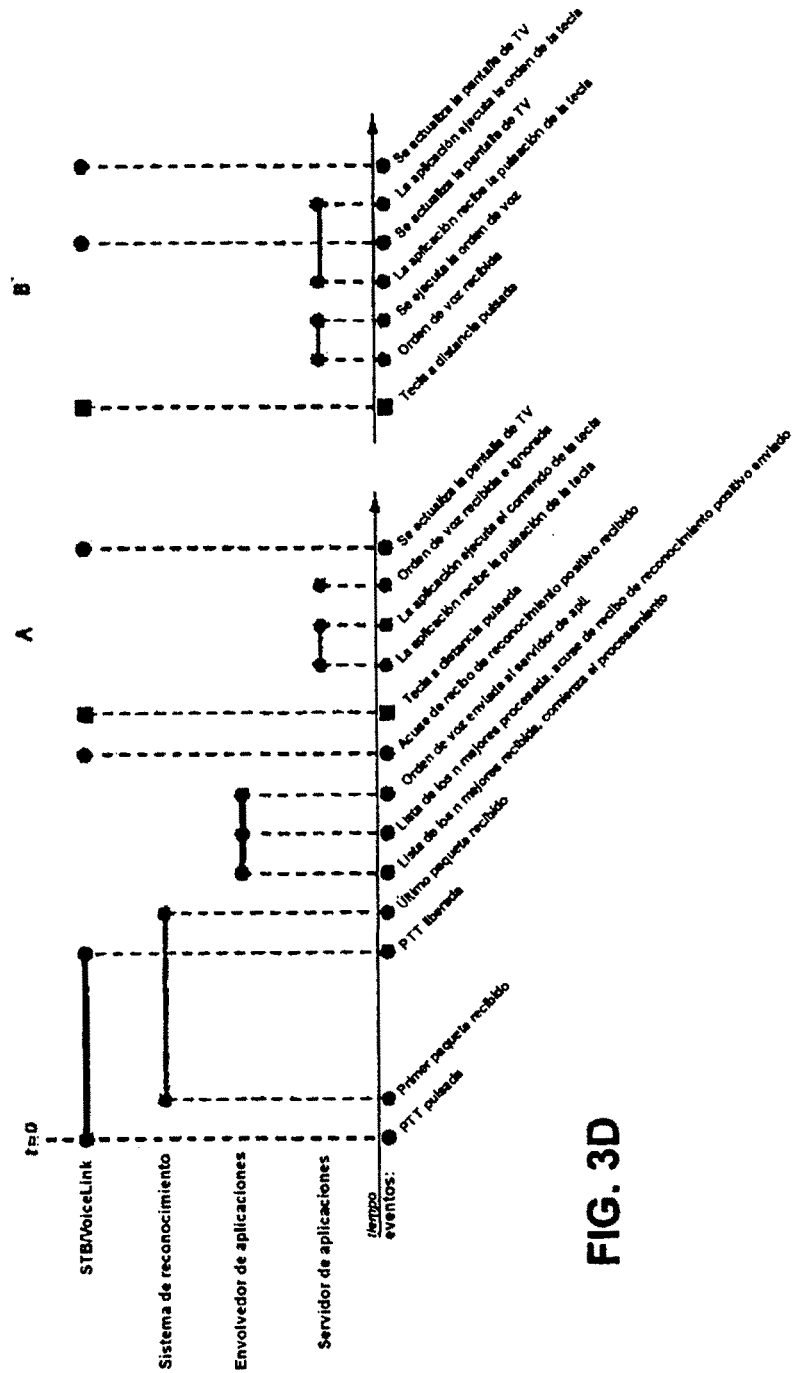


FIG. 3C



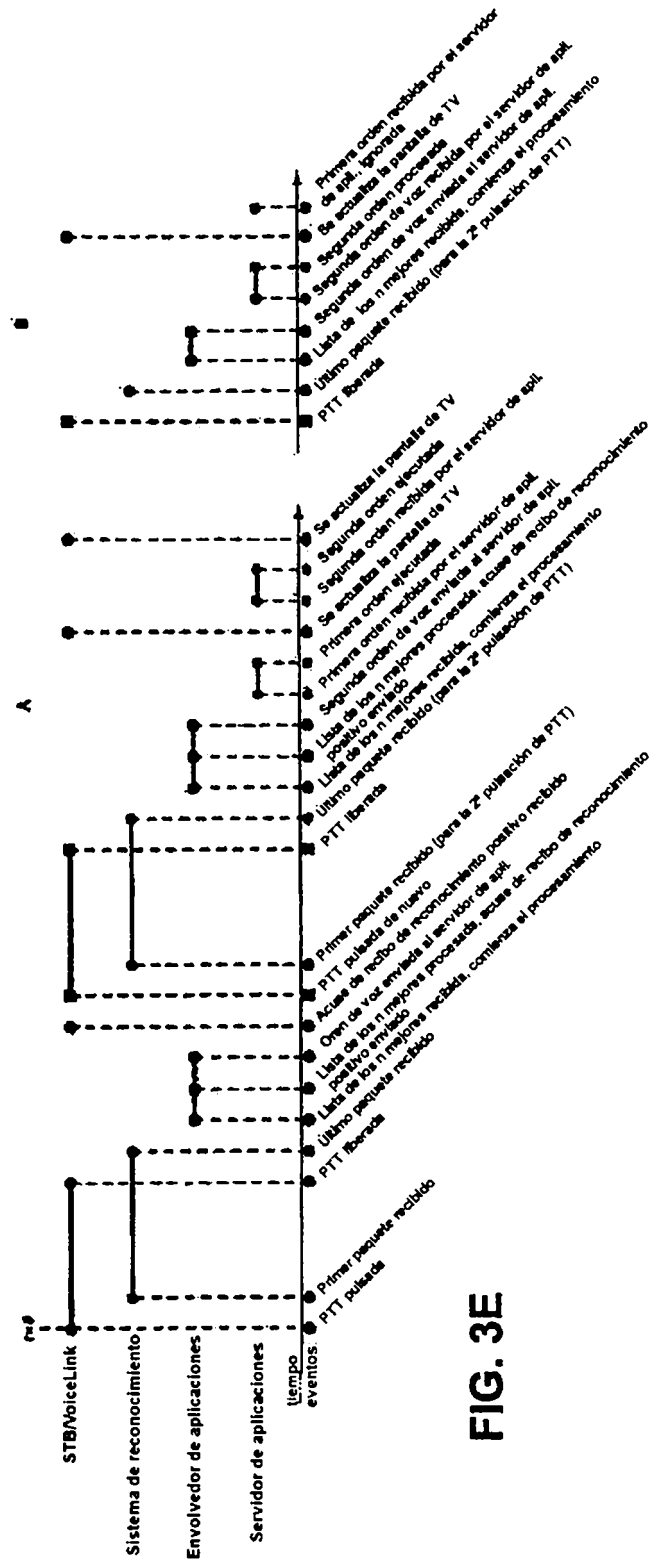
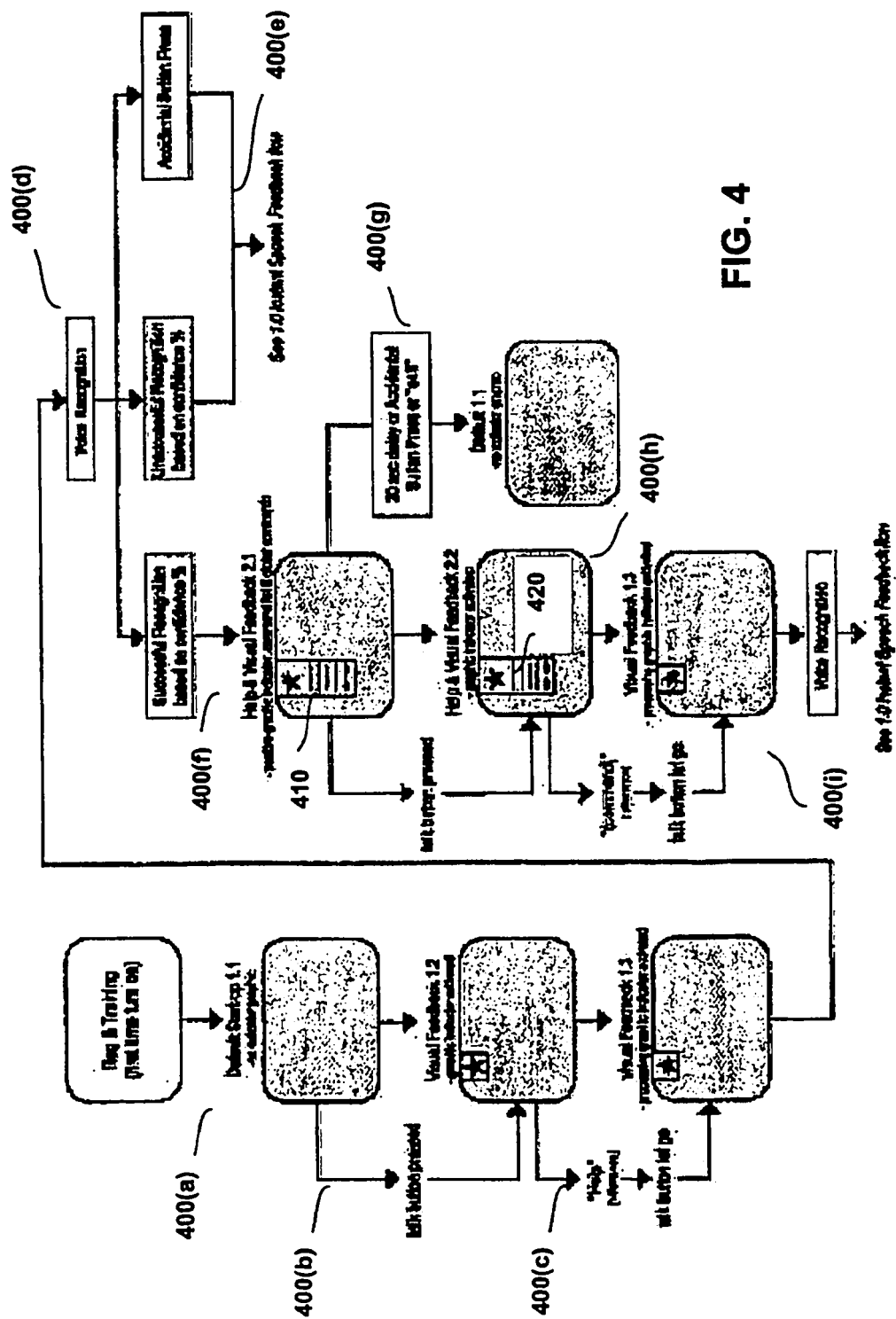


FIG. 3E



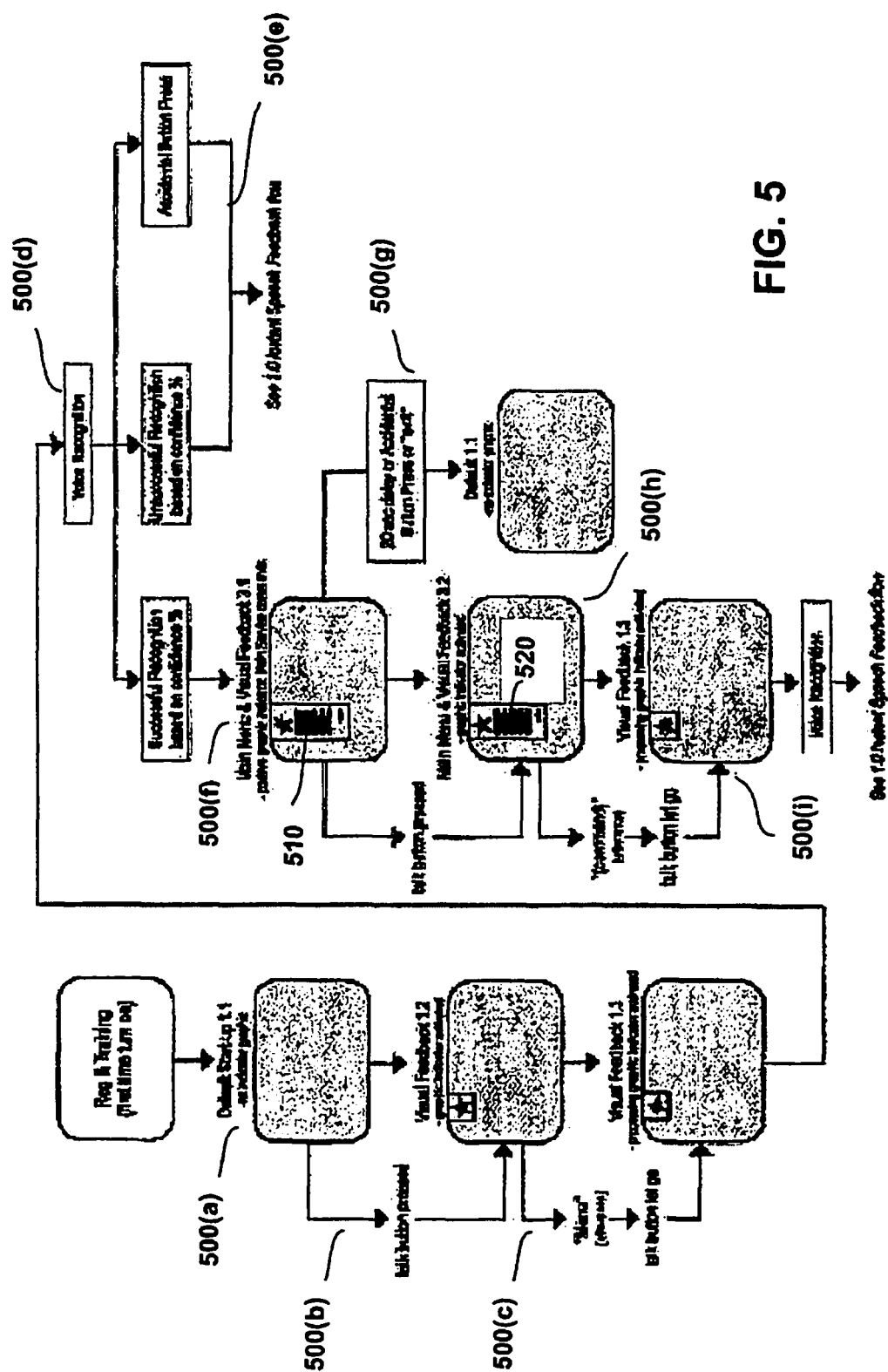


FIG. 5

FIG. 6A

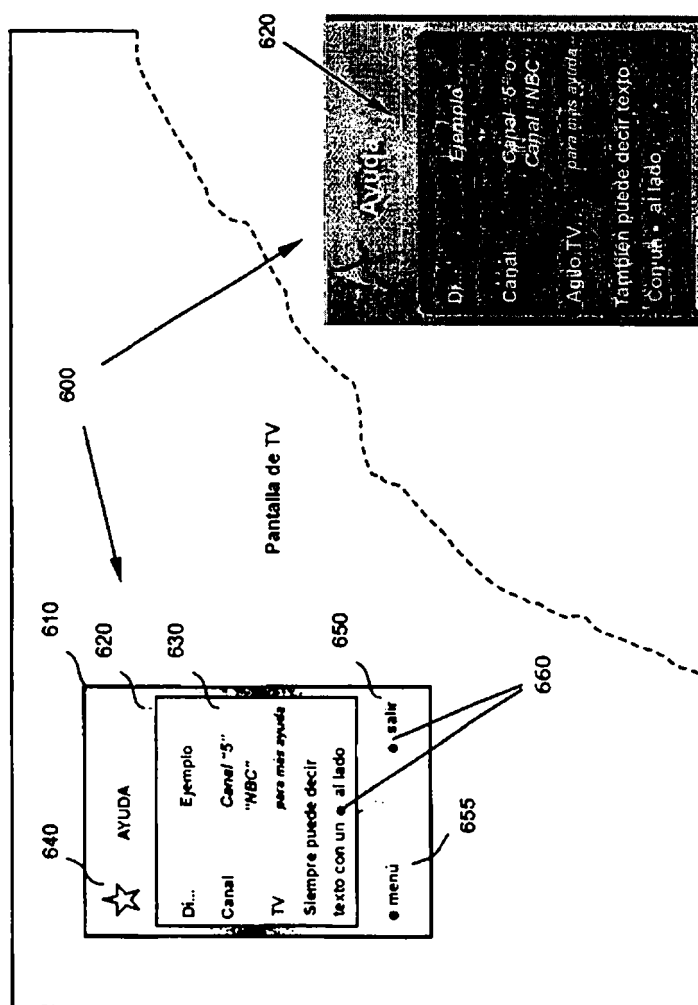


FIG. 6B