

TRADUCCIÓN DURANTE UNA LLAMADA

Antecedentes de la invención

Los sistemas de comunicación permiten a los usuarios comunicarse
5 entre sí a través de una red de comunicación, por ejemplo, mediante la
realización de una llamada por la red. La red puede ser, por ejemplo,
Internet o la red telefónica pública conmutada (PSTN por sus siglas en
inglés). Durante una llamada, las señales de audio y/o video se pueden
transmitir entre los nodos de la red, permitiendo a los usuarios transmitir y
10 recibir datos de audio (como un discurso) y/o datos de vídeo (como cámara
de vídeo) entre sí en una sesión de comunicación período a través de la
red de comunicación.

Estos sistemas de comunicación incluyen sistemas de protocolo de
voz o video a través de Internet (VoIP). Para utilizar un sistema de VoIP,
15 un usuario instala y ejecuta el software del cliente en un dispositivo de
usuario. El software de cliente establece conexiones de VoIP, así como
proporcionar otras funciones tales como el registro y la autenticación de
usuario. Además de la comunicación de voz, el cliente también puede
configurar las conexiones para los modos de comunicación, por ejemplo
20 para proporcionar mensajería instantánea ("MI"), mensajería SMS,
transferencia de archivos y servicios de correo de voz a los usuarios.

Breve descripción de la invención

La presente breve descripción se proporciona para introducir una
selección de conceptos en una forma simplificada que se describe más
25 adelante en la descripción detallada. Esta breve descripción no pretende

identificar las características claves o características esenciales de la materia reclamada, tampoco está destinada a ser utilizada para limitar el alcance de la materia reclamada.

De acuerdo con un primer aspecto, se describe un método
5 implementado por computadora que se lleva a cabo en un sistema de comunicación. El sistema de comunicación es destinado para efectuar una llamada de voz o vídeo entre al menos un usuario fuente que habla una lengua fuente y un usuario de destino que habla un idioma de destino. La llamada de audio se recibe, el audio de llamada comprende un discurso del
10 usuario fuente en el idioma fuente. Un procedimiento de traducción se lleva a cabo en el audio de la llamada para generar una traducción de audio del discurso del usuario fuente en el idioma de destino para emitirse al usuario de destino. Se señala un cambio en el comportamiento del procedimiento de traducción, el cambio relacionado con la generación de
15 la traducción, provocando de esta manera una notificación a ser emitida al usuario de destino para notificar al usuario de destino sobre el cambio.

De acuerdo con un segundo aspecto, se describe un sistema informático para su uso en un sistema de comunicación. El sistema de comunicación es para efectuar una llamada de voz o vídeo entre al menos
20 un usuario de origen que habla un idioma de origen y un usuario de destino que habla un idioma de destino. El sistema informático comprende uno o más componentes de salida de audio disponibles para el usuario de destino, un componente de emisión de la traducción y un componente de emisión de notificación. El componente de emisión de la traducción está
25 configurado para emitir una traducción de audio del discurso del usuario de

origen en el idioma de destino al usuario de destino a través de los componentes de emisión de audio. La traducción se genera mediante la realización de un procedimiento de traducción automática en el audio de llamada de la llamada que comprende el discurso del usuario de origen en el idioma de origen. El componente de emisión de notificación está configurado para emitir una notificación al usuario de destino para notificar al usuario de destino de un cambio en el comportamiento del procedimiento de traducción, el cambio está relacionado con la generación de la traducción.

De acuerdo con el tercer aspecto, se describe un producto de programa informático que comprende un código informático almacenado en un medio de almacenamiento legible configurado, cuando se ejecuta, para implementar cualquiera de los métodos o sistemas descritos en el presente documento.

Breve descripción de las figuras

Para una mejor comprensión de la materia y para mostrar cómo la misma se puede llevar a cabo, se hará referencia a continuación, por medio de ejemplo, solo a las siguientes figuras en las cuales:

La figura 1 es una ilustración esquemática de un sistema de comunicación;

La figura 2 es un diagrama de bloques esquemático de un dispositivo de usuario;

La figura 3 es un diagrama de bloques esquemático de un servidor;

La figura 4A es un diagrama de bloques de funciones que muestra la funcionalidad del sistema de comunicación;

La figura 4B es un diagrama de bloques de funciones que muestra algunos de los componentes de la figura 4A;

La Figura 5 es un diagrama de flujo de un método para facilitar la comunicación entre los usuarios como parte de una llamada;

5 La figura 6 es un diagrama de flujo de un método para operar un avatar de traducción a ser desplegado en la interfaz del usuario del cliente;

Las figuras 7A a 7E ilustran esquemáticamente el comportamiento del avatar de traducción en varios escenarios ejemplares;

10 La Figura 8 es un diagrama de bloques de funciones de un sistema de traducción basado en una notificación.

Descripción detallada

Ahora se describirán las modalidades por medio de ejemplo únicamente.

15 Se hace primero referencia a la figura 1, que ilustra un sistema de comunicación 100 que es un sistema de comunicación basado en paquetes en esta modalidad, pero que puede no estar basado en paquetes en otras modalidades. Un primer usuario 102a del sistema de comunicación (usuario A o "Alice") opera un dispositivo de usuario 104a, que se muestra
20 conectado a una red de comunicaciones 106. La primera usuario (Alice) también se denomina más adelante como un "usuario de origen" por razones que serán evidentes. La red de comunicaciones 106 puede ser por ejemplo Internet. El dispositivo de usuario 104a está dispuesto para recibir información de y la información de salida al usuario 102a del dispositivo.

25 El dispositivo de usuario 104 ejecuta un cliente de comunicación

118a, proporcionado por un proveedor de software asociado con el sistema de comunicación 100. El cliente de comunicación 108a es un programa de software ejecutado en un procesador local en el dispositivo de usuario 104 que permite que el dispositivo de usuario 104a establezca los eventos de comunicación - como llamadas de audio, llamadas de audio y vídeo (denominadas de forma equivalente como llamadas de vídeo), sesiones de comunicación de mensajería instantánea, etc. – en la red 106.

La figura 1 también muestra un segundo usuario 102b (usuario B o "Bob") que tiene un 104b dispositivo de usuario que ejecuta un cliente 118b para comunicarse a través de la red 106 de la misma manera que el dispositivo de usuario 104a realiza el cliente 118a para comunicarse a través de la red 106. Por lo tanto, los usuarios A y B (102a y 102b) se pueden comunicar entre sí a través de la red de comunicaciones 106. El segundo usuario (Bob) también se denomina más adelante como un "usuario de destino" por razones que estarán obvias más adelante.

Puede haber más usuarios conectados a la red de comunicaciones 106, pero para mayor claridad sólo los dos usuarios 102a y 102b se muestran conectados a la red 106 en la figura 1.

Se debe tomar en cuenta que en modalidades alternativas, los dispositivos de usuario 104a y/o 104b se pueden conectar a la red de comunicación 106 a través de redes intermedias adicionales que no se muestran en la figura 1. Por ejemplo, si uno de los dispositivos de usuario es un tipo particular de dispositivo móvil, entonces se puede conectar a la red de comunicación 106 a través de una red móvil celular (que no se muestra en la figura 1), por ejemplo una red GSM o UMTS.

Los eventos de comunicación entre Alice y Bob se pueden establecer utilizando los clientes 118a, 118b de diversas maneras. Por ejemplo, una llamada se puede ser establecida por uno de los usuarios Alice y Bob que envían una invitación de llamada al otro usuario (ya sea directamente o indirectamente por medio de una entidad de red intermediaria, tal como un servidor o controlador) que el otro acepta, y puede ser terminada por uno de los usuarios Alice y Bob que decide finalizar la llamada a su cliente. Alternativamente, como se explica en más detalle a continuación, una llamada se puede establecer mediante la solicitud que otra entidad en el sistema 100 establece una llamada con Alice y Bob como participantes, siendo una llamada en conferencia (específicamente de 3 vías) entre Alice, Bob y esa entidad en este evento.

Cada instancia de cliente de comunicación 118a, 118b tiene una instalación de acceso/autenticación que asocia los dispositivos 104a de usuario, 104b con sus respectivos usuarios 102a, 102b por ejemplo por el usuario que introduce un nombre de usuario (u otro identificador de usuario adecuado que transmite una identidad de dicho usuario dentro del sistema 100) y la contraseña en el cliente y que se ha verificado frente a los datos de cuenta del usuario almacenados en un servidor (o similar) del sistema de comunicación 100 como parte de un procedimiento de autenticación. Los usuarios son identificados de forma única por los identificadores de usuario asociados (por ejemplo los nombres de usuario) dentro del sistema de comunicación 100, con cada nombre de usuario que se asigna a la(s) instancia(s) respectiva(s) del cliente para el cual se pueden transmitir los datos (por ejemplo, llamada de audio/vídeo) para ese

usuario identificado.

Los usuarios pueden tener instancias de cliente de comunicación que se ejecutan en otros dispositivos asociados con los mismos detalles de acceso/registro. En el caso cuando el mismo usuario, que tiene una
5 nombre de usuario particular, puede estar conectado simultáneamente en varias instancias de la misma aplicación del cliente en diferentes dispositivos, un servidor (o similar) está dispuesto para asignar el nombre de usuario (ID de usuario) para todas esas varias instancias, pero también asignar un sub-identificador separado (sub-ID) para cada instancia
10 individual particular. Por lo tanto, el sistema de comunicación es capaz de distinguir entre las diferentes instancias, mientras todavía mantiene una identidad coherente para el usuario dentro del sistema de comunicación.

El usuario 102a (Alice) accede (se autentica) en el cliente 118a del dispositivo 104a como "Usuario 1". El usuario 102b (Bob) accede (se
15 autentica) en el cliente 118b del dispositivo 104b como "Usuario 2".

La Figura 2 ilustra una vista detallada de un dispositivo de usuario 104 (por ejemplo 104a, 104b) en el cual se ejecuta una instancia de cliente de comunicaciones 118 (por ejemplo, 118a, 118b). El dispositivo de usuario 104 comprende al menos un procesador 202 en la forma de una o
20 más unidades centrales de proceso ("CPU"), a la cual se conecta una memoria (almacenamiento informático) 214 para almacenamiento de datos, un dispositivo de salida en forma de una pantalla 222 (por ejemplo, 222a, 222b), que tiene un área de visualización disponible, tal como una pantalla de visualización, un teclado numérico (o un teclado) 218 y una cámara 216
25 para capturar los datos de vídeo (que son ejemplos de dispositivos de

entrada). La pantalla 222 puede comprender una pantalla táctil para la entrada de datos al procesador 202 y, por lo tanto, también constituyen un dispositivo de entrada del dispositivo de usuario 104. Un dispositivo de salida de audio 210 (por ejemplo, uno o más altavoces) y un dispositivo de audio de entrada 212 (por ejemplo, uno o más micrófonos) están conectados a la CPU 202. La pantalla 222, el teclado 218, la cámara 216, el dispositivo de salida de audio 210 y el dispositivo de entrada de audio 212 pueden estar integrados en el dispositivo de usuario 104 o uno o más pantallas 222, el teclado 218, la cámara 216, el dispositivo de salida de audio 210 y el dispositivo de entrada de audio 212 no pueden estar integrados en el dispositivo de usuario 104 y pueden estar conectados a la CPU 202 a través de sus respectivas interfaces. Un ejemplo de una tal interfaz es una interfaz USB. Por ejemplo un auricular de audio (es decir, un dispositivo único que contiene tanto un componente de salida de audio como un componente de entrada de audio) o cascos//auriculares del tipo que se introducen en los oídos (o similar) se pueden conectar a un dispositivo de usuario a través de una interfaz adecuado, tal como USB o interfaz basada en un conector de audio.

La CPU 202 está conectada a una interfaz de red 220 tal como un módem para la comunicación con la red de comunicaciones 106 para comunicarse a través del sistema de comunicación 100. La interfaz de red 220 puede o no estar integrada en el dispositivo de usuario 104.

El dispositivo de usuario 104 puede ser, por ejemplo, un teléfono móvil (por ejemplo, teléfono inteligente), una computadora personal ("PC") (incluyendo, por ejemplo, PC con sistema Windows®, Mac OS® y Linux®),

un dispositivo de juego, dispositivo de televisión (TV) (por ejemplo, TV inteligente), dispositivo informático de tableta u otro dispositivo integrado capaz de conectarse a la red 106.

Algunos de los componentes mencionados anteriormente pueden no
5 estar presentes en algunos dispositivos de usuario, por ejemplo, un dispositivo de usuario puede tomar la forma de un auricular de teléfono (VoIP o diferente) o un dispositivo de conferencia telefónica (VoIP o diferente).

La Figura 2 ilustra también un sistema operativo ("OS") 204
10 ejecutado en la CPU 202. El sistema operativo 204 gestiona los recursos de hardware del ordenador y se ocupa de los datos que se transmiten hacia y desde la red a través de la interfaz de red 220. El cliente 118 se muestra ejecutándose en la parte superior del sistema operativo 204. El cliente y el sistema operativo se pueden almacenar en la memoria 214
15 para su ejecución en el procesador 202.

El cliente 118 tiene una interfaz de usuario (UI) para presentar la información a y recibir información de un usuario del dispositivo de usuario 104. La interfaz de usuario comprende una interfaz gráfica de usuario (GUI) para la visualización de información en el área disponible de la
20 pantalla 222.

Volviendo a la figura 1, Alice 102, el usuario de origen, habla un idioma de origen; Bob, el usuario de destino, habla un idioma de destino que no es el idioma de origen (es decir, diferente al idioma de origen) y no entiende el idioma de origen (o tiene una comprensión limitada del mismo).
25 Por lo tanto, es probable que Bob sea incapaz de comprender o al menos

tener dificultades para entender lo que Alice dice en una llamada entre los dos usuarios. En los ejemplos a continuación, Bob se presenta como un hablante del idioma chino y Alice hablante del idioma inglés, como se apreciará, este es sólo un ejemplo y el usuario puede hablar cualquiera de los dos idiomas de cualquier país o región. Además, "diferentes idiomas" como se usa en este documento también se utilizan para significar diferentes dialectos de una misma lengua.

Para este fin, un sistema transmisor de traducción de idiomas (sistema de retransmisión traductor) 108 se proporciona en el sistema de comunicación 100. El propósito del transmisor traductor es traducir el audio de una llamada de audio o vídeo entre Alice y Bob. Es decir, el transmisor traductor es para la traducción del audio de la llamada de una llamada de voz o vídeo entre Alice y Bob del idioma de origen al idioma de destino para facilitar durante la comunicación durante la llamada entre Alice y Bob (es decir, para ayudar a Bob a comprender a Alice durante la llamada y viceversa). El transmisor traductor genera una traducción del audio de la llamada recibida de Alice en el idioma de origen, siendo la traducción al idioma de destino. La traducción puede comprender una traducción audible codificada como una señal de audio para la salida a Bob a través del (de los) altavoz/altavoces de su dispositivo y/o una traducción basada en texto para mostrarla a Bob a través de su pantalla.

Como se explica en más detalle a continuación, el sistema transmisor traductor 108 actúa tanto como traductor como un transmisor en el sentido de que recibe el audio de la llamada no traducida de Alice a través de la red 106, lo traduce y transmite la versión traducida de audio

de la llamada de Alice a Bob (es decir, transmite la traducción directamente a Bob a través de la red 106 para la salida durante la llamada, por ejemplo, en contraste con, digamos, el dispositivo de usuario de Alice o Bob, que actúa como un solicitante mediante la solicitud de una

5 traducción de un servicio de traductor, que se devuelve al solicitante para ser transmitido a otro dispositivo por el solicitante en sí). Esto representa una ruta de acceso rápido y eficiente a través de la red, lo que reduce al mínimo la carga de los clientes en términos de recursos de la red y el aumento de la velocidad total a la cual la traducción llega a Bob.

10 El traductor realiza un procedimiento de traducción automática "en vivo" de una llamada de voz o vídeo entre Alice y Bob en el sentido de que la traducción es, en cierta medida sincrónica con el discurso natural de Alice y Bob. Por ejemplo, el discurso normalmente natural durante la conversación implicará los intervalos de actividad de voz de Alice (es

15 decir, intervalos en los que Alice está hablando de manera intercalada con intervalos de inactividad de conversación de Alice, por ejemplo, cuando Alice hace una pausa para pensar o está escuchando a Bob. Un intervalo de actividad de voz, por ejemplo, puede corresponder a una frase o un pequeño número de frases precedidas y seguidas de una pausa en el

20 discurso de Alice. La traducción simultánea se puede llevar a cabo por tales intervalos de actividad de conversación por las cuales una traducción de intervalo inmediatamente anterior de Alice de la actividad de voz se desencadena por un intervalo suficiente (por ejemplo predeterminado) de inactividad de conversación ("inmediatamente anterior", con referencia al

25 intervalo de expresión más reciente de la actividad que no haya sido

traducido). En este caso, cuando la traducción esté completa, se puede transmitir a Bob para su emisión para que Bob la escuche tan pronto como sea posible después de escuchar el período de actividad más reciente de conversación natural de Alice, es decir, de manera que un período de actividad de conversación de Alice se escuche por Bob, seguido de una pausa corta (mientras se llevan a cabo la traducción y la transmisión de los mismos), seguido por Bob que escucha y/o ve la traducción del discurso de Alice en ese intervalo. La realización de la traducción con base en un tal intervalo puede resultar en una mayor calidad de la traducción ya que el procedimiento de traducción puede hacer uso del contexto en el que aparecen las palabras en una frase para efectuar una traducción más exacta. Debido a que el servicio de traductor está actuando como un transmisor, la longitud de esta breve pausa se reduce al mínimo, lo que resulta en una experiencia de usuario más natural para Bob.

De manera alternativa, la traducción automática se puede realizar con base en una palabra o varias palabras y, por ejemplo se emite mientras el discurso de Alice está todavía en curso y está escuchado, por ejemplo, por Bob como subtítulos que se muestran en el dispositivo de Bob y/o como audio que se reproduce en la parte superior de la conversación natural de Alice (por ejemplo, con el volumen de expresión reducido de Alice con respecto a la traducción audible). Esto puede resultar en una experiencia de usuario más sensible para Bob mientras la traducción se genera en tiempo casi real (por ejemplo, con un menos de aprox. 2 segundos el tiempo de respuesta). Los dos también se pueden combinar; por ejemplo, los resultados intermedios del sistema de reconocimiento de

la conversación (traducidos) se pueden visualizar en la pantalla, lo que les permite ser editados como los mejores cambios de hipótesis mientras la frase sigue y la traducción de la mejor hipótesis después se traducen en audio (ver a continuación).

5 La Figura 3 es una vista detallada del sistema de transmisor traductor 108. El sistema transmisor traductor 110 comprende al menos un procesador 304, que ejecuta el código 110. Al procesador 304 se conecta el almacenamiento informático (memoria) 302 para almacenar el código 110 para dicha ejecución y de datos y una interfaz de red 306 para la
10 conexión a la red 106. Aunque se muestra como un único dispositivo informático, la funcionalidad del sistema de transmisión 108 se puede distribuir alternativamente a través de múltiples dispositivos informáticos, por ejemplo, varios servidores, por ejemplo situados en el mismo centro de datos. Es decir, la funcionalidad del sistema de transmisión puede ser
15 implementado por cualquier sistema informático que comprende uno o más dispositivos informáticos y uno o más procesadores (por ejemplo, uno más núcleos de procesador). El sistema informático puede ser "localizado" en el sentido en el cual toda la funcionalidad de procesamiento y memoria se localiza sustancialmente a la misma ubicación geográfica (por ejemplo, en
20 el mismo centro de datos que comprende uno o más servidores de red a nivel local, que se ejecutan en el mismo o diferentes dispositivos de servidor del centro de datos). Como será evidente, esto puede ayudar a aumentar aún más la velocidad a la cual la traducción se transmite a Bob (el cual, en el ejemplo anterior reduce la longitud de la breve pausa entre
25 Alice que termina un intervalo de conversación y el comienzo de la salida

de traducción posteriormente, lo que resulta en una mejor experiencia de usuario para Bob).

Como parte del código 110, la memoria 302 contiene el código informático configurado para implementar un agente de traductor. Como se
5 explica con más detalle a continuación, el agente traductor también se asocia con su propio identificador de usuario (nombre de usuario) dentro del sistema de comunicación 100 de la misma manera que los usuarios están asociados con los correspondientes nombres de usuario. De este modo, el agente traductor también está identificado por un identificador de
10 usuario asociado y aparece, por lo tanto, en algunas modalidades, como otro usuario del sistema de comunicación 100, por ejemplo, que parece ser constantemente un usuario en línea que los usuarios "reales" 104a, 104b pueden agregar un contacto y transmitir datos a/recibir datos de uso de sus respectivos clientes 118a, 118b; en otras modalidades, el hecho de
15 que un robot que tiene un identificador de usuario puede estar oculto (o al menos disfrazado de manera que se oculta sustancialmente) a los usuarios, por ejemplo, con los UI del cliente configurados de tal manera que los usuarios podrían no conocer las identidades del robot (discutido anteriormente).

20 Como se apreciará, múltiples robots pueden compartir la misma identidad (es decir, estar asociados con el mismo nombre de usuario) y los robots se pueden distinguir utilizando diferentes identificadores que pueden ser invisibles para los usuarios finales.

El sistema transmisor traductor 108 también puede realizar otras
25 funciones que no están necesariamente directamente relacionadas con la

traducción, tales como la mezcla de flujos de audio de la llamada como en las modalidades ejemplares descritas a continuación.

La figura 4A es un diagrama de bloques que ilustra la función de las interacciones y la señalización entre los dispositivos de usuario 104a, 104b y un componente de gestión de llamadas 400. De acuerdo con los diversos métodos descritos a continuación, el sistema de gestión de llamadas 400 facilita la comunicación interpersonal entre personas que no comparten un idioma común (por ejemplo, Alice y Bob). La figura 4B es otra ilustración de algunos componentes mostrados en la figura 4A.

El componente de gestión de la llamada 400 representa la funcionalidad implementada mediante la ejecución del código 110 en el sistema transmisor traductor 108. El componente de gestión de llamadas se muestra que comprende bloques funcionales (componentes) 402-412 que representan diferentes funciones realizadas por dicho código 110 cuando se ejecuta. Específicamente, el componente de gestión de llamada 400 comprende los siguientes componentes: una instancia 402 del agente traductor mencionado cuya funcionalidad se describe con más detalle a continuación, un traductor de audio 404 configurado para traducir la conversación de audio del idioma de origen en texto en el idioma de destino, un convertidor de texto a conversación 410 configurado para convertir el texto en el idioma de destino de conversación sintetizada en el idioma de destino y un mezclador de audio 412 configurado para mezclar múltiples señales de audio de entrada para generar una única corriente de audio mixta que comprende audio de cada una de esas señales. El traductor de audio comprende un componente de reconocimiento

automático de conversación 406 configurado para el idioma de origen. Es decir, están configuradas para el reconocimiento del idioma de origen en el audio recibido, es decir, para identificar estas porciones particulares de sonido corresponden a las palabras en el idioma de origen

5 (específicamente para convertir la conversación de audio en el idioma origen a un texto en el idioma de origen de esta modalidad; en otras modalidades, no es necesario que sea texto - por ejemplo, el traductor puede traducir un conjunto completo de hipótesis proporcionadas por el motor de voz, representada como una red, que se puede codificar de

10 diversas maneras). El reconocimiento de voz también se puede configurar para identificar el idioma que el usuario de origen está hablando en la marcha (y configurado para el idioma de origen en respuesta por ejemplo, configurado para un modo "francés a ..." en respuesta a la detección del francés), o puede estar pre-configurado para el idioma de origen (por

15 ejemplo, a través de una interfaz de usuario o perfil de configuración, o mediante la señalización basada en mensajería instantánea, etc., que preconfigura el robot para, digamos, un modo "francés-a ..."). El componente 400 comprende también un traductor de texto 408 configurado para traducir el texto en el idioma de origen a un texto en el idioma de

20 destino. Colectivamente los componentes 404, 408 implementan la funcionalidad de traducción del traductor de audio 404. Los componentes 402, 404 y 410 constituyen un subsistema de traducción de extremo posterior (servicio de traducción) 401, con los componentes 404 y 410 que constituyen de un subsistema de traducción de voz a voz (S2ST) mismo y

25 el agente que opera como un intermediario entre los clientes 118a/118b y

ese subsistema.

Como se ha indicado, los componentes de las figuras 4A/4B pueden representar procesos que se ejecutan en la misma máquina o procesos distintos que se ejecutan en máquinas diferentes (por ejemplo, el reconocimiento de voz y la traducción del texto se pueden implementar como dos procesos distintos que se ejecutan en máquinas diferentes).

El agente traductor tiene una primera entrada conectada para recibir el audio de la llamada desde el dispositivo de usuario de Alice 104a través de la red 106, una primera salida conectada a una entrada del traductor de audio 404 (específicamente, del componente de reconocimiento de voz 406), una segunda entrada conectada a una salida del componente de reconocimiento de voz 406 (que es una primera salida del traductor de audio 404), una tercera entrada conectada a una salida del traductor de texto 408 (que es una segunda salida del traductor de audio 404), una segunda salida conectada a una primera entrada del mezclador 412, una tercera salida conectada para transmitir el texto traducido en el idioma de destino al dispositivo de usuario 104b de Bob y una cuarta salida configurada para transmitir el texto reconocido en el idioma fuente del dispositivo de usuario 104a de Alice y también al dispositivo de usuario 104b de Bob. El agente 402 también tiene una cuarta entrada conectada a una salida del convertidor de texto a voz 410 y una quinta salida conectada a una entrada del convertidor de texto a voz. El mezclador 412 tiene una segunda entrada conectada para recibir el audio de la llamada del dispositivo 104a de Alice y una salida conectada para transmitir el flujo de audio mezclado a Bob través de la red 106. La salida del componente

de reconocimiento de voz 406 también está conectada a una entrada del traductor de texto 408. El agente 402 tiene una quinta entrada conectada para recibir datos de realimentación desde el dispositivo de usuario de Alice 104a a través de la red 106 que transmite la realimentación de los resultados del usuario de origen del procedimiento de reconocimiento de origen (por ejemplo, que indica la precisión de los mismos), la información de realimentación que ha sido seleccionada por Alice a través de su interfaz de usuario del cliente y la información de transporte perteneciente al texto reconocido para su uso en la configuración del reconocedor de voz 406 para mejorar sus resultados. Alice se encuentra en una posición para proporcionar esta información cuando se recibe la información que pertenece a los resultados del reconocimiento de voz que se puede emitir a través de su interfaz de usuario del cliente.

Las entradas/salidas que representan las señales de audio se muestran como flechas de trazo grueso como en la figura 4A; las entradas/salidas que representan señales basados en texto se muestran como flechas delgadas.

La instancia del agente traductor 402 funciona como una interfaz entre los clientes Alice y Bob 118 y el subsistema de traducción 401 y funciona como un "agente de software" independiente. La computación basada en agentes se conoce en la técnica. Un agente de software es un programa informático autónomo que lleva a cabo tareas en nombre de los usuarios en una relación de agencia. Al actuar como un agente de software, el agente traductor 402 funciona como una entidad de software autónomo, el cual, una vez iniciado (por ejemplo, en respuesta a una

iniciación de una llamada o sesión relacionada) se extiende sustancialmente de manera continua durante la llamada o sesión (en lugar de ser ejecutado sobre demanda, es decir en lugar de ser ejecutado sólo cuando requerido para realizar algunas tareas específicas), esperando las
5 entradas, las cuales, cuando se detectan, disparan las tareas automáticas para realizarlas en esas entradas por el agente traductor 402.

En modalidades particulares, la instancia de agente traductor 402 tiene una identidad dentro del sistema de comunicación 100 justo como los usuarios del sistema 100 tienen identidades dentro del sistema. En este
10 sentido, el agente traductor puede ser considerado como un "robot"; que es una entidad de software de inteligencia artificial (AI) que aparece como un usuario normal (miembro) del sistema de comunicación 100 en virtud de su nombre de usuario y el comportamiento asociado (ver anteriormente). En algunas implementaciones, una instancia diferente respectiva de un
15 robot se puede ser asignar a cada llamada (es decir, con base en instancia por llamada), por ejemplo, TraductorInglésEspañol1, TraductorInglésEspañol2. Es decir, en algunas implementaciones el robot está asociado a una sesión única (por ejemplo, llamada entre dos o más usuarios). Por otra parte, el servicio de traducción al cual el robot ofrece
20 una interfaz se puede ser compartir entre varios robots (y también otros clientes).

En otras implementaciones, un ejemplo de Robot que es capaz de llevar múltiples conversaciones al mismo tiempo se puede configurar de una manera directa.

25 En particular, los usuarios humanos 104a, 104b del sistema de

comunicación 100 puede incluir el robot como participante en las llamadas de voz o de vídeo entre dos o más usuarios humanos por ejemplo, invitando al robot para unirse a una llamada establecida como participante, o solicitando que el robot iniciara una llamada de conferencia entre los dos
5 o más participantes humanos deseados y el propio robot. La solicitud es instigada por la interfaz de usuario del cliente de uno de los clientes 118a, 118b, que proporciona opciones para seleccionar el robot y cualquier usuario humano deseado como, por ejemplo, participantes de la enlistando los seres humanos y los robots como contactos en una lista de contactos
10 que se muestra a través de la interfaz de usuario del cliente.

Las modalidades basadas en robots no requieren dispositivos de hardware especializados o software especializado para su instalación en las máquinas de los usuarios y o requieren que los parlantes (es decir, participantes) estén físicamente cerca entre sí, para que el robot se puede
15 integrar perfectamente en la arquitectura del sistema de comunicación existentes sin la necesidad de, por ejemplo, clientes redistribuidos de software actualizado.

El agente 402 (robot) aparece en el sistema de comunicación 100 (denominado alternativamente como la red de chat) como miembro regular
20 de la red. Los participantes de la conversación de voz pueden conseguir que los discursos de sus interlocutores se traduzcan a su idioma invitando al robot apropiado a una llamada de voz o de video (también conocido como sesión de chat o conversación por ejemplo) un hablante del idioma chino con un hablante del idioma Inglés podría invitar a un agente llamado
25 (es decir, que tiene un nombre de usuario) "Traductor-Inglés-Chino" en la

conversación. El robot entonces toma el papel del traductor o intérprete durante el resto de la conversación, traduciendo cualquier discurso de su idioma de origen al idioma de destino. El mismo se puede presentar en forma de texto (para la visualización en el dispositivo de destino, por ejemplo, a través de subtítulos o en una ventana de chat de la interfaz de usuario del cliente objetivo) y/o como discurso de idioma de destino (para reproducirlo a través de los altavoces en el dispositivo de destino, generado mediante el componente de conversión de texto a voz 410).

Por lo tanto, las modalidades proporcionan:

- Perfecta integración en los servicios de llamadas/chat multimedia (no requiere instalación independiente)
- Comunicación remota (los participantes no tienen que estar físicamente cerca)
- Implementaciones basadas en servidor independiente del dispositivo (por ejemplo, que los clientes de servicios, por ejemplo, 104a, 104b para las nuevas plataformas no requieren software independiente alguno), lo que permite un despliegue más transparente de mejoras y nuevas características.

En algunas modalidades, el robot tiene acceso a la separación de los flujos de audio separados por hablante, lo que permite el reconocimiento de voz de mayor calidad.

En tales modalidades, en el nivel superior se encuentra el "robot", que aparece a los usuarios del sistema de chat, tal como un miembro regular de la red humana lo haría. El robot intercepta los flujos de audio de todos los usuarios que hablan su idioma de origen (por ejemplo, 104a), y

los pasa a un sistema de traducción de voz a texto (traductor de audio 404). La salida del sistema de traducción de voz a texto es el texto en el idioma de destino. El robot comunica la información del idioma de destino a los usuarios del idioma de destino 104b. El robot también puede

5 comunicar los resultados de reconocimiento de voz de la señal de audio de origen al hablante de origen 104a y/o el oyente de destino 104b. El hablante de origen posteriormente puede corregir los resultados del reconocimiento por información de corrección de retroalimentación al robot a través de la red 106 para obtener una mejor traducción, o tratar de

10 repetir o reformular su enunciado (o partes del mismo) para lograr un mejor reconocimiento y la traducción. Alternativamente, al hablante se le puede presentar una lista de los n mejores o la representación de la red de hablantes (es decir, un gráfico que reprime visualmente distintas hipótesis posibles para el discurso de origen reconocido), lo que les permite aclarar

15 o corregir los reconocimientos de 1 mejor imperfecto retroalimentando la selección al identificar la mejor hipótesis. La información de reconocimiento (por ejemplo, el texto propio en el idioma de origen) también se podría transmitir al usuario objetivo – esto puede ser útil para los oyentes que tienen un pequeño grado de competencia en el idioma de

20 origen o cuya comprensión lectora en tal idioma es mejor que su comprensión auditiva. Teniendo acceso al texto de origen también podría permitir que el usuario de destino tenga más sentido de la traducción incorrecta o ambigua; entidades nombradas como los nombres de personas o lugares, por ejemplo, podrían ser reconocidos correctamente por el

25 sistema de reconocimiento de voz pero traducidos incorrectamente.

Los detalles de implementación del robot dependen de la arquitectura y el nivel de acceso a la red de chat.

Las implementaciones de sistemas que proporcionan SDK ("kits para desarrolladores de software") dependerán de las características proporcionadas por el SDK. Normalmente, estas ofrecen acceso de lectura a secuencias de video y audio independientes para cada participante de la conversación, y escribir el acceso a las secuencias de video y audio para el propio robot.

Algunos sistemas proporcionan los SDK del robot lateral del servidor, que permite el acceso completo a todas las secuencias y habilita los escenarios tales como la imposición de los subtítulos de vídeo a través de la señal de vídeo de los hablantes de origen y/o reemplaza o mezcla la señal de salida de audio del hablante de destino. Por último, cuando se dispone de un control completo del sistema, la traducción se puede integrar en cualquier forma, incluidos los cambios en la interfaz de usuario del cliente para hacer que la experiencia de conversación entre idiomas sea más fácil para los usuarios.

En el nivel más débil, las redes "cerradas" sin protocolos definidos públicamente y/o del SDK se pueden servir por los robots que interceptan y modifican las señales desde y hacia los dispositivos de micrófono, cámara y altavoz en un equipo del cliente (por ejemplo, 104a, 104b -en lugar de un transmisor separado). En este caso, el robot puede realizar la detección del idioma para averiguar qué partes de la señal están en su idioma de origen (por ejemplo, para distinguir entre el discurso en otros idiomas en una secuencia de audio mixta).

La comunicación del texto en el idioma de destino puede ocurrir de varias maneras; el texto se puede comunicar en público (universalmente visible/audible para todos los participantes de la llamada por ejemplo, Alice, así como Bob) o canal de chat en privado (únicamente entre el robot y el usuario de destino) y/o como subtítulos de vídeo superpuestos, ya sea en la secuencia de video del parlante del idioma de origen o de destino. El texto también se puede pasar a un componente de conversión de texto a voz (convertidor de texto a voz 410), lo que hace que el texto en el idioma de destino sea una señal de audio que puede reemplazar la señal de audio original del hablante o de lo contrario ser mezclado con el mismo. En una modalidad alternativa, sólo el texto traducido se envía a través de la red y la síntesis de texto a voz se realiza en el lado del cliente (ahorro de recursos de la red).

La traducción puede ser o bien a base de turnos (el robot espera hasta que el usuario pause o indique de alguna otra manera que su expresión es completa, como, por ejemplo, hacer clic en el botón, después, comunicar la información en el idioma de destino) o simultánea - es decir, sustancialmente contemporánea con el discurso de origen (el Robot comienza a comunicar la información de idioma de destino el momento en que tiene suficiente texto para producir la salida coherente semántica y sintáctica). El primero utiliza la detección de actividad de voz para determinar cuándo iniciar la traducción de una parte anterior del discurso (la traducción se realiza por intervalos de actividad de voz detectada); esta última utiliza la detección de activación de voz y un componente de segmentación automática (estando llevado a cabo, para cada intervalo de

actividad de discurso detectado, en un intervalo por segmento, que puede tener uno o más segmentos). Como se apreciará, los componentes para llevar a cabo tales funciones son fácilmente disponibles. En el escenario basado en turnos, el uso del robot que actúa como traductor virtual de
5 terceros en la llamada ayudaría a los usuarios enmarcándolos en un escenario del mundo real común con un traductor (como uno podría tener en un tribunal); la traducción simultánea es análoga a un intérprete simultáneo humano (por ejemplo, como los que se pueden encontrar en el Parlamento Europeo o la ONU). Por lo tanto, ambos proporcionan una
10 experiencia de traducción intuitiva para los usuarios de destino.

Se debe señalar que las referencias a la "traducción automática" (o similar) como se usa en el presente documento cubre tanto la traducción a base de turnos como traducción simultánea (entre otros). Es decir, la "traducción automática" (o similar) cubre tanto la emulación automatizada
15 de traductores humanos e intérpretes humanos.

Como se apreciará, el tema no se limita a cualquiera de los componentes particulares de reconocimiento de discurso o traducción - para todos los intentos y propósitos, estas se pueden tratar como un cuadro negro. Las técnicas para realizar una traducción de una señal de
20 discurso se conocen en la técnica, y hay varios componentes disponibles para llevar a cabo tales funciones.

Aunque las figuras 4A/4B muestran sólo una traducción de una sola vía en aras de la simplicidad, se apreciará fácilmente que el robot 402 puede realizar las funciones de traducción equivalente en audio de la
25 llamada de Bob para el beneficio de Alice. Del mismo modo, mientras los

métodos a continuación se describen en relación a la traducción de un solo sentido por simplicidad, se apreciará que tales métodos se pueden aplicar a traducciones de dos vías (o de múltiples vías).

Un método para facilitar la comunicación entre los usuarios durante una llamada de voz o vídeo entre los usuarios se describirá ahora con referencia a la figura 5, que es un diagrama de flujo para el método. La Figura 5 describe un procedimiento de traducción durante la llamada desde el idioma de Alice al idioma de Bob sólo por simplicidad; se apreciará que un proceso equivalente y separado se puede realizar para traducir del idioma de Bob al idioma de Alice de forma simultánea en la misma llamada (de esta perspectiva, Alice puede ser vista como el objetivo y Bob como el destino).

En la etapa S502, la solicitud de un servicio de traductor es recibida por el sistema transmisor traductor 108, solicitando que el robot realice un servicio de traducción durante una llamada de voz o de video en la cual Alice, Bob y el robot serán participantes. Por lo tanto, la llamada constituye una llamada de conferencia (grupo) - específicamente llamada de tres vías. En el paso S504, se establece la llamada. La solicitud puede ser una solicitud para el agente 402 para establecer una llamada de conferencia entre el robot 402 y al menos Alice y Bob en cuyo caso el robot establece la llamada (con S502 siendo antes S504) estableciendo invitaciones de llamada a Alice y Bob, o la solicitud podrá ser una invitación para que el robot 402 se una a una llamada ya establecida entre al menos Alice y Bob (con S504 siendo después S502), en cuyo caso Alice (o Bob) establece la llamada instigando las invitaciones de llamada a Bob

(o Alice) y el robot). La misma puede ser instigada a través de la interfaz de usuario del cliente o de forma automática, ya sea por el cliente o alguna otra entidad (por ejemplo, un servicio de calendario configurado para instigar automáticamente una llamada en un tiempo pre-especificado).

5 En el paso S506, el robot 402 recibe la llamada de audio de Alice como una secuencia de audio a través de la red 106 del cliente de Alice 118. El audio de la llamada es de audio captado por el micrófono de Alice, y comprende el discurso de Alice, que está en el idioma de origen. El robot 402 suministra el audio de la llamada al componente de reconocimiento de
10 discurso 406.

 En el paso S508, el componente de reconocimiento de voz 406 realiza un procedimiento de reconocimiento de voz en el audio de la llamada. El procedimiento de reconocimiento de voz está configurado para el reconocimiento del idioma de origen. Específicamente, el procedimiento
15 de reconocimiento de voz detecta patrones particulares en el audio de la llamada que coinciden con los patrones de voz conocidos del idioma de origen con el fin de generar una representación alternativa de este discurso. Esto puede ser, por ejemplo, una representación textual del discurso como una cadena de caracteres en el idioma de origen, con el
20 procedimiento que constituye un procedimiento de reconocimiento de discurso de origen a texto de origen, o alguna otra representación, como una representación de entidad vectorial. Los resultados del procedimiento de reconocimiento de voz (por ejemplo, vectores de cadena/función) se introducen en el traductor de texto 408 y también se suministran de nuevo
25 al robot 402.

En el paso S510, el traductor de voz 408 realiza un procedimiento de traducción de los resultados de entrada en el texto en el idioma de destino (o alguna otra representación similar).

La traducción se realiza 'sustancialmente en vivo, por ejemplo, por frases (o par de frases), por segmento detectado o función de cada palabra (o algunas palabras) como se ha mencionado anteriormente. Por lo tanto, el texto traducido se emite de forma semi-continua mientras el audio de la llamada se recibe por Alice. El texto en el idioma de destino se suministra de nuevo al robot 402.

En el paso S512, el texto en el idioma de destino se suministra por el robot al convertidor de texto a voz, que convierte el texto en el idioma de destino en voz artificial que se habla en el idioma de destino. La voz sintética se suministra de nuevo al robot 402.

Debido a que tanto la salida del texto de audio del traductor 404 y la voz sintética están en el idioma de destino, son comprensibles para Bob que hable el idioma de destino.

En el paso S514, el audio sintético se suministra al mezclador 412 donde se mezcla con el audio original de Alice (que comprende su discurso original, natural) para generar una secuencia de audio mixto que comprende tanto el discurso sintético traducido en el idioma de destino como el discurso natural original en el idioma de origen, que se transmite a Bob través de la red 106 (S516) para emitirlo a través del dispositivo de salida de audio de su dispositivo de usuario como parte de la llamada. Por lo tanto, Bob puede calibrar el tono de Alice etc., del discurso natural (aunque él no lo entiende), mientras capta el significado de la voz sintética

que resulta en una comunicación más natural. Es decir, el sistema también puede transmitir el audio no traducido de Alice, así como el audio traducido. Además, incluso cuando el usuario de destino no entiende el idioma de origen, todavía hay información que se debe recolectar del mismo, por ejemplo, entonación (que puede ser capaz de decir si el hablante de origen está haciendo una pregunta, por ejemplo).

Alternativamente, la señal original del discurso de Alice no se puede transmitir a Bob de manera que sólo el discurso sintético, traducido se puede transmitir a Bob.

Como se ha mencionado, el texto en el idioma de destino también se puede transmitir a Bob por el robot (y mostrado a través de su interfaz de usuario del cliente, por ejemplo, en una interfaz de chat o como subtítulos). Como también se ha mencionado, el texto en el idioma de origen obtenido por el procedimiento de reconocimiento de voz en el cual se basa la traducción - y/u otra información de reconocimiento que pertenece al proceso de reconocimiento de voz realizado en su discurso, tales como reconocimientos alternativos posibles (por ejemplo, donde no se reconoce la ambigüedad en la realización del procedimiento de reconocimiento) - del procedimiento de reconocimiento también se puede transmitir a Alice y desplegar a través de su interfaz de usuario para que pueda medir la precisión de dicho procedimiento de reconocimiento. La interfaz de usuario del cliente puede presentar varias opciones de retroalimentación mediante las cuales Alice puede alimentar información de vuelta al robot a través de la red para mejorar y perfeccionar el procedimiento de reconocimiento de voz como se realizó en su discurso. El

texto de la lengua de origen también se puede transmitir a Bob (por ejemplo, si Bob selecciona una opción para recibirlo a través de su interfaz de usuario del cliente), por ejemplo si Bob es más competente en la lectura del idioma de origen hablado por Alice de lo que es en la interpretación
5 auditiva.

En modalidades, el componente 406 de discurso a texto puede emitir una versión de texto de cada palabra mientras se reconoce (por ejemplo, en función de cada palabra) o algunos otros resultados de reconocimiento de voz intermedia parcial, que se pueden visualizar en el dispositivo de
10 usuario de Alice mientras ella está hablando. Es decir, el procedimiento de reconocimiento de voz puede estar configurado, por al menos un intervalo de actividad de voz por el usuario de origen, para generar resultados "provisional" parciales de reconocimiento de voz mientras que la actividad de voz está en curso antes de generar los resultados finales de
15 reconocimiento de voz cuando la actividad de voz es completada (es decir, cuando Alice deja de hablar al menos temporalmente). La traducción se genera, en última instancia, utilizando los resultados finales (no los resultados parciales, que pueden estar sujetos a cambios antes de que se lleve a cabo la traducción – ver a continuación), pero se genera la
20 información relativa a los resultados parciales, no obstante, transmitidos y emitidos para a Alice antes de generar la traducción. Esto invita al usuario de origen (Alice) influir en la traducción posterior, por ejemplo, mediante la modificación de su actividad de voz (por ejemplo, mediante la repetición de una parte que ellos pueden ver que se ha interpretado erróneamente)
25 siempre y cuando observan que las inexactitudes están presentes en los

resultados parciales.

Mientras Alice continúa hablando, el procedimiento de reconocimiento se refina posteriormente, de manera que el componente 406 puede efectivamente "cambiar de opinión" sobre la(s) palabra(s) que
5 previamente se ha(n) reconocido en su caso, en vista del contexto proporcionado por las palabras siguientes. En general, el componente 406 puede generar resultados iniciales (y efectivamente provisionales) de reconocimiento de voz en tiempo sustancialmente real (por ejemplo, con los resultados que se actualizan en una escala de tiempo del orden de 2
10 segundos), que se puede visualizar por Alice sustancialmente en tiempo real para que pueda tener una idea de la precisión con la cual está siendo reconocido su discurso - incluso si los resultados provisionales están sujetos a cambios antes de producir los resultados finales de la cual se genera realmente el audio, todavía pueden proporcionar una parte
15 suficiente de una idea para serle útil a Alice. Por ejemplo, si Alice puede ver que el procedimiento de reconocimiento ha interpretado su discurso de una manera muy imprecisa (y por lo tanto sabe que, mientras ella simplemente sigue hablando, la traducción resultante posteriormente emitida para Bob sería ilegible o sin sentido), ella puede cortar su
20 secuencia actual del discurso y repetir lo que acaba de decir en lugar de tener que terminar una porción entera del discurso antes de que el error se vuelva evidente (por ejemplo, que de otro modo sólo sería el caso después de que Bob haya escuchado y no comprende la traducción confusa o sin sentido). Como se apreciará, esto ayudará a facilitar una secuencia natural
25 de la conversación entre Alice y Bob. Otra posibilidad es tener un botón u

otro mecanismo de interfaz de usuario, que Alice podría utilizar para detener el reconocimiento actual y empezar de nuevo.

En esta modalidad, el mezclador 412 de la figura 4A, también es aplicado por el sistema de transmisión 108 en sí. Es decir, además de
5 implementar funciones de traductor, el sistema de transmisión 108 también implementa funciones de mezcla de audio de la llamada. La implementación de la funcionalidad de la mezcla (por lo que, para cada participante humano, se mezclan múltiples secuencias de audio individuales en una secuencia de audio única respectiva para transmitir a
10 dicho usuario) en el sistema de transmisión 108 en sí más que en otras partes del sistema (por ejemplo, en uno de los dispositivos de usuario 104a, 104) proporciona un cómodo acceso a la secuencia de audio individual al Robot - como se ha mencionado anteriormente, al tener acceso a las secuencias de audio de llamadas individuales puede resultar
15 en una mejor calidad de traducción. Cuando el sistema de transmisión 108 también se localiza, esto también asegura que el robot tiene acceso inmediato, rápido a las secuencias de audio individuales que minimiza aún más los retrasos de traducción.

Cuando los usuarios adicionales participan en una llamada (además
20 de Alice, Bob y el robot en sí), las secuencias de audio de la llamada de estos usuarios pueden también, realizar traducciones separadas en cada secuencia de audio por el robot 402. Cuando más de dos usuarios humanos participan en una llamada, las secuencias de audio para todos aquellos usuarios se pueden recibir de forma individual en el sistema de
25 transmisión 108 para su mezcla, lo que también proporciona un cómodo

acceso a todas las secuencias de audio individuales para su uso por el robot. Cada usuario puede recibir después una secuencia de audio mixta que contiene todas las traducciones necesarias (es decir discurso sintético traducid para cada usuario que habla una lengua diferente a la del usuario). Un sistema con tres (o más) usuarios podría tener usuarios que hablan cada uno un idioma diferente, donde su voz se traduciría en ambos (o más) idiomas de destino, y el discurso de los dos (o más) hablantes objetivo se traducirá a su idioma. Cada usuario se puede presentarse a través de su interfaz de usuario del cliente con el texto original y su propia traducción. Por ejemplo, el usuario A habla inglés, el usuario B habla italiano y el usuario C habla francés. El usuario A y el usuario B verán el inglés y el italiano, mientras que el usuario C verá inglés y francés.

En algunos sistemas de comunicación existentes, el usuario que inicia una llamada en grupo se asigna automáticamente para organizar la convocatoria mencionada, con el audio de la llamada que se mezclan en el dispositivo de ese usuario por defecto y otros clientes en la llamada enviando automáticamente las secuencias de audio a ese usuario por defecto para la mezcla. Se espera que el anfitrión a continuación, genere una secuencia de audio mixta respectiva para cada usuario, la secuencia de audio respectiva para ese usuario siendo una mezcla de audio de todos los demás participantes (es decir, todo el audio que no sea el audio propio de ese usuario). En tales sistemas, una solicitud para que el robot inicie la llamada se asegurará de que el robot se asigne como anfitrión, por lo tanto asegurando que el cliente de cada otro participante transmita su secuencia de audio individual al sistema de transmisión 108 para mezclarse por

defecto, permitiendo de este modo el acceso a las secuencias de audio individuales para el robot de forma predeterminada. El robot después proporciona una secuencia de audio mixta respectiva a cada participante que incluye no sólo el audio de los otros participantes humanos, sino
5 también cualquier archivo de audio (por ejemplo, discurso sintetizado traducido) para ser transportado por el propio robot.

En algunas implementaciones basadas en robot, el software de cliente puede ser modificado (en particular, la interfaz gráfica de usuario de cliente puede ser modificada) para disfrazar el hecho de que un robot
10 está realizando la traducción. Es decir, desde la perspectiva de la arquitectura subyacente del sistema de comunicación, el robot aparece sustancialmente como si se tratara de otro miembro del sistema de comunicación para que el robot se integre perfectamente en dicho sistema de comunicación sin modificación de la arquitectura subyacente; sin
15 embargo, esto se puede ocultar para los usuarios de manera que el hecho de que las traducciones que están recibiendo durante las llamadas están siendo transportadas por un robot que es un participante en la llamada (al menos en términos de los protocolos subyacentes) es sustancialmente invisible al nivel de interfaz de usuario.

20 Mientras que la anterior se describe con referencia a la implementación de un robot, es decir, con referencia a un agente traductor que está integrado en el sistema de comunicación 100 mediante la asociación de ese agente con su propio identificador de usuario de tal manera que aparece como un usuario normal del sistema de comunicación
25 100 - es posible que otras modalidades no tengan la implementación del

robot. Por ejemplo, el transmisor traductor 108 puede, a su vez, integrarse en un sistema de comunicación como parte de la arquitectura del propio sistema de comunicación, con la comunicación entre el sistema 108 y los diferentes clientes que se efectúa mediante los protocolos de comunicación adaptados a tales interacciones. Por ejemplo, el agente traductor se puede alojar en una nube como un servicio en la nube (por ejemplo, que se ejecuta en una o más máquinas virtuales implementadas por una plataforma de hardware de nube subyacente).

Es decir, el traductor podría, por ejemplo, ser un dispositivo informático/sistema de este tipo de dispositivos que ejecutan un robot con un identificador de usuario, o un servicio de traductor que se ejecuta en la nube, etc. De cualquier forma, el audio de una llamada se recibe desde el usuario de origen, pero la traducción se envía directamente al usuario de destino desde el sistema traductor (no transmitido a través del cliente del usuario de origen), es decir, en cada caso, el sistema traductor actúa como un transmisor eficaz entre la fuente y los usuarios de destino. Un servicio de nube (o similar), se podría, por ejemplo, acceder, directamente a partir de un navegador web (por ejemplo, mediante la descarga de un complemento, por ejemplo, sin complemento en el navegador de comunicación basado en JavaScript), desde un cliente de software dedicado (aplicación o incrustado), marcando desde un teléfono fijo o móvil, etc.

Un método para proporcionar una traducción del discurso de un usuario de origen a un usuario de destino se describirá ahora con referencia a las figuras 6, 7A-E y 8.

La Figura 8 muestra un sistema de traducción basado en notificaciones 800 que comprende los siguientes bloques funcionales (componentes): un traductor de voz a voz (S2ST) 802 (que puede implementar la funcionalidad similar a la del sistema S2ST formado de los componentes 404 y 410 en la figura 4A/B), que lleva a cabo un procedimiento de traducción de voz a voz para generar el discurso traducido sintético en el idioma de destino de audio de la llamada de Alice que comprende el discurso de Alice en el idioma de origen que se traduce posteriormente y un componente de generación de notificaciones (componente de notificación) 804 configurado para generar una o más notificaciones, separadas del audio traducido, para emitirlas al usuario de destino que transmiten los cambios en el comportamiento de la traducción (es decir, un cambio en la naturaleza de las operaciones relacionadas con la traducción realizadas al proporcionar un servicio de traducción durante la llamada entrante) del procedimiento de traducción cuando se detecta por el componente de notificación. Estos componentes representan funcionalidad implementada, por ejemplo, mediante la ejecución del código 110 en el transmisor traductor 108 (o mediante la ejecución del código en algunos otros sistemas informáticos de servicios integrales), mediante la ejecución del cliente 118 en el dispositivo 104a, mediante la ejecución del cliente 118b en el dispositivo 104b, o cualquier combinación de los mismos (es decir, con la funcionalidad distribuida a través de múltiples dispositivos). En general, el sistema 800 se puede implementar por cualquier sistema informático de uno o más dispositivos informáticos de una manera localizada o distribuida

El procedimiento de traducción emite la traducción del audio como una secuencia de audio que es enviada al usuario de destino a través de los altavoces del dispositivo de destino, ya que se emite por el procedimiento de traducción (por ejemplo, que se transmite al dispositivo de destino a través de la red cuando se traduce de forma remota o por emisión directamente por los altavoces cuando se traduce a nivel local). Por lo tanto, la emisión de la traducción de audio por el procedimiento de traducción es sustancialmente contemporáneo con la emisión de esa traducción en el dispositivo de destino (es decir, con los únicos retrasos significativos siendo aquellos introducidas como resultado de la latencia de la red y/o en el dispositivo de destino, etc.)

Además, el sistema 800 comprende un componente de emisión de notificación 806 y un componente de emisión de la traducción 808 que están separados entre sí (recibiendo entradas separadas y diferentes) implementadas en el dispositivo del usuario de destino 104b y que representan funcionalidad implementada mediante la ejecución del cliente 118b en el dispositivo del usuario de destino 104b. Los componentes 806 y 808 reciben (a partir de los componentes 804 y 802, respectivamente) y emiten al usuario de destino las notificaciones generadas y el audio traducida (que se emite a través de los altavoces del dispositivo de destino), respectivamente. La notificaciones (audio traducido respectivo) se puede recibir a través de la red 106 en la cual el componente de generación de notificación 804 (traductor 802 respectivo) se implementa a distancia desde el dispositivo de usuario de destino (por ejemplo, en el dispositivo de origen y/o un servidor etc.), o localmente si el componente

804 de generación de notificaciones (traductor 802 respectivo) se implementa en el dispositivo de destino en sí.

El traductor de voz a voz tiene una entrada conectada para recibir el audio de la llamada de Alice (por ejemplo, a través de la red 106, o de forma local en la cual el componente 802 se implementa en el dispositivo de Alice), una primera salida conectada a una entrada del componente de emisión de la traducción 808 para proporcionar el audio traducido a Bob (por ejemplo, a través de la red 106, o directamente a los altavoces de Bob cuando se implementa en el dispositivo de Bob), y una segunda salida conectada a una primera entrada del componente de notificación 804. Esta segunda salida comunica las señales al componente de notificación que señala los cambios en el comportamiento del procedimiento de traducción (por ejemplo, a través de la red 106 cuando estos componentes se implementan en diferentes dispositivos, o por medio de, por ejemplo, de dispositivo local, señalización interna cuando se implementa en el mismo dispositivo). El componente de generación de notificación tiene una salida conectada a una entrada del componente de salida de la notificación 806, cuya entrada causa que la notificación antes mencionada se emita a Bob (por el componente de emisión de notificación) para notificarle cuando se detectan tales cambios. El componente de notificación tiene al menos una primera salida conectada al menos a un dispositivo respectivo de salida (pantalla, altavoces, y/u otro dispositivo de salida) del dispositivo del usuario de destino 118b para la emisión de las notificaciones. El componente de salida de la traducción 808 tiene una salida conectada a los altavoces del dispositivo de usuario de destino 104b para la emisión de

la traducción del audio.

Además, el componente de emisión de notificación 806 tiene una segunda salida conectada a una segunda entrada del componente de notificación, que suministra información relacionada con la emisión acerca
5 de la manera en la que las notificaciones se deben emitir en dispositivo de usuario de destino 10 para su uso en la generación de las notificaciones. Es decir, el componente de emisión de notificaciones 806 retroalimenta la información al componente de generación de notificaciones 804 sobre la forma en la cual la(s) notificación(es) se deben emitir emitida en el
10 dispositivo de destino, el componente de generación de notificaciones se utiliza para determinar cómo se generan las notificaciones. Por lo tanto, la manera en la que se generan las notificaciones puede depender de la manera en la cual se emiten realmente en el dispositivo. Esta información se puede alimentar de nuevo de forma remota a través de la red 106 donde
15 se implementa el componente de generación de la notificación 804 o la retroalimentación puede ser un proceso localizado (interno) en el dispositivo de destino en el cual el componente de generación de notificación 804 se implementa localmente en el dispositivo de destino.

En el caso de las notificaciones visuales que se mostrarán en la
20 pantalla del dispositivo de destino, la información relacionada con la emisión comprende información de diseño que transmite la manera de colocación de las notificaciones de salida en el área disponible de la pantalla del dispositivo de destino.

En los ejemplos descritos a continuación, el componente de
25 notificación 804 genera datos de vídeo sintéticos de un "avatar" animado

para mostrar a Bob en el dispositivo del usuario (que se puede transmitir a través de la red 106 o se comunicará directamente a la pantalla cuando el componente 804 se implementa en el dispositivo de Bob. En estos ejemplos, el componente de notificación 804 genera un video sintético del avatar animado, dicho video contiene las notificaciones - por ejemplo, como un cambio en el comportamiento visual del avatar. La información de diseño comprende la información sobre dónde se va a mostrar el video del avatar en el área de la pantalla disponible del dispositivo de destino durante una llamada de video con relación a los videos mostrados del usuario de destino (Bob) y/o el usuario de origen (Alice) para su uso en la determinación del comportamiento visual del avatar.

La Figura 6 es un diagrama de flujo para el método. El método de la figura 6 se lleva a cabo durante, y como parte de, una llamada de voz o vídeo establecida entre un usuario de origen (por ejemplo, Alice), utilizando un dispositivo de usuario de origen (por ejemplo, 104a), y un usuario de destino (por ejemplo, Bob), utilizando un dispositivo de usuario de destino (por ejemplo, 104b), en el cual se lleva a cabo un procedimiento de traducción en el audio de la llamada, el audio de la llamada comprende el discurso del usuario de origen en el idioma de origen, para generar una traducción de audio del discurso del usuario de origen en el idioma de destino para emitirlo al usuario de destino. El procedimiento de traducción se puede realizar en un relevador de traducción en la forma descrita anteriormente, o puede o no, por ejemplo llevar a cabo en uno de los dispositivos de usuario o en algún otro componente del sistema (por ejemplo un servidor que realiza el procedimiento de traducción, pero no

actúa como un relevador, por ejemplo, que regresa la traducción en el dispositivo de usuario de origen para la transmisión indirecta al dispositivo del usuario de destino). El método es un método implementado por computadora implementado por el código programado de manera adecuada cuando se ejecuta - p.ej. el código 110 cuando se ejecuta en el procesador 304 de la figura 3 y/o el código de cliente del cliente 118a y/o 118b. Esto es, el método se puede realizar en cualquier sistema de comunicación adecuado para efectuar una llamada de voz o vídeo entre un usuario de origen que habla un idioma de origen y un usuario de destino que habla un idioma de destino implementando alguna forma del discurso durante la llamada al procedimiento de traducción de discurso para la generación del discurso sintético, traducido en el idioma de destino para emitirlo al usuario de destino.

En una traducción de voz a voz que implica un procedimiento de traducción de voz a voz, el procedimiento de traducción global puede funcionar de la siguiente manera: un usuario de origen, por ejemplo, Alice está hablando en su propio idioma (de origen) – el sistema reconoce la voz, la traduce y envía el texto al traductor de voz para el oyente. Cuando está soportado por el vídeo, puede haber una demora (por ejemplo hasta unos pocos segundos) entre el momento en que la otra persona deja de hablar y se envía el audio traducido. Esto crea una gran confusión, haciendo difícil para el oyente entender cuando es seguro empezar a hablar sin interrumpir la conversación de su compañero.

En otras palabras, el discurso de Alice se compone generalmente de los intervalos de actividad de voz - en la cual Alice está hablando en el

idioma de origen - intercalados con intervalos de inactividad de voz en la cual Alice no está hablando, por ejemplo, porque ella está esperando a que Bob hable o porque ella está escuchando en ese momento lo que Bob está diciendo.

5 Con este fin, el método comprende la señalización de un cambio en el comportamiento del procedimiento de traducción, el cambio relativo a la generación de la traducción, y provocando de esta manera que una notificación sea emitida al usuario de destino para notificar al usuario de destino de la variación cuando la misma se detecta. La señalización puede
10 ser remota a través de la red 106 (si el procedimiento de traducción no se realiza en el dispositivo de destino). Asimismo puede haber algún beneficio para la misma o una notificación similar está siendo emitida al hablante de origen también - por ejemplo, si ven que el componente de traducción está ocupado realizando la traducción, pueden hacer una pausa, permitiendo
15 que su interlocutor se ponga al tanto, antes de continuar con el resto de lo que ellos están diciendo.

En el siguiente ejemplo, los posibles cambios señalados en el comportamiento incluyen procedimiento que comprende:

un estado de "escucha" ("espera"), en el cual no se genera o emite
20 actualmente una traducción, por ejemplo, porque no tiene nada que traducir (por ejemplo, entra cuando se ha terminado de traducir todo el discurso en el último intervalo de actividad de voz de Alice, y Alice se encuentra todavía en un intervalo de inactividad de conversación, es decir, no se ha reanudado el discurso, de manera que el procedimiento no tiene
25 nada que hacer en ese momento);

un estado "pendiente" ("traducción pasiva") en el cual Alice está hablando actualmente y el procedimiento está monitoreando (es decir, escucha) el discurso con el propósito de traducirlo (por ejemplo, entra, desde el estado de escucha cuando Alice vuelve a hablar) - también se
 5 puede generar una traducción provisional, parcial en este punto (ver anteriormente);

un estado de "pensamiento" ("traducción activa "), en el cual Alice puede no hablar actualmente, pero habló suficientemente recientemente para que el procedimiento procese su discurso más reciente con el fin de
 10 traducirlo (por ejemplo, entra del estado pendiente cuando Alice deja de hablar);

un estado "de conversación" ("de emisión") en el cual se emite actualmente la traducción de audio generada (por ejemplo, entra después de alcanzar un punto en el cual es posible, por ejemplo, cuando el
 15 procedimiento acaba de terminar la generación de una traducción del discurso de Alice hablado durante el último intervalo de actividad de voz de Alice).

un estado "de confusión" ("error"), en el cual el procedimiento no está en condiciones de proceder, por ejemplo, porque ha sido incapaz de
 20 realizar una traducción de voz o algún ha ocurrido otro error (entra en el punto cuando se reconoce un tal error).

En modalidades particulares, con acceso a la secuencia de vídeo de Bob (que no se ilustra en la figura 4A/B), el robot puede asumir la personalidad de un avatar de "presentación", que está animado de tal
 25 manera que es evidente cuando se está hablando, escuchando (en

espera), etc. Un avatar es una representación gráfica, generada artificialmente, de un personaje animado, por ejemplo, capaz de ser animado para transmitir el significado a través de señales visuales, tales como expresiones faciales, lenguaje corporal, otros gestos, etc. En este caso, el comportamiento del avatar se controla para que coincida con el comportamiento del procedimiento de traducción, es decir, el avatar imita efectivamente las señales visuales de un traductor humano real (cuando se realiza la traducción basada en turnos) o intérprete (al realizar la traducción continua), proporcionando una experiencia de usuario atractiva e intuitiva para el usuario de destino, así como haciendo que la información de la imagen de usuario está tratando de transmitir información fácilmente comprensible al usuario de destino. Por ejemplo, en una conversación con un traductor humano, el oyente prestará atención al traductor hasta que termine y después empezará a hablar; por medio de la señalización antes mencionada, el avatar se puede hacer para imitar este comportamiento haciendo que adopte una pose visual que indica que están escuchando a Alice cuando el procedimiento pasa al estado pendiente y causando el movimiento de sus labios después de que el procedimiento de traducción entra en el estado de conversación para coincidir con el inicio de la emisión de la traducción de audio.

Por lo tanto, el avatar se comporta como un traductor humano y proporciona señales visuales. Por ejemplo, una señal visual para indicar al oyente cuando es seguro empezar a hablar mediante la adopción de una pose para escuchar al entrar en el estado de escuchar. Por lo tanto, el cliente del usuario puede emitir, a través de los componentes del altavoz

una traducción audible en la lengua de destino del discurso del usuario de origen durante ese intervalo (es decir, una porción de traducción del discurso traducido correspondiente al discurso de origen en ese intervalo), y la emisión de una indicación para el usuario de destino (notificación)

5 cuando la emisión de la traducción audible (es decir, esta parte de la traducción) ha finalizado sustancialmente para indicar que el usuario de destino está libre para responder al usuario de origen. Aquí, "sustancialmente terminado" incluye cualquier punto en el tiempo suficientemente cerca para terminar la emisión de tal manera que es

10 seguro para que Bob comience a hablar sin interrumpir la secuencia natural de la conversación.

Como será evidente, los cambios en el estado del procedimiento de traducción (a base de turnos) mencionados anteriormente en el hecho de que refleja de cerca los cambios reales en el modo de pensar de un

15 traductor humano o intérprete (traducción simultánea) en una traducción en vivo en la vida real o escenario interpretación. Es decir, al igual que un procedimiento automatizado que funcionará en un estado de escuchar, en espera, pendiente, de conversación o confundido, también lo puede realizar la mentalidad de un ser humano de la vida real que hace lo mismo.

20 Esto se realiza por la configuración del avatar para aproximar diversas acciones que se deben realizar por un traductor humano cuando se comunica un cambio en su estado de ánimo en un escenario de traducción en la vida real, que cambia correspondiente al cambio en el comportamiento del procedimiento de traducción. Esto se explica con más

25 detalle a continuación con referencia, en particular, a las figuras 7A-E que

ilustran el comportamiento visual del avatar.

El avatar puede ser, por ejemplo, una representación de un ser humano, un animal o de otro carácter que tiene por lo menos una característica visual (por ejemplo, características faciales, partes corporales y/o aproximaciones de los mismos) que se pueden adoptar para 5 transmitir los señales visuales de una manera que al menos parcialmente imita el comportamiento humano esperado de un traductor humano.

En una conversación de vídeo de tres partes con discurso basado en el robot a la traducción del discurso, donde el robot está integrado en un 10 sistema de comunicación existente, puede haber dos videos y una imagen que se muestran en la pantalla por "defecto" (debido a que el sistema de comunicación simplemente trata el robot como si fuera otro usuario que no tiene capacidades de video, pero tiene una imagen estática asociada en el sistema de comunicación con su nombre de usuario): el video de la 15 persona que llama, el video de la persona llamada y una imagen estática que representa el robot de traducción.

Por ejemplo, en un discurso basado en vídeo al sistema de traducción de discurso (S2ST) incluyendo el vídeo de la interfaz de usuario del cliente de Bob podría mostrar el video del usuario integral (Alice), el 20 video del usuario integral (por ejemplo, en una porción más pequeña del área de visualización disponible para el video de Alice) y, por defecto, alguna imagen asociada con el nombre de usuario, por ejemplo un gráfico estático del robot automatizado. Cuando Alice habla en su propio idioma Bob puede ver visualmente el movimiento de los labios de Alice y espera 25 hasta que Alice termine de hablar. Posteriormente, el robot traductor

procesa el audio (reconocimiento y traducción) y empieza a hablar el idioma de Bob. Durante ese tiempo la persona que llama no tendría ninguna pista visual, siempre y cuando el procedimiento de traducción es completo y es seguro empezar a hablar. Esto es probable que genere
5 confusión para Bob.

De acuerdo con modalidades particulares, la idea es efectivamente reemplazar la imagen del robot traductor con un avatar, permitiendo de este modo:

- el uso de un avatar para el sistema de traducción de discurso a
10 discurso
- gestos del avatar para imitar lo que un traductor humano o intérprete haría.

Es decir, para evitar una tal confusión, la imagen estática se sustituye con un avatar, que se comporta de forma visual como un
15 traductor humano. Esto se puede llevar a cabo, por ejemplo, mediante la transmisión de una secuencia de vídeo de vídeo generado sintéticamente (generado en la forma descrita a continuación) del robot al usuario de destino como si se tratara de una secuencia de vídeo de otro usuario humano en la llamada de vídeo, y que se mostrará de forma automática a
20 través de la interfaz de usuario del cliente de forma automática (que no necesitaría modificación alguna al software de cliente y sería compatible con clientes de versiones anteriores). Alternativamente, el vídeo se puede generar en el dispositivo de destino en sí pero, sin embargo, muestra como si fuera un vídeo entrante de otro usuario (que puede necesitar
25 modificación alguna del software del cliente, pero que sería más eficiente

en términos de recursos de la red, ya que no sería necesario que el video de avatar se transmita a través de la red 106).

Las figuras 7A-E ilustran la pantalla del dispositivo del usuario de Bob 104b en varios puntos durante la llamada de video. Como se muestra, en cada uno de estos puntos de vídeo de Alice 702 como se refleja en su dispositivo 104a se muestra en una primera porción del área de visualización disponible junto con el vídeo del avatar sintético 704 que se muestra en una segunda porción del área de visualización disponible (las primera y segunda porciones siendo de tamaño similar), con el vídeo de Bob 706 capturado en el dispositivo 104b de Bob (y también se transmite a Alice) mostrado en la tercera parte del área de visualización disponible, debajo del video del avatar 704 (la tercera porción más pequeña que las primera y segunda porciones en este ejemplo). En este ejemplo, el avatar tiene una forma aproximadamente humana-masculina para propósitos de ilustración.

Volviendo a la figura 6, en el paso S600 de la figura 6, comienza el procedimiento de traducción durante una llamada. El procedimiento de traducción durante una llamada es de manera que el discurso de Alice se traduce del idioma de origen en un discurso sintético en el idioma de destino para mostrarlo a Bob durante, y como parte de, una llamada de voz o vídeo en la cual están participando al menos Alice y Bob.

En este ejemplo, el procedimiento de traducción comienza en el estado "escuchando", que está señalizado al componente de notificación 804 (S602). En este caso, el avatar se controla mediante el componente de notificación 804 en el vídeo sintético a adoptar una pose cuando se está

escuchando tal como se ilustra en la figura 7A.

En el paso S604, el componente traductor detecta si Alice ha empezado a hablar, por ejemplo, mediante la supervisión del audio de la llamada recibido de Alice y la realización de la detección de actividad de voz (VAD) del mismo. El avatar permanece en la pose escuchando durante el tiempo que el procedimiento de traducción permanece en el estado escuchando, que será el caso hasta que Alice empiece a hablar. Cuando se detecta que Alice ha comenzado hablar, el traductor 802 señala al componente de notificación 804 que el procedimiento de traducción ha entrado en el "estado pendiente" (S606), por ejemplo, en el cual se monitorea el discurso de Alice para traducirlo, comenzando los preparativos para la traducción, o realizando una traducción del discurso que puede estar sujeto a modificaciones una vez que se recibe el discurso (mientras el discurso posteriormente puede proporcionar un contexto que tiene impacto en el reconocimiento o la traducción del discurso anterior). En respuesta, el componente de notificación controla el comportamiento del avatar para adoptar un comportamiento visual de escucha, por ejemplo, cuando el usuario integral está hablando, el avatar presta atención a Alice, por ejemplo, girando a su vez su rostro hacia el vídeo de Alice. Esto se ilustra en la figura 7B.

La figura 7B ilustra un ejemplo en el cual la información de diseño integral con relación a las posiciones relativas de Alice y videos del avatar en el área de pantalla disponible del dispositivo de destino se puede utilizar para influir en la generación del vídeo del avatar mismo. En el ejemplo de la figura 7B, el video del avatar se muestra a la derecha del

vídeo de Alice – la información de diseño que transmite este posicionamiento relativo se retroalimenta desde el componente de emisión de notificaciones 806 al componente de generación de notificación 804. A partir de esa información, el componente de generación de notificación 804
5 controla el vídeo del avatar en el traductor que entra en el modo "pendiente" para mover los ojos del avatar a la izquierda, asegurando de esta manera para que se dirijan a la parte de la pantalla en la cual se muestra el vídeo de Alice en el dispositivo de destino para dar la impresión de que el avatar mira a Alice y le presta atención. Por lo tanto, la
10 información relacionada de diseño se utiliza para hacer la experiencia de usuario más natural para Bob, haciendo que el comportamiento del avatar sea natural e intuitivo.

En el paso S606, se determina si Alice sigue hablando (es decir, si es o no se ha hecho una pausa para un tiempo suficiente (por ejemplo
15 predeterminado) transcurrido desde el comienzo de su más reciente intervalo de actividad de voz), por ejemplo, usando el VAD. Durante el tiempo que Alice sigue hablando, el procedimiento de traducción permanece en el estado "pendiente" y el avatar de este modo continúa mostrando el comportamiento de escucha. Cuando Alice para de hablar, el
20 procedimiento de traducción pasa al estado "pensando" durante el cual se realiza el procesamiento para los propósitos de emitir la traducción final de audio del intervalo de expresión más reciente de Alice. Este se indica en el componente de notificación (S610) y, en respuesta, el componente de notificación hace que el avatar adopte un comportamiento visual para
25 transmitir las acciones de pensamiento, por ejemplo, el avatar podría

adoptar una actitud de pensamiento, por ejemplo, poniendo su mano cerca de su barbilla o imitando una cara de pensamiento - esto se ilustra en la figura 7C.

El avatar se mantiene en esta postura mientras que el procedimiento de traducción realiza el procesamiento; cuando el procesamiento se ha completado, el proceso de traducción pasa al estado "hablando" y comienza a emitir el audio traducido ahora listo (ver S610). Esto se señala en el paso S616, y en respuesta el avatar se controla para adoptar un estado visual de habla, por ejemplo, cuando se expone la traducción, el avatar podría prestar atención (girar la cara hacia) al usuario del extremo cercano (es decir, mirar directamente fuera de la pantalla) y mostrar los labios que hablan (es decir, movimiento de los labios). Esto se ilustra en la figura 7D. El avatar permanece en este estado durante el tiempo que el traductor permanece en estado de habla (es decir, durante el tiempo en el que se emite el audio traducido); tras la finalización de dicha emisión, el traductor vuelve a entrar en el estado de escucha (ver S620).

Si algo va mal durante el procesamiento, este traductor vuelve a entrar en estado "confundido", que es rascándose la cabeza o algún otro estado visual de confusión. Esto se ilustra en la figura 7E. Además, cuando el avatar también se muestra en el dispositivo de Alice, el avatar puede "pedir" Alice que repita (es decir, que lo diga de nuevo, Lo siento, yo no lo entiendo, etc., - es decir, una solicitud de audio se puede emitir para Alice en el idioma de origen pidiéndole que repita lo que acaba de decir.

Una pieza de información que se transmite por el avatar utilizando la

información visual es para indicar cuando el usuario de destino está libre para empezar a hablar - el punto en el cual los labios de los avatares dejan de moverse constituye una indicación visual que lo transmite.

El comportamiento del avatar también puede ser influenciado por
5 otros comportamientos, por ejemplo, otros eventos. Por ejemplo, el
componente de generación de notificaciones 804 puede también recibir
información relativa a Bob, por ejemplo, en relación con el comportamiento
de Bob (además de recibir la información relativa a Alice, la cual, en este
caso se recibe por medio de la información relacionada con el
10 procedimiento de traducción se realiza en el discurso de Alice). Por
ejemplo, el discurso de Bob también se puede analizar para detectar
cuando Bob empieza a hablar, en dicho punto el avatar puede ser el
controlador para mirar el vídeo de Bob 706 como se muestra en la pantalla
de Bob. La información de diseño retroalimentada relacionada con la
15 posición del vídeo de Bob en su pantalla también se puede utilizar para
controlar el comportamiento del avatar - por ejemplo, en los ejemplos de
las figuras 7A-E, el vídeo de Bob se muestra encima del vídeo del avatar
704 y con base en esto, el avatar puede ser controlado para mirar hacia
abajo cuando Bob está hablando de este modo que parece que mira hacia
20 Bob.

Si bien se ha descrito con referencia a un robot, hay que señalar que
la materia descrita en relación con las figuras 6, 7A-E y 8 también se
aplica a los sistemas que no se basan en robots - aquí el avatar se puede
configurar para comportarse de la misma manera, pero estará
25 representando efectivamente otro servicio de traducción (por ejemplo,

servicio de traducción basado en la nube) diferente a un robot (con un identificador de usuario asignado y que aparece, por lo tanto, como un usuario del sistema de comunicación) *per se*.

Por otra parte, mientras en la anterior, una notificación constituye
5 una notificación visual transmitida por un avatar animado (es decir, incorporado en el vídeo del avatar), la notificación, en otras modalidades, puede tomar cualquier forma deseada, por ejemplo, en forma de un icono que cambia de forma, color, etc. en la pantalla (por ejemplo, por medio de una representación animada de una luz que cambia de rojo a verde cuando
10 sea seguro para Bob hablar), o una emisión de indicación audible (por ejemplo, un tono u otro icono sónico) a través de los altavoces, o una notificación táctil efectuada mediante el accionamiento, por ejemplo, de un componente de vibración que provoca la vibración física, táctil del dispositivo de usuario de Bob, y/u otro componente mecánico de dicho
15 dispositivo. Alas notificaciones de audio y/o táctiles pueden ser particularmente útiles para dispositivos móviles.

Como se ha mencionado, mientras que anteriormente se ha descrito en relación con la traducción de un solo sentido para la simplicidad, la traducción de dos vías se puede realizar con traducciones separadas e
20 independientes que se realizan en cada secuencia de audio de la llamada individual. Además, mientras anteriormente se ha descrito con referencia a las llamadas con dos participantes humanos, también se contemplan las llamadas entre cualquier número ($n < 2$) de los participantes humanos con hasta n vías de traducción (por ejemplo, si todos los usuarios n hablan
25 diferentes idiomas). Las traducciones independientes para cada uno de

varios seres humanos entre sí durante una llamada de n vías en las secuencias de audio separadas de los diferentes participantes humanos en beneficio (por ejemplo, para transmitir a) uno o más de los otros participantes humanos. Además, la traducción en el idioma de destino se
5 puede transferir a múltiples usuarios objetivo que hablan el idioma de destino.

Las referencias a secuencias de medios (por ejemplo, audio/video) (o similares) se refieren a una transmisión de medios (por ejemplo, audio/video) a través de una red de comunicación a un dispositivo para su
10 emisión en ese dispositivo, ya que se recibe, en contraste con los medios que se reciben en su totalidad antes de comenzar la emisión de los mismos. Por ejemplo, cuando se genera una secuencia de audio o vídeo sintético, los medios se transmiten mientras se generan en el dispositivo para la emisión, ya que es como se recibe (y por lo tanto, a veces,
15 mientras se está generando).

De acuerdo con otro aspecto de la materia, la descripción considera un método realizado en un sistema de comunicación en el cual los usuarios se identifican de forma única por identificadores de usuario asociados, el sistema de comunicaciones para efectuar una llamada de voz o vídeo entre
20 un usuario de origen que habla un idioma de origen y un usuario de destino que habla un idioma de destino, el código informático soportado por el sistema de comunicación configurado para implementar un agente de traductor, el agente de traductor identificado de forma única por un identificador de usuario asociado, facilitando sustancialmente de este
25 modo la comunicación con el agente como si fuera otro usuario del sistema

de comunicación, comprendiendo el método: recibir una solicitud de traducción que solicita que el agente del traductor participe en la llamada; responder a la recepción de la solicitud, incluyendo una instancia del agente del traductor como un participante en la llamada, en la cual la

5 instancia del agente de traductor está configurada cuando se incluye para causar las operaciones de: recibir un audio de la llamada desde el usuario de origen, el audio de la llamada comprende el discurso del usuario de origen en el idioma de origen, realizar un procedimiento automático de reconocimiento de voz en el audio de la llamada, el procedimiento de

10 reconocimiento de voz está configurado para el reconocimiento del idioma de origen y el uso de los resultados del procedimiento de reconocimiento de voz para proporcionar al usuario de destino una traducción del discurso del usuario de destino en el idioma de destino.

El agente puede aparecer como otro miembro del sistema de comunicación (en virtud de su identificador de usuario asociado), por

15 ejemplo, en las listas de contactos del usuario o la naturaleza del robot puede estar oculta en el nivel de interfaz de usuario.

De acuerdo con todavía otro aspecto de la materia, se describe un sistema informático para su uso en un sistema de comunicación, el sistema

20 de comunicación para efectuar una llamada de voz o de vídeo entre al menos un usuario de origen que habla un idioma de origen y un usuario de destino que habla un idioma de destino, el sistema informático comprende: uno o más componentes de salida disponibles para el usuario de destino; un componente de emisión de traducción configurado, al menos por un

25 intervalo de actividad de voz de usuario de origen, a la salida a través de

los componentes de salida del audio de una traducción audible en el idioma de destino del discurso del usuario de origen durante ese intervalo; y un componente de emisión de notificaciones configurado para emitir una notificación al usuario de destino cuando la emisión de la traducción audible ha terminado sustancialmente para indicar que el usuario de destino está libre para responder al usuario de origen.

De acuerdo con otro aspecto todavía de la materia, un dispositivo de usuario comprende: uno o más componentes de salida de audio; un componente de visualización para la salida de información visual a un usuario de destino del dispositivo de usuario; almacenamiento informático que contiene el software de cliente para efectuar una llamada de voz o de vídeo entre el usuario de destino y un usuario de origen de otro dispositivo de usuario, el usuario de origen habla un idioma de origen y el usuario de destino habla un idioma de destino; una interfaz de red configurada para recibir el audio de la llamada a través de una red de comunicaciones el audio de la llamada de voz comprende el discurso del usuario de origen en el idioma de origen durante los intervalos de actividad de voz del usuario de origen; uno o más procesadores configurados para ejecutar el software de cliente, el software de cliente está configurado cuando se ejecuta para realizar las operaciones de: emitir la llamada de audio recibida a través de los componentes de salida de audio, durante al menos un intervalo de actividad de voz del usuario de origen, emitiendo a través de los componentes de salida del audio una traducción audible en el idioma de destino del discurso de usuario de origen durante ese intervalo, emitir para el usuario de destino una indicación cuando la emisión de esa traducción

sonora ha terminado sustancialmente para indicar que el usuario de destino está libre para responder al usuario de origen.

En general, cualquiera de las funciones descritas en este documento se pueden implementar utilizando software, firmware, hardware (por ejemplo, circuitería lógica fija) o una combinación de estas implementaciones. Los términos "módulo", "funcionalidad", "componente" y "lógica", como se usa en este documento representan generalmente software, firmware, hardware, o una combinación de los mismos (por ejemplo, los bloques funcionales de las figuras 4A, 4B y 8). En el caso de una implementación de software, el módulo, la funcionalidad, o lógica representan el código de programa que realiza tareas especificadas (por ejemplo, los pasos del método de las figuras 5 y 6) cuando se ejecuta en un procesador (por ejemplo CPU o varios CPU). El código de programa se puede almacenar en dispositivos de memoria legibles por una o más computadoras. Las características de las técnicas descritas a continuación son independientes de la plataforma, lo que significa que las técnicas se pueden implementar en una variedad de plataformas informáticas comerciales que tienen una variedad de procesadores.

Por ejemplo, los dispositivos de usuario pueden incluir también una entidad (por ejemplo, software, tales como el cliente 118) que causa que el hardware de los dispositivos de usuario realicen las operaciones, por ejemplo, bloques funcionales de procesadores, y así sucesivamente. Por ejemplo, los dispositivos de usuario pueden incluir un medio legible por computadora que puede ser configurado para mantener instrucciones que hacen que los dispositivos de usuario, y más particularmente el sistema

operativo y hardware asociado de los dispositivos de usuario realicen las operaciones. Por lo tanto, las instrucciones funcionan para configurar el sistema operativo y el hardware asociado para realizar las operaciones y de esta manera resultan en la transformación del estado del sistema operativo y hardware asociado para realizar dichas funciones. Las instrucciones se pueden proporcionar por el medio legible por computadora para los dispositivos del usuario a través de una variedad de configuraciones diferentes.

Una tal configuración de un medio legible por computadora es el medio que soporta la señal y, por lo tanto, está configurado para transmitir las instrucciones (por ejemplo, como una onda portadora) al dispositivo informático, tal como a través de una red. El medio legible por computadora también puede estar configurado como un medio de almacenamiento legible por computadora y, por lo tanto, no es un medio que soporta la señal. Ejemplos de un medio de almacenamiento legible por computadora incluyen una memoria de acceso aleatorio (RAM), memoria de sólo lectura (ROM), un disco óptico, memoria flash, memoria de disco duro y otros dispositivos de memoria que pueden utilizar técnicas magnéticas, ópticas y otras para almacenar las instrucciones y otros datos.

En las modalidades del primer aspecto como se establece en la sección de breve descripción, el cambio en el comportamiento puede ser uno de:

- el procedimiento de traducción entra en un estado de escucha, en el cual está actualmente en espera de la futura actividad de discurso por el usuario de origen durante un intervalo actual de inactividad de

conversación por el usuario de origen.

- el procedimiento de traducción entra en un estado pasivo de traducción sensible al usuario de origen que comienza un período de actividad de voz, en el cual el procedimiento de traducción está
5 monitoreando la actividad de voz actual por el usuario de origen en el audio de la llamada

- el procedimiento de traducción entra en un estado de traducción activo sensible al usuario de origen que termina un intervalo de actividad de voz, en el cual el procedimiento de traducción está generando
10 actualmente una traducción de audio del discurso del usuario de origen en ese intervalo a ser emitido cuando se haya completado esa generación.

- el procedimiento de traducción entra en un estado de emisión sensible al procedimiento de traducción que completa la generación de una traducción de audio del discurso de usuario de origen durante un intervalo
15 anterior de actividad de habla del usuario de origen, en el cual la traducción de audio generada está siendo enviada por el procedimiento de traducción para que sea emitido al usuario de destino.

- el procedimiento de traducción entra en un estado de error sensible al procedimiento que encuentra un error en la generación de la traducción.

20 El audio traducido puede ser transmitido a través de una red de comunicación, ya que se genera en un dispositivo de destino del usuario de destino para su emisión a través de uno o más componentes de salida del audio de este dispositivo tal y como se recibe.

El vídeo sintético se puede generar en función del cambio señalado
25 en el comportamiento, el video sintético para la visualización en un

dispositivo de usuario de destino del usuario de destino y que incorpora la notificación. El vídeo sintético puede ser un avatar animado que realiza acciones visuales, la notificación realizada como una acción visual por el avatar. La acción del avatar puede aproximar una acción esperada para
5 ser realizada por un traductor o intérprete humano cuando se comunica un cambio en su estado de ánimo en un escenario de traducción o interpretación de la vida real, dicho cambio corresponde al cambio en el comportamiento del procedimiento de traducción.

La notificación puede comprender una notificación visual para la
10 visualización en un dispositivo de usuario de destino del usuario de destino y/o una notificación de audio para la reproducción en el dispositivo del usuario de destino y/o una notificación táctil emitida por el accionamiento de un componente mecánico del dispositivo del usuario de destino.

15 En las modalidades del segundo aspecto, el audio de la llamada puede comprender el discurso del usuario de origen en el idioma de origen durante los intervalos de actividad de habla del usuario de origen intercalados con intervalos de inactividad de habla en los cuales el usuario de origen no está hablando; al menos durante un intervalo de actividad de
20 habla del usuario de origen, el componente de emisión de la traducción se puede configurar para emitir a través de los componentes de salida de audio una traducción de audio del discurso del usuario de origen durante ese intervalo y el componente de emisión de notificación se puede configurar para emitir la notificación cuando la emisión de la traducción ha
25 terminado sustancialmente para indicar que el usuario de destino está libre

de responder al usuario de origen.

El sistema informático puede ser realizado por un dispositivo del usuario de destino del usuario de destino o por una combinación del dispositivo del usuario de destino y al menos un otro dispositivo informático para que el dispositivo de usuario de destino se conecte a través de una red de comunicación.

El sistema informático puede comprender una entrada configurada para recibir una señal que apunta al cambio en el comportamiento de procedimiento de traducción; y un componente de generación de notificaciones configurado para generar la notificación en función de la señal recibida.

El componente de emisión de notificaciones se puede configurar para generar la información relacionada con la emisión definiendo la forma en la cual se debe emitir la notificación para el usuario de destino; y el componente de generación de notificación puede estar configurado para generar la notificación en función de la información relacionada con los resultados.

El sistema informático puede comprender una pantalla disponible para el usuario de destino y la notificación puede comprender una notificación visual que se mostrará en la pantalla y la información relacionada con la emisión comprende información de diseño relacionada. El componente de generación de notificaciones puede estar configurado para generar el vídeo sintético que incorpora la notificación, el vídeo sintético generado en función de la información de diseño. El vídeo sintético puede ser de un avatar animado que lleva a cabo acciones

visuales, la notificación configurada como una acción de avatar visual que se controla en función de la información de diseño.

Aunque la materia ha sido descrita en un lenguaje específico para las características estructurales y/o actos metodológicos, se debe entender
5 que la materia definida en las reivindicaciones adjuntas no se limita necesariamente a las características o actos específicos descritos anteriormente. Por el contrario, las características y los actos específicos descritos anteriormente se describen como formas ejemplares de aplicación de las reivindicaciones.

10

15

20

25