



- (51) **International Patent Classification:**
G10L 15/18 (2013.01) *G06F 17/30* (2006.01)
G10L 15/22 (2006.01)
- (21) **International Application Number:**
PCT/US2015/042048
- (22) **International Filing Date:**
24 July 2015 (24.07.2015)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
14/447,737 31 July 2014 (31.07.2014) US
- (71) **Applicant:** **GOOGLE INC.** [US/US]; 1600 Amphitheatre Parkway, Mountain View, California 94043 (US).
- (72) **Inventors:** **CHEN, Johnny**; 1600 Amphitheatre Parkway, Mountain View, California 94043 (US). **DEAN, Thomas L.**; 1600 Amphitheatre Parkway, Mountain View, California 94043 (US). **LAU, Qiangfeng Peter**; 1600 Amphitheatre Parkway, Mountain View, California 94043 (US). **GANDHE, Sudeep**; 1600 Amphitheatre Parkway, Mountain View, California 94043 (US). **SCHINE, Gabriel**;

1600 Amphitheatre Parkway, Mountain View, California 94043 (US).

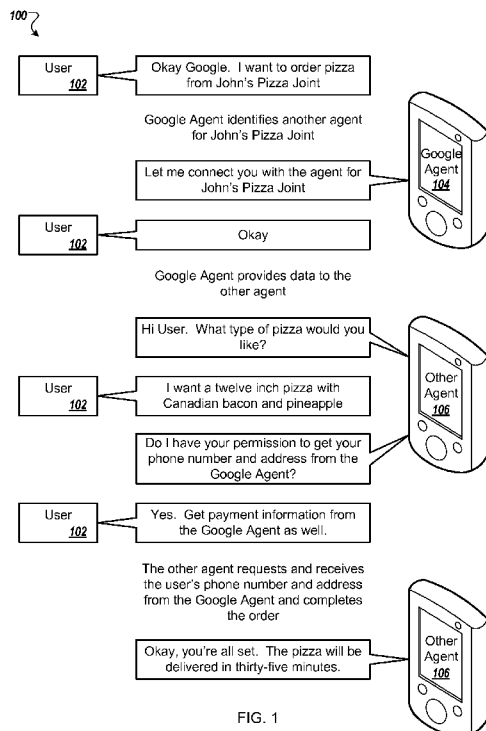
- (74) **Agents:** **COUENHOVEN, Joel D.** et al.; Fish & Richardson P.C., P.O. Box 1022, Minneapolis, Minnesota 55440-1022 (US).

- (81) **Designated States** (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

- (84) **Designated States** (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK,

[Continued on next page]

- (54) **Title:** CONVERSATIONAL AGENTS



- (57) **Abstract:** Methods, systems, and apparatus, including computer programs encoded on computer storage media, for handing off a user conversation between computer-implemented agents. One of the methods includes receiving, by a computer-implemented agent specific to a user device, a digital representation of speech encoding an utterance, determining, by the computer-implemented agent, that the utterance specifies a requirement to establish a communication with another computer-implemented agent, and establishing, by the computer-implemented agent, a communication between the other computer-implemented agent and the user device.



SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,
GW, KM, ML, MR, NE, SN, TD, TG).

— *as to the applicant's entitlement to claim the priority of
the earlier application (Rule 4.17(iii))*

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted
a patent (Rule 4.17(ii))*

Published:

— *with international search report (Art. 21(3))*

CONVERSATIONAL AGENTS

BACKGROUND

[0001] This specification relates to computer-implemented conversational agents.

[0002] A computer-implemented agent may interact with a user to provide the user with an interface for features of a device on which the computer-implemented agent executes or with which the computer-implemented agent is associated. For example, a computer implemented agent may execute on a mobile device, receive voice input from a user, and execute one or more actions in response to the voice input, such as retrieving search results responsive to the voice input and providing the search results to the user.

SUMMARY

[0003] In general, one innovative aspect of the subject matter described in this specification can be embodied in methods that include the actions of receiving, by a computer-implemented agent specific to a user device, a digital representation of speech encoding an utterance, determining, by the computer-implemented agent, that the utterance specifies a requirement to establish a communication with another computer-implemented agent, and establishing, by the computer-implemented agent, a communication between the other computer-implemented agent and the user device. Other embodiments of this aspect include corresponding computer systems, apparatus, and computer programs recorded on one or more computer storage devices, each configured to perform the actions of the methods. A system of one or more computers can be configured to perform particular operations or actions by virtue of having software, firmware, hardware, or a combination of them installed on the system that in operation causes or cause the system to perform the actions. One or more computer programs can be configured to perform particular operations or actions by virtue of including instructions that, when executed by data processing apparatus, cause the apparatus to perform the actions.

[0004] The foregoing and other embodiments can each optionally include one or more of the following features, alone or in combination. The other computer-implemented agent might not be specific to the user device. The other computer-implemented agent may be different from the computer-implemented agent. Establishing the communication with the other computer-implemented agent and the user device may include establishing the communication with the other computer-implemented agent

executing on another device and the user device. Establishing the communication with the other computer-implemented agent and the user device may include establishing the communication with the other computer-implemented agent and the computer-implemented agent. The computer-implemented agent may be for a first application on a user device and the other computer-implemented agent may be for a second, different application on the user device.

[0005] In some implementations, the method includes providing a representation of the utterance to the other computer-implemented agent. The representation may include a text representation of the utterance. The method may include receiving, by the computer-implemented agent from the other computer-implemented agent, another representation of a response to the utterance, and generating, using the other representation, data to cause a presentation of the response. The method may include determining, by the computer-implemented agent, settings, associated with the other computer-implemented agent, for the presentation of the response. The representation may include the digital representation of speech. Providing the representation of the utterance to the other computer-implemented agent may include providing data for a user who spoke the utterance to the other computer-implemented agent. Providing data for the user who spoke the utterance to the other computer-implemented agent may include determining privacy settings for the user, and providing the data for the user to the other computer-implemented agent according to the privacy settings.

[0006] In some implementations, the method may include receiving data representing properties of the other computer-implemented agent prior to receiving the digital representation of speech encoding the utterance. Receiving the data representing the properties of the other computer-implemented agent may include receiving a definition of a data flow for the other computer-implemented agent that identifies a response for the other computer-implemented agent given a particular type of input. Receiving the data representing the properties of the other computer-implemented agent may include receiving an identification of a particular style of speech for the other computer-implemented agent. Receiving the data representing the properties of the other computer-implemented agent may include receiving text-to-speech parameters for responses provided by the other computer-implemented agent. Receiving the data representing the properties of the other computer-implemented agent may include receiving a selection of a template that identifies a data flow for computer-implemented agents that identifies a type of response for the other computer-implemented agent given

a particular type of input. The template may be for computer-implemented agents that provide online ordering services. The online ordering services may include online pizza delivery services.

[0007] The subject matter described in this specification can be implemented in particular embodiments so as to realize one or more of the following advantages. In some implementations, a system enables third parties to build voice input and output interfaces, e.g., computer-implemented conversational agents, for user devices, e.g., using an application programming interface. In some implementations, a system enables a computer-implemented agent to automatically initiate conversation between another computer-implemented agent and a user, e.g., to hand off a conversation, based on input received from the user, e.g., specific to the other computer-implemented agent. In some implementations, a hand off of a conversation between agents enables users who cannot, or do not wish to, physically interact with a device to interact with or receive information from different entities while using the device. For instance, a user may be disabled or a device may not include a touch interface, e.g., a touch screen or a keyboard.

[0008] The details of one or more embodiments of the subject matter of this specification are set forth in the accompanying drawings and the description below. Other features, aspects, and advantages of the subject matter will become apparent from the description, the drawings, and the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0009] FIG. 1 is an example environment of a hand off between computer-implemented conversational agents.

[0010] FIG. 2 is an example system in which an entity creates an agent that may receive a conversation hand off from a device agent.

[0011] FIG. 3 is a flow diagram of a process for establishing communication between a user device and a computer-implemented conversational agent.

[0012] FIG. 4 is a block diagram of a computing system that can be used in connection with computer-implemented methods described in this document.

[0013] Like reference numbers and designations in the various drawings indicate like elements.

DETAILED DESCRIPTION

Overview

[0014] A system may provide an extensible, feature rich, computer-implemented conversational agent platform that enables entities to quickly build conversational agents that can interact with users via a voice-in/voice-out interface. Each entity may use common features to create a custom conversational agent that helps users accomplish various tasks and achieve various objectives, e.g., associated with the entity.

[0015] The system may provide a user interface that enables third parties to author conversational agents, e.g., a set of conversation dialogs the conversational agent will use when interacting with a user. The third parties may create a conversational agent to describe a business of the third party, to interact with the users or customers of the third party, or both. The conversational dialogs may specify particular responses or questions for the conversational agent based on input received from the user.

[0016] Each of the third parties may specify a purpose for their conversational agent, e.g., to make airline reservations, tell jokes, allow a user to purchase movie tickets, etc. A third party may select a template, e.g., a default conversational flow, for a conversational agent that corresponds with the purpose of the conversational agent. For instance, a company that provides airline reservations may select a template for reservation conversational agents or airline reservation conversational agents and customize the template for their specific use or company, e.g., by customizing prompts and responses provided to a user.

[0017] Some of the customizations may include changing settings of the conversational agent, such as a particular style of speech, e.g., a particular language use or ways the agent responds to particular situations with the user, a particular voice, e.g., whether the agent has a male or female voice or a sound of speech, a particular language, e.g., as text-to-speech parameters, certain idioms and vocabulary, and/or custom commands for the agent, e.g., voice commands. In some examples, settings for an airline agent may indicate that the airline agent is very professional and business oriented and settings for a comedy club's agent may indicate that the comedy club agent is bubbly, edgy, and funny.

[0018] The system may include a platform that converts an authored conversational flow for a third party agent into a natural conversational user interface with a user. For instance, the platform may use the text-to-speech parameters for the

agent and the agent's voice settings to convert a text response specified by the agent's conversational flow to an audio stream and audibly present the audio stream to a user.

[0019] The system may include automated error correction techniques that enable the system to interact with a user to determine an appropriate action in response to input that may not be clear. Some error correction techniques may prompt the user with yes-or-no questions or by offering a set of options from which the user can select. The questions and multiple choice queries may enable the system to resolve ambiguous references, e.g., "what does 'that' refer to?", assess the user's sentiment, e.g., "does that mean you like the last song?", and otherwise clarify the user's intent, e.g., "what did you mean by 'LOL'".

[0020] For example, one potential source of error may include unintelligible speech from a user, e.g., when the system's speech recognition system is unable to determine what the user said. In this example, one error correction technique may be to prompt the user to repeat what he said. Another potential source of error may include receipt of a command by the system with which the system cannot determine an appropriate response. In this example, the system may provide the user with some options from which he can select, e.g., options commonly selected by other users.

[0021] The system may provide error correction techniques to third parties so that each third party does not have to create error correct techniques themselves. For instance, a third party may select one or more error correction techniques specific to their industry, specific to a particular conversation flow template, or from a list.

[0022] The platform may allow a conversational agent for a user device to "hand off" a user conversation to another conversational agent, e.g., a specific third party conversational agent. For instance, the user may ask the conversational agent for the user device to recommend a restaurant and walk the user through the process of selecting a restaurant. Once the conversational agent for the user device receives a restaurant selection from the user, the conversational agent for the user device may ask the user whether the user would like to connect with a conversational agent for the selected restaurant or with a restaurant reservation conversational agent.

[0023] The third parties may make third party data available to the platform to allow the platform to use the data during a conversation flow. For example, a pizza delivery third party may provide the platform with access to the types of pizzas currently available to ensure that the platform does not allow a user to order a pizza that is not currently available, e.g., when some ingredients for specific types of pizzas are not available.

[0024] The system may also allow third party agents to hand off conversations to other third party agents. In some examples, the conversational agent for a user device may receive a request for a restaurant recommendation and may hand off the conversation to a restaurant's agent that provides the user with menu information. The restaurant's agent may then receive a request to make a reservation and hand off the conversation to a restaurant reservation agent, created by the same third party restaurant or by another third party, e.g., that manages restaurant reservations.

Example Operating Environment

[0025] FIG. 1 is an example environment 100 of a hand off between computer-implemented conversational agents. For example, a user 102 may initiate a conversation with a computer-implemented conversational agent 104 by saying "okay Google. I want to order pizza from John's Pizza Joint." The agent 104 receives the input, analyzes the input, and determines that another computer-implemented conversational agent 106 for John's Pizza Joint is available to assist the user in ordering a pizza from John's Pizza Joint. The agent 104 responds to the user 102 stating "let me connect you with the agent for John's Pizza Joint."

[0026] The agent 104 receives input from the user 102 saying "okay" and provides data to the other agent 106 about the conversation, e.g., that the user 102 wants to order a pizza, as part of a process to hand off the conversation to the other agent 106. The data does not include any information that identifies the user 102 unless the user 102 previously gave the agent 104 permission to release the user's identity to any other agents or to the specific other agent 106 for John's Pizza Joint, e.g., explicitly or in a setting.

[0027] The other agent 106 then communicates with the user 102, e.g., by asking the user 102 "what type of pizza would you like?" The other agent 106, and the agent 104, may interact with the user 102 through a platform, executing on a user device operated by the user 102 or on a server, that provides a text-to-speech interface for the agents, as described in more detail below.

[0028] The other agent 106 may receive a response from the user 102 indicating that the user 102 "wants a twelve inch pizza with Canadian bacon and pineapple" and prompt the user 102 whether the other agent 106 "has your permission to get your phone number and address from the Google Agent." When the other agent 106 receives a response "yes. Get payment information from the Google Agent as well" from the user 102, the other agent 106 retrieves the phone number, address, and payment information

from the agent 104 and completes the order for the user 102. The other agent 106 may then inform the user 102 “okay, you’re all set. The pizza will be delivered in thirty-five minutes.”

[0029] When the other agent 106 does not receive permission from the user 102 to retrieve the user’s phone number and address from the agent 104, the other agent 106 prompts the user 102 for his address, and phone number. The other agent 106 may prompt the user 102 for payment information. The other agent 106 may then provide the user 102 with delivery information, such as when the pizza is expected to be delivered.

Example System

[0030] FIG. 2 is an example system 200 in which an entity creates an agent that may receive a conversation hand off from a device agent. For instance, an administrator for John’s Pizza Joint may access a conversation server 202, and a conversational agent platform 203 on the conversation server 202, to create a computer-implemented conversational agent 208. The administrator may access a conversational agent application programming interface (API) 204 that includes templates 206 for different types of conversational agents. The administrator may use the agent API 204 to create the conversational agent 208 using a template or without using a template. Some templates may include conversation data flows for online ordering, for any type of food service or particular types of food services, e.g., pizza delivery, ordering tickets, e.g., to a game or show or for an airline, getting movie recommendations or watching movies or television shows, e.g., online, etc.

[0031] The administrator may define a conversation data flow 210 and agent settings 212 for the conversational agent 208. The conversation data flow may include a series of questions and responses for the conversational agent 208 to use during a conversation with a user, e.g., based on expected prompt types or expected response types from the user. The conversation data flow 210 may include multiple different conversation data flows, e.g., for different interactions with a user. A first conversation data flow may allow the conversational agent 208 to provide a user with a menu for John’s Pizza Joint, a second conversation data flow may provide a user with an interface to order a pizza from John’s Pizza Joint, and a third conversation data flow may include multiple different flows depending on different types of input from the user, e.g., whether or not the user would like to order a side.

[0032] The agent settings 212 may include a particular style of speech, a particular voice, a particular language, custom commands, or a combination of any two or more of these. For example, the administrator may select a deep male voice with a certain accent as the particular voice. Some examples of styles of speech include serious, formal, bubbly, funny, and casual.

[0033] Data representing the conversational agent 208 may be stored on the conversation server 202, e.g., one or more computers that include data for multiple conversational agents, a server 214 for John's Pizza Joint, or both. The server 214 may be administered by the John's Pizza Joint entity, or may be managed on behalf of the entity by a third party.

[0034] When a user device 226 accesses a web page or an application 216 for John's Pizza Joint, a device agent 222, associated with the user device 226, may ask a user of the user device 226 if they would like to speak with the conversational agent 208, e.g., instead of using a visual user interface. When the user device 226 accesses an application 216 for John's Pizza Joint, the user device 226 may communicate with the application 216 executing on John's Pizza Joint's server 214 or may install the application, e.g., as John's Pizza Joint application 232 on the user device 226. The installation of the application 232 on the user device 226 may include data for the conversational agent 208 and when the user launches the application 232, e.g., by instructing the device agent 222 to launch the application 232, the conversational agent 208 may initiate a conversation with the user, e.g., the device agent 222 may hand off the user conversation to the conversational agent 208.

[0035] When the conversational agent 208 interacts with the user device 226, the conversational agent 208 may access user accounts 218 and order history data 220 for John's Pizza Joint. For example, the conversational agent 208 may ask a user of the user device 226 for account information, access that user's account using the account information, and determine recent orders for that user which are listed in the order history data 220. The conversational agent 208 may use the order history data 220 and the user accounts 218 to provide a recommendation on which pizza to order, to facilitate the order process, e.g., by determining the user's address rather than prompting the user for his address, or both.

[0036] The conversational agent 208 may gain access to the user's account or the order history data 220 by communicating with a computer-implemented device conversational agent 222 specific to the user device 226. The device agent 222 may

execute on the conversation server 202, and be associated with the user device 226 by a device identifier 224 for the user device 226, or may execute on the user device 226.

When the device agent 222 receives a request for data, e.g., user profile data 228, from the conversational agent 208, the device agent 222 may access privacy settings 230 for the user device 226, which are stored on the user device 226, the conversation server 202, or both, to determine what information may be shared with the conversational agent 208.

[0037] In some examples, the privacy settings 230 may indicate that the device agent 222 may provide the conversational agent 208 with the user's name, address, and phone number. The conversational agent 208 may use the user's name, address, and phone number to determine an account for the user and the user's order history.

[0038] The privacy settings 230 may indicate that particular data may be shared with some conversational agents and not shared with other conversational agents. In some examples, one of the conversational agents may prompt the user to determine what information may be shared between the conversational agents. The privacy settings 230, or other settings, may be updated using the response from the user.

[0039] The device agent 222 may initiate communications between the user device 226, e.g., the device agent 222, and the conversational agent 208. For instance, the device agent 222 may receive a request from a user of the user device 226 asking for a restaurant recommendation and identify John's Pizza Joint in response, or receive a request specifically asking to order a pizza from or view the menu of John's Pizza Joint. The device agent 222 connects the user device 226 with the conversational agent 208 for John's Pizza Joint to facilitate communication between the user and the conversational agent 208. For example, the communication may be between the user device 226 and the conversation server 202, or between the user device 226 and John's Pizza Joint's server 214.

[0040] Separate software, e.g., on the conversation server 202, may be used for the device agent 222 and the conversational agent 208 or the same software may be used for both agents. For instance, the server may have a single application, e.g., which is part of the conversational agent platform 203, that accesses different data, e.g., settings, for each of the agents and that provides an interface for communication with the user device 226 and the conversational agents. The single application may access the conversation data flow 210 to determine an appropriate response to a question or statement from the user and the agent settings 212 to determine how that response should be presented. The

single application then generates data representing the response and provides that data to the user device 226 which uses the data to provide an audible response to the user.

[0041] In some examples, the device agent 222 may provide representations of input, e.g., text representations of utterances, to the conversational agent 208 to facilitate communication between the user and the conversational agent. The device agent 222 receives responses from the conversational agent 208, e.g., either text or audible responses, and provides the responses to the user. When the device agent 222 receives text responses from the conversational agent 208, the device agent 222 may use settings for the conversational agent 208 to generate and provide audible responses to the user.

[0042] In some implementations, an application for the third party, e.g., John's Pizza Joint, does not have to be installed on the user device 226 for the user to interact with a service provided by the third party. For example, the device agent 222 may determine that there is an agent for the third party service, e.g., the conversational agent 208. The device agent 222 may then communicate with the conversational agent 208, or the user device 226 may communicate with the conversational agent 208, to provide an audible interface to the user.

[0043] In these implementations data for the conversational agent 208 may be stored on John's Pizza Joint's server 214, the conversation server 202, or a combination of the two. For instance, the conversational agent 208 may execute on the conversation server 202 or data for the conversational agent 208 may reside on the conversation server 202, e.g., and be used by another application to generate response that will be presented to a user. The conversational agent 208 on the conversation server 202 may include information about data stored on John's Pizza Joint's server 214, such as how to access the user accounts 218, the order history data 220, or both. The executing conversational agent 208, or an application using data representing the conversational agent 208, may use this information stored on John's Pizza Joint's server 214 when communicating with the user device 226 and a respective user.

[0044] The user device 226 may include personal computers, mobile communication devices, and other devices that can send and receive data over a network 234. The network 234, such as a local area network (LAN), wide area network (WAN), the Internet, or a combination thereof, connects the conversation server 202, John's Pizza Joint's server 214, and the user device 226.

Example Process Flow

[0045] FIG. 3 is a flow diagram of a process 300 for establishing communication between a user device and a computer-implemented conversational agent. For example, the process 300 can be used by the conversation server 202 or the user device 226 from the system 200.

[0046] The process receives data representing properties of another computer-implemented agent (302). For example, a server receives a conversation data flow and other settings for the other agent from an administrator. In some examples, the server receives the settings for the other agent from another computer.

[0047] The process receives, by a computer-implemented agent specific to a user device, a digital representation of speech encoding an utterance (304). For instance, the agent receives a request from a user for a restaurant recommendation, asks the user if they would like to make a reservation at John's Pizza Joint, and receives a digital representation of the utterance "yes."

[0048] The process determines, by the computer-implemented agent, that the utterance specifies a requirement to establish a communication with the other computer-implemented agent (306). The agent determines that the other agent, specific to John's Pizza Joint or a restaurant reservation system, can assist the user in making a reservation at John's Pizza Joint.

[0049] The process establishes, by the computer-implemented agent, a communication between the other computer-implemented agent and the user device (308). For example, the agent hands off communication with the user to the other agent and creates a connection between the other agent and the user device. When the other agent executes on a server, the connection may be between the user device and the server. When the agent executes on a server and the other agent executes on another server, the connection may include multiple points, e.g., the user device, the server, and the other server. In some examples, both the agent and the other agent may execute on the same server and the agent establishes a connection between the user device and the server, e.g., the application that represents the other agent.

[0050] The process provides a representation of the utterance to the other computer-implemented agent (310). For example, the agent or the user device may provide the other agent with the digital representation of speech or a text representation of the utterance. The representation may include the specific utterance or may include

content, e.g., words, related to the utterance. For instance, in this example the representation may indicate that the user is looking to make a reservation at John's Pizza Joint and may include a day, e.g., today, and/or a time, e.g., 6 PM, for the reservation, e.g., when the representation is provided to another agent for a restaurant reservation system.

[0051] The process receives, by the computer-implemented agent, another representation of a response to the utterance (312). For instance, the other agent determines a conversation data flow that corresponds with the representation of the utterance, identifies an action to perform, if any, from the conversation data flow, e.g., to determine whether or not the user can make a reservation at John's Pizza Joint, and generates an appropriate response to the utterance using the conversation data flow and data from the action. The other agent may determine that the user cannot make a reservation at 6 PM, e.g., because there is no availability at John's Pizza Joint at that time, but can make a reservation for 7 PM and provide a text representation of "we don't have any openings at 6 PM tonight, would 7 PM work for you?" to a server, e.g., executing the agent, or the user device.

[0052] In some implementations, an application uses a conversational data flow for the other agent to determine the response to the utterance. For example, the application uses data about the utterance, e.g., that the user is looking to make a reservation, to select the conversational data flow from multiple conversational data flows, e.g., a reservation conversational data flow, and generates a response to the user's utterance.

[0053] The process generates, using the other representation, data to cause a presentation of the response (314). For example, when the server receives the text representation, the server may generate an audio signal of the statement "we don't have any openings at 6 PM tonight, would 7 PM work for you?" using the settings for the other agent. In some examples, the user device receives the text representation, or an audible representation, from the other agent and audibly presents the response to the user.

[0054] One or more steps of the process 300 may repeat as the user requests additional information from the other agent and the other agent responds to the user's requests. For instance, the process may repeat steps 304, and 310-314 until the other agent completes a transaction with the user, e.g., finishes providing the user with information required by the user, or until the other agent determines that the conversation

with the user should be passed to a different agent, e.g., and the other agent repeats steps 306-308 to hand off the conversation to the different agent.

[0055] The order of steps in the process 300 described above is illustrative only, and establishing communication between the user device and the computer-implemented conversational agent can be performed in different orders. For example, the process 300 may establish communication between the other agent and the user device and receive a representation of a response from the other agent and then receive a digital representation of speech encoding another utterance and provide a representation of the other utterance to the other agent.

[0056] In some implementations, the process 300 can include additional steps, fewer steps, or some of the steps can be divided into multiple steps. For example, the process 300 may include steps 304-308 and might not include one or more of the steps 302 and 310-314.

Optional Implementation Details

[0057] In some implementations, the other agent accesses one or more systems of a third party to provide responses or other information to a user. For instance, the other agent may receive a reservation request from a user or from an agent associated with the user's device, access a database, stored on the same server executing the other agent or a different server, to determine whether the user may make the requested reservation or if a different time is available, and generates a response based on the determination.

[0058] In some implementations, different agents, e.g., associated with the same entity or third party, may be used for different conversation data flows or different types of interactions with a user. For instance, a hotel may have a reservation agent, a concierge agent, and a room service agent. When a user requests information from the hotel, an agent on his device may determine a type of interaction for which the user is looking and the hotel agent that is most likely to assist the user with that interaction. In one example, when the device agent receives a request stating that "I'm thinking of staying at Hotel Fancy, but I want to know what's around Hotel Fancy. Let me talk to an agent at Hotel Fancy that can tell me what restaurants and activities are nearby." The device agent may then use keywords in the request to determine that the concierge agent is the agent most likely to have the information the user requested and create communication between the user device and the concierge agent to enable the concierge agent to communicate with the user.

[0059] In some implementations, a third party may have multiple agents for the same type of conversation data flow but different types of users. For example, a first agent may be used to interact with users that have a first type of demographic and a second agent, e.g., with a different voice or style of speech, may be used to interact with users that have a second type of demographic, different from the first type of demographic.

Additional Implementation Details

[0060] In situations in which the systems or agents discussed here collect personal information about users, or may make use of personal information, the users may be provided with an opportunity to control whether programs or features collect user information (e.g., information about a user's social network, social actions or activities, profession, a user's preferences, or a user's current location), or to control whether and/or how to receive content from the agents that may be more relevant to the user. In addition, certain data may be treated in one or more ways before it is stored or used, so that personally identifiable information is removed. For example, a user's identity may be treated so that no personally identifiable information can be determined for the user, or a user's geographic location may be generalized where location information is obtained (such as to a city, ZIP code, or state level), so that a particular location of a user cannot be determined. Thus, the user may have control over how information is collected about the user and used by an agent.

[0061] Embodiments of the subject matter and the functional operations described in this specification can be implemented in digital electronic circuitry, in tangibly-embodied computer software or firmware, in computer hardware, including the structures disclosed in this specification and their structural equivalents, or in combinations of one or more of them. Embodiments of the subject matter described in this specification can be implemented as one or more computer programs, i.e., one or more modules of computer program instructions encoded on a tangible non-transitory program carrier for execution by, or to control the operation of, data processing apparatus. Alternatively or in addition, the program instructions can be encoded on an artificially-generated propagated signal, e.g., a machine-generated electrical, optical, or electromagnetic signal, that is generated to encode information for transmission to suitable receiver apparatus for execution by a data processing apparatus. The computer storage medium can be a

machine-readable storage device, a machine-readable storage substrate, a random or serial access memory device, or a combination of one or more of them.

[0062] The term “data processing apparatus” refers to data processing hardware and encompasses all kinds of apparatus, devices, and machines for processing data, including by way of example a programmable processor, a computer, or multiple processors or computers. The apparatus can also be or further include special purpose logic circuitry, e.g., an FPGA (field programmable gate array) or an ASIC (application-specific integrated circuit). The apparatus can optionally include, in addition to hardware, code that creates an execution environment for computer programs, e.g., code that constitutes processor firmware, a protocol stack, a database management system, an operating system, or a combination of one or more of them.

[0063] A computer program, which may also be referred to or described as a program, software, a software application, a module, a software module, a script, or code, can be written in any form of programming language, including compiled or interpreted languages, or declarative or procedural languages, and it can be deployed in any form, including as a stand-alone program or as a module, component, subroutine, or other unit suitable for use in a computing environment. A computer program may, but need not, correspond to a file in a file system. A program can be stored in a portion of a file that holds other programs or data, e.g., one or more scripts stored in a markup language document, in a single file dedicated to the program in question, or in multiple coordinated files, e.g., files that store one or more modules, sub-programs, or portions of code. A computer program can be deployed to be executed on one computer or on multiple computers that are located at one site or distributed across multiple sites and interconnected by a communication network.

[0064] The processes and logic flows described in this specification can be performed by one or more programmable computers executing one or more computer programs to perform functions by operating on input data and generating output. The processes and logic flows can also be performed by, and apparatus can also be implemented as, special purpose logic circuitry, e.g., an FPGA (field programmable gate array) or an ASIC (application-specific integrated circuit).

[0065] Computers suitable for the execution of a computer program include, by way of example, general or special purpose microprocessors or both, or any other kind of central processing unit. Generally, a central processing unit will receive instructions and data from a read-only memory or a random access memory or both. The essential

elements of a computer are a central processing unit for performing or executing instructions and one or more memory devices for storing instructions and data.

Generally, a computer will also include, or be operatively coupled to receive data from or transfer data to, or both, one or more mass storage devices for storing data, e.g., magnetic, magneto-optical disks, or optical disks. However, a computer need not have such devices. Moreover, a computer can be embedded in another device, e.g., a mobile telephone, a personal digital assistant (PDA), a mobile audio or video player, a game console, a Global Positioning System (GPS) receiver, or a portable storage device, e.g., a universal serial bus (USB) flash drive, to name just a few.

[0066] Computer-readable media suitable for storing computer program instructions and data include all forms of non-volatile memory, media and memory devices, including by way of example semiconductor memory devices, e.g., EPROM, EEPROM, and flash memory devices; magnetic disks, e.g., internal hard disks or removable disks; magneto-optical disks; and CD-ROM and DVD-ROM disks. The processor and the memory can be supplemented by, or incorporated in, special purpose logic circuitry.

[0067] To provide for interaction with a user, embodiments of the subject matter described in this specification can be implemented on a computer having a display device, e.g., a CRT (cathode ray tube) or LCD (liquid crystal display) monitor, for displaying information to the user and a keyboard and a pointing device, e.g., a mouse or a trackball, by which the user can provide input to the computer. Other kinds of devices can be used to provide for interaction with a user as well; for example, feedback provided to the user can be any form of sensory feedback, e.g., visual feedback, auditory feedback, or tactile feedback; and input from the user can be received in any form, including acoustic, speech, or tactile input. In addition, a computer can interact with a user by sending documents to and receiving documents from a device that is used by the user; for example, by sending web pages to a web browser on a user's device in response to requests received from the web browser.

[0068] Embodiments of the subject matter described in this specification can be implemented in a computing system that includes a back-end component, e.g., as a data server, or that includes a middleware component, e.g., an application server, or that includes a front-end component, e.g., a client computer having a graphical user interface or a Web browser through which a user can interact with an implementation of the subject matter described in this specification, or any combination of one or more such back-end,

middleware, or front-end components. The components of the system can be interconnected by any form or medium of digital data communication, e.g., a communication network. Examples of communication networks include a local area network (LAN) and a wide area network (WAN), e.g., the Internet.

[0069] The computing system can include clients and servers. A client and server are generally remote from each other and typically interact through a communication network. The relationship of client and server arises by virtue of computer programs running on the respective computers and having a client-server relationship to each other. In some embodiments, a server transmits data, e.g., an HTML page, to a user device, e.g., for purposes of displaying data to and receiving user input from a user interacting with the user device, which acts as a client. Data generated at the user device, e.g., a result of the user interaction, can be received from the user device at the server.

[0070] An example of one such type of computer is shown in FIG. 4, which shows a schematic diagram of a generic computer system 400. The system 400 can be used for the operations described in association with any of the computer-implement methods described previously, according to one implementation. The system 400 includes a processor 410, a memory 420, a storage device 430, and an input/output device 440. Each of the components 410, 420, 430, and 440 are interconnected using a system bus 450. The processor 410 is capable of processing instructions for execution within the system 400. In one implementation, the processor 410 is a single-threaded processor. In another implementation, the processor 410 is a multi-threaded processor. The processor 410 is capable of processing instructions stored in the memory 420 or on the storage device 430 to display graphical information for a user interface on the input/output device 440.

[0071] The memory 420 stores information within the system 400. In one implementation, the memory 420 is a computer-readable medium. In one implementation, the memory 420 is a volatile memory unit. In another implementation, the memory 420 is a non-volatile memory unit.

[0072] The storage device 430 is capable of providing mass storage for the system 400. In one implementation, the storage device 430 is a computer-readable medium. In various different implementations, the storage device 430 may be a floppy disk device, a hard disk device, an optical disk device, or a tape device.

[0073] The input/output device 440 provides input/output operations for the system 400. In one implementation, the input/output device 440 includes a keyboard

and/or pointing device. In another implementation, the input/output device 440 includes a display unit for displaying graphical user interfaces.

[0074] While this specification contains many specific implementation details, these should not be construed as limitations on the scope of what may be claimed, but rather as descriptions of features that may be specific to particular embodiments. Certain features that are described in this specification in the context of separate embodiments can also be implemented in combination in a single embodiment. Conversely, various features that are described in the context of a single embodiment can also be implemented in multiple embodiments separately or in any suitable subcombination. Moreover, although features may be described above as acting in certain combinations and even initially claimed as such, one or more features from a claimed combination can in some cases be excised from the combination, and the claimed combination may be directed to a subcombination or variation of a subcombination.

[0075] Similarly, while operations are depicted in the drawings in a particular order, this should not be understood as requiring that such operations be performed in the particular order shown or in sequential order, or that all illustrated operations be performed, to achieve desirable results. In certain circumstances, multitasking and parallel processing may be advantageous. Moreover, the separation of various system modules and components in the embodiments described above should not be understood as requiring such separation in all embodiments, and it should be understood that the described program components and systems can generally be integrated together in a single software product or packaged into multiple software products.

[0076] Particular embodiments of the subject matter have been described. Other embodiments are within the scope of the following claims. For example, the actions recited in the claims can be performed in a different order and still achieve desirable results. As one example, the processes depicted in the accompanying figures do not necessarily require the particular order shown, or sequential order, to achieve desirable results. In some cases, multitasking and parallel processing may be advantageous.

CLAIMS

1. A computer-implemented method comprising:
 - receiving, by a computer-implemented agent specific to a user device, a digital representation of speech encoding an utterance;
 - 5 determining, by the computer-implemented agent, that the utterance specifies a requirement to establish a communication with another computer-implemented agent; and
 - establishing, by the computer-implemented agent, a communication between the other computer-implemented agent and the user device.
- 10 2. The method of claim 1 wherein the other computer-implemented agent is not specific to the user device.
3. The method of claim 1 wherein establishing the communication with the other computer-implemented agent and the user device comprises establishing the communication with the other computer-implemented agent executing on another device and the user device.
4. The method of claim 1 wherein establishing the communication with the other computer-implemented agent and the user device comprises establishing the communication with the other computer-implemented agent and the computer-implemented agent.
- 15 5. The method of claim 1 wherein the computer-implemented agent is for a first application on a user device and the other computer-implemented agent is for a second, different application on the user device.
6. The method of claim 1 comprising:
 - 20 providing a representation of the utterance to the other computer-implemented agent.
7. The method of claim 6 wherein the representation comprises a text representation of the utterance.

8. The method of claim 7 comprising:

receiving, by the computer-implemented agent from the other computer-implemented agent, another representation of a response to the utterance; and
generating, using the other representation, data to cause a presentation of the
5 response.

9. The method of claim 8 comprising:

determining, by the computer-implemented agent, settings, associated with the
other computer-implemented agent, for the presentation of the response.

10. The method of claim 6 wherein the representation comprises the digital representation
10 of speech.

11. The method of claim 6 wherein providing the representation of the utterance to the
other computer-implemented agent comprises providing data for a user who spoke the
utterance to the other computer-implemented agent.

12. The method of claim 11 wherein providing data for the user who spoke the utterance
to the other computer-implemented agent comprises:

determining privacy settings for the user; and
providing the data for the user to the other computer-implemented agent according
to the privacy settings.

13. The method of claim 1 comprising:

receiving data representing properties of the other computer-implemented agent
prior to receiving the digital representation of speech encoding the utterance.

14. The method of claim 13 wherein receiving the data representing the properties of the
other computer-implemented agent comprises receiving a definition of a data flow for the
other computer-implemented agent that identifies a response for the other computer-
implemented agent given a particular type of input.

15. The method of claim 13 wherein receiving the data representing the properties of the
other computer-implemented agent comprises receiving an identification of a particular
style of speech for the other computer-implemented agent.

16. The method of claim 13 wherein receiving the data representing the properties of the other computer-implemented agent comprises receiving text-to-speech parameters for responses provided by the other computer-implemented agent.

17. The method of claim 13 wherein receiving the data representing the properties of the other computer-implemented agent comprises receiving a selection of a template that identifies a data flow for computer-implemented agents that identifies a type of response for the other computer-implemented agent given a particular type of input.

18. The method of claim 17 wherein the template is for computer-implemented agents that provide online ordering services.

19. A system, comprising:

- a data processing apparatus; and

- a non-transitory computer readable storage medium in data communication with the data processing apparatus and storing instructions executable by the data processing apparatus and upon such execution cause the data processing to perform operations comprising:

- receiving, by a computer-implemented agent specific to a user device, a digital representation of speech encoding an utterance;

- determining, by the computer-implemented agent, that the utterance specifies a requirement to establish a communication with another computer-implemented agent; and

- establishing, by the computer-implemented agent, a communication between the other computer-implemented agent and the user device.

20. A non-transitory computer readable storage medium storing instructions executable by a data processing apparatus and upon such execution cause the data processing to perform operations comprising:

- receiving, by a computer-implemented agent specific to a user device, a digital representation of speech encoding an utterance;

- determining, by the computer-implemented agent, that the utterance specifies a requirement to establish a communication with another computer-implemented agent; and

- establishing, by the computer-implemented agent, a communication between the other computer-implemented agent and the user device.

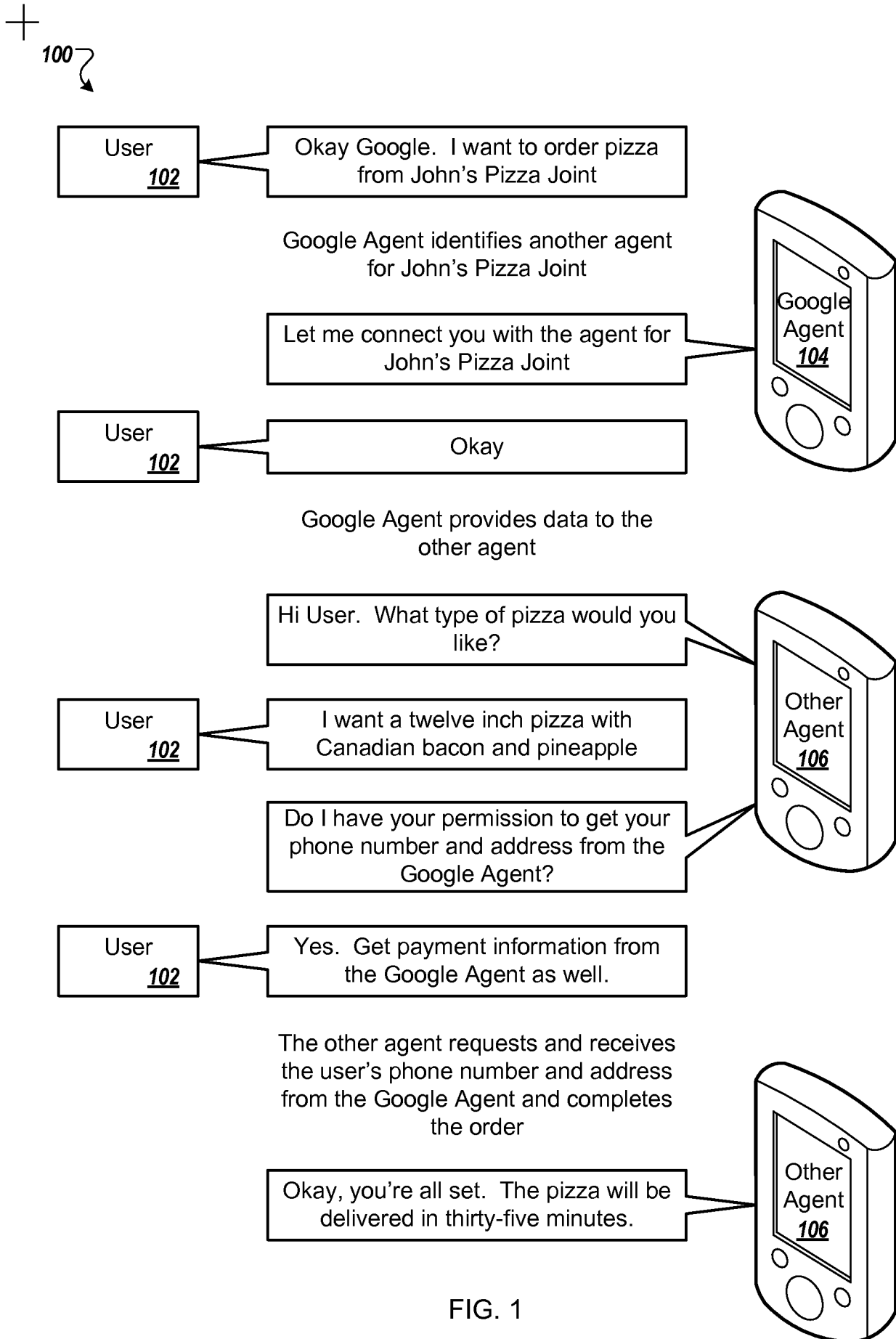


FIG. 1

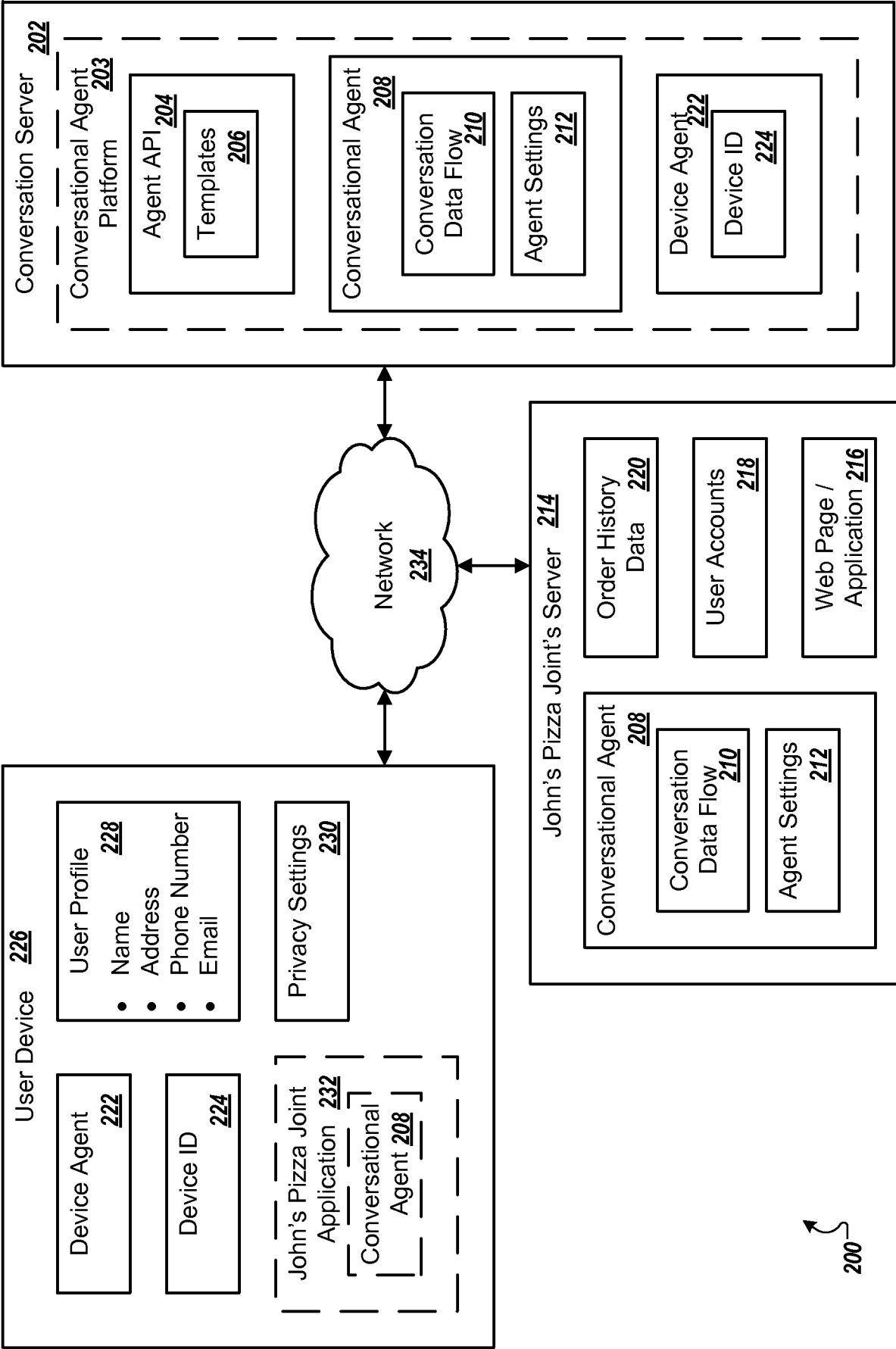


FIG. 2

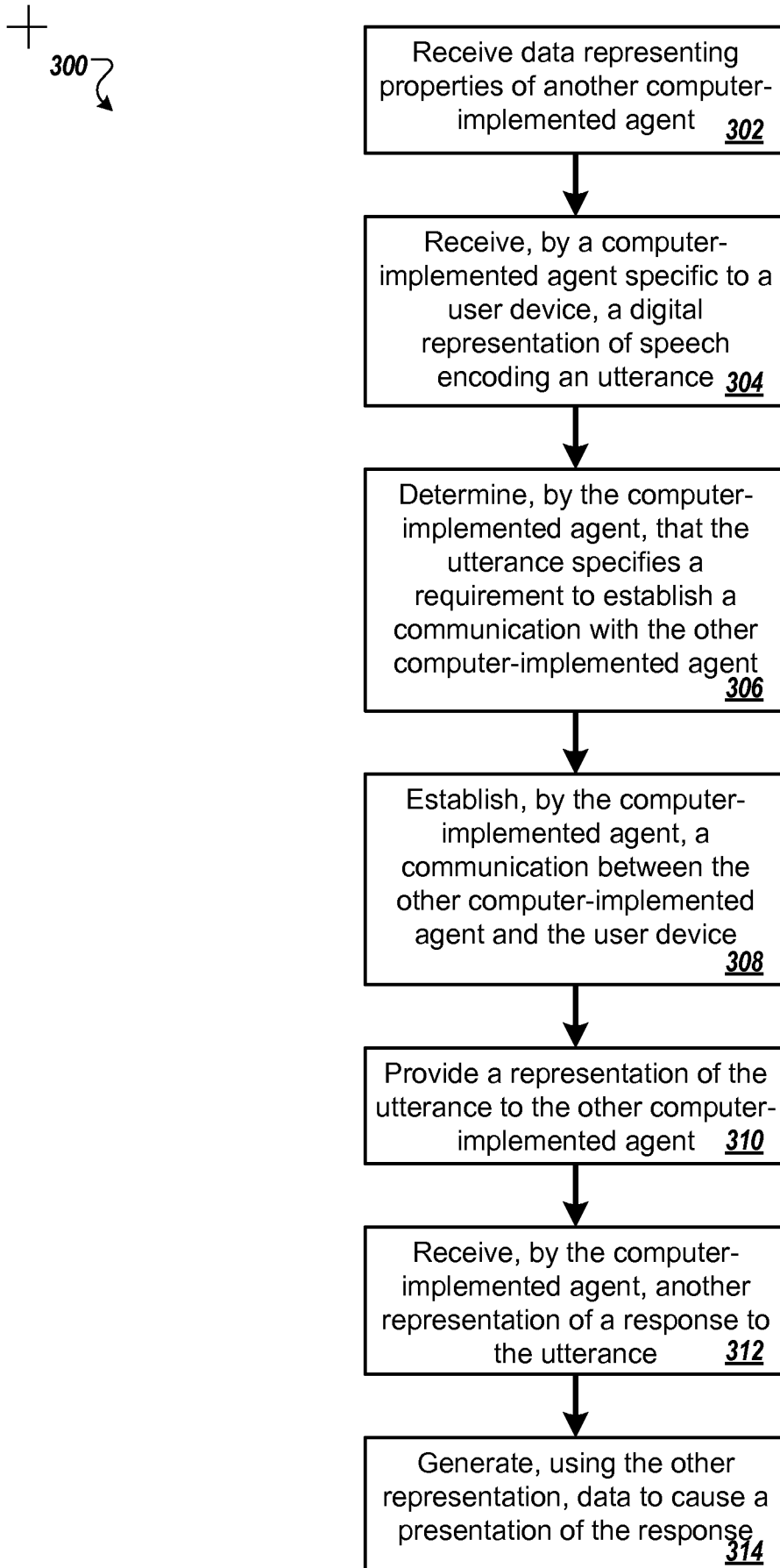


FIG. 3

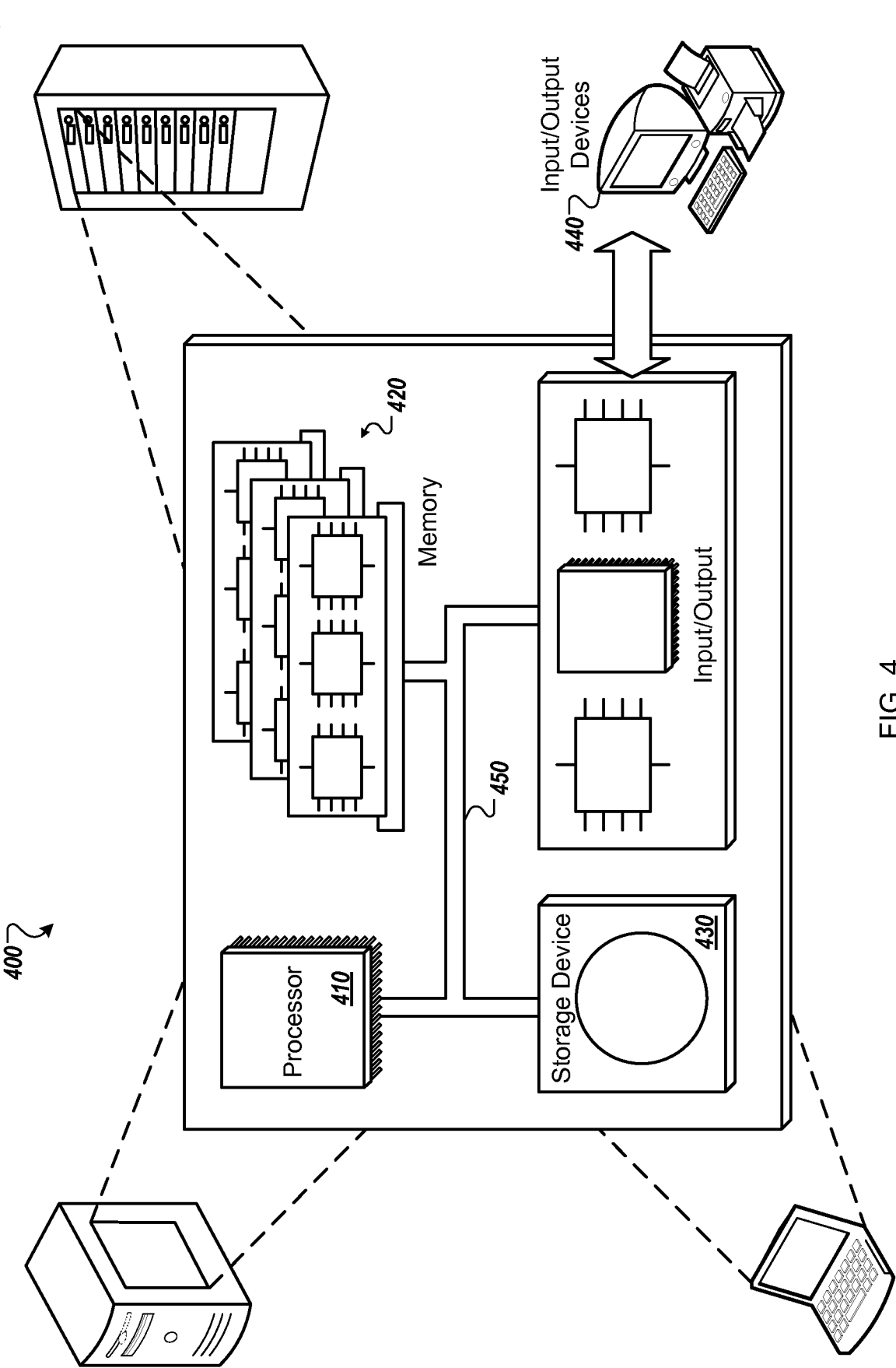


FIG. 4

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2015/042048

A. CLASSIFICATION OF SUBJECT MATTER
INV. G10L15/18 G10L15/22 G06F17/30
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G10L G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 2 273 491 A1 (VOICEBOX TECHNOLOGIES INC [US]) 12 January 2011 (2011-01-12) page 9, paragraph 38 - paragraph 40 page 10, paragraph 43 page 13, paragraph 62 - page 14 page 6, paragraph 21 page 5, paragraph 18 page 11, paragraph 50 ----- -/--	1-20



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

9 October 2015

Date of mailing of the international search report

26/10/2015

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Ziegler, Stefan

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2015/042048

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2004/193420 A1 (KENNEWICK ROBERT A [US] ET AL) 30 September 2004 (2004-09-30) page 2, paragraph 14 page 4, paragraph 28 page 16, paragraph 160 page 2, paragraph 17 - paragraph 18 page 3, paragraph 22 page 12, paragraph 126 page 2, paragraph 23 - paragraph 27 page 13, paragraph 134 page 16, paragraph 164 page 18, paragraph 178 -----	1-20
X	WO 2013/155619 A1 (PASUPALAK SAM [CA]; PANTONY JOSHUA R [CA]; HSU WILSON [CA]; WU ZHIYUAN) 24 October 2013 (2013-10-24) page 7, paragraph 42 page 11, paragraph 59 - page 12 page 14, paragraph 69 page 27, paragraph 120 page 28, paragraph 122 page 127, paragraph 120 - page 128, paragraph 122 page 8, paragraph 47 page 13, paragraph 64 - paragraph 65 page 52, paragraph 247 page 50, paragraph 241 - page 51 page 31, paragraph 142 -----	1-20
A	WO 2005/038775 A1 (METAPHOR SOLUTIONS INC [US]; KUPERSTEIN MICHAEL [US]) 28 April 2005 (2005-04-28) figure 1 page 8, line 24 - line 30 page 42 - page 44 -----	1-20

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2015/042048

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 2273491	A1	12-01-2011	CN 101939740 A 05-01-2011
			EP 2232388 A1 29-09-2010
			EP 2273491 A1 12-01-2011
			US 2009150156 A1 11-06-2009
			US 2012101809 A1 26-04-2012
			US 2012101810 A1 26-04-2012
			US 2012109753 A1 03-05-2012
			US 2013211710 A1 15-08-2013
			US 2014156278 A1 05-06-2014
			US 2014288934 A1 25-09-2014
			US 2015073910 A1 12-03-2015
			US 2015095159 A1 02-04-2015
			WO 2009075912 A1 18-06-2009
US 2004193420	A1	30-09-2004	US 2004193420 A1 30-09-2004
			US 2010145700 A1 10-06-2010
WO 2013155619	A1	24-10-2013	EP 2839391 A1 25-02-2015
			US 2015066479 A1 05-03-2015
			WO 2013155619 A1 24-10-2013
WO 2005038775	A1	28-04-2005	NONE