INTERRUPTING VIRTUAL ASSISTANTS TO INITIATE NEW TASKS

Google Inc.
INTERRUPTING VIRTUAL ASSISTANTS TO INITIATE NEW TASKS

ABSTRACT

An interactive assistant, referred to herein as “an interactive assistant,” “a virtual assistant,” a “computational assistant,” or simply an “assistant,” is described that allows a user to interrupt presentation of information concerning a first task and initiate a second task without having to wait until the virtual assistant presents all of the information concerning a first task or otherwise issue a “stop” or other cease command with regard to the first task. The user may, in other words, stop presentation of the information concerning the first task by the virtual assistant and initiate the second task in a manner that better adheres to conversational norms.

DESCRIPTION

Virtual, intelligent, or computational assistants (e.g., also referred to simply “assistants”) execute on counter-top devices, mobile phones, automobiles, and many other type of computing devices. Assistants output useful information, responds to users’ needs, or otherwise performs certain operations to help users complete real-world and/or virtual tasks. Some assistants are configured to record reminders, notes, lists, and other information as persistent records that a user can later retrieve or modify during a subsequent interaction with the assistant.

Some assistants may further respond to certain tasks with information relevant to a task initiated by a query issued by the user. The amount of information returned in response to the query may vary with some queries resulting in large amounts of information. The assistant may, in some examples, audibly (or, via text on a display) present the large amounts of information, some of which may be irrelevant to the user’s current context or needs. The assistant may be unable to process further tasks or queries until either all of the information has been presented or...
until the user issues a particular command, such as “stop” or “cease,” either verbally or via keyboard or some other physical or digital button (e.g., such as pressing the power button briefly).

A virtual assistant is described that allows a user to interrupt presentation of information concerning a first task and initiate a second task without having to wait until the virtual assistant presents all of the information concerning a first task or otherwise issue a “stop” or other cease command with regard to the first task. The user may, in other words, stop presentation of the information concerning the first task by the virtual assistant and initiate the second task in a manner that better adheres to conversational norms.

In operation, a device may provide an interface with which to interact with the virtual assistant. The device may include a cellular phone (such as a so-called “smart phone”), a tablet computer, a laptop computer, a desktop computer, a workstation, a home audio device (such as a so-called “smart speaker”), a gaming console, a portable gaming console, an audio/visual (AV) receiver, a digital disc player, and the like.

The device may include, in some examples, a display by which to present the interface for interacting with the digital assistant. The device may further include one or more sensors by which to receiver data input by a user of the device, the data specifying interactions with the virtual assistant. The sensors may include capacitive touch sensors integrated with the display to capture text interactions, one or more transducers (e.g., a microphone) for capturing audio interactions, cameras for capturing image and/or video interactions, a global positioning system (GPS) sensor for capturing location-based interactions, a compass for capturing positional-based interactions, etc.
In some examples, a server or dedicated device (which may be separate from the device providing the interface by which to interact with the digital assistant - which itself may be referred to as a “client device”) may provide some or all of the functionality of the digital assistant. The server may receive the data indicative of the interaction (e.g., any combination of the above text, audio, image, video, location-based, or position-based interactions) from the client device.

As a user interacts with the assistant, the assistant may obtain personal information about the user. Examples of personal information include: habits, routines, preferences, notes, lists, contacts, communications, interests, location histories, and other types of user information. After receiving explicit permission from the user, the assistant may store, the personal information at user information data stores and, in the course of providing assistant services, make use of the personal information stored at the user information data stores.

A user may be provided with controls allowing the user to make an election as to both if and when the assistant, the computing device, or the computing systems described herein can collect or make use of supplemental data (e.g., user information or contextual information about a user’s social network, social actions or activities, profession, a user’s preferences, or a user’s current location), and if and when the user is sent content or communications from a server. 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 what
supplemental data is collected about the user, how that supplemental data is used, and what supplemental data is provided to the user.

Furthermore, the external systems, the computing devices, and the assistant treat the information stored at the information stores so that the information is protected, encrypted, or otherwise not susceptible to unauthorized use. The information stored at the information data stores may be stored locally at the computing devices and/or remotely (e.g., in a cloud computing environment provided by the external systems and which is accessible via a network).

In any event, the server may include one or more processors configured to process the data indicative of the interactions in accordance with the digital assistant to generate a result. The server may provide the result to the client device responsive to the data indicative of the interactions.

The digital assistant may represent a software agent trained through, at least in some examples, machine learning to process data indicative of natural language text and audio, and possibly images and video to perform tasks or services for individuals. The digital assistant may, to provide a few examples, answer queries, provide directions to destinations, schedule meetings and/or appointments, generate electronic mail (e-mail) and/or text messages, initiate a telephone call, and the like.

In some instances, the digital assistant may present large amounts of information concerning a first task initiated by the user. To illustrate, consider that a user may enter an interaction requesting that the digital assistant determine the operating hours of a business or other enterprise, park, or venue. The digital assistant may retrieve a large amount of information, indicating that a park, for example, is open on Mondays from 10 AM to 6 PM, Tuesdays from 8 AM to 5 PM, Wednesdays from 9 AM to 6 PM, Fridays from 12 PM to 10 PM,
Saturdays from 6 AM to 6 PM and Sundays from 11 AM to 5 PM, while being closed on Thursdays. Much of this information may be irrelevant to the user given that the user is considering going to the park today (which may for purposes of example assumed to be a Tuesday).

Rather than wait for the digital assistant to finish this lengthy list of information concerning the operating hours of the park, the user may interrupt the presentation of the information and initiate a second task. In the above example, the user may initiate a second task to the digital assistant involving the determination of a quickest bus route to get from the here to the park. The user may initiate the second task by interrupting the presentation of the operating hours of the park and without using any special command to halt or otherwise cease presentation of the information concerning the first task related to the operating hours of the park.
As shown in the above flowchart, the client device may receive an interaction from the user (e.g., in the form of a query) concerning a first (1st) task. The client device may transmit the query to the server. The server may process the query. When the result is determined, the server may return the result to the client as information concerning the first task.

The client device presents the result to the user. When audibly presented, the client
device may attempt to determine when the user is interrupting the audible presentation in an attempt to better adhere to social norms, and thereby better emulate human assistants. When no interruption is detected, the client device may continue to present the result to the user. When an interruption is detected (e.g., possibly via touch or proximity gestures or other touch interactions, audio interactions – where the user may start speaking the next command while the device is still outputting results of the previous command, physical button interactions – including keyboard interactions, etc., including combinations of interactions), the client device may halt the presentation of result to the user, and process the interruption, allowing the user to specify an interaction (e.g., a query) concerning a second (2nd) task. In some examples, the interruption does not otherwise contain a statement or clue as to the need to stop the audible presentation, other than it is a statement, command, request or the like that contextually suggests stopping and, for example, starting on the different statement, command, request or the like. That is, no “stop”, “pause” or the like need be issued by the user. The client device may transmit the indication concerning the second task to the server, which may process the indication concerning the second task and return a result to the client device.

The client device may then present the result to the user (where the result in the example above would involve information concerning the quickest bus route to the park, initiating a navigation application showing the identified quickest bus route, etc.). That is, the presentation of the result may not necessarily involve actual presentation of some item to the user, but that the server or the client device may perform some operation related to the query, such as opening the navigation application.

Furthermore, although described as providing the indication to the server, the client device may locally process the indication of the intended option and provide the result to the
user. That is, the client device need not interface with the server hosting the digital assistant to return the result to the client when clarification is required. Furthermore, in some examples, the client device may host the digital assistant itself, thereby avoiding network expense, power consumption and the like associated with interfacing with the remotely hosted virtual assistant.