



(51) International Patent Classification:

G06F 3/16 (2006.01) G10L 15/00 (2013.01)
G06F 17/28 (2006.01) G10L 15/22 (2006.01)

(21) International Application Number:

PCT/US2016/021410

(22) International Filing Date:

8 March 2016 (08.03.2016)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/129,932 8 March 2015 (08.03.2015) US
14/841,449 31 August 2015 (31.08.2015) US

(71) Applicant: APPLE INC. [US/US]; 1 Infinite Loop, Cupertino, CA 95014 (US).

(72) Inventors: LEMAY, Stephen, O.; c/o Apple Inc., 1 Infinite Loop, Cupertino, CA 95014 (US). NEWENDORP, Brandon, J.; c/o Apple Inc., 1 Infinite Loop, Cupertino, CA 95014 (US). DASCOLA, Jonathan, R.; c/o Apple Inc., 1 Infinite Loop, Cupertino, CA 95014 (US).

(74) Agents: EIDE, Christopher, B. et al.; Morrison & Foerster LLP, 755 Page Mill Road, Palo Alto, CA 94304-1018 (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, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))

(54) Title: VIRTUAL ASSISTANT ACTIVATION

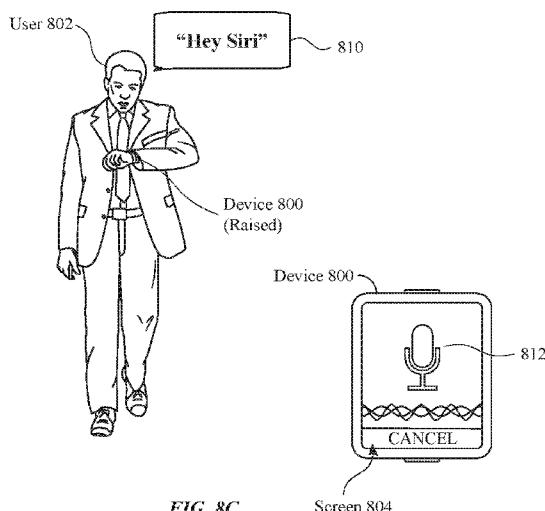


FIG. 8C

(57) Abstract: At an electronic device with a display, a microphone, and an input device: while the display is on, receiving user input via the input device, the user input meeting a predetermined condition; in accordance with receiving the user input meeting the predetermined condition, sampling audio input received via the microphone; determining whether the audio input comprises a spoken trigger; and in accordance with a determination that audio input comprises the spoken trigger, triggering a virtual assistant session.

VIRTUAL ASSISTANT ACTIVATION

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of priority of U.S. Provisional Patent Application Serial No. 62/129,932, titled “VIRTUAL ASSISTANT ACTIVATION,” filed March 8, 2015; and U.S. Non-provisional Patent Application Serial No. 14/841,449, titled “VIRTUAL ASSISTANT ACTIVATION,” filed August 31, 2015. The content of these applications is hereby incorporated by reference in its entirety.

[0002] This application is related to U.S. Provisional Patent Application Serial No. 62/026,532, titled “Raise Gesture Detection in a Device,” filed July 18, 2014; and U.S. Patent Application Serial No. 12/987,982, titled “Intelligent Automated Assistant,” filed January 10, 2011. The content of these applications is hereby incorporated by reference.

BACKGROUND

1. Field

[0003] The present disclosure relates generally to intelligent automated assistants and, more specifically, to techniques for triggering intelligent automated assistants.

2. Description of Related Art

[0004] Intelligent automated assistants (or virtual assistants) provide beneficial interfaces between human users and electronic devices. Exemplary virtual assistants allow users to interact with devices or systems using natural language in spoken and/or text forms. A user can provide a spoken input containing a user request to a virtual assistant operating on an electronic device. The virtual assistant interprets the user’s intent from the spoken input and operationalizes the user’s intent into one or more tasks. The tasks can then be performed by executing one or more services of the electronic device, thereby returning relevant output responsive to the user request.

BRIEF SUMMARY

[0005] As virtual assistant become increasingly sophisticated, it can be desirable to call upon their assistance under increasing numbers of usage scenarios. Efficient techniques for triggering a virtual assistant are thus helpful. Additionally, in situations where virtual

assistants are executing on battery-powered portable electronic devices, efficient techniques for triggering virtual assistant while conserving battery-power are particularly helpful.

[0006] In accordance with some embodiments, a method of triggering a virtual assistant on an electronic device comprises: at an electronic device with a display, a microphone, and an input device: while the display is on, receiving user input via the input device, the user input meeting a predetermined condition; in accordance with receiving the user input meeting the predetermined condition, sampling audio input received via the microphone; determining whether the audio input comprises a spoken trigger; and in accordance with a determination that audio input comprises the spoken trigger, triggering a virtual assistant session.

[0007] In accordance with some embodiments, a method of triggering a virtual assistant on an electronic device comprises: at an electronic device with a display and a microphone: detecting a software event meeting a predetermined condition; in accordance with a determination that the software event meeting the condition is detected, sampling audio input received via the microphone; determining whether audio input received via the microphone comprises a spoken trigger; and in accordance with a determination that the audio input comprises the spoken trigger, triggering a virtual assistant session.

[0008] In accordance with some embodiments, a method of triggering a virtual assistant on an electronic device comprises: at an electronic device with a display, a microphone, and an input device: receiving user input via the input device, the user input meeting a predetermined condition; in accordance with receiving the user input meeting the predetermined condition, sampling audio input received via the microphone; determining whether the audio input comprises a spoken trigger; in accordance with a determination that audio input comprises the spoken trigger: triggering a virtual assistant session, and providing a haptic output.

[0009] In accordance with some embodiments, a method of triggering a virtual assistant on an electronic device comprises: at an electronic device with a display and a microphone: sampling audio input received via the microphone and determining whether the audio input represents a task; while sampling audio input, receiving instruction to cease sampling of audio input; providing output acknowledging the received instruction to cease the sampling of audio input; and after providing the output, continuing the sampling of audio input and

determining whether the audio input comprises a task, for a duration, then ceasing the sampling of audio input.

[0010] In accordance with some embodiments, an electronic device comprises: a display; one or more processors; a memory; and one or more programs, wherein the one or more programs are stored in memory and configured to be executed by the one or more processors, the one or more programs including instructions for: while the display is on, receiving user input via an input device, the user input meeting a predetermined condition; in accordance with receiving the user input meeting the predetermined condition, sampling audio input received via a microphone; determining whether the audio input comprises a spoken trigger; and in accordance with a determination that audio input comprises the spoken trigger, triggering a virtual assistant session.

[0011] In accordance with some embodiments, an electronic device comprises: a display; one or more processors; a memory; and one or more programs, wherein the one or more programs are stored in memory and configured to be executed by the one or more processors, the one or more programs including instructions for: detecting a software event meeting a predetermined condition; in accordance with a determination that the software event meeting the condition is detected, sampling audio input received via a microphone; determining whether audio input received via the microphone comprises a spoken trigger; and in accordance with a determination that the audio input comprises the spoken trigger, triggering a virtual assistant session.

[0012] In accordance with some embodiments, an electronic device comprises: a display; one or more processors; a memory; and one or more programs, wherein the one or more programs are stored in memory and configured to be executed by the one or more processors, the one or more programs including instructions for: receiving user input via the input device, the user input meeting a predetermined condition;

[0013] In accordance with receiving the user input meeting the predetermined condition, sampling audio input received via a microphone; determining whether the audio input comprises a spoken trigger; in accordance with a determination that audio input comprises the spoken trigger: triggering a virtual assistant session, and providing a haptic output.

[0014] In accordance with some embodiments, an electronic device comprises: a display; one or more processors; a memory; and one or more programs, wherein the one or more

programs are stored in memory and configured to be executed by the one or more processors, the one or more programs including instructions for: sampling audio input received via a microphone and determine whether the audio input represents a task; while sampling audio input, receiving instruction to cease sampling of audio input; providing output acknowledging the received instruction to cease the sampling of audio input; and after providing the output, continuing the sampling of audio input and determining whether the audio input comprises a task, for a duration, then cease the sampling of audio input.

[0015] In accordance with some embodiments, a non-transitory computer readable storage medium stores one or more programs, the one or more programs comprising instructions, which when executed by an electronic device with a display, a microphone, and an input device, cause the electronic device to: while the display is on, receive user input via the input device, the user input meeting a predetermined condition; in accordance with receiving the user input meeting the predetermined condition, sample audio input received via the microphone; determine whether the audio input comprises a spoken trigger; and in accordance with a determination that audio input comprises the spoken trigger, trigger a virtual assistant session.

[0016] In accordance with some embodiments, a non-transitory computer readable storage medium stores one or more programs, the one or more programs comprising instructions, which when executed by an electronic device with a display, a microphone, and an input device, cause the electronic device to: detect a software event meeting a predetermined condition; in accordance with a determination that the software event meeting the condition is detected, sample audio input received via the microphone; determine whether audio input received via the microphone comprises a spoken trigger; and in accordance with a determination that the audio input comprises the spoken trigger, trigger a virtual assistant session.

[0017] In accordance with some embodiments, a non-transitory computer readable storage medium stores one or more programs, the one or more programs comprising instructions, which when executed by an electronic device with a display, a microphone, and an input device, cause the electronic device to: receive user input via the input device, the user input meeting a predetermined condition; in accordance with receiving the user input meeting the predetermined condition, sample audio input received via the microphone; determine whether the audio input comprises a spoken trigger; in accordance with a

determination that audio input comprises the spoken trigger: trigger a virtual assistant session, and provide a haptic output.

[0018] In accordance with some embodiments, a non-transitory computer readable storage medium stores one or more programs, the one or more programs comprising instructions, which when executed by an electronic device with a display, a microphone, and an input device, cause the electronic device to: sample audio input received via the microphone and determine whether the audio input represents a task; while sampling audio input, receive instruction to cease sampling of audio input; provide output acknowledging the received instruction to cease the sampling of audio input; and after providing the output, continue the sampling of audio input and determine whether the audio input comprises a task, for a duration, then cease the sampling of audio input.

[0019] In accordance with some embodiments, an electronic device comprises: while a displaying means is on, means for receiving user input via the input device, the user input meeting a predetermined condition; means for, in accordance with receiving the user input meeting the predetermined condition, sampling audio input received via a microphone; means for determining whether the audio input comprises a spoken trigger; and means for, in accordance with a determination that audio input comprises the spoken trigger, triggering a virtual assistant session.

[0020] In accordance with some embodiments, an electronic device comprises: means for detecting a software event meeting a predetermined condition; means for, in accordance with a determination that the software event meeting the condition is detected, sampling audio input received via a microphone; means for determining whether audio input received via the microphone comprises a spoken trigger; and means for, in accordance with a determination that the audio input comprises the spoken trigger, triggering a virtual assistant session.

[0021] In accordance with some embodiments, an electronic device comprises: means for receiving user input via an input device, the user input meeting a predetermined condition; means for, in accordance with receiving the user input meeting the predetermined condition, sampling audio input received via a microphone; means for determining whether the audio input comprises a spoken trigger; means for, in accordance with a determination that audio input comprises the spoken trigger: triggering a virtual assistant session, and providing a haptic output.

[0022] In accordance with some embodiments, an electronic device comprises: means for sampling audio input received via a microphone and determining whether the audio input represents a task; means for, while sampling audio input, receiving instruction to cease sampling of audio input; means for providing output acknowledging the received instruction to cease the sampling of audio input; and means for, after providing the output, continuing the sampling of audio input and determining whether the audio input comprises a task, for a duration, then ceasing the sampling of audio input.

[0023] In accordance with some embodiments, an electronic device comprises: a display; a microphone; an input device; and a processing unit coupled to the display, the microphone, and the input device, the processing unit configured to: while the display is on, receive user input via the input device, the user input meeting a predetermined condition; in accordance with receiving the user input meeting the predetermined condition, sample audio input received via the microphone; determine whether the audio input comprises a spoken trigger; and in accordance with a determination that audio input comprises the spoken trigger, trigger a virtual assistant session.

[0024] In accordance with some embodiments, an electronic device comprises: a display; a microphone; and a processing unit coupled to the display, the microphone, and the input device, the processing unit configured to: detect a software event meeting a predetermined condition; in accordance with a determination that the software event meeting the condition is detected, sample audio input received via the microphone; determine whether audio input received via the microphone comprises a spoken trigger; and in accordance with a determination that the audio input comprises the spoken trigger, trigger a virtual assistant session.

[0025] In accordance with some embodiments, an electronic device comprises: a display; a microphone; an input device; and a processing unit coupled to the display, the microphone, and the input device, the processing unit configured to: receive user input via the input device, the user input meeting a predetermined condition; in accordance with receiving the user input meeting the predetermined condition, sample audio input received via the microphone; determine whether the audio input comprises a spoken trigger; in accordance with a determination that audio input comprises the spoken trigger: trigger a virtual assistant session, and provide a haptic output.

[0026] In accordance with some embodiments, an electronic device comprises: a display; a microphone; a input device and a processing unit coupled to the display, the microphone, and the input device, the processing unit configured to: sample audio input received via the microphone and determine whether the audio input represents a task; while sampling audio input, receive instruction to cease sampling of audio input; provide output acknowledging the received instruction to cease the sampling of audio input; and after providing the output, continue the sampling of audio input and determine whether the audio input comprises a task, for a duration, then cease the sampling of audio input.

BRIEF DESCRIPTION OF THE FIGURES

[0027] For a better understanding of the various described embodiments, reference should be made to the Description of Embodiments below, in conjunction with the following drawings in which like reference numerals refer to corresponding parts throughout the figures.

[0028] **FIG. 1** illustrates a system and environment for implementing a digital assistant in accordance with some embodiments.

[0029] **FIG. 2A** is a block diagram illustrating a portable multifunction device implementing the client-side portion of a digital assistant in accordance with some embodiments.

[0030] **FIG. 2B** is a block diagram illustrating exemplary components for event handling in accordance with some embodiments.

[0031] **FIG. 3** illustrates a portable multifunction device implementing the client-side portion of a digital assistant in accordance with some embodiments.

[0032] **FIG. 4** is a block diagram of an exemplary multifunction device with a display and a touch-sensitive surface in accordance with some embodiments.

[0033] **FIG. 5A** illustrates an exemplary user interface for a menu of applications on a portable multifunction device in accordance with some embodiments.

[0034] **FIG. 5B** illustrates an exemplary user interface for a multifunction device with a touch-sensitive surface that is separate from the display in accordance with some embodiments.

[0035] **FIG. 6A** illustrates a personal electronic device in accordance with some embodiments.

[0036] **FIG. 6B** is a block diagram illustrating a personal electronic device in accordance with some embodiments.

[0037] **FIG. 7A** illustrates a digital assistant system or a server portion thereof in accordance with some embodiments.

[0038] **FIG. 7B** illustrates the functions of the digital assistant shown in FIG. 8A in accordance with some embodiments.

[0039] **FIG. 7C** illustrates a portion of an ontology in accordance with some embodiments.

[0040] **FIG. 8A-8C** illustrate exemplary techniques for triggering a virtual assistant in accordance with some embodiments.

[0041] **FIG. 9A-9C** illustrate exemplary techniques for triggering a virtual assistant in accordance with some embodiments.

[0042] **FIG. 10A-10D** illustrate exemplary techniques for triggering a virtual assistant in accordance with some embodiments.

[0043] **FIG. 11A-11C** illustrate exemplary techniques for triggering a virtual assistant in accordance with some embodiments.

[0044] **FIG. 12** is a flow diagram illustrating a method of triggering a virtual assistant in accordance with some embodiments.

[0045] **FIG. 13** is a flow diagram illustrating a method of triggering a virtual assistant in accordance with some embodiments.

[0046] **FIG. 14** is a flow diagram illustrating a method of triggering a virtual assistant in accordance with some embodiments.

[0047] **FIG. 15** is a flow diagram illustrating a method of triggering a virtual assistant in accordance with some embodiments.

[0048] **FIG. 16** is a functional block diagram of an electronic device in accordance with some embodiments.

[0049] **FIG. 17** is a functional block diagram of an electronic device in accordance with some embodiments.

[0050] **FIG. 18** is a functional block diagram of an electronic device in accordance with some embodiments.

[0051] **FIG. 19** is a functional block diagram of an electronic device in accordance with some embodiments.

[0052] **FIG. 20A-20E** illustrate exemplary techniques for interacting with a virtual assistant in accordance with some embodiments.

[0053] **FIG. 21** is a flow diagram illustrating a method of interacting with a virtual assistant in accordance with some embodiments.

DESCRIPTION OF EMBODIMENTS

[0054] In the following description of the disclosure and embodiments, reference is made to the accompanying drawings in which it is shown by way of illustration specific embodiments that can be practiced. It is to be understood that other embodiments and examples can be practiced and changes can be made without departing from the scope of the disclosure.

[0055] Virtual assistant services can be computationally intensive. For example, a virtual assistant may sample audio input from a user by activating a microphone, its associated circuitry, and corresponding software processes, thereby drawing computing resources. There is a need for electronic devices that provide virtual assistant services to employ techniques that reduce processor and associated battery power draw, such that computing resources otherwise wasted on processing unnecessary user inputs are conserved while providing an effective virtual assistant interface.

[0056] Below, FIGS. 1, 2A, 7A-7B, and 16-19 provide a description of exemplary devices for performing the techniques for transitioning between computing devices. Figures 8A-8C, 9A-9C, 10A-10D, and 11A-11C illustrate exemplary user interfaces for transitioning

between computing devices. The user interfaces in the figures are also used to illustrate the processes described below, including the processes in FIGS. 12-15.

[0057] Although the following description uses terms “first,” “second,” etc. to describe various elements, these elements should not be limited by the terms. These terms are only used to distinguish one element from another. For example, a first input could be termed a second input, and, similarly, a second input could be termed a first input, without departing from the scope of the various described examples. The first input and the second input can both be outputs and, in some cases, can be separate and different inputs.

[0058] The terminology used in the description of the various described examples herein is for the purpose of describing particular examples only and is not intended to be limiting. As used in the description of the various described examples and the appended claims, the singular forms “a,” “an,” and “the” are intended to include the plural forms as well, unless the context clearly indicates otherwise. It will also be understood that the term “and/or” as used herein refers to and encompasses any and all possible combinations of one or more of the associated listed items. It will be further understood that the terms “includes,” “including,” “comprises,” and/or “comprising,” when used in this specification, specify the presence of stated features, integers, steps, operations, elements, and/or components, but do not preclude the presence or addition of one or more other features, integers, steps, operations, elements, components, and/or groups thereof.

[0059] The term “if” may be construed to mean “when” or “upon” or “in response to determining” or “in response to detecting,” depending on the context. Similarly, the phrase “if it is determined” or “if [a stated condition or event] is detected” may be construed to mean “upon determining” or “in response to determining” or “upon detecting [the stated condition or event]” or “in response to detecting [the stated condition or event],” depending on the context.

1. System and Environment

[0060] FIG. 1 illustrates a block diagram of system 100 according to various examples. In some examples, system 100 can implement a digital assistant. The terms “digital assistant,” “virtual assistant,” “intelligent automated assistant,” or “automatic digital assistant” can refer to any information processing system that interprets natural language input in spoken and/or textual form to infer user intent, and performs actions based on the

inferred user intent. For example, to act on an inferred user intent, the system can perform one or more of the following: identifying a task flow with steps and parameters designed to accomplish the inferred user intent, inputting specific requirements from the inferred user intent into the task flow; executing the task flow by invoking programs, methods, services, APIs, or the like; and generating output responses to the user in an audible (e.g., speech) and/or visual form.

[0061] Specifically, a digital assistant can be capable of accepting a user request at least partially in the form of a natural language command, request, statement, narrative, and/or inquiry. Typically, the user request can seek either an informational answer or performance of a task by the digital assistant. A satisfactory response to the user request can be a provision of the requested informational answer, a performance of the requested task, or a combination of the two. For example, a user can ask the digital assistant a question, such as “Where am I right now?” Based on the user’s current location, the digital assistant can answer, “You are in Central Park near the west gate.” The user can also request the performance of a task, for example, “Please invite my friends to my girlfriend’s birthday party next week.” In response, the digital assistant can acknowledge the request by saying “Yes, right away,” and then send a suitable calendar invite on behalf of the user to each of the user’s friends listed in the user’s electronic address book. During performance of a requested task, the digital assistant can sometimes interact with the user in a continuous dialogue involving multiple exchanges of information over an extended period of time. There are numerous other ways of interacting with a digital assistant to request information or performance of various tasks. In addition to providing verbal responses and taking programmed actions, the digital assistant can also provide responses in other visual or audio forms, e.g., as text, alerts, music, videos, animations, etc.

[0062] As shown in FIG. 1, in some examples, a digital assistant can be implemented according to a client-server model. The digital assistant can include client-side portion 102 (hereafter “DA client 102”) executed on user device 104 and server-side portion 106 (hereafter “DA server 106”) executed on server system 108. DA client 102 can communicate with DA server 106 through one or more networks 110. DA client 102 can provide client-side functionalities such as user-facing input and output processing and communication with DA server 106. DA server 106 can provide server-side functionalities for any number of DA clients 102 each residing on a respective user device 104.

[0063] In some examples, DA server 106 can include client-facing I/O interface 112, one or more processing modules 114, data and models 116, and I/O interface to external services 118. The client-facing I/O interface 112 can facilitate the client-facing input and output processing for DA server 106. One or more processing modules 114 can utilize data and models 116 to process speech input and determine the user's intent based on natural language input. Further, one or more processing modules 114 perform task execution based on inferred user intent. In some examples, DA server 106 can communicate with external services 120 through network(s) 110 for task completion or information acquisition. I/O interface to external services 118 can facilitate such communications.

[0064] User device 104 can be any suitable electronic device. For example, user devices can be a portable multifunctional device (e.g., device 200, described below with reference to FIG. 2A), a multifunctional device (e.g., device 400, described below with reference to FIG. 4), or a personal electronic device (e.g., device 600, described below with reference to FIG. 6A-B.) A portable multifunctional device can be, for example, a mobile telephone that also contains other functions, such as PDA and/or music player functions. Specific examples of portable multifunction devices can include the iPhone®, iPod Touch®, and iPad® devices from Apple Inc. of Cupertino, California. Other examples of portable multifunction devices can include, without limitation, laptop or tablet computers. Further, in some examples, user device 104 can be a non-portable multifunctional device. In particular, user device 104 can be a desktop computer, a game console, a television, or a television set-top box. In some examples, user device 104 can include a touch-sensitive surface (e.g., touch screen displays and/or touchpads). Further, user device 104 can optionally include one or more other physical user-interface devices, such as a physical keyboard, a mouse, and/or a joystick. Various examples of electronic devices, such as multifunctional devices, are described below in greater detail.

[0065] Examples of communication network(s) 110 can include local area networks (LAN) and wide area networks (WAN), e.g., the Internet. Communication network(s) 110 can be implemented using any known network protocol, including various wired or wireless protocols, such as, for example, Ethernet, Universal Serial Bus (USB), FIREWIRE, Global System for Mobile Communications (GSM), Enhanced Data GSM Environment (EDGE), code division multiple access (CDMA), time division multiple access (TDMA), Bluetooth,

Wi-Fi, voice over Internet Protocol (VoIP), Wi-MAX, or any other suitable communication protocol.

[0066] Server system 108 can be implemented on one or more standalone data processing apparatus or a distributed network of computers. In some examples, server system 108 can also employ various virtual devices and/or services of third-party service providers (e.g., third-party cloud service providers) to provide the underlying computing resources and/or infrastructure resources of server system 108.

[0067] In some examples, user device 104 can communicate with DA server 106 via second user device 122. Second user device 122 can be similar or identical to user device 104. For example, second user device 122 can be similar to any one of devices 200, 400, or 600, described below with reference to FIGS. 2A, 4, and 6A-B, respectively. User device 104 can be configured to communicatively couple to second user device 122 via a direct communication connection, such as Bluetooth, NFC, BTLE, or the like, or via a wired or wireless network, such as a local Wi-Fi network. In some examples, second user device 122 can be configured to act as a proxy between user device 104 and DA server 106. For example, DA client 102 of user device 104 can be configured to transmit information (e.g., a user request received at user device 104) to DA server 106 via second user device 122. DA server 106 can process the information and return relevant data (e.g., data content responsive to the user request) to user device 104 via second user device 122.

[0068] In some examples, user device 104 can be configured to communicate abbreviated requests for data to second user device 122 to reduce the amount of information transmitted from user device 104. Second user device 122 can be configured to determine supplemental information to add to the abbreviated request to generate a complete request to transmit to DA server 106. This system architecture can advantageously allow user device 104 having limited communication capabilities and/or limited battery power (e.g., a watch or a similar compact electronic device) to access services provided by DA server 106 by using second user device 122, having greater communication capabilities and/or battery power (e.g., a mobile phone, laptop computer, tablet computer, or the like), as a proxy to DA server 106. While only two user devices 104 and 122 are shown in FIG. 1, it should be appreciated that system 100 can include any number and type of user devices configured in this proxy configuration to communicate with DA server system 106.

[0069] Although the digital assistant shown in FIG. 1 can include both a client-side portion (e.g., DA client 102) and a server-side portion (e.g., DA server 106), in some examples, the functions of a digital assistant can be implemented as a standalone application installed on a user device. In addition, the divisions of functionalities between the client and server portions of the digital assistant can vary in different implementations. For instance, in some examples, the DA client can be a thin-client that provides only user-facing input and output processing functions, and delegates all other functionalities of the digital assistant to a backend server.

2. Electronic Devices

[0070] Attention is now directed toward embodiments of electronic devices for implementing the client-side portion of a digital assistant. FIG. 2A is a block diagram illustrating portable multifunction device 200 with touch-sensitive display system 212 in accordance with some embodiments. Touch-sensitive display 212 is sometimes called a “touch screen” for convenience and is sometimes known as or called a “touch-sensitive display system.” Device 200 includes memory 202 (which optionally includes one or more computer-readable storage mediums), memory controller 222, one or more processing units (CPUs) 220, peripherals interface 218, RF circuitry 208, audio circuitry 210, speaker 211, microphone 213, input/output (I/O) subsystem 206, other input control devices 216, and external port 224. Device 200 optionally includes one or more optical sensors 264. Device 200 optionally includes one or more contact intensity sensors 265 for detecting intensity of contacts on device 200 (e.g., a touch-sensitive surface such as touch-sensitive display system 212 of device 200). Device 200 optionally includes one or more tactile output generators 267 for generating tactile outputs on device 200 (e.g., generating tactile outputs on a touch-sensitive surface such as touch-sensitive display system 212 of device 200 or touchpad 455 of device 400). These components optionally communicate over one or more communication buses or signal lines 203.

[0071] As used in the specification and claims, the term “intensity” of a contact on a touch-sensitive surface refers to the force or pressure (force per unit area) of a contact (e.g., a finger contact) on the touch-sensitive surface, or to a substitute (proxy) for the force or pressure of a contact on the touch-sensitive surface. The intensity of a contact has a range of values that includes at least four distinct values and more typically includes hundreds of distinct values (e.g., at least 256). Intensity of a contact is, optionally, determined (or

measured) using various approaches and various sensors or combinations of sensors. For example, one or more force sensors underneath or adjacent to the touch-sensitive surface are, optionally, used to measure force at various points on the touch-sensitive surface. In some implementations, force measurements from multiple force sensors are combined (e.g., a weighted average) to determine an estimated force of a contact. Similarly, a pressure-sensitive tip of a stylus is, optionally, used to determine a pressure of the stylus on the touch-sensitive surface. Alternatively, the size of the contact area detected on the touch-sensitive surface and/or changes thereto, the capacitance of the touch-sensitive surface proximate to the contact and/or changes thereto, and/or the resistance of the touch-sensitive surface proximate to the contact and/or changes thereto are, optionally, used as a substitute for the force or pressure of the contact on the touch-sensitive surface. In some implementations, the substitute measurements for contact force or pressure are used directly to determine whether an intensity threshold has been exceeded (e.g., the intensity threshold is described in units corresponding to the substitute measurements). In some implementations, the substitute measurements for contact force or pressure are converted to an estimated force or pressure, and the estimated force or pressure is used to determine whether an intensity threshold has been exceeded (e.g., the intensity threshold is a pressure threshold measured in units of pressure). Using the intensity of a contact as an attribute of a user input allows for user access to additional device functionality that may otherwise not be accessible by the user on a reduced-size device with limited real estate for displaying affordances (e.g., on a touch-sensitive display) and/or receiving user input (e.g., via a touch-sensitive display, a touch-sensitive surface, or a physical/mechanical control such as a knob or a button).

[0072] As used in the specification and claims, the term “tactile output” refers to physical displacement of a device relative to a previous position of the device, physical displacement of a component (e.g., a touch-sensitive surface) of a device relative to another component (e.g., housing) of the device, or displacement of the component relative to a center of mass of the device that will be detected by a user with the user’s sense of touch. For example, in situations where the device or the component of the device is in contact with a surface of a user that is sensitive to touch (e.g., a finger, palm, or other part of a user’s hand), the tactile output generated by the physical displacement will be interpreted by the user as a tactile sensation corresponding to a perceived change in physical characteristics of the device or the component of the device. For example, movement of a touch-sensitive surface (e.g., a touch-sensitive display or trackpad) is, optionally, interpreted by the user as a “down click” or “up

click” of a physical actuator button. In some cases, a user will feel a tactile sensation such as an “down click” or “up click” even when there is no movement of a physical actuator button associated with the touch-sensitive surface that is physically pressed (e.g., displaced) by the user’s movements. As another example, movement of the touch-sensitive surface is, optionally, interpreted or sensed by the user as “roughness” of the touch-sensitive surface, even when there is no change in smoothness of the touch-sensitive surface. While such interpretations of touch by a user will be subject to the individualized sensory perceptions of the user, there are many sensory perceptions of touch that are common to a large majority of users. Thus, when a tactile output is described as corresponding to a particular sensory perception of a user (e.g., an “up click,” a “down click,” “roughness”), unless otherwise stated, the generated tactile output corresponds to physical displacement of the device or a component thereof that will generate the described sensory perception for a typical (or average) user.

[0073] It should be appreciated that device 200 is only one example of a portable multifunction device, and that device 200 optionally has more or fewer components than shown, optionally combines two or more components, or optionally has a different configuration or arrangement of the components. The various components shown in FIG. 2A are implemented in hardware, software, or a combination of both hardware and software, including one or more signal processing and/or application-specific integrated circuits.

[0074] Memory 202 may include one or more computer-readable storage mediums. The computer-readable storage mediums may be tangible and non-transitory. Memory 202 may include high-speed random access memory and may also include non-volatile memory, such as one or more magnetic disk storage devices, flash memory devices, or other non-volatile solid-state memory devices. Memory controller 222 may control access to memory 202 by other components of device 200.

[0075] In some examples, a non-transitory computer-readable storage medium of memory 202 can be used to store instructions (e.g., for performing aspects of methods 1200, 1300, 1400, or 1500, described below) for use by or in connection with an instruction execution system, apparatus, or device, such as a computer-based system, processor-containing system, or other system that can fetch the instructions from the instruction execution system, apparatus, or device and execute the instructions. In other examples, the instructions (e.g., for performing aspects of methods 1200, 1300, 1400, or 1500, described

below) can be stored on a non-transitory computer-readable storage medium (not shown) of the server system 108 or can be divided between the non-transitory computer-readable storage medium of memory 202 and the non-transitory computer-readable storage medium of server system 108. In the context of this document, a “non-transitory computer-readable storage medium” can be any medium that can contain or store the program for use by or in connection with the instruction execution system, apparatus, or device.

[0076] Peripherals interface 218 can be used to couple input and output peripherals of the device to CPU 220 and memory 202. The one or more processors 220 run or execute various software programs and/or sets of instructions stored in memory 202 to perform various functions for device 200 and to process data. In some embodiments, peripherals interface 218, CPU 220, and memory controller 222 may be implemented on a single chip, such as chip 204. In some other embodiments, they may be implemented on separate chips.

[0077] RF (radio frequency) circuitry 208 receives and sends RF signals, also called electromagnetic signals. RF circuitry 208 converts electrical signals to/from electromagnetic signals and communicates with communications networks and other communications devices via the electromagnetic signals. RF circuitry 208 optionally includes well-known circuitry for performing these functions, including but not limited to an antenna system, an RF transceiver, one or more amplifiers, a tuner, one or more oscillators, a digital signal processor, a CODEC chipset, a subscriber identity module (SIM) card, memory, and so forth. RF circuitry 208 optionally communicates with networks, such as the Internet, also referred to as the World Wide Web (WWW), an intranet and/or a wireless network, such as a cellular telephone network, a wireless local area network (LAN) and/or a metropolitan area network (MAN), and other devices by wireless communication. The RF circuitry 208 optionally includes well-known circuitry for detecting near field communication (NFC) fields, such as by a short-range communication radio. The wireless communication optionally uses any of a plurality of communications standards, protocols, and technologies, including but not limited to Global System for Mobile Communications (GSM), Enhanced Data GSM Environment (EDGE), high-speed downlink packet access (HSDPA), high-speed uplink packet access (HSUPA), Evolution, Data-Only (EV-DO), HSPA, HSPA+, Dual-Cell HSPA (DC-HSPDA), long term evolution (LTE), near field communication (NFC), wideband code division multiple access (W-CDMA), code division multiple access (CDMA), time division multiple access (TDMA), Bluetooth, Bluetooth Low Energy (BTLE), Wireless Fidelity (Wi-Fi) (e.g.,

IEEE 802.11a, IEEE 802.11b, IEEE 802.11g, IEEE 802.11n, and/or IEEE 802.11ac), voice over Internet Protocol (VoIP), Wi-MAX, a protocol for e mail (e.g., Internet message access protocol (IMAP) and/or post office protocol (POP)), instant messaging (e.g., extensible messaging and presence protocol (XMPP), Session Initiation Protocol for Instant Messaging and Presence Leveraging Extensions (SIMPLE), Instant Messaging and Presence Service (IMPS)), and/or Short Message Service (SMS), or any other suitable communication protocol, including communication protocols not yet developed as of the filing date of this document.

[0078] Audio circuitry 210, speaker 211, and microphone 213 provide an audio interface between a user and device 200. Audio circuitry 210 receives audio data from peripherals interface 218, converts the audio data to an electrical signal, and transmits the electrical signal to speaker 211. Speaker 211 converts the electrical signal to human-audible sound waves. Audio circuitry 210 also receives electrical signals converted by microphone 213 from sound waves. Audio circuitry 210 converts the electrical signal to audio data and transmits the audio data to peripherals interface 218 for processing. Audio data may be retrieved from and/or transmitted to memory 202 and/or RF circuitry 208 by peripherals interface 218. In some embodiments, audio circuitry 210 also includes a headset jack (e.g., 312, FIG. 3). The headset jack provides an interface between audio circuitry 210 and removable audio input/output peripherals, such as output-only headphones or a headset with both output (e.g., a headphone for one or both ears) and input (e.g., a microphone).

[0079] I/O subsystem 206 couples input/output peripherals on device 200, such as touch screen 212 and other input control devices 216, to peripherals interface 218. I/O subsystem 206 optionally includes display controller 256, optical sensor controller 258, intensity sensor controller 259, haptic feedback controller 261, and one or more input controllers 260 for other input or control devices. The one or more input controllers 260 receive/send electrical signals from/to other input control devices 216. The other input control devices 216 optionally include physical buttons (e.g., push buttons, rocker buttons, etc.), dials, slider switches, joysticks, click wheels, and so forth. In some alternate embodiments, input controller(s) 260 are, optionally, coupled to any (or none) of the following: a keyboard, an infrared port, a USB port, and a pointer device such as a mouse. The one or more buttons (e.g., 308, FIG. 3) optionally include an up/down button for volume control of speaker 211

and/or microphone 213. The one or more buttons optionally include a push button (e.g., 306, FIG. 3).

[0080] A quick press of the push button may disengage a lock of touch screen 212 or begin a process that uses gestures on the touch screen to unlock the device, as described in U.S. Patent Application 11/322,549, “Unlocking a Device by Performing Gestures on an Unlock Image,” filed December 23, 2005, U.S. Pat. No. 7,657,849, which is hereby incorporated by reference in its entirety. A longer press of the push button (e.g., 306) may turn power to device 200 on or off. The user may be able to customize a functionality of one or more of the buttons. Touch screen 212 is used to implement virtual or soft buttons and one or more soft keyboards.

[0081] Touch-sensitive display 212 provides an input interface and an output interface between the device and a user. Display controller 256 receives and/or sends electrical signals from/to touch screen 212. Touch screen 212 displays visual output to the user. The visual output may include graphics, text, icons, video, and any combination thereof (collectively termed “graphics”). In some embodiments, some or all of the visual output may correspond to user-interface objects.

[0082] Touch screen 212 has a touch-sensitive surface, sensor, or set of sensors that accepts input from the user based on haptic and/or tactile contact. Touch screen 212 and display controller 256 (along with any associated modules and/or sets of instructions in memory 202) detect contact (and any movement or breaking of the contact) on touch screen 212 and convert the detected contact into interaction with user-interface objects (e.g., one or more soft keys, icons, web pages, or images) that are displayed on touch screen 212. In an exemplary embodiment, a point of contact between touch screen 212 and the user corresponds to a finger of the user.

[0083] Touch screen 212 may use LCD (liquid crystal display) technology, LPD (light emitting polymer display) technology, or LED (light emitting diode) technology, although other display technologies may be used in other embodiments. Touch screen 212 and display controller 256 may detect contact and any movement or breaking thereof using any of a plurality of touch sensing technologies now known or later developed, including but not limited to capacitive, resistive, infrared, and surface acoustic wave technologies, as well as other proximity sensor arrays or other elements for determining one or more points of contact

with touch screen 212. In an exemplary embodiment, projected mutual capacitance sensing technology is used, such as that found in the iPhone® and iPod Touch® from Apple Inc. of Cupertino, California.

[0084] A touch-sensitive display in some embodiments of touch screen 212 may be analogous to the multi-touch sensitive touchpads described in the following U.S. Patents: 6,323,846 (Westerman et al.), 6,570,557 (Westerman et al.), and/or 6,677,932 (Westerman), and/or U.S. Patent Publication 2002/0015024A1, each of which is hereby incorporated by reference in its entirety. However, touch screen 212 displays visual output from device 200, whereas touch-sensitive touchpads do not provide visual output.

[0085] A touch-sensitive display in some embodiments of touch screen 212 may be as described in the following applications: (1) U.S. Patent Application No. 11/381,313, “Multipoint Touch Surface Controller,” filed May 2, 2006; (2) U.S. Patent Application No. 10/840,862, “Multipoint Touchscreen,” filed May 6, 2004; (3) U.S. Patent Application No. 10/903,964, “Gestures For Touch Sensitive Input Devices,” filed July 30, 2004; (4) U.S. Patent Application No. 11/048,264, “Gestures For Touch Sensitive Input Devices,” filed January 31, 2005; (5) U.S. Patent Application No. 11/038,590, “Mode-Based Graphical User Interfaces For Touch Sensitive Input Devices,” filed January 18, 2005; (6) U.S. Patent Application No. 11/228,758, “Virtual Input Device Placement On A Touch Screen User Interface,” filed September 16, 2005; (7) U.S. Patent Application No. 11/228,700, “Operation Of A Computer With A Touch Screen Interface,” filed September 16, 2005; (8) U.S. Patent Application No. 11/228,737, “Activating Virtual Keys Of A Touch-Screen Virtual Keyboard,” filed September 16, 2005; and (9) U.S. Patent Application No. 11/367,749, “Multi-Functional Hand-Held Device,” filed March 3, 2006. All of these applications are incorporated by reference herein in their entirety.

[0086] Touch screen 212 may have a video resolution in excess of 100 dpi. In some embodiments, the touch screen has a video resolution of approximately 160 dpi. The user may make contact with touch screen 212 using any suitable object or appendage, such as a stylus, a finger, and so forth. In some embodiments, the user interface is designed to work primarily with finger-based contacts and gestures, which can be less precise than stylus-based input due to the larger area of contact of a finger on the touch screen. In some embodiments, the device translates the rough finger-based input into a precise pointer/cursor position or command for performing the actions desired by the user.

[0087] In some embodiments, in addition to the touch screen, device 200 may include a touchpad (not shown) for activating or deactivating particular functions. In some embodiments, the touchpad is a touch-sensitive area of the device that, unlike the touch screen, does not display visual output. The touchpad may be a touch-sensitive surface that is separate from touch screen 212 or an extension of the touch-sensitive surface formed by the touch screen.

[0088] Device 200 also includes power system 262 for powering the various components. Power system 262 may include a power management system, one or more power sources (e.g., battery, alternating current (AC)), a recharging system, a power failure detection circuit, a power converter or inverter, a power status indicator (e.g., a light-emitting diode (LED)) and any other components associated with the generation, management and distribution of power in portable devices.

[0089] Device 200 may also include one or more optical sensors 264. FIG. 2A shows an optical sensor coupled to optical sensor controller 258 in I/O subsystem 206. Optical sensor 264 may include charge-coupled device (CCD) or complementary metal-oxide semiconductor (CMOS) phototransistors. Optical sensor 264 receives light from the environment, projected through one or more lenses, and converts the light to data representing an image. In conjunction with imaging module 243 (also called a camera module), optical sensor 264 may capture still images or video. In some embodiments, an optical sensor is located on the back of device 200, opposite touch screen display 212 on the front of the device so that the touch screen display may be used as a viewfinder for still and/or video image acquisition. In some embodiments, an optical sensor is located on the front of the device so that the user's image may be obtained for video conferencing while the user views the other video conference participants on the touch screen display. In some embodiments, the position of optical sensor 264 can be changed by the user (e.g., by rotating the lens and the sensor in the device housing) so that a single optical sensor 264 may be used along with the touch screen display for both video conferencing and still and/or video image acquisition.

[0090] Device 200 optionally also includes one or more contact intensity sensors 265. FIG. 2A shows a contact intensity sensor coupled to intensity sensor controller 259 in I/O subsystem 206. Contact intensity sensor 265 optionally includes one or more piezoresistive strain gauges, capacitive force sensors, electric force sensors, piezoelectric force sensors,

optical force sensors, capacitive touch-sensitive surfaces, or other intensity sensors (e.g., sensors used to measure the force (or pressure) of a contact on a touch-sensitive surface). Contact intensity sensor 265 receives contact intensity information (e.g., pressure information or a proxy for pressure information) from the environment. In some embodiments, at least one contact intensity sensor is collocated with, or proximate to, a touch-sensitive surface (e.g., touch-sensitive display system 212). In some embodiments, at least one contact intensity sensor is located on the back of device 200, opposite touch screen display 212, which is located on the front of device 200.

[0091] Device 200 may also include one or more proximity sensors 266. FIG. 2A shows proximity sensor 266 coupled to peripherals interface 218. Alternately, proximity sensor 266 may be coupled to input controller 260 in I/O subsystem 206. Proximity sensor 266 may perform as described in U.S. Patent Application Nos. 11/241,839, “Proximity Detector In Handheld Device”; 11/240,788, “Proximity Detector In Handheld Device”; 11/620,702, “Using Ambient Light Sensor To Augment Proximity Sensor Output”; 11/586,862, “Automated Response To And Sensing Of User Activity In Portable Devices”; and 11/638,251, “Methods And Systems For Automatic Configuration Of Peripherals,” which are hereby incorporated by reference in their entirety. In some embodiments, the proximity sensor turns off and disables touch screen 212 when the multifunction device is placed near the user’s ear (e.g., when the user is making a phone call).

[0092] Device 200 optionally also includes one or more tactile output generators 267. FIG. 2A shows a tactile output generator coupled to haptic feedback controller 261 in I/O subsystem 206. Tactile output generator 267 optionally includes one or more electroacoustic devices such as speakers or other audio components and/or electromechanical devices that convert energy into linear motion such as a motor, solenoid, electroactive polymer, piezoelectric actuator, electrostatic actuator, or other tactile output generating component (e.g., a component that converts electrical signals into tactile outputs on the device). Contact intensity sensor 265 receives tactile feedback generation instructions from haptic feedback module 233 and generates tactile outputs on device 200 that are capable of being sensed by a user of device 200. In some embodiments, at least one tactile output generator is collocated with, or proximate to, a touch-sensitive surface (e.g., touch-sensitive display system 212) and, optionally, generates a tactile output by moving the touch-sensitive surface vertically (e.g., in/out of a surface of device 200) or laterally (e.g., back and forth in the same plane as a

surface of device 200). In some embodiments, at least one tactile output generator sensor is located on the back of device 200, opposite touch screen display 212, which is located on the front of device 200.

[0093] Device 200 may also include one or more accelerometers 268. FIG. 2A shows accelerometer 268 coupled to peripherals interface 218. Alternately, accelerometer 268 may be coupled to an input controller 260 in I/O subsystem 206. Accelerometer 268 may perform as described in U.S. Patent Publication No. 20050190059, “Acceleration-based Theft Detection System for Portable Electronic Devices,” and U.S. Patent Publication No. 20060017692, “Methods And Apparatuses For Operating A Portable Device Based On An Accelerometer,” both of which are incorporated by reference herein in their entirety. In some embodiments, information is displayed on the touch screen display in a portrait view or a landscape view based on an analysis of data received from the one or more accelerometers. Device 200 optionally includes, in addition to accelerometer(s) 268, a magnetometer (not shown) and a GPS (or GLONASS or other global navigation system) receiver (not shown) for obtaining information concerning the location and orientation (e.g., portrait or landscape) of device 200.

[0094] In some embodiments, the software components stored in memory 202 include operating system 226, communication module (or set of instructions) 228, contact/motion module (or set of instructions) 230, graphics module (or set of instructions) 232, text input module (or set of instructions) 234, Global Positioning System (GPS) module (or set of instructions) 235, Digital Assistant Client Module 229, and applications (or sets of instructions) 236. Further, memory 202 can store data and models, such as user data and models 231. Furthermore, in some embodiments, memory 202 (FIG. 2A) or 470 (FIG. 4) stores device/global internal state 257, as shown in FIGS. 2A and 4. Device/global internal state 257 includes one or more of: active application state, indicating which applications, if any, are currently active; display state, indicating what applications, views or other information occupy various regions of touch screen display 212; sensor state, including information obtained from the device’s various sensors and input control devices 216; and location information concerning the device’s location and/or attitude.

[0095] Operating system 226 (e.g., Darwin, RTXC, LINUX, UNIX, OS X, iOS, WINDOWS, or an embedded operating system such as VxWorks) includes various software components and/or drivers for controlling and managing general system tasks (e.g., memory

management, storage device control, power management, etc.) and facilitates communication between various hardware and software components.

[0096] Communication module 228 facilitates communication with other devices over one or more external ports 224 and also includes various software components for handling data received by RF circuitry 208 and/or external port 224. External port 224 (e.g., Universal Serial Bus (USB), FIREWIRE, etc.) is adapted for coupling directly to other devices or indirectly over a network (e.g., the Internet, wireless LAN, etc.). In some embodiments, the external port is a multi-pin (e.g., 30-pin) connector that is the same as, or similar to and/or compatible with, the 30-pin connector used on iPod® (trademark of Apple Inc.) devices.

[0097] Contact/motion module 230 optionally detects contact with touch screen 212 (in conjunction with display controller 256) and other touch-sensitive devices (e.g., a touchpad or physical click wheel). Contact/motion module 230 includes various software components for performing various operations related to detection of contact, such as determining if contact has occurred (e.g., detecting a finger-down event), determining an intensity of the contact (e.g., the force or pressure of the contact or a substitute for the force or pressure of the contact), determining if there is movement of the contact and tracking the movement across the touch-sensitive surface (e.g., detecting one or more finger-dragging events), and determining if the contact has ceased (e.g., detecting a finger-up event or a break in contact). Contact/motion module 230 receives contact data from the touch-sensitive surface. Determining movement of the point of contact, which is represented by a series of contact data, optionally includes determining speed (magnitude), velocity (magnitude and direction), and/or an acceleration (a change in magnitude and/or direction) of the point of contact. These operations are, optionally, applied to single contacts (e.g., one finger contacts) or to multiple simultaneous contacts (e.g., “multitouch”/multiple finger contacts). In some embodiments, contact/motion module 230 and display controller 256 detect contact on a touchpad.

[0098] In some embodiments, contact/motion module 230 uses a set of one or more intensity thresholds to determine whether an operation has been performed by a user (e.g., to determine whether a user has “clicked” on an icon). In some embodiments, at least a subset of the intensity thresholds are determined in accordance with software parameters (e.g., the intensity thresholds are not determined by the activation thresholds of particular physical actuators and can be adjusted without changing the physical hardware of device 200). For example, a mouse “click” threshold of a trackpad or touch screen display can be set to any of

a large range of predefined threshold values without changing the trackpad or touch screen display hardware. Additionally, in some implementations, a user of the device is provided with software settings for adjusting one or more of the set of intensity thresholds (e.g., by adjusting individual intensity thresholds and/or by adjusting a plurality of intensity thresholds at once with a system-level click “intensity” parameter).

[0099] Contact/motion module 230 optionally detects a gesture input by a user. Different gestures on the touch-sensitive surface have different contact patterns (e.g., different motions, timings, and/or intensities of detected contacts). Thus, a gesture is, optionally, detected by detecting a particular contact pattern. For example, detecting a finger tap gesture includes detecting a finger-down event followed by detecting a finger-up (liftoff) event at the same position (or substantially the same position) as the finger-down event (e.g., at the position of an icon). As another example, detecting a finger swipe gesture on the touch-sensitive surface includes detecting a finger-down event followed by detecting one or more finger-dragging events, and subsequently followed by detecting a finger-up (liftoff) event.

[0100] Graphics module 232 includes various known software components for rendering and displaying graphics on touch screen 212 or other display, including components for changing the visual impact (e.g., brightness, transparency, saturation, contrast, or other visual property) of graphics that are displayed. As used herein, the term “graphics” includes any object that can be displayed to a user, including, without limitation, text, web pages, icons (such as user-interface objects including soft keys), digital images, videos, animations, and the like.

[0101] In some embodiments, graphics module 232 stores data representing graphics to be used. Each graphic is, optionally, assigned a corresponding code. Graphics module 232 receives, from applications etc., one or more codes specifying graphics to be displayed along with, if necessary, coordinate data and other graphic property data, and then generates screen image data to output to display controller 256.

[0102] Haptic feedback module 233 includes various software components for generating instructions used by tactile output generator(s) 267 to produce tactile outputs at one or more locations on device 200 in response to user interactions with device 200. In some embodiments, event monitor 271 sends requests to the peripherals interface 218 at predetermined intervals. In response, peripherals interface 218 transmits event information.

In other embodiments, peripherals interface 218 transmits event information only when there is a significant event (e.g., receiving an input above a predetermined noise threshold and/or for more than a predetermined duration).

[0103] Text input module 234, which may be a component of graphics module 232, provides soft keyboards for entering text in various applications (e.g., contacts 237, e mail 240, IM 241, browser 247, and any other application that needs text input).

[0104] GPS module 235 determines the location of the device and provides this information for use in various applications (e.g., to telephone 238 for use in location-based dialing; to camera 243 as picture/video metadata; and to applications that provide location-based services such as weather widgets, local yellow page widgets, and map/navigation widgets).

[0105] Digital assistant client module 229 can include various client-side digital assistant instructions to provide the client-side functionalities of the digital assistant. For example, digital assistant client module 229 can be capable of accepting voice input (e.g., speech input), text input, touch input, and/or gestural input through various user interfaces (e.g., microphone 213, accelerometer(s) 268, touch-sensitive display system 212, optical sensor(s) 229, other input control devices 216, etc.) of portable multifunction device 200. Digital assistant client module 229 can also be capable of providing output in audio (e.g., speech output), visual, and/or tactile forms through various output interfaces (e.g., speaker 211, touch-sensitive display system 212, tactile output generator(s) 267, etc.) of portable multifunction device 200. For example, output can be provided as voice, sound, alerts, text messages, menus, graphics, videos, animations, vibrations, and/or combinations of two or more of the above. During operation, digital assistant client module 229 can communicate with DA server 106 using RF circuitry 208.

[0106] User data and models 231 can include various data associated with the user (e.g., user-specific vocabulary data, user preference data, user-specified name pronunciations, data from the user's electronic address book, to-do lists, shopping lists, etc.) to provide the client-side functionalities of the digital assistant. Further, user data and models 231 can includes various models (e.g., speech recognition models, statistical language models, natural language processing models, ontology, task flow models, service models, etc.) for processing user input and determining user intent.

[0107] In some examples, digital assistant client module 229 can utilize the various sensors, subsystems, and peripheral devices of portable multifunction device 200 to gather additional information from the surrounding environment of the portable multifunction device 200 to establish a context associated with a user, the current user interaction, and/or the current user input. In some examples, digital assistant client module 229 can provide the contextual information or a subset thereof with the user input to DA server 106 to help infer the user's intent. In some examples, the digital assistant can also use the contextual information to determine how to prepare and deliver outputs to the user. Contextual information can be referred to as context data.

[0108] In some examples, the contextual information that accompanies the user input can include sensor information, e.g., lighting, ambient noise, ambient temperature, images or videos of the surrounding environment, etc. In some examples, the contextual information can also include the physical state of the device, e.g., device orientation, device location, device temperature, power level, speed, acceleration, motion patterns, cellular signals strength, etc. In some examples, information related to the software state of DA server 106, e.g., running processes, installed programs, past and present network activities, background services, error logs, resources usage, etc., and of portable multifunction device 200 can be provided to DA server 106 as contextual information associated with a user input.

[0109] In some examples, the digital assistant client module 229 can selectively provide information (e.g., user data 231) stored on the portable multifunction device 200 in response to requests from DA server 106. In some examples, digital assistant client module 229 can also elicit additional input from the user via a natural language dialogue or other user interfaces upon request by DA server 106. Digital assistant client module 229 can pass the additional input to DA server 106 to help DA server 106 in intent deduction and/or fulfillment of the user's intent expressed in the user request.

[0110] A more detailed description of a digital assistant is described below with reference to FIGS. 7A-C. It should be recognized that digital assistant client module 229 can include any number of the sub-modules of digital assistant module 726 described below.

[0111] Applications 236 may include the following modules (or sets of instructions), or a subset or superset thereof:

- Contacts module 237 (sometimes called an address book or contact list);

- Telephone module 238;
- Video conference module 239;
- E-mail client module 240;
- Instant messaging (IM) module 241;
- Workout support module 242;
- Camera module 243 for still and/or video images;
- Image management module 244;
- Video player module;
- Music player module;
- Browser module 247;
- Calendar module 248;
- Widget modules 249, which may include one or more of: weather widget 249-1, stocks widget 249-2, calculator widget 249-3, alarm clock widget 249-4, dictionary widget 249-5, and other widgets obtained by the user, as well as user-created widgets 249-6;
- Widget creator module 250 for making user-created widgets 249-6;
- Search module 251;
- Video and music player module 252, which merges video player module and music player module;
- Notes module 253;
- Map module 254; and/or
- Online video module 255.

[0112] Examples of other applications 236 that may be stored in memory 202 include other word processing applications, other image editing applications, drawing applications, presentation applications, JAVA-enabled applications, encryption, digital rights management, voice recognition, and voice replication.

[0113] In conjunction with touch screen 212, display controller 256, contact/motion module 230, graphics module 232, and text input module 234, contacts module 237 may be used to manage an address book or contact list (e.g., stored in application internal state 292 of contacts module 237 in memory 202 or memory 470), including: adding name(s) to the address book; deleting name(s) from the address book; associating telephone number(s), e-mail address(es), physical address(es) or other information with a name; associating an image with a name; categorizing and sorting names; providing telephone numbers or e-mail addresses to initiate and/or facilitate communications by telephone 238, video conference module 239, e-mail 240, or IM 241; and so forth.

[0114] In conjunction with RF circuitry 208, audio circuitry 210, speaker 211, microphone 213, touch screen 212, display controller 256, contact/motion module 230, graphics module 232, and text input module 234, telephone module 238 may be used to enter a sequence of characters corresponding to a telephone number, access one or more telephone numbers in contacts module 237, modify a telephone number that has been entered, dial a respective telephone number, conduct a conversation, and disconnect or hang up when the conversation is completed. As noted above, the wireless communication may use any of a plurality of communications standards, protocols, and technologies.

[0115] In conjunction with RF circuitry 208, audio circuitry 210, speaker 211, microphone 213, touch screen 212, display controller 256, optical sensor 264, optical sensor controller 258, contact/motion module 230, graphics module 232, text input module 234, contacts module 237, and telephone module 238, video conference module 239 includes executable instructions to initiate, conduct, and terminate a video conference between a user and one or more other participants in accordance with user instructions.

[0116] In conjunction with RF circuitry 208, touch screen 212, display controller 256, contact/motion module 230, graphics module 232, and text input module 234, e-mail client module 240 includes executable instructions to create, send, receive, and manage e-mail in response to user instructions. In conjunction with image management module 244, e-mail

client module 240 makes it very easy to create and send e-mails with still or video images taken with camera module 243.

[0117] In conjunction with RF circuitry 208, touch screen 212, display controller 256, contact/motion module 230, graphics module 232, and text input module 234, the instant messaging module 241 includes executable instructions to enter a sequence of characters corresponding to an instant message, to modify previously entered characters, to transmit a respective instant message (for example, using a Short Message Service (SMS) or Multimedia Message Service (MMS) protocol for telephony-based instant messages or using XMPP, SIMPLE, or IMPS for Internet-based instant messages), to receive instant messages, and to view received instant messages. In some embodiments, transmitted and/or received instant messages may include graphics, photos, audio files, video files and/or other attachments as are supported in an MMS and/or an Enhanced Messaging Service (EMS). As used herein, “instant messaging” refers to both telephony-based messages (e.g., messages sent using SMS or MMS) and Internet-based messages (e.g., messages sent using XMPP, SIMPLE, or IMPS).

[0118] In conjunction with RF circuitry 208, touch screen 212, display controller 256, contact/motion module 230, graphics module 232, text input module 234, GPS module 235, map module 254, and music player module, workout support module 242 includes executable instructions to create workouts (e.g., with time, distance, and/or calorie burning goals); communicate with workout sensors (sports devices); receive workout sensor data; calibrate sensors used to monitor a workout; select and play music for a workout; and display, store, and transmit workout data.

[0119] In conjunction with touch screen 212, display controller 256, optical sensor(s) 264, optical sensor controller 258, contact/motion module 230, graphics module 232, and image management module 244, camera module 243 includes executable instructions to capture still images or video (including a video stream) and store them into memory 202, modify characteristics of a still image or video, or delete a still image or video from memory 202.

[0120] In conjunction with touch screen 212, display controller 256, contact/motion module 230, graphics module 232, text input module 234, and camera module 243, image management module 244 includes executable instructions to arrange, modify (e.g., edit), or

otherwise manipulate, label, delete, present (e.g., in a digital slide show or album), and store still and/or video images.

[0121] In conjunction with RF circuitry 208, touch screen 212, display controller 256, contact/motion module 230, graphics module 232, and text input module 234, browser module 247 includes executable instructions to browse the Internet in accordance with user instructions, including searching, linking to, receiving, and displaying web pages or portions thereof, as well as attachments and other files linked to web pages.

[0122] In conjunction with RF circuitry 208, touch screen 212, display controller 256, contact/motion module 230, graphics module 232, text input module 234, e-mail client module 240, and browser module 247, calendar module 248 includes executable instructions to create, display, modify, and store calendars and data associated with calendars (e.g., calendar entries, to-do lists, etc.) in accordance with user instructions.

[0123] In conjunction with RF circuitry 208, touch screen 212, display controller 256, contact/motion module 230, graphics module 232, text input module 234, and browser module 247, widget modules 249 are mini-applications that may be downloaded and used by a user (e.g., weather widget 249-1, stocks widget 249-2, calculator widget 249-3, alarm clock widget 249-4, and dictionary widget 249-5) or created by the user (e.g., user-created widget 249-6). In some embodiments, a widget includes an HTML (Hypertext Markup Language) file, a CSS (Cascading Style Sheets) file, and a JavaScript file. In some embodiments, a widget includes an XML (Extensible Markup Language) file and a JavaScript file (e.g., Yahoo! Widgets).

[0124] In conjunction with RF circuitry 208, touch screen 212, display controller 256, contact/motion module 230, graphics module 232, text input module 234, and browser module 247, the widget creator module 250 may be used by a user to create widgets (e.g., turning a user-specified portion of a web page into a widget).

[0125] In conjunction with touch screen 212, display controller 256, contact/motion module 230, graphics module 232, and text input module 234, search module 251 includes executable instructions to search for text, music, sound, image, video, and/or other files in memory 202 that match one or more search criteria (e.g., one or more user-specified search terms) in accordance with user instructions.

[0126] In conjunction with touch screen 212, display controller 256, contact/motion module 230, graphics module 232, audio circuitry 210, speaker 211, RF circuitry 208, and browser module 247, video and music player module 252 includes executable instructions that allow the user to download and play back recorded music and other sound files stored in one or more file formats, such as MP3 or AAC files, and executable instructions to display, present, or otherwise play back videos (e.g., on touch screen 212 or on an external, connected display via external port 224). In some embodiments, device 200 optionally includes the functionality of an MP3 player, such as an iPod (trademark of Apple Inc.).

[0127] In conjunction with touch screen 212, display controller 256, contact/motion module 230, graphics module 232, and text input module 234, notes module 253 includes executable instructions to create and manage notes, to-do lists, and the like in accordance with user instructions.

[0128] In conjunction with RF circuitry 208, touch screen 212, display controller 256, contact/motion module 230, graphics module 232, text input module 234, GPS module 235, and browser module 247, map module 254 may be used to receive, display, modify, and store maps and data associated with maps (e.g., driving directions, data on stores and other points of interest at or near a particular location, and other location-based data) in accordance with user instructions.

[0129] In conjunction with touch screen 212, display controller 256, contact/motion module 230, graphics module 232, audio circuitry 210, speaker 211, RF circuitry 208, text input module 234, e-mail client module 240, and browser module 247, online video module 255 includes instructions that allow the user to access, browse, receive (e.g., by streaming and/or download), play back (e.g., on the touch screen or on an external, connected display via external port 224), send an e-mail with a link to a particular online video, and otherwise manage online videos in one or more file formats, such as H.264. In some embodiments, instant messaging module 241, rather than e-mail client module 240, is used to send a link to a particular online video. Additional description of the online video application can be found in U.S. Provisional Patent Application No. 60/936,562, "Portable Multifunction Device, Method, and Graphical User Interface for Playing Online Videos," filed June 20, 2007, and U.S. Patent Application No. 11/968,067, "Portable Multifunction Device, Method, and Graphical User Interface for Playing Online Videos," filed December 31, 2007, the contents of which are hereby incorporated by reference in their entirety.

[0130] Each of the above-identified modules and applications corresponds to a set of executable instructions for performing one or more functions described above and the methods described in this application (e.g., the computer-implemented methods and other information processing methods described herein). These modules (e.g., sets of instructions) need not be implemented as separate software programs, procedures, or modules, and thus various subsets of these modules may be combined or otherwise rearranged in various embodiments. For example, video player module may be combined with music player module into a single module (e.g., video and music player module 252, FIG. 2A). In some embodiments, memory 202 may store a subset of the modules and data structures identified above. Furthermore, memory 202 may store additional modules and data structures not described above.

[0131] In some embodiments, device 200 is a device where operation of a predefined set of functions on the device is performed exclusively through a touch screen and/or a touchpad. By using a touch screen and/or a touchpad as the primary input control device for operation of device 200, the number of physical input control devices (such as push buttons, dials, and the like) on device 200 may be reduced.

[0132] The predefined set of functions that are performed exclusively through a touch screen and/or a touchpad optionally include navigation between user interfaces. In some embodiments, the touchpad, when touched by the user, navigates device 200 to a main, home, or root menu from any user interface that is displayed on device 200. In such embodiments, a “menu button” is implemented using a touchpad. In some other embodiments, the menu button is a physical push button or other physical input control device instead of a touchpad.

[0133] FIG. 2B is a block diagram illustrating exemplary components for event handling in accordance with some embodiments. In some embodiments, memory 202 (FIG. 2A) or 470 (FIG. 4) includes event sorter 270 (e.g., in operating system 226) and a respective application 236-1 (e.g., any of the aforementioned applications 237-251, 255, 480-490).

[0134] Event sorter 270 receives event information and determines the application 236-1 and application view 291 of application 236-1 to which to deliver the event information. Event sorter 270 includes event monitor 271 and event dispatcher module 274. In some embodiments, application 236-1 includes application internal state 292, which indicates the current application view(s) displayed on touch-sensitive display 212 when the application is

active or executing. In some embodiments, device/global internal state 257 is used by event sorter 270 to determine which application(s) is (are) currently active, and application internal state 292 is used by event sorter 270 to determine application views 291 to which to deliver event information.

[0135] In some embodiments, application internal state 292 includes additional information, such as one or more of: resume information to be used when application 236-1 resumes execution, user interface state information that indicates information being displayed or that is ready for display by application 236-1, a state queue for enabling the user to go back to a prior state or view of application 236-1, and a redo/undo queue of previous actions taken by the user.

[0136] Event monitor 271 receives event information from peripherals interface 218. Event information includes information about a sub-event (e.g., a user touch on touch-sensitive display 212, as part of a multi-touch gesture). Peripherals interface 218 transmits information it receives from I/O subsystem 206 or a sensor, such as proximity sensor 266, accelerometer(s) 268, and/or microphone 213 (through audio circuitry 210). Information that peripherals interface 218 receives from I/O subsystem 206 includes information from touch-sensitive display 212 or a touch-sensitive surface.

[0137] In some embodiments, event sorter 270 also includes a hit view determination module 272 and/or an active event recognizer determination module 273.

[0138] Hit view determination module 272 provides software procedures for determining where a sub-event has taken place within one or more views when touch-sensitive display 212 displays more than one view. Views are made up of controls and other elements that a user can see on the display.

[0139] Another aspect of the user interface associated with an application is a set of views, sometimes herein called application views or user interface windows, in which information is displayed and touch-based gestures occur. The application views (of a respective application) in which a touch is detected may correspond to programmatic levels within a programmatic or view hierarchy of the application. For example, the lowest level view in which a touch is detected may be called the hit view, and the set of events that are recognized as proper inputs may be determined based, at least in part, on the hit view of the initial touch that begins a touch-based gesture.

[0140] Hit view determination module 272 receives information related to sub events of a touch-based gesture. When an application has multiple views organized in a hierarchy, hit view determination module 272 identifies a hit view as the lowest view in the hierarchy which should handle the sub-event. In most circumstances, the hit view is the lowest level view in which an initiating sub-event occurs (e.g., the first sub-event in the sequence of sub-events that form an event or potential event). Once the hit view is identified by the hit view determination module 272, the hit view typically receives all sub-events related to the same touch or input source for which it was identified as the hit view.

[0141] Active event recognizer determination module 273 determines which view or views within a view hierarchy should receive a particular sequence of sub-events. In some embodiments, active event recognizer determination module 273 determines that only the hit view should receive a particular sequence of sub-events. In other embodiments, active event recognizer determination module 273 determines that all views that include the physical location of a sub-event are actively involved views, and therefore determines that all actively involved views should receive a particular sequence of sub-events. In other embodiments, even if touch sub-events were entirely confined to the area associated with one particular view, views higher in the hierarchy would still remain as actively involved views.

[0142] Event dispatcher module 274 dispatches the event information to an event recognizer (e.g., event recognizer 280). In embodiments including active event recognizer determination module 273, event dispatcher module 274 delivers the event information to an event recognizer determined by active event recognizer determination module 273. In some embodiments, event dispatcher module 274 stores in an event queue the event information, which is retrieved by a respective event receiver 282.

[0143] In some embodiments, operating system 226 includes event sorter 270. Alternatively, application 236-1 includes event sorter 270. In yet other embodiments, event sorter 270 is a stand-alone module, or a part of another module stored in memory 202, such as contact/motion module 230.

[0144] In some embodiments, application 236-1 includes a plurality of event handlers 290 and one or more application views 291, each of which includes instructions for handling touch events that occur within a respective view of the application's user interface. Each application view 291 of the application 236-1 includes one or more event recognizers 280.

Typically, a respective application view 291 includes a plurality of event recognizers 280. In other embodiments, one or more of event recognizers 280 are part of a separate module, such as a user interface kit (not shown) or a higher level object from which application 236-1 inherits methods and other properties. In some embodiments, a respective event handler 290 includes one or more of: data updater 276, object updater 277, GUI updater 278, and/or event data 279 received from event sorter 270. Event handler 290 may utilize or call data updater 276, object updater 277, or GUI updater 278 to update the application internal state 292. Alternatively, one or more of the application views 291 include one or more respective event handlers 290. Also, in some embodiments, one or more of data updater 276, object updater 277, and GUI updater 278 are included in a respective application view 291.

[0145] A respective event recognizer 280 receives event information (e.g., event data 279) from event sorter 270 and identifies an event from the event information. Event recognizer 280 includes event receiver 282 and event comparator 284. In some embodiments, event recognizer 280 also includes at least a subset of: metadata 283, and event delivery instructions 288 (which may include sub-event delivery instructions).

[0146] Event receiver 282 receives event information from event sorter 270. The event information includes information about a sub-event, for example, a touch or a touch movement. Depending on the sub-event, the event information also includes additional information, such as location of the sub-event. When the sub-event concerns motion of a touch, the event information may also include speed and direction of the sub-event. In some embodiments, events include rotation of the device from one orientation to another (e.g., from a portrait orientation to a landscape orientation, or vice versa), and the event information includes corresponding information about the current orientation (also called device attitude) of the device.

[0147] Event comparator 284 compares the event information to predefined event or sub-event definitions and, based on the comparison, determines an event or sub event, or determines or updates the state of an event or sub-event. In some embodiments, event comparator 284 includes event definitions 286. Event definitions 286 contain definitions of events (e.g., predefined sequences of sub-events), for example, event 1 (287-1), event 2 (287-2), and others. In some embodiments, sub-events in an event (287) include, for example, touch begin, touch end, touch movement, touch cancellation, and multiple touching. In one example, the definition for event 1 (287-1) is a double tap on a displayed object. The double

tap, for example, comprises a first touch (touch begin) on the displayed object for a predetermined phase, a first liftoff (touch end) for a predetermined phase, a second touch (touch begin) on the displayed object for a predetermined phase, and a second liftoff (touch end) for a predetermined phase. In another example, the definition for event 2 (287-2) is a dragging on a displayed object. The dragging, for example, comprises a touch (or contact) on the displayed object for a predetermined phase, a movement of the touch across touch-sensitive display 212, and liftoff of the touch (touch end). In some embodiments, the event also includes information for one or more associated event handlers 290.

[0148] In some embodiments, event definition 287 includes a definition of an event for a respective user-interface object. In some embodiments, event comparator 284 performs a hit test to determine which user-interface object is associated with a sub-event. For example, in an application view in which three user-interface objects are displayed on touch-sensitive display 212, when a touch is detected on touch-sensitive display 212, event comparator 284 performs a hit test to determine which of the three user-interface objects is associated with the touch (sub-event). If each displayed object is associated with a respective event handler 290, the event comparator uses the result of the hit test to determine which event handler 290 should be activated. For example, event comparator 284 selects an event handler associated with the sub-event and the object triggering the hit test.

[0149] In some embodiments, the definition for a respective event (287) also includes delayed actions that delay delivery of the event information until after it has been determined whether the sequence of sub-events does or does not correspond to the event recognizer's event type.

[0150] When a respective event recognizer 280 determines that the series of sub-events do not match any of the events in event definitions 286, the respective event recognizer 280 enters an event impossible, event failed, or event ended state, after which it disregards subsequent sub-events of the touch-based gesture. In this situation, other event recognizers, if any, that remain active for the hit view continue to track and process sub-events of an ongoing touch-based gesture.

[0151] In some embodiments, a respective event recognizer 280 includes metadata 283 with configurable properties, flags, and/or lists that indicate how the event delivery system should perform sub-event delivery to actively involved event recognizers. In some

embodiments, metadata 283 includes configurable properties, flags, and/or lists that indicate how event recognizers may interact, or are enabled to interact, with one another. In some embodiments, metadata 283 includes configurable properties, flags, and/or lists that indicate whether sub-events are delivered to varying levels in the view or programmatic hierarchy.

[0152] In some embodiments, a respective event recognizer 280 activates event handler 290 associated with an event when one or more particular sub-events of an event are recognized. In some embodiments, a respective event recognizer 280 delivers event information associated with the event to event handler 290. Activating an event handler 290 is distinct from sending (and deferred sending) sub-events to a respective hit view. In some embodiments, event recognizer 280 throws a flag associated with the recognized event, and event handler 290 associated with the flag catches the flag and performs a predefined process.

[0153] In some embodiments, event delivery instructions 288 include sub-event delivery instructions that deliver event information about a sub-event without activating an event handler. Instead, the sub-event delivery instructions deliver event information to event handlers associated with the series of sub-events or to actively involved views. Event handlers associated with the series of sub-events or with actively involved views receive the event information and perform a predetermined process.

[0154] In some embodiments, data updater 276 creates and updates data used in application 236-1. For example, data updater 276 updates the telephone number used in contacts module 237, or stores a video file used in video player module. In some embodiments, object updater 277 creates and updates objects used in application 236-1. For example, object updater 277 creates a new user-interface object or updates the position of a user-interface object. GUI updater 278 updates the GUI. For example, GUI updater 278 prepares display information and sends it to graphics module 232 for display on a touch-sensitive display.

[0155] In some embodiments, event handler(s) 290 includes or has access to data updater 276, object updater 277, and GUI updater 278. In some embodiments, data updater 276, object updater 277, and GUI updater 278 are included in a single module of a respective application 236-1 or application view 291. In other embodiments, they are included in two or more software modules.

[0156] It shall be understood that the foregoing discussion regarding event handling of user touches on touch-sensitive displays also applies to other forms of user inputs to operate multifunction devices 200 with input devices, not all of which are initiated on touch screens. For example, mouse movement and mouse button presses, optionally coordinated with single or multiple keyboard presses or holds; contact movements such as taps, drags, scrolls, etc. on touchpads; pen stylus inputs; movement of the device; oral instructions; detected eye movements; biometric inputs; and/or any combination thereof are optionally utilized as inputs corresponding to sub-events which define an event to be recognized.

[0157] FIG. 3 illustrates a portable multifunction device 200 having a touch screen 212 in accordance with some embodiments. The touch screen optionally displays one or more graphics within user interface (UI) 300. In this embodiment, as well as others described below, a user is enabled to select one or more of the graphics by making a gesture on the graphics, for example, with one or more fingers 302 (not drawn to scale in the figure) or one or more styluses 303 (not drawn to scale in the figure). In some embodiments, selection of one or more graphics occurs when the user breaks contact with the one or more graphics. In some embodiments, the gesture optionally includes one or more taps, one or more swipes (from left to right, right to left, upward and/or downward), and/or a rolling of a finger (from right to left, left to right, upward and/or downward) that has made contact with device 200. In some implementations or circumstances, inadvertent contact with a graphic does not select the graphic. For example, a swipe gesture that sweeps over an application icon optionally does not select the corresponding application when the gesture corresponding to selection is a tap.

[0158] Device 200 may also include one or more physical buttons, such as “home” or menu button 304. As described previously, menu button 304 may be used to navigate to any application 236 in a set of applications that may be executed on device 200. Alternatively, in some embodiments, the menu button is implemented as a soft key in a GUI displayed on touch screen 212.

[0159] In one embodiment, device 200 includes touch screen 212, menu button 304, push button 306 for powering the device on/off and locking the device, volume adjustment button(s) 308, subscriber identity module (SIM) card slot 310, headset jack 312, and docking/charging external port 224. Push button 306 is, optionally, used to turn the power on/off on the device by depressing the button and holding the button in the depressed state for

a predefined time interval; to lock the device by depressing the button and releasing the button before the predefined time interval has elapsed; and/or to unlock the device or initiate an unlock process. In an alternative embodiment, device 200 also accepts verbal input for activation or deactivation of some functions through microphone 213. Device 200 also, optionally, includes one or more contact intensity sensors 265 for detecting intensity of contacts on touch screen 212 and/or one or more tactile output generators 267 for generating tactile outputs for a user of device 200.

[0160] FIG. 4 is a block diagram of an exemplary multifunction device with a display and a touch-sensitive surface in accordance with some embodiments. Device 400 need not be portable. In some embodiments, device 400 is a laptop computer, a desktop computer, a tablet computer, a multimedia player device, a navigation device, an educational device (such as a child's learning toy), a gaming system, or a control device (e.g., a home or industrial controller). Device 400 typically includes one or more processing units (CPUs) 410, one or more network or other communications interfaces 460, memory 470, and one or more communication buses 420 for interconnecting these components. Communication buses 420 optionally include circuitry (sometimes called a chipset) that interconnects and controls communications between system components. Device 400 includes input/output (I/O) interface 430 comprising display 440, which is typically a touch screen display. I/O interface 430 also optionally includes a keyboard and/or mouse (or other pointing device) 450 and touchpad 455, tactile output generator 457 for generating tactile outputs on device 400 (e.g., similar to tactile output generator(s) 267 described above with reference to FIG. 2A), sensors 459 (e.g., optical, acceleration, proximity, touch-sensitive, and/or contact intensity sensors similar to contact intensity sensor(s) 265 described above with reference to FIG. 2A). Memory 470 includes high-speed random access memory, such as DRAM, SRAM, DDR RAM, or other random access solid state memory devices; and optionally includes non-volatile memory, such as one or more magnetic disk storage devices, optical disk storage devices, flash memory devices, or other non-volatile solid state storage devices. Memory 470 optionally includes one or more storage devices remotely located from CPU(s) 410. In some embodiments, memory 470 stores programs, modules, and data structures analogous to the programs, modules, and data structures stored in memory 202 of portable multifunction device 200 (FIG. 2A), or a subset thereof. Furthermore, memory 470 optionally stores additional programs, modules, and data structures not present in memory 202 of portable multifunction device 200. For example, memory 470 of device 400 optionally stores drawing

module 480, presentation module 482, word processing module 484, website creation module 486, disk authoring module 488, and/or spreadsheet module 490, while memory 202 of portable multifunction device 200 (FIG. 2A) optionally does not store these modules.

[0161] Each of the above-identified elements in FIG. 4 may be stored in one or more of the previously mentioned memory devices. Each of the above-identified modules corresponds to a set of instructions for performing a function described above. The above-identified modules or programs (e.g., sets of instructions) need not be implemented as separate software programs, procedures, or modules, and thus various subsets of these modules may be combined or otherwise rearranged in various embodiments. In some embodiments, memory 470 may store a subset of the modules and data structures identified above. Furthermore, memory 470 may store additional modules and data structures not described above.

[0162] Attention is now directed towards embodiments of user interfaces that may be implemented on, for example, portable multifunction device 200.

[0163] FIG. 5A illustrates an exemplary user interface for a menu of applications on portable multifunction device 200 in accordance with some embodiments. Similar user interfaces may be implemented on device 400. In some embodiments, user interface 500 includes the following elements, or a subset or superset thereof:

[0164] Signal strength indicator(s) 502 for wireless communication(s), such as cellular and Wi-Fi signals;

- Time 504;
- Bluetooth indicator 505;
- Battery status indicator 506;
- Tray 508 with icons for frequently used applications, such as:
 - Icon 516 for telephone module 238, labeled “Phone,” which optionally includes an indicator 514 of the number of missed calls or voicemail messages;

- Icon 518 for e-mail client module 240, labeled “Mail,” which optionally includes an indicator 510 of the number of unread e-mails;
- Icon 520 for browser module 247, labeled “Browser;” and
- Icon 522 for video and music player module 252, also referred to as iPod (trademark of Apple Inc.) module 252, labeled “iPod;” and
- Icons for other applications, such as:
 - Icon 524 for IM module 241, labeled “Messages;”
 - Icon 526 for calendar module 248, labeled “Calendar;”
 - Icon 528 for image management module 244, labeled “Photos;”
 - Icon 530 for camera module 243, labeled “Camera;”
 - Icon 532 for online video module 255, labeled “Online Video;”
 - Icon 534 for stocks widget 249-2, labeled “Stocks;”
 - Icon 536 for map module 254, labeled “Maps;”
 - Icon 538 for weather widget 249-1, labeled “Weather;”
 - Icon 540 for alarm clock widget 249-4, labeled “Clock;”
 - Icon 542 for workout support module 242, labeled “Workout Support;”
 - Icon 544 for notes module 253, labeled “Notes;” and
 - Icon 546 for a settings application or module, labeled “Settings,” which provides access to settings for device 200 and its various applications 236.

[0165] It should be noted that the icon labels illustrated in FIG. 5A are merely exemplary. For example, icon 522 for video and music player module 252 may optionally be labeled “Music” or “Music Player.” Other labels are, optionally, used for various application icons. In some embodiments, a label for a respective application icon includes a name of an application corresponding to the respective application icon. In some embodiments, a label

for a particular application icon is distinct from a name of an application corresponding to the particular application icon.

[0166] FIG. 5B illustrates an exemplary user interface on a device (e.g., device 400, FIG. 4) with a touch-sensitive surface 551 (e.g., a tablet or touchpad 455, FIG. 4) that is separate from the display 550 (e.g., touch screen display 212). Device 400 also, optionally, includes one or more contact intensity sensors (e.g., one or more of sensors 457) for detecting intensity of contacts on touch-sensitive surface 551 and/or one or more tactile output generators 459 for generating tactile outputs for a user of device 400.

[0167] Although some of the examples which follow will be given with reference to inputs on touch screen display 212 (where the touch-sensitive surface and the display are combined), in some embodiments, the device detects inputs on a touch-sensitive surface that is separate from the display, as shown in FIG. 5B. In some embodiments, the touch-sensitive surface (e.g., 551 in FIG. 5B) has a primary axis (e.g., 552 in FIG. 5B) that corresponds to a primary axis (e.g., 553 in FIG. 5B) on the display (e.g., 550). In accordance with these embodiments, the device detects contacts (e.g., 560 and 562 in FIG. 5B) with the touch-sensitive surface 551 at locations that correspond to respective locations on the display (e.g., in FIG. 5B, 560 corresponds to 568 and 562 corresponds to 570). In this way, user inputs (e.g., contacts 560 and 562, and movements thereof) detected by the device on the touch-sensitive surface (e.g., 551 in FIG. 5B) are used by the device to manipulate the user interface on the display (e.g., 550 in FIG. 5B) of the multifunction device when the touch-sensitive surface is separate from the display. It should be understood that similar methods are, optionally, used for other user interfaces described herein.

[0168] Additionally, while the following examples are given primarily with reference to finger inputs (e.g., finger contacts, finger tap gestures, finger swipe gestures), it should be understood that, in some embodiments, one or more of the finger inputs are replaced with input from another input device (e.g., a mouse-based input or stylus input). For example, a swipe gesture is, optionally, replaced with a mouse click (e.g., instead of a contact) followed by movement of the cursor along the path of the swipe (e.g., instead of movement of the contact). As another example, a tap gesture is, optionally, replaced with a mouse click while the cursor is located over the location of the tap gesture (e.g., instead of detection of the contact followed by ceasing to detect the contact). Similarly, when multiple user inputs are

simultaneously detected, it should be understood that multiple computer mice are, optionally, used simultaneously, or a mouse and finger contacts are, optionally, used simultaneously.

[0169] FIG. 6A illustrates exemplary personal electronic device 600. Device 600 includes body 602. In some embodiments, device 600 can include some or all of the features described with respect to devices 200 and 400 (e.g., FIGS. 2A-4B). In some embodiments, device 600 has touch-sensitive display screen 604, hereafter touch screen 604. Alternatively, or in addition to touch screen 604, device 600 has a display and a touch-sensitive surface. As with devices 200 and 400, in some embodiments, touch screen 604 (or the touch-sensitive surface) may have one or more intensity sensors for detecting intensity of contacts (e.g., touches) being applied. The one or more intensity sensors of touch screen 604 (or the touch-sensitive surface) can provide output data that represents the intensity of touches. The user interface of device 600 can respond to touches based on their intensity, meaning that touches of different intensities can invoke different user interface operations on device 600.

[0170] Techniques for detecting and processing touch intensity may be found, for example, in related applications: International Patent Application Serial No. PCT/US2013/040061, titled “Device, Method, and Graphical User Interface for Displaying User Interface Objects Corresponding to an Application,” filed May 8, 2013, and International Patent Application Serial No. PCT/US2013/069483, titled “Device, Method, and Graphical User Interface for Transitioning Between Touch Input to Display Output Relationships,” filed November 11, 2013, each of which is hereby incorporated by reference in their entirety.

[0171] In some embodiments, device 600 has one or more input mechanisms 606 and 608. Input mechanisms 606 and 608, if included, can be physical. Examples of physical input mechanisms include push buttons and rotatable mechanisms. In some embodiments, device 600 has one or more attachment mechanisms. Such attachment mechanisms, if included, can permit attachment of device 600 with, for example, hats, eyewear, earrings, necklaces, shirts, jackets, bracelets, watch straps, chains, trousers, belts, shoes, purses, backpacks, and so forth. These attachment mechanisms may permit device 600 to be worn by a user.

[0172] FIG. 6B depicts exemplary personal electronic device 600. In some embodiments, device 600 can include some or all of the components described with respect to FIGS. 2A,

2B, and 4. Device 600 has bus 612 that operatively couples I/O section 614 with one or more computer processors 616 and memory 618. I/O section 614 can be connected to display 604, which can have touch-sensitive component 622 and, optionally, touch-intensity sensitive component 624. In addition, I/O section 614 can be connected with communication unit 630 for receiving application and operating system data, using Wi-Fi, Bluetooth, near field communication (NFC), cellular, and/or other wireless communication techniques. Device 600 can include input mechanisms 606 and/or 608. Input mechanism 606 may be a rotatable input device or a depressible and rotatable input device, for example. Input mechanism 608 may be a button, in some examples.

[0173] Input mechanism 608 may be a microphone, in some examples. Personal electronic device 600 can include various sensors, such as GPS sensor 632, accelerometer 634, directional sensor 640 (e.g., compass), gyroscope 636, motion sensor 638, and/or a combination thereof, all of which can be operatively connected to I/O section 614.

[0174] Memory 618 of personal electronic device 600 can be a non-transitory computer-readable storage medium, for storing computer-executable instructions, which, when executed by one or more computer processors 616, for example, can cause the computer processors to perform the techniques described below, including methods 1200, 1300, 1400, or 1500 (FIGS. 12-15). The computer-executable instructions can also be stored and/or transported within any non-transitory computer-readable storage medium for use by or in connection with an instruction execution system, apparatus, or device, such as a computer-based system, processor-containing system, or other system that can fetch the instructions from the instruction execution system, apparatus, or device and execute the instructions. For purposes of this document, a “non-transitory computer-readable storage medium” can be any medium that can tangibly contain or store computer-executable instructions for use by or in connection with the instruction execution system, apparatus, or device. The non-transitory computer-readable storage medium can include, but is not limited to, magnetic, optical, and/or semiconductor storages. Examples of such storage include magnetic disks, optical discs based on CD, DVD, or Blu-ray technologies, as well as persistent solid-state memory such as flash, solid-state drives, and the like. Personal electronic device 600 is not limited to the components and configuration of FIG. 6B, but can include other or additional components in multiple configurations.

[0175] As used here, the term “affordance” refers to a user-interactive graphical user interface object that may be displayed on the display screen of devices 200, 400, and/or 600 (FIGS. 2, 4, and 6). For example, an image (e.g., icon), a button, and text (e.g., hyperlink) may each constitute an affordance.

[0176] As used herein, the term “focus selector” refers to an input element that indicates a current part of a user interface with which a user is interacting. In some implementations that include a cursor or other location marker, the cursor acts as a “focus selector” so that when an input (e.g., a press input) is detected on a touch-sensitive surface (e.g., touchpad 455 in FIG. 4 or touch-sensitive surface 551 in FIG. 5B) while the cursor is over a particular user interface element (e.g., a button, window, slider or other user interface element), the particular user interface element is adjusted in accordance with the detected input. In some implementations that include a touch screen display (e.g., touch-sensitive display system 212 in FIG. 2A or touch screen 212 in FIG. 5A) that enables direct interaction with user interface elements on the touch screen display, a detected contact on the touch screen acts as a “focus selector” so that when an input (e.g., a press input by the contact) is detected on the touch screen display at a location of a particular user interface element (e.g., a button, window, slider, or other user interface element), the particular user interface element is adjusted in accordance with the detected input. In some implementations, focus is moved from one region of a user interface to another region of the user interface without corresponding movement of a cursor or movement of a contact on a touch screen display (e.g., by using a tab key or arrow keys to move focus from one button to another button); in these implementations, the focus selector moves in accordance with movement of focus between different regions of the user interface. Without regard to the specific form taken by the focus selector, the focus selector is generally the user interface element (or contact on a touch screen display) that is controlled by the user so as to communicate the user’s intended interaction with the user interface (e.g., by indicating, to the device, the element of the user interface with which the user is intending to interact). For example, the location of a focus selector (e.g., a cursor, a contact, or a selection box) over a respective button while a press input is detected on the touch-sensitive surface (e.g., a touchpad or touch screen) will indicate that the user is intending to activate the respective button (as opposed to other user interface elements shown on a display of the device).

[0177] As used in the specification and claims, the term “characteristic intensity” of a contact refers to a characteristic of the contact based on one or more intensities of the contact. In some embodiments, the characteristic intensity is based on multiple intensity samples. The characteristic intensity is, optionally, based on a predefined number of intensity samples, or a set of intensity samples collected during a predetermined time period (e.g., 0.05, 0.1, 0.2, 0.5, 1, 2, 5, 10 seconds) relative to a predefined event (e.g., after detecting the contact, prior to detecting liftoff of the contact, before or after detecting a start of movement of the contact, prior to detecting an end of the contact, before or after detecting an increase in intensity of the contact, and/or before or after detecting a decrease in intensity of the contact). A characteristic intensity of a contact is, optionally based on one or more of: a maximum value of the intensities of the contact, a mean value of the intensities of the contact, an average value of the intensities of the contact, a top 10 percentile value of the intensities of the contact, a value at the half maximum of the intensities of the contact, a value at the 90 percent maximum of the intensities of the contact, or the like. In some embodiments, the duration of the contact is used in determining the characteristic intensity (e.g., when the characteristic intensity is an average of the intensity of the contact over time). In some embodiments, the characteristic intensity is compared to a set of one or more intensity thresholds to determine whether an operation has been performed by a user. For example, the set of one or more intensity thresholds may include a first intensity threshold and a second intensity threshold. In this example, a contact with a characteristic intensity that does not exceed the first threshold results in a first operation, a contact with a characteristic intensity that exceeds the first intensity threshold and does not exceed the second intensity threshold results in a second operation, and a contact with a characteristic intensity that exceeds the second threshold results in a third operation. In some embodiments, a comparison between the characteristic intensity and one or more thresholds is used to determine whether or not to perform one or more operations (e.g., whether to perform a respective operation or forgo performing the respective operation) rather than being used to determine whether to perform a first operation or a second operation.

[0178] In some embodiments, a portion of a gesture is identified for purposes of determining a characteristic intensity. For example, a touch-sensitive surface may receive a continuous swipe contact transitioning from a start location and reaching an end location, at which point the intensity of the contact increases. In this example, the characteristic intensity of the contact at the end location may be based on only a portion of the continuous swipe

contact, and not the entire swipe contact (e.g., only the portion of the swipe contact at the end location). In some embodiments, a smoothing algorithm may be applied to the intensities of the swipe contact prior to determining the characteristic intensity of the contact. For example, the smoothing algorithm optionally includes one or more of: an unweighted sliding-average smoothing algorithm, a triangular smoothing algorithm, a median filter smoothing algorithm, and/or an exponential smoothing algorithm. In some circumstances, these smoothing algorithms eliminate narrow spikes or dips in the intensities of the swipe contact for purposes of determining a characteristic intensity.

[0179] The intensity of a contact on the touch-sensitive surface may be characterized relative to one or more intensity thresholds, such as a contact-detection intensity threshold, a light press intensity threshold, a deep press intensity threshold, and/or one or more other intensity thresholds. In some embodiments, the light press intensity threshold corresponds to an intensity at which the device will perform operations typically associated with clicking a button of a physical mouse or a trackpad. In some embodiments, the deep press intensity threshold corresponds to an intensity at which the device will perform operations that are different from operations typically associated with clicking a button of a physical mouse or a trackpad. In some embodiments, when a contact is detected with a characteristic intensity below the light press intensity threshold (e.g., and above a nominal contact-detection intensity threshold below which the contact is no longer detected), the device will move a focus selector in accordance with movement of the contact on the touch-sensitive surface without performing an operation associated with the light press intensity threshold or the deep press intensity threshold. Generally, unless otherwise stated, these intensity thresholds are consistent between different sets of user interface figures.

[0180] An increase of characteristic intensity of the contact from an intensity below the light press intensity threshold to an intensity between the light press intensity threshold and the deep press intensity threshold is sometimes referred to as a “light press” input. An increase of characteristic intensity of the contact from an intensity below the deep press intensity threshold to an intensity above the deep press intensity threshold is sometimes referred to as a “deep press” input. An increase of characteristic intensity of the contact from an intensity below the contact-detection intensity threshold to an intensity between the contact-detection intensity threshold and the light press intensity threshold is sometimes referred to as detecting the contact on the touch-surface. A decrease of characteristic

intensity of the contact from an intensity above the contact-detection intensity threshold to an intensity below the contact-detection intensity threshold is sometimes referred to as detecting liftoff of the contact from the touch-surface. In some embodiments, the contact-detection intensity threshold is zero. In some embodiments, the contact-detection intensity threshold is greater than zero.

[0181] In some embodiments described herein, one or more operations are performed in response to detecting a gesture that includes a respective press input or in response to detecting the respective press input performed with a respective contact (or a plurality of contacts), where the respective press input is detected based at least in part on detecting an increase in intensity of the contact (or plurality of contacts) above a press-input intensity threshold. In some embodiments, the respective operation is performed in response to detecting the increase in intensity of the respective contact above the press-input intensity threshold (e.g., a “down stroke” of the respective press input). In some embodiments, the press input includes an increase in intensity of the respective contact above the press-input intensity threshold and a subsequent decrease in intensity of the contact below the press-input intensity threshold, and the respective operation is performed in response to detecting the subsequent decrease in intensity of the respective contact below the press-input threshold (e.g., an “up stroke” of the respective press input).

[0182] In some embodiments, the device employs intensity hysteresis to avoid accidental inputs sometimes termed “jitter,” where the device defines or selects a hysteresis intensity threshold with a predefined relationship to the press-input intensity threshold (e.g., the hysteresis intensity threshold is X intensity units lower than the press-input intensity threshold or the hysteresis intensity threshold is 75%, 90%, or some reasonable proportion of the press-input intensity threshold). Thus, in some embodiments, the press input includes an increase in intensity of the respective contact above the press-input intensity threshold and a subsequent decrease in intensity of the contact below the hysteresis intensity threshold that corresponds to the press-input intensity threshold, and the respective operation is performed in response to detecting the subsequent decrease in intensity of the respective contact below the hysteresis intensity threshold (e.g., an “up stroke” of the respective press input). Similarly, in some embodiments, the press input is detected only when the device detects an increase in intensity of the contact from an intensity at or below the hysteresis intensity threshold to an intensity at or above the press-input intensity threshold and, optionally, a

subsequent decrease in intensity of the contact to an intensity at or below the hysteresis intensity, and the respective operation is performed in response to detecting the press input (e.g., the increase in intensity of the contact or the decrease in intensity of the contact, depending on the circumstances).

[0183] For ease of explanation, the descriptions of operations performed in response to a press input associated with a press-input intensity threshold or in response to a gesture including the press input are, optionally, triggered in response to detecting either: an increase in intensity of a contact above the press-input intensity threshold, an increase in intensity of a contact from an intensity below the hysteresis intensity threshold to an intensity above the press-input intensity threshold, a decrease in intensity of the contact below the press-input intensity threshold, and/or a decrease in intensity of the contact below the hysteresis intensity threshold corresponding to the press-input intensity threshold. Additionally, in examples where an operation is described as being performed in response to detecting a decrease in intensity of a contact below the press-input intensity threshold, the operation is, optionally, performed in response to detecting a decrease in intensity of the contact below a hysteresis intensity threshold corresponding to, and lower than, the press-input intensity threshold.

3. Digital Assistant System

[0184] FIG. 7A illustrates a block diagram of digital assistant system 700 in accordance with various examples. In some examples, digital assistant system 700 can be implemented on a standalone computer system. In some examples, digital assistant system 700 can be distributed across multiple computers. In some examples, some of the modules and functions of the digital assistant can be divided into a server portion and a client portion, where the client portion resides on one or more user devices (e.g., devices 104, 122, 200, 400, or 600) and communicates with the server portion (e.g., server system 108) through one or more networks, e.g., as shown in FIG. 1. In some examples, digital assistant system 700 can be an implementation of server system 108 (and/or DA server 106) shown in FIG. 1. It should be noted that digital assistant system 700 is only one example of a digital assistant system, and that digital assistant system 700 can have more or fewer components than shown, may combine two or more components, or may have a different configuration or arrangement of the components. The various components shown in FIG. 7A can be implemented in hardware, software instructions for execution by one or more processors, firmware, including

one or more signal processing and/or application specific integrated circuits, or a combination thereof.

[0185] Digital assistant system 700 can include memory 702, one or more processors 704, input/output (I/O) interface 706, and network communications interface 708. These components can communicate with one another over one or more communication buses or signal lines 710.

[0186] In some examples, memory 702 can include a non-transitory computer-readable medium, such as high-speed random access memory and/or a non-volatile computer-readable storage medium (e.g., one or more magnetic disk storage devices, flash memory devices, or other non-volatile solid-state memory devices).

[0187] In some examples, I/O interface 706 can couple input/output devices 716 of digital assistant system 700, such as displays, keyboards, touch screens, and microphones, to user interface module 722. I/O interface 706, in conjunction with user interface module 722, can receive user inputs (e.g., voice input, keyboard inputs, touch inputs, etc.) and processes them accordingly. In some examples, e.g., when the digital assistant is implemented on a standalone user device, digital assistant system 700 can include any of the components and I/O communication interfaces described with respect to devices 200, 400, or 600 in FIGS. 2A, 4, 6A-B, respectively. In some examples, digital assistant system 700 can represent the server portion of a digital assistant implementation, and can interact with the user through a client-side portion residing on a user device (e.g., devices 104, 200, 400, or 600).

[0188] In some examples, the network communications interface 708 can include wired communication port(s) 712 and/or wireless transmission and reception circuitry 714. The wired communication port(s) can receive and send communication signals via one or more wired interfaces, e.g., Ethernet, Universal Serial Bus (USB), FIREWIRE, etc. The wireless circuitry 714 can receive and send RF signals and/or optical signals from/to communications networks and other communications devices. The wireless communications can use any of a plurality of communications standards, protocols, and technologies, such as GSM, EDGE, CDMA, TDMA, Bluetooth, Wi-Fi, VoIP, Wi-MAX, or any other suitable communication protocol. Network communications interface 708 can enable communication between digital assistant system 700 with networks, such as the Internet, an intranet, and/or a wireless

network, such as a cellular telephone network, a wireless local area network (LAN), and/or a metropolitan area network (MAN), and other devices.

[0189] In some examples, memory 702, or the computer-readable storage media of memory 702, can store programs, modules, instructions, and data structures including all or a subset of: operating system 718, communications module 720, user interface module 722, one or more applications 724, and digital assistant module 726. In particular, memory 702, or the computer-readable storage media of memory 702, can store instructions for performing methods 1200, 1300, 1400, or 1500, described below. One or more processors 704 can execute these programs, modules, and instructions, and reads/writes from/to the data structures.

[0190] Operating system 718 (e.g., Darwin, RTXC, LINUX, UNIX, iOS, OS X, WINDOWS, or an embedded operating system such as VxWorks) can include various software components and/or drivers for controlling and managing general system tasks (e.g., memory management, storage device control, power management, etc.) and facilitates communications between various hardware, firmware, and software components.

[0191] Communications module 720 can facilitate communications between digital assistant system 700 with other devices over network communications interface 708. For example, communications module 720 can communicate with RF circuitry 208 of electronic devices such as devices 200, 400, and 600 shown in FIG. 2A, 4, 6A-B, respectively. Communications module 720 can also include various components for handling data received by wireless circuitry 714 and/or wired communications port 712.

[0192] User interface module 722 can receive commands and/or inputs from a user via I/O interface 706 (e.g., from a keyboard, touch screen, pointing device, controller, and/or microphone), and generate user interface objects on a display. User interface module 722 can also prepare and deliver outputs (e.g., speech, sound, animation, text, icons, vibrations, haptic feedback, light, etc.) to the user via the I/O interface 706 (e.g., through displays, audio channels, speakers, touch-pads, etc.).

[0193] Applications 724 can include programs and/or modules that are configured to be executed by one or more processors 704. For example, if the digital assistant system is implemented on a standalone user device, applications 724 can include user applications, such as games, a calendar application, a navigation application, or an email application. If

digital assistant system 700 is implemented on a server, applications 724 can include resource management applications, diagnostic applications, or scheduling applications, for example.

[0194] Memory 702 can also store digital assistant module 726 (or the server portion of a digital assistant). In some examples, digital assistant module 726 can include the following sub-modules, or a subset or superset thereof: input/output processing module 728, speech-to-text (STT) processing module 730, natural language processing module 732, dialogue flow processing module 734, task flow processing module 736, service processing module 738, and speech synthesis module 740. Each of these modules can have access to one or more of the following systems or data and models of the digital assistant module 726, or a subset or superset thereof: ontology 760, vocabulary index 744, user data 748, task flow models 754, service models 756, and ASR systems.

[0195] In some examples, using the processing modules, data, and models implemented in digital assistant module 726, the digital assistant can perform at least some of the following: converting speech input into text; identifying a user's intent expressed in a natural language input received from the user; actively eliciting and obtaining information needed to fully infer the user's intent (e.g., by disambiguating words, games, intentions, etc.); determining the task flow for fulfilling the inferred intent; and executing the task flow to fulfill the inferred intent.

[0196] In some examples, as shown in FIG. 7B, I/O processing module 728 can interact with the user through I/O devices 716 in FIG. 7A or with a user device (e.g., devices 104, 200, 400, or 600) through network communications interface 708 in FIG. 7A to obtain user input (e.g., a speech input) and to provide responses (e.g., as speech outputs) to the user input. I/O processing module 728 can optionally obtain contextual information associated with the user input from the user device, along with or shortly after the receipt of the user input. The contextual information can include user-specific data, vocabulary, and/or preferences relevant to the user input. In some examples, the contextual information also includes software and hardware states of the user device at the time the user request is received, and/or information related to the surrounding environment of the user at the time that the user request was received. In some examples, I/O processing module 728 can also send follow-up questions to, and receive answers from, the user regarding the user request. When a user request is received by I/O processing module 728 and the user request can

include speech input, I/O processing module 728 can forward the speech input to STT processing module 730 (or speech recognizer) for speech-to-text conversions.

[0197] STT processing module 730 can include one or more ASR systems. The one or more ASR systems can process the speech input that is received through I/O processing module 728 to produce a recognition result. Each ASR system can include a front-end speech pre-processor. The front-end speech pre-processor can extract representative features from the speech input. For example, the front-end speech pre-processor can perform a Fourier transform on the speech input to extract spectral features that characterize the speech input as a sequence of representative multi-dimensional vectors. Further, each ASR system can include one or more speech recognition models (e.g., acoustic models and/or language models) and can implement one or more speech recognition engines. Examples of speech recognition models can include Hidden Markov Models, Gaussian-Mixture Models, Deep Neural Network Models, n-gram language models, and other statistical models. Examples of speech recognition engines can include the dynamic time warping based engines and weighted finite-state transducers (WFST) based engines. The one or more speech recognition models and the one or more speech recognition engines can be used to process the extracted representative features of the front-end speech pre-processor to produce intermediate recognitions results (e.g., phonemes, phonemic strings, and sub-words), and ultimately, text recognition results (e.g., words, word strings, or sequence of tokens). In some examples, the speech input can be processed at least partially by a third-party service or on the user's device (e.g., device 104, 200, 400, or 600) to produce the recognition result. Once STT processing module 730 produces recognition results containing a text string (e.g., words, or sequence of words, or sequence of tokens), the recognition result can be passed to natural language processing module 732 for intent deduction.

[0198] More details on the speech-to-text processing are described in U.S. Utility Application Serial No. 13/236,942 for "Consolidating Speech Recognition Results," filed on September 20, 2011, the entire disclosure of which is incorporated herein by reference.

[0199] In some examples, STT processing module 730 can include and/or access a vocabulary of recognizable words via phonetic alphabet conversion module 731. Each vocabulary word can be associated with one or more candidate pronunciations of the word represented in a speech recognition phonetic alphabet. In particular, the vocabulary of recognizable words can include a word that is associated with a plurality of candidate

pronunciations. For example, the vocabulary may include the word “tomato” that is associated with the candidate pronunciations of /tə'meɪrou/ and /tə'matou/. Further, vocabulary words can be associated with custom candidate pronunciations that are based on previous speech inputs from the user. Such custom candidate pronunciations can be stored in STT processing module 730 and can be associated with a particular user via the user's profile on the device. In some examples, the candidate pronunciations for words can be determined based on the spelling of the word and one or more linguistic and/or phonetic rules. In some examples, the candidate pronunciations can be manually generated, e.g., based on known canonical pronunciations.

[0200] In some examples, the candidate pronunciations can be ranked based on the commonness of the candidate pronunciation. For example, the candidate pronunciation /tə'meɪrou/ can be ranked higher than /tə'matou/, because the former is a more commonly used pronunciation (e.g., among all users, for users in a particular geographical region, or for any other appropriate subset of users). In some examples, candidate pronunciations can be ranked based on whether the candidate pronunciation is a custom candidate pronunciation associated with the user. For example, custom candidate pronunciations can be ranked higher than canonical candidate pronunciations. This can be useful for recognizing proper nouns having a unique pronunciation that deviates from canonical pronunciation. In some examples, candidate pronunciations can be associated with one or more speech characteristics, such as geographic origin, nationality, or ethnicity. For example, the candidate pronunciation /tə'meɪrou/ can be associated with the United States, whereas the candidate pronunciation /tə'matou/ can be associated with Great Britain. Further, the rank of the candidate pronunciation can be based on one or more characteristics (e.g., geographic origin, nationality, ethnicity, etc.) of the user stored in the user's profile on the device. For example, it can be determined from the user's profile that the user is associated with the United States. Based on the user being associated with the United States, the candidate pronunciation /tə'meɪrou/ (associated with the United States) can be ranked higher than the candidate pronunciation /tə'matou/ (associated with Great Britain). In some examples, one of the ranked candidate pronunciations can be selected as a predicted pronunciation (e.g., the most likely pronunciation).

[0201] When a speech input is received, STT processing module 730 can be used to determine the phonemes corresponding to the speech input (e.g., using an acoustic model),

and then attempt to determine words that match the phonemes (e.g., using a language model). For example, if STT processing module 730 can first identify the sequence of phonemes /tə'meɪrou/ corresponding to a portion of the speech input, it can then determine, based on vocabulary index 744, that this sequence corresponds to the word “tomato.”

[0202] In some examples, STT processing module 730 can use approximate matching techniques to determine words in an utterance. Thus, for example, the STT processing module 730 can determine that the sequence of phonemes /tə'meɪrou/ corresponds to the word “tomato,” even if that particular sequence of phonemes is not one of the candidate sequence of phonemes for that word.

[0203] Natural language processing module 732 (“natural language processor”) of the digital assistant can take the sequence of words or tokens (“token sequence”) generated by STT processing module 730, and attempt to associate the token sequence with one or more “actionable intents” recognized by the digital assistant. An “actionable intent” can represent a task that can be performed by the digital assistant, and can have an associated task flow implemented in task flow models 754. The associated task flow can be a series of programmed actions and steps that the digital assistant takes in order to perform the task. The scope of a digital assistant’s capabilities can be dependent on the number and variety of task flows that have been implemented and stored in task flow models 754, or in other words, on the number and variety of “actionable intents” that the digital assistant recognizes. The effectiveness of the digital assistant, however, can also be dependent on the assistant’s ability to infer the correct “actionable intent(s)” from the user request expressed in natural language.

[0204] In some examples, in addition to the sequence of words or tokens obtained from STT processing module 730, natural language processing module 732 can also receive contextual information associated with the user request, e.g., from I/O processing module 728. The natural language processing module 732 can optionally use the contextual information to clarify, supplement, and/or further define the information contained in the token sequence received from STT processing module 730. The contextual information can include, for example, user preferences, hardware, and/or software states of the user device, sensor information collected before, during, or shortly after the user request, prior interactions (e.g., dialogue) between the digital assistant and the user, and the like. As described herein, contextual information can be dynamic, and can change with time, location, content of the dialogue, and other factors.

[0205] In some examples, the natural language processing can be based on, e.g., ontology 760. Ontology 760 can be a hierarchical structure containing many nodes, each node representing either an “actionable intent” or a “property” relevant to one or more of the “actionable intents” or other “properties.” As noted above, an “actionable intent” can represent a task that the digital assistant is capable of performing, i.e., it is “actionable” or can be acted on. A “property” can represent a parameter associated with an actionable intent or a sub-aspect of another property. A linkage between an actionable intent node and a property node in ontology 760 can define how a parameter represented by the property node pertains to the task represented by the actionable intent node.

[0206] In some examples, ontology 760 can be made up of actionable intent nodes and property nodes. Within ontology 760, each actionable intent node can be linked to one or more property nodes either directly or through one or more intermediate property nodes. Similarly, each property node can be linked to one or more actionable intent nodes either directly or through one or more intermediate property nodes. For example, as shown in FIG. 7C, ontology 760 can include a “restaurant reservation” node (i.e., an actionable intent node). Property nodes “restaurant,” “date/time” (for the reservation), and “party size” can each be directly linked to the actionable intent node (i.e., the “restaurant reservation” node).

[0207] In addition, property nodes “cuisine,” “price range,” “phone number,” and “location” can be sub-nodes of the property node “restaurant,” and can each be linked to the “restaurant reservation” node (i.e., the actionable intent node) through the intermediate property node “restaurant.” For another example, as shown in FIG. 7C, ontology 760 can also include a “set reminder” node (i.e., another actionable intent node). Property nodes “date/time” (for setting the reminder) and “subject” (for the reminder) can each be linked to the “set reminder” node. Since the property “date/time” can be relevant to both the task of making a restaurant reservation and the task of setting a reminder, the property node “date/time” can be linked to both the “restaurant reservation” node and the “set reminder” node in ontology 760.

[0208] An actionable intent node, along with its linked concept nodes, can be described as a “domain.” In the present discussion, each domain can be associated with a respective actionable intent, and refers to the group of nodes (and the relationships there between) associated with the particular actionable intent. For example, ontology 760 shown in FIG. 7C can include an example of restaurant reservation domain 762 and an example of reminder

domain 764 within ontology 760. The restaurant reservation domain includes the actionable intent node “restaurant reservation,” property nodes “restaurant,” “date/time,” and “party size,” and sub-property nodes “cuisine,” “price range,” “phone number,” and “location.” Reminder domain 764 can include the actionable intent node “set reminder,” and property nodes “subject” and “date/time.” In some examples, ontology 760 can be made up of many domains. Each domain can share one or more property nodes with one or more other domains. For example, the “date/time” property node can be associated with many different domains (e.g., a scheduling domain, a travel reservation domain, a movie ticket domain, etc.), in addition to restaurant reservation domain 762 and reminder domain 764.

[0209] While FIG. 7C illustrates two example domains within ontology 760, other domains can include, for example, “find a movie,” “initiate a phone call,” “find directions,” “schedule a meeting,” “send a message,” and “provide an answer to a question,” “read a list,” “providing navigation instructions,” “provide instructions for a task” and so on. A “send a message” domain can be associated with a “send a message” actionable intent node, and may further include property nodes such as “recipient(s),” “message type,” and “message body.” The property node “recipient” can be further defined, for example, by the sub-property nodes such as “recipient name” and “message address.”

[0210] In some examples, ontology 760 can include all the domains (and hence actionable intents) that the digital assistant is capable of understanding and acting upon. In some examples, ontology 760 can be modified, such as by adding or removing entire domains or nodes, or by modifying relationships between the nodes within the ontology 760.

[0211] In some examples, nodes associated with multiple related actionable intents can be clustered under a “super domain” in ontology 760. For example, a “travel” super-domain can include a cluster of property nodes and actionable intent nodes related to travel. The actionable intent nodes related to travel can include “airline reservation,” “hotel reservation,” “car rental,” “get directions,” “find points of interest,” and so on. The actionable intent nodes under the same super domain (e.g., the “travel” super domain) can have many property nodes in common. For example, the actionable intent nodes for “airline reservation,” “hotel reservation,” “car rental,” “get directions,” and “find points of interest” can share one or more of the property nodes “start location,” “destination,” “departure date/time,” “arrival date/time,” and “party size.”

[0212] In some examples, each node in ontology 760 can be associated with a set of words and/or phrases that are relevant to the property or actionable intent represented by the node. The respective set of words and/or phrases associated with each node can be the so-called “vocabulary” associated with the node. The respective set of words and/or phrases associated with each node can be stored in vocabulary index 744 in association with the property or actionable intent represented by the node. For example, returning to FIG. 7B, the vocabulary associated with the node for the property of “restaurant” can include words such as “food,” “drinks,” “cuisine,” “hungry,” “eat,” “pizza,” “fast food,” “meal,” and so on. For another example, the vocabulary associated with the node for the actionable intent of “initiate a phone call” can include words and phrases such as “call,” “phone,” “dial,” “ring,” “call this number,” “make a call to,” and so on. The vocabulary index 744 can optionally include words and phrases in different languages.

[0213] Natural language processing module 732 can receive the token sequence (e.g., a text string) from STT processing module 730, and determine what nodes are implicated by the words in the token sequence. In some examples, if a word or phrase in the token sequence is found to be associated with one or more nodes in ontology 760 (via vocabulary index 744), the word or phrase can “trigger” or “activate” those nodes. Based on the quantity and/or relative importance of the activated nodes, natural language processing module 732 can select one of the actionable intents as the task that the user intended the digital assistant to perform. In some examples, the domain that has the most “triggered” nodes can be selected. In some examples, the domain having the highest confidence value (e.g., based on the relative importance of its various triggered nodes) can be selected. In some examples, the domain can be selected based on a combination of the number and the importance of the triggered nodes. In some examples, additional factors are considered in selecting the node as well, such as whether the digital assistant has previously correctly interpreted a similar request from a user.

[0214] User data 748 can include user-specific information, such as user-specific vocabulary, user preferences, user address, user’s default and secondary languages, user’s contact list, and other short-term or long-term information for each user. In some examples, natural language processing module 732 can use the user-specific information to supplement the information contained in the user input to further define the user intent. For example, for a user request “invite my friends to my birthday party,” natural language processing module

732 can be able to access user data 748 to determine who the “friends” are and when and where the “birthday party” would be held, rather than requiring the user to provide such information explicitly in his/her request.

[0215] Other details of searching an ontology based on a token string is described in U.S. Utility Application Serial No. 12/341,743 for “Method and Apparatus for Searching Using An Active Ontology,” filed December 22, 2008, the entire disclosure of which is incorporated herein by reference.

[0216] In some examples, once natural language processing module 732 identifies an actionable intent (or domain) based on the user request, natural language processing module 732 can generate a structured query to represent the identified actionable intent. In some examples, the structured query can include parameters for one or more nodes within the domain for the actionable intent, and at least some of the parameters are populated with the specific information and requirements specified in the user request. For example, the user may say “Make me a dinner reservation at a sushi place at 7.” In this case, natural language processing module 732 can be able to correctly identify the actionable intent to be “restaurant reservation” based on the user input. According to the ontology, a structured query for a “restaurant reservation” domain may include parameters such as {Cuisine}, {Time}, {Date}, {Party Size}, and the like. In some examples, based on the speech input and the text derived from the speech input using STT processing module 730, natural language processing module 732 can generate a partial structured query for the restaurant reservation domain, where the partial structured query includes the parameters {Cuisine = “Sushi”} and {Time = “7pm”}. However, in this example, the user’s utterance contains insufficient information to complete the structured query associated with the domain. Therefore, other necessary parameters such as {Party Size} and {Date} may not be specified in the structured query based on the information currently available. In some examples, natural language processing module 732 can populate some parameters of the structured query with received contextual information. For example, in some examples, if the user requested a sushi restaurant “near me,” natural language processing module 732 can populate a {location} parameter in the structured query with GPS coordinates from the user device.

[0217] In some examples, natural language processing module 732 can pass the generated structured query (including any completed parameters) to task flow processing module 736 (“task flow processor”). Task flow processing module 736 can be configured to receive the

structured query from natural language processing module 732, complete the structured query, if necessary, and perform the actions required to “complete” the user’s ultimate request. In some examples, the various procedures necessary to complete these tasks can be provided in task flow models 754. In some examples, task flow models 754 can include procedures for obtaining additional information from the user and task flows for performing actions associated with the actionable intent.

[0218] As described above, in order to complete a structured query, task flow processing module 736 may need to initiate additional dialogue with the user in order to obtain additional information, and/or disambiguate potentially ambiguous utterances. When such interactions are necessary, task flow processing module 736 can invoke dialogue flow processing module 734 to engage in a dialogue with the user. In some examples, dialogue flow processing module 734 can determine how (and/or when) to ask the user for the additional information and receives and processes the user responses. The questions can be provided to and answers can be received from the users through I/O processing module 728. In some examples, dialogue flow processing module 734 can present dialogue output to the user via audio and/or visual output, and receives input from the user via spoken or physical (e.g., clicking) responses. Continuing with the example above, when task flow processing module 736 invokes dialogue flow processing module 734 to determine the “party size” and “date” information for the structured query associated with the domain “restaurant reservation,” dialogue flow processing module 734 can generate questions such as “For how many people?” and “On which day?” to pass to the user. Once answers are received from the user, dialogue flow processing module 734 can then populate the structured query with the missing information, or pass the information to task flow processing module 736 to complete the missing information from the structured query.

[0219] Once task flow processing module 736 has completed the structured query for an actionable intent, task flow processing module 736 can proceed to perform the ultimate task associated with the actionable intent. Accordingly, task flow processing module 736 can execute the steps and instructions in the task flow model according to the specific parameters contained in the structured query. For example, the task flow model for the actionable intent of “restaurant reservation” can include steps and instructions for contacting a restaurant and actually requesting a reservation for a particular party size at a particular time. For example, using a structured query such as: {restaurant reservation, restaurant = ABC Café, date =

3/12/2012, time = 7pm, party size = 5}, task flow processing module 736 can perform the steps of: (1) logging onto a server of the ABC Café or a restaurant reservation system such as OPENTABLE®, (2) entering the date, time, and party size information in a form on the website, (3) submitting the form, and (4) making a calendar entry for the reservation in the user's calendar.

[0220] In some examples, task flow processing module 736 can employ the assistance of service processing module 738 ("service processing module") to complete a task requested in the user input or to provide an informational answer requested in the user input. For example, service processing module 738 can act on behalf of task flow processing module 736 to make a phone call, set a calendar entry, invoke a map search, invoke or interact with other user applications installed on the user device, and invoke or interact with third-party services (e.g., a restaurant reservation portal, a social networking website, a banking portal, etc.). In some examples, the protocols and application programming interfaces (API) required by each service can be specified by a respective service model among service models 756. Service processing module 738 can access the appropriate service model for a service and generate requests for the service in accordance with the protocols and APIs required by the service according to the service model.

[0221] For example, if a restaurant has enabled an online reservation service, the restaurant can submit a service model specifying the necessary parameters for making a reservation and the APIs for communicating the values of the necessary parameter to the online reservation service. When requested by task flow processing module 736, service processing module 738 can establish a network connection with the online reservation service using the web address stored in the service model, and send the necessary parameters of the reservation (e.g., time, date, party size) to the online reservation interface in a format according to the API of the online reservation service.

[0222] In some examples, natural language processing module 732, dialogue flow processing module 734, and task flow processing module 736 can be used collectively and iteratively to infer and define the user's intent, obtain information to further clarify and refine the user intent, and finally generate a response (i.e., an output to the user, or the completion of a task) to fulfill the user's intent. The generated response can be a dialogue response to the speech input that at least partially fulfills the user's intent. Further, in some examples, the generated response can be output as a speech output. In these examples, the generated

response can be sent to speech synthesis module 740 (e.g., speech synthesizer) where it can be processed to synthesize the dialogue response in speech form. In yet other examples, the generated response can be data content relevant to satisfying a user request in the speech input.

[0223] Speech synthesis module 740 can be configured to synthesize speech outputs for presentation to the user. Speech synthesis module 740 synthesizes speech outputs based on text provided by the digital assistant. For example, the generated dialogue response can be in the form of a text string. Speech synthesis module 740 can convert the text string to an audible speech output. Speech synthesis module 740 can use any appropriate speech synthesis technique in order to generate speech outputs from text, including, but not limited, to concatenative synthesis, unit selection synthesis, diphone synthesis, domain-specific synthesis, formant synthesis, articulatory synthesis, hidden Markov model (HMM) based synthesis, and sinewave synthesis. In some examples, speech synthesis module 740 can be configured to synthesize individual words based on phonemic strings corresponding to the words. For example, a phonemic string can be associated with a word in the generated dialogue response. The phonemic string can be stored in metadata associated with the word. Speech synthesis model 740 can be configured to directly process the phonemic string in the metadata to synthesize the word in speech form.

[0224] In some examples, instead of (or in addition to) using speech synthesis module 740, speech synthesis can be performed on a remote device (e.g., the server system 108), and the synthesized speech can be sent to the user device for output to the user. For example, this can occur in some implementations where outputs for a digital assistant are generated at a server system. And because server systems generally have more processing power or resources than a user device, it can be possible to obtain higher quality speech outputs than would be practical with client-side synthesis.

[0225] Additional details on digital assistants can be found in the U.S. Utility Application No. 12/987,982, entitled “Intelligent Automated Assistant,” filed January 10, 2011, and U.S. Utility Application No. 13/251,088, entitled “Generating and Processing Task Items That Represent Tasks to Perform,” filed September 30, 2011, the entire disclosures of which are incorporated herein by reference.

[0226] Attention is now directed towards embodiments of techniques and associated user interfaces (“UI”) for triggering (e.g., initiating) a virtual assistant that may be implemented on an electronic device with a display and a touch-sensitive surface.

4. Exemplary techniques for triggering a virtual assistant

[0227] FIGS. 8A-8C illustrate exemplary techniques for triggering a virtual assistant based on whether input at an electronic device meets a predetermined condition, in accordance with some embodiments. These figures are also used to illustrate the processes described below, including the processes in FIGS. 12.

[0228] FIG. 8A illustrates electronic device 800 of user 802. Device 800 is one of devices 104, 122, 200, 400, 600, and 1700 (FIGS. 1, 2, 4, 6, and 17) in some embodiments. Device 800 has touch screen 804. In the illustrated (e.g., lowered) position, touch screen 804 is powered off to conserve battery usage.

[0229] FIG. 8B illustrates user 802 lifting electronic device 800 into a raised position for viewing. In response to the movement, device 800 powers on touch screen 804 to present a user interface. Techniques for determining device movement such as the lifting movement seen in FIGS. 8A and 8B are further described with reference to FIG. 12, below.

[0230] In accordance with touch screen 804 powering on, device 800 also begins to sample audio input through its microphone to listen for spoken input from user 802. FIG. 8C illustrates user 802 providing spoken instructions 810 while device 800 is sampling audio input. As device 800 detects spoken instruction 810 while sampling audio input, the device determines whether the sampled audio input contains a phrase for triggering a virtual assistant service. If device 800 determines that the sampled input contains a spoken trigger for triggering a virtual assistant, device 800 triggers the virtual assistant. If device 800 determines that the sampled audio input does not contain the spoken trigger, it may continue to sample audio input for the spoken trigger for a limited time.

[0231] In the illustrated example, the spoken trigger is the phrase of “Hey Siri.” Thus, spoken instruction 810, when sampled by device 800, causes device 800 to trigger a virtual assistant service. However, instruction 810 does not contain additional language representing a user request, beyond the trigger phrase. Thus, device 800 awaits further spoken input upon trigger the virtual assistant service by continuing to sample audio input. While awaiting

additional input, device 800 optionally displays indication 812 prompting user 802 for further instructions.

[0232] In this way, device 800 activates the microphone, associated circuitry, and corresponding software processes for sampling audio input for a virtual assistant based on whether the device has detected a particular usage condition. This technique allows device 800 to disable certain electronic circuitry (e.g., microphone circuitry) and/or reduce execution of computer instructions (e.g., an associated processor power consumption) at times when a user is less likely to be providing active, cognitive input to the device—such as when the device is lowered and its display is powered off—thereby reducing overall power consumption by the device. Restated, the technical benefit of battery conservation is achieved through the notion that a user is less likely to provide spoken input while the device is outside certain usage conditions, in some embodiments.

[0233] FIGS. 9A-9D illustrate exemplary techniques for triggering a virtual assistant session in accordance with some embodiments. These figures are also used to illustrate the processes described below, including the processes in FIG. 13.

[0234] FIG. 9A illustrates electronic device 900 of user 902. Device 900 is one of devices 104, 122, 200, 400, 600, and 1600 (FIGS. 1, 2, 4, 6, and 16) in some embodiments. Device 900 has touch screen 904. Touch screen 904 is displaying notification 906 notifying user 902 of an incoming message just received by device 900. The display of notification 906 is triggered by one or more software events associated with the incoming message. In accordance with detecting one or more relevant software events, device 900 begins to sample audio input via its microphone to identify spoken input. Techniques for determining relevant software events such as message notifications as seen in FIGS. 9A and 9B are further described with reference to FIG. 13, below.

[0235] FIG. 9B illustrates user 902 providing spoken instructions 910 while device 900 is sampling audio input. When device 900 detects spoken instruction 910 in accordance with its sampling of audio input, the device determines whether the sampled input contains a spoken trigger for a virtual assistant. In the illustrated example, the spoken trigger is the phrase of “Hey Siri.” If device 900 determines that the sampled input contains a spoken trigger for triggering a virtual assistant, device 900 triggers the virtual assistant. If device 900

determines that the sampled audio input does not contain the spoken trigger, it may continue to sample audio input for the spoken trigger for a limited time.

[0236] In the illustrated example, spoken instructions 910 contains both a spoken trigger (e.g., “hey Siri”) and natural language input (e.g., “tell the sender I can’t chat right now”) that the virtual assistant can operationalize into computing tasks. For example, the task is to send a brief reply to the sender of incoming message 912. Turning to FIG. 9C, responsive to the sampling of audio input containing instructions 910, electronic device 900 initiates and transmits message 916 to sender 918 of incoming message 912, informing the sender that user 902 is currently unavailable.

[0237] This technique of powering-on the microphone of device 900 in accordance with software events detected by the device, as illustrated by FIGS. 9A-9C, allows electronic device 900 to disable certain electronic circuitry (e.g., microphone circuitry) and/or reduce execution of computer instructions (e.g., an associated processor power consumption) at times when a user is less likely to be providing active, cognitive input to the device, thereby reducing overall power consumption by the device. Restated, the technical benefit of battery conservation is achieved through the notion that a user is less likely to provide spoken input while the device is idle from software events that cause output to the user, in some embodiments.

[0238] FIGS. 10A-10D illustrate exemplary techniques for triggering a virtual assistant session in accordance with some embodiments. These figures are also used to illustrate the processes described below, including the processes in FIG. 14.

[0239] FIG. 10A illustrates electronic device 1000 of user 1002. Device 1000 is one of 104, 122, 200, 400, 600, and 1700 (FIGS. 1, 2, 4, 6, and 17) in some embodiments. Device 1000 has touch screen 1004. In its illustrated position (e.g., lowered), touch screen 1004 is powered off to conserve the battery life of device 1000. FIG. 10B illustrates user 1002 lifting electronic device 1000 into a raised position for viewing. In response to the movement, device 1000 begins to sample audio input for spoken input from user 1002. In some embodiments, device 1000 begins this sampling of audio input even though touch screen 1004 remains off. In some embodiments, device 1000 begins this sampling of audio input regardless of whether touch screen 1004 is on or off.

[0240] FIG. 10C illustrates user 1002 providing spoken instructions 1010 while device 1000 is sampling audio input with touch screen 1004 off. When device 1000 detects spoken instruction 1010 in accordance of its sampling of audio input, the device determines whether the sampled input contains a spoken trigger for a virtual assistant. In the illustrated example, the spoken trigger is the phrase of “Hey Siri.” If device 1000 determines that the sampled input contains a spoken trigger for triggering a virtual assistant, device 1000 triggers the virtual assistant without powering on touch screen 1004, if the screen is powered off. If device 1000 determines that the sampled audio input does not contain the spoken trigger, it may continue to sample audio input for the spoken trigger for some time, as further described with reference to FIG. 14, below.

[0241] In the illustrated example of FIG. 10C, the sampled audio input contains a spoken trigger (e.g., “hey Siri”) and a natural language input (e.g., “download my e-mail”) that the virtual assistant can operationalize into tasks. In some embodiments, device 1000 provides output (e.g., haptic and/or audio output) acknowledging the received spoken input. In some embodiments, the acknowledgement output is provided without turning on touch screen 1004.

[0242] Turning to FIG. 10D, responsive to spoken instructions 1010, device 1000 initiates the requested task of downloading e-mail. In some embodiments, the task is performed while touch screen 1004 remains powered off. In some embodiments, device 1000 provides output (e.g., haptic and/or audio output) indicating to the user that the spoken input is being performed and/or has been performed. In some embodiments, the output is provided without turning on touch screen 1004.

[0243] This technique of powering-on a device’s microphone to process a user’s spoken input under some usage conditions while the touch screen remains off, described with reference to FIGS. 10A-10D, permits reductions in overall power consumption by device 1000, while maintaining the device’s ability to process a user’s spoken input. Restated, the technical benefit of battery conservation is achieved through efficient controls on the activation of the display on a display-equipped device, in some embodiments.

[0244] FIGS. 11A-11C illustrate exemplary techniques for triggering a virtual assistant session in accordance with some embodiments. These figures are also used to illustrate the processes described below, including the processes in FIGS. 15.

[0245] FIG. 11A illustrates electronic device 1100 of user 1102. Device 1100 is one of 104, 122, 200, 400, 600, and 1800 (FIGS. 1, 2, 4, 6, and 18) in some embodiments. Device 1100 has touch screen 1104. Touch screen 1104 is displaying indication 1106 indicating that a virtual assistant session is active and is awaiting the user's spoken instructions. While the virtual assistant session is active, a microphone of device 1100 is sampling audio input for user instructions. Other user interfaces (not shown) may also indicate that a virtual assistant session is active, such as a virtual assistant transcript listing the previous user inputs and virtual assistant responses. Touch screen 1104 also displays includes cancel affordance 1108 for ending the virtual assistant session.

[0246] As seen in FIG. 11B, responsive to detecting user selection of cancel affordance 1108, device 1100 reports to the user that the virtual assistant session has ended by replacing virtual assistant indication 1106 with a non-virtual assistant user interface. In the illustrated example, device 1100 displays digital clock face 1110. In some embodiments (not expressly illustrated in FIG. 11B), device 1100 powers off touch screen 1104 in acknowledging the user's cancellation of the virtual assistant session.

[0247] In some embodiments, however, even as device 1100 indicates to the user that the virtual assistant session has ended, its microphone continues to sample audio input for the user's spoken instructions for some time. In some embodiments, the on-going audio sampling occurs for a predetermined amount of time (e.g., eight seconds). In some embodiments, the on-going audio sampling occurs until an additional user input is received, such as a touch input on touch screen 1104, activation of an application, so forth. Additional techniques for identifying a duration for which the sampling of audio input continues are described with reference to FIG. 15, below.

[0248] Turning to FIG. 11C, during its on-going sampling of audio input, device 1100 triggers the virtual assistant responsive to sampled audio input that includes a spoken trigger, such as spoken input 1112 which includes the phrase "hey Siri."

[0249] This technique of continuing to sample audio input for a time after a user has ended a virtual assistant (as opposed to turning off the microphone immediately) is beneficial. Consider, for instance, the situation in which a user's cancellation of the virtual assistant (e.g., at FIG. 11B) is erroneous, and has the effect of placing the device into a state where it is would have placed awaiting an intermediate input such as device movement before sampling

audio input again. That is, the user's erroneous cancellation would have precluded the device's ability to process the user's immediate subsequent spoken input. Under this situation, the above-described techniques permit the device to respond as if the erroneous user input did not occur.

[0250] As another example, consider the situation in which a user's cancellation of the virtual assistant would have delayed the device's ability to process the user's subsequent spoken input by requiring the device to power-cycle its microphone and associated circuitry. The above-described techniques permit electronic device 1100 to respond to the user while avoiding the delay caused by power cycle, thereby improving response times. These benefits impart a higher-quality user experience to device 1100, and are particularly helpful for portable devices where design compromises are made between processing power and battery consumption.

[0251] FIGS. 20A-20F illustrate exemplary techniques for handling a user's request to a virtual assistant, in accordance with some embodiments. These figures are also used to illustrate the processes described below, including the processes in FIG. 21.

[0252] FIG. 20A illustrates electronic device 2000 of user 2002. Device 2000 is one of 104, 122, 200, 400, 600, and 1900 (FIGS. 1, 2, 4, 6, and 19) in some embodiments. Device 2000 has touch screen 2004. As depicted in FIG. 20A, device 2000 is displaying a virtual assistant user interface 2006, indicating that it is sampling audio input for user instructions 2010, which instructs the device to initiate a message to a recipient.

[0253] Turning to FIG. 20B, device 2000 begins to process spoken instructions 2010 to identify computing tasks that can be performed to fulfill the user's instructions. The amount of time that is required to operationalize a user's instruction varies under different usage scenarios. For example, a slow data service provider may slow down the device's response time.

[0254] As seen in FIG. 20C, while the virtual assistant is processing the user's instructions, user 2002 may lower device 2000 into a non-viewing position. In the absence of user input, touch screen 2004 turns off after some time to conserve battery power, even as virtual assistant continues to process user instructions 2010, in the illustrated example.

[0255] Turning to FIG. 20D, when device 2000 has operationalized the user's instructions to a point where output should be displayed to the user, device 200 provides a haptic and/or audible output such that the user is notified that additional information is now available. For example, in response to user instructions 2010, device 2000 may display a confirmation after it has launched a messaging application and populated an outgoing message, but before sending the message to the recipient.

[0256] As seen in FIG. 20E, alerted by the haptic and/or audible output, user 2002 may raise device 2000 into viewing position, which powers on touch screen 2004 to show message 2012 addressed to recipient "Jen" and containing message body 2012, both of which correspond to the user's natural language input 2010.

[0257] Although the examples of FIGS. 20A-20E depict a user's instructions to create a message, the techniques described are not so limited. The virtual assistant of device 2000 may use a haptic and/or audible alert to inform the user whenever the virtual assistant has obtained useful results while the device's screen is off. For instance, device 2000 may vibrate to alert the user that navigation directions have been downloaded and are available for viewing. In this way, device 2000 may relieve the user from having to focus on the device while the user's input is being processed. These techniques also permit device 2000 to power down its touch screen momentarily to conserve battery power while its virtual assistant continues to perform tasks.

6. Exemplary processes for triggering a virtual assistant

[0258] FIGS. 12 is a flow diagram illustrating method 1200 of triggering a virtual assistant on an electronic device in accordance with some embodiments. Method 1200 is performed at one of devices 104, 200, 400, 800, and 1600 (FIGS. 1, 2, 4, 8, and 16) in some embodiments. In some embodiments, method 1200 is performed at a device having a touch screen. In some embodiments, method 1200 is performed at a device with a separate display screen and a touch-sensitive surface. Operations in method 1200 are, optionally, combined and/or the order of some operations is, optionally, changed.

[0259] At block 1202, while the electronic device's display is on, the device detects user input via an input component of the device. At block 1204, the device determines if the input meets a predetermined condition. If the input meets a predetermined condition, processing proceeds to 1206 where the device samples audio input received via its microphone. If the

input does not meet the predetermined condition, processing returns to block 1202, where the device may detect another user input.

[0260] In some embodiments, the predetermined condition is that the display screen of the device is on. The device determines whether the display is on, for example, by determining whether its backlight is lit.

[0261] In some embodiments, the predetermined condition is a movement of the device, such as a lifting of the device into a viewing position (as seen in FIGS. 8A and 8B). Whether device movement constitutes a lifting movement is based on accelerometer readings over time, in some embodiments. Whether device movement constitutes a lifting movement is based on the smoothness of accelerometer readings over time, in some embodiments. Whether device movement constitutes a lifting movement is based on sensing minimum period of dwell time (e.g., a period of relative non-movement or steadiness), in some embodiments.

[0262] In the exemplary lifting movement depicted in FIGS. 8A-8B (as well as in FIGS. 10A-10B), user 802 raises device 800 from its lowered position in FIG. 8A to its raised position in FIG. 8B. This movement includes acceleration over time in one direction (e.g., in the vertical direction along the y-axis), and a rotational acceleration over time in another direction (e.g., a horizontal rotation about the x-axis), both of which can be determined by the acceleration sensors of device 800. Optionally, the smoothness of such a lifting movement can be determined by comparing fluctuations the acceleration sensors' readings over time. Optionally, the lifting movement ends with the user glancing on the device's touch screen in the raised position. The period of relative steadiness as user 802 focuses on the display of device 800 can be determined using acceleration sensor readings. Device 800 uses at least some of these accelerometer readings in determining whether a predetermined condition for sampling audio input via microphone is met, in some embodiments.

[0263] Additional detail regarding techniques for determining raising gestures of an electronic device are described in, e.g., U.S. Provisional Patent Application Serial No. 62/026,532, titled "Raise Gesture Detection in a Device," filed July 18, 2014, the content of which is hereby incorporated by reference. For brevity, the content of the application is not repeated here.

[0264] In some embodiments, the predetermined condition is activation of an input device such as a mechanical button, touch-sensitive button, rotatable input device, so forth. In some embodiments, the predetermined condition is a touch on touch-sensitive surface or touch screen (such as touch screen 804 shown in FIG. 8A).

[0265] Optionally, at block 1204, the device determines whether it is operating in a predetermined mode where its audio output is muted (e.g., a “do not disturb”). If the device is operating in such a mode, processing returns to block 1202, meaning that the device foregoes sampling audio for triggering a virtual assistant.

[0266] In some embodiments, sampling audio at block 1206 involves powering on a microphone of the device. In some examples, sampling audio involves powering on additional electronics circuitry to perform signal processing. In some examples, sampling audio involves passing audio input received from the microphone to software recognition algorithms that are being executed on one or more processors of the device.

[0267] Having sampled audio input at block 1206, the device proceeds to block 1208, where it determines whether the sampled audio input comprises a spoken trigger for triggering a virtual assistant. If the sampled audio input contains the spoken trigger, processing proceeds to block 1210 where the device triggers a virtual assistant session (such as by displaying user interface 812 shown in FIG. 8C). In some embodiments, the spoken trigger is a predetermined phrase, such as “hey Siri.”

[0268] If the sampled audio input does not contain the spoken trigger, processing proceeds to block 1212, where the device determines whether to continue to sample audio input. If the device is to continue sampling audio input, processing returns to block 1206 where audio sampling continues. If the device is to cease sampling of audio input, processing proceeds to block 1214 where audio sampling is stopped.

[0269] In some embodiments, the device ceases to sample audio input after the device’s display is turned off. In some embodiments, the device ceases to sample audio input after sampling for a threshold duration of time, such as a duration of eight seconds. In some embodiments, the device ceases to sample audio after detecting activation of an input device such as a mechanical button, touch-sensitive button, rotatable input device, so forth. In some embodiments, the device ceases to sample audio after detecting activation of an input device such as a touch on touch-sensitive surface or touch screen.

[0270] In some embodiments, the device ceases to sample audio input after an intervening input is received, such as a user input to invoke another application installed on the device. In some embodiments, the device ceases to sample audio based on an amount of lowering of the device during a time interval, the amount of lowering determined from an accelerometer of the electronic device. In some embodiment, the device ceases to sample audio based on smoothness of a lowering of the electronic device during a time interval, the smoothness of the lowering determined from an accelerometer of the device. In some embodiment, the device ceases to sample audio input after identifying an audio endpoint in the audio input. Exemplary end points include those indicating that the received audio input is too low in volume to be near-field speech (as opposed to background noise), unlikely to be human speech, a stop in the user's spoken input, so forth.

[0271] In some embodiments, the device ceases sampling of audio ceases at block 1214 by powering off the microphone of the device. In some embodiments, sampling of audio ceases by powering off associated circuitry that performs microphone signal processing. In some embodiments, sampling audio ceases by not processing audio input received from the microphone through software recognition algorithms. Optionally, at block 1214, the device provides one or more of a haptic, audio, and visual output indicating the end of the audio sampling.

[0272] It should be understood that the particular order in which the operations in FIGS. 12 has been described is merely exemplary and is not intended to indicate that the described order is the only order in which the operations could be performed. One of ordinary skill in the art would recognize various ways to reorder the operations described herein. Additionally, it should be noted that details of other processes described herein with respect to other methods described herein (e.g., methods 1300 and 1500) are also applicable in an analogous manner to method 1200 described above with respect to FIG. 12. For example, it is possible for the device performing method 1200 to also perform method 1500 (FIG. 15) for overcoming erroneous user cancellation of the virtual assistant. For brevity, these details are not repeated here.

[0273] FIGS. 13 is a flow diagram illustrating method 1300 of triggering a virtual assistant on an electronic device in accordance with some embodiments. Method 1300 is performed at one of devices 104, 200, 400, 900, and 1700 (FIGS. 1, 2, 4, 8, and 17) in some embodiments. In some embodiments, method 1300 is performed at a device having a touch

screen. In some embodiments, method 1300 is performed at a device with a separate display screen and a touch-sensitive surface. Operations in method 1300 are, optionally, combined and/or the order of some operations is, optionally, changed.

[0274] At block 1302, the electronic device detects a software event. At block 1304, the device determines whether the software event meets a predetermined condition. If the software event meets the predetermined condition, processing proceeds to block 1306, and the device begins sampling audio input received via its microphone. If the software event does not meet the predetermined condition, processing returns to block 1302, where the device may detect another software event.

[0275] In the illustration of FIG. 9A, notification 906 of incoming text message 912 meets the predetermined condition for causing device 900 to sample audio input. In some embodiments, the predetermined condition is that the software event is an operating system event triggered by the operating system of the device. Exemplary software events triggered by the operating system includes an alert about a hardware component of device 900, such as an alert that the device has been connected to a computer via cable. In some embodiments, the predetermined condition is that the software event is an application-based event triggered by an application executing on the device. Exemplary software events triggered by an application of the device include, calendar reminders, task reminders, so forth. Exemplary software events that may be considered to have been triggered by the operating system and/or an application or device 900 include clock events such as an alarm or a timer expiration.

[0276] In some embodiments, the predetermined condition is that the software event is an event triggered by an external source. Exemplary software events triggered by external sources include incoming calls (e.g., voice- or video-based, cellular or WiFi-based calls), messages (e.g., e-mail, text message SMS, multimedia message, iMessage, so forth) calendar invitations, so forth. Exemplary software events triggered by external sources also include application-based notifications and alerts, such as alerts indicating the availability of newly available information on a web site or service.

[0277] In some embodiments, the predetermined condition is that the software event indicates that a virtual assistant session is active on the device.

[0278] Optionally, at block 1304, the device determines whether it is operating in a predetermined mode where its audio output is muted (e.g., a “do not disturb”). If the device

is operating in such a mode, processing returns to block 1302, meaning that the device foregoes sampling audio for triggering a virtual assistant.

[0279] Optionally, at block 1304, the device determines its display screen is on. The device determines whether the display is on, for example, by determining whether its backlight is lit. If the device's display is off, processing returns to block 1302, meaning that the device foregoes sampling audio for triggering a virtual assistant.

[0280] Having sampled audio input at block 1306, the device proceeds to block 1308, where it determines whether the sampled audio input includes a spoken trigger for triggering a virtual assistant. If the sampled audio input contains the spoken trigger, processing proceeds to block 1310 where the device triggers a virtual assistant session and, optionally, executes tasks based on the user's spoken input (as can be seen in the exemplary illustrations of FIGS. 9B and 9C). In some embodiments, the spoken trigger is a predetermined phrase, such as "hey Siri."

[0281] If the sampled audio input does not contain the spoken trigger, processing proceeds to block 1312, where the device determines whether to continue to sample audio input. If the device is to continue sampling audio input, processing returns to block 1306 where audio sampling continues. If the device is to cease sampling of audio input, processing proceeds to block 1314 where audio sampling is stopped.

[0282] In some embodiments, the device ceases to sample audio input after the device's display is turned off. In some embodiments, the device ceases to sample audio input after sampling for a threshold duration of time, such as a duration of eight seconds. In some embodiments, the device ceases to sample audio after detecting activation of an input device such as a mechanical button, touch-sensitive button, rotatable input device, so forth. In some embodiments, the device ceases to sample audio after detecting activation of an input device such as a touch on touch-sensitive surface or touch screen.

[0283] In some embodiments, the device ceases to sample audio input after an intervening input is received, such as a user input to invoke another application installed on the device. In some embodiments, the device ceases to sample audio based on an amount of lowering of the device during a time interval, the amount of lowering determined from an accelerometer of the electronic device. In some embodiments, the device ceases to sample audio based on smoothness of a lowering of the electronic device during a time interval, the

smoothness of the lowering determined from an accelerometer of the device. In some embodiments, the device ceases to sample audio input after identifying an audio endpoint in the audio input. Exemplary end points include those indicating that the received audio input is too low in volume to be near-field speech (as opposed to background noise), unlikely to be human speech, a stop in the user's spoken input, so forth.

[0284] In some embodiments, the device ceases sampling of audio ceases at block 1314 by powering off the microphone of the device. In some embodiments, sampling of audio ceases by powering off associated circuitry that performs microphone signal processing. In some embodiments, sampling audio ceases by not processing audio input received from the microphone through software recognition algorithms. Optionally, at block 1314, the device provides one or more of a haptic, audio, and visual output indicating the end of the audio sampling.

[0285] It should be understood that the particular order in which the operations in FIGS. 13 has been described is merely exemplary and is not intended to indicate that the described order is the only order in which the operations could be performed. One of ordinary skill in the art would recognize various ways to reorder the operations described herein. Additionally, it should be noted that details of other processes described herein with respect to other methods described herein (e.g., methods 1200 and 1500) are also applicable in an analogous manner to method 1300 described above with respect to FIG. 13. For example, it is possible for the device performing method 1300 to also perform method 1500 (FIG. 15) for overcoming erroneous user cancellation of the virtual assistant. For brevity, these details are not repeated here.

[0286] FIGS. 14 is a flow diagram illustrating method 1400 of triggering a virtual assistant on an electronic device in accordance with some embodiments. Method 1400 is performed at one of devices 104, 200, 400, 1000, and 1800 (FIGS. 1, 2, 4, 10, and 18) in some embodiments. In some embodiments, method 1400 is performed at a device having a touch screen. In some embodiments, method 1400 is performed at a device with a separate display screen and a touch-sensitive surface. Operations in method 1400 are, optionally, combined and/or the order of some operations is, optionally, changed.

[0287] At block 1402, the device detects user input via an input component of the device. At block 1404, the device determines if the input meets a predetermined condition. If the

input meets a predetermined condition, processing proceeds to 1406 where the device samples audio input received via its microphone. If the input does not meet the predetermined condition, processing returns to block 1402, where the device may detect another user input.

[0288] In some embodiments, the predetermined condition is that the device receives the user input while its display is off. The device determines whether the display is off, for example, by determining whether its backlight is lit.

[0289] In some embodiments, the predetermined condition is a movement of the device, such as a lifting of the device into a viewing position (e.g., as seen in FIGS. 8A-8B and 10A-10B). In some embodiments, the predetermined condition is based on an amount of lifting of the electronic device during a time interval, the amount of lifting determined from an accelerometer of the electronic device. In some embodiments, the predetermined condition is based on smoothness of a lifting of the electronic device during a time interval, the smoothness of the lifting determined from an accelerometer of the electronic device. In some embodiments, the predetermined condition comprises a minimum period of dwell time in a position in accordance with the lifting of the electronic device.

[0290] In some embodiments, the predetermined condition is activation of an input device such as a mechanical button, touch-sensitive button, rotatable input device, so forth. In some examples, the predetermined condition is a touch on touch-sensitive surface or touch screen.

[0291] Optionally, at block 1404, the device determines whether it is operating in a predetermined mode where its audio output is muted (e.g., a “do not disturb”). If the device is operating in such a mode, processing returns to block 1402, meaning that the device foregoes sampling audio for triggering a virtual assistant.

[0292] In some embodiments, sampling audio at block 1406 involves powering on a microphone of the device. In some embodiments, sampling audio involves powering on additional electronics circuitry to perform signal processing. In some embodiments, sampling audio involves passing audio input received from the microphone to software recognition algorithms that are being executed on one or more processors of the device.

[0293] In some embodiments, the device proceeds from block 1402 (where the input is detected) to block 1406 (where audio sampling occurs) without turning on its display (e.g., by not turning on the backlight of the display). Optionally, the device provides at least one of a haptic and audio output at block 1406 without powering on its display. In some embodiments, the device provides a visual output from a light source other than its display as it starts to sample audio input.

[0294] Having sampled audio input at block 1406, the device proceeds to block 1408, where it determines whether sampled audio input comprises a spoken trigger for triggering a virtual assistant. If the sampled audio input contains the spoken trigger, processing proceeds to block 1410 where the device triggers a virtual assistant session. In some embodiments, the spoken trigger is a predetermined phrase, such as “hey Siri.”

[0295] In some embodiments, the device proceeds from block 1402 (where the input is detected) to block 1408 (where the virtual assistant is triggered) while the display of the device remains off (e.g., by not turning on the backlight of the display). In some embodiments, the device provides at least one of a haptic and audio output as it triggers the virtual assistant at block 1410. In some embodiments, the device provides a visual output from a light source other than its display as it triggers the virtual assistant at block 1410.

[0296] If the sampled audio input does not contain the spoken trigger, processing proceeds to block 1412, where the device determines whether to continue to sample audio input. If the device is to continue sampling audio input, processing returns to block 1406 where audio sampling continues. If the device is to cease sampling of audio input, processing proceeds to block 1414 where audio sampling is stopped.

[0297] In some embodiments, the device ceases to sample audio input after sampling for a threshold duration of time, such as a duration of eight seconds. In some embodiments, the device ceases to sample audio after detecting activation of an input device such as a mechanical button, touch-sensitive button, rotatable input device, so forth. In some embodiments, the device ceases to sample audio after detecting activation of an input device such as a touch on touch-sensitive surface or touch screen.

[0298] In some embodiments, the device ceases to sample audio input after an intervening input is received, such as a user input to invoke another application installed on the device. In some embodiments, the device ceases to sample audio based on an amount of

lowering of the device during a time interval, the amount of lowering determined from an accelerometer of the electronic device. In some embodiments, the device ceases to sample audio based on smoothness of a lowering of the electronic device during a time interval, the smoothness of the lowering determined from an accelerometer of the device. In some embodiments, the device ceases to sample audio input after identifying an audio endpoint in the audio input. Exemplary end points include those indicating that the received audio input is too low in volume to be near-field speech (as opposed to background noise), unlikely to be human speech, a stop in the user's spoken input, so forth.

[0299] In some embodiments, the device ceases sampling of audio at block 1414 by powering off the microphone of the device. In some embodiments, sampling of audio ceases by powering off associated circuitry that performs microphone signal processing. In some embodiments, sampling audio ceases by not processing audio input received from the microphone through software recognition algorithms.

[0300] Optionally, the device provides at least one of a haptic and audio output as it ceases to sample audio input without turning on its display. In some embodiments, the device provides a visual output from a light source other than its display as it ceases to sample audio input.

[0301] It should be understood that the particular order in which the operations in FIGS. 14 has been described is merely exemplary and is not intended to indicate that the described order is the only order in which the operations could be performed. One of ordinary skill in the art would recognize various ways to reorder the operations described herein. Additionally, it should be noted that details of other processes described herein with respect to other methods described herein (e.g., methods 1300 and 1500) are also applicable in an analogous manner to method 1400 described above with respect to FIG. 14. For example, it is possible for the device performing method 1400 to also perform method 1500 (FIG. 15) for overcoming erroneous user cancellation of the virtual assistant. For brevity, these details are not repeated here.

[0302] FIGS. 15 is a flow diagram illustrating method 1500 of triggering a virtual assistant on an electronic device in accordance with some embodiments. Method 1500 is performed at one of devices 104, 200, 400, 1100, and 1900 (FIGS. 1, 2, 11, and 19) in some embodiments. In some embodiments, method 1500 is performed at a device having a touch

screen. In some embodiments, method 1500 is performed at a device with a separate display screen and a touch-sensitive surface. Operations in method 1500 are, optionally, combined and/or the order of some operations is, optionally, changed.

[0303] At block 1502, the electronic device samples audio input and determines whether the audio input represents an executable task, for example, a user request that is provided in spoken natural language. While sampling audio input, the device can provide a UI indicating of the sampling, such as screen 1106 of FIG. 11A.

[0304] In some embodiments, sampling of audio input occurs when the device's display is powered on, for example, when the backlight of the display is lit. In some embodiments, the device samples audio input responsive to a movement meeting a predetermined condition, such as a lifting of the device into a viewing position (as seen in FIGS. 8A-8B, 10A-10B, and 11A). In some embodiments, the predetermined condition is based on an amount of lifting of the electronic device during a time interval, the amount of lifting determined from an accelerometer of the electronic device. In some embodiments, the predetermined condition is based on smoothness of a lifting of the electronic device during a time interval, the smoothness of the lifting determined from an accelerometer of the electronic device. In some embodiments, the predetermined condition comprises a minimum period of dwell time in a position in accordance with the lifting of the electronic device.

[0305] In some embodiments, the predetermined condition is activation of an input device such as a mechanical button, touch-sensitive button, rotatable input device, so forth. In some embodiments, the predetermined condition is a touch on touch-sensitive surface or touch screen.

[0306] At block 1504, while sampling audio input, the electronic device receives an instruction to cease sampling of audio input. In some embodiments, the device is adapted to receive instruction to cease sampling of audio input from a user of the device. The instruction can be received as a spoken instruction via the microphone of the device, for example. Also, the instruction can be received by detecting activation of an input device such as a touch on touch-sensitive surface or touch screen, such as a touch on affordance 1108 in FIG. 11A, for example. Further, the instruction can be received by detecting activation of a mechanical button, touch-sensitive button, rotatable input device, so forth, for example.

[0307] In some embodiments, the device is adapted to receive instruction to cease sampling of audio input from another electronic device external to the device. The instruction can be received through a communication medium such as cellular communication, Bluetooth communication, WiFi communication, or the like.

[0308] In some embodiments, at block 1504, the device prompts the user for spoken input. In some embodiments, at block 1504, the device determines whether the audio input represents a request, and in accordance with determining whether the audio input represents a task, the device: determines a user intent based on at least a portion of the audio input; identifies and executes the task based on the user intent; and provides at least one of a haptic, audio, and visual output representing execution of the task.

[0309] At block 1506, responsive to the instruction to cease sampling of audio input, the device provides an output acknowledgement. The acknowledgement reports to the user that the virtual assistant has been canceled, for example. In some embodiments, the acknowledgement involves one or more of a haptic, audio, and visual output. In some embodiments, the acknowledgement involves turning the display off, as seen in FIG. 11B.

[0310] In some embodiments, after receiving the instruction to cease sampling of audio input at block 1504 and after providing the acknowledgement at block 1506, processing proceeds to block 1508, where the device continues to sample audio input and determines whether the audio input includes a user task, for some limited time (after which the device ceases to sample audio input).

[0311] In some embodiments, the device ceases to sample audio input after the device's display is turned off. In some embodiments, the device ceases to sample audio input after continuing to sample for a threshold duration of time, such as a duration of eight seconds. In some embodiments, the device ceases to sample audio after detecting activation of an input device such as a mechanical button, touch-sensitive button, rotatable input device, so forth. In some embodiments, the device ceases to sample audio after detecting activation of an input device such as a touch on touch-sensitive surface or touch screen.

[0312] In some embodiments, the device ceases to sample audio input after an intervening input is received, such as a user input to invoke another application installed on the device. In some embodiments, the device ceases to sample audio based on an amount of lowering of the device during a time interval, the amount of lowering determined from an

accelerometer of the electronic device. In some embodiments, the device ceases to sample audio based on smoothness of a lowering of the electronic device during a time interval, the smoothness of the lowering determined from an accelerometer of the device. In some embodiments, the device ceases to sample audio input after identifying an audio endpoint in the audio input. Exemplary end points include those indicating that the received audio input is too low in volume to be near-field speech (as opposed to background noise), unlikely to be human speech, a stop in the user's spoken input, so forth.

[0313] In some embodiments, after receiving the instruction to cease sampling of audio input 1504, the device determines whether the instruction to cease sampling of audio input originates from a user or an external device. In accordance with a determination that the audio input originated from an external device, the device ceases to sample audio input immediately.

[0314] In some embodiments, the device ceases sampling of audio ceases at block 1506 by powering off the microphone of the device. In some embodiments, sampling of audio ceases by powering off associated circuitry that performs microphone signal processing. In some embodiments, sampling audio ceases by not processing audio input received from the microphone through software recognition algorithms. Optionally, at block 1506, the device provides one or more of a haptic, audio, and visual output indicating the end of the audio sampling.

[0315] It should be understood that the particular order in which the operations in FIGS. 15 has been described is merely exemplary and is not intended to indicate that the described order is the only order in which the operations could be performed. One of ordinary skill in the art would recognize various ways to reorder the operations described herein. Additionally, it should be noted that details of other processes described herein with respect to other methods described herein (e.g., methods 1200, 1300, 1400, 2100) are also applicable in an analogous manner to method 1500 described above with respect to FIG. 15. For example, it is possible for the device performing method 1500 to also perform method 1400 for triggering virtual assistant sessions. For brevity, these details are not repeated here.

[0316] FIGS. 21 is a flow diagram illustrating method 2100 of triggering a virtual assistant on an electronic device in accordance with some embodiments. Method 2100 is performed at one of devices 104, 200, 400, and 2000 (FIGS. 1, 2, 4, 8, and 20) in some

embodiments. In some embodiments, method 2100 is performed at a device having a touch screen. In some embodiments, method 2100 is performed at a device with a separate display screen and a touch-sensitive surface. Operations in method 1200 are, optionally, combined and/or the order of some operations is, optionally, changed.

[0317] At block 2102, while a predetermined service such as a virtual assistant is executing on the device, the device detects a user's natural language input such as spoken input. At block 2104, the device operationalizes the user's input by identifying one or more computing tasks to be performed. At block 2106, while performing the one or more tasks, the device powers down its display screen (e.g., by turning off the backlight of the screen). At block 2108, after completing the one or more computing tasks, the device provides a haptic and/or audio output.

[0318] It should be understood that the particular order in which the operations in FIGS. 21 has been described is merely exemplary and is not intended to indicate that the described order is the only order in which the operations could be performed. One of ordinary skill in the art would recognize various ways to reorder the operations described herein. Additionally, it should be noted that details of other processes described herein with respect to other methods described herein (e.g., methods 1200, 1300, 1400, and 1500) are also applicable in an analogous manner to method 2100 described above with respect to FIG. 21. For example, it is possible for the device performing method 2100 to also perform method 1400 for triggering virtual assistant sessions. For brevity, these details are not repeated here.

[0319] It is noted that the sampling of audio input at blocks 1206, 1306, 1406, and 1502 (FIGS. 12, 13, 14, and 15) can occur independent of whether the electronic device is being charged. However, such sampling of audio input can depend on the remaining battery level of the device. In some examples, a device may sample audio input consistent with the above-described processes, but forego sampling of audio input when its remaining battery level is below a threshold level.

[0320] In accordance with some embodiments, FIG. 16 shows a functional block diagram of an electronic device 1600 configured in accordance with the principles of the various described embodiments, including those described with reference to FIGS. 8 and 12. The functional blocks of the device are, optionally, implemented by hardware, software, or a combination of hardware and software to carry out the principles of the various described

embodiments. It is understood by persons of skill in the art that the functional blocks described in FIG. 16 are, optionally, combined or separated into sub-blocks to implement the principles of the various described embodiments. Therefore, the description herein optionally supports any possible combination or separation or further definition of the functional blocks described herein.

[0321] As shown in FIG. 16, an electronic device 1600 includes a display unit 1602 configured to display a graphical user interface, optional touch-sensitive surface unit 1604 to receive contacts, a microphone unit 1606 to detect audio input, optional movement sensor unit 1608 to obtain movement readings such as acceleration in a time interval, optional input device unit 1610 such as a button, optional feedback unit 1612 to provide feedback such as audio, haptic, and/or visual feedback, and processing unit 1614 coupled to the above-described units. In some embodiments, processing unit 1614 includes display enabling unit 1616, input detection unit 1618, audio sampling unit 1620, audio processing unit 1622, virtual assistant service unit 1624, and, optionally, feedback enabling unit 1626.

[0322] In some embodiments, processing unit 1614 is configured to: while the display is on, receive (e.g., with display enabling unit 1616) user input via the input device, the user input meeting a predetermined condition; in accordance with receiving the user input meeting the predetermined condition (e.g., with input detection unit 1618), sample audio input (e.g., with audio sampling unit 1620) received via the microphone (e.g., from microphone unit 1606); determine whether the audio input comprises a spoken trigger (e.g., with audio processing unit 1622); and in accordance with a determination that audio input comprises the spoken trigger (e.g., with audio processing unit 1622), trigger a virtual assistant session (e.g., with virtual assistant service unit 1624).

[0323] In some embodiments, the predetermined condition is based on an amount of lifting of the electronic device during a time interval, the amount of lifting determined from an accelerometer of the electronic device (e.g., movement sensor unit 1608). In some embodiments, the predetermined condition is based on smoothness of a lifting of the electronic device during a time interval, the smoothness of the lifting determined from an accelerometer of the electronic device (e.g., movement sensor unit 1608). In some embodiments, the predetermined condition comprises a minimum period of dwell time in a position in accordance with the lifting of the electronic device. In some embodiments, the

predetermined condition comprises detecting (e.g., input detection unit 1618) activation of a button of the electronic device (e.g., input device unit 1610).

[0324] In some embodiments, the predetermined condition comprises detecting touch input on a touch-sensitive surface of the electronic device (e.g., touch-sensitive surface unit 1604).

[0325] In some embodiments, the processing unit is further configured to: determine (e.g., with display enabling unit 1616) whether the display is on; the sampling of audio input (e.g., with audio sampling unit 1620) occurs in accordance with a determination that the display is on. In some embodiments, the processing unit is further configured to: in accordance with determining whether the display is on, determine whether the backlight of the display is on (e.g., with display enabling unit 1616).

[0326] In some embodiments, the processing unit is further configured to: cease the sampling of audio input after the sampling of audio input has occurred for a predetermined duration of time (e.g., with audio sampling unit 1620).

[0327] In some embodiments, the user input meeting a predetermined condition is a first user input meeting a first predetermined condition, and processing unit 1614 is further configured to: while sampling audio input (e.g., with audio sampling unit 1620) in accordance with receiving the first user input, receiving a second user input (e.g., with input detection unit 1618) meeting a second predetermined condition; and in accordance with receiving the second user input meeting the second predetermined condition, ceasing (e.g., with audio sampling unit 1620) the sampling of audio input.

[0328] In some embodiments, the second predetermined condition is based on an amount of lowering of the electronic device during a time interval, the amount of lowering determined from an accelerometer of the electronic device (e.g., with movement sensor unit 1608).

[0329] In some embodiments, the first predetermined condition is based on smoothness of a lowering of the electronic device during a time interval, the smoothness of the lowering determined from an accelerometer of the electronic device (e.g., with movement sensor unit 1608).

[0330] In some embodiments, the second predetermined condition comprises detecting (e.g., with input detection unit 1618) activation of a button of the electronic device. In some embodiments, the second predetermined condition comprises detecting touch input on a touch-sensitive surface of the electronic device.

[0331] In some embodiments, processing unit 1614 is further configured to: while sampling audio input in accordance with receiving the user input meeting the predetermined condition (e.g., with audio sampling unit 1620), determining whether the display is off (e.g., with display enabling unit 1616); and in accordance with a determination that the display is off, ceasing (e.g., with audio sampling unit 1620) the sampling of audio input.

[0332] In some embodiments, processing unit 1614 is further configured to: while sampling audio input in accordance with receiving the user input meeting the predetermined condition (e.g., with audio sampling unit 1620), identifying an audio endpoint in the audio input (e.g., with audio processing unit 1622); and in response to identifying the audio endpoint, ceasing (e.g., with audio sampling unit 1620) the sampling of audio input.

[0333] In some embodiments, processing unit 1614 is further configured to: in accordance with ceasing to sample audio input, causing a haptic output (e.g., with feedback enabling unit 1626 and/or feedback unit 1612).

[0334] In some embodiments, processing unit 1614 is further configured to: in accordance with triggering the virtual assistant session (e.g., with virtual assistant service unit 1624), prompting (e.g., with display enabling unit 1616 and/or feedback enabling unit 1626) a user for spoken input (e.g., with display unit 1602 and/or feedback unit 1612).

[0335] In some embodiments, the audio input comprises the spoken trigger and additional input, and processing unit 1614 is further configured to, in accordance with triggering the virtual assistant session: determining a user intent based on at least the additional input in the audio input, and executing a task associated with the user intent (e.g., with virtual assistant service unit 1624).

[0336] In some embodiments, processing unit 1614 is further configured to: in accordance with a determination that the electronic device is operating in a predetermined mode (e.g., with virtual assistant service unit 1624), forgoing sampling of audio input (e.g.,

with audio sampling unit 1620), even after receiving the user input meeting the predetermined condition while the display is on.

[0337] In some embodiments, operation in the predetermined mode comprises muting audio output of the electronic device. In some embodiments, the spoken trigger comprises a predetermined phrase.

[0338] The operations described above with respect to FIG. 12, optionally, implemented by components depicted in FIGS. 1, 2, 7A-7B, or FIG. 16. For example, sampling audio operation 1202 is optionally implemented by audio circuitry 210 (FIG. 2). When audio having a spoken trigger is detected, the device uses display controller 256 (FIG. 2) and/or haptic feedback controller 261 (FIG. 2) to provide corresponding output. It would be clear to a person of ordinary skill in the art how other processes can be implemented based on the components depicted in FIGS. 1, 2, 7A-7B, or FIG. 16.

[0339] It is understood by persons of skill in the art that the functional blocks described in FIG. 16 are, optionally, combined or separated into sub-blocks to implement the principles of the various described embodiments. Therefore, the description herein optionally supports any possible combination or separation or further definition of the functional blocks described herein. For example, units 1602-1612 can have associated “controller” units that are operatively coupled with the respective unit and processing unit 1614 to enable operation. These controller units are not separately illustrated in FIG. 16 but are understood to be within the grasp of one of ordinary skill in the art who is designing a device having units 1602-1612 such as device 1600. The description herein thus optionally supports combination, separation, and/or further definition of the functional blocks described herein.

[0340] In accordance with some embodiments, FIG. 17 shows a functional block diagram of an electronic device 1700 configured in accordance with the principles of the various described embodiments, including those described with reference to FIGS. 9 and 13. The functional blocks of the device are, optionally, implemented by hardware, software, or a combination of hardware and software to carry out the principles of the various described embodiments. It is understood by persons of skill in the art that the functional blocks described in FIG. 17 are, optionally, combined or separated into sub-blocks to implement the principles of the various described embodiments. Therefore, the description herein optionally

supports any possible combination or separation or further definition of the functional blocks described herein.

[0341] As shown in FIG. 17, an electronic device 1700 includes a display unit 1702 configured to display a graphical user interface, optional touch-sensitive surface unit 1704 to receive contacts, a microphone unit 1706 to detect audio input, optional movement sensor unit 1708 to obtain movement readings such as acceleration in a time interval, optional input device unit 1710 such as a button, optional feedback unit 1712 to provide feedback such as audio, haptic, and/or visual feedback, and processing unit 1714 coupled to the above-described units. In some embodiments, processing unit 1714 includes display enabling unit 1716, software event detection input 1728, input detection unit 1718, audio sampling unit 1720, audio processing unit 1722, virtual assistant service unit 1724, and, optionally, feedback enabling unit 1726.

[0342] In some embodiments, processing unit 1714 is configured to: detect a software event (e.g., with software event detection unit 1728) meeting a predetermined condition; in accordance with a determination that the software event meeting the condition is detected, sample audio input (e.g., with audio sampling unit 1620) received via the microphone (e.g., microphone unit 1706); determine (e.g., with audio processing unit 1722) whether audio input received via the microphone comprises a spoken trigger; and in accordance with a determination that the audio input comprises the spoken trigger, trigger a virtual assistant session (e.g., with virtual assistant service unit 1724).

[0343] In some embodiments, the predetermined condition is that the software event is an event triggered by an application of the electronic device. In some embodiments, the predetermined condition comprises the software event representing a notification. In some embodiments, the predetermined condition comprises the software event representing a calendar notification.

[0344] In some embodiments, the predetermined condition comprises the software event representing an incoming calendar invitation from an external device. In some embodiments, the predetermined condition comprises the software event representing an incoming transmission, the transmission triggered from an external device. In some embodiments, the predetermined condition comprises the software event representing an incoming phone call

from an external device. In some embodiments, the predetermined condition comprises the software event representing an incoming text or multimedia message from an external device.

[0345] In some embodiments, processing unit 1714 is further configured to: in accordance with triggering a virtual assistant session, initiating (e.g., with virtual assistant service unit 1624) a response addressed to the external device. An exemplary response is a reply to an incoming message, such as text message 916 in FIG. 9C. Other exemplary responses include e-mail messages, multi-media messages, calendar invitation responses, so forth. As used herein, initiation In the example of FIG. 9C, responsive text message 916 was transmitted by device 900 automatically based on user instructions 910. Optionally, after initiating message 916 (e.g., populating the message with recipient information and content), the device can prompt the user for a final confirmation before sending the message to the recipient.

[0346] In some embodiments, processing unit 1714 is further configured to: determining (e.g., with display enabling unit 1716) whether the display is on, and wherein the sampling of audio input (e.g., with audio sampling unit 1720) occurs in accordance with a determination that the display is on.

[0347] In some embodiments, processing unit 1714 is further configured to: in accordance with determining whether the display is on (e.g., with display enabling unit 1716), determining (e.g., with display enabling unit 1716) whether the backlight of the display is on.

[0348] In some embodiments, processing unit 1714 is further configured to: cease the sampling of audio input (e.g., with audio sampling unit 1720) after the sampling of audio input has occurred for a predetermined duration of time.

[0349] In some embodiments, the user input meeting a predetermined condition is a first user input meeting a first predetermined condition, and processing unit 1714 is further configured to: while sampling audio input (e.g., with audio sampling unit 1720) in accordance with receiving the first user input, receiving (e.g., with input detection unit 1718) a second user input meeting a second predetermined condition; and in accordance with receiving the second user input meeting the second predetermined condition, ceasing (e.g., with audio sampling unit 1720) the sampling of audio input.

[0350] In some embodiments, the second predetermined condition is based on an amount of lowering of the electronic device during a time interval, the amount of lowering determined from an accelerometer (e.g., with movement sensor unit 1708) of the electronic device. In some embodiments, the first predetermined condition is based on smoothness of a lowering of the electronic device during a time interval, the smoothness of the lowering determined from an accelerometer (e.g., with movement sensor unit 1708) of the electronic device.

[0351] In some embodiments, the second predetermined condition comprises detecting activation (e.g., with input detection unit 1718) of a button (e.g., with input device unit 1710) of the electronic device. In some embodiments, the second predetermined condition comprises detecting touch input on a touch-sensitive surface (e.g., with touch-sensitive surface unit 1704) of the electronic device.

[0352] In some embodiments, processing unit 1714 is further configured to: while sampling audio input (e.g., with audio sampling unit 1720) in accordance with receiving the user input meeting the predetermined condition, determining (e.g., with display enabling unit 1716) whether the display is off; and in accordance with a determination that the display is off, ceasing the sampling of audio input (e.g., with audio sampling unit 1720).

[0353] In some embodiments, processing unit 1714 is further configured to: while sampling audio input (e.g., with audio sampling unit 1720) in accordance with receiving the user input meeting the predetermined condition, identifying (e.g., with audio processing unit 1722) an audio endpoint in the audio input; and in response to identifying the audio endpoint, ceasing the sampling of audio input (e.g., with audio sampling unit 1720).

[0354] In some embodiments, processing unit 1714 is in accordance with ceasing to sample audio input, causing (e.g., with feedback enabling unit 1726) a haptic output (e.g., with feedback unit 1712).

[0355] In some embodiments, processing unit 1714 is in accordance with triggering the virtual assistant session, prompting (e.g., with display enabling unit 1716 and/or feedback enabling unit 1726) a user for spoken input (e.g., with display unit 1702 and/or feedback unit 1712).

[0356] In some embodiments, the audio input comprises the spoken trigger and additional input, and processing unit 1714 is further configured to: in accordance with triggering the virtual assistant session (e.g., with virtual assistant service unit 1724): determining (e.g., with virtual assistant service unit 1724) a user intent based on at least the additional input in the audio input, and executing (e.g., with virtual assistant service unit 1724) a task associated with the user intent.

[0357] In some embodiments, processing unit 1714 is further configured to: in accordance with a determination that the electronic device is operating in a predetermined mode (e.g., with virtual assistant service unit 1724), forgoing sampling of audio input (e.g., with audio sampling unit 1720), even after receiving (e.g., with input detection unit 1718) user input meeting the predetermined condition while the display is on.

[0358] In some embodiments, operation in the predetermined mode comprises muting audio output of the electronic device. In some embodiments, the spoken trigger comprises a predetermined phrase, such as “hey Siri.”

[0359] The operations described above with respect to FIG. 13, optionally, implemented by components depicted in FIGS. 1, 2, 7A-7B, or FIG. 17. For example, sampling audio operation 1202 is optionally implemented by audio circuitry 210 (FIG. 2). When audio having a spoken trigger is detected, the device uses display controller 256 (FIG. 2) and/or haptic feedback controller 261 (FIG. 2) to provide corresponding output. It would be clear to a person of ordinary skill in the art how other processes can be implemented based on the components depicted in FIGS. 1, 2, 7A-7B, or FIG. 17.

[0360] It is understood by persons of skill in the art that the functional blocks described in FIG. 17 are, optionally, combined or separated into sub-blocks to implement the principles of the various described embodiments. Therefore, the description herein optionally supports any possible combination or separation or further definition of the functional blocks described herein. For example, units 1702-1712 can have associated “controller” units that are operatively coupled with the respective unit and processing unit 1714 to enable operation. These controller units are not separately illustrated in FIG. 17 but are understood to be within the grasp of one of ordinary skill in the art who is designing a device having units 1702-1712 such as device 1700. The description herein thus optionally supports combination, separation, and/or further definition of the functional blocks described herein.

[0361] In accordance with some embodiments, FIG. 18 shows a functional block diagram of an electronic device 1800 configured in accordance with the principles of the various described embodiments, including those described with reference to FIGS. 10 and 14. The functional blocks of the device are, optionally, implemented by hardware, software, or a combination of hardware and software to carry out the principles of the various described embodiments. It is understood by persons of skill in the art that the functional blocks described in FIG. 18 are, optionally, combined or separated into sub-blocks to implement the principles of the various described embodiments. Therefore, the description herein optionally supports any possible combination or separation or further definition of the functional blocks described herein.

[0362] As shown in FIG. 18, an electronic device 1800 includes a display unit 1802 configured to display a graphical user interface, optional touch-sensitive surface unit 1804 to receive contacts, a microphone unit 1806 to detect audio input, optional movement sensor unit 1808 to obtain movement readings such as acceleration in a time interval, optional input device unit 1810 such as a button, optional feedback unit 1812 to provide feedback such as audio, haptic, and/or visual feedback, and processing unit 1814 coupled to the above-described units. In some embodiments, processing unit 1814 includes display enabling unit 1816, input detection unit 1818, audio sampling unit 1820, audio processing unit 1822, virtual assistant service unit 1824, and, optionally, feedback enabling unit 1826.

[0363] In some embodiments, processing unit 1814 is configured to: receive (e.g., with input detection unit 1818) user input via the input device, the user input meeting a predetermined condition; in accordance with receiving the user input meeting the predetermined condition, sampling audio input (e.g., with audio sampling unit 1820) received via the microphone; determining (e.g., with audio processing unit 1822) whether the audio input comprises a spoken trigger; in accordance with a determination that audio input comprises the spoken trigger: triggering (e.g., with virtual assistant service unit 1824) a virtual assistant session and causing a haptic output (e.g., with feedback unit 1826).

[0364] In some embodiments, processing unit 1814 is further configured to: in accordance with receiving the user input meeting a predetermined condition, receiving the user input (e.g., with input detection unit 1818) while the display is off.

[0365] In some embodiments, processing unit 1814 is further configured to: in accordance with triggering the virtual assistant session, triggering the virtual assistant session (e.g., with virtual assistant service unit 1824) without turning on a backlight of the display

[0366] In some embodiments, the predetermined condition is based on an amount of lifting of the electronic device during a time interval, the amount of lifting determined from an accelerometer of the electronic device (e.g., with movement unit 1808). In some embodiments, the predetermined condition is based on smoothness of a lifting of the electronic device during a time interval, the smoothness of the lifting determined from an accelerometer of the electronic device (e.g., with movement unit 1808). In some embodiments, the predetermined condition comprises a minimum period of dwell time in a position in accordance with the lifting of the electronic device.

[0367] In some embodiments, the predetermined condition comprises detecting activation of a button of the electronic device (e.g., with input detection unit 1818). In some embodiments, the predetermined condition comprises detecting touch input on a touch-sensitive surface of the electronic device.

[0368] In some embodiments, processing unit 1814 is further configured to, ceasing the sampling of audio input (e.g., with audio sampling unit 1820) after the sampling of audio input has occurred for a predetermined duration of time.

[0369] In some embodiments, the user input meeting a predetermined condition is a first user input meeting a first predetermined condition, and processing unit 1814 is further configured to: while sampling audio input (e.g., with audio sampling unit 1820) in accordance with receiving the first user input, receiving a second user input (e.g., with input detection unit 1818) meeting a second predetermined condition; and in accordance with receiving the second user input meeting the second predetermined condition, ceasing the sampling of audio input (e.g., with audio sampling unit 1820).

[0370] In some embodiments, the second predetermined condition is based on an amount of lowering of the electronic device during a time interval, the amount of lowering determined from an accelerometer (e.g., with movement unit 1808) of the electronic device. In some embodiments, the first predetermined condition is based on smoothness of a lowering of the electronic device during a time interval, the smoothness of the lowering determined from an accelerometer (e.g., with movement unit 1808) of the electronic device.

[0371] In some embodiments, the second predetermined condition comprises detecting (e.g., with input detection unit 1818) activation of a button (e.g., with input device unit 1810) of the electronic device. In some embodiments, the second predetermined condition comprises detecting (e.g., with input detection unit 1818) touch input on a touch-sensitive surface (e.g., with touch-sensitive surface unit 1804) of the electronic device.

[0372] In some embodiments, processing unit 1814 is further configured to, while sampling audio input (e.g., with audio sampling unit 1820) in accordance with receiving the user input meeting the predetermined condition, identifying (e.g., with audio processing unit 1822) an audio endpoint in the audio input; and in response to identifying the audio endpoint, ceasing the sampling of audio input (e.g., with audio sampling unit 1820).

[0373] In some embodiments, processing unit 1814 is further configured to, in accordance with ceasing to sample audio input (e.g., with audio sampling unit 1820), providing a causing output (e.g., with feedback unit 1826). In some embodiments, processing unit 1814 is further configured to, in accordance with triggering the virtual assistant session (e.g., with virtual assistant service unit 1824), prompting (e.g., with display enabling unit 1816 and/or feedback enabling unit 1826) a user for spoken input (e.g., with display unit 1802 and/or feedback unit 1812).

[0374] In some embodiments, the audio input comprises the spoken trigger and additional input, and processing unit 1814 is further configured to: in accordance with triggering the virtual assistant session (e.g., with virtual assistant service unit 1824): determining (e.g., with virtual assistant service unit 1824) a user intent based on at least the additional input in the audio input, and executing a task associated with the user intent (e.g., with virtual assistant service unit 1824).

[0375] In some embodiments, processing unit 1814 is further configured to, in accordance with a determination that the electronic device is operating in a predetermined mode, forgoing sampling of audio input (e.g., with audio sampling unit 1820), even after receiving (e.g., with input detection unit 1818) the user input (e.g., with touch-sensitive surface unit 1804 or input device unit 1810) meeting the predetermined condition while the display is on.

[0376] In some embodiments, operation in the predetermined mode comprises muting audio output of the electronic device. In some embodiments, the spoken trigger comprises a predetermined phrase, such as “hey Siri.”

[0377] The operations described above with respect to FIG. 14, optionally, implemented by components depicted in FIGS. 1, 2, 7A-7B, or FIG. 18. For example, sampling audio operation 1202 is optionally implemented by audio circuitry 210 (FIG. 2). When audio having a spoken trigger is detected, the device uses display controller 256 (FIG. 2) and/or haptic feedback controller 261 (FIG. 2) to provide corresponding output. It would be clear to a person of ordinary skill in the art how other processes can be implemented based on the components depicted in FIGS. 1, 2, 7A-7B, or FIG. 18.

[0378] It is understood by persons of skill in the art that the functional blocks described in FIG. 19 are, optionally, combined or separated into sub-blocks to implement the principles of the various described embodiments. Therefore, the description herein optionally supports any possible combination or separation or further definition of the functional blocks described herein. For example, units 1802-1812 can have associated “controller” units that are operatively coupled with the respective unit and processing unit 1814 to enable operation. These controller units are not separately illustrated in FIG. 18 but are understood to be within the grasp of one of ordinary skill in the art who is designing a device having units 1802-1812 such as device 1800. The description herein thus optionally supports combination, separation, and/or further definition of the functional blocks described herein.

[0379] In accordance with some embodiments, FIG. 19 shows a functional block diagram of an electronic device 1900 configured in accordance with the principles of the various described embodiments, including those described with reference to FIGS. 8 and 12. The functional blocks of the device are, optionally, implemented by hardware, software, or a combination of hardware and software to carry out the principles of the various described embodiments. It is understood by persons of skill in the art that the functional blocks described in FIG. 19 are, optionally, combined or separated into sub-blocks to implement the principles of the various described embodiments. Therefore, the description herein optionally supports any possible combination or separation or further definition of the functional blocks described herein.

[0380] As shown in FIG. 19, an electronic device 1900 includes a display unit 1902 configured to display a graphical user interface, optional touch-sensitive surface unit 1904 to receive contacts, a microphone unit 1906 to detect audio input, optional movement sensor unit 1908 to obtain movement readings such as acceleration in a time interval, optional input device unit 1910 such as a button, optional feedback unit 1912 to cause feedback such as audio, haptic, and/or visual feedback, and processing unit 1914 coupled to the above-described units. In some embodiments, processing unit 1914 includes display enabling unit 1916, input detection unit 1918, audio sampling unit 1920, audio processing unit 1922, virtual assistant service unit 1924, and, optionally, feedback enabling unit 1926.

[0381] In some embodiments, processing unit 1914 is configured to: sample (e.g., with audio sampling unit 1920) audio input received via the microphone and determine (e.g., with audio processing unit 1922) whether the audio input represents a task; while sampling audio input (e.g., with audio sampling unit 1920), receive (e.g., with input detection unit 1918) instruction to cease sampling of audio input; provide (e.g., with feedback unit 1926) output acknowledging the received instruction to cease the sampling of audio input; and after providing the output, continue to sample audio input (e.g., with audio sampling unit 1920) and determine (e.g., with audio processing unit 1922) whether the audio input comprises a task, for a duration, then cease the sampling of audio input (e.g., with audio sampling unit 1920).

[0382] In some embodiments, processing unit 1914 is further configured to, in accordance with sampling audio input (e.g., with audio sampling unit 1920), sample (e.g., with audio sampling unit 1920) audio input while the display is on.

[0383] In some embodiments, processing unit 1914 is further configured to, in accordance with providing (e.g., with feedback unit 1926) the output acknowledging the received instruction to end the sampling of audio input, cause (e.g., with feedback unit 1926) a haptic output (e.g., with feedback unit 1912).

[0384] In some embodiments, processing unit 1914 is further configured to, in accordance with providing (e.g., with feedback unit 1926) the output acknowledging the received instruction to end the sampling of audio input, provide (e.g., with feedback unit 1926) an audio output (e.g., with feedback unit 1912).

[0385] In some embodiments, processing unit 1914 is further configured to, in accordance with the sampling (e.g., with audio sampling unit 1920) of audio received via the microphone, prompt a user for spoken input (e.g., with display enabling unit 1916 and display unit 1902, and/or with feedback enabling unit 1926 and feedback unit 1912).

[0386] In some embodiments, processing unit 1914 is further configured to, in accordance with determining (e.g., with virtual assistant service unit 1924) whether the audio input represents a task, determine (e.g., with virtual assistant service unit 1924) a user intent based on at least a portion of the audio input; identify and execute the task based on the user intent (e.g., with virtual assistant service unit 1924); and cause (e.g., with feedback unit 1926) haptic output (e.g., with feedback unit 1912) representing execution of the task.

[0387] In some embodiments, processing unit 1914 is further configured to, determine (e.g., with input detection unit 1918) whether the received instruction to cease sampling of audio input is received from a user via the electronic device or is received from an external device; where the continued sampling of audio input (e.g., with audio sampling unit 1920) and determining (e.g., with audio processing unit 1922) of whether the audio input comprises a task for a duration after receiving the instruction to cease sampling occurs in accordance with a determination that the received instruction is received from a user via the electronic device; and in accordance with a determination that the received instruction is received from an external device, the device ceases the sampling of audio input (e.g., with audio sampling unit 1920) responsive to the received instruction.

[0388] In some embodiments, the spoken trigger comprises comprise a predetermined phrase, such as “hey Siri.”

[0389] The operations described above with respect to FIG. 15, optionally, implemented by components depicted in FIGS. 1, 2, 7A-7B, or FIG. 19. For example, sampling audio operation 1202 is optionally implemented by audio circuitry 210 (FIG. 2). When audio having a spoken trigger is detected, the device uses display controller 256 (FIG. 2) and/or haptic feedback controller 261 (FIG. 2) to provide corresponding output. It would be clear to a person of ordinary skill in the art how other processes can be implemented based on the components depicted in FIGS. 1, 2, 7A-7B, or FIG. 19.

[0390] It is understood by persons of skill in the art that the functional blocks described in FIG. 19 are, optionally, combined or separated into sub-blocks to implement the principles

of the various described embodiments. Therefore, the description herein optionally supports any possible combination or separation or further definition of the functional blocks described herein. For example, units 1902-1912 can have associated “controller” units that are operatively coupled with the respective unit and processing unit 1914 to enable operation. These controller units are not separately illustrated in FIG. 19 but are understood to be within the grasp of one of ordinary skill in the art who is designing a device having units 1902-1912 such as device 1900. The description herein thus optionally supports combination, separation, and/or further definition of the functional blocks described herein.

[0391] Although the disclosure and examples have been fully described with reference to the accompanying figures, it is to be noted that various changes and modifications will become apparent to those skilled in the art. Such changes and modifications are to be understood as being included within the scope of the disclosure and examples as defined by the appended claims.

CLAIMS

What is claimed is:

1. A method, comprising:

at an electronic device with a display, a microphone, and an input device:

while the display is on, receiving user input via the input device, the user input meeting a predetermined condition;

in accordance with receiving the user input meeting the predetermined condition, sampling audio input received via the microphone;

determining whether the audio input comprises a spoken trigger; and

in accordance with a determination that audio input comprises the spoken trigger, triggering a virtual assistant session.

2. The method of claim 1, wherein the predetermined condition is based on an amount of lifting of the electronic device during a time interval, the amount of lifting determined from an accelerometer of the electronic device.

3. The method of any one of claims 1 and 2, wherein the predetermined condition is based on smoothness of a lifting of the electronic device during a time interval, the smoothness of the lifting determined from an accelerometer of the electronic device.

4. The method of any one of claims 2 and 3, wherein the predetermined condition comprises a minimum period of dwell time in a position in accordance with the lifting of the electronic device.

5. The method of claim 1, wherein the predetermined condition comprises detecting activation of a button of the electronic device.

6. The method of claim 1, wherein the predetermined condition comprises detecting touch input on a touch-sensitive surface of the electronic device.

7. The method of any one of claims 1-6, further comprising:

determining whether the display is on; and

wherein the sampling of audio input occurs in accordance with a determination that the display is on.

8. The method of claim 6, further comprising:

in accordance with determining whether the display is on, determining whether the backlight of the display is on.

9. The method of any one of claims 1-8, further comprising:

ceasing the sampling of audio input after the sampling of audio input has occurred for a predetermined duration of time.

10. The method of any one of claims 1-8, wherein the user input meeting a predetermined condition is a first user input meeting a first predetermined condition, the method further comprising:

while sampling audio input in accordance with receiving the first user input, receiving a second user input meeting a second predetermined condition; and

in accordance with receiving the second user input meeting the second predetermined condition, ceasing the sampling of audio input.

11. The method of claim 10, wherein the second predetermined condition is based on an amount of lowering of the electronic device during a time interval, the amount of lowering determined from an accelerometer of the electronic device.

12. The method of any one of claims 10 and 11, wherein the first predetermined condition is based on smoothness of a lowering of the electronic device during a time interval, the smoothness of the lowering determined from an accelerometer of the electronic device.

13. The method of claim 10, wherein the second predetermined condition comprises detecting activation of a button of the electronic device.

14. The method of claim 10, wherein the second predetermined condition comprises detecting touch input on a touch-sensitive surface of the electronic device.

15. The method of any one of claims 1-8, further comprising:

while sampling audio input in accordance with receiving the user input meeting the predetermined condition, determining whether the display is off; and

in accordance with a determination that the display is off, ceasing the sampling of audio input.

16. The method of any one of claims 1-8, further comprising:

while sampling audio input in accordance with receiving the user input meeting the predetermined condition, identifying an audio endpoint in the audio input; and

in response to identifying the audio endpoint, ceasing the sampling of audio input.

17. The method of any one of claims 9-16, further comprising:

in accordance with ceasing to sample audio input, providing a haptic output.

18. The method of any one of claims 1-17, further comprising:

in accordance with triggering the virtual assistant session, prompting a user for spoken input.

19. The method of any one of claims 1- 17, wherein the audio input comprises the spoken trigger and additional input, the method further comprising:

in accordance with triggering the virtual assistant session:

determining a user intent based on at least the additional input in the audio input, and

executing a task associated with the user intent.

20. The method of any one of claims 1-19, further comprising:

in accordance with a determination that the electronic device is operating in a predetermined mode, forgoing sampling of audio input, even after receiving the user input meeting the predetermined condition while the display is on,

wherein operation in the predetermined mode comprises muting audio output of the electronic device.

21. The method of any one of claims 1-20, wherein the spoken trigger comprises comprise a predetermined phrase.

22. A method, comprising:

at an electronic device with a display and a microphone:

detecting a software event meeting a predetermined condition;

in accordance with a determination that the software event meeting the condition is detected, sampling audio input received via the microphone;

determining whether audio input received via the microphone comprises a spoken trigger; and

in accordance with a determination that the audio input comprises the spoken trigger, triggering a virtual assistant session.

23. The method of claim 22, wherein the predetermined condition is that the software event is an event triggered by an application of the electronic device.

24. The method of claim 22, wherein the predetermined condition comprises the software event representing a notification.

25. The method of claim 22, wherein the predetermined condition comprises the software event representing a calendar notification.

26. The method of claim 22, wherein the predetermined condition comprises the software event representing an incoming calendar invitation from an external device.

27. The method of claim 22, wherein the predetermined condition comprises the software event representing an incoming transmission, the transmission triggered from an external device.

28. The method of claim 27, wherein the predetermined condition comprises the software event representing an incoming phone call from an external device.

29. The method of claim 27, wherein the predetermined condition comprises the software event representing an incoming text or multimedia message from an external device.

30. The method of any one of claims 26-29 further comprising:

in accordance with triggering a virtual assistant session, initiating a response addressed to the external device.

31. The method of any one of claims 22-30, further comprising:

determining whether the display is on, and

wherein the sampling of audio input occurs in accordance with a determination that the display is on.

32. The method of claim 31, further comprising:

in accordance with determining whether the display is on, determining whether the backlight of the display is on.

33. The method of any one of claims 22-32, further comprising:

ceasing the sampling of audio input after the sampling of audio input has occurred for a predetermined duration of time.

34. The method of any one of claims 22-32, wherein the user input meeting a predetermined condition is a first user input meeting a first predetermined condition, the method further comprising:

while sampling audio input in accordance with receiving the first user input, receiving a second user input meeting a second predetermined condition; and

in accordance with receiving the second user input meeting the second predetermined condition, ceasing the sampling of audio input.

35. The method of claim 34, wherein the second predetermined condition is based on an amount of lowering of the electronic device during a time interval, the amount of lowering determined from an accelerometer of the electronic device.

36. The method of any one of claims 34 and 35, wherein the first predetermined condition is based on smoothness of a lowering of the electronic device during a time interval, the smoothness of the lowering determined from an accelerometer of the electronic device.

37. The method of claim 34, wherein the second predetermined condition comprises detecting activation of a button of the electronic device.

38. The method of claim 34, wherein the second predetermined condition comprises detecting touch input on a touch-sensitive surface of the electronic device.

39. The method of any one of claims 22-29, further comprising:

while sampling audio input in accordance with receiving the user input meeting the predetermined condition, determining whether the display is off; and

in accordance with a determination that the display is off, ceasing the sampling of audio input.

40. The method of any one of claims 22-29, further comprising:

while sampling audio input in accordance with receiving the user input meeting the predetermined condition, identifying an audio endpoint in the audio input; and

in response to identifying the audio endpoint, ceasing the sampling of audio input.

41. The method of any one of claims 30-40, further comprising:

in accordance with ceasing to sample audio input, providing a haptic output.

42. The method of any one of claims 22-41, further comprising:

in accordance with triggering the virtual assistant session, prompting a user for spoken input.

43. The method of any one of claims 22-41, wherein the audio input comprises the spoken trigger and additional input, the method further comprising:

in accordance with triggering the virtual assistant session:

determining a user intent based on at least the additional input in the audio input, and

executing a task associated with the user intent.

44. The method of any one of claims 22-43, further comprising:

in accordance with a determination that the electronic device is operating in a predetermined mode, forgoing sampling of audio input, even after receiving the user input meeting the predetermined condition while the display is on,

wherein operation in the predetermined mode comprises muting audio output of the electronic device.

45. The method of any one of claims 22-44, wherein the spoken trigger comprises a predetermined phrase.

46. A method, comprising:

at an electronic device with a display, a microphone, and an input device:

receiving user input via the input device, the user input meeting a predetermined condition;

in accordance with receiving the user input meeting the predetermined condition, sampling audio input received via the microphone;

determining whether the audio input comprises a spoken trigger;

in accordance with a determination that audio input comprises the spoken trigger:

triggering a virtual assistant session, and

providing a haptic output.

47. The method of claim 46, further comprising:

in accordance with receiving the user input meeting a predetermined condition, receiving the user input while the display is off.

48. The method of any one of claims 46 and 47, further comprising:

in accordance with triggering the virtual assistant session, triggering the virtual assistant session without turning on a backlight of the display

49. The method of any one of claims 46-48, wherein the predetermined condition is based on an amount of lifting of the electronic device during a time interval, the amount of lifting determined from an accelerometer of the electronic device.

50. The method of any one of claims 46-48, wherein the predetermined condition is based on smoothness of a lifting of the electronic device during a time interval, the smoothness of the lifting determined from an accelerometer of the electronic device.

51. The method of any one of claims 49 and 50, wherein the predetermined condition comprises a minimum period of dwell time in a position in accordance with the lifting of the electronic device.

52. The method of any one of claims 46-48, wherein the predetermined condition comprises detecting activation of a button of the electronic device.

53. The method of any one of claims 46-48, wherein the predetermined condition comprises detecting touch input on a touch-sensitive surface of the electronic device.

54. The method of any one of claims 46-53, further comprising:

ceasing the sampling of audio input after the sampling of audio input has occurred for a predetermined duration of time.

55. The method of any one of claims 46-53, wherein the user input meeting a predetermined condition is a first user input meeting a first predetermined condition, the method further comprising:

while sampling audio input in accordance with receiving the first user input, receiving a second user input meeting a second predetermined condition; and

in accordance with receiving the second user input meeting the second predetermined condition, ceasing the sampling of audio input.

56. The method of claim 55, wherein the second predetermined condition is based on an amount of lowering of the electronic device during a time interval, the amount of lowering determined from an accelerometer of the electronic device.

57. The method of any one of claims 55 and 56, wherein the first predetermined condition is based on smoothness of a lowering of the electronic device during a time interval, the smoothness of the lowering determined from an accelerometer of the electronic device.

58. The method of claim 55, wherein the second predetermined condition comprises detecting activation of a button of the electronic device.

59. The method of claim 55, wherein the second predetermined condition comprises detecting touch input on a touch-sensitive surface of the electronic device.

60. The method of any one of claims 46-53, further comprising:

while sampling audio input in accordance with receiving the user input meeting the predetermined condition, identifying an audio endpoint in the audio input; and

in response to identifying the audio endpoint, ceasing the sampling of audio input.

61. The method of any one of claims 54-60, further comprising:

in accordance with ceasing to sample audio input, providing a haptic output.

62. The method of any one of claims 46-61, further comprising:

in accordance with triggering the virtual assistant session, prompting a user for spoken input.

63. The method of any one of claims 46-61, wherein the audio input comprises the spoken trigger and additional input, the method further comprising:

in accordance with triggering the virtual assistant session:

determining a user intent based on at least the additional input in the audio input, and

executing a task associated with the user intent.

64. The method of any one of claims 46-63, further comprising:

in accordance with a determination that the electronic device is operating in a predetermined mode, forgoing sampling of audio input, even after receiving the user input meeting the predetermined condition while the display is on.

wherein operation in the predetermined mode comprises muting audio output of the electronic device.

65. The method of any one of claims 46-64, wherein the spoken trigger comprises comprise a predetermined phrase.

66. A method, comprising:

at an electronic device with a display and a microphone:

sampling audio input received via the microphone and determining whether the audio input represents a task;

while sampling audio input, receiving instruction to cease sampling of audio input;

providing output acknowledging the received instruction to cease the sampling of audio input; and

after providing the output, continuing the sampling of audio input and determining whether the audio input comprises a task, for a duration, then ceasing the sampling of audio input.

67. The method of claim 66, further comprising:

in accordance with sampling audio input, sampling audio input while the display is on.

68. The method of any one of claims 66 and 67, further comprising:

in accordance with providing the output acknowledging the received instruction to end the sampling of audio input, providing a haptic output.

69. The method of any one of claims 66-68, further comprising:

in accordance with providing the output acknowledging the received instruction to end the sampling of audio input, providing an audio output.

70. The method of any one of claims 66-69, further comprising:

in accordance with the sampling of audio received via the microphone, prompting a user for spoken input.

71. The method of any one of claims 66-70, further comprising:

in accordance with determining whether the audio input represents a request, determining a user intent based on at least a portion of the audio input;

identifying and executing a task based on the user intent; and

providing haptic output representing execution of the task.

72. The method of any one of claims 66-71, wherein the spoken trigger comprises comprise a predetermined phrase.

73. The method of any one of claims 66-72, further comprising:

determining whether the received instruction to cease sampling of audio input is received from a user via the electronic device or is received from an external device;

wherein the continuing of the sampling of audio input and the determining of whether the audio input comprises a task, for a duration, occurs in accordance with a determination that the received instruction is received from a user via the electronic device; and

in accordance with a determination that the received instruction is received from an external device, ceasing the sampling of audio input responsive to the received instruction.

74. A non-transitory computer readable storage medium storing one or more programs, the one or more programs comprising instructions, which when executed by an electronic device with a display, a microphone, and an input device, cause the electronic device to:

while the display is on, receive user input via the input device, the user input meeting a predetermined condition;

in accordance with receiving the user input meeting the predetermined condition, sample audio input received via the microphone;

determine whether the audio input comprises a spoken trigger; and

in accordance with a determination that audio input comprises the spoken trigger, trigger a virtual assistant session.

75. A non-transitory computer readable storage medium storing one or more programs, the one or more programs comprising instructions, which when executed by an electronic device with a display, a microphone, and an input device, cause the electronic device to:

perform the methods of any one of claims 1-21.

76. A non-transitory computer readable storage medium storing one or more programs, the one or more programs comprising instructions, which when executed by an electronic device with a display and a microphone, cause the electronic device to:

detect a software event meeting a predetermined condition;

in accordance with a determination that the software event meeting the condition is detected, sample audio input received via the microphone;

determine whether audio input received via the microphone comprises a spoken trigger; and

in accordance with a determination that the audio input comprises the spoken trigger, trigger a virtual assistant session.

77. A non-transitory computer readable storage medium storing one or more programs, the one or more programs comprising instructions, which when executed by an electronic device with a display, a microphone, and an input device, cause the electronic device to:

perform the methods of any one of claims 22-46.

78. A non-transitory computer readable storage medium storing one or more programs, the one or more programs comprising instructions, which when executed by an electronic device with a display, a microphone, and an input device, cause the electronic device to:

receive user input via the input device, the user input meeting a predetermined condition;

in accordance with receiving the user input meeting the predetermined condition,
sample audio input received via the microphone;

determine whether the audio input comprises a spoken trigger;

in accordance with a determination that audio input comprises the spoken trigger:

trigger a virtual assistant session, and

provide a haptic output.

79. A non-transitory computer readable storage medium storing one or more programs, the one or more programs comprising instructions, which when executed by an electronic device with a display, a microphone, and an input device, cause the electronic device to:

perform the methods of any one of claims 46-65.

80. A non-transitory computer readable storage medium storing one or more programs, the one or more programs comprising instructions, which when executed by an electronic device with a display and a microphone, cause the electronic device to:

sample audio input received via the microphone and determine whether the audio input represents a task;

while sampling audio input, receive instruction to cease sampling of audio input;

provide output acknowledging the received instruction to cease the sampling of audio input; and

after providing the output, continue the sampling of audio input and determine whether the audio input comprises a task, for a duration, then cease the sampling of audio input.

81. A non-transitory computer readable storage medium storing one or more programs, the one or more programs comprising instructions, which when executed by an electronic device with a display, a microphone, and an input device, cause the electronic device to:

perform the methods of any one of claims 66-73.

82. An electronic device comprising:

a display;

one or more processors;

a memory; and

one or more programs, wherein the one or more programs are stored in memory and configured to be executed by the one or more processors, the one or more programs including instructions for:

while the display is on, receiving user input via an input device, the user input meeting a predetermined condition;

in accordance with receiving the user input meeting the predetermined condition, sampling audio input received via a microphone;

determining whether the audio input comprises a spoken trigger; and

in accordance with a determination that audio input comprises the spoken trigger, triggering a virtual assistant session.

83. An electronic device comprising:

a display;

one or more processors;

a memory; and

one or more programs, wherein the one or more programs are stored in memory and configured to be executed by the one or more processors, the one or more programs including instructions for perform the methods of any one of claims 1-21.

84. An electronic device comprising:

a display;

one or more processors;

a memory; and

one or more programs, wherein the one or more programs are stored in memory and configured to be executed by the one or more processors, the one or more programs including instructions for:

detecting a software event meeting a predetermined condition;

in accordance with a determination that the software event meeting the condition is detected, sampling audio input received via a microphone;

determining whether audio input received via the microphone comprises a spoken trigger; and

in accordance with a determination that the audio input comprises the spoken trigger, triggering a virtual assistant session.

85. An electronic device comprising:

a display;

one or more processors;

a memory; and

one or more programs, wherein the one or more programs are stored in memory and configured to be executed by the one or more processors, the one or more programs including instructions for perform the methods of any one of claims 22-45.

86. An electronic device comprising:

a display;

one or more processors;

a memory; and

one or more programs, wherein the one or more programs are stored in memory and configured to be executed by the one or more processors, the one or more programs including instructions for:

receiving user input via the input device, the user input meeting a predetermined condition;

in accordance with receiving the user input meeting the predetermined condition, sampling audio input received via a microphone;

determining whether the audio input comprises a spoken trigger;

in accordance with a determination that audio input comprises the spoken trigger:

triggering a virtual assistant session, and

enabling a haptic output.

87. An electronic device comprising:

a display;

one or more processors;

a memory; and

one or more programs, wherein the one or more programs are stored in memory and configured to be executed by the one or more processors, the one or more programs including instructions for perform the methods of any one of claims 46-65.

88. An electronic device comprising:

a display;

one or more processors;

a memory; and

one or more programs, wherein the one or more programs are stored in memory and configured to be executed by the one or more processors, the one or more programs including instructions for:

sampling audio input received via a microphone and determine whether the audio input represents a task;

while sampling audio input, receiving instruction to cease sampling of audio input;

enabling output acknowledging the received instruction to cease the sampling of audio input; and

after enabling the output, continuing the sampling of audio input and determining whether the audio input comprises a task, for a duration, then cease the sampling of audio input.

89. An electronic device comprising:

a display;

one or more processors;

a memory; and

one or more programs, wherein the one or more programs are stored in memory and configured to be executed by the one or more processors, the one or more programs including instructions for perform the methods of any one of claims 66-73.

90. An electronic device comprising:

while a displaying means is on, means for receiving user input via the input device, the user input meeting a predetermined condition;

means for, in accordance with receiving the user input meeting the predetermined condition, sampling audio input received via a microphone;

means for determining whether the audio input comprises a spoken trigger; and

means for, in accordance with a determination that audio input comprises the spoken trigger, triggering a virtual assistant session.

91. An electronic device comprising:

means for performing the methods of any one of claims 1-21.

92. An electronic device comprising:

means for detecting a software event meeting a predetermined condition;

means for, in accordance with a determination that the software event meeting the condition is detected, sampling audio input received via a microphone;

means for determining whether audio input received via the microphone comprises a spoken trigger; and

means for, in accordance with a determination that the audio input comprises the spoken trigger, triggering a virtual assistant session.

93. An electronic device comprising:

means for performing the methods of any one of claims 22-45.

94. An electronic device comprising:

means for receiving user input via an input device, the user input meeting a predetermined condition;

means for, in accordance with receiving the user input meeting the predetermined condition, sampling audio input received via a microphone;

means for determining whether the audio input comprises a spoken trigger;

means for, in accordance with a determination that audio input comprises the spoken trigger:

triggering a virtual assistant session, and

providing a haptic output.

95. An electronic device comprising:

means for performing the methods of any one of claims 46-65.

96. An electronic device comprising:

means for sampling audio input received via a microphone and determining whether the audio input represents a task;

means for, while sampling audio input, receiving instruction to cease sampling of audio input;

means for providing output acknowledging the received instruction to cease the sampling of audio input; and

means for, after providing the output, continuing the sampling of audio input and determining whether the audio input comprises a task, for a duration, then ceasing the sampling of audio input.

97. An electronic device comprising:

means for performing the methods of any one of claims 66-73.

98. An electronic device comprising:

a display; a microphone; an input device; and

a processing unit coupled to the display, the microphone, and the input device, the processing unit configured to:

while the display is on, receive user input via the input device, the user input meeting a predetermined condition;

in accordance with receiving the user input meeting the predetermined condition, sample audio input received via the microphone;

determine whether the audio input comprises a spoken trigger; and

in accordance with a determination that audio input comprises the spoken trigger, trigger a virtual assistant session.

99. The electronic device of claim 98, wherein the predetermined condition is based on an amount of lifting of the electronic device during a time interval, the amount of lifting determined from an accelerometer of the electronic device.

100. The electronic device of any one of claims 98 and 99, wherein the predetermined condition is based on smoothness of a lifting of the electronic device during a time interval, the smoothness of the lifting determined from an accelerometer of the electronic device.

101. The electronic device of any one of claims 98 and 99, wherein the predetermined condition comprises a minimum period of dwell time in a position in accordance with the lifting of the electronic device.

102. The electronic device of claim 98, wherein the predetermined condition comprises detecting activation of a button of the electronic device.

103. The electronic device of claim 98, wherein the predetermined condition comprises detecting touch input on a touch-sensitive surface of the electronic device.

104. The electronic device of any one of claims 98-103, the processing unit further configured to:

determine whether the display is on; and

wherein the sampling of audio input occurs in accordance with a determination that the display is on.

105. The electronic device of claim 104, the processing unit further configured to:

in accordance with determining whether the display is on, determine whether the backlight of the display is on.

106. The electronic device of any one of claims 98-105, the processing unit further configured to:

cease the sampling of audio input after the sampling of audio input has occurred for a predetermined duration of time.

107. The electronic device of any one of claims 98-105, wherein the user input meeting a predetermined condition is a first user input meeting a first predetermined condition, wherein the processing unit is further configured to:

while sampling audio input in accordance with receiving the first user input, receive a second user input meeting a second predetermined condition; and

in accordance with receiving the second user input meeting the second predetermined condition, cease the sampling of audio input.

108. The electronic device of claim 107, wherein the second predetermined condition is based on an amount of lowering of the electronic device during a time interval, the amount of lowering determined from an accelerometer of the electronic device.

109. The electronic device of any one of claims 107 and 108, wherein the first predetermined condition is based on smoothness of a lowering of the electronic device during a time interval, the smoothness of the lowering determined from an accelerometer of the electronic device.

110. The electronic device of claim 107, wherein the second predetermined condition comprises detecting activation of a button of the electronic device.

111. The electronic device of claim 107, wherein the second predetermined condition comprises detecting touch input on a touch-sensitive surface of the electronic device.

112. The electronic device of any one of claims 98-105, the processing unit further configured to:

while sampling audio input in accordance with receiving the user input meeting the predetermined condition, determine whether the display is off; and

in accordance with a determination that the display is off, cease the sampling of audio input.

113. The electronic device of any one of claims 98-105, the processing unit further configured to:

while sampling audio input in accordance with receiving the user input meeting the predetermined condition, identify an audio endpoint in the audio input; and

in response to identifying the audio endpoint, cease the sampling of audio input.

114. The electronic device of any one of claims 106-113, the processing unit further configured to:

in accordance with ceasing to sample audio input, causing a haptic output.

115. The electronic device of any one of claims 98-114, the processing unit further configured to:

in accordance with triggering the virtual assistant session, prompt a user for spoken input.

116. The electronic device of any one of claims 98-114, wherein the audio input comprises the spoken trigger and additional input, wherein the processing unit is further configured to:

in accordance with triggering the virtual assistant session:

determine a user intent based on at least the additional input in the audio input, and

execute a task associated with the user intent.

117. The electronic device of any one of claims 98-116, the processing unit further configured to:

in accordance with a determination that the electronic device is operating in a predetermined mode, forgo sampling of audio input, even after receiving the user input meeting the predetermined condition while the display is on,

wherein operation in the predetermined mode comprises muting audio output of the electronic device.

118. The electronic device of any one of claims 98-117, wherein the spoken trigger comprises comprise a predetermined phrase.

119. An electronic device comprising:

a display; a microphone; and

a processing unit coupled to the display, the microphone, and the input device, the processing unit configured to:

detect a software event meeting a predetermined condition;

in accordance with a determination that the software event meeting the condition is detected, sample audio input received via the microphone;

determine whether audio input received via the microphone comprises a spoken trigger; and

in accordance with a determination that the audio input comprises the spoken trigger, trigger a virtual assistant session.

120. The electronic device of claim 119, wherein the predetermined condition is that the software event is an event triggered by an application of the electronic device.

121. The electronic device of claim 119, wherein the predetermined condition comprises the software event representing a notification.

122. The electronic device of claim 119, wherein the predetermined condition comprises the software event representing a calendar notification.

123. The electronic device of claim 119, wherein the predetermined condition comprises the software event representing an incoming calendar invitation from an external device.

124. The electronic device of claim 119, wherein the predetermined condition comprises the software event representing an incoming transmission, the transmission triggered from an external device.

125. The electronic device of claim 124, wherein the predetermined condition comprises the software event representing an incoming phone call from an external device.

126. The electronic device of claim 124, wherein the predetermined condition comprises the software event representing an incoming text or multimedia message from an external device.

127. The electronic device of any one of claims 119-126, the processing unit further configured to:

in accordance with triggering a virtual assistant session, initiate a response addressed to the external device.

128. The electronic device of any one of claims 119-127, the processing unit further configured to:

determine whether the display is on, and

wherein the sampling of audio input occurs in accordance with a determination that the display is on.

129. The electronic device of claim 128, the processing unit further configured to:

in accordance with determining whether the display is on, determine whether the backlight of the display is on.

130. The electronic device of any one of claims 119-129, the processing unit further configured to:

cease the sampling of audio input after the sampling of audio input has occurred for a predetermined duration of time.

131. The electronic device of any one of claims 119-129, wherein the user input meeting a predetermined condition is a first user input meeting a first predetermined condition, wherein the processing unit is further configured to:

while sampling audio input in accordance with receiving the first user input, receive a second user input meeting a second predetermined condition; and

in accordance with receiving the second user input meeting the second predetermined condition, cease the sampling of audio input.

132. The electronic device of claim 131, wherein the second predetermined condition is based on an amount of lowering of the electronic device during a time interval, the amount of lowering determined from an accelerometer of the electronic device.

133. The electronic device of any one of claims 131 and 132, wherein the first predetermined condition is based on smoothness of a lowering of the electronic device during a time interval, the smoothness of the lowering determined from an accelerometer of the electronic device.

134. The electronic device of claim 131, wherein the second predetermined condition comprises detecting activation of a button of the electronic device.

135. The electronic device of claim 131, wherein the second predetermined condition comprises detecting touch input on a touch-sensitive surface of the electronic device.

136. The electronic device of any one of claims 119-135, the processing unit further configured to:

while sampling audio input in accordance with receiving the user input meeting the predetermined condition, determine whether the display is off; and

in accordance with a determination that the display is off, cease the sampling of audio input.

137. The electronic device of any one of claims 119-136, the processing unit further configured to:

while sampling audio input in accordance with receiving the user input meeting the predetermined condition, identify an audio endpoint in the audio input; and

in response to identifying the audio endpoint, cease the sampling of audio input.

138. The electronic device of any one of claims 130-137, the processing unit further configured to:

in accordance with ceasing to sample audio input, causing a haptic output.

139. The electronic device of any one of claims 119-138, the processing unit further configured to:

in accordance with triggering the virtual assistant session, prompt a user for spoken input.

140. The electronic device of any one of claims 119-139, wherein the audio input comprises the spoken trigger and additional input, wherein the processing unit is further configured to:

in accordance with triggering the virtual assistant session:

determine a user intent based on at least the additional input in the audio input,
and

execute a task associated with the user intent.

141. The electronic device of any one of claims 119-140, the processing unit further configured to:

in accordance with a determination that the electronic device is operating in a predetermined mode, forgo sampling of audio input, even after receiving the user input meeting the predetermined condition while the display is on,

wherein operation in the predetermined mode comprises muting audio output of the electronic device.

142. The electronic device of any one of claims 119-141, wherein the spoken trigger comprises comprise a predetermined phrase.

143. An electronic device comprising:

a display; a microphone; an input device; and

a processing unit coupled to the display, the microphone, and the input device, the processing unit configured to:

receive user input via the input device, the user input meeting a predetermined condition;

in accordance with receiving the user input meeting the predetermined condition, sample audio input received via the microphone;

determine whether the audio input comprises a spoken trigger;
in accordance with a determination that audio input comprises the spoken trigger:
trigger a virtual assistant session, and
causing a haptic output.

144. The electronic device of claim 143, the processing unit further configured to:

in accordance with receiving the user input meeting a predetermined condition, receive the user input while the display is off.

145. The electronic device of any one of claims 143 and 144, the processing unit further configured to:

in accordance with triggering the virtual assistant session, trigger the virtual assistant session without turning on a backlight of the display.

146. The electronic device of any one of claims 143-145, wherein the predetermined condition is based on an amount of lifting of the electronic device during a time interval, the amount of lifting determined from an accelerometer of the electronic device.

147. The electronic device of any one of claims 140-142, wherein the predetermined condition is based on smoothness of a lifting of the electronic device during a time interval, the smoothness of the lifting determined from an accelerometer of the electronic device.

148. The electronic device of any one of claims 143 and 144, wherein the predetermined condition comprises a minimum period of dwell time in a position in accordance with the lifting of the electronic device.

149. The electronic device of any one of claims 143-145 wherein the predetermined condition comprises detecting activation of a button of the electronic device.

150. The electronic device of any one of claims 143-145, wherein the predetermined condition comprises detecting touch input on a touch-sensitive surface of the electronic device.

151. The electronic device of any one of claims 143-150, the processing unit further configured to:

cease the sampling of audio input after the sampling of audio input has occurred for a predetermined duration of time.

152. The electronic device of any one of claims 130-151, wherein the user input meeting a predetermined condition is a first user input meeting a first predetermined condition, wherein the processing unit is further configured to:

while sampling audio input in accordance with receiving the first user input, receive a second user input meeting a second predetermined condition; and

in accordance with receiving the second user input meeting the second predetermined condition, cease the sampling of audio input.

153. The electronic device of claim 149, wherein the second predetermined condition is based on an amount of lowering of the electronic device during a time interval, the amount of lowering determined from an accelerometer of the electronic device.

154. The electronic device of any one of claims 149 and 150, wherein the first predetermined condition is based on smoothness of a lowering of the electronic device during a time interval, the smoothness of the lowering determined from an accelerometer of the electronic device.

155. The electronic device of claim 149, wherein the second predetermined condition comprises detecting activation of a button of the electronic device.

156. The electronic device of claim 149, wherein the second predetermined condition comprises detecting touch input on a touch-sensitive surface of the electronic device.

157. The electronic device of any one of claims 143-156, the processing unit further configured to:

while sampling audio input in accordance with receiving the user input meeting the predetermined condition, identify an audio endpoint in the audio input; and

in response to identifying the audio endpoint, cease the sampling of audio input.

158. The electronic device of any one of claims 151-157, the processing unit further configured to:

in accordance with ceasing to sample audio input, causing a haptic output.

159. The electronic device of any one of claims 143-158, the processing unit further configured to:

in accordance with triggering the virtual assistant session, prompt a user for spoken input.

160. The electronic device of any one of claims 143-158, wherein the audio input comprises the spoken trigger and additional input, wherein the processing unit is further configured to:

in accordance with triggering the virtual assistant session:

determine a user intent based on at least the additional input in the audio input,
and

execute a task associated with the user intent.

161. The electronic device of any one of claims 143-160 the processing unit further configured to:

in accordance with a determination that the electronic device is operating in a predetermined mode, forgo sampling of audio input, even after receiving the user input meeting the predetermined condition while the display is on,

wherein operation in the predetermined mode comprises muting audio output of the electronic device.

162. The electronic device of any one of claims 140-158, wherein the spoken trigger comprises comprise a predetermined phrase.

163. An electronic device comprising:

a display; a microphone; a input device and

a processing unit coupled to the display, the microphone, and the input device, the processing unit configured to:

sample audio input received via the microphone and determine whether the audio input represents a task;

while sampling audio input, receive instruction to cease sampling of audio input;

causing output acknowledging the received instruction to cease the sampling of audio input; and

after providing the output, continue the sampling of audio input and determine whether the audio input comprises a task, for a duration, then cease the sampling of audio input.

164. The electronic device of claim 163, the processing unit further configured to:

in accordance with sampling audio input, sample audio input while the display is on.

165. The electronic device of any one of claims 163 and 164, the processing unit further configured to:

in accordance with providing the output acknowledging the received instruction to end the sampling of audio input, provide a haptic output.

166. The electronic device of any one of claims 163-165 the processing unit further configured to:

in accordance with providing the output acknowledging the received instruction to end the sampling of audio input, provide an audio output.

167. The electronic device of any one of claims 163-166, the processing unit further configured to:

in accordance with the sampling of audio received via the microphone, prompt a user for spoken input.

168. The electronic device of any one of claims 163-167, the processing unit further configured to:

in accordance with determining whether the audio input represents a task, determine a user intent based on at least a portion of the audio input;

identify and execute the task based on the user intent; and

provide haptic output representing execution of the task.

169. The electronic device of any one of claims 163-168, wherein the spoken trigger comprises comprise a predetermined phrase.

170. The electronic device of any one of claims 163-169, the processing unit further configured to:

determine whether the received instruction to cease sampling of audio input is received from a user via the electronic device or is received from an external device;

wherein the continuing of the sampling of audio input and the determining of whether the audio input comprises a task, for a duration, occurs in accordance with a determination that the received instruction is received from a user via the electronic device; and

in accordance with a determination that the received instruction is received from an external device, cease the sampling of audio input responsive to the received instruction.

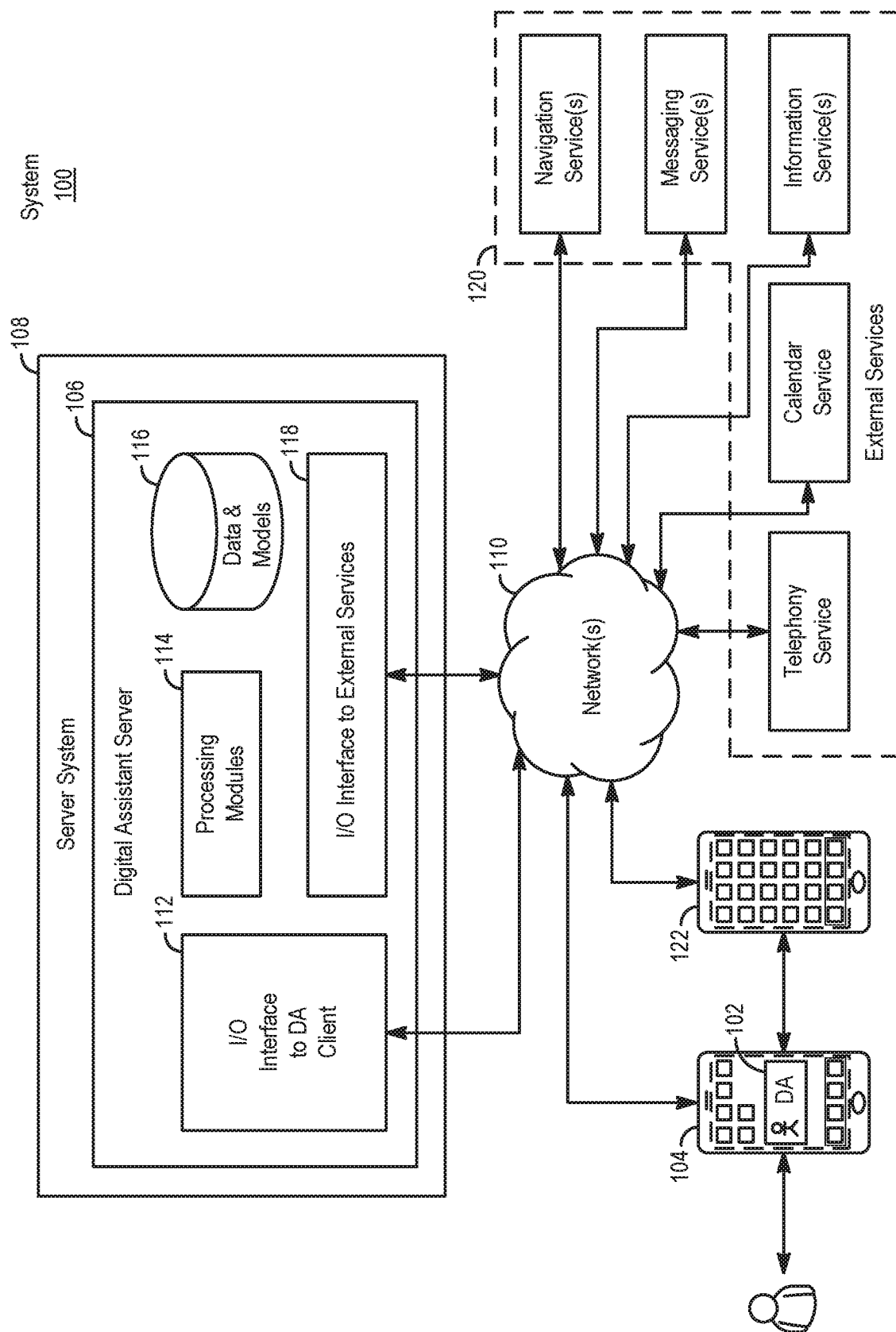


FIG. 1

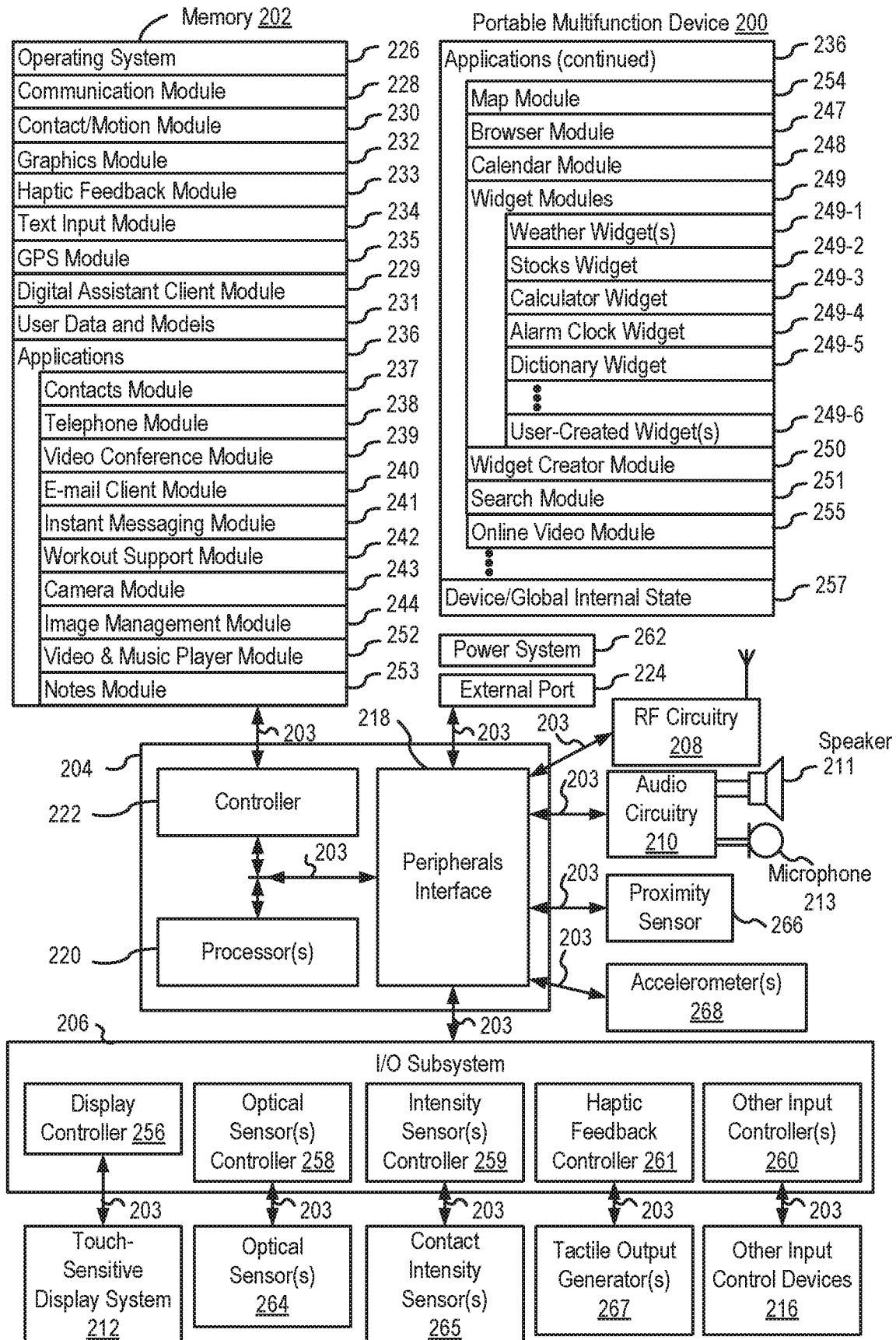


FIG. 2A

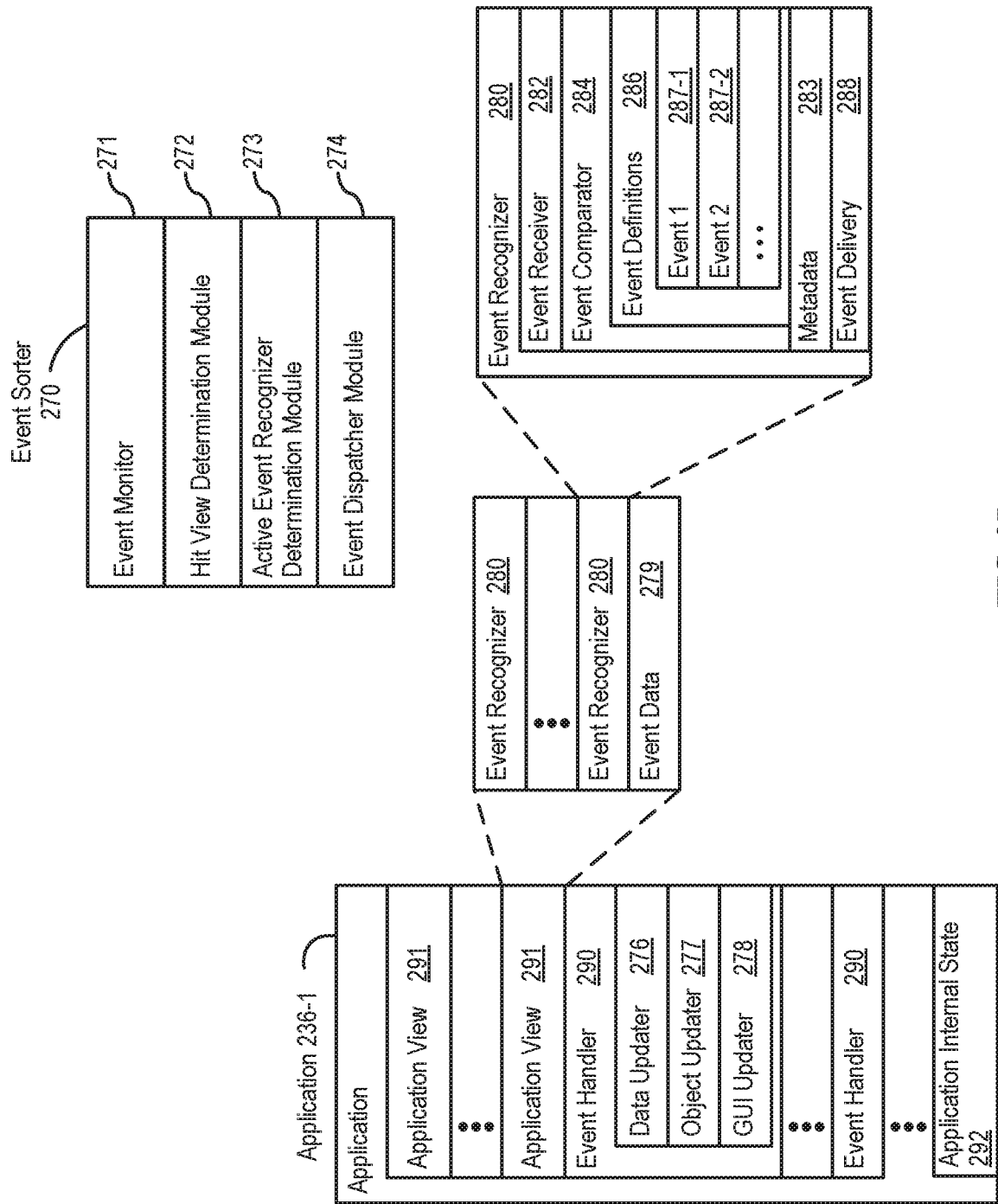


FIG. 2B

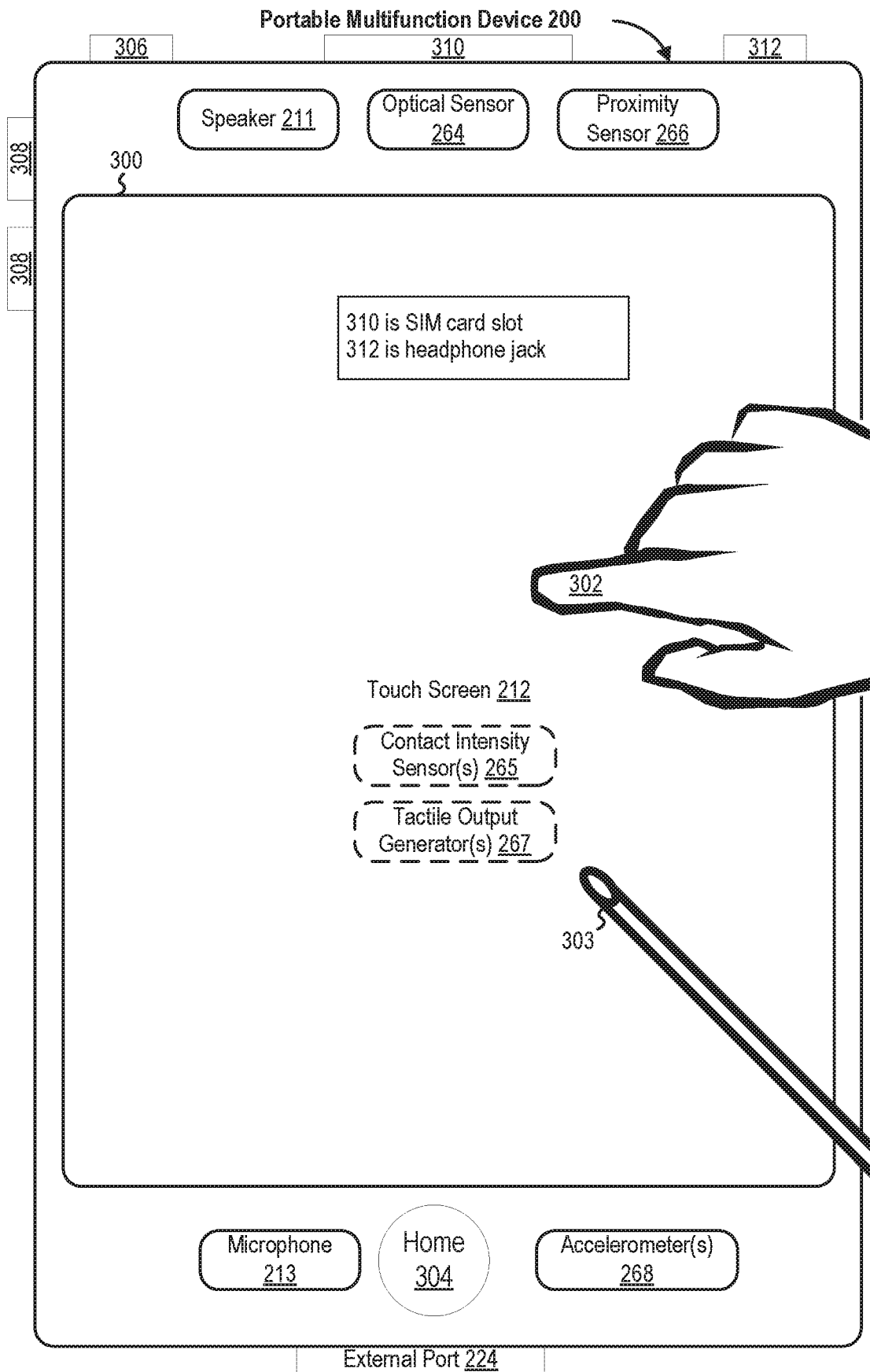


FIG. 3

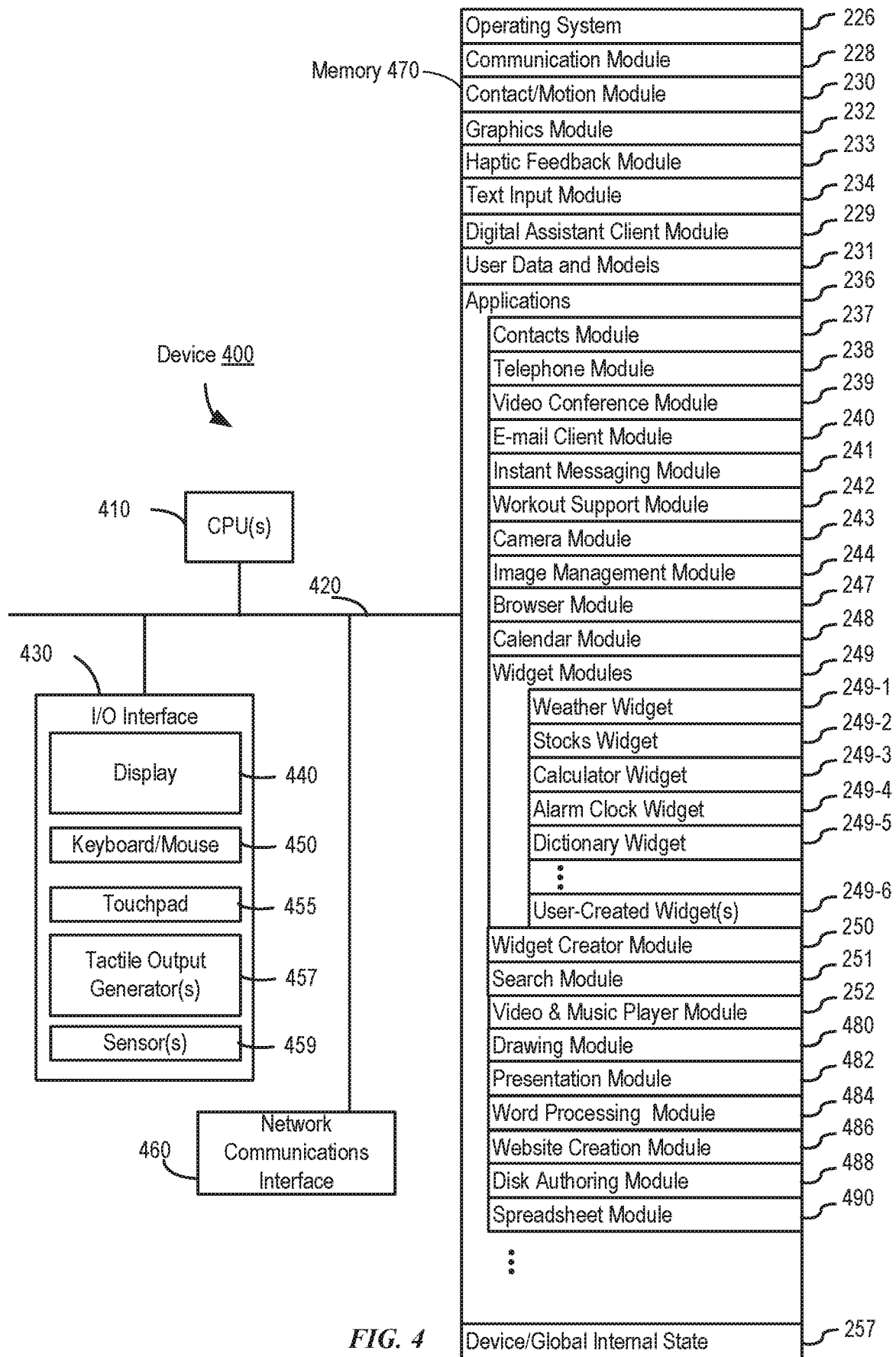


FIG. 4

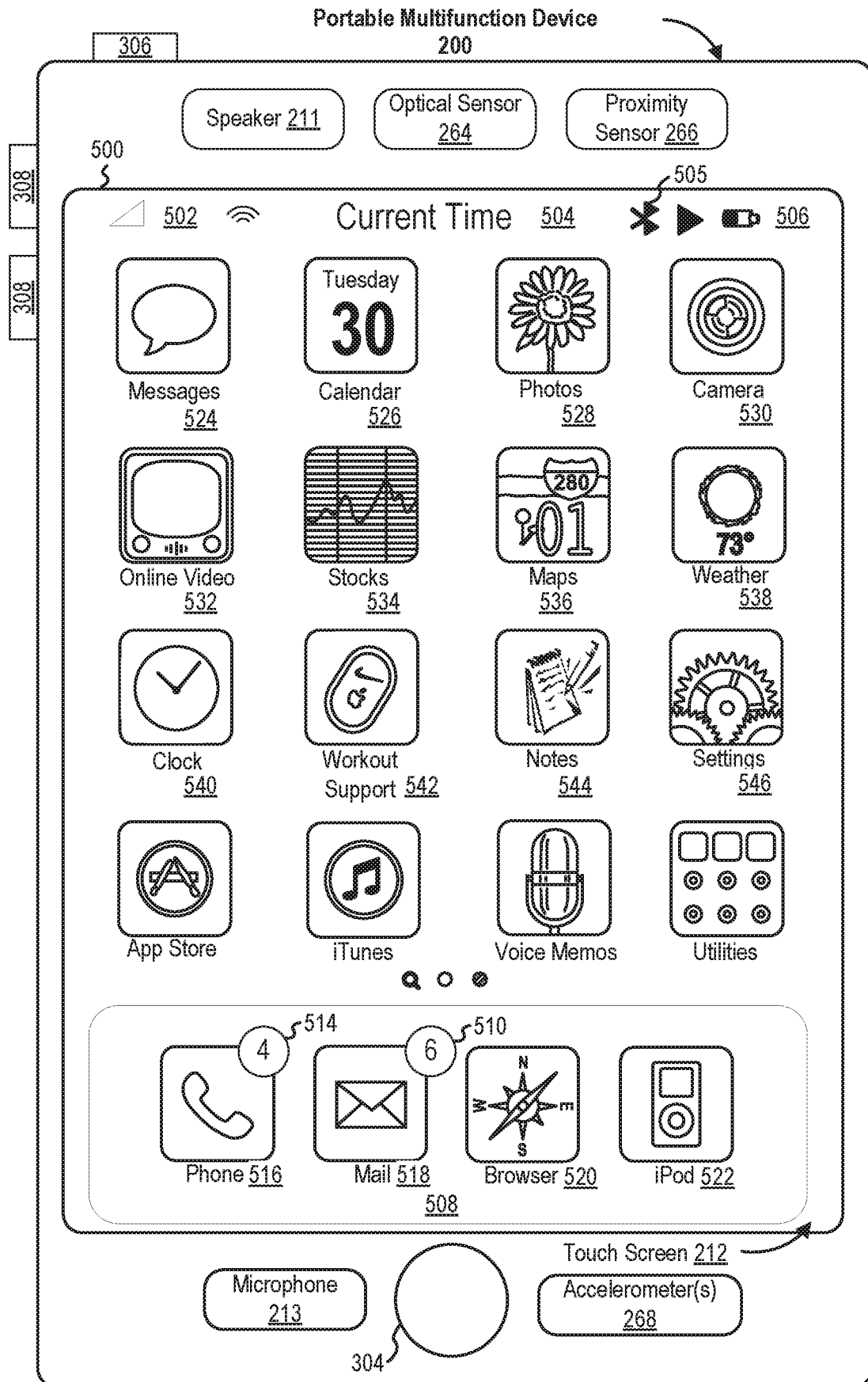


FIG. 5A

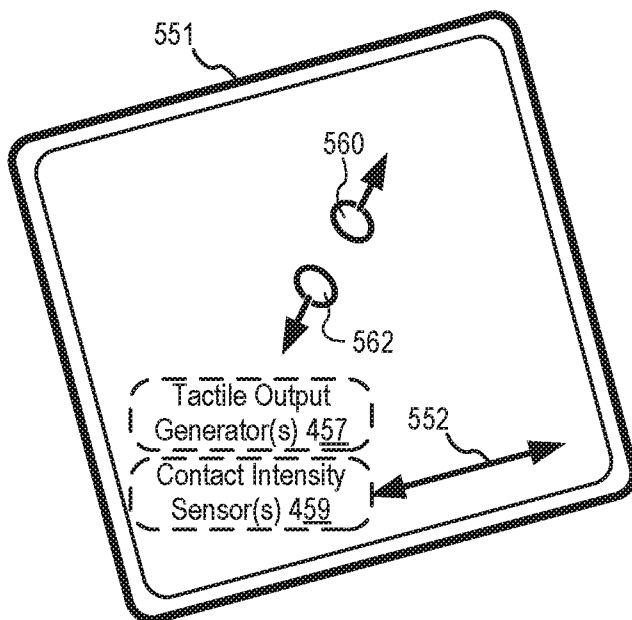
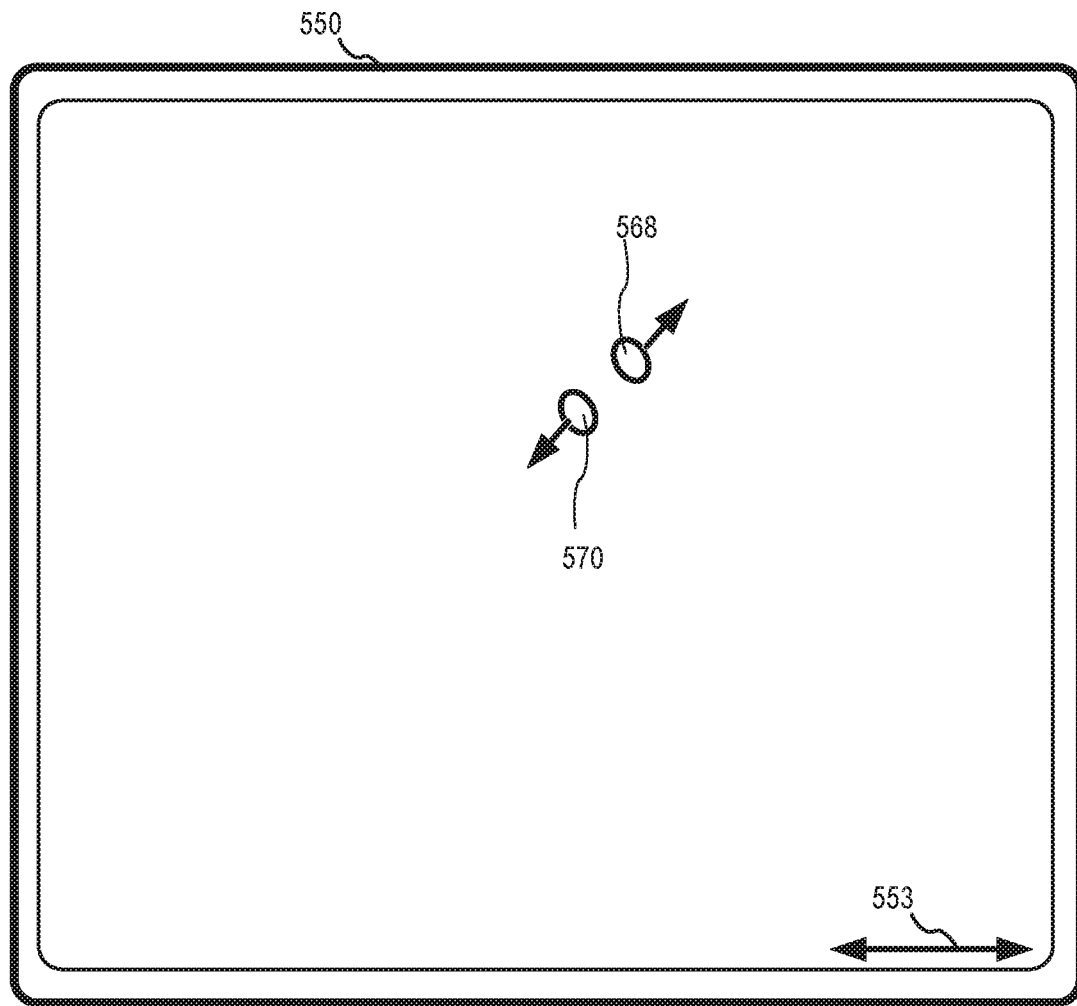
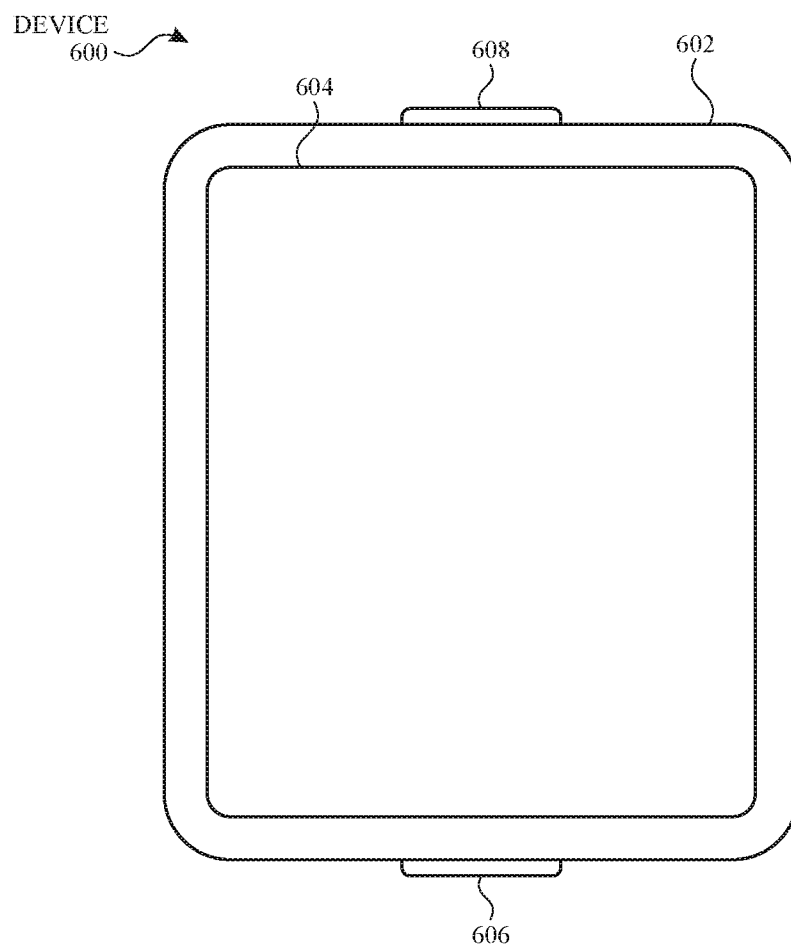


FIG. 5B

**FIG. 6A**

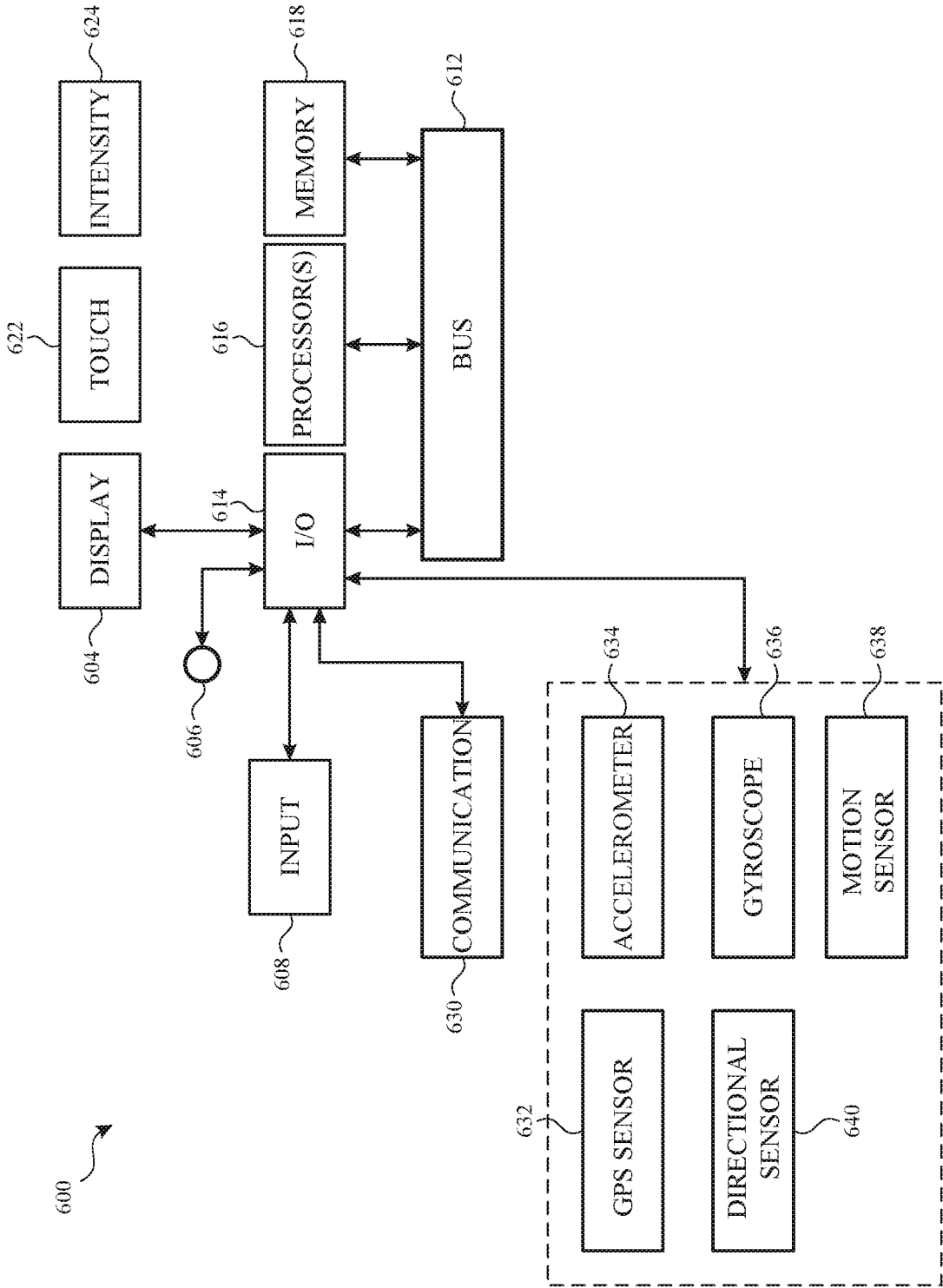


FIG. 6B

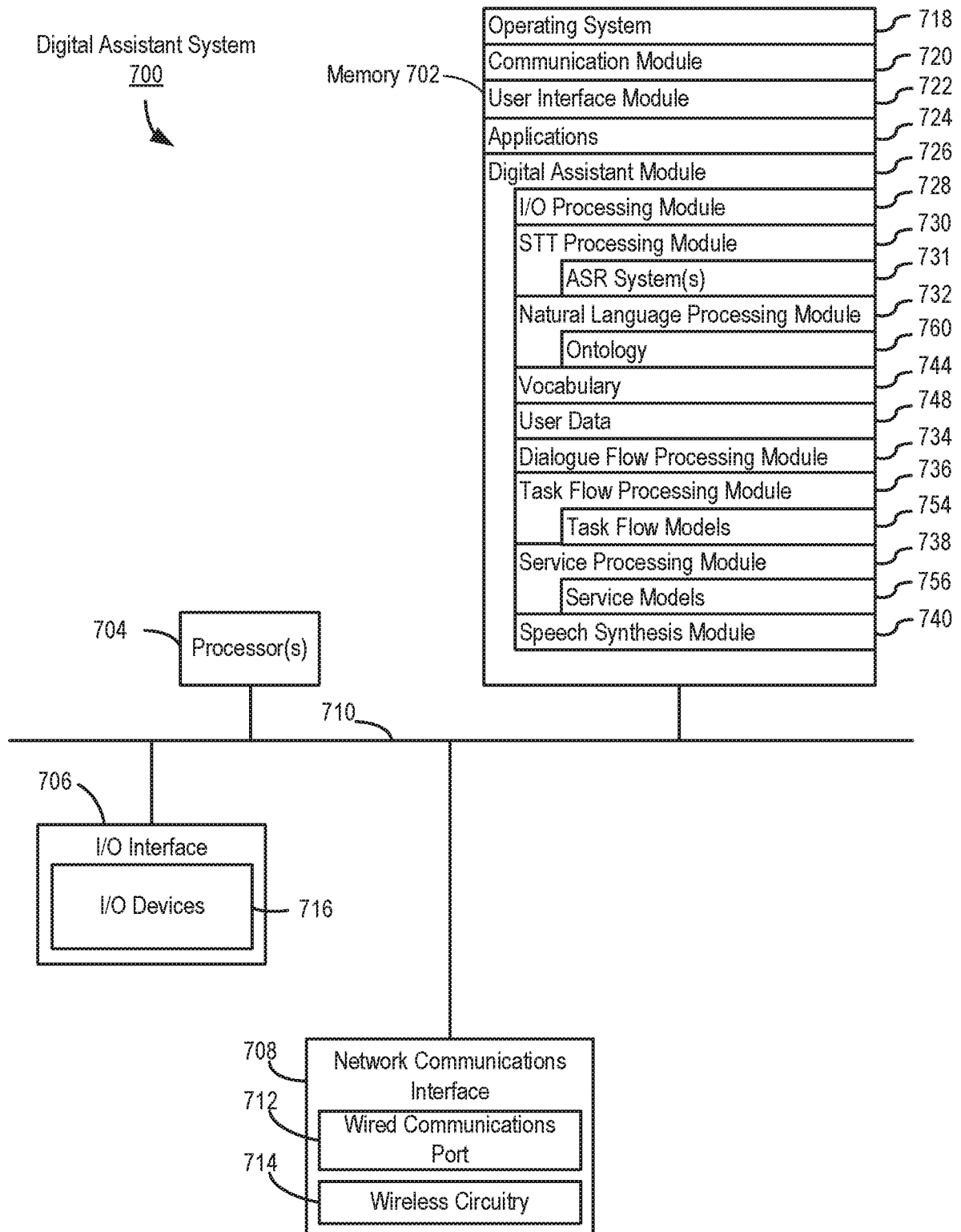


FIG. 7A

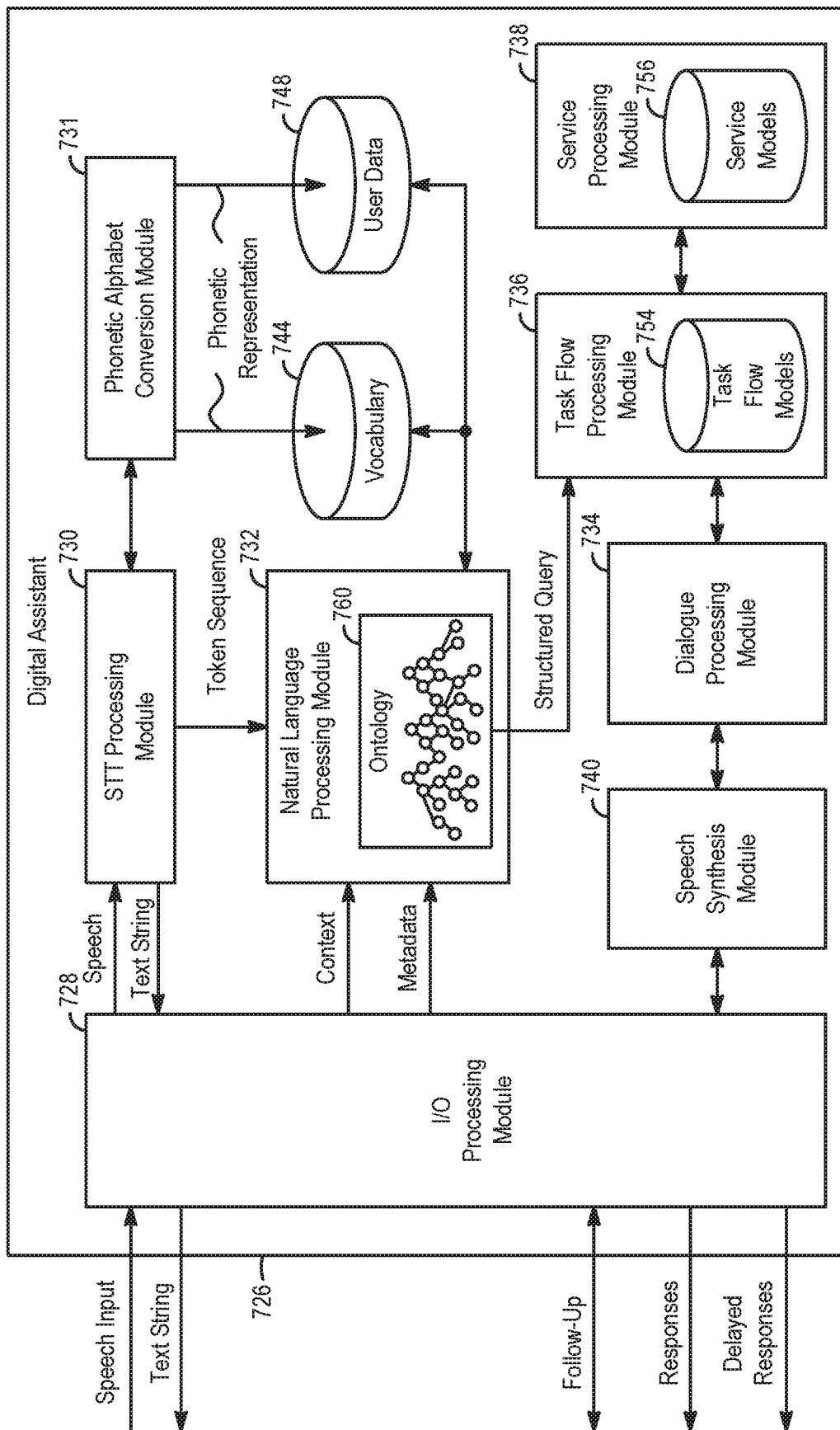


FIG. 7B

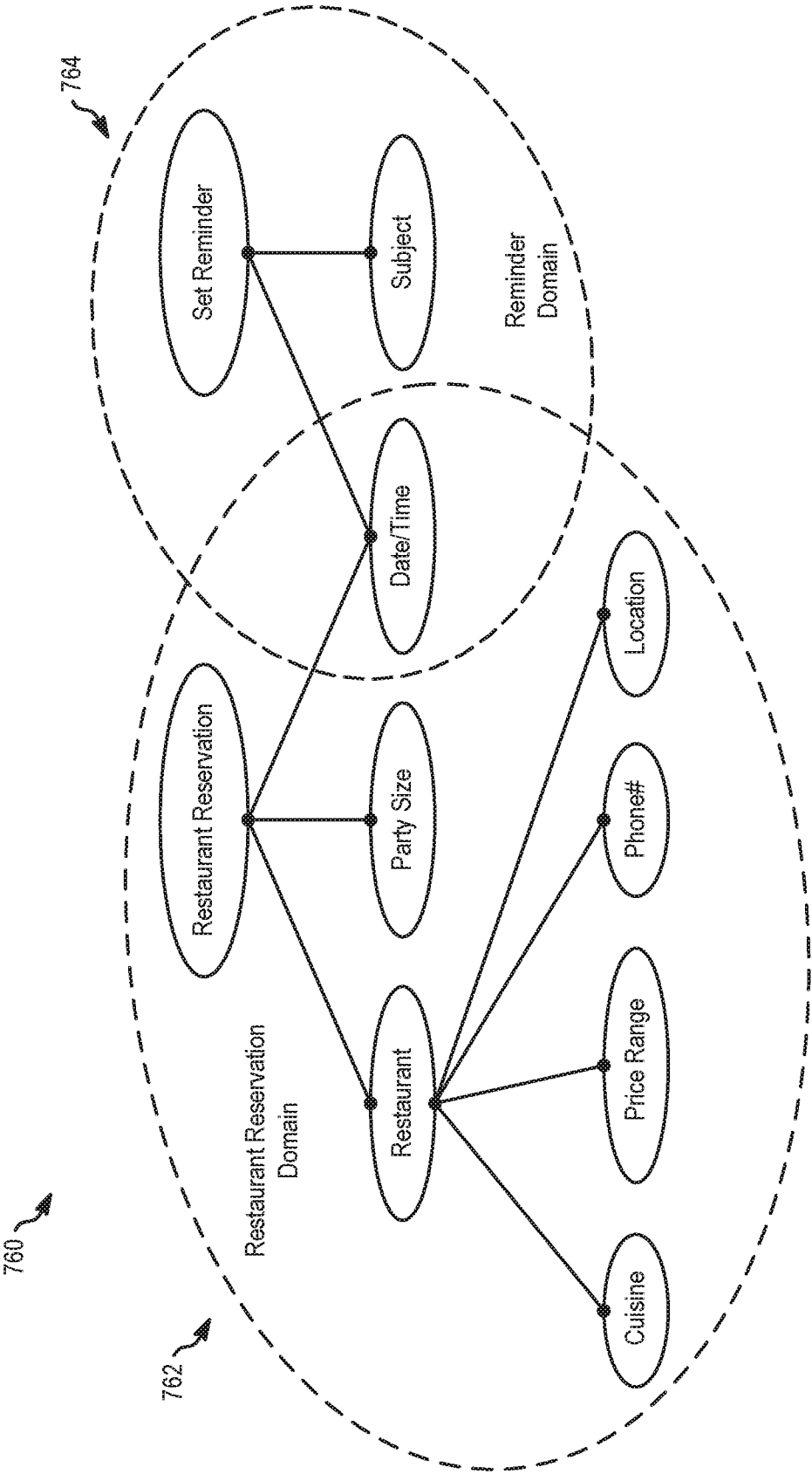
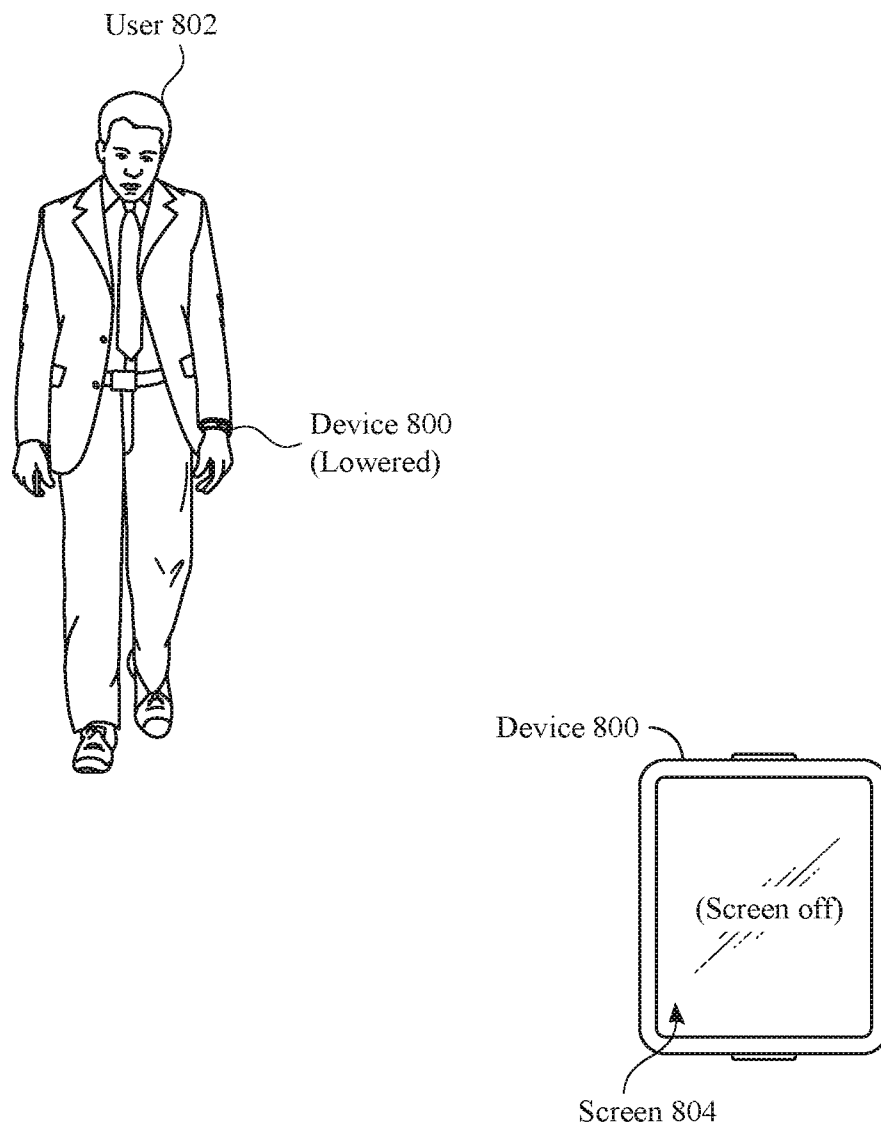


FIG. 7C

**FIG. 8A**

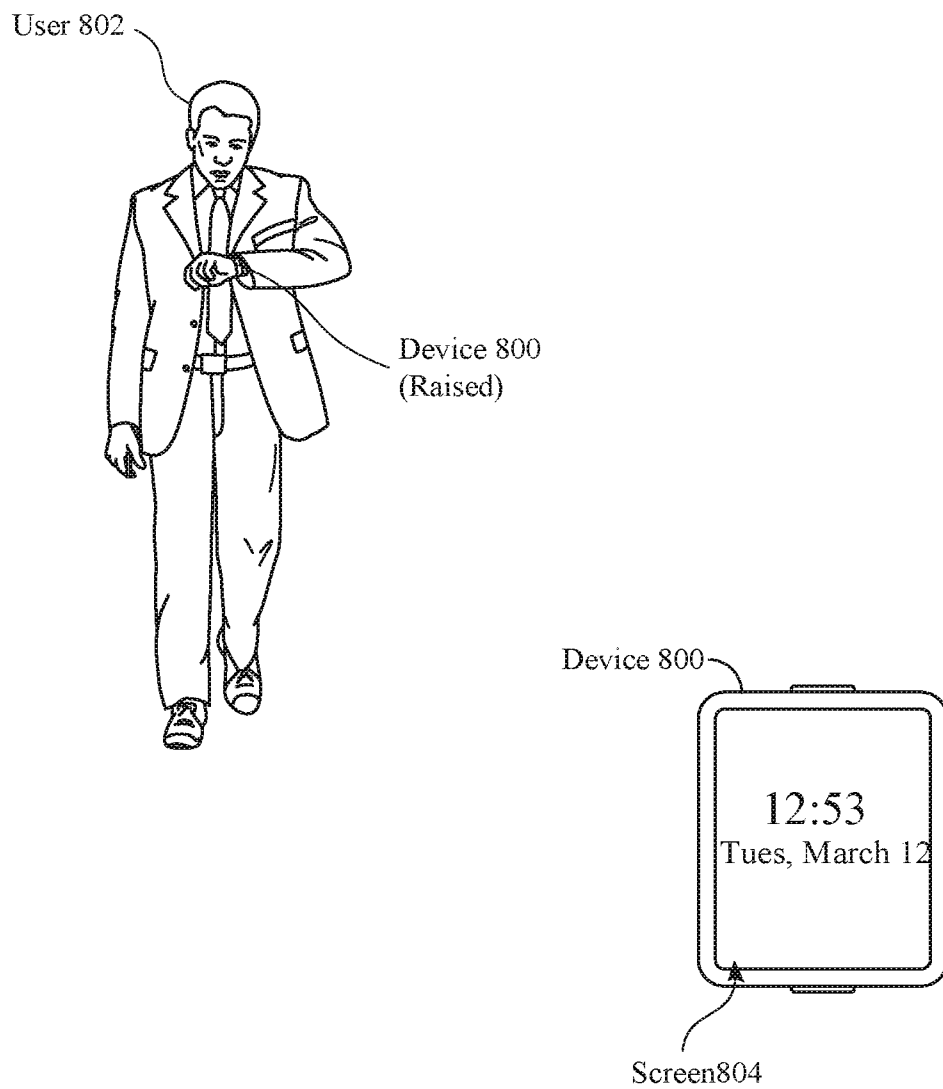
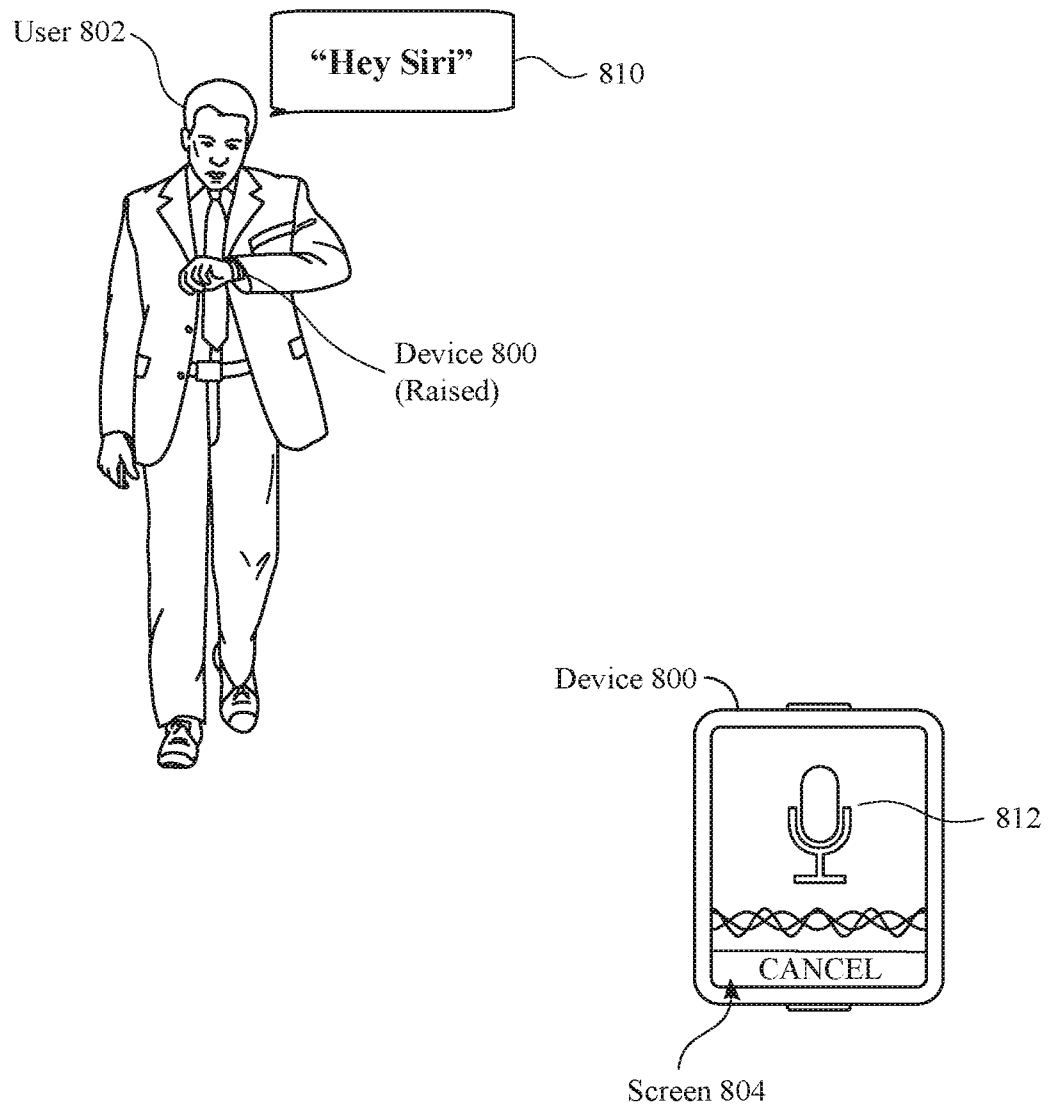
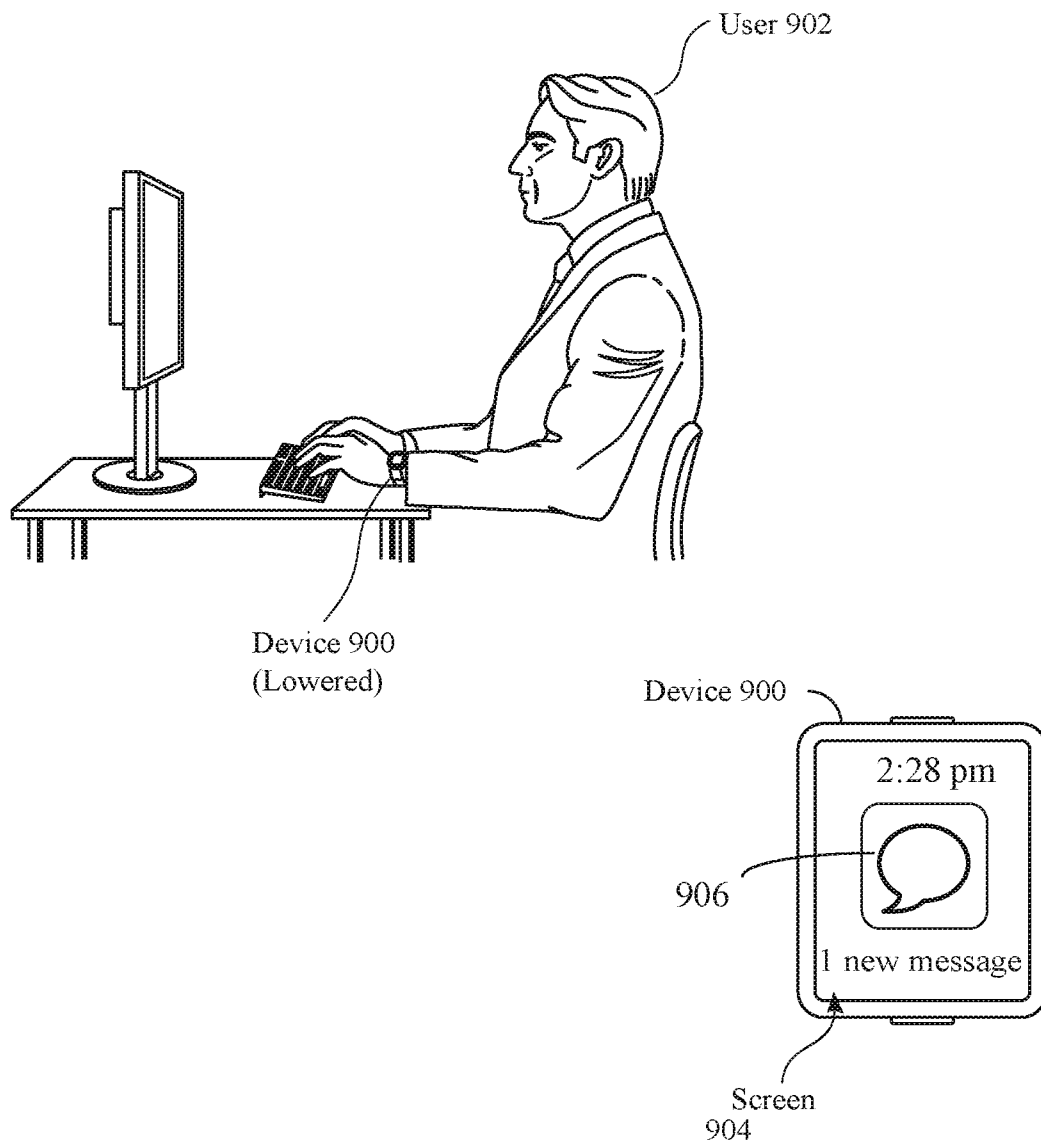
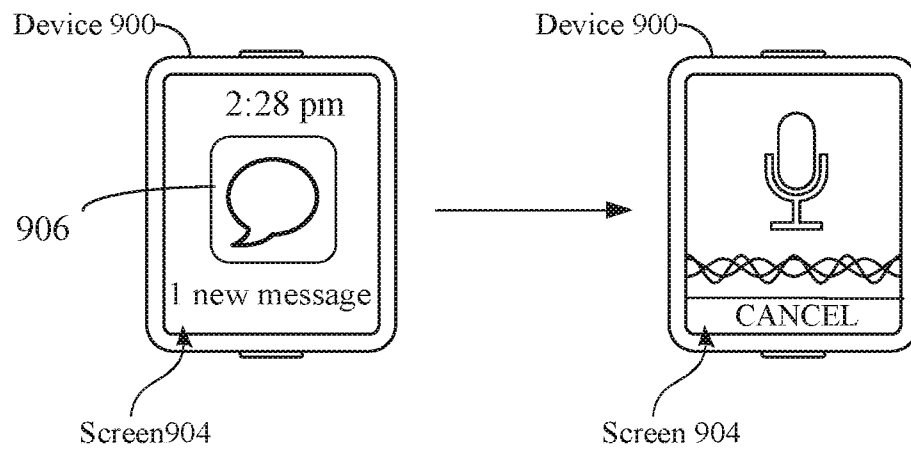
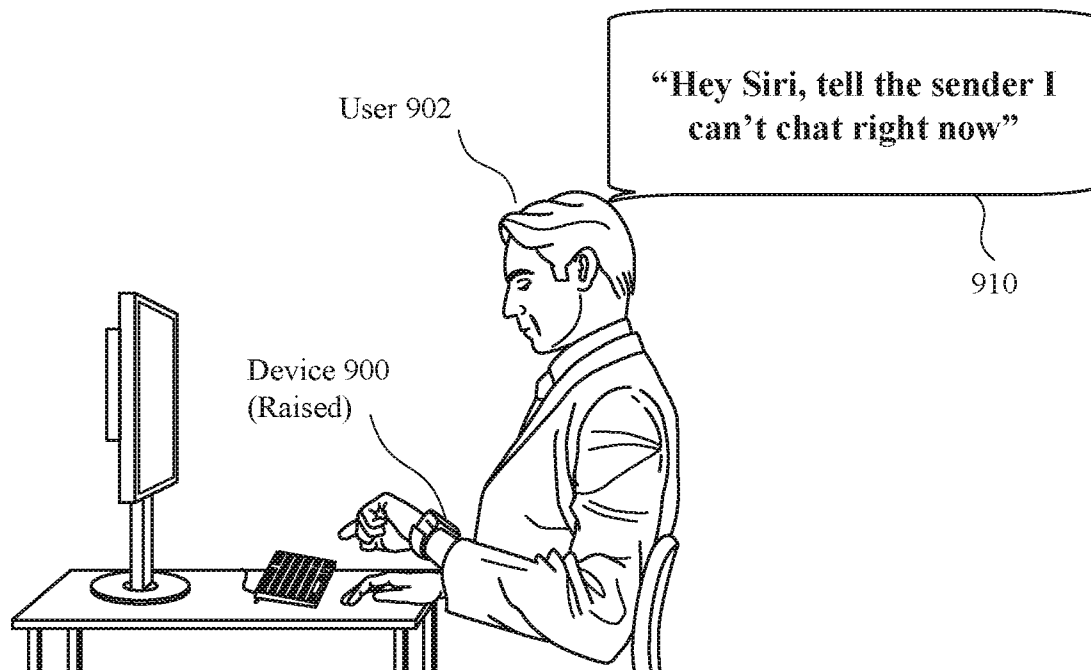
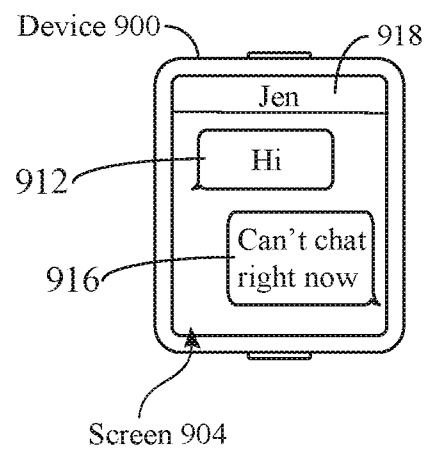
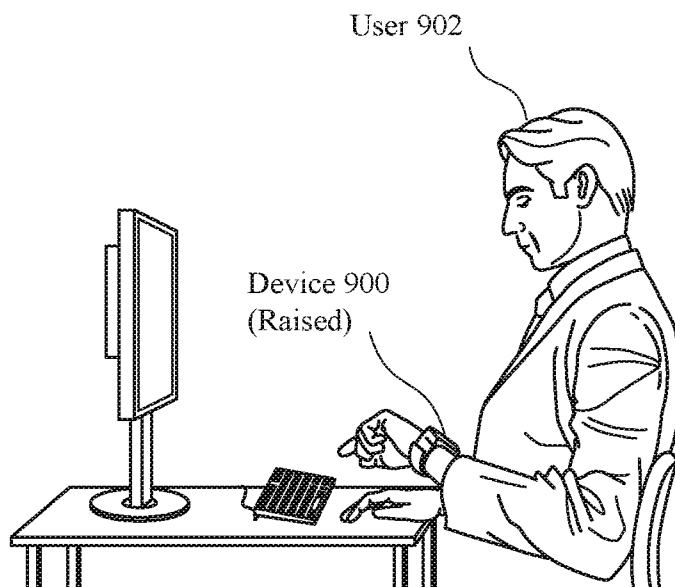


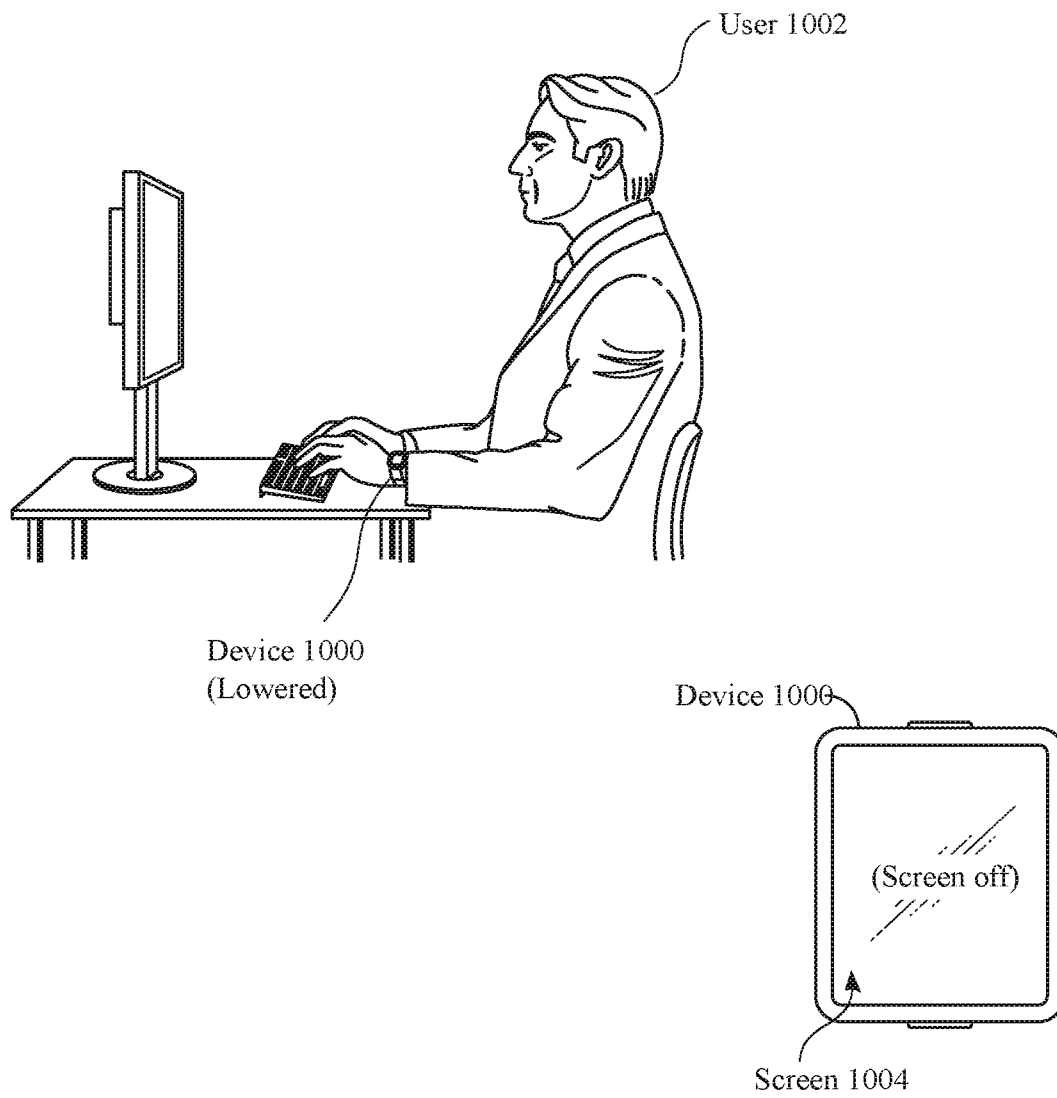
FIG. 8B

**FIG. 8C**

**FIG. 9A**

**FIG. 9B**

**FIG. 9C**

**FIG. 10A**

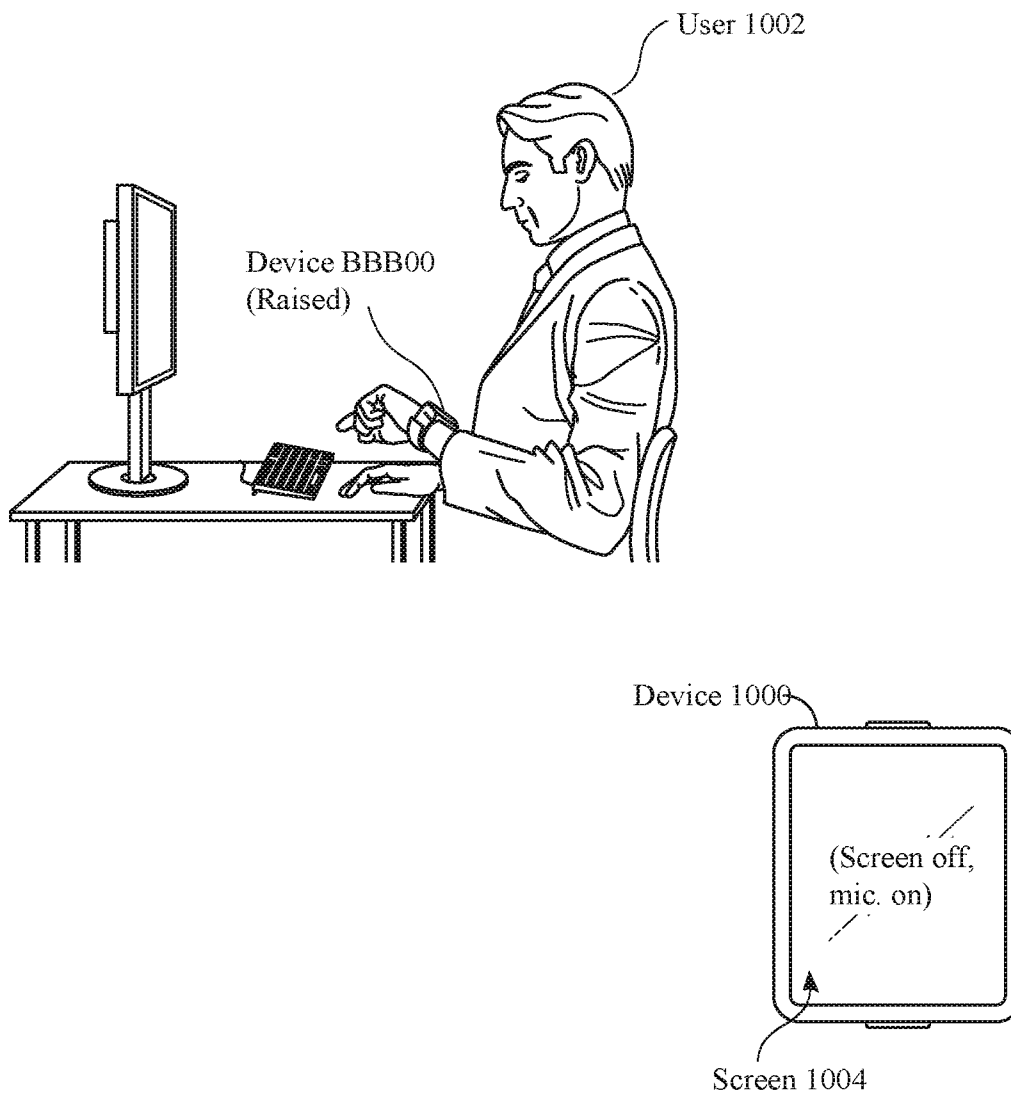
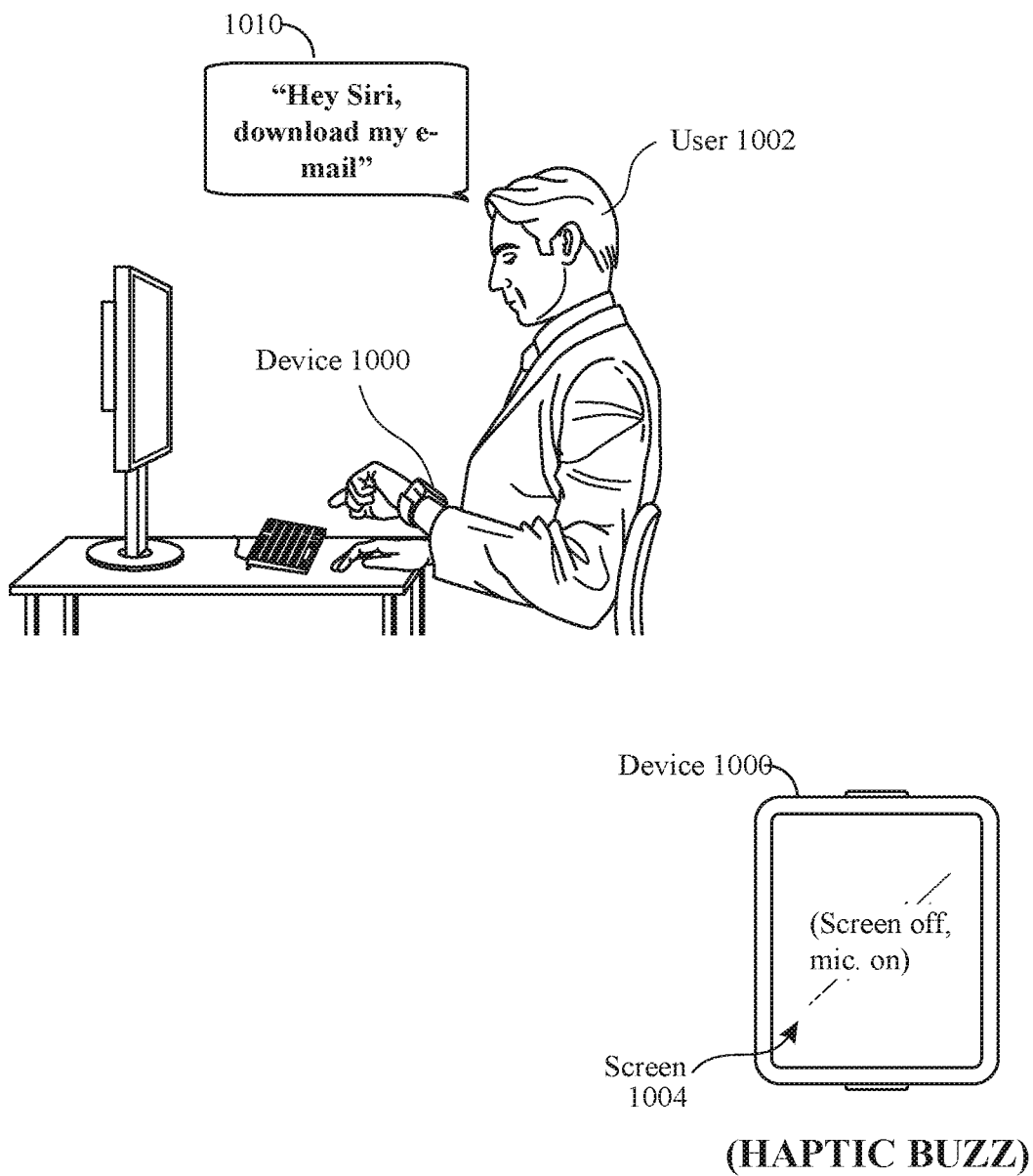


FIG. 10B

**FIG. 10C**

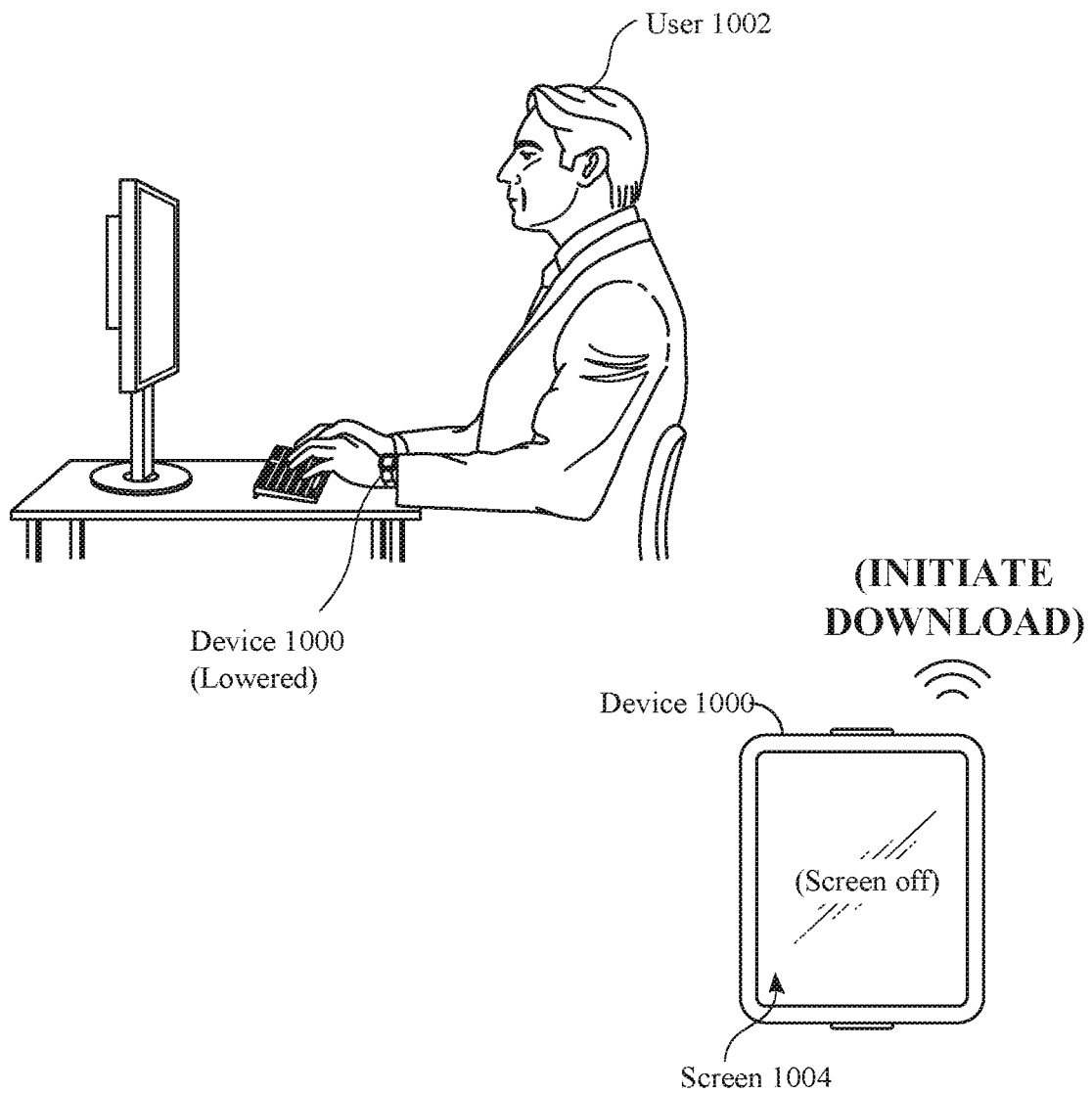
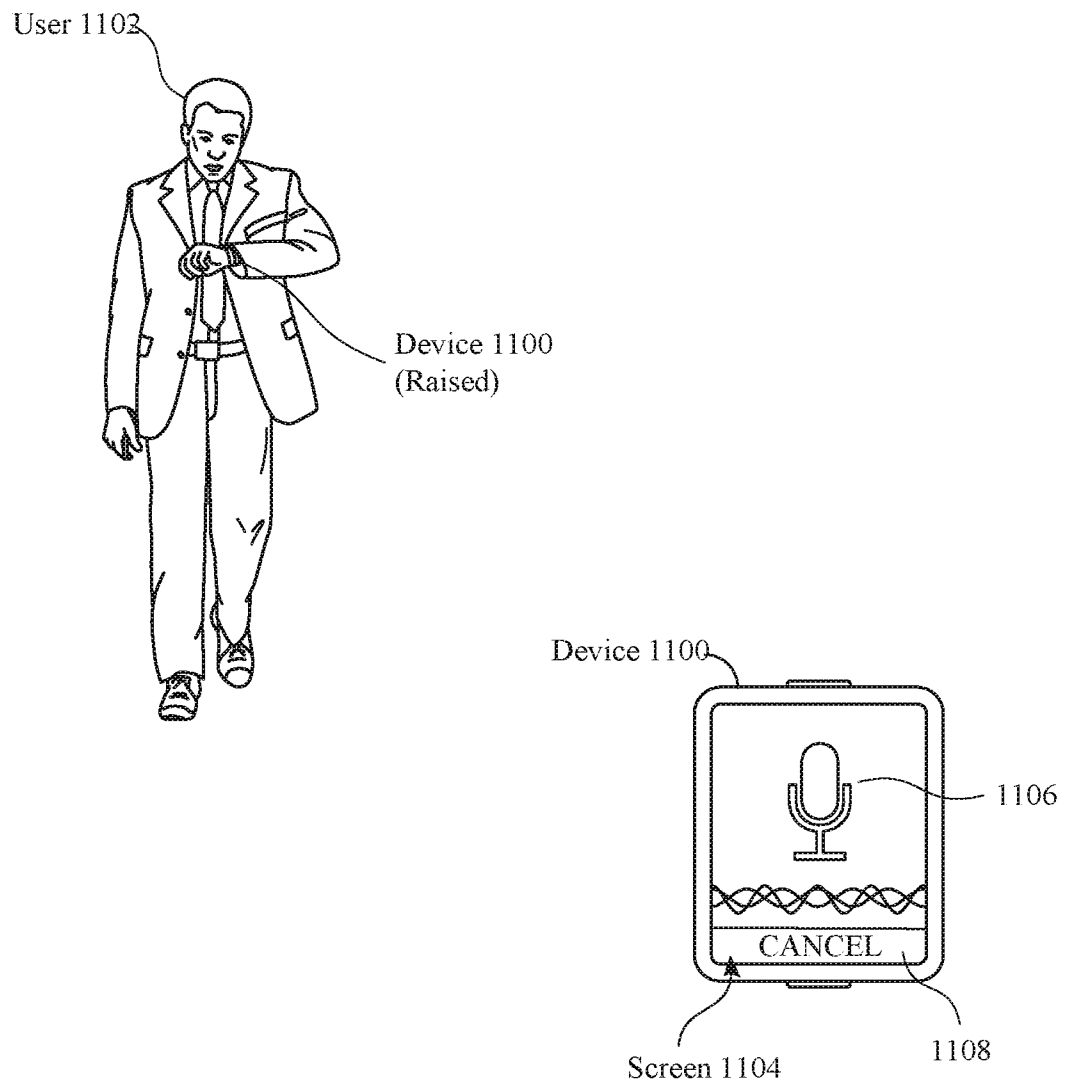
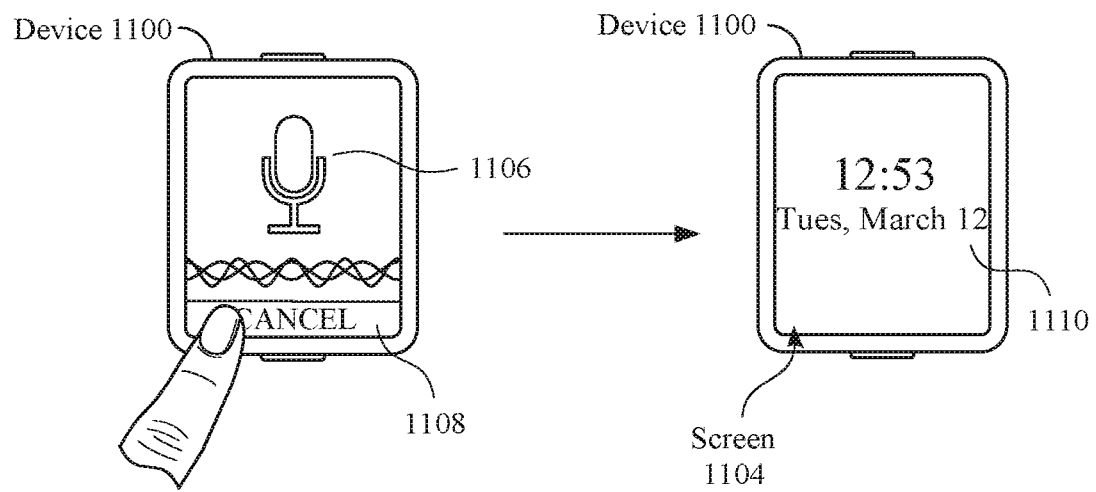


FIG. 10D

**FIG. 11A**

**FIG. 11B**

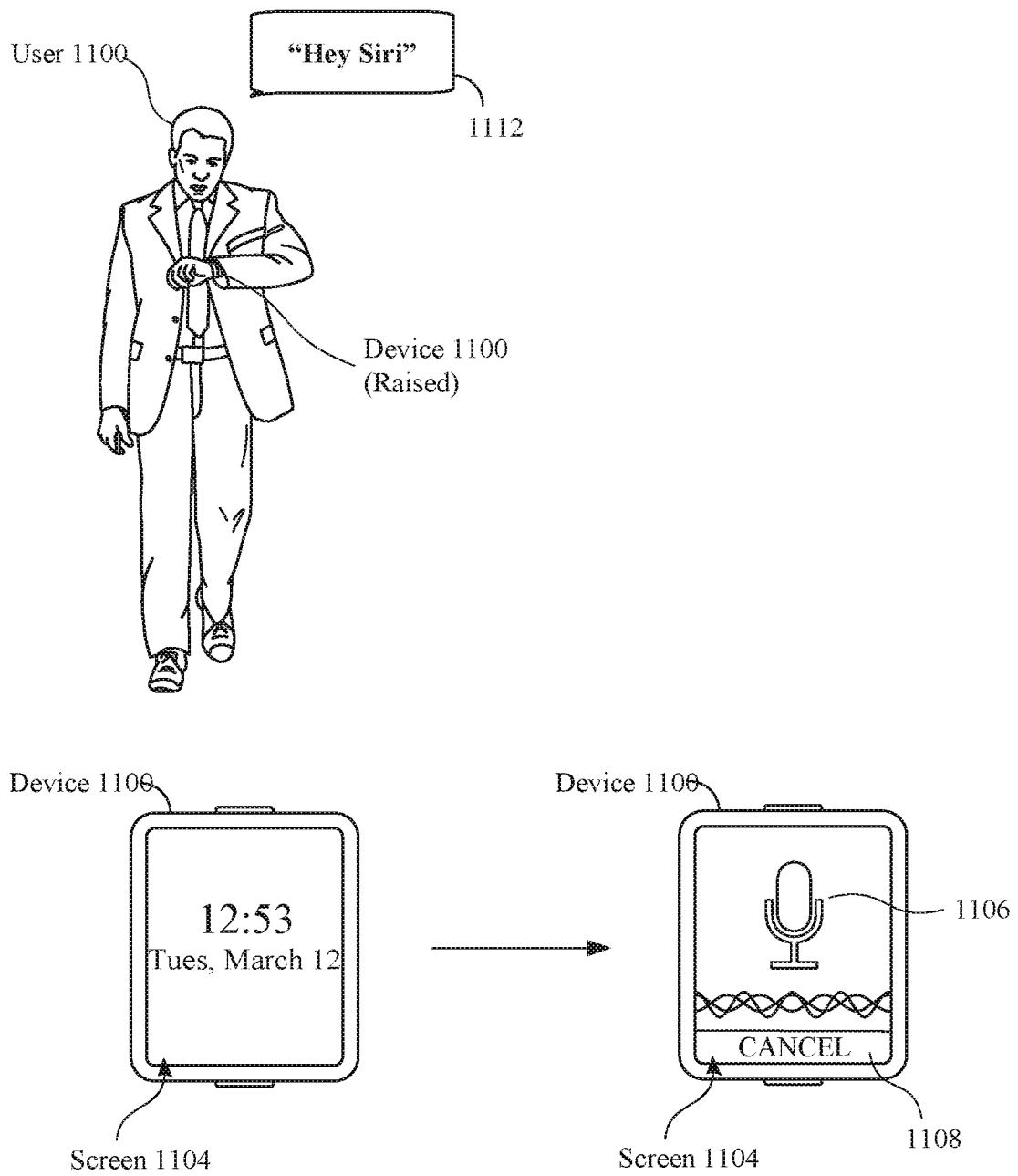
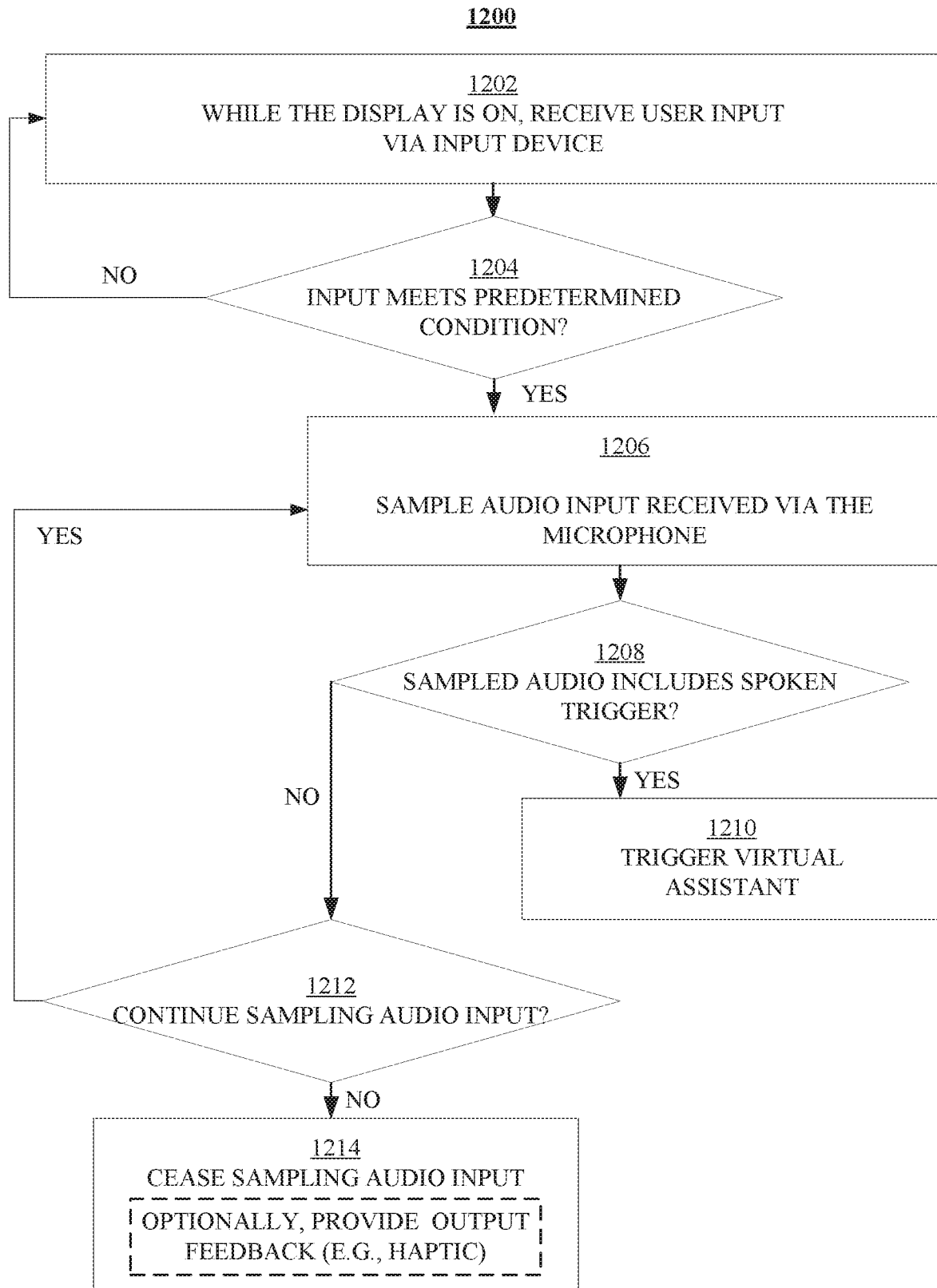
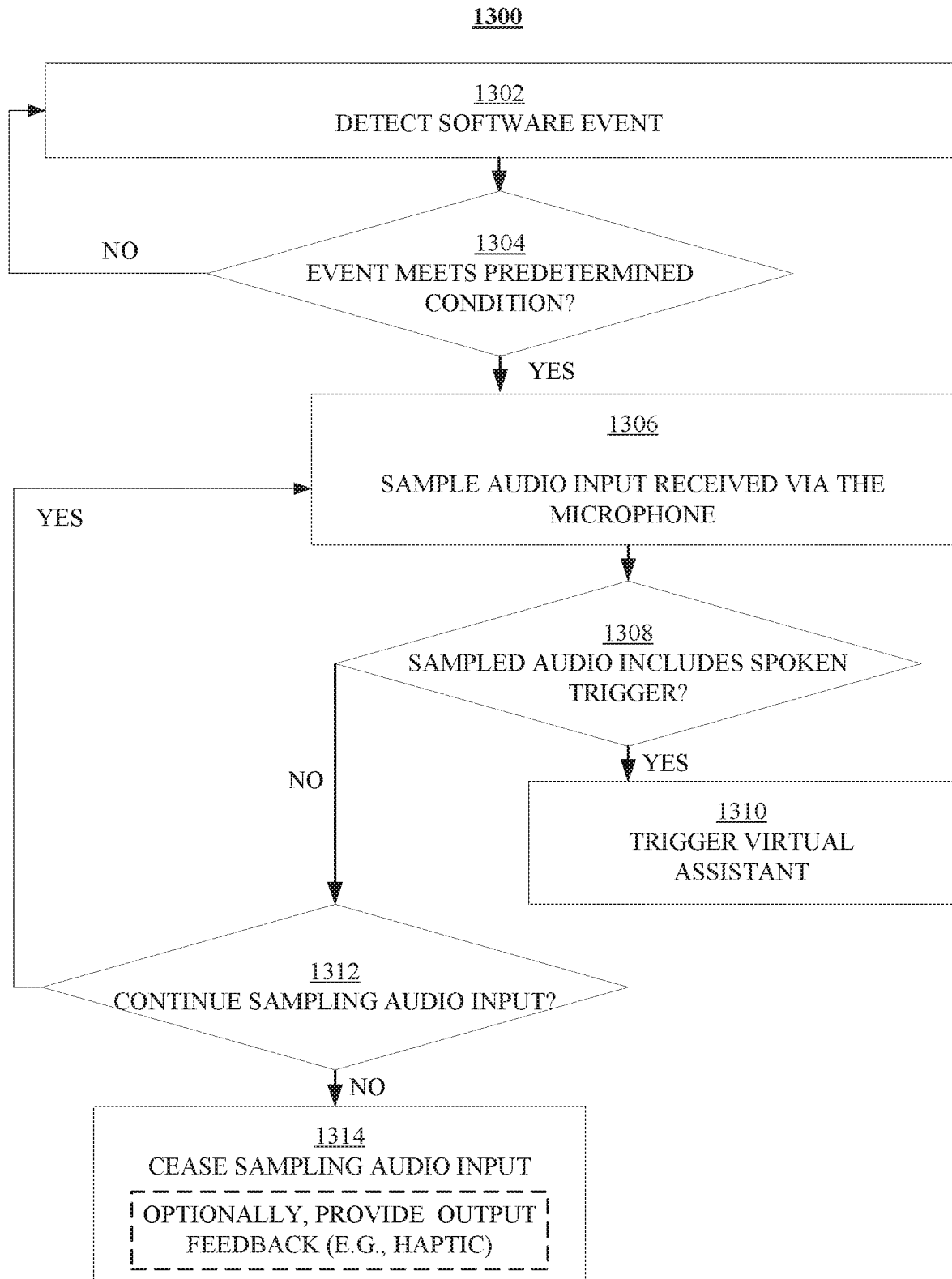
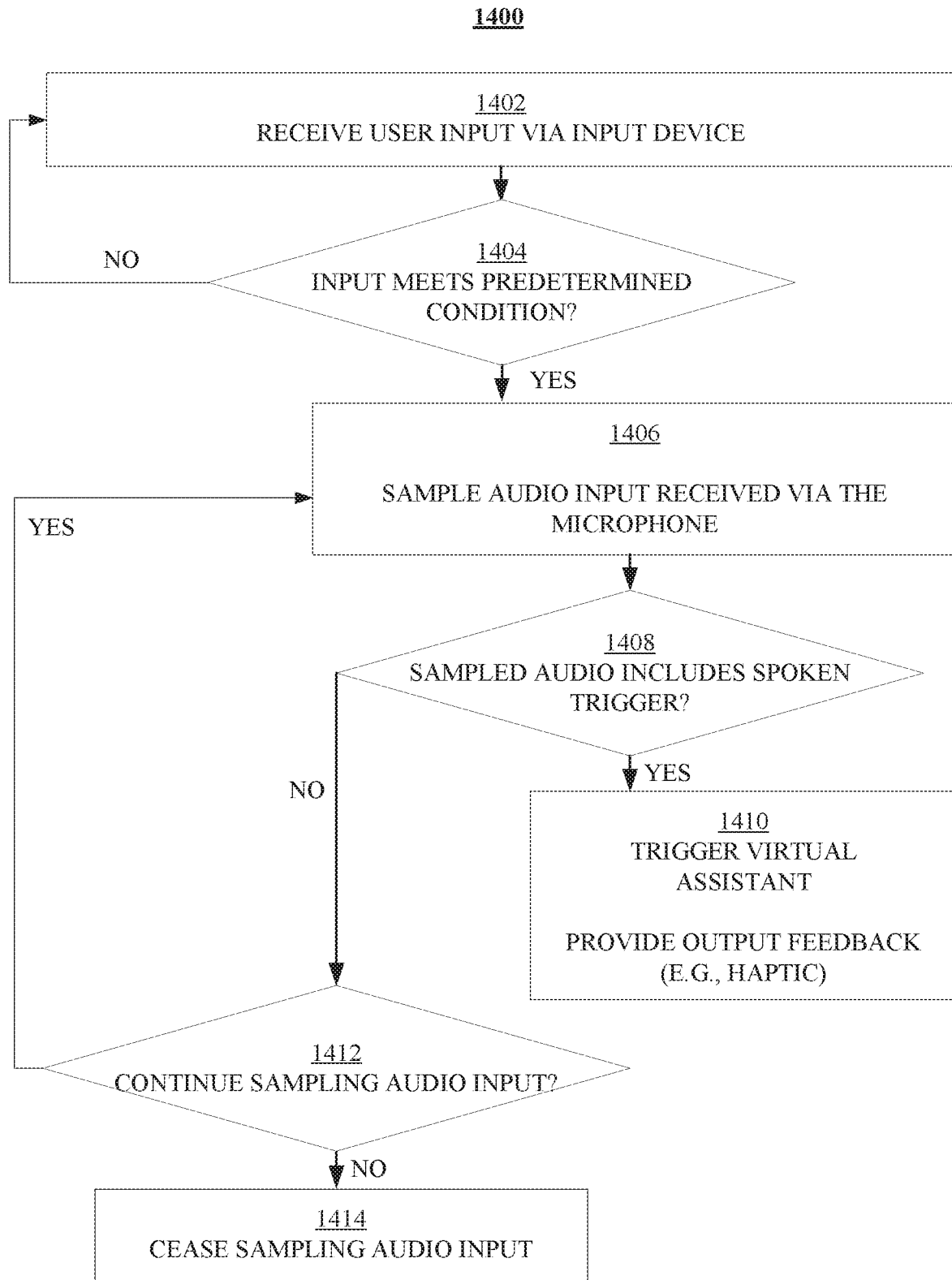


FIG. 11C



**FIG. 13**

**FIG. 14**

1500

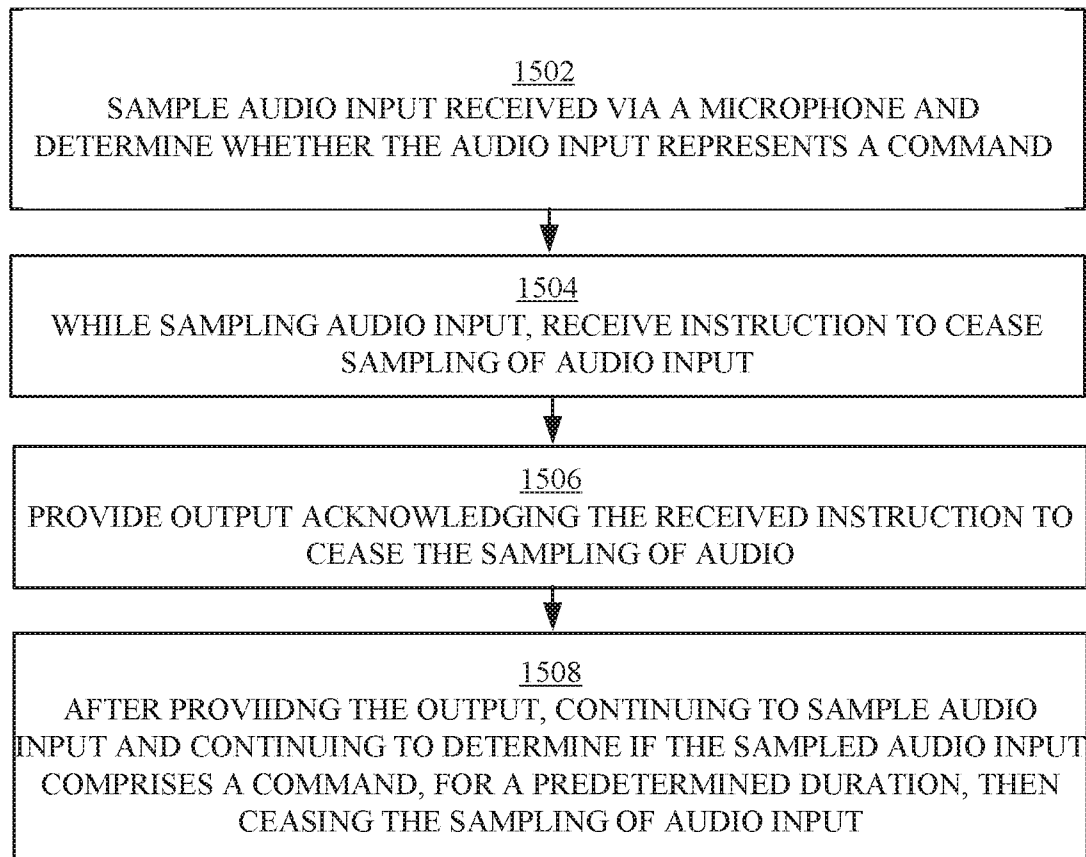


FIG. 15

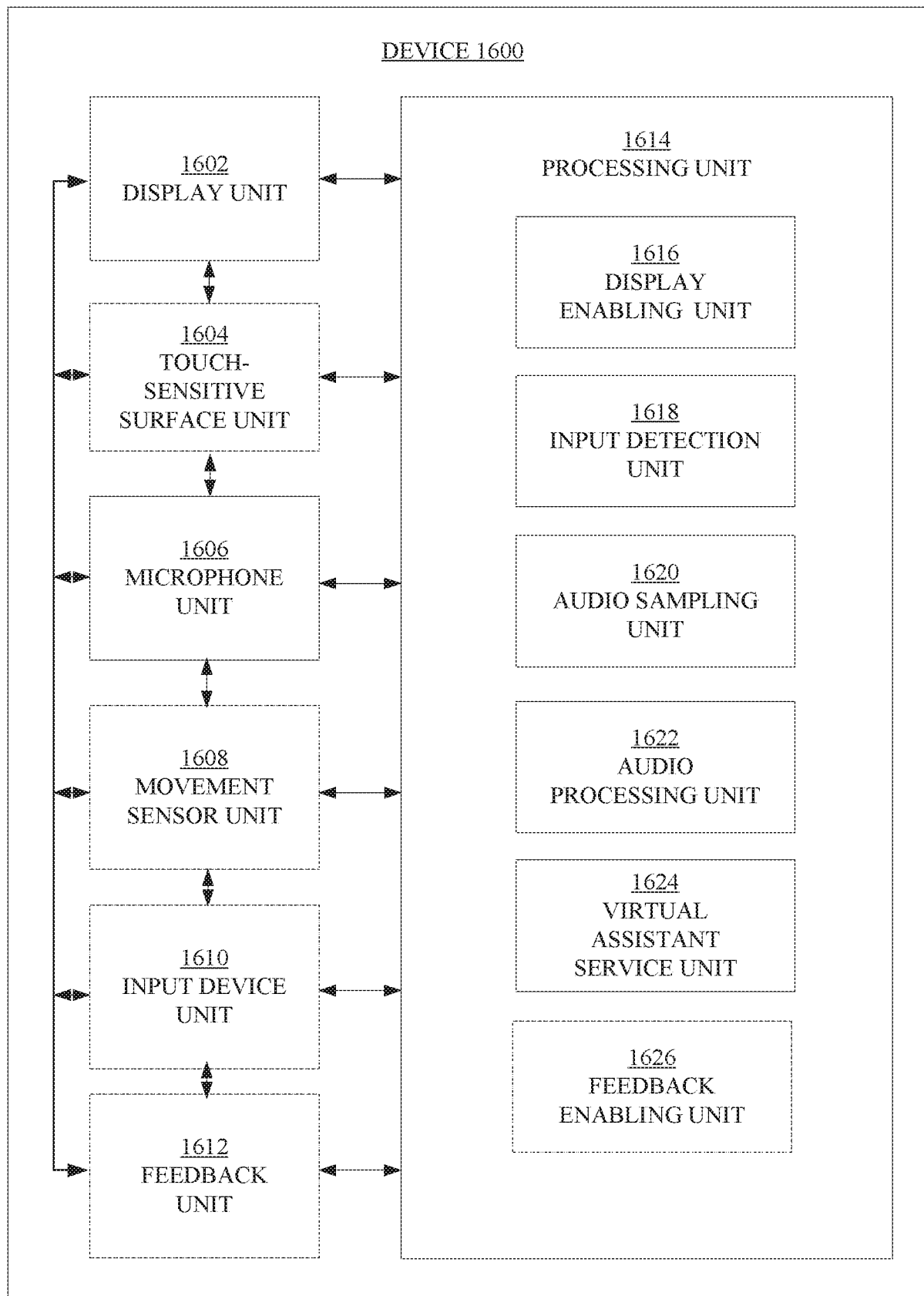


FIG. 16

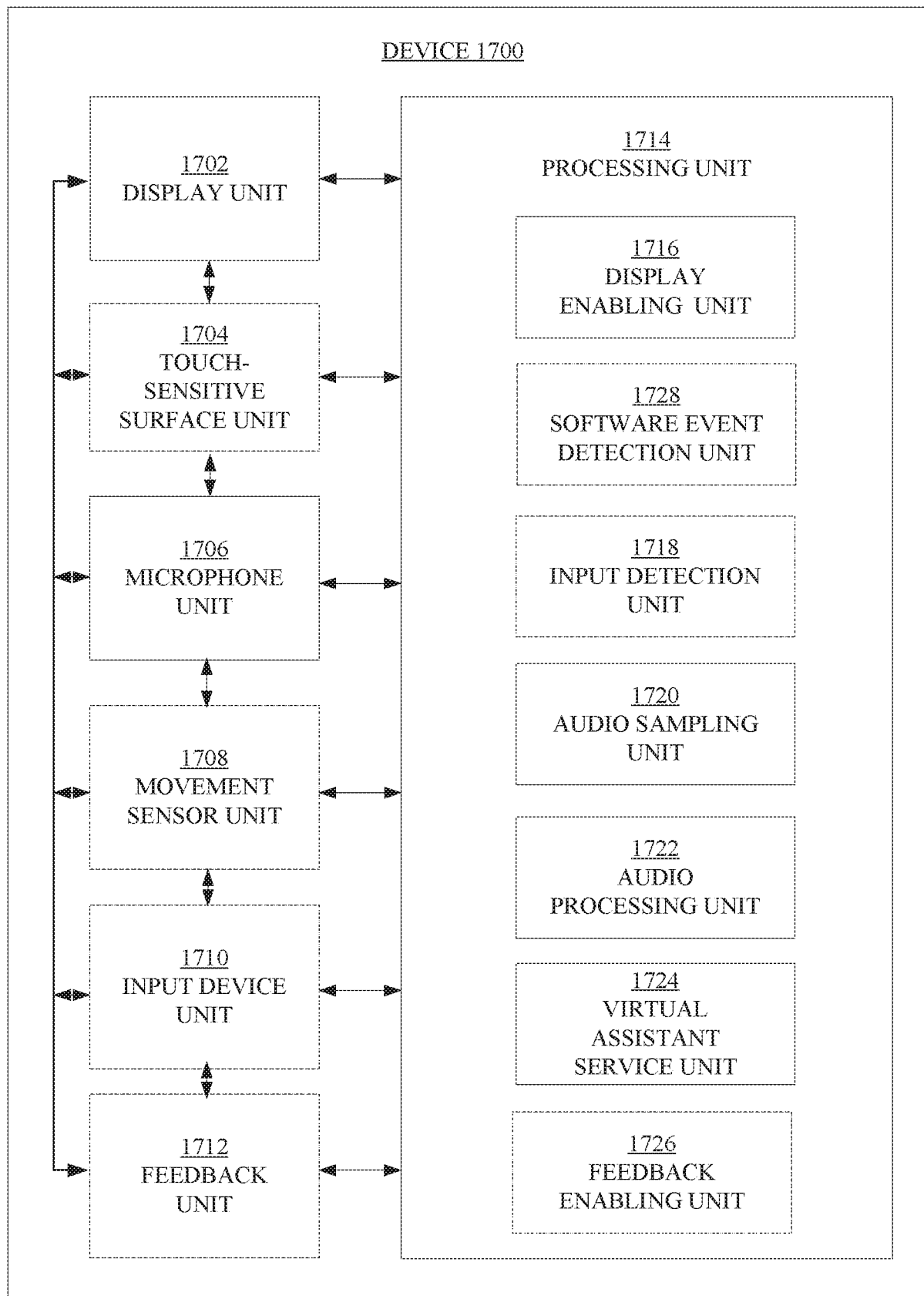


FIG. 17

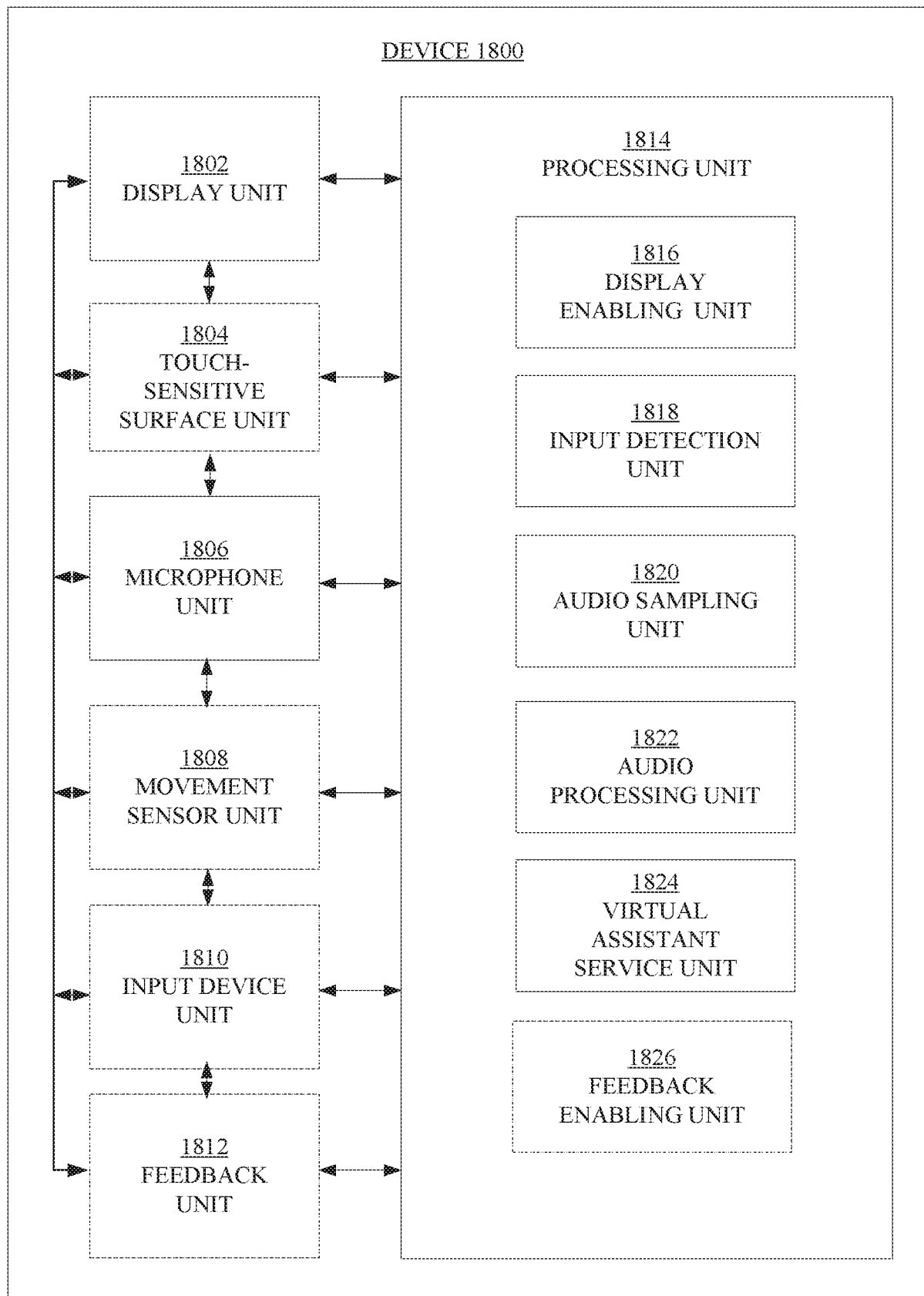


FIG. 18

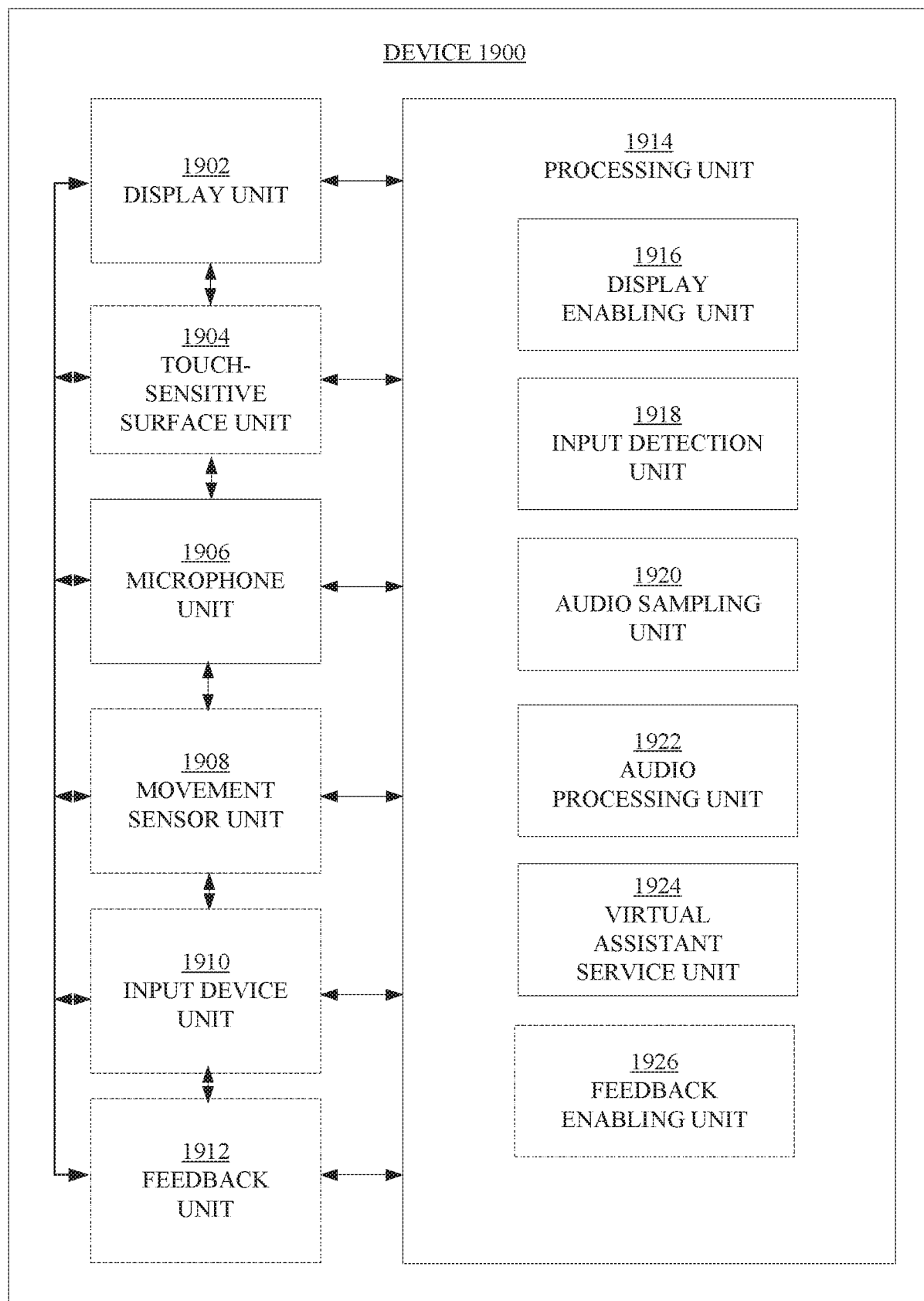
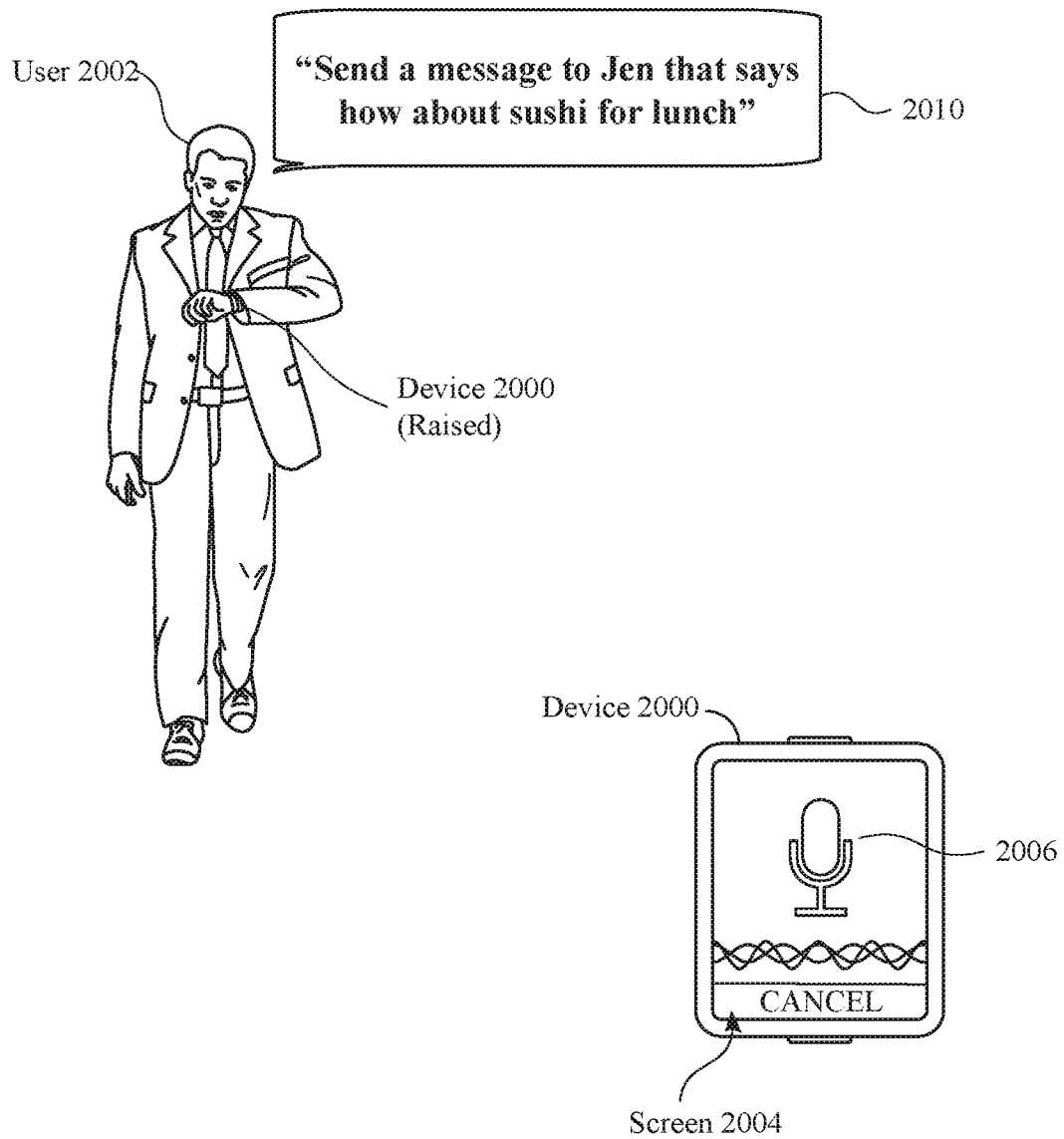
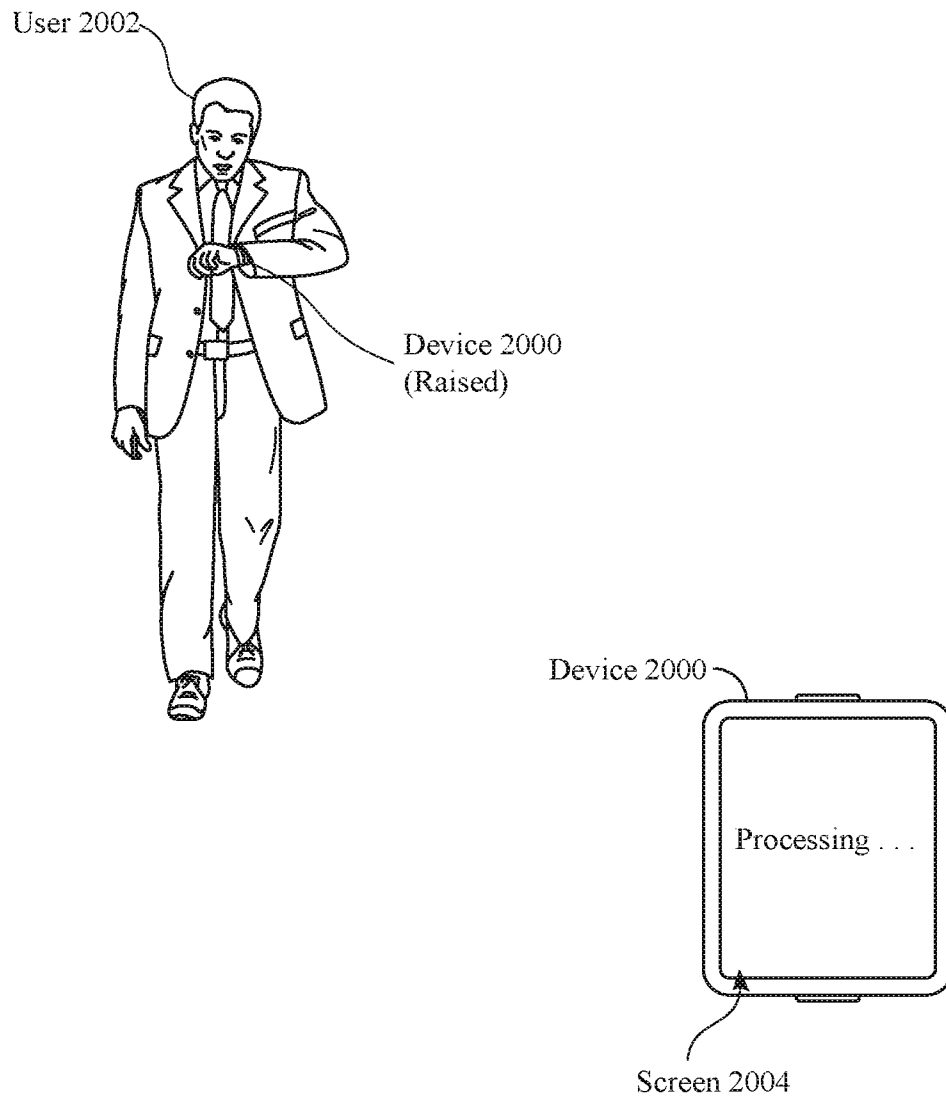


FIG. 19

**FIG. 20A**

**FIG. 20B**

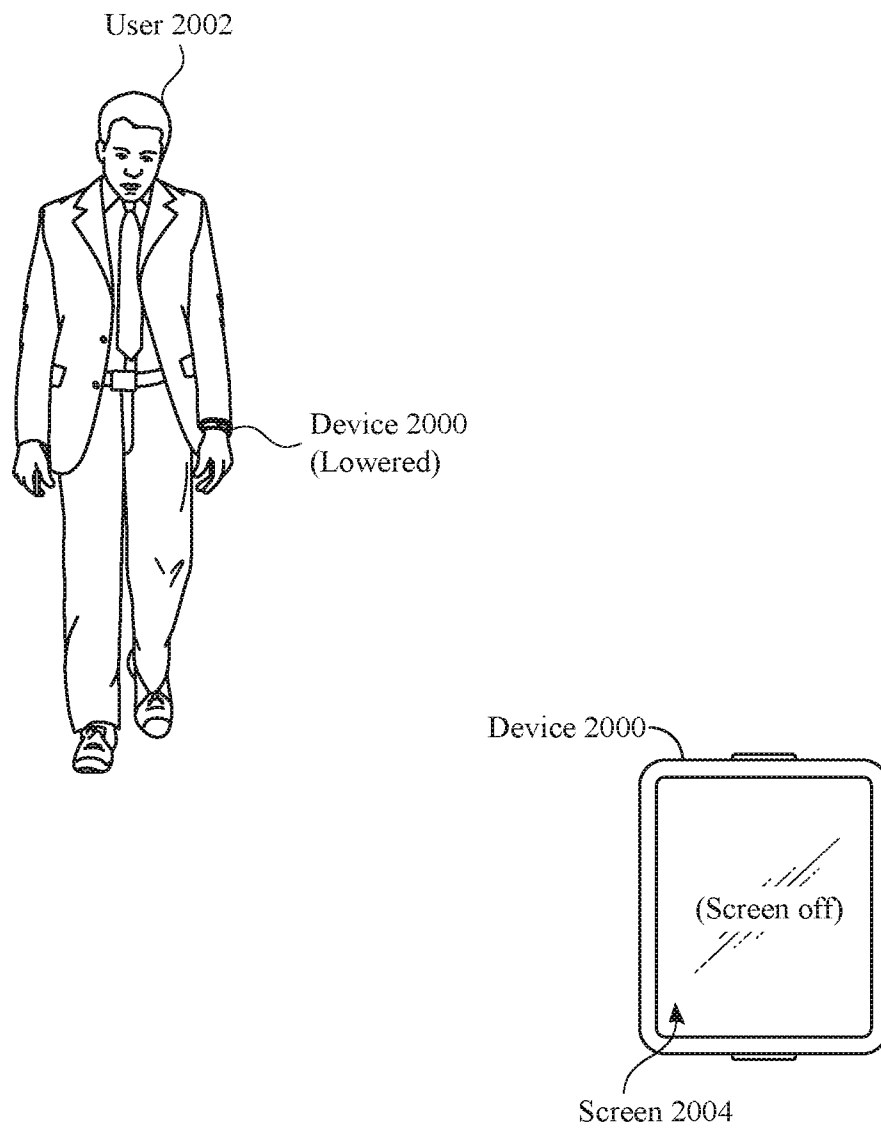


FIG. 20C

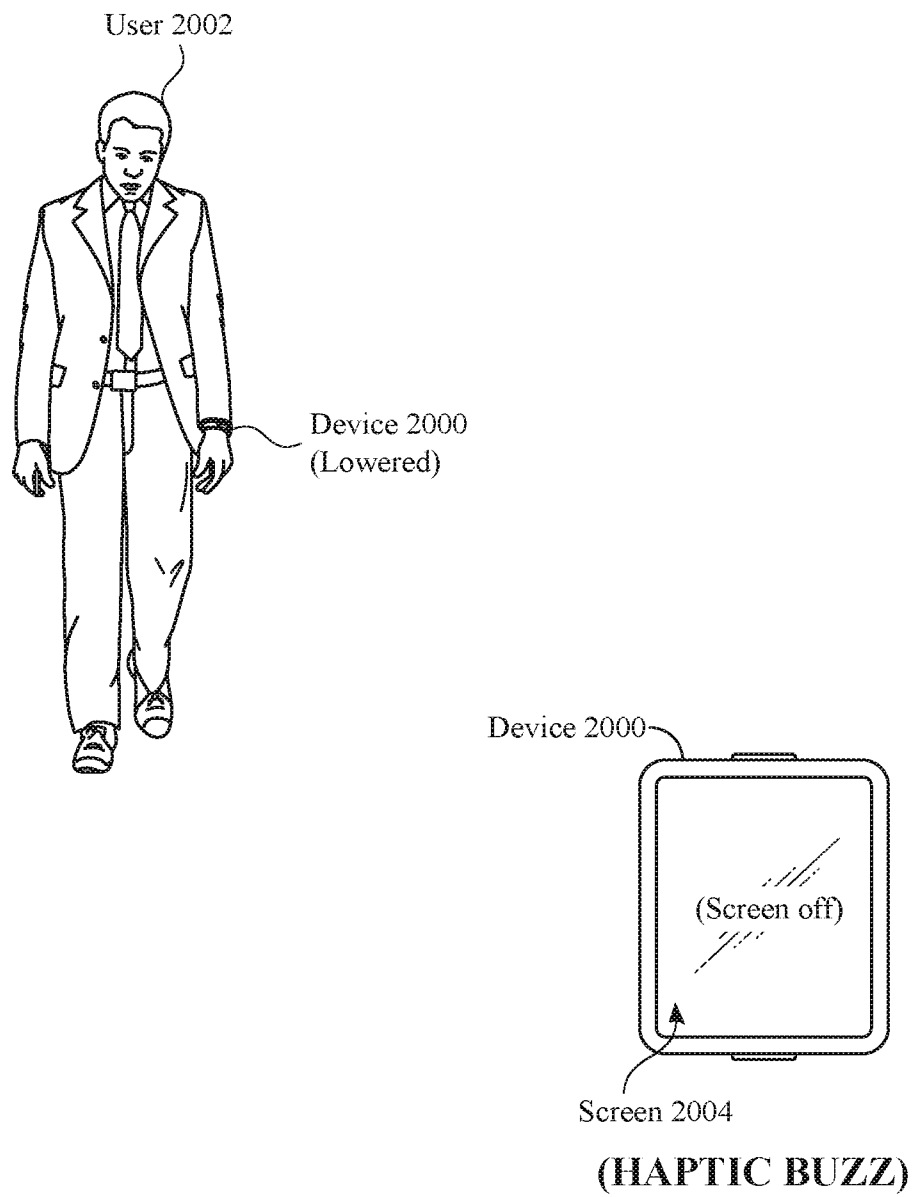
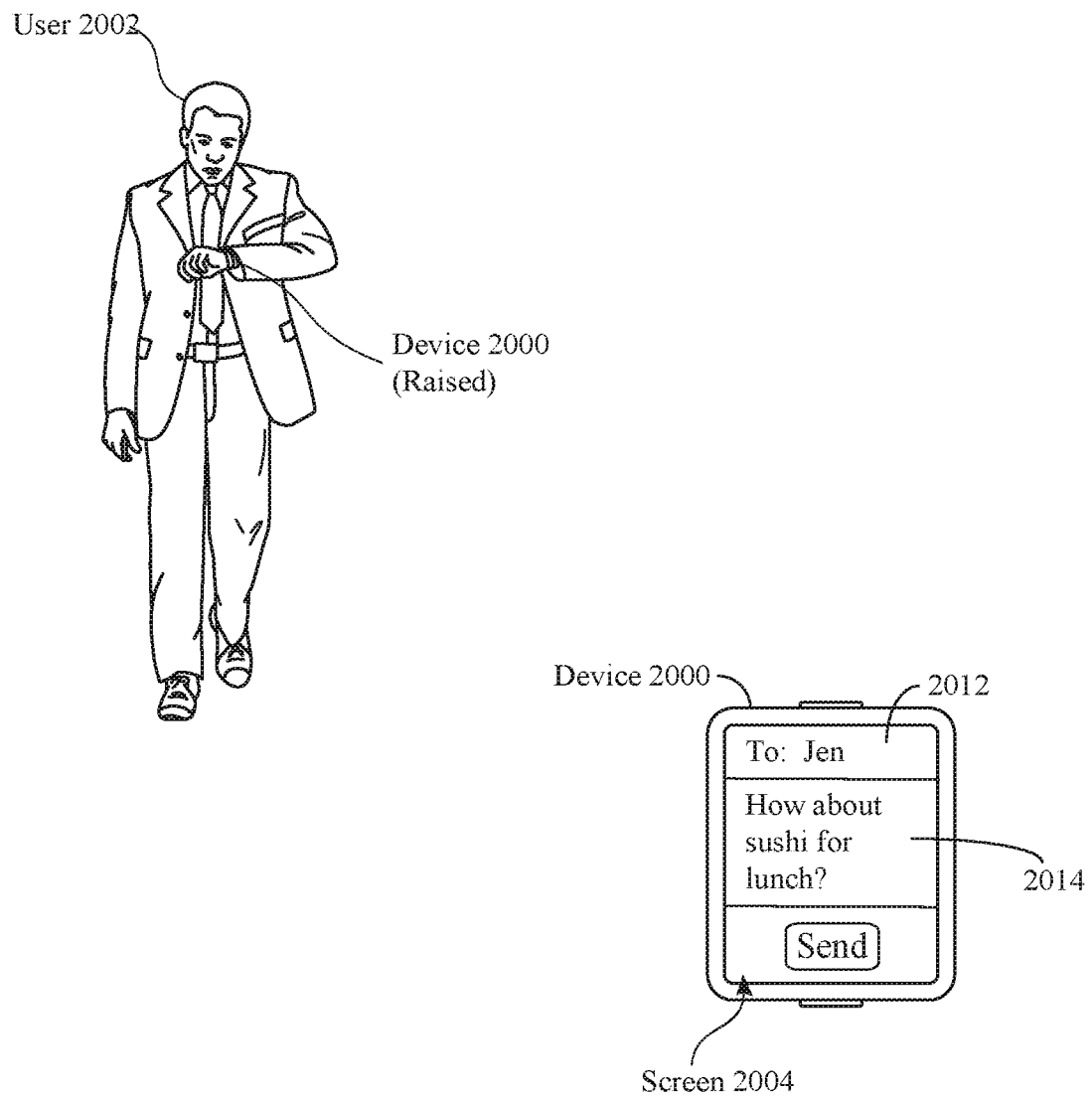
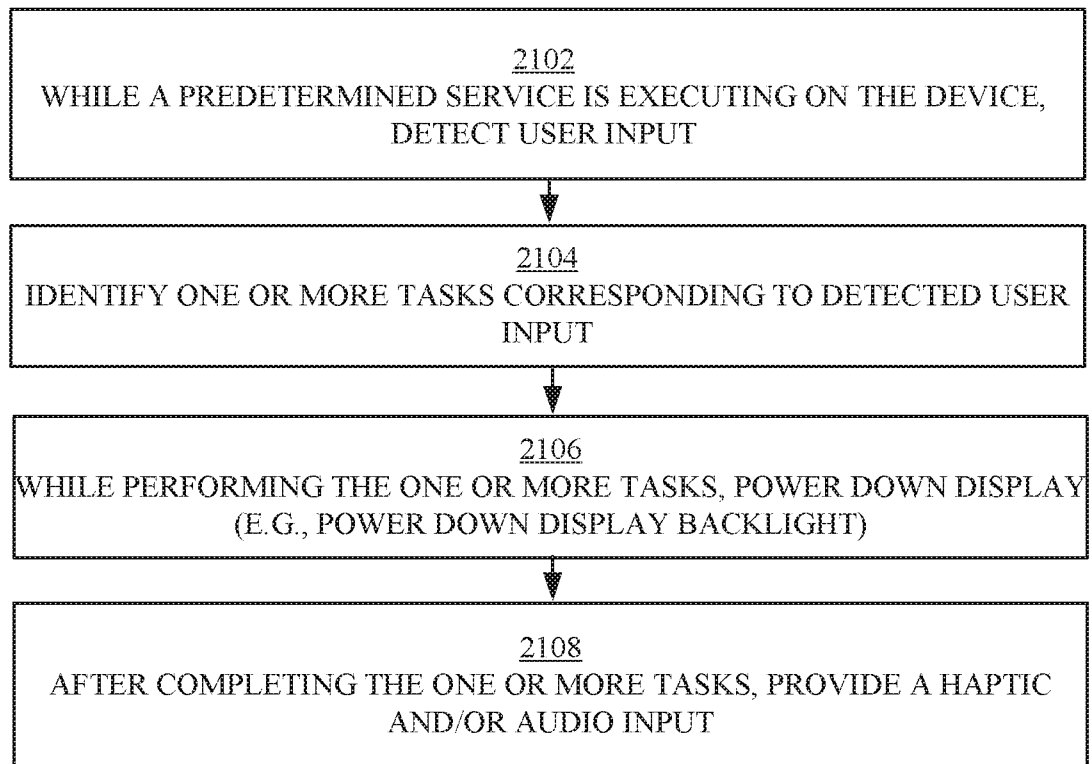


FIG. 20D

**FIG. 20E**

2100**FIG. 21**

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2016/021410

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - G06F 3/16; G06F 17/28; G10L 15/00; G10L 15/22 (2016.01)

CPC - G06F 3/167; G06F 17/28; G10L 15/1822; G10L 15/22; G10L 15/265 (2016.01)

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)

IPC - G06F 3/16; G06F 17/28; G10L 15/00; G10L 15/22

CPC - G06F 3/167; G06F 17/28; G10L 15/1822; G10L 15/22; G10L 15/265

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

USPC - 704/251; 704/275; 715/728 (keyword delimited)

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

PatBase, Google Patents, Google Scholar, Google

Search terms used: virtual assistant, trigger, word, task

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	US 2013/0275899 A1 (APPLE INC.) 17 October 2013 (17.10.2013) entire document	1, 5, 6, 22-30, 39, 40, 46-48, 54, 74, 76, 78, 82, 84, 86, 90, 92, 94, 98, 99, 102, 103, 119-127, 143-145, 148 --- 2, 3, 67, 68, 100, 101, 164, 165
X --- Y	US 2008/0201306 A1 (COOPER et al) 21 August 2008 (21.08.2008) entire document	66, 80, 88, 96, 163 --- 67, 68, 164, 165
Y	US 20100082286 A1 (LEUNG) 01 April 2010 (01.04.2010) entire document	2, 3, 100, 101
A	US 2014/0222436 A1 (APPLE INC) 07 August 2014 (07.08.2014) entire document	1-3, 5, 6, 22-30, 39, 40, 46-48, 54, 66-68, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98-103, 119-127, 143-145, 148, 163-165

☒ 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

28 June 2016

Date of mailing of the international search report

26 JUL 2016

Name and mailing address of the ISA/

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, VA 22313-1450

Facsimile No. 571-273-8300

Authorized officer

Blaine R. Copenheaver

PCT Helpdesk: 571-272-4300
PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2016/021410

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2013/0085755 A1 (BRINGERT et al) 04 April 2013 (04.04.2013) entire document	1-3, 5, 6, 22-30, 39, 40, 46-48, 54, 66-68, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98-103, 119-127, 143-145, 148, 163-165
A	US 2012/0022872 A1 (GRUBER et al) 26 January 2012 (26.01.2012) entire document	1-3, 5, 6, 22-30, 39, 40, 46-48, 54, 66-68, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98-103, 119-127, 143-145, 148, 163-165
A	US 200/80186960 A1 (KOCHEISEN et al) 07 August 2008 (07.08.2008) entire document	1-3, 5, 6, 22-30, 39, 40, 46-48, 54, 66-68, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98-103, 119-127, 143-145, 148, 163-165

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2016/021410

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☒ Claims Nos.: see extra sheet
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This international Searching Authority found multiple inventions in this international application, as follows:
See supplemental page

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☒ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2016/021410

Continued from Box No. III Observations where unity of invention is lacking

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I, claims 1-3,5-6,8,22-30,39-40,46-48,54,74,76,78,82,84,86,90,92,94,98-103,119-127,143-145,148, drawn to a method comprising determining whether the audio input comprises a spoken trigger.

Group II, claims 66-68,80,88,96,163-165, drawn to a method comprising determining whether the audio input represents a task.

The inventions listed as Groups I-II do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: the special technical feature of the Group I invention: determining whether the audio input comprises a spoken trigger; and in accordance with a determination that audio input comprises the spoken trigger, triggering a virtual assistant session as claimed therein is not present in the invention of Group II. The special technical feature of the Group II invention: determining whether the audio input represents a task; determining whether the audio input comprises a task, for a duration, then ceasing the sampling of audio input as claimed therein is not present in the invention of Group I.

Groups I and II lack unity of invention because even though the inventions of these groups require the technical feature of an electronic device with a display and a microphone: sampling audio input received via the microphone, this technical feature is not a special technical feature as it does not make a contribution over the prior art.

Specifically, US 2008/0186960 A1 (KOCHEISEN et al) 07 August 2008 (07.08.2008) teaches an electronic device with a display and a microphone (Para. 20): sampling audio input received via the microphone (Paras. 28 and 30).

Since none of the special technical features of the Group I or II inventions are found in more than one of the inventions, unity of invention is lacking.

Box II continuation :

Claims 4,7,9-21,31-38,41-45,49-53,55-65,69-73,75,77,79,81,83,85,87,89,91,93,95,97,104-118,128-142,146-147,149-162,166-170 are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).